cenicaña

Biotecnología para el mejoramiento de la caña de azúcar

Agroindustria de la caña de azúcar en Colombia

Trujillo Montenegro, Jhon Henry

Biotecnología para el mejoramiento de la caña de azúcar / Jhon Henry Trujillo Montenegro; María Camila Martínez Villa; John Jaime Riascos Arcos. Centro de Investigación de la Caña de Azúcar de Colombia (Ed.) -- Cali: Centro de Investigación de la Caña de Azúcar de Colombia, 2024.

26 p. (Agroindustria de la caña de azúcar en Colombia)

Incluye referencias bibliográficas

ISBN 978-958-8449-38-8

1. Caña de azúcar. 2. Marcadores moleculares. 3. Mejoramiento genético. 4. Genoma. 5. Biotecnología. 6. CC 01-1940.

I. Martínez Villa, María Camila. II. Riascos Arcos, John Jaime. III. Título. IV. Agroindustria de la caña de azúcar en Colombia

633.61 CDD 23 ed.
T866

Cenicaña – Biblioteca Guillermo Ramos Núñez

Centro de Investigación de la Caña de Azúcar de Colombia
Calle 38 norte No. 3CN-75. Cali, Valle del Cauca, Colombia
Estación experimental: San Antonio de los Caballeros, vía Cali-Florida km 26
www.cenicana.org

Producción editorial: Servicio de Cooperación Técnica y Transferencia de Tecnología
Coordinación editorial: Victoria Carrillo C.
Corrección de textos: Ernesto Fernández R.
Diseño e ilustración: Alcira Arias V.

Cita bibliográfica

Trujillo Montenegro, J. H., Martínez Villa, M. C. & Riascos Arcos, J. J. (2024). Biotecnología para el mejoramiento de la caña de azúcar. En: Centro de Investigación de la Caña de Azúcar de Colombia (Ed). Agroindustria de la caña de azúcar en Colombia. Cenicaña. https://www.cenicana.org/coleccion-agroindustria-de-la-cana-de-azucar/

BIOLOGÍA

Biotecnología para el mejoramiento de la caña de azúcar

Jhon Henry Trujillo Montenegro

María Camila Martínez Villa

John Jaime Riascos Arcos

Contenido

Introducción 4

Complejidad genética de la caña de azúcar 6

Secuenciación de alto rendimiento, marcadores moleculares y mejoramiento genético 8

El genoma de la caña de azúcar 13

Referencias 22

Introducción

En Colombia, al igual que en todos los programas de mejoramiento genético de la caña de azúcar alrededor del mundo, este proceso se realiza de manera convencional o clásica, mediante la selección de caracteres fenotípicos de interés agronómico para el cultivo. Si bien es cierto esta estrategia ha producido resultados invaluables para la industria de la caña de azúcar de Colombia —más del 90% del área cultivada para la producción azucarera y alcoholera en el valle del río Cauca está sembrada con variedades Cenicaña Colombia, CC— el ciclo productivo del cultivo (13 meses) hace que este proceso de mejoramiento pueda tomar entre 10 y 12 años. Adicionalmente, problemas como la escasa floración en las condiciones ambientales del valle del río Cauca dificultan realizar los cruzamientos. En tal sentido, con el objetivo de agilizar el proceso de mejoramiento genético de la caña de azúcar y ampliar el rango de las características favorables seleccionadas, Cenicaña, al igual que sus pares en otros países, ha incorporado la biotecnología en el esquema de mejoramiento y selección de variedades CC.

Entre las herramientas biotecnológicas claves para lograr mayor eficiencia en el proceso de mejoramiento genético clásico se destaca la utilización de marcadores moleculares para la genotipificación de germoplasma y la transformación genética. Los marcadores moleculares son regiones del genoma (secuencias de ADN) que a menudo se utilizan para rastrear la herencia (segregación) de una posición en particular. A la fecha hay suficiente evidencia que demuestra que el uso de estos marcadores ayuda a acelerar los procesos de mejoramiento genético (Foiada et al., 2015; Jiang et al., 2012; Steele et al., 2006). En el caso de la caña de azúcar, cuyo genoma es altamente complejo, los adelantos recientes en las tecnologías de secuenciación de ADN (secuenciación de segunda y tercera generación), que favorecen la identificación de marcadores SNP (*Single Nucleotide Polymorphism,* polimorfismo de un solo nucleótido), los cuales son abundantes en el genoma de la caña de azúcar (Davey et al., 2011), sugieren altas probabilidades de éxito. Adicionalmente, el desarrollo de metodologías eficientes de transformación genética hace posible la introducción de caracteres individuales con capacidad de impactar positivamente el desarrollo agronómico del cultivo, lo que además permite rebasar el obstáculo que representa la escasa floración de variedades de caña mejoradas.

Complejidad genética de la caña de azúcar

El genoma de la caña de azúcar se caracteriza por ser altamente complejo. Los genotipos comerciales sembrados hoy provienen de un proceso de domesticación reciente que involucra principalmente las especies *Saccharum officinarum* y *S. spontaneum,* y algunas contribuciones de especies como *S. sinense*, *S. barberi* y *S. robustum*. Las especies *S. officinarum* y *S. spontaneum* poseen genomas poliploides, aunque su constitución cromosómica es diferente. Los clones de *S. spontaneum* son octoploides (x = 8) y poseen un número de cromosomas (constitución $2n$) que varía entre 40 y 128; mientras que los clones de *S. officinarum* son decaploides (x = 10), con una constitución menos variable: 80 cromosomas (Price & Daniels, 1968; Sreenivasan et al., 1987). Los híbridos actuales son poliploides, con una distribución no uniforme de cromosomas en un mismo grupo (aneuploidía), además de una constitución cromosómica muy variable (x = 10-13, $2n$ = 100-130). El tamaño del genoma de los híbridos comerciales de caña es variable y se estima cercano a los 10 Gbp (Moore et al., 2014).

Además de su origen interespecífico, la complejidad genética de la caña de azúcar está ligada a su historia de domesticación reciente. El mejoramiento genético data de 1888 y fue motivado por hallazgos previos que demostraban que esta planta era capaz de producir semilla viable (revisado en Moore et al., 2014). Esto, sumado a un ciclo de cultivo relativamente largo y, consecuentemente, a una menor tasa de ciclos de selección, ha contribuido a que la mayoría de las aneuploidías resultantes de los cruzamientos interespecíficos *S. officinarum* x *S. spontaneum* se mantengan en los híbridos modernos.

La secuenciación y el ensamblaje del primer genoma vegetal se dieron en la planta modelo *Arabidopsis thaliana* (AGI, 2000). Comparativamente, el genoma de *A. thaliana* es de los más simples, con una constitución diploide total de 10 cromosomas $2n = 2x = 10$ y un tamaño de 135 Mbp. Debido a que los genomas de las especies cultivadas, tales como la caña de azúcar, son en general poliploides y de mayor tamaño, lo que consecuentemente implica una mayor cantidad de información repetida, se dificulta el ensamblaje. No obstante, gracias a los avances en las tecnologías de secuenciación de alto rendimiento (*High Throughput Sequencing*, HTS) hoy es posible llevar a cabo la secuenciación de genomas complejos. En 2009, Schnable y colaboradores reportaron la finalización de la secuenciación del genoma del maíz, y Peterson y su equipo, la del sorgo. Sin embargo, el genoma del maíz y el del sorgo no conllevan tantas aneuploidías como el genoma de la caña de azúcar. Debido a esta complejidad genética, secuenciar el genoma de la caña de azúcar ha sido un desafío mayor (**Figura 1**).

El genoma del maíz y el del sorgo no conllevan tantas aneuploidías como el genoma de la caña de azúcar. Debido a esta complejidad genética, secuenciar el genoma de la caña de azúcar ha sido un desafío mayor.

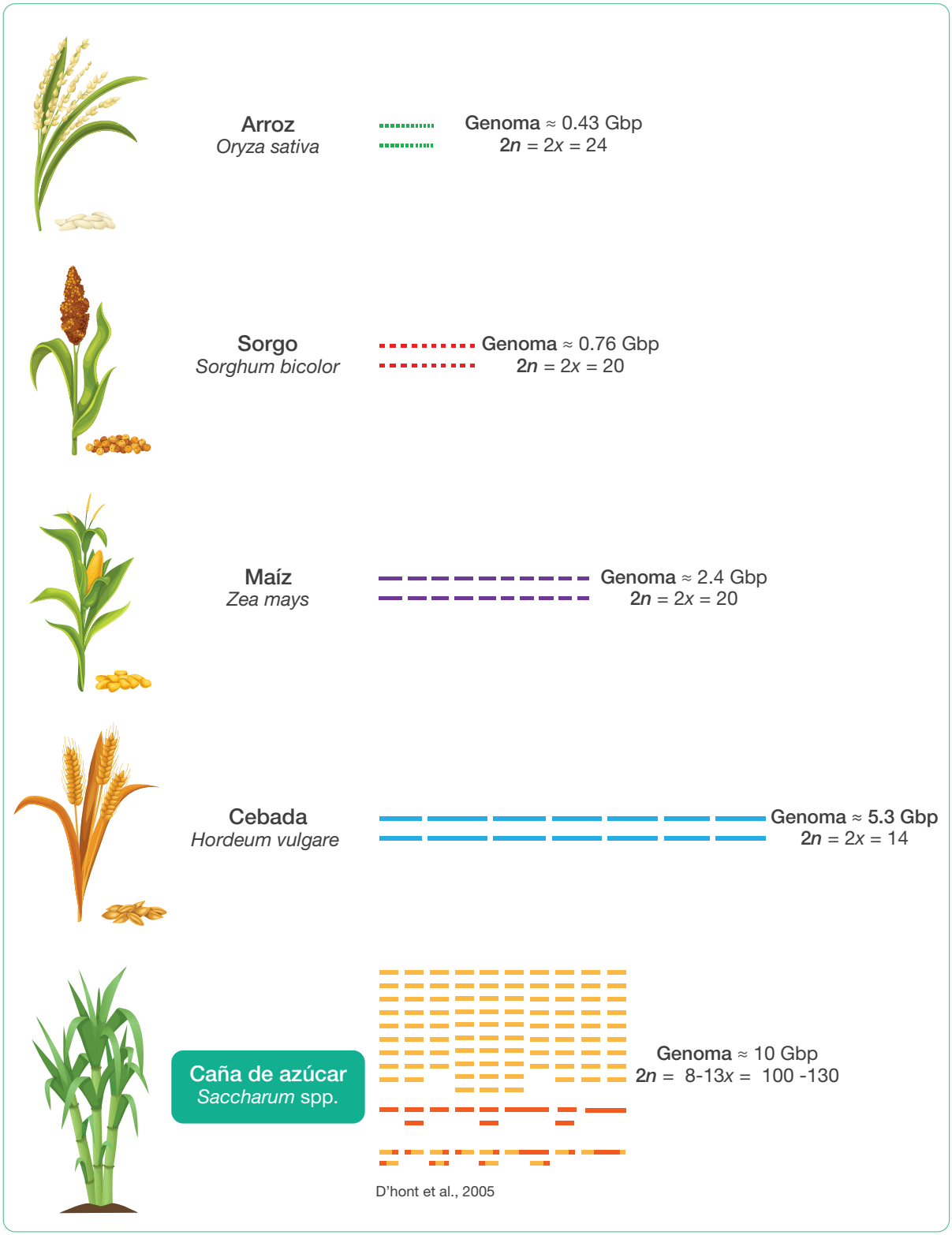


Figura 1. Representación gráfica del genoma de la caña de azúcar y otras especies relacionadas. Las estimaciones del tamaño del genoma de cada especie están dadas según el contenido monoploide.

Secuenciación de alto rendimiento, marcadores moleculares y mejoramiento genético

Secuenciación del ADN y nuevas tecnologías

La secuenciación del ADN consiste en identificar el orden de los nucleótidos que componen el material genético provisto. Establecer el orden de estos nucleótidos en diferentes tipos de organismos vivos, como las plantas, permite conocer con mayor profundidad el genotipo de estos individuos y la información genética que se transporta en un segmento de ADN en particular. Existen diferentes tipos de tecnologías de secuenciación, y entre ellas las tradicionales se conocen como tecnologías de secuenciación de primera generación, que incluyen la secuenciación química propuesta por Maxam y Gilbert (1977), la cual utiliza métodos químicos para fragmentar secuencias de ADN ya marcadas radioactivamente; y el método de secuenciación enzimática de Sanger et al. (1977), que se basa en el uso de la enzima ADN polimerasa para sintetizar cadenas de ADN con una terminación específica. Las tecnologías de secuenciación de alto rendimiento (HTS), también conocidas como de próxima generación (*Next Generation Sequencing*, NGS), se basan en el concepto de secuenciación en paralelo, es decir, permiten descifrar de manera precisa el orden exacto de los nucleótidos de millones de moléculas de ADN al mismo tiempo. Esta capacidad hace posible producir grandes volúmenes de información a un menor costo en comparación con la tecnología Sanger, conocida como secuenciación clásica (Metzker, 2010).

Actualmente existen distintas plataformas comerciales de NGS, las más comunes de las cuales son Illumina, Ion Torrent y PacBio. Cada plataforma usa una combinación de procedimientos técnicos distintos, principalmente los tres siguientes: (1) la preparación de colecciones de moléculas de ADN, denominadas librerías de ADN; (2) la secuenciación y (3) el análisis de los datos. En general, la preparación de librerías y el análisis de los datos se hacen de distinta manera en cada plataforma para ajustarse a la técnica de secuenciación utilizada, pero las tres plataformas realizan la secuenciación mediante la detección de señales producidas durante la reacción de síntesis de ADN por la incorporación de nucleótidos en el ADN molde. Sin embargo, cada plataforma efectúa este paso de manera diferente. Illumina detecta la fluorescencia producida en grupos de millones de fragmentos del ADN molde; Ion Torrent detecta los cambios de pH por liberación de iones de hidrógeno durante la síntesis, y PacBio detecta la fluorescencia en tiempo real y sobre una única molécula de ADN molde. Para profundizar en los detalles de las distintas tecnologías y plataformas se recomienda consultar los trabajos de Metzker (2010) y Mardis (2013).

Los avances en las tecnologías de secuenciación de alto rendimiento (HTS), que favorecen la identificación de diversos marcadores moleculares como los SNP (polimorfismo de un solo nucleótido), abren un nuevo abanico de posibilidades para los estudios de mapeo genético en especies con genomas complejos como el de la caña de azúcar. Por basarse en las diferencias de un solo nucleótido, los marcadores SNP son más abundantes que cualquier otro tipo de marcador en el genoma. Esto representa una ventaja con respecto a marcadores de ADN como los SSR (*Single Sequence Repeats*), RFLP (*Restriction Length Polymorphisms*) y AFLP (*Amplified Length Polymosrphisms*), entre otros, para discernir la diversidad genética de una especie que presenta una base genética estrecha.

Marcadores moleculares y su aplicación en el mejoramiento genético en Cenicaña

Con base en los recientes avances en la secuenciación de ADN, Cenicaña inició una línea de investigación encaminada a buscar marcadores moleculares asociados a características de interés agronómico para el cultivo de la caña de azúcar, apoyando con ello el esquema de mejoramiento genético. Para el estudio se tomaron 130 individuos representativos de la diversidad genética del banco de germoplasma de Cenicaña. Adicionalmente, se seleccionó un conjunto de 90 genotipos que incluían variedades comerciales, parentales elites y otros genotipos de interés para el Programa de Variedades. En total, este estudio contó con 220 genotipos, entre los que se encuentran quince representantes de la especie *S. officinarum*, uno de *S. spontaneum*, cinco de *S. sinense*, siete de *S. barberi*, tres del género *Erianthus* y 183 híbridos modernos provenientes de dieciocho programas de mejoramiento genético alrededor del mundo. Utilizando la tecnología Illumina, se secuenció una porción del genoma de cada uno de estos individuos para ser utilizada en la identificación de marcadores moleculares de secuencia única, SNP. La porción del genoma a secuenciar se seleccionó con base en las tecnologías GBS (*Genotyping by Sequencing)* y RADSeq (*Restriction Amplified DNA Sequencing*), diseñadas para reducir la complejidad de los genomas al enfocarse solo en una porción significativa de ellos. Las secuencias obtenidas fueron procesadas mediante software bioinformático especializado, con lo cual se logró identifcar más de 50,728 marcadores SNPs altamente polimórfcos, distribuidos a lo largo de todo el genoma. En la **Figura 2** se pueden observar de manera gráfica estos marcadores ubicados sobre el genoma de CC 01-1940. Para profundizar en los detalles de las distintas tecnologías y

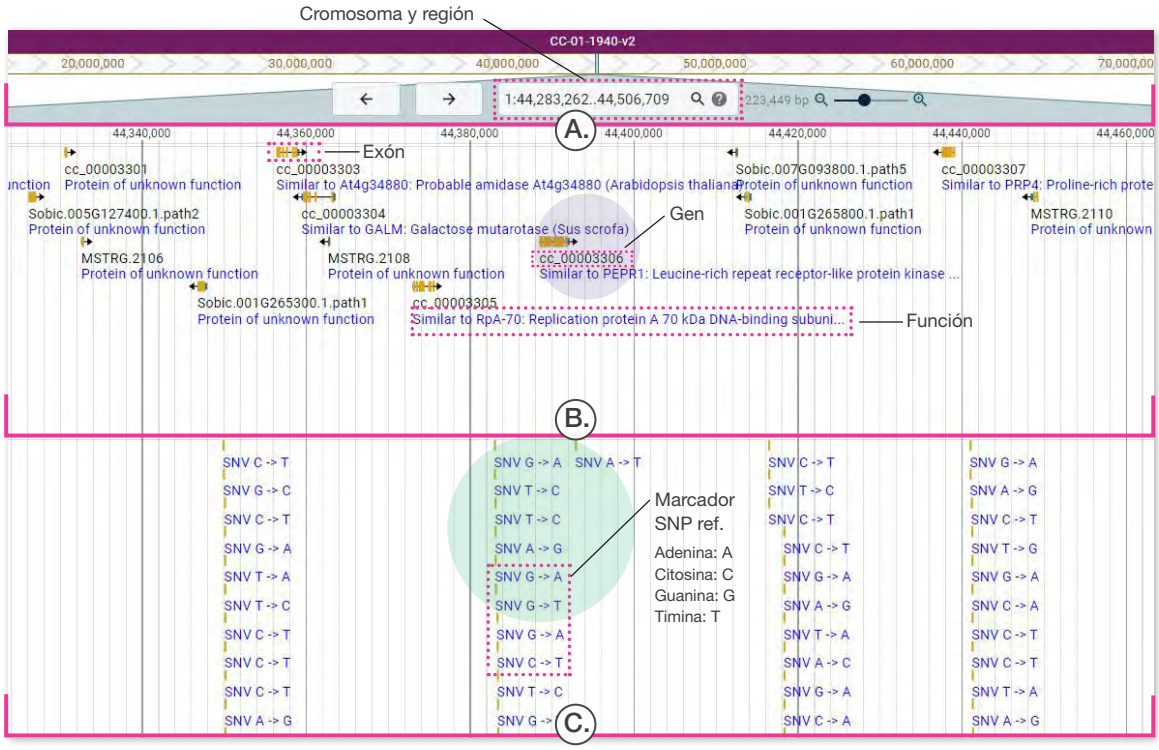


Figura 2. Representación gráfica correspondiente a la caracterización genética del ensamblaje del genoma de caña de azúcar, variedad CC 01-1940. Cromosoma 1 y región seleccionada de manera aleatoria entre las posiciones 44'283,262 y 44'506,709 (A). Genes en la región de interés: identidad, exones, función (B). Marcadores mononucleótidos SNP de referencia (próximos a los genes identificados) (C). Ventanas del navegador JBrowse (Diesh et al., 2023).

plataformas se recomienda consultar a Davey et al. (2011).

Así como el objetivo del estudio ha estado centrado en la identificación y validación de marcadores asociados a fenotipos de interés para el cultivo de la caña de azúcar, como producción de sacarosa, altura de la planta y resistencia a enfermedades y plagas, entre otros, el mismo conjunto de marcadores ha servido para conocer de manera más detallada la diversidad genética existente en la población en estudio. En general, se ha observado que al incrementar el número de marcadores moleculares se incrementa también la precisión en la estimación de las relaciones entre individuos, debido a la posibilidad de establecer un mayor número de comparaciones entre diferentes regiones del genoma de la caña de azúcar. En la **Figura 3** se puede observar cómo es posible separar los 220 genotipos en siete grupos independientes, tres de los cuales incluyen *S. spontaneum*, las especies *S. officinarum, barberi* y *sinensi* y, *Erianthus*, y los otros cuatro (todos ubicados en la misma rama del dendrograma) corresponden a híbridos modernos. Esto es importante, puesto que de las especies del género *Saccharum* incluidas en el estudio, las elegidas son las más distantes evolutivamente, lo cual no se había observado en estudios previos. Los híbridos modernos también muestran una separación que en algunos casos permite discernir los aportes de los diversos programas de mejoramiento. Por ejemplo, se puede identificar un grupo conformado exclusivamente por híbridos desarrollados en Cenicaña (CC), al igual que agrupamientos donde se ubican mayormente individuos pertenecientes a los programas de Australia, Brasil, Estados Unidos o Puerto Rico. Finalmente, y como era de esperar, se observó que los individuos pertenecientes al género *Erianthus* se encuentran más distantes de los otros grupos de individuos.

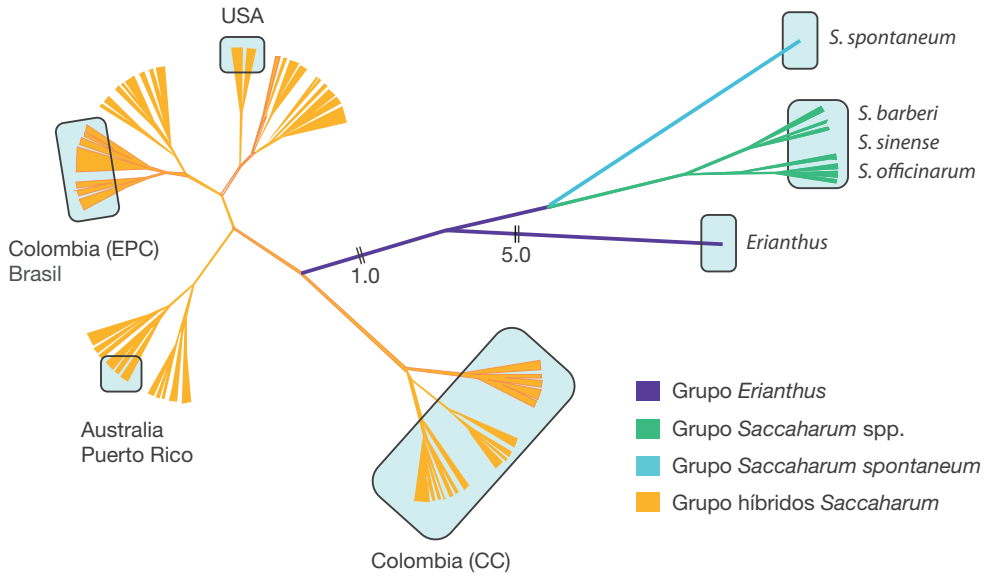


Figura 3. Dendrograma que ilustra la diversidad genética de 220 variedades de caña de azúcar y sus agrupaciones respectivas.

Se distinguen cuatro grupos mayores: en violeta, el '*outgroup*', conformado por individuos del género *Erianthus;* en azul, el grupo con un único individuo perteneciente a la especie *S. spontaneum*; en verde, un tercer grupo conformado por individuos de la especie *S. officinarum* e híbridos tradicionales como *S. barberi* y *S. sinense*; y en naranja, un grupo mayor, donde se ubican la gran mayoría de los híbridos desarrollados en diferentes programas de mejoramiento alrededor del mundo.

Estos análisis también confirmaron la premisa de que la diversidad disponible en el cultivo de la caña de azúcar es escasa. En este estudio, el promedio de similitud genética, cuyo máximo es 1 cuando todos los individuos son iguales, fue de 0.87 (Loaiza et al., 2017), mayor que el encontrado en caracterizaciones previas realizadas en Cenicaña utilizando marcadores clásicos de ADN (Macea, 2009; Espinosa, 2008; Riascos et al., 2003). Las implicaciones de estos resultados son importantes porque apuntan a que se requieren mayores esfuerzos en el esquema de mejoramiento para ampliar la base genética del cultivo, por ejemplo, a través de introgresiones silvestres en el germoplasma moderno.

Los marcadores moleculares SNP (polimorfismo de un solo nucleótido) también han servido para entender la distribución de las variantes genéticas de la población en estudio, lo que a su vez permite separar los individuos en diferentes grupos que comparten patrones de distribución similares. Los resultados obtenidos con esta técnica mostraron cuatro subpoblaciones en el interior de los 220 individuos (**Figura 4**). Como se suponía, debido a su naturaleza silvestre los individuos pertenecientes al género *Erianthus* y a las especies *Saccharum* forman dos poblaciones independientes. Los dos grupos restantes los conforman mayormente híbridos desarrollados a partir de cruzamientos entre especies como *S. officinarum* y *S. spontaneum* e híbridos mejorados. Algunos individuos pertenecientes a *S. officinarum* fueron incluidos dentro de estos dos grupos. Este comportamiento era el esperado, ya que los híbridos modernos han heredado aproximadamente el 80% de su material genético de *S. officinarum*. Finalmente, el resultado más sobresaliente de este análisis muestra que las variedades CP 57-603 y POJ 2878 han tenido una gran influencia en el programa de mejoramiento genético colombiano. La variedad CP 57-603 fue el genotipo más representativo del grupo Q3, y la variedad POJ 2078 lo fue del grupo Q4 (**Figura 4**). Estos resultados concuerdan con la historia de cruzamientos de Cenicaña, que ha usado ambas variedades de manera recurrente para generar nuevos híbridos.

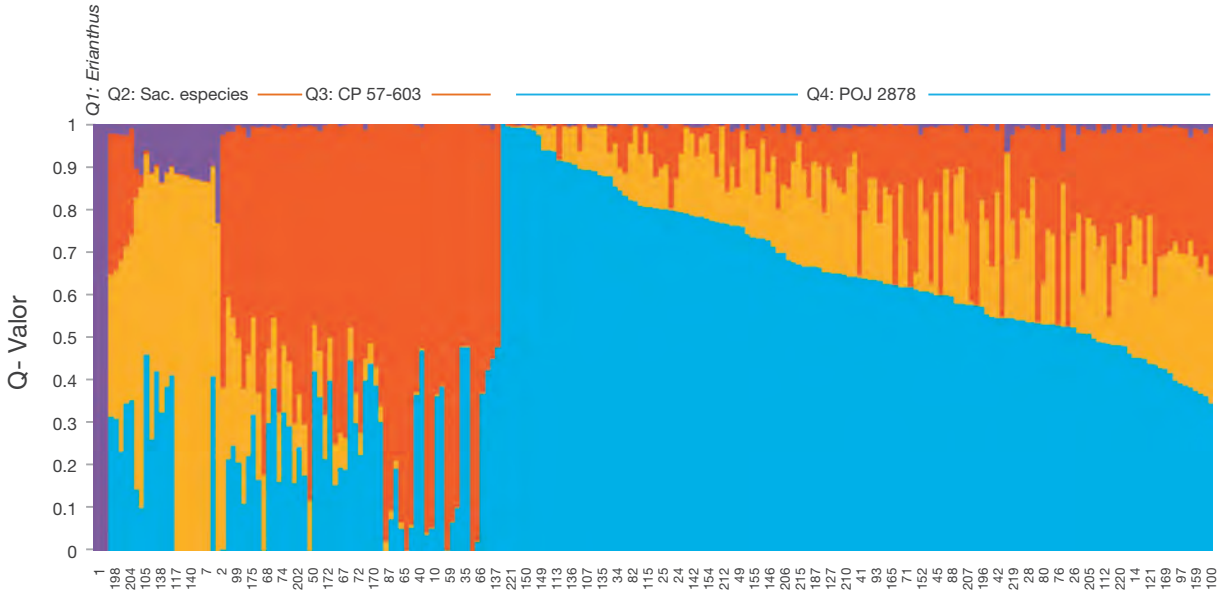
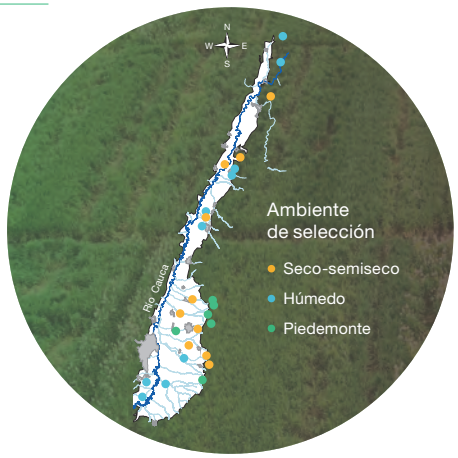


Figura 4. Representación de la estructura genética de 220 variedades de caña de azúcar luego de usar STRUCTURE V2.3.4 (Pritchard et al., 2000) con K = 4.

Los genotipos como CP 57-603 (anaranjado) y POJ 2878 (azul) tuvieron una gran influencia en los híbridos modernos desarrollados por Cenicaña.

Los estudios de asociación genotipo-fenotipo se han abordado con el concepto de asociación de todo el genoma, o GWAS por sus siglas en inglés (*Genome Wide Associations Studies*). Este concepto se diferencia de los estudios de asociación clásicos de QTL (*Quantitative Trait Loci*) porque utiliza poblaciones diversas en lugar de poblaciones biparentales; por la falta de conocimiento sobre la localización genómica de las variantes genéticas, y por el uso de una gran cantidad de marcadores polimórficos (Hirschhorn y Daly, 2005). En los estudios clásicos de QTL, usualmente dos padres contrastantes para una o varias características de interés se cruzan entre sí y su progenie, por lo común poblaciones F_2 en la mayoría de especies y F_1 en el caso de la caña de azúcar, es caracterizada fenotípica y genotípicamente. Debido a que la población utilizada proviene de un cruzamiento biparental, los marcadores moleculares empleados evalúan los eventos de recombinación entre dichos parentales. Cuantos más marcadores se usen, mayor será el cubrimiento de dichos eventos de recombinación y mayor la probabilidad de encontrar un marcador molecular asociado a una característica fenotípica de interés. Cuando se utilizan poblaciones diversas y no biparentales en análisis GWAS (estudios de asociación de todo el genoma), son más numerosos los eventos de recombinación que deben evaluarse, porque la población en estudio proviene de diversos cruzamientos e implica diversos parentales. Al aumentar los eventos de recombinación se debe utilizar un número mayor de marcadores moleculares a fin de tener mayores posibilidades de identificar las asociaciones con los fenotipos de interés, porque se incrementa la probabilidad de encontrar marcadores ligados genéticamente con las variantes genéticas. Debido a que los avances en las tecnologías de secuenciación de ADN han facilitado la identificación a gran escala de variantes de un solo nucleótido, en la actualidad es posible llevar a cabo estrategias de asociación GWAS.

Con el objetivo de contar con un fenotipo de alta calidad, la población de 220 individuos está siendo caracterizada en el valle del río Cauca en los ambientes seco-semiseco, húmedo y de piedemonte, midiendo sus variables agronómicas de interés tales como sacarosa (% caña), toneladas de caña por hectárea (TCH), resistencia a enfermedades y plagas en condiciones controladas; y uso eficiente del nitrógeno, entre muchas otras. A la fecha, los análisis de asociación GWAS en Cenicaña se han centrado en la optimización de las rutinas de asociación actualmente disponibles, ya que la mayoría de los estudios han sido desarrollados para especies diploides y no poliploides como es el caso de la caña de azúcar. A pesar de esto, los primeros ensayos realizados con las herramientas existentes y datos fenotípicos históricos colectados para casi la totalidad de los 220 individuos, han mostrado un amplio potencial para la identificación de marcadores asociados con la producción de sacarosa (Riascos et al., 2016).



Con el objetivo de contar con un fenotipo de alta calidad, la población de 220 individuos está siendo caracterizada en el valle del río Cauca en los ambientes seco-semiseco, húmedo y de piedemonte.

El genoma de la caña de azúcar

¿Qué es un genoma y para qué sirve?

Todas las características de un ser vivo responden a la acción de sus genes y a la influencia del ambiente. La capacidad de generar raíces, tolerar estres biótico o abiótico, la coloración de los jugos en el caso de la caña de azúcar, entre otras, son características controladas por la información almacenada en el material genético (conocido como ADN) y moldeadas en mayor o menor medida por las condiciones del entorno. La estructura del ADN es una cadena en forma de doble hélice, constituida por enlaces exclusivos entre adenina y timina (A:T) y guanina y citosina (G:C), y sostenida por una estructura de ácidos fosfóricos. Debido a su gran longitud, la doble hélice de ADN está sujeta a diversos niveles de compactación, de manera que pueda ser almacenada de manera eficiente en los cromosomas (Annunziato, 2008). En organismos eucariotas (con compartimentos celulares), el ADN se almacena en su mayoría en el núcleo; sin embargo, también existen porciones del genoma en las mitocondrias y los cloroplastos.

Secuenciar y ensamblar un genoma consiste en descifrar el orden de todos los nucleótidos que este almacena. A partir de aquí es posible predecir qué porciones de ese genoma son genes, qué porciones son secuencias regulatorias y cuáles tienen función desconocida. Por medio de análisis comparativos llevados a cabo con herramientas computacionales especializadas se puede asignar identidades a cada gen, lo que facilita la definición de familias génicas y sus posibles funciones. Los análisis comparativos entre genomas de diferentes especies permiten estimar el grado de conservación o divergencia evolutiva de una especie. Con ello es posible saber si la caña de azúcar es más cercana genéticamente al sorgo o al maíz o a cualquier especie cuyo genoma esté secuenciado. En el interior de una especie, sobre todo cuando se cuenta con una caracterización del fenotipo, las comparaciones entre diferentes genomas facilitan la identificación de regiones asociadas con características de interés. Un genoma ensamblado de manera detallada permite determinar el número de copias (ploidía) y el número de cromosomas de una especie. En resumen, la secuenciación y el ensamblaje de un genoma son una base sólida para la investigación básica y aplicada.

Ensamblaje de genomas

Los recientes avances en las tecnologías de secuenciación de alto rendimiento han permitido avanzar en el ensamblaje de genomas complejos, generando herramientas prometedoras en la búsqueda del mejoramiento genético. El aumento en el tamaño de las secuencias, los bajos costos y la implementación de nuevos algoritmos acoplados a estas tecnologías han servido como base para definir nuevas estrategias que permitan la reconstrucción de genomas complejos como el de la caña de azúcar. Entre estas nuevas tecnologías y en el caso específico del ensamblaje de genomas, destacan la tecnología de secuenciación de PacBio (Rhoads & Au, 2015) y la tecnología de secuenciación Illumina (Bentley et al., 2008). La tecnología PacBio se caracteriza por generar millones de fragmentos largos con un tamaño promedio de 15 kbp, lo cual es ideal para reconstruir genomas altamente repetitivos; por su parte, la tecnología Illumina se caracteriza por secuenciar millones de fragmentos pequeños, con un tamaño promedio de 150 pb pero con una alta calidad en sus bases secuenciadas, muy útil para corregir errores de secuenciación que puedan presentar las lecturas largas de PacBio.

Tradicionalmente se distinguen dos estrategias mayores para el ensamblaje de genomas: *clone by clone* (también conocida como *BAC by BAC*) y *Whole Genome Shotgun Sequencing* (WGSS). La tecnología *clone by clone* parte de la creación de un mapa genético para el individuo de interés. Como se mencionó previamente, los mapas genéticos son representaciones del genoma que se construyen a partir de la organización posicional de un número considerable de marcadores de ADN. Posteriormente, dicho genoma es fragmentado en regiones de aproximadamente 150 kb y clonado en vectores de ADN (*Bacterial Artificial Chromosomes*, BAC), los cuales sirven para multiplicar (clonar) cada uno de los fragmentos clonados. Dichas regiones de 150 kb son adicionalmente fraccionadas en regiones más pequeñas, de aproximadamente 500 pb, y secuenciadas con el fin de reconstruir los fragmentos más grandes. La reconstrucción final del genoma tiene en cuenta el mapa genético del cual se partió, lo cual facilita la organización de las regiones de 150 kb en los cromosomas respectivos (**Figura 5**). Debido a lo laborioso de la estrategia *clone by clone,* hoy en día se utiliza cada vez menos. En el caso de *whole genome shotgun*, el ADN es extraído, fragmentado y secuenciado sin seguir un orden en particular. El

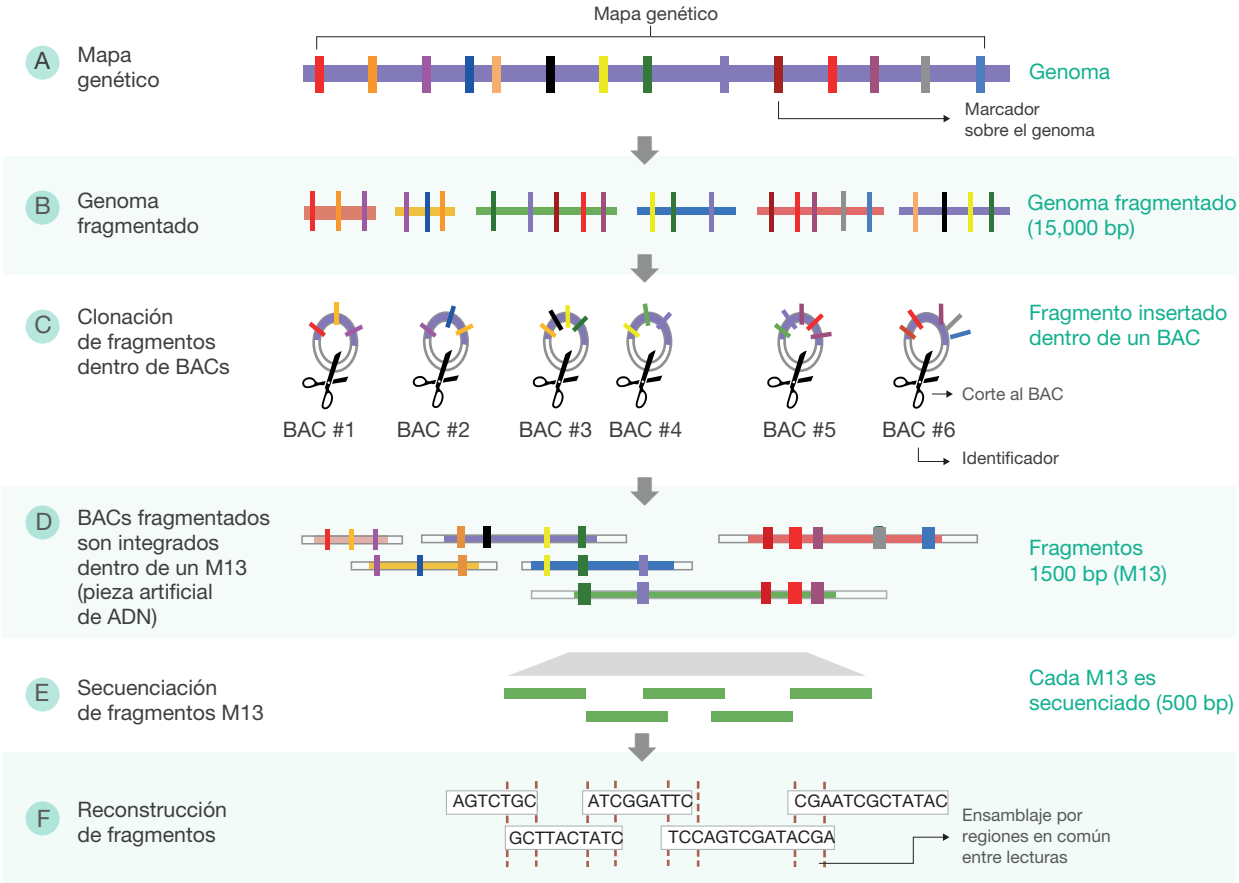


Figura 5. Esquema general de la estrategia de secuenciación *clone by clone.* Mapa genético *(*A). Genoma fragmentado (B). Clonación de fragmentos dentro de BACs (C). BACs fragmentados son integrados dentro de un M13 (pieza artificial de ADN) (D). Secuenciación de fragmentos M13 (E). Reconstrucción de fragmentos (F).

éxito de esta estrategia depende mayormente de una secuenciación robusta, que represente varias veces el genoma de interés, y una buena capacidad computacional que facilite la ejecución de múltiples alineamientos entre dichas secuencias, a fin de decidir cuál es el mejor ensamblaje (**Figura 6**). Debido a su simplicidad conceptual y al hecho de que con ella se pueden conseguir avances de manera rápida, comúnmente se prefiere esta estrategia a la *clone by clone;* sin embargo, no es fácil de implementar cuando se trata de genomas complejos con altos contenidos de secuencias repetidas. Esto debido a que las regiones repetidas dificultan el trabajo de los algoritmos de ensamblaje, que no logran distinguir la posición específica (única) de una secuencia de ADN que se repite varias veces en el genoma.

> Los recientes avances en las tecnologías de secuenciación de alto rendimiento han conducido a que el ensamblaje de genomas completos sea una metodología prometedora en la búsqueda del mejoramiento genético.

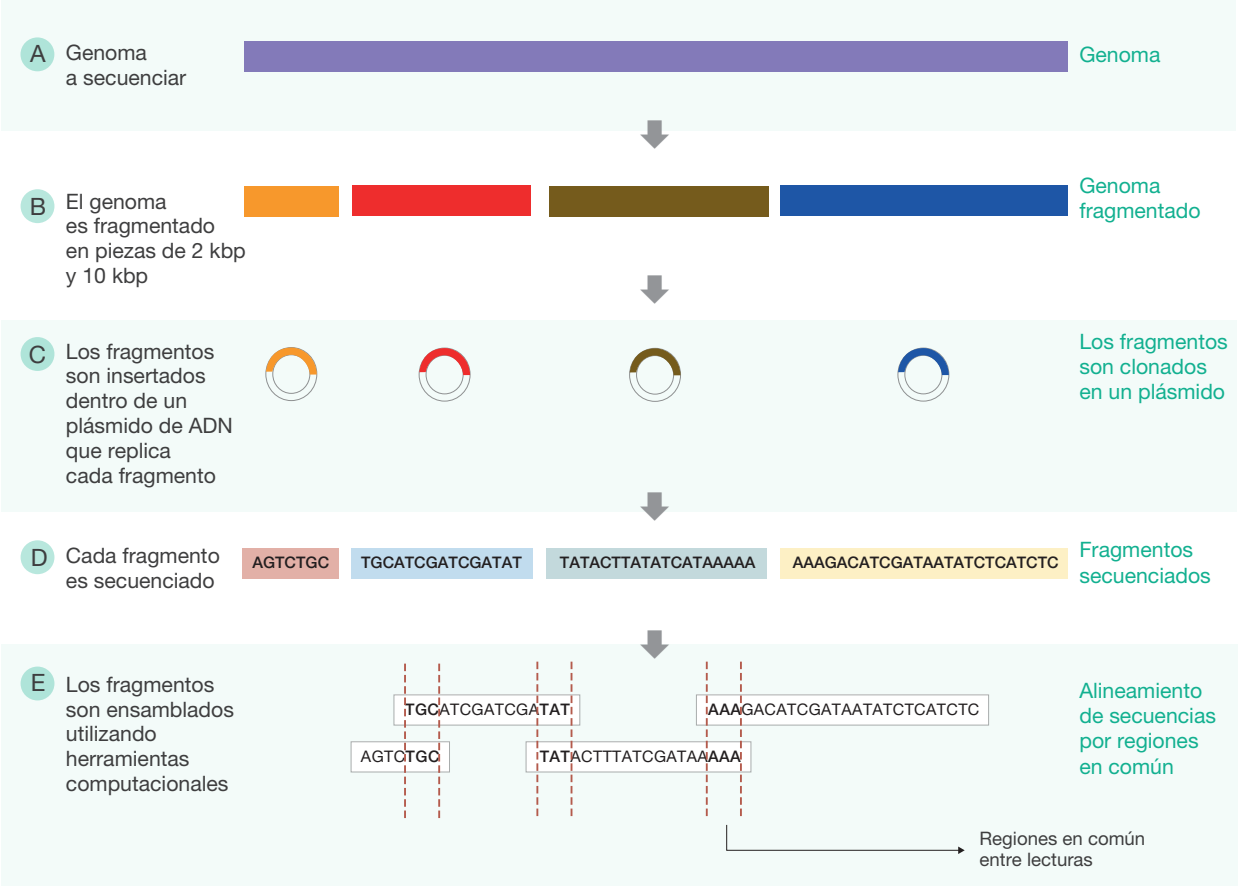


Figura 6. Esquema general de la estrategia de secuenciación *whole genome sequencing.* Genoma a secuenciar (A). El genoma es fragmentado en piezas de 2 kbp y 10 kbp (B). Los fragmentos son insertados dentro de un plásmido de ADN que replica cada fragmento (C). Cada fragmento es secuenciado (D). Los fragmentos son ensamblados utilizando herramientas computacionales (E).

Avances en la secuenciación del genoma de la caña de azúcar

En la actualidad varios estudios se orientan a secuenciar genomas híbridos, como los de las especies progenitoras *S. spontaneum* y *S. officinarum*. Dado que los genomas híbridos comportan un grado de complejidad mayor que el de las especies progenitoras, no es extraño que los mayores avances se hayan logrado en estas últimas. El profesor Ray Ming, de la Universidad de Illinois, ha liderado un proyecto de ensamblaje del genoma de la especie *S. spontaneum* (Zhang et al., 2018) y de la especie *S. officinarum* (Comunicación personal). Adicionalmente, la construcción del genoma de *S. spontaneum* ha sido financiada parcialmente con fondos del Consorcio Internacional de Biotecnología de la Caña de Azúcar (ICSB, por sus siglas en inglés), del cual Colombia es miembro y Cenicaña su representante oficial.

En el ensamblaje de ambos genomas se ha empleado una estrategia híbrida. En términos de secuenciación se han producido librerías BAC para cada una de las especies, que se han reconstruido a través de la secuenciación (usando la tecnología de secuenciación Illumina) de fragmentos de ADN cortos. En igual sentido, y con el objetivo de dar continuidad a los ensamblajes producidos a partir de secuencias cortas, se ha empleado la tecnología PacBio a manera de *shotgun,* sin utilizar los clones de las librerías BAC. A la fecha ha sido finalizado el ensamblaje de *S. spontaneum* y continúa en proceso el de *S. officinarum,* utilizando en ambos casos la tecnología de ensamblaje apoyada en la elaboración de mapas genéticos y la construcción de librerías Hi-C (**Cuadro 1**) (Belton et al., 2012). Esta última permite capturar información sobre la proximidad estructural entre dos segmentos de ADN, lo que a su vez facilita inferir el orden de los nucleótidos en el genoma.

El genoma de *S. spontaneum* se secuenció y se ensambló a partir del genotipo AP 85-441, un haploide originado a su vez en el genotipo SES208 y desarrollado por Moore et al. (1989). Por ser un genotipo haploide, AP 85-441 es tetraploide (x = 4), no octoploide (x = 8) como su progenitor original, y tiene un total de 32 cromosomas (ocho por cada subgenoma). Con este trabajo se logró identificar en *S. spontaneum* 35,525 genes, de los cuales 4289 (12.7%) tienen cuatro alelos; 9792 (27.6%), tres alelos; 14,797 (41.7%), dos alelos y 6647 (18.7%), un alelo. Por ser un genotipo tetraploide se espera un máximo de cuatro alelos (variantes de genes) por cada gen; sin embargo, no todos los genes mostrarán este número de alelos, ya que no todas las regiones del genoma tienen el mismo grado de relevancia para el mantenimiento de la especie y por tanto no se ven afectadas por una tasa similar de mutaciones. Los autores del estudio identificaron también 361 genes pertenecientes a la familia de genes de resistencia NBS (*nucleotide binding site*), característica heredada a los híbridos modernos del parental *S. spontaneum* (Zhang et al., 2018). Lo peculiar de esta familia de genes es que la mayoría de sus miembros (80%) se ubican en tan solo cuatro cromosomas, en regiones que han sufrido varios rearreglos en la historia evolutiva de *S. spontaneum*, lo que también ejemplifica la relevancia de ciertas regiones del genoma para esta especie.

A pesar de que la mayoría de sus representantes son decaploides (x = 10), el genotipo de la especie *S. officinarum,* LA-Purple, escogido para ser secuenciado y ensamblado, es un octoploide (x = 8) que posee 80 cromosomas (**Cuadro 1**). Los avances más recientes sobre este ensamblaje muestran un total de 78 moléculas de mayor tamaño o *super-scaffolds*, número bastante cercano a los 80 cromosomas de la LA-Purple y que representan cerca del 80% de su genoma. Además de las especies progenitoras, varios programas de mejoramiento alrededor del mundo han avanzado en sus esfuerzos por secuenciar cultivares híbridos de interés particular. Por ejemplo, en Brasil dos grupos de investigación, uno en el Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), liderado por Diego Riaño y otro en la Universidad de São Paulo, liderado por Glaucia Souza, se han enfocado en el híbrido SP 80-3280 de caña de azúcar, un cultivar de amplia relevan-

Cuadro 1. Descripción de los trabajos de secuenciación de organismos relacionados con caña de azúcar.

Organismo (genotipo)	Programa de mejoramiento	Constitución cromosómica	Tecnología de secuenciación	Software de ensamblaje	Nivel de ensamblaje	Calidad del ensamblaje[2]	Referencia
Sorghum bicolor	No aplica	Diploide (x = 2, 2n = 20)	Illumina / Sanger	SOAPdenovo	Cromosoma	Contig: 874 N50 = 64.2e6 bp	Zheng et al. (2011)
Saccharum officinarum (LA-Purple)	No aplica	Octoploide (x = 8, 2n = 80)	PacBio / Hi-C	Canu	Scaffold[1]	Contig: 49,710 N50 = 136,410 bp	Zhang et al. (2016)
Saccharum spontaneum (AP85-441)	No aplica	Tetraploide (x = 4, 2n = 32)	Hi-C, mapa genético	Canu	Cromosoma	Contig: 15,768 N50 = 91,359,291 bp	Chen et al. (2017)
Híbrido (SP-80-3280)	Brasil	Decaploide posiblemente	Illumina / TruSeq Synthetic	Celera-Assembler	Scaffold	Contig: 190,697 N50 = 8451 bp	Riaño et al. (2017)
Híbrido (KK3)	Tailandia	Decaploide posiblemente	PacBio	Canu	Contig	Contig: 104,407 N50 = 83,343 b	Shearman et al. (2018)
Híbrido (R570)	Francia	Decaploide posiblemente	BAC (WGP)	HGAP3	Pseudo cromosoma	Contig: 211 N50 = 41.25e6 bp	Garsmeur et al. (2018a)
Híbrido (CC 01-1940)	Colombia	Decaploide posiblemente	Illumina / PacBio, Hi-C	Canu	Contig	Contig: 35,089 N50 = 34.94e6 bp	Trujillo-Montenegro et al. (2021)

1. Scaffold: es una secuencia de consenso generada a partir de otras de menor tamaño o contigs.
2. Contig: es el consenso del traslape de lecturas en una región del ensamblaje.
3. N50: métrica de la calidad del ensamblaje, que se entiende como el mínimo tamaño necesario para tener el 50% del tamaño del genoma.

cia para los programas de mejoramiento genético de ese país. Aunque su materia de estudio es el mismo cultivar, cada grupo de investigación ha definido objetivos distintos en los ensamblajes. En el caso de CNPEM, en principio se buscó evaluar la tecnología de secuenciación *TruSeq Synthetic Long Read* (McCoy et al., 2014) para el ensamblaje de genomas complejos. Esta estrategia permite reconstruir fragmentos grandes de ADN, que son ensamblados utilizando tecnologías de secuenciación de fragmentos cortos de Illumina. Aunque los resultados fueron en principio alentadores, la información secuenciada solo permitió construir una versión preliminar o borrador (*draft*) de dicho genoma. De hecho, este fue el primer resultado publicado y sus datos son de uso libre (Riaño et al., 2017). En el caso de la Universidad de São Paulo, el ensamblaje utilizó el mismo tipo de tecnología de secuenciación que empleó CNPEM, pero en una mayor proporción y con el uso de secuencias BACs. Esta estrategia permitió generar un ensamblaje que logró identificar 373,869 genes (cifra que incluye el número de copias homólogas de cada gen) que constituyen el genoma total del híbrido SP 80-3280. Y aunque este híbrido no está conformado por grandes moléculas como los cromosomas, tales resultados indican que el ensamblaje contiene la mayor cantidad de información extraída de un híbrido de caña de azúcar, lo que representa un gran paso hacia el ensamblaje completo de su genoma.

En Francia los mayores esfuerzos se han orientado a la caracterización detallada del genoma del híbrido R570 (Garsmeur et al., 2018). Este híbrido de caña de azúcar fue utilizado para producir una población de mapeo genético, a través de autofecundación, que sirvió para identificar el marcador molecular *Bru1* que confiere resistencia al hongo *Puccinia melanocephala*, causante de la enfermedad roya café (Daugrois et al., 1996). Con el objetivo de clonar el gen asociado al marcador *Bru1*, Tompkins et al. (1999) desarrollaron una librería BAC a partir del genoma del híbrido R570. Esta misma librería ha sido usada de manera alternativa para seleccionar, con base en el genoma del sorgo, un conjunto de clones que representan una región rica en genes del genoma del híbrido R570. En otras palabras, el propósito de este trabajo no fue secuenciar la totalidad del genoma de R570, sino la porción rica en genes de una representación del genoma monoploide o conjunto único de cromosomas. Los resultados obtenidos muestran un tamaño de genoma monoploide de 426.7 Mbp y un total de 25,316 genes, número inferior al identificado por Ming y colaboradores en el genoma de *S. spontaneum* (Garsmeur et al., 2018a; 2018b). Sin embargo, estos resultados fueron los previstos, dado que en la construcción del genoma del hibrido R 570 se usó el genoma monoploide del sorgo (27,532 genes) como referencia. Es relevante mencionar que esta investigación fue parcialmente financiada por el Consorcio Internacional de Biotecnología (ICSB) y que actualmente, en colaboración con el Joint Genome Institute (JGI) de Estados Unidos, se avanza en el ensamblaje de la totalidad del genoma del híbrido R570 (Schmutz et al., 2018). Similares investigaciones se adelantan en Tailandia, donde el cultivar Khon Kaen 3 (KK3) es actualmente el de mayor prevalencia en ese país, cuyos esfuerzos en biotecnología están encaminados a secuenciar y ensamblar su genoma. Los avances más recientes en cada una de estas investigaciones se resumen en el **Cuadro 1**. Finalmente, valga subrayar que Cenicaña ha generado una primera versión del genoma de la variedad CC 01-1940 (Trujillo-Montenegro et al., 2021), un híbrido de gran productividad desarrollado recientemente y ampliamente adoptado por los cultivadores colombianos.

Cenicaña ha generado una primera versión del genoma de la variedad CC 01-1940, un híbrido de gran productividad desarrollado recientemente y ampliamente adoptado por los cultivadores colombianos.

Avances en la construcción del genoma de la variedad CC 01-1940

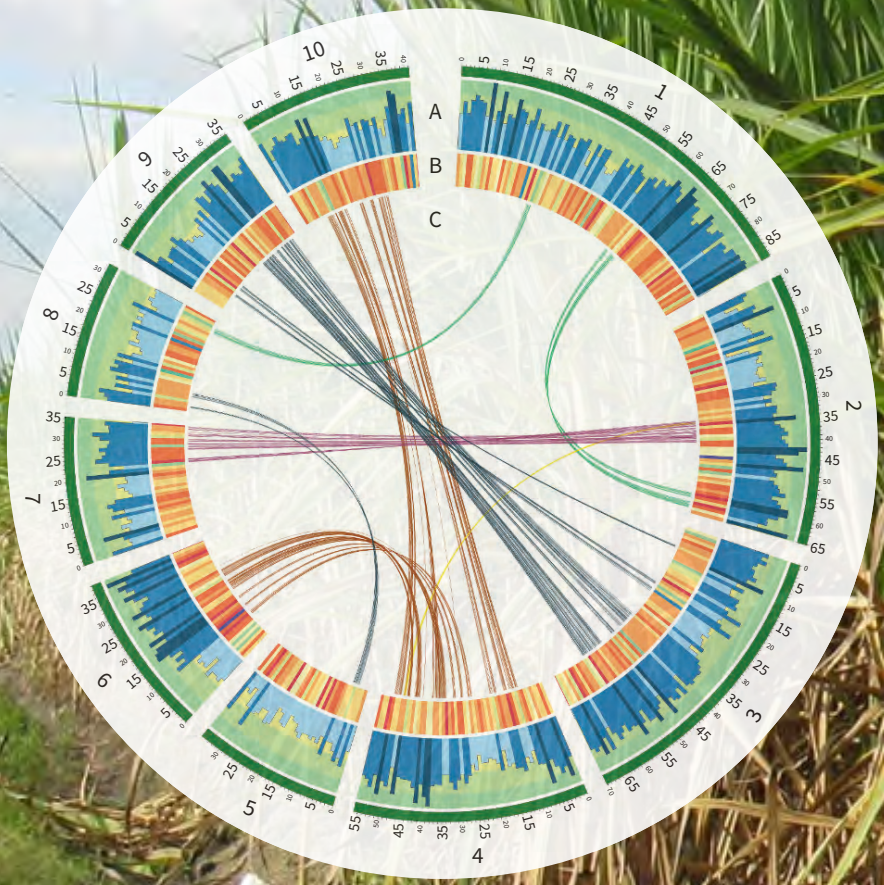
Hasta la fecha no existe un genoma de referencia completo para la caña de azúcar que permita avanzar de manera más precisa en los análisis bioinformáticos que se ejecutan en los programas de mejoramiento para este cultivo. Por esta razón, Cenicaña decidió implementar los últimos avances de estas nuevas tecnologías y definió una estrategia para ensamblar el genoma de la variedad CC 01-1940. Esta variedad se caracteriza por producir en promedio veinte toneladas más de caña por hectárea y dos toneladas más de sacarosa por hectárea que el testigo comercial CC 85-92, un genotipo históricamente sobresaliente. De la variedad CC 01-1940 se conoce, como resultado de un estudio de citometría, que su ADN total es 11.21 $\pm$ 0.374 Gbp y el tamaño de su genoma monoploide es de 1.019 Gb $\pm$ 0.033 con base en una ploidía de 10x.

La estrategia usada para la construcción del genoma de esta variedad consistió en secuenciar 100.5 Gbp de lecturas largas de PacBio, 100.7 Gbp de datos Illumina (lecturas cortas) y 102.8 Gbp de datos Illumina (lecturas cortas) utilizando la metodología de secuenciación Hi-C. Para el ensamblaje de las lecturas largas de PacBio se utilizó la herramienta Flye assembler (Kolmogorov et al., 2020), con la que se obtuvo una versión base del ensamblaje, cuyos eventuales errores fueron corregidos utilizando las lecturas Illumina y la herramienta NGSEP (V 4.0.1) (Ghurye et al., 2017); y para mejorar el orden y la posición de estas secuencias sobre el genoma se utilizaron las lecturas Hi-C y la herramienta ALLHIC Assembler tool (Zhang et al., 2019), con lo cual fue posible generar un ensamblaje de mejor continuidad, con 10 pseudo-cromosomas, 44 scaffolds, 35,035 contigs y un N50 34.94 Mbp, para un total ensamblado de 903.2 Mbp.

Para confirmar los resultados obtenidos en este ensamblaje fueron mapeados a esta referencia datos GBS y RADSeq generados previamente, lo que indicó un incremento del 30% en los porcentajes de mapeo que se tenían inicialmente con el genoma de sorgo a un 88%. Esto permitió encontrar un total de 50,728 marcadores moleculares SNP. Además, se analizó la presencia de contaminación dentro del ensamblaje, con el propósito de confirmar que lo ensamblado correspondía en muy alto porcentaje a caña de azúcar. Para tal efecto se utilizó la herramienta BlobTools, con la que fue posible determinar que el 99.8% de la información ensamblada pertenecía a caña de azúcar; el 0.2% restante provenía de hongos y bacterias, lo que es normal encontrar cuando se ensamblan genomas de plantas. Adicionalmente se evaluó qué tan completo estaba el ensamblaje, con base en la herramienta Busco v 4.0.6 (Simão et al., 2015), que da una medida cuantitativa de este factor teniendo en cuenta la información evolutiva esperada en el contenido de genes del genoma. Con tal recurso se comprobó que el ensamblaje contenía el 96% de la información genética esperada; hecho que superó en este aspecto a todos los ensamblajes de caña de azúcar publicados hasta el momento. Además, se encontró que el ensamblaje contenía aproximadamente 63,724 genes.

Hasta la fecha no existe un genoma de referencia completo para la caña de azúcar que permita avanzar de manera más precisa en los análisis bioinformáticos que se ejecutan en los programas de mejoramiento para este cultivo. Por esta razón, Cenicaña decidió implementar los últimos avances de estas nuevas tecnologías y definió una estrategia para ensamblar el genoma de la variedad CC 01-1940.

Este genoma está siendo utilizado como referencia para la identificación de marcadores moleculares con potencial en el programa de mejoramiento genético de Cenicaña.



Escala de intensidad

Alto

Medio

Bajo

Cromosomas

Frecuencia de genes dentro de cada cromosoma

Regiones repetitivas del genoma

Genoma ensamblado del híbrido de caña de azúcar CC 01-1940.

Variedad CC 01-1940

Referencias

AGI, The Arabidopsis Genome Initiative. (2000). Analysis of the Genome Sequence of the Flowering Plant Arabidopsis Thaliana. Nature 408 (6814), pp. 796–815.

Annunziato A. (2008). DNA packaging: nucleosomes and chromatin. Nature Education, 1:26.

Belton J.M., McCord R.P. Gibcus J.H., Naumova N., Zhan Y. & Dekker J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58, pp. 268-276.

Bentley D. R., Balasubramanian S., Swerdlow H. P., Smith G. P., Milton J., Brown C. G., Hall K. P., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature, 456(7218), pp. 53–59.

Chen Y., Zhang Q., Hu W. et al. (2017). Evolution and expression of the fructokinase gene family in *Saccharum*. BMC Genomics 18:197. https://doi.org/10.1186/s12864-017-3535-7

Daugrois J.H., Grivet L., Roques D., Hoarau J.Y., Lombard H., Glaszmann J.C. & D'Hont A. (1996). A putative major gene for rust resistance linked with a RFLP marker in sugar cane cultivar "R 570". Theoretical and Applied Genetics 92, pp.1059-1064.

Davey J., Hohenlohe P., Etter P., Boone J., Catchen J. & Blaxter, M. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. Nature Publishing Group 12, pp. 499–510.

D'Hont A. (2005). Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and bananas. Cytogenics and Genome Research, 109, pp. 27–33.

D'Hont A., Grivet L., Feldmann P., Glaszmann J.H., Rao S., Berding N. (1996). Characterization of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol. Gen. Genet. 250, pp. 405-413.

Diesh, C., Stevens, GJ, Xie, P. et al. (2023). JBrowse 2: un navegador de genoma modular con vistas de sintenia y a variación estructural. Genoma Biol 24, 74. https://doi.org/10.1186/s13059-023-02914-z

Espinosa, K. (2008). Evaluación de la diversidad genética de 136 variedades de caña de azúcar (*Saccharum* spp.) usando microsatélites. Documento de trabajo N.° 624. Centro de Investigación de la Caña de Azúcar de Colombia (Cenicaña), 32 pp.

Foiada F., Westermeier P., Kessel B., Ouzunova M., Wimmer V., Mayerhofer W. & Schön C.C. (2015). Improving resistance to the European corn borer: a comprehensive study in elite maize using QTL mapping and genome-wide prediction. Theoretical and Applied Genetics, 128(5), p. 875-891.

Garsmeur O., Droc G., Antonise R. et al. (2018a). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun 9, 2638 . https://doi.org/10.1038/s41467-018-05051-5

Garsmeur O. (2018b). A Reference Sequence of the Monoploid Genome of Sugarcane. Plant Animal Genome Conference XXVI: The Largest Ag-Genomics Meeting in the World. San Diego, CA.

Hirschhorn J. N. & Daly M. J. (2005). Genome-wide association studies for common diseases and complex traits. Nature reviews genetics, 6(2), pp. 95-108.

Jiang H., Feng Y., Bao L., Li X., Gao G., Zhang Q. & He Y. (2012). Improving blast resistance of Jin 23B and its hybrid rice by marker-assisted gene pyramiding. Molecular breeding, 30(4), pp. 1679-1688.

Kolmogorov M., Bickhart D.M., Behsaz B. et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17, 1103–1110 (2020). https://doi.org/10.1038/s41592-020-00971-x

Lee M. (2007). Maize breeding and genomics: an historical overview and perspectives. In: Varshney, R. K. and Tuberosa, R. eds. Genomics-Assisted Crop Improvement. Dordrecht: Springer Netherlands, pp. 129–146.

Li R., Chia-Ling H., Amanda Y., Zhihong Z., Xiaoliang R. & Zhongying Z. (2015). Illumina Synthetic Long Read Sequencing Allows Recovery of Missing Sequences Even in the 'Finished' C. Elegans Genome. Scientific Reports 5 (1). https://doi.org/10.1038/srep10814.

Loaiza C.D., Duitama J.A. & J. J. Riascos (2017). Calculation of genetic distances as a function of allelic dosage increases clustering accuracy in polyploids. IV congreso colombiano de Biología computacional y bioinformática. Presentación, VIII conferencia Iberoamericana de Bioinformática. Cali, Colombia.

Macea E. (2009). Análisis molecular de la diversidad genética de caña de azúcar del banco de germoplasma de Cenicaña usando microsatélites. Documento de trabajo N.° 704. Laboratorio de biotecnología. Centro de Investigación de la Caña de Azúcar de Colombia (Cenicaña), 38 pp.

Mardis E. R. (2013). Next-generation sequencing platforms. En Annu. Rev. Anal. Chem. (Palo Alto. Calif)., vol. 6, n.° March, pp. 287-303, pp. 2-5.

Maxam A. M. & Gilbert W. (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America, 74(2), pp. 560–564.

McCoy R.C., Taylor R.W., Blauwkamp T.A., Kelley J.L., Kertesz M., Pushkarev D., Petrov D.A. & Fiston-Lavier A.-S. (2014). Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. Plos One 9(9), p. e106689.

Metzker M. L. (2010). Sequencing technologies - The next generation. Nature Reviews. Genetics 11 (1), pp. 31-46.

Moore P. A., Paterson F.C. & Tew T. (2014). Sugarcane: The Crop, the Plant, and Domestication. En: P. H. Moore and F. C. Botha, Eds. Sugarcane: Physiology, Biochemistry and Functional Biology (pp. 1 -17). John Wiley & Sons, 765 pp.

Paterson A.H., Bowers J.E., Bruggmann R., Dubchak I., Grimwood J., Gundlach H., Haberer G., Hellsten U., Mitros T., Poliakov A., Schmutz J., Spannagl M., Tang H., Wang X., Wicker T., Bharti A.K. Chapman J., Feltus F.A., Gowik U., Grigoriev I.V. & Rokhsar D.S. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature 457 (7229), pp. 551–556.

Price D. & J. Daniels (1968). Cytology of South Pacific Sugarcane and Related Grasses: With special reference to Fiji, Journal of Heredity, Volume 59, Issue 2, pp. 141-145.

Rhoads A. & Au K.F. (2015). Pacbio sequencing and its applications. Genomics, proteomics & bioinformatics / Beijing Genomics Institute 13(5), pp. 278–289.

Riaño-Pachón D.M. and Mattiello L. (2017). Draft genome sequencing of the sugarcane hybrid SP80-3280. F1000Research 6:861. https://doi.org/10.12688/f1000research.11859.2

Riascos J.J. (2016). Advances in the Utilization of Next Generation Sequencing Data in the Colombian Sugarcane Breeding Program. In: Plant and Animal Genome XXIV Conference. Plant and Animal Genome.

Riascos J.J. (2003). Diversidad genética en variedades de caña de azúcar (*Saccharum* spp.) usando marcadores moleculares. Revista Colombiana de Biotecnología. 2003;5, pp. 6-15.

Schmutz J. (2018). Whole Genome Sequencing of Sugarcane: Building off the Foundation of the Single Haplotype Path. Plant Animal Genome Conference XXVI: The Larguest Ag-Genomics Meeting in the World. San Diego, CA.

Schnable P. S. et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science 326, pp. 1112–1115.

Shearman J., Sonthirod C., Naktang C. et al. (2018). Assembling a hybrid a sugarcane genome. https://pag.confex.com/pag/xxvi/meetingapp.cgi/Paper31646

Sreenivasan T.V. & Mohan Naidu K. (1987). Conservation of sugarcane germoplasm. In: Copersucar International Sugarcane Breeding Workshop, São Paulo: Copersucar, pp 33–53.

Steele K.A., Price A.H., Shashidhar H.E. & Witcombe J.R. (2006). Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theoretical and Applied Genetics, 112(2), pp. 208-221.

Trujillo-Montenegro J.H., Rodríguez Cubillos M.J., Loaiza Orduz C.D. et al. (2021). Unraveling the genome of a high yielding Colombian sugarcane hybrid. Frontiers in Plant Science, 12, 1311pp. https://www.frontiersin.org/articles/10.3389/fpls.2021.694859/full

Zhang J., Sharma A., Yu Q. et al. (2016). Comparative structural analysis of Bru1 region homeologs in Saccharum spontaneum and S. officinarum . BMC Genomics 17:446. https://doi.org/10.1186/s12864-016-2817-9

Zhang X., Zhang S., Zhao Q., Ming R. and Tang H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nature Plants 5, pp. 833–845. doi: 10.1038/s41477-019-0487-8

Zheng LY., Guo XS., He B. et al. (2011). Genome-wide patterns of genetic variation in sweet and grain sorghum (Sorghum bicolor). Genome Biol 12:R114. https://doi.org/10.1186/gb-2011-12-11-r114

Zimin A.V., Puiu D., Hall R., Kingan S., Clavijo B.J. & Salzberg S.L. (2017). The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum.* GigaScience 6(11), pp. 1–7.

LOS AUTORES

Jhon Henry Trujillo Montenegro

Ingeniero de Sistemas y Computación, doctorado en Ingeniería con énfasis en Ciencias de la Computación, egresado en 2009 de la Pontificia Universidad Javeriana, seccional Cali, y en 2021 de la Universidad del Valle, sede Cali. En el año 2016 inicia como estudiante de doctorado en el Centro de Investigación de la Caña de Azúcar de Colombia, donde lleva a cabo su tesis titulada: "Construcción de un genoma y una huella molecular de caña de azúcar utilizando secuenciación de alto rendimiento". Cuenta con más de siete años de experiencia en el área de la bioinformática aplicada al mejoramiento genético de la caña de azúcar, llevando a cabo proyectos relacionados con la caracterización genética de individuos de caña de azúcar, ensamblaje de genomas complejos, manejo de datos de secuenciación de nueva generación, selección genómica, estudios de asociación genotipo-fenotipo y diseño y programación de algoritmos.

María Camila Martínez Villa

Bióloga con maestría en Biología Computacional de la Universidad de los Andes. Durante 2018 y 2019 contribuyó al equipo de biotecnología de Cenicaña. Su labor incluyó la aplicación de modelos lineales mixtos utilizando herramientas avanzadas de consulta, visualización y análisis conectadas en un flujo de trabajo.

John Jaime Riascos Arcos

Nació en Cali, en 1978. Es biólogo de la Universidad del Valle y obtuvo su doctorado en la facultad de Agronomía, en North Carolina State University (NCSU). Ha sido parte de diversos grupos de investigación en instituciones como el CIAT, NCSU, Duke University y Cenicaña. En esta última institución trabajó como joven investigador en los años 2001 a 2003 y continuó de manera ininterrumpida desde noviembre de 2009 hasta la fecha. Sus intereses en investigación han estado enfocados en el desarrollo de herramientas biotecnológicas para el beneficio de los cultivos, especialmente para procesos de mejoramiento genético. En Cenicaña se ha desempeñado como investigador y líder del grupo de biotecnología, y en los últimos tres años como director del Programa de Variedades.

Biotecnología
para el mejoramiento
de la caña de azúcar

www.cenicana.org