



Publicación del Departamento de
Agronomía de la Universidad Nacional del Sur

agro UNS

- ▶ **Dormición y germinación de Malvaceae nativas ornamentales**
- ▶ **Aprovechamiento de residuos derivados de la producción de aceite de soja a baja escala**
- ▶ **Avispas galícolas que afectan a los eucaliptos en la región del sudoeste bonaerense**
- ▶ **Secuenciación y ensamblado del genoma Pasto llorón**



José Carballo
Ingrid Garbus
Diego Zappacosta
Juan P. Selva
Viviana Echenique

El Ing. Agr. Carballo es becario doctoral de CONICET. La doctora Garbus (investigadora CONICET) es docente del Depto. de Ciencias de la salud, UNS. El doctor Selva (inv. CONICET) es docente del Depto. de Biología, Bioquímica y Farmacia, UNS. Los doctores Zappacosta y Echenique son investigadores de CONICET y docentes del Depto. de Agronomía, UNS.
Contacto: jcarballo@cerzos-conicet.gob.ar

Secuenciación y ensamblado del genoma Pasto llorón

La obtención de la secuencia completa del primer genoma de pasto llorón, permitió individualizar cada uno de los 10 cromosomas que constituyen el número básico de la especie y obtener un catálogo completo de sus genes. Los genes relacionados con calidad forrajera, resistencia a estrés y modo reproductivo podrán ser utilizados en programas de mejoramiento.

El pasto llorón (*Eragrostis curvula*) es una poácea (gramínea) perenne que se desarrolla en áreas con escasas precipitaciones y suelos arenosos. En Argentina, es especialmente atractiva como forraje en regiones semiáridas tales como el sudoeste de la provincia de Buenos Aires, La Pampa y San Luis. A pesar de no tener un elevado valor nutritivo en términos de calidad de su forraje, el pasto llorón es muchas veces la única alternativa en campos de cría.

Otra característica importante del pasto llorón es el modo por el cual se reproduce, denominado apomixis. La apomixis es un modo de reproducción asexual por medio de semillas en el cual la progenie resulta en individuos genéticamente idénticos a la planta que le dio origen, es decir que las semillas son clones maternos. La transferencia por medio de ingeniería genética de esta característica a cultivos con mayor importancia económica ha sido referida como “el santo grial de la agricultura” por Sailer *et al.* (2016). Algunas ventajas que traerían aparejadas esta transferencia son la fijación y estabilización de características deseables por varias generaciones, el acortamiento de los programas de mejoramiento y la disminución de los costos en la producción de semillas híbridas. En especies como

maíz y girasol, los cultivares comerciales se obtienen cruzando una planta hembra con una planta macho a fin de producir una semilla híbrida heterocigota superior. Este fenómeno se denomina vigor híbrido o heterosis y la fijación por medio de apomixis de esta característica traería enormes ventajas económicas ya que los procesos de producción de semillas se acortarían significativamente evitando hacer cruzamientos todos los años. Todo esto redundaría en precios más competitivos para los productores y en mayores beneficios para las empresas semilleras.

Pasto llorón: ploidía y modo reproductivo

El número básico de cromosomas (ploidía) del pasto llorón es 10 (n), por lo tanto, los individuos diploides (2 juegos de cromosomas) tienen 20 cromosomas ($2n=2x=20$) y los tetraploides (4 juegos de cromosomas) 40 ($2n=4x=40$). Si bien los mecanismos que regulan la apomixis no han sido revelados aún, se propone que existe una relación entre la ploidía y el modo reproductivo, siendo sexuales los diploides y apomícticos los poliploides. Hasta el momento, solo fue descubierto un individuo tetraploide con el 100%



de sus sacos reproductivos sexuales mientras que el resto son completamente apomícticos o apomícticos facultativos, es decir, apomícticos con diferentes porcentajes de sexualidad y apomixis. Estas configuraciones de ploidía y modo reproductivo permiten la comparación de individuos muy cercanos genéticamente y así dilucidar cuales son las regiones genómicas y/o genes que están presentes/ausentes y relacionarlas a la forma en la cual se reproducen.

Secuenciado y ensamblado de un genoma

Para poder realizar trabajos de ingeniería genética sobre un organismo es necesario conocer cuáles son los genes involucrados en el proceso de interés y determinar su estructura y secuencia exacta. Se denomina secuenciación del ADN (ácido desoxirribonucleico) a los métodos y técnicas bioquímicas y bioinformáticas cuya finalidad es obtener la secuencia de nucleótidos (bases) que conforman el ADN de una especie, los cuales son ensamblados en forma ordenada por medio de programas. De este modo es posible obtener la colección completa de genes de un organismo dado y su ubicación en los distintos cromosomas de la especie.

El objetivo de la secuenciación de un genoma es obtener el mejor ensamblado del ADN genómico que se encuentra en los cromosomas. Por ejemplo, en el escenario perfecto en pasto llorón deberíamos obtener 10 cromosomas en los cultivares diploides y 20 en los tetraploides. Es interesante destacar que se obtienen la mitad de los cromosomas que componen la especie, ya que en la etapa de ensamblado los cromosomas homólogos son colapsados en un solo cromosoma debido a que la información genética que contienen, en términos de ADN, es muy similar.

En la etapa de secuenciado existen dos factores que afectan considerablemente el resultado final del ensamblado, ellos son la calidad y la longitud de la secuencia. Además, el ensamblado es afectado por la cobertura de secuenciación que se obtiene al dividir la cantidad de nucleótidos secuenciados por la longitud final del genoma ensamblado. Por ejemplo, el genoma diploide de pasto llorón tiene alrededor de 600 millones de pares de bases y se obtuvieron, luego de la secuenciación, aproximadamente 54.000 millones, es decir el equivalente a 90 genomas, lo cual se expresa como 90X.

Una vez ensamblado el genoma por medio de programas bioinformáticos se determina su calidad a tra-

vés de diferentes parámetros: la longitud total ensamblada, el porcentaje del genoma cubierto por esa longitud, la cantidad de secuencias finales obtenidas y el N50, que es definido como la longitud del contig (secuencia de ADN ensamblada) que se encuentra en la mitad de la longitud total ensamblada. La contigüidad es otro parámetro importante que se refiere al orden y a la cantidad de secuencias obtenidas al final del ensamblado. Debido a la alta complejidad de los genomas de las gramíneas, es muy difícil generar ensamblados con alta contigüidad. Existen técnicas post-ensamblado como Hi-C y Chicago que permiten aumentar la contigüidad del ensamblado y obtener secuencias del largo de los cromosomas.

Además de estos parámetros cuantitativos existen otros cualitativos, como los genes BUSCO (Por sus siglas en inglés, genes de referencia ortólogos de copia única). Estos son genes de referencia que se encuentran en todas las especies (ortólogos) en una sola copia y, por lo tanto, permiten determinar si el genoma se encuentra completamente ensamblado de acuerdo al porcentaje de genes BUSCO encontrados.

El genoma de pasto llorón

Nuestro trabajo consistió en la secuenciación del genoma diploide del cultivar Victoria utilizando la tecnología de secuenciación PacBio. El objetivo fue obtener un genoma de referencia de *Eragrostis curvula* que nos permitiera luego ensamblar los genomas más complejos de los poliploides y así poder identificar y caracterizar la/s región/es condicionantes de la apomixis. Luego de la extracción de ADN genómico se obtuvieron dos muestras con dos tamaños de fragmentos de 10.000 y 20.000 bases (Figura 1). Con la primera se buscó aumentar la cobertura, mientras que con la segunda el objetivo fue cubrir amplias regiones repetitivas. Las dos muestras fueron ensambladas utilizando el programa Falcon. Para lograr un ensamblado con fragmentos del tamaño de cromosomas se trabajó integrando este ensamblado por medio del programa Hi-Rise con Dovetail Chicago y Dovetail Hi-C. Con la aplicación de esta metodología fue posible obtener secuencias de pasto llorón del tamaño de los cromosomas. Con este ensamblado se obtuvieron 602 millones de bases distribuidas en 1.143 secuencias, una longitud cercana al tamaño total del genoma. Además, fue posible obtener el 96,4% de los genes BUSCO completos y un N50 de 43 millones de bases, siendo uno de los más largos entre los genomas de las gramíneas secuenciados hasta el momento.

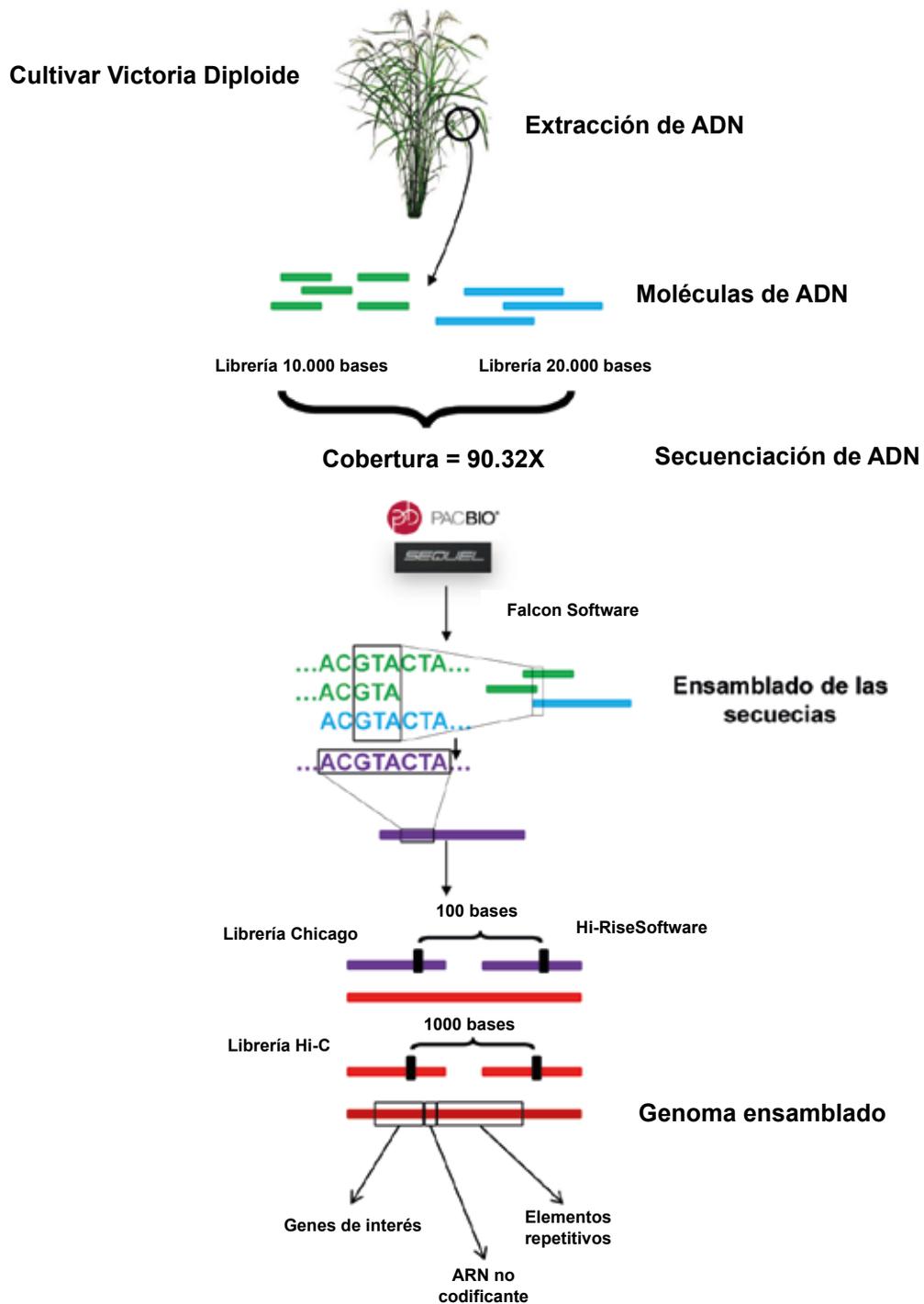


Figura 1. Representación del proceso de obtención de la secuencia genómica del pasto llorón, desde la extracción de ADN, ensamblado y elementos analizados. Cada barra representa una secuencia de ADN.

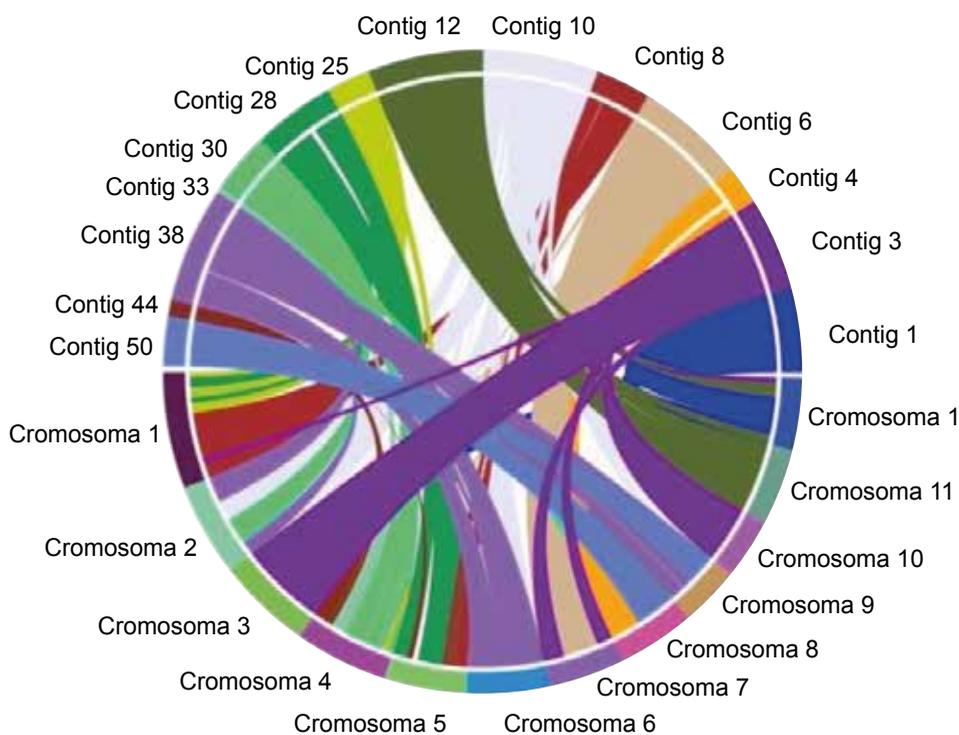


Figura 2. Comparación del genoma de pasto llorón con el genoma de arroz. El genoma de pasto llorón se encuentra en la parte superior, con cada secuencia identificada como un contig. Las franjas que unen los contigs con los cromosomas representan las partes comunes entre los genomas, por ejemplo el contig 3 de pasto llorón cubre completamente el cromosoma 3 de arroz, indicando que el contig 3 corresponde a un cromosoma completo y que la estructura se ha mantenido conservada durante la evolución de las gramíneas.

Qué buscar en un genoma

Los elementos que codifican para las características de los organismos son los genes. La manipulación de estos genes por medio de ingeniería genética permite la transferencia de caracteres deseados entre especies y la regulación diferencial de los mismos. Por este motivo y para conocer más acerca de sus funciones es que los genes son los blancos primarios en todo proyecto de genómica. Además de los genes, también componen los genomas diferentes tipos de estructuras no codificantes como el ARN no codificante y los elementos repetitivos. Estas estructuras codificantes y no codificantes son identificadas a través de análisis comparativos y *de novo* (sin referencia comparativa). A través de este trabajo encontramos en el pasto llorón 56.469 genes y un 28% del genoma cubierto por elementos repetitivos. El análisis de genómica comparativa realizado con el genoma de arroz permitió identificar los cromosomas y las estructuras conservadas entre las especies (Figura 2).

Debido a que el pasto llorón posee gran resistencia al estrés hídrico se buscaron factores de transcripción de tipo WRKY, relacionados con esta característica. Los factores de transcripción pueden asemejarse a “interruptores” que “encienden” o “apagan” determinadas vías metabólicas. Encontramos 74 de genes los cuales pudieron clasificarse de acuerdo a su estructura por el método propuesto por Eulgem *et al.* (2007). En cuanto a calidad de forraje, se identificaron todos los genes correspondientes a la vía de la síntesis de lignina, compuesto que afecta directamente a la digestibilidad del forraje y que es posible manipular por ingeniería genética. Estas dos características, resistencia al estrés hídrico y calidad forrajera, son cualidades interesantes para tener en cuenta en programas de mejoramiento genético de pasto llorón.

Consideraciones finales

El genoma diploide de alta calidad que hemos obtenido, nos posibilitará el ensamblado de genomas de

pasto llorón con mayor nivel de ploidía, portadores de la/s región/es determinantes de la apomixis. De esta manera se podrán hacer análisis comparativos determinando los genes que están vinculados a la sexualidad y a la apomixis. Este trabajo se está llevando a cabo por nuestro grupo de trabajo en colaboración con un grupo del Reino Unido, liderado por el Dr. Mario Cáccamo a través de un PICT categoría Raíces (FONCYT).

Bibliografía

Carballo, J., Santos, B. A. C. M., Zappacosta, D., Garbus, I., Selva, J. P., Gallo, C. A., & Echenique, V. (2019). A high-quality genome of *Eragrostis curvula* grass provides insights into Poaceae evolution and supports new strategies to enhance forage quality. *Scientific reports*, 9(1), 10250.

Eulgem, T. & Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Current Opinion in Plant Biology*, 10, 366-371.

Sailer, C., Schmid, B. & Grossniklaus, U. (2016). Apomixis allows the transgenerational fixation of phenotypes in hybrid plants. *Current Biology*, 26, 331-337.