



Uso de las ciencias ómicas para el mejoramiento genético de cultivos

The omics sciences used for crop improvement programs

Kelly Botero O.¹; Tatiana Arias²

¹ MSc. Centro de Bioinformática y Biología Computacional - BIOS, Manizales, Caldas, Colombia. kelly.botero@gmail.com.

² PhD. Corporación para Investigaciones Biológicas, Biotecnología Agrícola y Ambiental, Medellín, Antioquia, Colombia. tatiana.tatianaarias@gmail.com.

Citar: Botero, K. & Arias, T. (2018). Uso de las ciencias ómicas para el mejoramiento genético de cultivos. *Revista de Ciencias Agrícolas*. 35(2): 64-78 doi:<http://dx.doi.org/10.22267/rcia.183502.92>

Recibido: agosto 02 de 2018.

Aceptado: octubre 10 de 2018.

RESUMEN

El crecimiento de la población, el cambio climático y la pérdida de los servicios ecosistémicos, son algunos de los desafíos que enfrenta el sector agrícola en las últimas décadas para garantizar la seguridad alimentaria a largo plazo. Los programas de mejoramiento genético son un frente de acción que puede contribuir al desarrollo de materiales genéticos adaptados a nuevas condiciones ambientales. El desarrollo y la implementación de tecnologías de secuenciación de alto rendimiento han permitido acelerar dichos programas en cultivos que alimentan a la mayoría de la población mundial. El propósito de este artículo es revisar algunos de los desarrollos tecnológicos de las ciencias ómicas para el estudio de mejoramiento genético de cultivos. En este trabajo se discuten cuatro enfoques de las ciencias ómicas y sus aplicaciones en la agricultura: la filogenómica, la genómica comparativa, la transcriptómica comparativa y la selección asistida por marcadores moleculares. Estos enfoques permiten comprender la historia evolutiva de cultivos y sus parientes silvestres, identificar la estructura y función de los genes de interés en la agricultura, revelar la expresión de los genes importantes en el proceso de domesticación y caracterizar molecularmente individuos de especies agrícolas, con el fin de evidenciar variaciones genéticas que permitan agilizar procesos de selección. Es necesario implementar programas de mejoramiento genético que incluya el uso de algunas, o todas estas tecnologías con el propósito de acelerar los resultados de dichos procesos y contribuir a los desafíos del sector agrícola.

Palabras clave: filogenética, genómica, SAM, transcriptómica.

ABSTRACT

Population growth, climate change and the loss of ecosystem services are some of the challenges facing agriculture in recent decades to ensure long-term food security. Plant breeding programs can contribute to the development of genotypes adapted to new environmental conditions. The development and implementation of high-throughput sequencing technologies have accelerated breeding programs for crops that feed the world's population. This article reviews the main technological developments in "omics" used for genetic crop improvement. Here, we briefly discuss four technologies and their applications in the field of agriculture: phylogenomics, comparative genomics, comparative transcriptomics and marker assisted selection. These approaches allow to understand the evolutionary history of crops and their wild relatives, as well as identify the structure and function of genes of interest in agriculture. In addition, these help to reveal the expression of important genes in the domestication process and characterize individuals of agricultural species using molecular approaches that allow streamlining selection processes. It is necessary to implement genetic improvement programs that include the use of some or all of these technologies with the purpose of accelerating processes that can contribute to current agricultural challenges.

Key words: Phylogenetics, genomics, MAS, Transcriptomics.

INTRODUCCIÓN

La seguridad alimentaria se ve amenazada por el crecimiento de la población, el cambio climático y la pérdida generalizada de servicios ecosistémicos (Rosenzweig *et al.*, 2001; Mooney, 2010). Los desafíos ambientales en la agricultura son enormes y se volverán más relevantes a medida que se traten de asumir nuevas demandas de alimentos a nivel mundial. Se calcula que a mediados de este siglo, la población humana alcanzará los nueve billones de personas (Jackson *et al.*, 2011; 2011; Parker, 2011); por lo tanto, la búsqueda de soluciones para asegurar la alimentación será más urgente que en cualquier otro momento de la historia. Aunque el fitomejoramiento tradicional ha producido mejoras considerables en la producción mundial de alimentos, es poco probable, que sin ayuda de nuevas tecnologías, sea suficiente para enfrentar los desafíos actuales y venideros (Jackson *et al.*, 2011).

En las últimas décadas, la generación de nuevas tecnologías de secuenciación genómica ha transformado la manera como se estudia la diversidad genética. Existe un increíble aumento en el volumen de datos generados y el desarrollo de herramientas bioinformáticas ha sido necesario para procesarlos y analizarlos. El propósito de este artículo, es realizar una aproximación de como los

desarrollos de las ciencias ómicas (filogenómica, la genómica comparativa, la transcriptómica comparativa y la selección asistida por marcadores moleculares) son útiles para el estudio de la historia evolutiva, los genes y la variaciones genéticas de especies de importancia agrícola, y como estos pueden ser aprovechados dentro de programas de mejoramiento genético para obtener variedades con características de interés agronómico de una manera más rápida, que utilizando el mejoramiento tradicional (Figura 1).

Los programas de mejoramiento de cultivos, involucran en sus procesos, un ciclo que consta de la evaluación de fenotipos y genotipos, la selección y cruzamiento de fenotipos superiores, evaluación del fenotipo obtenido y reinicio del ciclo. Este proceso produce genotipos/fenotipos superiores, que son validados y desarrollados en nuevas variedades (Dempewolf *et al.*, 2017). Con el advenimiento de las tecnologías de secuenciación de alto rendimiento (NGS, del inglés *next generation sequencing*), las cuales tienen la capacidad de secuenciar millones de fragmentos de ADN de forma paralela (Buermans y den Dunnen, 2014), las ciencias ómicas han tomado mayor relevancia en estudios que buscan mejorar cultivos agrícolas debido a que pueden acelerar procesos de mejoramiento genético.

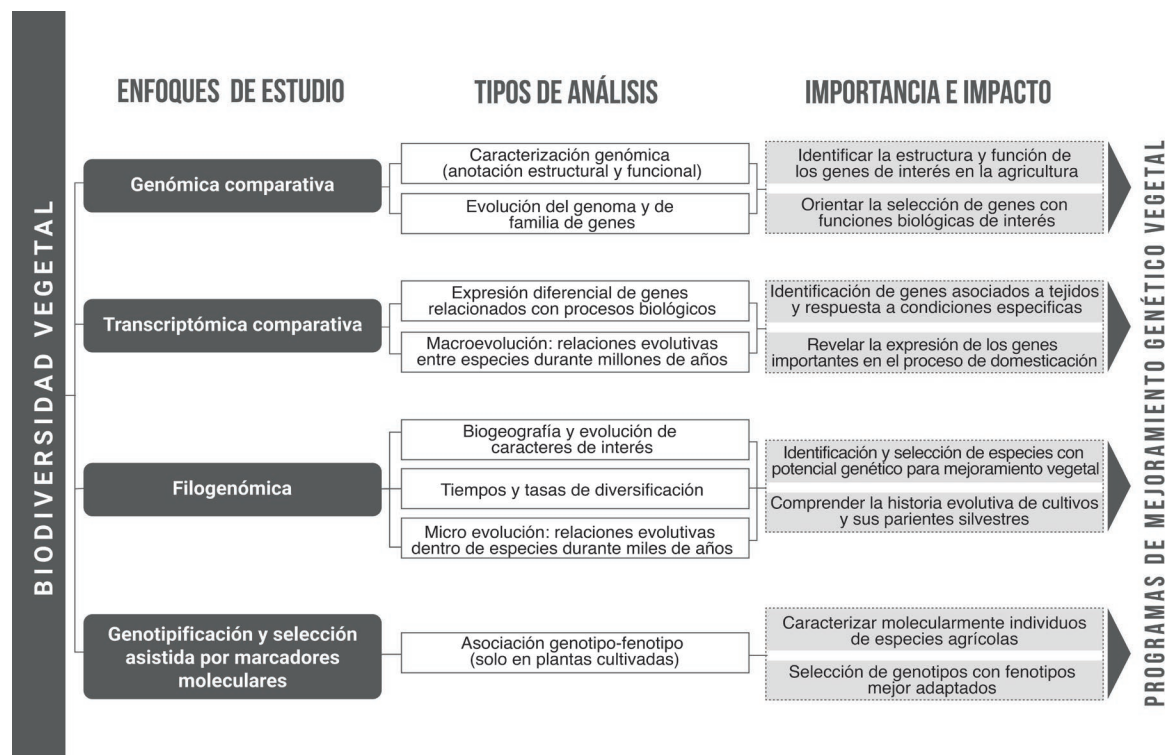


Figura 1. Ciencias ómicas útiles para orientar programas de mejoramiento genético vegetal.

A continuación se mencionan algunas aproximaciones metodológicas a nivel experimental y computacional, que permiten: 1) comprender la historia evolutiva de especies cultivadas y sus parientes silvestres, 2) identificar la estructura y función de los genes presentes en el acervo genético de una especie cultivada, 3) revelar la expresión de los genes de interés agronómico ante determinados eventos inductores y 4) evidenciar variaciones genéticas entre individuos con fenotipos diferentes.

Filogenómica. Para entender el grado y la distribución de la diversidad genética de los recursos vegetales, es necesario utilizar herramientas moleculares y computacionales que permitan reconstruir la historia evolutiva de especies cultivadas, parientes silvestres y otros parientes cercanos (Brozynska *et al.*, 2015; Smýkal *et al.*, 2015). Las filogenias son aproximaciones que ayudan a descifrar las relaciones evolutivas entre especies, estimar tiempos y tasas de diversificación e identificar puntos clave en la evolución de caracteres de interés, tales como, resistencia a patógenos y enfermedades, tolerancia a variables ambientales

desfavorables y aumento de la producción de frutos y semillas, entre otros (Smýkal *et al.*, 2015). Las reconstrucciones filogenéticas permiten entender fuentes de caracteres agronómicos de interés que pueden ser potencialmente incorporados en cultivos desde el acervo genético. Adicionalmente, han permitido entender el origen y la diversificación de los principales cultivos como el maíz, el trigo y el arroz a través del estudio filogenético.

Con el surgimiento de las NGS, los métodos genómicos son cada vez más comunes para reconstruir filogenias. Los más ampliamente utilizados para evaluar una gran cantidad de especies son la hibridación del ADN del cloroplasto o del núcleo con sondas específicas (Wicke y Schneeweiss, 2015) y la secuenciación del ADN total a una cobertura baja, con el fin de recuperar principalmente secuencias de regiones altamente abundantes del genoma, como las del cloroplasto, la mitocondria y el ribosoma (Straub *et al.*, 2012). El enriquecimiento selectivo vía hibridación, es una técnica de secuenciación que utiliza sondas con el propósito de capturar genes de copia única en el genoma de

las plantas (Kato *et al.*, 2002a). Las secuencias de estos genes sirven para reconstruir árboles filogenéticos que reflejan la historia biparental de las especies, que al ser combinada con filogenias uniparentales (reconstruidas con marcadores del cloroplasto o la mitocondria), permite no solo obtener reconstrucciones filogenéticas más confiables e identificar eventos de hibridación, transferencia horizontal de genes, recombinación, entre otros (Brozynska *et al.*, 2015; Wicke y Schneeweiss, 2015), sino también entender los procesos evolutivos y moleculares que han dado origen a la domesticación.

Los métodos bioinformáticos para la reconstrucción de filogenias, se han basado principalmente, en el concepto de homología de secuencias, aunque también se han desarrollado aproximaciones filogenéticas libres de alineamientos (Chan y Ragan, 2013). Existen varios programas de alineamiento de múltiples de secuencias (Notredame *et al.*, 2000; Kato *et al.*, 2002b y Edgar, 2004) para encontrar hipótesis de homología a partir de genes anotados que son utilizados como marcadores filogenéticos (Chan & Ragan, 2013). Este tipo de alineamientos son utilizados por diversos programas para reconstruir árboles filogenéticos a diferentes escalas taxonómicas. Tales programas, implementan métodos basados en parsimonia (Swofford, 2003), máxima verosimilitud (Stamatakis, 2006; Kubatko *et al.*, 2009) y en análisis bayesianos (Huelsenbeck y Ronquist, 2001; Drummond y Rambaut, 2007; Liu, 2008; Larget *et al.*, 2010). Algunos programas para reconstruir árboles filogenéticos, requieren la incorporación de modelos evolutivos que determinan las tasas de sustitución de las secuencias. Estos modelos son generados a priori (Posada, 2008) o posteriori (Venditti *et al.*, 2008).

Los resultados que arrojan los análisis filogenéticos son utilizados para informar programas de mejoramiento genético. Una filogenia, es una representación esquemática de las relaciones evolutivas entre las especies cultivadas y las especies silvestres relacionadas. Las filogenias sirven para enten-

der caracteres agronómicos de interés (e.i. mayor resistencia a patógenos y enfermedades, mayor productividad, mejor tolerancia al estrés) de manera comparativa, permitiendo que eventualmente dichos caracteres puedan ser introgresados desde las especies silvestres a la especie cultivada (Hajjar y Hodgkin, 2007; Martínez-Castillo *et al.*, 2014). Los análisis filogenéticos además, permiten elucidar mecanismos moleculares/biológicos detrás de caracteres de interés, su origen y lapso de tiempo en el que sucedió la domesticación.

Uno de los ejemplos clásicos del uso de filogenias para entender la domesticación e incorporar este conocimiento en mejoramiento, es el maíz. El estudio del acervo genético de este cultivo ha recibido bastante atención, particularmente para entender las relaciones evolutivas entre las especies cercanamente relacionadas y sus variedades silvestres y no domesticadas. El uso de filogenias permitió encontrar los *teosintes*, un complejo de variedades y especies cercanamente relacionadas al maíz domesticado, una de las cuales es el progenitor inmediato del maíz domesticado (Doebley, 2004). A través del uso de filogenias, también se ha inferido, que el maíz evolucionó rápidamente bajo la selección humana, dando lugar a notables transformaciones fenotípicas (Ramos-Madrugal *et al.*, 2016). El uso de la biología comparativa ha permitido identificar las bases moleculares de diferencias morfológicas entre el maíz y *teosinte*, los cuales, han orientado a investigaciones posteriores en genes específicos de *teosinte* que puedan contribuir al mejoramiento del maíz (Hoisington *et al.*, 1999; Doebley, 2004; Ramos-Madrugal *et al.*, 2016).

Genómica comparativa. El estudio de la genómica comparativa se centra en el análisis de las similitudes y diferencias entre genomas para inferir estructura y función de los genes, procesos de duplicación del genoma y evolución de familias de genes, de elementos repetitivos y del tamaño del genoma (Schmutz *et al.*, 2014; Tang *et al.*, 2015). En programas de mejoramiento vegetal, éste tipo de estudios, facilita la identificación de recursos genómicos útiles para procesos de selección de

caracteres de interés, dado que permite identificar funciones biológicas, comprender la variabilidad intra e interespecífica y entender elementos regulatorios (Bonham *et al.*, 2013; Borozan *et al.*, 2015).

La genómica comparativa implica el análisis de genomas previamente ensamblados y anotados, por lo tanto, una de las tareas más importantes en la bioinformática ha sido el desarrollo de herramientas de ensamblaje y anotación que se ajusten a las diversas características genómicas de los organismos y a los variados propósitos de estudio. En cuanto al proceso de ensamblaje de secuencias, algunas herramientas superan claramente a otras, pero es difícil predecir cuál podría ser la más apropiada en una situación dada. Existen una gran variedad de ensambladores (*de novo* y de referencia), pero la selección de uno en particular dependerá de las características de cada genoma, tales como el tamaño, la composición de bases, el contenido de elementos repetitivos y el nivel de polimorfismo (Ekblom y Wolf, 2014; Simpson y Pop, 2015).

Para el proceso de anotación de genes se han desarrollado múltiples métodos computacionales que incluyen, tanto algoritmos para realizar predicción de genes, como para inferir funciones biológicas a través de homologías de secuencias. Para la predicción de genes, comúnmente se utilizan métodos *ab initio* que permiten predecir genes basados en características presentes en las secuencias genómicas que pueden ser señales genéticas de longitud fija (codones de inicio y de parada) y/o de longitud variable (regiones codificantes y no codificantes) (Majoros *et al.*, 2004; Stanke *et al.*, 2004; Lomsdize *et al.*, 2005). Para inferir las funciones biológicas de los genes, se han desarrollado métodos heurísticos que generan soluciones sub-óptimas de alineamientos de secuencias (Altschul *et al.*, 1990).

En términos generales, estos métodos utilizan ciertas métricas para determinar la calidad del alineamiento, tales como, el porcentaje de similitud, que es la fracción de elementos iguales alineados entre las secuencias, el *e-value*, que determina la significancia estadística del alineamiento, el *score*

que representa la calidad del alineamiento, entre otros (Altschul *et al.*, 1990). Aunque, estos métodos no garantizan encontrar el mejor alineamiento, han sido optimizados para encontrar soluciones casi óptimas, con las cuales se pueden hacer inferencias de homología estadísticamente significativas y de esta forma, asignar funciones a los genes (Stormo, 2009).

Los estudios de genómica comparativa en plantas, han girado en torno a especies que son cultivadas para la alimentación (The Potato Genome Sequencing Consortium, 2011; Zhu *et al.*, 2005; Wang *et al.*, 2011), donde se identifican genes candidatos que están relacionados con características de interés agronómico. Explorar la variación de estos genes entre individuos y especies es de gran importancia para realizar gestiones eficientes en la agricultura (Bolger *et al.*, 2014), a través del mejoramiento genético de cultivos. Este tipo de análisis permiten identificar recursos genómicos como los marcadores moleculares, que una vez han sido asociados a una característica de interés, permiten la selección de fenotipos deseados (Varshney *et al.*, 2009) y facilitan el estudio de procesos de domesticación.

En programas de mejoramiento vegetal, es útil hacer distinción entre las bases genéticas asociadas a cambios durante la domesticación y otros rasgos que están sujetos al mejoramiento de variedades de cultivos siguientes a la domesticación inicial. Los cambios que ocurrieron durante la domesticación inicial, generalmente, se fijan dentro de las especies de cultivo y los fenotipos asociados a estos cambios son los que hacen que el cultivo sea fenológicamente congruente con prácticas de cultivo y que cuente con características importantes para procesos productivos. Por el contrario, los rasgos sujetos a mejoramiento, son típicamente variables entre poblaciones o cultivares. Estos incluyen entre muchos otros fenotipos, los relacionados con adaptación a climas específicos, contenido de almidón, pigmentación de los frutos y morfología de los frutos (Olsen y Wendel, 2013).

En maíz, la resecuenciación del genoma de plantas silvestres, landraces y líneas mejoradas permitió entender la domesticación del maíz moderno, donde se evidencia que existe una recuperación de la diversidad después de la domesticación, probablemente por introgresión genética de parientes silvestres (Hufford *et al.*, 2012). Lo que sugiere que la diversidad genética de plantas silvestres y criollas puede ser aprovechada en programas de mejoramiento de cultivares modernos. Por su parte, en arroz, la comparación de los genomas de variedades domesticadas y silvestres ha permitido identificar regiones del genoma que muestran marcas de selección y genes candidatos involucrados en procesos de domesticación, con los cuales se pueden hacer mejoramiento genético (He *et al.*, 2011; Xu *et al.*, 2011).

Transcriptómica comparativa. La transcriptómica es un área esencial para interpretar el funcionamiento del proceso transcripcional de los genes, el cual depende de estímulos intra o extracelulares que desencadenan cascadas de señalización en la que genes se expresan o reprimen de acuerdo con el tipo de estímulo inicial. Para el estudio de los transcriptomas, también llamados estudios funcionales del genoma, la tecnología RNA-seq permite obtener la secuencia de los transcritos, cuantificar el nivel de expresión génica de manera precisa y reproducible (Nookaew *et al.*, 2012; Sims *et al.*, 2014) y realizar análisis de empalme alternativo (Wang *et al.*, 2009; Sims *et al.*, 2014).

La transcriptómica comparativa en plantas, se ha utilizado para: 1) comparar la expresión génica entre especies domesticadas y silvestres, e identificar caracteres potencialmente valiosos en el mejoramiento de cultivos (Koenig *et al.*, 2013), 2) detectar cambios en los niveles de expresión génica bajo diferentes condiciones biológicas y/o tejidos celulares, para entender como los factores ambientales influyen en los estados de desarrollo y las funciones celulares (Mutz *et al.*, 2013) y 3) identificar y cuantificar transcritos nuevos y conocidos, relacionados con procesos bioquímicos como la fotosíntesis (Xu *et al.*, 2013), la asimilación

de nitrógeno (Gao *et al.*, 2014), la síntesis de metabolitos de interés productivo y/o comercial, entre muchos otros (Han *et al.*, 2013).

Un prerequisite para los análisis RNA-seq, es que los datos generados tengan el potencial de responder preguntas biológicas de interés. Para esto, es determinante hacer un buen diseño experimental que responda la pregunta de investigación con confiabilidad (Conesa *et al.*, 2016). Si el diseño no se realiza correctamente, los análisis pueden conducir a conclusiones erradas (Wang *et al.*, 2009; Mutz *et al.*, 2013). En el diseño experimental se recomienda incluir réplicas biológicas de los tratamientos (Schurch *et al.*, 2016) y procurar alta profundidad de secuenciación para disminuir la tasa de error de cada nucleótido en un ensamblaje de secuencias. En proyectos donde el presupuesto es limitado, es aconsejable elegir un mayor número de réplicas sobre una alta profundidad de secuenciación (Sims *et al.*, 2014).

Aunque no existe un sólo flujo de análisis bioinformático para todas las investigaciones basadas en RNA-seq, a grandes rasgos, el flujo de trabajo de los análisis RNA-seq consiste en cuatro etapas principales: 1) preprocesamiento de los datos crudos, 2) alineamiento de lecturas para ensamblaje y anotación, 3) cuantificación de la expresión génica y 4) análisis de expresión diferencial (Mutz *et al.*, 2013; Yang y Kim, 2015). Existen diversas herramientas bioinformáticas para cada una de las etapas de análisis (Love *et al.*, 2014; Patro *et al.*, 2014; Zhou *et al.*, 2014; Bray *et al.*, 2016;) que al ser combinadas permiten obtener un sistema apropiado de análisis, según los propósitos de cada investigación. Además, se han desarrollado herramientas capaces de realizar análisis de transcriptomas con (Trapnell *et al.*, 2009) o sin genoma de referencia (Haas *et al.*, 2013; Xie *et al.*, 2014).

Aproximaciones comparativas en la transcriptómica posibilitan reconstruir perfiles de expresión génica utilizados para comparar individuos de la misma especie, que presentan características contrastantes, como tejidos, estados de desarrollo y/o

condiciones particulares de interés. Sí se realiza un diseño experimental adecuado, las diferencias encontradas entre tratamientos en un análisis de RNA-seq, podrán detectar genes implicados en características o propiedades de interés. Algunos enfoques en el ámbito vegetal se han centrado en características fisiológicas y del desarrollo, como caracteres relacionados con la morfología de la planta y con las diferencias entre los sistemas de defensa y los mecanismos de producción de metabolitos primarios (Van de Mortel *et al.*, 2006).

Para los programas de mejoramiento vegetal, la principal importancia de los estudios transcriptómicos, consiste en la capacidad de identificar conjuntos de genes candidatos o loci claves que controlan respuestas biológicas de las plantas ante determinadas circunstancias y condiciones (Kudapa *et al.*, 2013). El conocimiento de la expresión de los genes también, puede ser utilizado para realizar modificaciones genéticas y mejoramiento variedades de cultivos (Valliyodan y Nguyen, 2006). Esto debido a que los análisis de transcritos son una herramienta que ayudan a validar la función de los genes (Kudapa *et al.*, 2013) y permite realizar selección de genotipos que responden de manera deseada a diversos factores ambientales. En cultivos de leguminosas, que son fuertemente impactados por diferentes factores de estrés, a través de estudios funcionales, se han descubierto genes relacionados a resistencia de estrés biótico y tolerancia a estrés abiótico. Esto ha ayudado a entender los mecanismos de respuesta de estas plantas al ambiente y seleccionar un conjunto de genes candidatos que pueden ser marcadores genéticos útiles en estrategias para obtener variedades con mejor tolerancia al estrés y mayor productividad ante condiciones adversas (Kudapa *et al.*, 2013).

Genotipificación y selección asistida por marcadores moleculares. Uno de los principales objetivos de la genómica funcional en la agricultura, es asociar características genéticas a características fenotípicas complejas para orientar programas de fitomejoramiento (Tester y Langridge, 2010). Generalmente, el proceso

se basa en el descubrimiento de marcadores moleculares, fenotipificación de rasgos de individuos y poblaciones, identificación de asociaciones genotipo – fenotipo y selección o creación de genotipos con rasgos agronómicos de interés, tales como, estabilidad, adaptabilidad y producción, entre otros (He *et al.*, 2014).

Debido al acelerado avance de las NGS, en los últimos años se han generado nuevos métodos para el descubrimiento, validación y evaluación de marcadores genéticos, que son útiles para caracterizar genéticamente poblaciones de interés, hacer cartografía genética, orientar programas de mejoramiento genético y seleccionar líneas isogénicas y/o de introgresión (Elshire *et al.*, 2011; Poland y Rife, 2012). Aquí nos enfocamos en el método de genotipificación por secuenciación (GBS, del inglés *genotyping by sequencing*), que es ampliamente utilizado para genotipificación en masa. Este método permite identificar miles de marcadores SNPs (polimorfismos de nucleótido simple) para revelar variaciones entre cultivares, y buscar marcas de selección relacionadas con domesticación y procesos de adaptación de las especies (Elshire *et al.*, 2011; Soto *et al.*, 2015).

El GBS se basa en reducir la complejidad del genoma, utilizando enzimas de restricción (ERs) que reconocen y cortan regiones específicas del genoma (Davey *et al.*, 2011; Elshire *et al.*, 2011). Este método es altamente reproducible y si se escogen las ERs de forma adecuada, se puede alcanzar regiones metiladas del genoma que son inaccesibles con otras aproximaciones, además permite detectar regiones codificantes y regiones de regulación génica (Elshire *et al.*, 2011) que frecuentemente son responsables de fenotipos agronómicamente importantes (Clark *et al.*, 2006).

Los análisis computacionales de los datos generados por GBS pueden ser *de novo* (Catchen *et al.*, 2011) o mediante el mapeo de las secuencias obtenidas a genomas de referencia (Glaubitz *et al.*, 2014). Ambos análisis permiten el descubrimiento de un amplio número SNPs. Torkamaneh *et al.*

(2016) realizaron una revisión detallada sobre los programas bioinformáticos que se pueden emplear para datos provenientes de GBS, de acuerdo con la especie de estudio, a la existencia de un genoma de referencia, a especificaciones técnicas, y a las ventajas y desventajas de utilizar diferentes metodologías (Torkamaneh *et al.*, 2016).

En importantes especies cultivadas como el arroz (Spindel *et al.*, 2013), la frambuesa (Ward *et al.*, 2013), el algodón (Gore *et al.*, 2014) y la yuca (Soto *et al.*, 2015) se han utilizado los SNPs derivados de GBS para crear mapas genéticos de alta resolución, utilizados para la construcción de mapas físicos, con los cuales se orienta la clonación de genes en plantas modelo o cultivadas (QTL, del inglés *quantitative trait locus*) (Poland y Rife, 2012; Soto *et al.*, 2015). Frecuentemente, en los programas de mejoramiento de cultivos, los mapeos genéticos y físicos, son útiles para los estudios de asociación genotipo-fenotipo, considerando que permiten asociar rasgos de interés con funciones génicas (Bakker *et al.*, 2011; Swamy *et al.*, 2011; Whankaew *et al.*, 2011).

Estos estudios de asociación utilizan metodologías enfocadas a capturar una porción sustancial de la variación genética de una especie. Dentro de estas metodologías, se tienen los estudios de asociación de genoma completo (GWAS, del inglés *Genome Wide Association Study*). El enfoque principal de los GWAS es evaluar a escala genómica las asociaciones entre variantes genéticas y rasgos fenotípicos de interés (He *et al.*, 2014; Pearson y Manolio, 2008). El estudio típico de GWAS consta de cuatro pasos: 1) selección de un gran número de individuos con el rasgo de interés y de un grupo de comparación adecuado; 2) genotipificación a través de técnicas como GBS para recuperar SNPs a lo largo de todo el genoma; 3) pruebas estadísticas para asociaciones entre los SNPs y el rasgo; y 4) replicación de asociaciones identificadas en una muestra de una población o examen de implicaciones funcionales a nivel experimental (Pearson y Manolio, 2008). GWAS se aplica de forma rutinaria en plantas modelo como *Arabidopsis*

(Atwell *et al.*, 2010) y cultivos como arroz (Huang *et al.*, 2011), tomate (Ranc *et al.*, 2012) y cebada (Wang *et al.*, 2012).

El último paso requerido en programas de mejoramiento es la selección asistida por marcadores moleculares (SAM). Esta estrategia permite realizar una selección dirigida a una característica de interés, de manera acelerada y sin necesariamente esperar al desarrollo completo de la planta, lo cual es ventajoso, si se trata de organismos con ciclos de vida largos. El método de selección genómica asistida por marcadores, combina marcadores moleculares, tales como los SNPs, a lo largo de todo el genoma con datos fenotípicos y de pedigrí para aumentar la precisión de reproducción en los valores genotípicos. Las precisiones en predicción por este método, junto con los continuos avances en las tecnologías para genotipificación de alto rendimiento, hacen que esta estrategia sea una herramienta prometedora para aumentar la eficiencia en la fitomejoramiento (Heffner *et al.*, 2010).

CONCLUSIONES

La caracterización molecular del material genético de especies cultivadas y su acervo genético, es el primer paso para iniciar programas de mejoramiento vegetal. Las técnicas de secuenciación genómica y el desarrollo de herramientas bioinformáticas han permitido acelerar los programas de mejoramiento de especies con interés agronómico y representan oportunidades para conocer, entender y clasificar más profunda y detalladamente, los recursos fitogenéticos que pueden ser incorporados en programas de mejoramiento de cultivos.

Aquí se discutieron cuatro enfoques y sus aplicaciones en el campo de la agricultura que pueden apoyar y acelerar la toma de decisiones para el mejoramiento de cultivos: 1) la filogenómica para entender las relaciones evolutivas de especies importantes en la agricultura y sus parientes silvestres, además de los procesos, el origen y el tiempo de la domesticación de las especies,

2) la genómica y transcriptómica comparativa que permiten la identificación de genes con funciones de interés agronómico para mejorar la productividad de un cultivo y la identificación de genes involucrados en la respuesta a diversas condiciones ambientales para orientar planes de mejoramiento genético de cultivos mejor adaptados a las condiciones climáticas actuales y futuras del planeta, por último 4) los estudios de genotipificación de especies y su asociación con el fenotipo para gestionar marcadores genéticos que permitan seleccionar más rápidamente individuos de una especie con genotipos ventajosos, acelerando así los programas de mejoramiento.

Para los programas de mejoramiento genético se recomienda la combinación de las diferentes ómicas expuestas, las cuales al ser complementarias permiten tomar decisiones más precisas en los procesos de selección de características de interés agronómico, donde es indispensable la evaluación holística de los factores que influyen en el resultado del mejoramiento de una especie. Además, los enfoques presentados permiten acelerar y optimizar la productividad y resistencia a factores adversos bióticos y abióticos de muchos cultivos que alimentan al mundo.

AGRADECIMIENTOS

Queremos agradecer a Diego Mauricio Riaño Pachón, Jorge Duitama, Roxana Yockten y Adriana Tofiño por la revisión y comentarios a los borradores del manuscrito.

Conflicto de intereses: Los autores declaran que no hay conflicto de interés.

REFERENCIAS BIBLIOGRÁFICAS

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. 215(3): 403-410.
- Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., Jiang, R., Mulyati, N. W., Zhang, X., Amer, M. A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., Ecker, J. R., Faure, N., Kniskern, J. M., Jones, J. D., Michael, T., Nemri, A., Roux, F., Salt, D. E., Tang, C., Todesco, M., Traw, M.B., Weigel, D., Marjoram, P., Borevitz, J.O., Bergelson, J. & Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*. 465(7298): 627-631. doi: <https://10.1038/nature08800>
- Bakker, E., Borm, T., Prins, P., van der Vossen, E., Uenk, G., Arens, M., Boer, J., Eck, H., Muskens, M., Vossen, J., Linden, G., Ham, R., Klein-Lankhorst, R., Visser, R., Smant, G., Bakker, J. & Goverse, A. (2011). A genome-wide genetic map of NB-LRR disease resistance loci in potato. *Theoretical and Applied Genetics*. 123(3): 493-508. doi: 10.1007/s00122-011-1602-z
- Bolger, M. E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B. & Mayer, K. F. (2014). Plant genome sequencing - applications for crop improvement. *Current Opinion in Biotechnology*. 26: 31-37. doi: 10.1016/j.copbio.2013.08.019
- Bonham-Carter, O., Steele, J. & Bastola, D. (2013). Alignment-free genetic sequence comparisons: a review of recent approaches by word analysis. *Briefings in Bioinformatics*. 15(6): 890-905. doi: <https://10.1093/bib/bbt052>
- Borozan, I., Watt, S. & Ferretti, V. (2015). Integrating alignment-based and alignment-free sequence similarity measures for biological sequence classification. *Bioinformatics*. 31(9): 1396-1404. doi: <https://10.1093/bioinformatics/btv006>
- Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 34(5): 525-527. doi: <https://10.1038/nbt.3519>
- Brozynska, M., Furtado, A. & Henry, R. J. (2015). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnology Journal*. 14(4): 1070-1085. doi: <https://10.1111/pbi.12454>

- Buermans, H. P. J. & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*. 1842(10): 1932-1941. doi: <https://10.1016/J.BBA-DIS.2014.06.015>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. (2011). Stacks: building and genotyping Loci de novo from short-read sequences. *G3: Genes, genomes, genetics*. 1(3): 171-182. doi: <https://10.1534/g3.111.000240>
- Chan, C. X. & Ragan, M. A. (2013). Next-generation phylogenomics. *Biology Direct*. 8(3):1-6. doi: 10.1186/1745-6150-8-3
- Clark, R. M., Wagler, T. N., Quijada, P. & Doebley, J. (2006). A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genetics*. 38(5): 594-597. doi: 10.1038/ng1784
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Wojciech, M., Gaffney, D., Elo, L., Zhang, X. & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*. 17(1): 13. doi: 10.1186/s13059-016-0881-8
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. 12: 499-510. doi: <https://doi.org/10.1038/nrg3012>
- Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C. & Guarino, L. (2017). Past and Future Use of Wild Relatives in Crop Breeding. *Crop Science*. 57(3): 1070-1082. doi: 10.2135/cropsci2016.10.0885
- Doebley, J. (2004). The genetics of maize evolution. *Annual review of genetics*. 38:37-59. doi: 10.1146/annurev.genet.38.072902.092425
- Drummond, A. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. 7(1):214. doi: 10.1186/1471-2148-7-214
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32(5): 1792-1797. doi: 10.1093/nar/gkh340
- Eklom, R. & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*. 7(9): 1026-1042. doi: 10.1111/eva.12178
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*. 6(5): e19379. doi: 10.1371/journal.pone.0019379
- Gao, J., Zhang, Y., Zhang, C., Qi, F., Li, X., Mu, S. & Peng, Z. (2014). Characterization of the Floral Transcriptome of Moso Bamboo (*Phyllostachys edulis*) at Different Flowering Developmental Stages by Transcriptome Sequencing and RNA-Seq Analysis. *PLoS ONE*. 9(6): e98910. doi: 10.1371/journal.pone.0098910
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*. 9(2): e90346. doi: 10.1371/journal.pone.0090346
- Gore, M. A., Fang, D. D., Poland, J. A., Zhang, J., Percy, R. G., Cantrell, R. G., Thyssen, G. & Lipka, A. E. (2014). Linkage Map Construction and Quantitative Trait Locus Analysis of Agronomic and Fiber Quality Traits in Cotton. *The Plant Genome*. 7(1):1-10. doi: 10.3835/plantgenome2013.07.0023
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R., Friedman, N. & Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*. 8(8): 1494-1512. doi: 10.1038/nprot.2013.084

- Hajjar, R. & Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica*. 156(1-2): 1-13. doi: 10.1007/s10681-007-9363-0
- Han, X.-J., Wang, Y.-D., Chen, Y.-C., Lin, L.-Y., & Wu, Q.-K. (2013). Transcriptome Sequencing and Expression Analysis of Terpenoid Biosynthesis Genes in *Litsea cubeba*. *PLoS ONE*. 8(10): e76890. doi: 10.1371/journal.pone.0076890
- He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H. & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*. 30(5): 484. doi: 10.3389/fpls.2014.00484
- He, Z., Zhai, W., Wen, H., Tang, T., Wang, Y., Lu, X., Greenberg, A. J., Hudson, R. R., Wu, C. I. & Shi, S. (2011). Two Evolutionary Histories in the Genome of Rice: the Roles of Domestication Genes. *PLoS Genetics*. 7(6): e1002100. doi: 10.1371/journal.pgen.1002100
- Heffner, E. L., Lorenz, A. J., Jannink, J. L., & Sorrells, M. E. (2010). Plant breeding with Genomic selection: Gain per unit time and cost. *Crop Science*. 50(5): 1681-1690. doi: 10.2135/cropsci2009.11.0662
- Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J. M., Skovmand, B., Taba, S., & Warburton, M. (1999). Plant genetic resources: what can they contribute toward increased crop productivity? *Proceedings of the National Academy of Sciences of the United States of America*. 96(11): 5937-5943. doi: 10.1073/PNAS.96.11.5937
- Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li W, Guo, Y., Deng, L., Zhu, C., Fan, D., Lu, Y., Weng, Q., Liu, K., Zhou, T., Jing, Y., Si, L., Dong, G., Huang, T., Lu, T., Feng, Q., Qian, Q., Li, J. & Han, B. (2011). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics*. 44(1): 32-39. doi: 10.1038/ng.1018
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17(8): 754-755. doi: https://doi.org/10.1093/bioinformatics/17.8.754
- Hufford, M. B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R. A., Elshire, R. J., Glaubitz, J. C., Guill, K. E., Kaeppler, S. M., Lai, J., Morrell, P. L., Shannon, L. M., Song, C., Springer, N. M., Swanson-Wagner, R. A., Tiffin, P., Wang, J., Zhang, G., Doebley, J., McMullen, M. D., Ware, D., Buckler, E.S., Yang, S. & Jeffrey, Ross-Ibarra (2012). Comparative population genomics of maize domestication and improvement. *Nature Genetics*. 44(7): 808-811. doi: 10.1038/ng.2309
- Jackson, S. A., Iwata, A., Lee, S.-H., Schmutz, J., & Shoemaker, R. (2011). Sequencing crop genomes: approaches and applications. *New Phytologist*. 191(4): 915-925. doi: 10.1111/j.1469-8137.2011.03804.x
- Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002a). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 30(14): 3059-3066.
- Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002b). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 30(14): 3059-3066.
- Koenig, D., Jimenez-Gomez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., Kumar, R., Covington, M. F., Devisetty, U.K., Tat, A.V., Tohge, T., Bolger, A., Schneeberger, K., Ossowski, S., Lanz, C., Xiong, G., Taylor-Teeple, M., Brady, S. M., Pauly, M., Weigel, D., Usadel, B., Fernie, A. R., Peng, J., Sinha, N.R. & Maloof, J. N. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proceedings of the National Academy of Sciences*. 110(28): E2655-E2662. doi: 10.1073/pnas.1309606110
- Kubatko, L. S., Carstens, B. C., & Knowles, L. L. (2009). STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*. 25(7): 971-973. doi: https://doi.org/10.1093/bioinformatics/btp079

- Kudapa, H., Ramalingam, A., Nayakoti, S., Chen, X., Zhuang, W.-J., Liang, X., Kahl, G., Edwards, D. & Varshney, R. K. (2013). Functional genomics to study stress responses in crop legumes: progress and prospects. *Functional Plant Biology*. 40(12):1221. doi: <https://10.1071/FP13191>
- Larget, B. R., Kotha, S. K., Dewey, C. N. & Ané, C. (2010). BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Phylogenetics*. 26(22): 2910-2911.
- Liu, L. (2008). BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*. 24(21): 2542-2543. doi: [10.1093/bioinformatics/btn484](https://doi.org/10.1093/bioinformatics/btn484)
- Lomsadze, A., Ter-hovhannisyan, V., Chernoff, Y. O. & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*. 33(20): 6494-6506. doi: [10.1093/nar/gki937](https://doi.org/10.1093/nar/gki937)
- Love, M. I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15(12): 550. doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8)
- Majoros, W., Pertea, M., & Salzberg, S. (2004). TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. *Bioinformatics*. 20(16): 2878-2879.
- Martinez-Castillo, J., Camacho-Perez, L., Villanueva-Viramontes, S., Andueza-Noh, R. H., & Chacon-Sanchez, M. I. (2014). Genetic structure within the Mesoamerican gene pool of wild *Phaseolus lunatus* (Fabaceae) from Mexico as revealed by microsatellite markers: Implications for conservation and the domestication of the species. *American Journal of Botany*. 101(5): 851-864. doi: [10.3732/ajb.1300412](https://doi.org/10.3732/ajb.1300412)
- Mooney, H. A. (2010). The ecosystem-service chain and the biological diversity crisis. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 365(1537): 31-39. doi: [10.1098/rstb.2009.0223](https://doi.org/10.1098/rstb.2009.0223)
- Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.G. & Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology*. 24(1): 22-30. doi: [10.1016/j.copbio.2012.09.004](https://doi.org/10.1016/j.copbio.2012.09.004)
- Nookaew, I., Papini, M., Pornputtapong, N., Scalcinati, G., Fagerberg, L., Uhlén, M. & Nielsen, J. (2012). A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Research*. 40(20): 10084-10097. doi: [10.1093/nar/gks804](https://doi.org/10.1093/nar/gks804)
- Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. Edited by J. Thornton. *Journal of Molecular Biology*. 302(1): 205-217. doi: [10.1006/jmbi.2000.4042](https://doi.org/10.1006/jmbi.2000.4042)
- Olsen, K. M. & Wendel, J. F. (2013). A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. *Annual Review of Plant Biology*. 64(1): 47-70. doi: [10.1146/annurev-arplant-050312-120048](https://doi.org/10.1146/annurev-arplant-050312-120048)
- Parker, J. (2011). The 9 billion-people question. Recuperada de <http://www.economist.com/node/18200618>
- Patro, R., Mount, S. M. & Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnology*. 32(5): 462-464. doi: [10.1038/nbt.2862](https://doi.org/10.1038/nbt.2862)
- Pearson, T. A. & Manolio, T. A. (2008). How to Interpret a Genome-wide Association Study. *JAMA*. 299(11):1335-1344. doi: [10.1001/jama.299.11.1335](https://doi.org/10.1001/jama.299.11.1335)
- Poland, J. A. & Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*. 5(5): 425-431. doi: [10.3835/plantgenome2012.05.0005](https://doi.org/10.3835/plantgenome2012.05.0005)
- Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*. 25(7): 1253-1256. doi: [10.1093/molbev/msn083](https://doi.org/10.1093/molbev/msn083)

- Ramos-Madrigal, J., Smith, B.D., Moreno-Mayar, J.V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M.T.P. & Wales, N. (2016). Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Current Biology*. 26(23):3195-3201. doi: 10.1016/j.cub.2016.09.036.
- Ranc, N., Muñoz, S., Xu, J., Le Paslier, M.-C., Chauveau, A., Bounon, R., Rolland, S., Bouchet, J. P., Brunel, D. & Causse, M. (2012). Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. *G3: Genes, Genomes, Genetics*. 2(8): 853-864. doi: 10.1534/g3.112.002667
- Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R. & Chivian, E. (2001). Climate change and extreme weather events - Implications for food production, plant diseases, and pests. *Global change & human health, nasa publications*. 2(2): 90-104.
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M.A., Chovatia, M., Gepts, P., Goodstein, D. M., Gonzales, M., Hellsten, U., Hyten, D. L., Jia, G., Kelly, J. D., Kudrna, D., Lee, R., Richard, M. M., Miklas, P. N., Osorno, J. M., Rodrigues, J., Thareau, V., Urrea, C. A., Wang, M., Yu, Y., Zhang, M., Wing, R. A., Cregan, P. B., Rokhsar, D. S. & Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*. 46(7): 707-713. doi: 10.1038/ng.3008
- Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M. & Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA Society*. 22(6): 839-851. doi: 10.1261/rna.053959.115
- Simpson, J. T. & Pop, M. (2015). The Theory and Practice of Genome Sequence Assembly. *Annual Review of Genomics and Human Genetics*. 16(1): 153-172. doi: 10.1146/annurev-genom-090314-050032
- Sims, D., Sudbery, I., Illott, N. E., Heger, A. & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*. 15(2): 121-132. doi: 10.1038/nrg3642
- Smýkal, P., Coyne, C. J., Ambrose, M. J., Maxted, N., Schaefer, H., Blair, M. W., Berger, J., Greene, S. L., Nelson, M. N., Besharat, N., Vymyslický, T., Toker, C., Saxena, R. K., Roorkiwal, M., Pandey, M. K., Hu, J., Li, Y. H., Wang, L. X., Guo, Y., Qiu, L. J., Redden, R.J. & Varshney, R. K. (2015). Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. *Critical Reviews in Plant Sciences*. 34(1-3): 43-104. doi: <https://doi.org/10.1080/07352689.2014.897904>
- Soto, J. C., Ortiz, J. F., Perlaza-Jiménez, L., Vásquez, A. X., Lopez-Lavalle, L. A. B., Mathew, B., León, J., Bernal, A. J., Ballvora, A. & López, C. E. (2015). A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. *BMC Genomics*. 16(1): 190. doi: 10.1186/s12864-015-1397-4
- Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., Lorieux, M., Ahmadi, N. & McCouch, S. (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theoretical and Applied Genetics*. 126(11): 2699-2716. doi: 10.1007/s00122-013-2166-x
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22(21): 2688-2690. doi: 10.1093/bioinformatics/btl446
- Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*. 32(2): W309-W312. doi: <https://doi.org/10.1093/nar/gkh379>
- Stormo, G. D. (2009). An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*. 27(1): 3.1.1-3.1.7 doi: <https://doi.org/10.1002/0471250953.bi0301s27>.

- Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*. 99(2): 349-364. doi: <https://doi.org/10.3732/ajb.1100335>
- Swamy, B. M., Vikram, P., Dixit, S., Ahmed, H., & Kumar, A. (2011). Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. *BMC Genomics*. 12(1): 319. doi: <https://doi.org/10.1186/1471-2164-12-319>
- Swofford, D. L. (2003). Phylogenetic analysis using parsimony. Recuperada de <https://paup.phylosolutions.com/>
- Tang, H., Bomhoff, M. D., Briones, E., Zhang, L., Schnable, J. C., & Lyons, E. (2015). SynFind: Compiling Syntenic Regions across Any Set of Genomes on Demand. *Genome Biology and Evolution*. 7(12): 3286-3298. doi: <https://doi.org/10.1093/gbe/evv219>
- Tester, M., & Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. *Science*. 327(5967): 818-822. doi: <https://doi.org/10.1126/science.1183700>
- The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature*. 475(7355): 189-195. doi: <https://doi.org/10.1038/nature10158>
- Torkamaneh, D., Laroche, J., & Belzile, F. (2016). Genome-Wide SNP Calling from Genotyping by Sequencing (GBS) Data: A Comparison of Seven Pipelines and Two Sequencing Technologies. *PLOS ONE*. 11(8): e0161333. doi: <https://doi.org/10.1371/journal.pone.0161333>
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 25(9): 1105-1111. doi: <https://doi.org/10.1093/bioinformatics/btp120>
- Valliyodan, B., & Nguyen, H. T. (2006). Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Current Opinion in Plant Biology*. 9(2): 189-195. doi: <https://doi.org/10.1016/j.pbi.2006.01.019>
- Varshney, R.K., Nayak, S.N., May, G.D. & Jackson, S.A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnology*. 27(9):522-530. doi: [10.1016/j.tibtech.2009.05.006](https://doi.org/10.1016/j.tibtech.2009.05.006)
- Van de Mortel, J. E., Almar Villanueva, L., Schat, H., Kwekkeboom, J., Coughlan, S., Moerland, P. D., Van Themaat, E. V. L., Koornneef, M. & Aarts, M. G. M. (2006). Large expression differences in genes for iron and zinc homeostasis, stress response, and lignin biosynthesis distinguish roots of *Arabidopsis thaliana* and the related metal hyperaccumulator *Thlaspi caerulescens*. *Plant Physiology*. 142(3): 1127-1147. doi: <https://doi.org/10.1104/pp.106.082073>
- Venditti, C., Meade, A. & Pagel, M. (2008). Phylogenetic mixture models can reduce node-density artifacts. *Systematic Biology*. 57(2): 286-293. doi: <https://doi.org/10.1080/10635150802044045>
- Wang, M., Jiang, N., Jia, T., Leach, L., Cockram, J., Comadrán, J., Shaw, P., Waugh, R. & Luo, Z. (2012). Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. *Theoretical and Applied Genetics*. 124(2): 233-246. doi: <https://doi.org/10.1007/s00122-011-1697-2>
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J. H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Wang, J., Wang, X., Freeling, M., Pires, J. C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.S., Weishaar, B., Liu, B., Li, B., Liu, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G. J., Bonnema, G., Tang, H., Wang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H., Jin, H., Parkin, I. A., Batley, J., Kim, J. S., Just, J., Li, J., Xu, J., Deng, J., Kim, J. A., Li, J., Yu, J., Meng, J., Wang, J., Min, J., Poulain, J., Wang, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M. G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P. J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S., Sato, S., Sun, S., Kwon, S. J., Choi, S. R., Lee, T. H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu,

- Y, Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y., Wang, Z., Li, Z., Wang, Z., Xiong, Z. & Zhang Z. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*. 43(10): 1035-1039. doi: <http://dx.doi.org/10.1038/ng.919>
- Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 10(1): 57-63. <https://doi.org/10.1038/nrg2484>
- Ward, J. A., Bhangoo, J., Fernández-Fernández, F., Moore, P., Swanson, J., Viola, R., Velasco, R., Bassil, N., Weber, C. A. & Sargent, D. J. (2013). Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genomics*. 14(1): 1-14. doi: <https://doi.org/10.1186/1471-2164-14-2>
- Whankaew, S., Poopear, S., Kanjanawattanawong, S., Tangphatsornruang, S., Boonseng, O., Lightfoot, D. A., & Triwitayakorn, K. (2011). A genome scan for quantitative trait loci affecting cyanogenic potential of cassava root in an outbred population. *BMC Genomics*, 12(1):266. doi: <https://doi.org/10.1186/1471-2164-12-266>
- Wicke, S., & Schneeweiss, G. M. (2015). Next-generation organellar genomics: Potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. In: E. Hörandl & M. S. Appelhans (Eds.), *Next-Generation Sequencing in Plant Systematics*. pp.1 - 42. Königstein, Germany : Koeltz Scientific Books. 298p.
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T. W., Li, Y., Xu, X., Wong, G. K. & Wang, J. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. 30(12): 1660-1666. doi: <https://doi.org/10.1093/bioinformatics/btu077>
- Xu, J., Li, Y., Ma, X., Ding, J., Wang, K., Wang, S., Tian, Y., Zhang, H. & Zhu, X. G. (2013). Whole transcriptome analysis using next-generation sequencing of model species *Setaria viridis* to support C4 photosynthesis research. *Plant Molecular Biology*. 83(1-2): 77-87. doi: <https://doi.org/10.1007/s11103-013-0025-4>
- Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., Dong, Y., Gutenkunst, R. N., Fang, L., Huang, L., Li, J., He, W., Zhang, G., Zheng, X., Zhang, F., Li, Y., Yu, C., Kristiansen, K., Zhang, X., Wang, J., Wright, M., McCouch, S., Nielsen, R., Wang, J. & Wang, W. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology*. 30(1): 105-111. doi: <https://doi.org/10.1038/nbt.2050>
- Yang, I. S., & Kim, S. (2015). Analysis of Whole Transcriptome Sequencing Data: Workflow and Software. *Genomics & Informatics*. 13(4): 119-125. doi: <https://doi.org/10.5808/GI.2015.13.4.119>
- Zhou, X., Lindsay, H., & Robinson, M. D. (2014). Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Research*. 42(11): e91-e91. doi: <https://doi.org/10.1093/nar/gku310>
- Zhu, H., Choi, H.-K., Cook, D. R. & Shoemaker, R. C. (2005). Bridging model and crop legumes through comparative genomics. *Plant Physiology*. 137(4): 1189-1196.