Genomic and transcriptomic approaches toward plant selection

Enfoques genómicos y transcriptómicos hacia la selección de plantas

Seyed Mehdi Jazayeri[1,*], Ronald Villamar-Torres[2,3,4].
[1]Universidad Nacional de Colombia, Bogotá, Colombia.
[2]Université de Montpellier, Montpellier, France.
[3]Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador – SENESCYT, Ecuador.
[4]Facultad de Ciencias Agrarias - Universidad Técnica Estatal de Quevedo, Ecuador.
smjazayeri@unal.edu.co, villamartorresronaldoswaldo@yahoo.es

Fecha de recepción: 18 de octubre de 2017 — **Fecha de aceptación:** 08 de diciembre de 2017

DOI: https://doi.org/10.26910/issn.2528-8083vol2iss8.2017pp54-64

***Abstract*—Omics era has opened a new window to biology. Genomics and transcriptomics are two well-known fields by which plant selection and breeding are fulfilled more easily and accurately. They provide useful information about genes, transcripts, their functions those are the principal data for other subsequent approaches. Reference genomes of various plants are available and facilitate genome-based studies. The complex of genomic, transcriptomic data and the findings from variant methods like QTLs (quantitative trait loci), SNPs (single nucleotide polymorphism), CNVs (copy number variant), resequencing, GBS (genome-by-sequencing) are extremely important for plant selection in terms of price and time. The new workflows are routinely using different approaches and mixing them based on the genomic/transcriptomic information in their subsequent steps. They, however, are validated during the whole process toward screening genotypes possessing agronomically important desired trait. SNP-Seq presented hereinafter is a new approach for analyzing plants toward selection and screening by SNP sequencing in various genotypes simultaneously. It can accelerate the cycle of plant selection from genotypes to phenotypes in a reverse engineering way.**

***Keywords:*—Genomics, Omics, Plant Selection, Plant Improvement, Transcriptomics**

***Resumen*—La era Omica ha abierto una nueva ventana a la biología. La genómica y la transcriptómica son dos campos conocidos, con los cuales, la selección y el mejoramiento de plantas se cumplen con mayor facilidad y precisión. Proporcionan información útil sobre los genes, las transcripciones, sus funciones y sirven como datos primordiales para otros enfoques posteriores. Los genomas de referencia de varias plantas han sido secuenciados, y están disponibles, facilitando así el acceso a información ómica indispensable para llevar a cabo estudios basados en estos mismos genomas. El total de datos genómicos, transcriptómicos y los hallazgos de métodos variantes que van desde QTL (rasgo cuantitativo), PSN (polimorfismo de un solo nucleótido), NCV (número de copias variante), GBS (genoma por secuencia) son extremadamente importantes para la selección y el mejoramiento de plantas en términos de precio y tiempo. Los nuevos flujos de trabajo utilizan diferentes enfoques basados en la información genómica / transcriptómica en pasos posteriores mezclándolos y se validan durante todo el proceso para seleccionar genotipos que posean un rasgo deseado agronómicamente importante. SNP-Seq, que se presenta en este artículo, es un nuevo enfoque para analizar las plantas hacia la selección y la detección mediante secuenciación de SNP en varios genotipos simultáneamente. Este proceso puede acelerar el ciclo de selección de plantas desde los genotipos a los fenotipos en una forma de ingeniería inversa.**

***Palabras claves:*—Genómica, Ómicas, Selección Vegetal, Fitomejoramiento, Transcriptómica**

INTRODUCTION

Sustainability of agriculture, including the increase of crop output, as wellas reduction of production costs and adaptation to stricter standards for the health of farmers, populations and for the environment in the growing regions, are major modern challenges taken into consideration by plant geneticists nowadays (Villamar et al., 2016). Consequently, it is important to elucidate the capacity of generating information that contributes as basis for carrying out programs of plant improvement. The domain of certain techniques like genomics and transcriptomics will allow getting equilibrium between cost and efficiency for the modern researchers, avoiding the use of tedious and long classic breeding process up to obtaining improved genotypes (Viot, 2016).

The suffixe–ome originating from Greek meaning “body”, appeared in chromosome and then genome, refer to an object in biology that possesses a character/feature/wholeness. It means that whole part of one object in living organisms; genome is the whole genes, transcriptome is the whole transcripts, proteome is the whole proteins and metabolome is the whole metabolites of an organism. The suffix–omics address to the study of one–ome. Nowadays, Omics is the field of research analyzing and inter/multidisciplinary studies of many different-omes generally using bioinformatics and computational biology.

Omics has revolutionized biological and life sciences through the tremendously increasing number of research and scientific outputs during the last decades, which have disclosed many issues and generated various outputs in biology, agriculture and life sciences and as an example, better-performance crops. Genomics, transcriptomics,

[*]Ingeniero Agrónomo

proteomics and metabolomics those are the field of study of genome, transcriptome, proteome and metabolome respectively, are under more consideration as the better-known omics (for more information see https://omics.org/ and http://www.genomicglossaries.com/content/omes.asp). Other Omics have also been derived from the principal -omics to solve different biological issues or reveal undiscovered matters like Epigenomics that studies epigenome, which is the regulatory chemical compounds controlling whole gene expression or Pangenomics that is the field of study of pangenome, which is the whole genome of individuals of a species.

Genomic and transcriptomic data help to find a genotype possessing better performance comparing with its neighbor, relative or parental genotypes or to unveil phylogenic and evolutionary relativeness among genotypes. Genomics has discovered some evolutionary issues like polyploidy in wheat Borrill et al. (2015) and family relation in bred and wild plants (Kang et al., 2016). These can lead us to a better understanding of plant kingdom toward plants with better potential yield or special characters in terms of agronomic traits. A workflow mixed of genomics and transcriptomics with bioinformatics tools allows novel gene discovery and unveils regulatory sequences. In other hand, they lead QTL to gene discovery and chromosomal position, and make available large collections of molecular markers such as SNPs. The gene expression is another output of genomics and transcriptomics by which the candidate genes are presented as the genes involved in specific pathways or regulatory function related to an agronomic trait. All these results and outputs can help the scientists to screen and discovery the more efficient genotypes toward plant selection.

Thanks to Next Generation Sequencing (NGS) technologies, mass sequencing of genomes and transcriptomes have been generated, those are capable to output a vast array of genomic information. The contribution of NGS technologies because of their relatively low cost and speed in generation and analyses is dramatically enhancing our ability to comprehensively understand how the gene-based information in a cell exist and how the genes acts and regulate biological processes and molecular functions in a living organism. This technology has engaged to study genome sequence variation, ancient DNA, cytosine DNA, DNA methylation, protein-DNA interactions, whole transcriptomes (RNA) sequencing, alternative-splicing, small RNA populations and mRNA regulation, microRNA sequencing using a number of these applications used in plant systems (Lister et al., 2009).

The genomics and transcriptomics methods such as DNA-Seq, RNA-Seq, ChIP-Seq, Methyl-Seq, MAINE-Seq, RIP-Seq and PhIT-Seq have currently improved our knowledge about gene expression and regulatory in plants (Shendure and Aiden, 2012). These methods-of-choice are increasingly used as their costs and complexity decrease and are generating many millions of sequencing reads as a routine. The new data based on genomics and transcriptomics like RNA-Seq outputs have this nature of big size and need high performance computation systems to analyze them (Jazayeri et al., 2015). However, nowadays as bioinformatics and computation facilities are advancing in better and higher performance and less price, these challenges are ignorable as one can find suitable computation unities to perform the required analyses because of availability of adequately suitable system.

Because of increasing global demand, climate change, and environmental issues, novel crops possessing higher production potential, more tolerance to environmental cues, and agro-biological features toward less chemical usage are highly required. Genomics supplies new foundations for crop-breeding systems, while combining with improved and automated phenotyping assays and functional genomic studies (Bevan et al., 2017). Advances in genomics propose the potential to accelerate the process of developing crops with promising agronomic traits. Application of genomics in agriculture enhances the productivity and sustainability in crop production. A tremendous increase in genomic resources has become available by mixing classic and high-throughput sequencing platforms. Some of the resources are expressed sequence tags (ESTs), BAC end sequence, genetic sequence polymorphisms (SNP, CVN), gene expression profiling (differentially expressed DE analysis), whole-genome (re)sequencing (WGS), whole transcriptome sequencing (WTS) and genome-wide association studies (GWAS). Due to availability of genomic sequencing facilities and expansion of bioinformatics tools, the studies from individual gene are switching to whole-genome analysis, which provides a wider view of how the complex of all genes work together (Wang et al., 2017). This can lead us to a better understanding of gene function and the probable regulatory matters among networks of genes and generating more easily and faster improved plants. Bioinformatics is the field that makes omics data meaningful and help us to find their powerful ability to discover novel issues. It integrates the data from different omics from genome to phenome (Edwards and Batley, 2004) and allows to mine data across various types of scientific discipline using many bioinformatics tools (Skuse and Du, 2008). Applications of bioinformatics in Omics studies is an essential part like what the software do in a computer; without the suitable programs no output will be generated. Nevertheless, the biologists, bioanalysts and omists by cooperation with bioinformaticiens are able to use omics data and interpret them in a scientific manner. In other hand, bioinformatics tools are those have advanced omics field, as they are the media for unrevealing the biological secrets hidden in the genes and biomolecules from the living level to a human-understandable interpreting output.

This article is aimed to present a review on genomics and transcriptomics as two most known and used Omics and their contribution to plant selection. This manuscript briefly explains genomics and transcriptomics and then the methods and approaches by which the principal information for plant selection is provided. As a matter of fact due to various available methods, the authors have tried to mention the core of the correlation between two different biological fields as plant genomics and selection. In addition, the authors have attempted to bring different examples from crops, model and non model plants to elucidate this reality that the methods can be applied for all plants in general while focusing on crop improvement and selection. Finally, a new approach called

"SNP-Seq" is proposed and presented here theoretically but need to be applied in the new studies to confirm its usefulness. **How do plant genomes and the omics approaches help plant improvement?** The data from interdisciplinary tools mainly as genomics and transcriptomics are subject to a very good assistance in plant breeding (Rival, 2017). Genomes and transcriptomes are two main core of Omics and provide the input data for subsequent bioanalysis tools/approaches by which many more results are obtainable and reachable. Thanks to bioinformatics, the outputs of genomics and transcriptomics are not just lists of sequences or genes with their annotation to be used in classic plant improvement programs. They are the data that give principal information on modern genomic based selection, which use the whole set of genes to evaluate a genotype as a favorite one considering the under study agronomic trait. The consequences in evolution, phylogeny, taxonomy, family relation and genetic relativeness can be deduced from the output of genomic and transcriptomic studies. The variants and polymorphism are also predicted by bioinformatics algorithms and can be used in plant populations in order to screen plant genotypes with desired performance and trait (Figure 1).

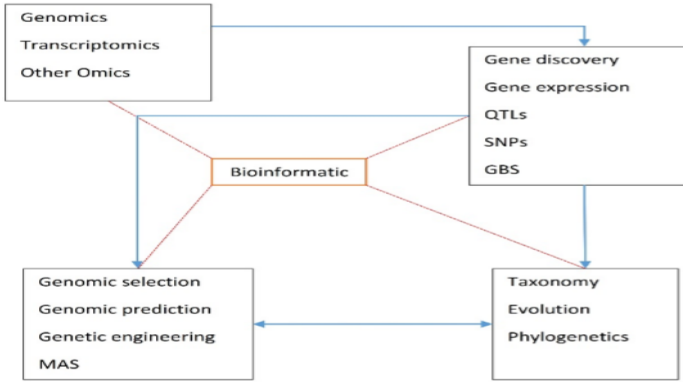

Figure 1. The omics and the approaches toward plant improvement. Bioinformatics serves in all steps as the actor of performing the task (shown by red dashed connector). The results of such pipeline from genomics and transcriptomics can be used in subsequent fields like evolution, taxonomy and phylogeny.

Source: Prepared by the authors.

The study of genome and transcriptome in the green alga Chromochloris zofingiensis has reported to enhance production of carotenoids and commercial production of this alga (Roth et al., 2017). In one study on rice, Zhang et al. 2016 showed that two elite indica rice varieties Zhenshan 97 and Minghui 63 are the parents of a leading rice hybrid while 1,059–2,217 more genes expressed in the hybrid than the parents did. The results of this study uncovered structural differences of the two genotypes that showed how the hybrid could be formed (Zhang et al., 2016). This research leads the scientists to find the well-performance genotypes by studying the genome of their parents in order to generate new hybrids with more production and with more accuracy. However, such study and studies between the relatives of a crop with desired agronomic traits will improve our understanding of how the crops diverge and how we can produce better performance crops using the whole gene or differentially expressed gene (DEG) catalog of each crop.

Tobacco genome revealed its ancestral origin, familial and taxonomical relation with tomato and potato and different biochemical pathways involved in alkaloid biosynthesis. In addition, the study confirmed that disease resistance to different viruses in tobacco is caused by genomic deletion of the S-form eIF4E1 locus (Sierro et al., 2014). The findings are useful to screen the resistant genotypes to virus-caused diseases not only in tobacco but also in other plants while these data are used as the reference.

The divergence among plants has been disclosed by comparative analyses among different genera or species. These results are subject to use in downstream steps and researches. The data can be of help to understand if a new cross (or backcross) can perform as expected. Oil palm genome disclosed that the palm trees are ancient tetraploids and American and African oil palms (i.e. Elaeis oleifera and E. guineensis respectively) diverged in the Old and New Worlds (Singh et al., 2013). This comparison between two species is very important for oil palm breeding and selection as the hybrids between two African and American species are more productive and less susceptible to diseases. In addition, transcriptomic results showed that the genes expressed in fruit and kernel of oil palm follow different patterns according to the properties of fruit and kernel oils.

One approach in genomic field is studying the genome of crop wild relatives (CWRs) in comparative studies. Genomics of CWR provides data to expand the genetic diversity of crop plants. Analysis of the nuclear genome, transcriptome and maternal (chloroplast and mitochondrial) genome of CWR eases their use in plant improvement. The analyses based on genome/transcriptome develop discovery of useful alleles in both domesticated and wild relatives and disclosing regions of the genome where likely diversity between domesticated and wild plants has been modified in domestication process (Brozynska et al., 2016).

Such pairwise comparative studies between different genera, species, contrasting genotypes, phenotypes based on their -omes reveal several data those enable us to find the answers to our questions and undiscovered issues about genes involved in productivity, tolerance, regulation while two set of genes are compared and generate the candidate genes in each item. It is a routine to elaborate novel studies, create new technologies, propose pipelines and workflows and generate pioneer programs and software as a chain of various tasks to find a solution for a problem.

Genomics and transcriptomics Genomics is the systematic study of an organism's genome. The genome is the total DNA of a cell or organism. In other hand, transcriptome is the total transcripts or mRNA in a cell or organisms. Genome transcribes to transcriptome and transcriptome translates to proteome. The genomic data give us this opportunity to find the genes whose subsequent transcripts show different expression pattern that direct us to the main core of central dogma in biology. Genome is more stable than transcriptome that is why genome is used in variant detection studies where a firm

base needed to compare how the natural or artificial selection has improved the plants during the time. Transcriptome is employed to determine expression profile where a flexible response to environmental cues as ephemera phenomenon is desirable. However, it should be taken into account that genomes and transcriptomes are subject to be the eligible items in G x E as both temporal and permanent effects of environment on plants are studied through evolutionary or adaptive patterns.

In order to perform a genomic, transcriptomic or better-to-say Omics study, experimental design is needed Horgan and Kenny (2011) considering 1) the use and choice of suitable biological samples depending on the experiment goals and the availability of the sample 2) the biological variation: considering that what needed and in which level to ensure that changes are not due to confounding factors and 3) the technical/analytical variation: this is a little bit tricky as currently the platforms and methods supported by bioinformatics tools are able to generate reproducible results with the least standard deviation for a specific experimental technique.
Advanced genomics permits us to recognize the fragments of the genome responsible for a trait (Khan et al., 2016). It can enhance our understanding of microevolution and family relations among crops through a better understanding of natural selection, mutation, and recombination. It is important to understand the structure, organization, and dynamics of genomes in plant species as it can provide insights into how genes have been adapted by natural and artificial selection to respond to environmental constraints and the potential of their manipulation for crop improvement (Wang et al., 2017).

The first reference genome published for Arabidopsis started the new genomic era that later on changed to a new period of Omics as the data of genomics, transcriptomics and then other Omics have served in millions of research projects. The valuable information published by the genome of Arabidopsis changed and revolutionized the scientific world of plant biology and agriculture in a very short time. However, now the genomic data and functional genomic outputs of almost all plants are based on the Arabidopsis structural and annotation genome.

The most highlighted genome case, Arabidopsis thaliana with its famous website https://www.arabidopsis.org/, is a highly studied model organism due to its reduced number of chromosomes (five chromosomes) that make it top in more straightforward information search resulting in further interpretation and data set management (Thaliana, 2000). For example, to date for Arabidopsis thaliana genome through molecular genetics and following bioinformatics supports, it has been reported that terpene synthesis belong to a total of 40 TPS-genes, of which thirty-two are considered as functional genes while eight are pseudogenes. According to their phylogeny, they are highly related to monoterpenes, sesquiterpenes and diterpenes, and highly expressed after insect attack (Aubourg et al., 2002). Other organisms of alimentary or economic interest have been sequenced as well and their genomes have been placed to the researcher disposal, among them: https://www.maizegdb.org/; http://www.cacaogenomedb.org/; https://shigen.nig.ac.jp/rice/oryzabase/ (Zea mais, Theobroma cacao and Oryza sativa, respectively).

However, reference genome of thousands plants now is available as the base of all other genomic related studies and as the first necessary data for other Omics like transcriptomics ad proteomics. A very handy and useful list based on the time of publication of plant genomes and the related reference is accessible in http://www.plabipd.de/timeline_view.ep. Another resource for plant published genomes with their explanation is available in https://genomevolution.org/wiki/index.php/Sequenced_plant_genomes.

Pan-genomics and pan-transcriptomics

The pan-genome (pangenome as well) is the set of all the genes and sequences (coding and non-coding) found in all individuals of a species. Pangenome or pantranscriptome is a compilation of various individual genomes or transcriptomes by which the probable variant effects and structural variation can be presented. It can ameliorate the variation analyses in plants as a supplementary for other types of polymorphism like SNPs or CNVs and is useful to disclose presence/absence of variants. It can be of help for studying the plants with big genomes and reveal the undiscovered genes and their potential functions as done in hexaploid wheat (Montenegro et al., 2017) and in maize (Jin et al., 2016) (Lu et al., 2015).

In addition, pangenome cans determine the events (insertion, deletion, mutation) among vast individuals of a species those have been distributed in the plant genome during the evolution processes. It can lead us to understand the processes generating genetic diversity and phenotypic variation (Morgante, 2013). As a viewpoint of plant selection, pangenome leads us to understand germplasm diversity toward a screening process. Pangenomics has helped to find the new genes that were lost during domestication in Brassica comparing wild and domesticated plants (Golicz et al., 2016). Such different and novel genes whose function is considered as a special trait between wild and domesticated plants can be used as the genic markers for screening plants among a very vast population consisting wild and domesticated or bred plants. This can advance the process of plant selection.

Bioinformatics

During the last years outstanding advances have been achieved in the manner of analyzing omics data, based on several methods frequently using informatics tools. Piecemeal, biologists have taken into advantages of this interdisciplinary field of the biological sciences to inquire deeply on dogma of molecular biology and their complex procedures (Bayat, 2002). Bioinformatics develops methods and software tools for understanding biological data, allowing that its wide field of action is linked to other sciences, such as: statistics, mathematics, computer, and engineering sciences (Robinson et al., 2010). Genomic and proteomic based sequence databases of different organisms with own bioinformatics tools integrated in their platforms are conforming currently an uncountable offer available for biologists and geneticists (Hamid et al., 2009).

The use of computer programming also makes part of the methodologies used by bioinformatics, as well as a reference to specific analysis "pipelines", which are performed, mainly for genomic and transcriptomic analyses. In the field of plant selection, since the insertion of this set of tools several investigations have been performed including its use, particularly, related to the identification of candidate genes and single nucleotide polymorphism (SNPs) with the aim of enhancing our understanding of genetic basis of certain diseases, as well as adaptation and tolerance to different biotic and abiotic stresses (Bayat, 2002).

In the last decade a very common essential tool as core of bioinformatics used by biologists, geneticist and other researchers linked to omics studies, has been Linux overall because of its rapidity, simplicity and facility to obtain results avoiding longer and tedious procedures. Nowadays its utilization has permitted to accelerate the speed of analysis of different genomic and transcriptomic data set. Therefore, it is not astonishing that presently many biological research centers worldwide have commenced to incorporate such bioinformatics type of profiles due to its wide demand by their projects. To all this, the researchers have nominated it like the era of the management, deduction and analysis of genomic and transcriptomic information thank to bioinformatics, which has as last aim of use this information for plant selection. Other tools exist those established by R language program available its bioinformatics version in the website: https://www.bioconductor.org/. Bioconductor, which is a package database, provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and development.

GWAS and TWAS

Genome-wide association study and transcriptome-wide association study are two method to find the relationship between molecular markers and QTL based on linkage disequilibrium. They can be used to identify the genes with complex traits as well as their expression-trait associations and to detect adaptive genetic variation in structured background in an ecological context as reported by Brachi et al. (Brachi et al., 2011) or in plant responses to biological factors as pathogen (Pathosystems) (Bartoli and Roux, 2017) or in plant to stress (Thoen et al., 2017). Association mapping on originating parents and their descendants can disclose some important QTL and favorable allelic variations, which can be further used to generate more favorable varieties using the breeding and selection methods like marker-assisted selection. GWAS identifies candidate genes related to agronomically important traits and then the results will be applied in the subsequent screening based on SNP of the candidate genes. GWAS can detect new genes those are not detectable by standard SNP analysis that accelerates crop improvement as reported by Yano et al. for rice (Yano et al., 2016).

Gene expression profile analysis

Differentially expression (DE) analysis is one promising way allowing discovering plants with an agronomic favorite trait based on a DE profile covering the whole gene/transcript set expressed differentially in such condition of comparison. One of the well-known and done approaches in transcriptomics is to study differentially expressed genes/transcripts (DEG) to pairwise-compare plants (i.e. two contrasting characters; susceptible versus tolerant, control versus treated, etc.). Gene expression profile (actually RNA-Seq or whole transcriptome sequencing) outputs a table of DEGs in under/over expression or up/down regulation manners while an agronomically important trait is studied such as tolerance to stresses (biotic and abiotic) (Fracasso et al., 2017) (Jazayeri et al., 2015), resistance to diseases and pests (Gong et al., 2015) (Li et al., 2016), metabolite biosynthesis like alkaloid and photosynthesis efficiency (Ding et al., 2015) (Zhao et al., 2016). From the DEGs, a list of candidate genes is extracted and used in subsequent steps where the plants possessing the similar expression pattern are screened for the trait. The advantage of this approach is to use the whole set of genes not just an individual gene as it has been shown and undiscovered that a gene set including functional and regulatory genes is active under a given condition.

QTLs

Being quantitative, agronomically important crop traits are under more attention as they are controlled by a total of effects coming from multiple genes each with a small effect. Each gene complex (a series of genes located on a QTL) can determine a quantitative trait due to genic cooperation among the genes. QTL analysis is a statistical method connecting phenotypic data (trait measurements) and genotypic data (usually molecular markers like SNPs, SSRs, etc.) in order to disclose the genetic information of variation and variants for desirable agronomic traits (Miles et al., 2008). The desired traits vary among plants depending on the plant under study and the interest to a special or specific trait and the pertinent QTLs. QTLs have increasingly been reported in several studies such as plant height Brown et al. (2008) Li et al. (2015) Xu et al. (2017), resistance to diseases Ayala et al. (2002) Zhou et al. (2015), seed size Zhang et al. (2015), grain yield Vikram et al. (2016), wood production in trees like Eucalyptus Rocha et al. (2007), and higher yield Habyarimana et al. (2017) Lu et al. (2017).

For an efficient crop improvement program using marker-assisted selection (MAS), mapping and isolation of QTLs are decisive and for a better understanding of the molecular mechanisms underlying the traits, they are crucial. However, QTL analysis is time-consuming and labor-intensive as the development and selection of DNA markers for linkage analysis needs such long labor task.

Expression Quantitative Trait Locus or eQTL, as a QTL-based field, is a chromosomal region possessing the variation in abundance of a mRNA transcript observed between individuals in a genetic mapping population (Druka et al., 2010). A single gene may possess one or many eQTLs. By eQTL, genetic regulatory networks, potential coexpression among

genes are better understood (Wang et al., 2014). Phenotypic variation and candidate genes of favorite trait QTLs are analyzed more easily using a combination of classic and trait QTL with the expression profiles of such QTLs. mRNA expression profiles then can infer chromosomal positions of several genes those by eQTL are detectable and can be used in plant selection according to their function and expression pattern. eQTL can be based on a specific gene or trait to find relevant QTL position on chromosome and then used as a screening income. Quantitative trait nucleotide (QTN) and quantitative trait transcript (QTT) are of the polymorphism predictors providing fruitful information combining gene function, phenotype and QTL architecture (Heidaritabar, 2015). They are of genetic approaches of quantitative traits allowing to observe quantitative trait effects and to map phenotype to genotype in the absence of biological context. (Mackay et al., 2009).

SNPs Thanks to high-throughput sequencing (HTS), Single nucleotide polymorphisms (SNPs) have rapidly improved molecular genetics during the recent years due to their abundance in the genomes and their manageability for data processing and compatibility for formats and platforms. Computational bioinformatics approaches are those that overcome SNP discovery methods due to the ever-increasing sequence information that is added to public databases; nevertheless, genome complexity in some plants can make it challengeable in obtaining informative SNPs (Mammadov et al., 2012). Thus, other alternative strategies in those crops are probably needed to combine with the undergoing methods.

Nowadays several million SNPs are reported for one species or variety thanks to powerful programs those are very helpful to detect SNPs in a genome-based data. The programs are able to distinguish sequencing error and real SNPs while comparing the sequences with a reference genome and this make the SNP detection and exploration a reliable method toward plant screening by using SNPs found in the interest gene(s) related to an agronomically important trait.

It is possible to generate and process millions of SNPs at the same time and generate SNP arrays and chips applicable for plant selection. Large SNP arrays including millions and thousands of SNPs compiled in a chip or array are reported for plants and of help for subsequent plant selection and improvement programs (Ganal et al., 2012). The user can find the gene of interest related to a desired trait and find the SNPs for such genes (if available and reported in the chip). Finally the information of gene-specific SNPs will be used to screen the genotypes those have these favorite SNPs and this opens the ports for downstream research toward plant improvement by mining the SNPs in different genotypes and selecting the candidate genotypes possessing these SNPs as those are expected to behave in the similar manner for the trait.

In one article a single-copy gene 50K SNP chip has been reported for rice that has facilitated phylogenetic study and genetic diversity of cultivated and wild rice and has been validated for its efficacy in plant breeding and making mega rice varieties (Singh et al., 2015). A multi-species 60K SNP chip has developed for Eucalyptus from 12 species that is an outstanding tool to address population genomics questions and empower genomic selection, GWAS and complex trait variant studies in Eucalyptus (Silva-Junior et al., 2015). In a study on oil palm, Kwong et al. have reported a study outputting an array of 200K SNPs for two oil palm species (Elaeis guineensis and E.oleifera) with more than 170K successful probe (Kwong et al., 2016). They revealed that the generated array can be used in differentiating oil palm origins and for population diversity; it is a robust approach with potential for developing early trait prediction to shorten the oil palm breeding cycle.

Genotyping-by-sequencing (GBS)

Array-based approaches to single nucleotide polymorphism (SNP) screening usually have been the method used by biologists when it is about analyzing and associating traits corresponding with specific regions in the genome. However, nowadays this has changed due to the access to new methods of NGS for genotyping, which is a powerful method for genetic screening uncovering SNPs in plants; which allows carrying out the studies based on genotyping. It seems the declining cost of the use of this technology is also a plus point so that researchers have spared no any effort to choose it Crossa et al. (2013) Elshire et al. (2011).

Among different advantages presented by genotyping by sequencing, some are: A) It allows comparative analyzes between samples in the absence of a reference genome, B) it provides a low cost per sample, C) it permits to identify sequences in prearranged zones of genetic variation for many samples, D) it identifies non-SNP variants, including small insertions, removals and microsatellites and E) it allows comparisons among samples even when they do not appear on reference genome Elshire et al. (2011).

For organisms that contain complex genomes or when researchers have limited resources, GBS is highly cost-effective. Therefore, some techniques including targeted sequencing based on amplicons, enrichment sequencing based on hybridization, and genotyping based on the restricted sequence representation by restriction enzymes are available Elshire et al. (2011).

Illumina to date is the most used platform for genotyping by sequencing; therefore, their inventors for guarantee optimal results give several advices. For example: it is necessary to account with a reference genome and high-diversity samples (i.e. biological and technical issues) and in order to minimize false positives, it is imperative to have high tolerance of ambiguity in heterozygote screening, avoiding the redundancy (Sandmann et al., 2017).

Genomic selection

Genomic selection (GS) is a method to foresee the genetic candidates by the markers distributed throughout the genome. Unlike marker-assisted selection that searches to identify individual candidate loci significantly associated with a desired trait, GS uses all available marker data as performance foretellers and consequently delivers more accurate predictions.

Selection based on GS data, can potentially lead us to obtain more rapid and lower cost gains from plant improvement.
The genomic estimated breeding value (GEBV) is used as a feature of GS. To obtain GEBV, all markers and their effects are used. Therefore, the GEBV have this potential ability to capture more of the genetic variation for the particular trait under selection (Newell and Jannink, 2014). Due to its more ability to detect the markers based on whole-genome predictions and to generate more accurate results, GS can replace phenotypic selection or marker-assisted breeding protocols (Desta and Ortiz, 2014). However, the combination of these methods may result in better conclusion.

Actually, GS is deployed to study a population different from the reference population, which possesses the estimated marker effects. There are two types of datasets in GS: a training set and a validation set. The training set is the reference population in which the marker effects are estimated. While marker effects are estimated based on the training set using certain statistical methods to incorporate this information; GBEV of new genotypes are predicted only based on the marker effect. The selection candidates of validation set obtained from the reference population that have been genotyped and selected based on marker effects estimated in the training set.

The marker effects of quantitative traits have considerably altered standard practices used in plant breeding. Computer simulation is the medium that disclose the benefits of GS in plant breeding. As NGS and HTS technologies are continuously reducing the cost of marker data in terms of price and increasing the available markers in terms of number, currently plant genotyping is less costly than phenotyping. In addition, GS speeds up the selection cycles and makes it easier to increase the selection outputs in a given time. Hence, progeny test in field for selection based on markers using the genomic information might enhance the genetic gains per unit. As there might be doubt on gene deletion and less variant in domesticated generation, it is suggested to consider wild and domesticated crops at the same time to empower usage of such genomic markers.

As GS is based on genomic data that cover more complete information from various sources and possess info from whole genome to individual genes, it is considered as more accurate and better pathway to get more reliable and outstanding results. This preference for GS has been reported in wheat breeding, while the average prediction accuracies for GS was reported 28% greater than MAS and 14% greater than phenotypic selection (PS) (Heffner et al., 2011).

In order to improve yield in Brassica napus, a study reported that an integration of different methods as GWAS, DEGs and SNPs generate a promising output by which selection of rapeseed accelerates (Lu et al., 2017). The authors of this article have shown that candidate gene discovery is more accurate while using various datasets outputted by different methods and approaches and it is possible to determine which genes are key responsible genes involved in yield improvement of rapeseed.

In pea, strategies of using SNPs, QTLs and GWAS toward genomic selection seems to be as a proficient approach that can help to generate more efficient peas (Tayeh et al., 2015). In rice, genomic selection has been applied successfully to breeding programs in an efficient pattern by using GWAS, QTLs and GBS method. This study has revealed that interpretation of GWAS in genetic architecture and population structure is a useful tool to enhance rice breeding through genomic selection (Spindel et al., 2015). In potato Slater et al. (2016) reported that for four key traits with varying heritability, genomic selection is more accurate using genome-wide SNPs and can improve genetic gain (Slater et al., 2016).

Drought tolerance in maize has been revealed by comparing different models and SNPs for plant transcription factors involved in the processes of drought response and tolerance were validated (Shikha et al., 2017). The authors of this study have shown the accuracy of models and SNPs those are useful for the selection of superior genotypes and tolerant hybrid.

A combination of SNPs and GBS can be employed to plant screening based on identification SNP and different programs are available to process genomics and transcriptomics data to a meaningful application. The SNPlex genotyping system represents a good compromise to investigate several hundred SNPs in a hundred or more samples at the same time (Tobler et al., 2005).

QTL-Seq is able to detect QTLs over wide ranges of experimental variables, and the method can be generally applied in population genomics studies to rapidly identify genomic regions that underwent artificial or natural selective sweeps (Takagi et al., 2013).

However, as shown in above examples, an approach of using different mentioned methods such as GBS+SNP+QTL, WGAS+SNP+GS, RNA-Seq+SNP-Seq+GBS, etc. covers more completely the different aspects of genomics. Mixing different approaches is somehow a good way to get more accurate and reliable results while they are based on genomic data and validated by different methods at the same time. In a study on Eucalyptus, Müller et al. (2017) showed that how genomic data can be useful for discovering heritable variation and how the genome wide analysis validate the results of genomic prediction (Müller et al., 2017).

We suggest a new strategy hereinafter we called it as "SNP-Seq" as a reverse engineering method from genotype to phenotype. The target of performing SNP-Seq is to unveil the relationship between genetically and phenotypically close plants in different scales from varieties to genera. It is a process for analyzing plant populations, different species, close genera and number of varieties in a genomic manner to discover any relationship among under study plants. The results will serve to different biological fields such as population genetics, evolution, phylogeny, taxonomy, plant selection, breeding in order to answer likely doubts and questions on biological relations among the under study complex of plants.

SNP-Seq

is the study of various SNPs simultaneously by sequencing different genotypes and populations, either the whole genomic or transcriptomic sequences or the partial ones obtained from

the previously available information for a targeted trait, in order to screen the genotypes with desired traits predicted by SNP presence. In this approach, several genotypes are sequenced at the same time and then will be compared with each other in order to find their relation based on the studied SNP. Then the plant genotypes with the same SNP sequence will be grouped and by applying statistical methods, the best genotypes will be targeted and screened. GBS is used after that to validate the SNPs. Finally, the chosen genotypes, which are suspected to possess a special desirable trait, will be validated for their potential trait in field and under a real situation. This approach can decrease the time and price of plant selection drastically as sequencing and analyzing such genomic/transcriptomic-based data takes less time as some months compared to a classic selection approach of some years. Another outstanding point about SNP-Seq is that it can be combined with other methods like genomic prediction or QTL to enhance its ability and accelerate and validate the results (Figure 2).

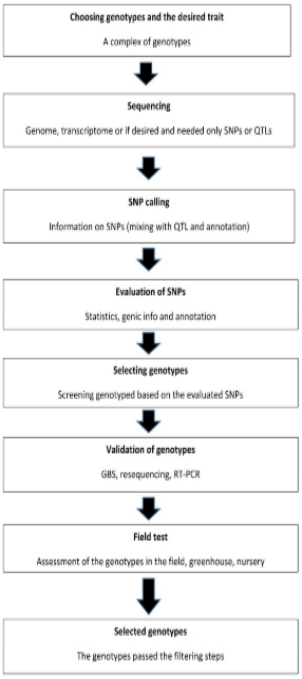

Figure 2. SNP-Seq workflow. SNP-Seq starts from a complex of genotypes those are screened for a desired trait. The SNP/genomic data are produced and then analyzed to find SNPs related to such trait using the genes, annotation, QTLs (the SNPs located on the QTLs of the trait that previously predicted and proved), and any other type of available information. If no previous data are available, the samples are used by their sequences and then the generated genomic data will be used to screen them according to the findings. The samples will be screened and validated by statistical and other types of methods like GBS or resequencing. Tests in field as real proof are done in order to evaluate the genotypes. Finally, the selected genotypes passed the different validation and evaluation steps are presented as the genotypes possessing the desired trait. The length of fleshes show the likely time that takes to pass to the next step.

Source: Prepared by the authors.

Due to concerns about genetically modified crops, molecular design in plant improvement using genomic data leads us to widen use of genetic diversity as the natural variation models in wild progenitors is higher than domesticated plants. It means that wild plants still provide better and more useful resource for genetics and plant selection and facilitate crop genomic and population genetic studies provided the genome-wide data are mined (Huang and Han, 2014).

However, natural-variant plants will be the feasible solution by which human can provide more food for future needs providing that we could generate the right and enough data from enhanced genomic approaches those can screen the wild and domesticated crops toward better-performance genotypes and varieties.

CONCLUSION

Omics tools, mainly genomics and transcriptomics, have opened a new window and started new era for plant improvement. The potential ability of their outputs let us know more and more about undiscovered points in plants and therefore plant selection. Combination of classic and modern technologies, phenotypic and genotypic data, genic and genomic profiles is the best solution for forthcoming plant selection programs if the data from individual genes and whole genomes are employed in a joint research. Plant selection and improvement will see better advances as Omics tools are more available in the biological and agricultural research gaining time and cost dramatically due to their ability of data outputting and analyzing. SNP-Seq as a new approach using different genomic methods is provided conceptually and the authors recommend it to be used in the future studies to prove its potential. However, it seems that due to the need of human being in the upcoming years to more alimentary vegetal products, it is mandatory to benefit from potential power of omics and their relevant approaches in biological plant studies and plant improvement and advance it by other interdisciplinary tools like bioinformatics.

BIBLIOGRAPHIC REFERENCES

Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (attps) gene family of arabidopsis thaliana. *Molecular Genetics and Genomics*, 267(6):730–745.

Ayala, L., Henry, M., Van Ginkel, M., Singh, R., Keller, B., and Khairallah, M. (2002). Identification of qtls for bydv tolerance in bread wheat. *Euphytica*, 128(2):249–259.

Bartoli, C. and Roux, F. (2017). Genome-wide association studies in plant pathosystems: Toward an ecological genomics approach. *Frontiers in plant science*, 8.

Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. *BMJ: British Medical Journal*, 324(7344):1018.

Bevan, M. W., Uauy, C., Wulff, B. B., Zhou, J., Krasileva, K., and Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature*, 543(7645):346–354.

Borrill, P., Adamski, N., and Uauy, C. (2015). Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist*, 208(4):1008–1022.

Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology*, 12(10):232.

Brown, P. J., Rooney, W. L., Franks, C., and Kresovich, S. (2008). Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. *Genetics*, 180(1):629–637.

Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant biotechnology journal*, 14(4):1070–1085.

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., De Los Campos, G., Burgueño, J., Windhausen, V. S., Buckler, E., et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3: Genes, Genomes, Genetics*, 3(11):1903–1926.

Desta, Z. A. and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science*, 19(9):592–601.

Ding, Z., Weissmann, S., Wang, M., Du, B., Huang, L., Wang, L., Tu, X., Zhong, S., Myers, C., Brutnell, T. P., et al. (2015). Identification of photosynthesis-associated c4 candidate genes through comparative leaf gradient transcriptome in multiple lineages of c3 and c4 species. *PLoS One*, 10(10):e0140629.

Druka, A., Potokina, E., Luo, Z., Jiang, N., Chen, X., Kearsey, M., and Waugh, R. (2010). Expression quantitative trait loci analysis in plants. *Plant biotechnology journal*, 8(1):10–27.

Edwards, D. and Batley, J. (2004). Plant bioinformatics: from genome to phenome. *Trends in biotechnology*, 22(5):232–237.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PloS one*, 6(5):e19379.

Fracasso, A., Magnanini, E., Marocco, A., and Amaducci, S. (2017). Real-time determination of photosynthesis, transpiration, water-use efficiency and gene expression of two sorghum bicolor (moench) genotypes subjected to dry-down. *Frontiers in Plant Science*, 8.

Ganal, M. W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H., and Durstewitz, G. (2012). Large snp arrays for genotyping in crop plants. *Journal of biosciences*, 37(5):821–828.

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., Chan, C. K. K., Severn-Ellis, A., McCombie, W. R., Parkin, I. A., et al. (2016). The pangenome of an agronomically important crop plant brassica oleracea. *Nature communications*, 7:13390.

Gong, S., Hao, Z., Meng, J., Liu, D., Wei, M., and Tao, J. (2015). Digital gene expression analysis to screen disease resistance-relevant genes from leaves of herbaceous peony (paeonia lactiflora pall.) infected by botrytis cinerea. *PloS one*, 10(7):e0133305.

Habyarimana, E., Parisi, B., and Mandolino, G. (2017). Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (solanum tuberosum l.). *Plant Breeding*, 136(2):245–252.

Hamid, J. S., Hu, P., Roslin, N. M., Ling, V., Greenwood, C. M., and Beyene, J. (2009). Data integration in genetics and genomics: methods and challenges. *Human genomics and proteomics: HGP*, 2009.

Heffner, E. L., Jannink, J.-L., and Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome*, 4(1):65–75.

Heidaritabar, M. (2015). Accuracy of quantitative trait nucleotide (qtn) prediction by surrounding snps.

Horgan, R. P. and Kenny, L. C. (2011). 'omic'technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, 13(3):189–195.

Huang, X. and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology*, 65:531–551.

Jazayeri, S. M., Melgarejo Munoz, L. M., and Romero, H. M. (2015). Rna-seq: A glance at technologies and methodologies. *Acta Biológica Colombiana*, 20(2):23–35.

Jin, M., Liu, H., He, C., Fu, J., Xiao, Y., Wang, Y., Xie, W., Wang, G., and Yan, J. (2016). Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. *Scientific reports*, 6.

Kang, Y. J., Lee, T., Lee, J., Shim, S., Jeong, H., Satyawan, D., Kim, M. Y., and Lee, S.-H. (2016). Translational genomics for plant breeding with the genome sequence explosion. *Plant biotechnology journal*, 14(4):1057–1069.

Khan, S., Nabi, G., Ullah, M. W., Yousaf, M., Manan, S., Siddique, R., and Hou, H. (2016). Overview on the role of advance genomics in conservation biology of endangered species. *International journal of genomics*, 2016.

Kwong, Q. B., Teh, C. K., Ong, A. L., Heng, H. Y., Lee, H. L., Mohamed, M., Low, J. Z.-B., Apparow, S., Chew, F. T., Mayes, S., et al. (2016). Development and validation of a high-density snp genotyping array for african oil palm. *Molecular plant*, 9(8):1132–1141.

Li, H., Chen, G., and Yan, W. (2015). Molecular characterization of barley 3h semi-dwarf genes. *PloS one*, 10(3):e0120558.

Li, J., Zhu, L., Hull, J. J., Liang, S., Daniell, H., Jin, S., and Zhang, X. (2016). Transcriptome analysis reveals a comprehensive insect resistance response mechanism in cotton to infestation by the phloem feeding insect bemisia tabaci (whitefly). *Plant biotechnology journal*, 14(10):1956–1975.

Lister, R., Gregory, B. D., and Ecker, J. R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current opinion in plant biology*, 12(2):107–118.

Lu, F., Romay, M. C., Glaubitz, J. C., Bradbury, P. J., Elshire, R. J., Wang, T., Li, Y., Li, Y., Semagn, K., Zhang, X., et al. (2015). High-resolution genetic mapping of maize pan-genome sequence anchors. *Nature communications*, 6.

Lu, K., Peng, L., Zhang, C., Lu, J., Yang, B., Xiao, Z., Liang, Y., Xu, X., Qu, C., Zhang, K., et al. (2017). Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in brassica napus. *Frontiers in plant science*, 8.

Mackay, T. F., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, 10(8):565–577.

Mammadov, J., Aggarwal, R., Buyyarapu, R.,

and Kumpatla, S. (2012). Snp markers and@articlemontenegro2017pangenome, title=The pangenome of hexaploid bread wheat, author=Montenegro, Juan D and Golicz, Agnieszka A and Bayer, Philipp E and Hurgobin, Bhavna and Lee, HueyTyng and Chan, Chon-Kit Kenneth and Visendi, Paul and Lai, Kaitao and Doležel, Jaroslav and Batley, Jacqueline and others, journal=The Plant Journal, volume=90, number=5, pages=1007–1013, year=2017, publisher=Wiley Online Library their impact on plant breeding. *International journal of plant genomics*, 2012.

Miles, C., Wayne, M., et al. (2008). Quantitative trait locus (qtl) analysis. *Nature Education*, 1(1):208.

Morgante, M. (2013). Structural variation and the plant pan genomes. *EMBnet. journal*, 19(A):p–11.

Müller, B. S., Neves, L. G., de Almeida Filho, J. E., Resende, M. F., Muñoz, P. R., dos Santos, P. E., Paludzyszyn Filho, E., Kirst, M., and Grattapaglia, D. (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of eucalyptus. *BMC genomics*, 18(1):524.

Newell, M. A. and Jannink, J.-L. (2014). Genomic selection in plant breeding. *Crop Breeding: Methods and Protocols*, pages 117–130.

Rival, A. (2017). Breeding the oil palm (elaeis guineensis jacq.) for climate change. *OCL*, 24(1):D107.

Robinson, G. E., Banks, J. A., Padilla, D. K., Burggren, W. W., Cohen, C. S., Delwiche, C. F., Funk, V., Hoekstra, H. E., Jarvis, E. D., Johnson, L., et al. (2010). Empowering 21st century biology. *BioScience*, 60(11):923–930.

Rocha, R. B., Barros, E. G., Cruz, C. D., Rosado, A. M., and Araújo, E. F. d. (2007). Mapping of qtls related with wood quality and developmental characteristics in hybrids (eucalyptus grandis x eucalyptus urophylla). *Revista Árvore*, 31(1):13–24.

Roth, M. S., Cokus, S. J., Gallaher, S. D., Walter, A., Lopez, D., Erickson, E., Endelman, B., Westcott, D., Larabell, C. A., Merchant, S. S., et al. (2017). Chromosome-level genome assembly and transcriptome of the green alga chromochloris zofingiensis illuminates astaxanthin production. *Proceedings of the National Academy of Sciences*, 114(21):E4296–E4305.

Sandmann, S., de Graaf, A. O., van der Reijden, B. A., Jansen, J. H., and Dugas, M. (2017). Glm-based optimization of ngs data analysis: A case study of roche 454, ion torrent pgm and illumina nextseq sequencing data. *PloS one*, 12(2):e0171983.

Shendure, J. and Aiden, E. L. (2012). The expanding scope of dna sequencing. *Nature biotechnology*, 30(11):1084–1094.

Shikha, M., Kanika, A., Rao, A. R., Mallikarjuna, M. G., Gupta, H. S., and Nepolean, T. (2017). Genomic selection for drought tolerance using genome-wide snps in maize. *Frontiers in plant science*, 8.

Sierro, N., Battey, J. N., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M. C., and Ivanov, N. V. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nature communications*, 5.

Silva-Junior, O. B., Faria, D. A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60k snp chip developed from pooled resequencing of 240 eucalyptus tree genomes across 12 species. *New Phytologist*, 206(4):1527–1540.

Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V., Singh, B., Mishra, S., Singh, Y., Singh, R., Rai, V., et al. (2015). Single-copy gene based 50 k snp chip for genetic studies and molecular breeding in rice. *Scientific reports*, 5.

Singh, R., Ong-Abdullah, M., Low, E.-T. L., Manaf, M. A. A., Rosli, R., Nookiah, R., Ooi, L. C.-L., Ooi, S.-E., Chan, K.-L., Halim, M. A., et al. (2013). Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature*, 500(7462):335–339.

Skuse, G. R. and Du, C. (2008). Bioinformatics tools for plant genomics. *International journal of plant genomics*, 2008.

Slater, A. T., Cogan, N. O., Forster, J. W., Hayes, B. J., and Daetwyler, H. D. (2016). Improving genetic gain with genomic selection in autotetraploid potato. *The plant genome*, 9(3).

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., and McCouch, S. R. (2015). Genomic selection and association mapping in rice (oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS genetics*, 11(2):e1004982.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., et al. (2013). Qtl-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of dna from two bulked populations. *The Plant Journal*, 74(1):174–183.

Tayeh, N., Aubert, G., Pilet-Nayel, M.-L., Lejeune-Hénaut, I., Warkentin, T. D., and Burstin, J. (2015). Genomic tools in pea breeding programs: status and perspectives. *Frontiers in plant science*, 6.

Thaliana, A. (2000). Genome. *Nature*, 408(6814):791–791.

Thoen, M. P., Davila Olivas, N. H., Kloth, K. J., Coolen, S., Huang, P.-P., Aarts, M. G., Bac-Molenaar, J. A., Bakker, J., Bouwmeester, H. J., Broekgaarden, C., et al. (2017). Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *New Phytologist*, 213(3):1346–1362.

Tobler, A. R., Short, S., Andersen, M. R., Paner, T. M., Briggs, J. C., Lambert, S. M., Wu, P. P., Wang, Y., Spoonde, A. Y., Koehler, R. T., et al. (2005). The snplex genotyping system: a flexible and scalable platform for snp genotyping. *Journal of biomolecular techniques: JBT*, 16(4):398.

Vikram, P., Swamy, B. M., Dixit, S., Trinidad, J., Cruz, M. T. S., Maturan, P. C., Amante, M., and Kumar, A. (2016). Linkages and interactions analysis of major effect drought grain yield qtls in rice. *PloS one*, 11(3):e0151532.

Villamar, R., Liu-Ba, G., Legavre, T., and Viot, C. (2016). Los compuestos orgánicos volátiles, defensa natural del algodón# gossypium hirsutum# mejorable gracias a la biotecnología. cibb-ba-eo-061.

Viot, C. (2016). La biotecnología y su utilización en con-

servación y caracterización de los recursos genéticos de# gossypium barbadense# y en mejoramiento del algodón.

Wang, J., Yu, H., Weng, X., Xie, W., Xu, C., Li, X., Xiao, J., and Zhang, Q. (2014). An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population. *Journal of experimental botany*, 65(4):1069–1079.

Wang, W., Cao, X. H., Miclăuș, M., Xiong, W., and Xu, J.-H. (2017). The promise of agriculture genomics. *International journal of genomics*, 2017.

Xu, Y., Jia, Q., Zhou, G., Zhang, X.-Q., Angessa, T., Broughton, S., Yan, G., Zhang, W., and Li, C. (2017). Characterization of the sdw1 semi-dwarf gene in barley. *BMC plant biology*, 17(1):11.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.-c., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature genetics*, 48(8):927–934.

Zhang, D., Li, J., Compton, R. O., Robertson, J., Goff, V. H., Epps, E., Kong, W., Kim, C., and Paterson, A. H. (2015). Comparative genetics of seed size traits in divergent cereal lineages represented by sorghum (panicoidae) and rice (oryzoidae). *G3: Genes, Genomes, Genetics*, 5(6):1117–1128.

Zhang, J., Chen, L.-L., Xing, F., Kudrna, D. A., Yao, W., Copetti, D., Mu, T., Li, W., Song, J.-M., Xie, W., et al. (2016). Extensive sequence divergence between the reference genomes of two elite indica rice varieties zhenshan 97 and minghui 63. *Proceedings of the National Academy of Sciences*, 113(35):E5163–E5171.

Zhao, H., Lou, Y., Sun, H., Li, L., Wang, L., Dong, L., and Gao, Z. (2016). Transcriptome and comparative gene expression analysis of phyllostachys edulis in response to high light. *BMC plant biology*, 16(1):34.

Zhou, T., Du, L., Wang, L., Wang, Y., Gao, C., Lan, Y., Sun, F., Fan, Y., Wang, G., and Zhou, Y. (2015). Genetic analysis and molecular mapping of qtls for resistance to rice black-streaked dwarf disease in rice. *Scientific reports*, 5.