Development of Single Nucleotide Polymorphism (SNP) Markersin Tropical Crops

Yheni Dwiningsih*, Miranti Rahmaningsih, Jawaher Alkahtani

Department of Crop, Soil and Environmental Sciences, University of Arkansas, Fayetteville, Arkansas, USA

*ydwining@uark.edu

Abstract. Understanding genetic diversity, association studies, evolution analysis, quantitative trait loci, marker-assisted selection and genome-wide association in tropical crops are important for improving plant characteristics in order to increase food sustainability in tropical countries. Single nucleotide polymorphism (SNP) marker is becoming the most popular molecular marker for those studies. By using SNP marker, genes associated with important traits can be identified efficiently compared to the other molecular markers. This review describes about how SNP can be discovered in the plant genomes and the application of SNP in plant breeding, especially in tropical crops such as rice, maize, peas, potato, tomato, cassava, taro, etc.

Keywords: food sustainability, plant breeding, SNP marker, tropical crops

1. Introduction

Gene position or genomic regions that regulate important traits in plants are discovered by using molecular markers. These markers are Single Nucleotide Polymorphisms (SNPs), Simple Sequence Repeats (SSRs), Amplified Fragment Length Polymorphisms (AFLPs), Restriction Fragment Length Polymorphisms (RFLPs), and Random Amplified Polymorphic DNAs (RAPDs). SNPs are widely used DNA markers to identify genomic regions for important traits that effectively to speed up the plant breeding. These SNPs characterized as the most abundant variations in the plant genome that are very useful in the high-resolution genotyping and produce the highest map resolution [1]. Additionally, using SNPs are more efficient and cost effective [2]. SNPs become the most popular DNA markers in the 21st century due to the development of genotyping by sequencing (GBS) technique [3]. Furthermore, important traits for plants including agronomic traits (grain yield, seed size, maturity day, etc.), morphological traits (plant height, seed color, root length, etc.), physiological traits (photosynthetic

activity, stomatal conducting, etc.), abiotic (drought, heat, flooding, salt resistance, etc.) and biotic resistance (bacterial blight, mosaic virus, root-knot resistance).

Genomic information of plants obtained from genetic diversity, association studies, evolution analysis, quantitative trait loci (QTL), marker-assisted selection (MAS) and genome-wide association studies (GWAS) are very useful for increasing food sustainability. To achieve food sustainability in this changing climate condition, plants with superior characteristics must be developed by using genetic information. Since 1900s, plant breeding has made a positive impact in increasing food production by developing superior cultivars [4, 5]. Furthermore, these superior cultivars have been developed through application effective molecular breeding by using MAS for desirable genes. These superior cultivars including increasing yield, adaptation in unstable environment, resistance to biotic and abiotic stresses, and quality improvement.
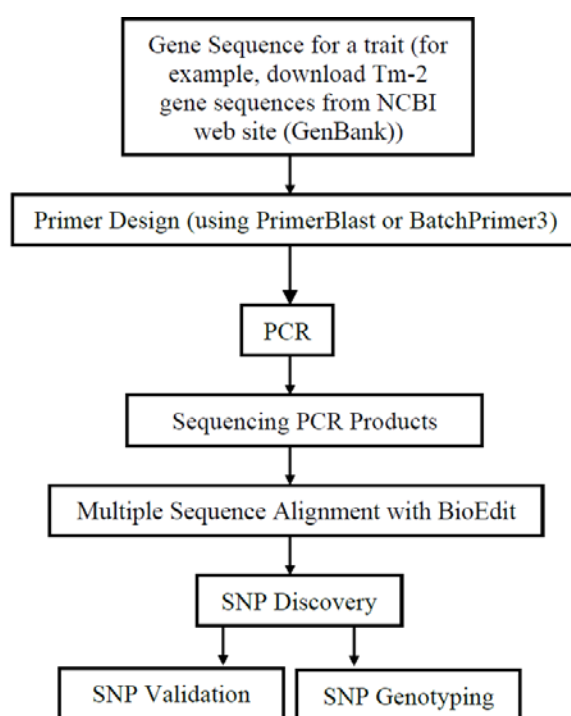
2. SNP discovery

SNP markers discovered by two methods, including SNP discovery from PCR and SNP discovery from High Throughput-Next generation sequencing (NGS) – RNA-Seq, RAD-Seq, GBS (Genome by Sequencing), WGS (Whole-Genome Sequencing), WGR (Whole-Genome Regression), etc. These approaches reduce complexity to lower the cost and simplify the discovery of SNP markers. In the past few years, NGS technologies have led to the discovery of thousands, even millions of SNPs. Moreover, there are many tools for SNP discovery such as BioEdit, DNASTAR Lasergene Genomics Suite, SAMtools, SOAPsnp, Stacks, Ddocent, Py RAD and GATK. SNPs are usually biallelic and thus easily assayed. In theory, a SNP is identified when a nucleotide from an accession read differs from the reference genome at the same nucleotide position. In the absence of a reference genome, this is achieved by comparing reads from different genotypes using *de novo* assembly strategies. Read assembly files generated by mapping programs are used to perform SNP calling. In practice, various empirical and statistical criteria are used to call SNPs, such as a minimum and maximum number of reads considering the read depth, the quality score and the consensus base ratio for examples. Moreover, SNP discovery is more robust when multiple and divergent genotypes are used simultaneously, creating the necessary basis to capture the genetic variability of a species.

Three kinds of SNP Genotyping platform: single SNP genotyping (using PCR with Taqman – Life Technologies or KASP genotyping – LGC Genomics), multiple SNP genotyping (with SNP chip – Illumina and Multiplex - Sequenom), and SNP genotyping by next generation sequencing (Genotyping by Sequencing/GBS and Restriction site associated DNA sequencing/RADSeq). In single SNP genotyping with Taqman – Life Technologies use forward primer, reverse primer and allele specific probes (VIC® labeled and 6-FAMTM labeled). Furthermore, in single SNP genotyping with Kompetitive Allele Specific PCR (KASP) genotyping – LGC Genomics are based in competitive allele-specific PCR and enable bi-allelic scoring of SNPs and insertions and deletions (Indels) at specific loci. The process in KASP genotyping is the SNP-specific KASP Assay mix and the universal KASP Master mix are added to DNA samples, a thermal cycling reaction in then performed, followed by an end-point fluorescent read. Moreover, in SNP genotyping by next generation sequencing with Genotyping by Sequencing/GBS, DNA sample info entered to webform, if the project approved, DNA samples shipped then GBS libraries made, followed by DNA sequencing data then compare with reference genome and getting data file SNPs. SNP identification in crops with GBS is more effective due to low cost and having simple methodology [6]. In SNP genotyping by next generation sequencing with Restriction site associated DNA sequencing/RADSeq can identify and score thousands of SNPs, randomly distributed across the target genome, from a group of individuals using Illumina technology (the use of restriction enzymes to cut DNA into fragments) and molecular identifiers (MID) to associate sequence reads to particular individuals. RADSeq can be used to detect restriction site presence-absence polymorphisms (by identifying a SNP that is present in one set of individuals but absent in another, indicating a variation in the restriction site).

Recent emergence of NGS technologies such as 454 Life Sciences (Roche Applied Science, Indianapolis, IN), HiSeq (Illumina, San Diego, CA), SOLiD and Ion Torrent (Life Technologies Corporation, Carlsbad, CA) has eliminated the problems associated with low throughput and high cost of SNP discovery. In addition, transcriptome resequencing using NGS technologies allows rapid and inexpensive SNP discovery within genes and avoids highly repetitive regions of a genome.

Unlike the above-mentioned techniques, recently developed genome complexity reduction technologies such as Complexity Reduction of Polymorphic Sequences (CRoPS) (Keygene N.V., Wageningen, The Netherlands) and Restriction Site Associated DNA (RAD) (Floragenics, Eugene, OR, USA) are computationally well equipped and capable of filtering out duplicated SNPs. These systems were successfully applied to discover SNPs in crops with and without reference genome sequences. Although GBS has the potential to discover several million SNPs, one of the major drawbacks of this technique is large numbers of missing data. To solve this problem, computational biologists developed data imputation models such as BEAGLE v3.0.2 and IMPUTE v2, to bring imputed data as close as possible to the real data.

Figure 1. The procedure of SNP Discovery by PCR

Table 1. Summary of commercially available SNP genotyping platforms [7]

Platform	Provider	Assay type	Technology	Throughput	Multiplexing
Taqman	Applied Biosystem	PCR	Taqman probe	~1536/day	~256
SNPlex	Applied Biosystem	PCR	Capillary electrophoresis	~1536/3 days	~48
BioMark HD	Fluidigm	PCR	Microfluid-based chips	~96/3 h	~96
KASPar	LGC	PCR	FRET quenching oligos	~96/day	Not vailable

					~96/5 days	~650K
Axiom Biobank	Affymetrix	Hybridization	Oligo nucleotide array		~96/5 days	~650K
Infinium II	Illumina	Hybridization	Bead array		~128/5 days	~700K
GoldenGate	Illumina	Primer extension	Bead array		~172/3 days	~1536
iPlex	Sequenome	Primer extension	Mass spectrometry		~3840/2.5 days	~40

**Table 2.** Representative GBS protocols published in peer-reviewed journals

Method	Restriction Enzyme	Insert size	Barcodes	Sequencing platform	Sequencing mode	Ref.
RAD-seq (Restriction association DNA sequencing)	Sbf1 or EcoR1	Size-selection	~96	Illumina	Paired-end	[8]
MSG (Multiplex shotgun genotyping)	MseI	Size-selection	~384	Illumina	Single-end	[9]
GBS (Genotypeby sequencing)	ApeKI	<350bp	~384	Illumina	Paired-end	[10]
Double-digested RAD-seq	EcoR1 and Msp1	Size-selection	~48	Illumina	Paired-end	[11]
Double-digestedGBS	Pst1 and Msp1	<350bp	~384	Illumina	Paired-end	[12]
Ion TorrentGBS	Pst1 and Msp1	<350bp	~384	Ion Torrent	Paired-end	[13]
SBG (Sequence-based genotyping)	EcoR1 and Mse1 Pst1 and Mse1	Size-selection	~32	Illumina	Paired-end	[14]
REST-seq (Restriction fragment sequencing)	Taq1 and Tru1	Size-selection	~305	Ion Torrent	Paired-end	[15]
Restrictionenzyme sequence comparativeanalysis	Mse1 or NlaIII	Size-selection	~96	Illumina	Paired-end	[16]

3. SNP Validation

The discovered SNPs must be validated to identify the true SNPs and get an idea of the percentage of potentially false SNPs resulting from an SNP discovery exercise. Validation can serve as an iterative and informative process to modify and optimize the SNP filtering criteria to improve SNP calling. For example, a subset of 144 SNPs from a total of 2,113,120 SNPs were validated using the Ilumina GoldenGate assay on 160 accessions in apple.

4. SNP Genotyping

SNP genotyping is the downstream application of SNP discovery to identify genetic variations. SNP applications including phylogenic analysis, marker-assisted selection, genetic mapping of quantitative trait loci (QTL), bulked segregant analysis, genome selection, and genome-wide association studies (GWAS).
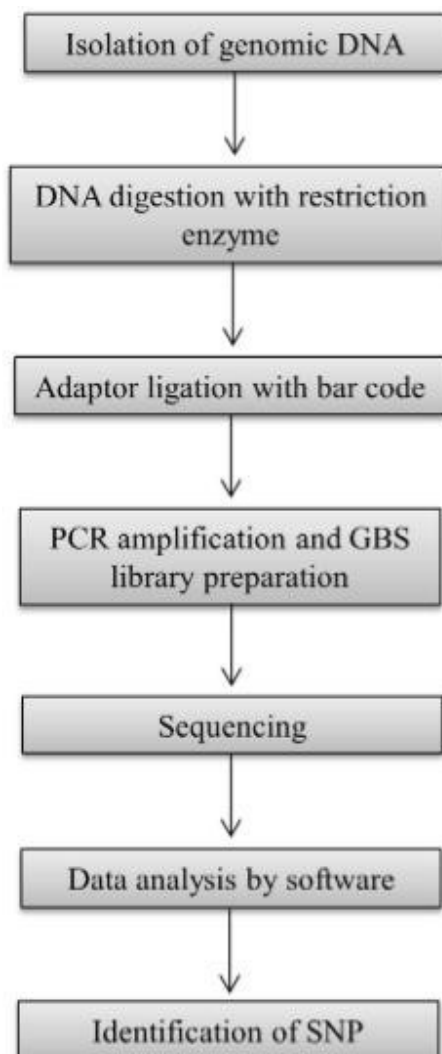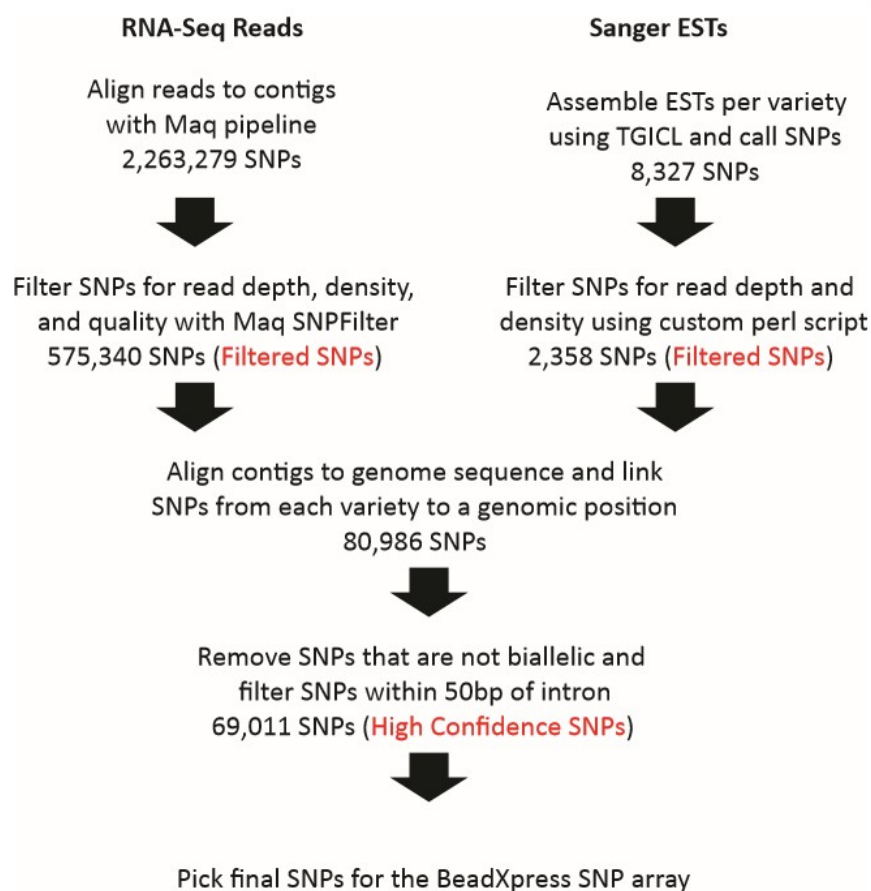

Figure 2. Overview of SNP discovery in plants through genotyping by sequencing (GBS) system [6]

In addition, analytical methods to discover novel SNPs and detect known SNPs include: capillary electrophoresis; mass spectrometry; single-strand conformation polymorphism (SSCP); single-base

extension; electrochemical analysis; denaturing HPLC (High-performance liquid chromatography) and gel electrophoresis; restriction fragment length polymorphism; and hybridization analysis.

5. SNP discovery in tropical crops

In potato (*Solanum tuberosum*), 575,340 SNPs were identified within three cultivars: 'Atlantic', 'Premier Russet' and 'Snowden' using Illumina sequencing. DNA was extracted from 248 potato lines using the Qiagen Qiaxtractor DX system (Qiagen Inc., Valencia, CA). Samples were loaded at 50 ng/µl on an Illumina BeadXpress Analyzer (Illumina inc., San Diego, CA) and data were analyzed using the Illumina GenomeStudio software. Cluster positions for three marker classes (AA, AB, and BB) were manually determined for each marker within the Illumina GenomeStudio software. The SNPs identified in this study will enable more efficient marker-assisted breeding efforts in potato.

RNA-Seq Reads

Align reads to contigs
with Maq pipeline
2,263,279 SNPs

Filter SNPs for read depth, density,
and quality with Maq SNPFilter
575,340 SNPs (Filtered SNPs)

Sanger ESTs

Assemble ESTs per variety
using TGICL and call SNPs
8,327 SNPs

Filter SNPs for read depth and
density using custom perl script
2,358 SNPs (Filtered SNPs)

Align contigs to genome sequence and link
SNPs from each variety to a genomic position
80,986 SNPs

Remove SNPs that are not biallelic and
filter SNPs within 50bp of intron
69,011 SNPs (High Confidence SNPs)

Pick final SNPs for the BeadXpress SNP array

Figure 3. The workflow of SNP discovery in potato transcriptomes and design of the BeadXpressSNP array [17]

The discovery of ~10,000 SNPs and nearly 1,000 indels in eggplant (Solanum melongena L.), equivalent to a SNP frequency of 0.8 per Kb and an indel frequency of 0.07 per Kb using RAD tag sequencing and the Illumina GoldenGate assay. This is the workflow for eggplant RAD tags sequencing and gene annotation based on Barchi et al. (2011) [18] (Figure 4).
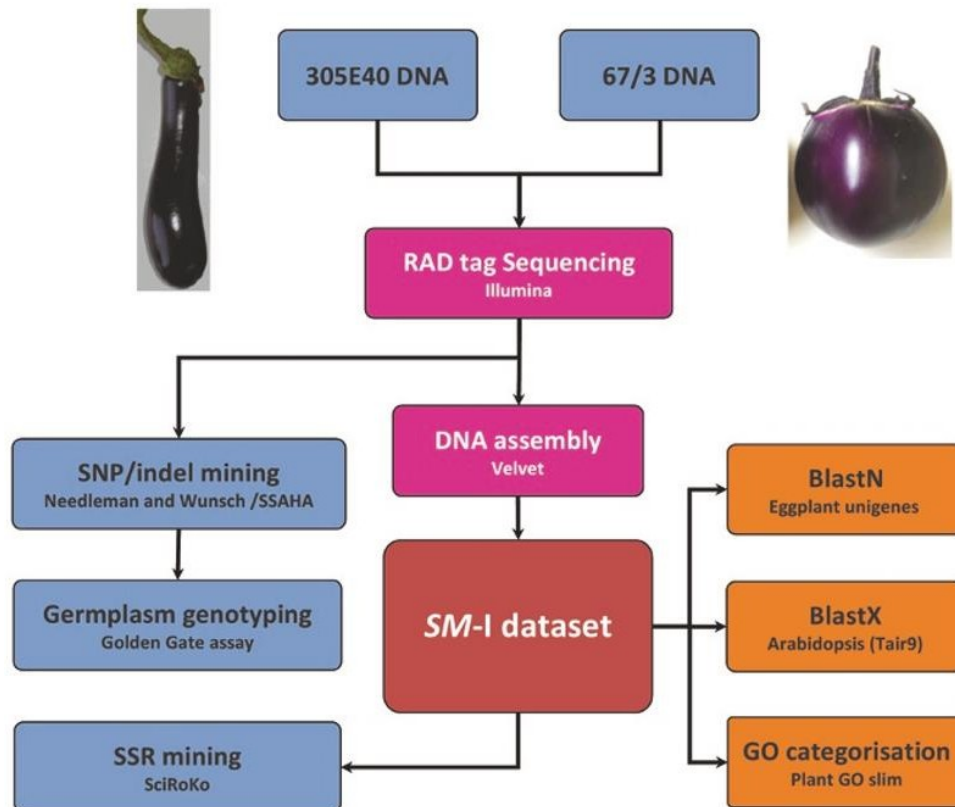
Figure 4. The workflow for eggplant RAD tags sequencing and gene annotation based on Barchi et al. (2011) [18]

A set of 100,000 to 250,000 SNPs were discovered in sorghum (Sorghum bicolor L.) by High Throughput-Next generation sequencing (NGS) – RAD-Seq [19]. In common bean (Phaseolus vulgaris L.), 827 non-genic SNPs were discovered using the GoldenGate technology of Illumina [20]. Moreover, in chickpea (Cicer arietinum L.), 1022 SNPs were identified based on Illumina GoldenGate Genotyping Technology [21]. About 533 SNPs were discovered in Camelina sativa by Illumina GoldenGate [22]. In quinoa (Chenopodium quinoa Willd.), 14,178 SNPs were identified based on KASPar genotyping chemistry and were detected using the Fluidigm dynamic array platform [23]. In rice (Oryza sativa), more than four million SNPs from around 500 rice landraces were identified by Yu et al. (2014) [24] using GBS. In Nipponbare genome 0.64 SNP was found per one kb, while in Dongjin genome contains0.45 SNP/kb. Moreover, Huang et al. (2009) [25] detected 122,791 SNPs in indica cv."9311" and japonica cv. "Nipponbare" that was average 3.2 SNPs/kb.

6. SNP Identification for a trait

SNPs are involving in genetic and genome mapping, association studies, genetic diversity analysis, and tagging important genes. In figure 5 is one of the example of association SNPs with a trait (plant height). In this example, each SNP conducts a test of association with a trait (plant height). In here, SNP (A/G) associated with plant height. In this significant SNP/trait association suggests SNP has direct biological function (functional polymorphism) and SNP in Linkage Disequilibrium with functional polymorphisms. Furthermore, association studies can determine whether a genetic variant is associated with a trait (for example, disease). Based on Shi et al. (2011) [26], SNPs have been discovered and verified in tomato and successfully used in selection of resistance to viruses, bacterial speck and bacterial spot. Below is the example of four tomatoes that resistance and susceptible to root-knot nematode (figure

6). SNP (G/C) associated with root-knot nematode. Mi is root-knot nematode resistance and mi is root-knot nematode susceptible.
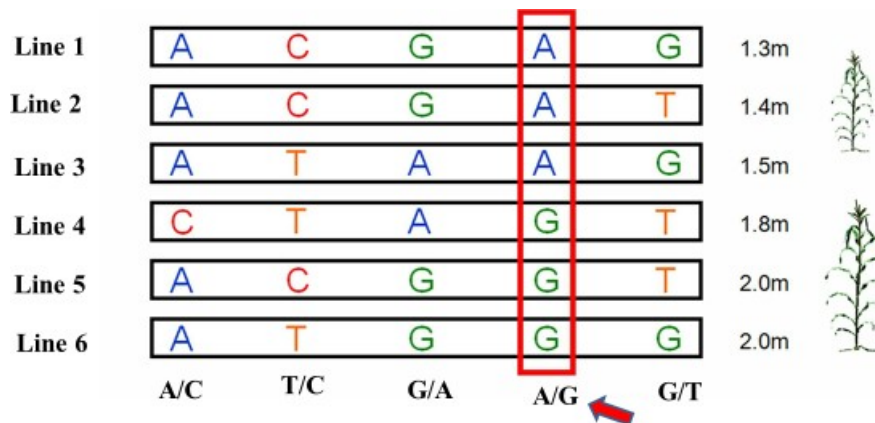
Figure 5. SNP identification for plant height

Motella/LA2823   Mi
    AAGTAGACGA**G**GTTAGTAAAAT
Mogeor/LA3471   Mi
    AAGTAGACGA**G**GTTAGTAAAAT NY07-
464   mi
    AAGTAGACGA**C**GTTAGTAAAAT
LA3130   mi
    AAGTAGACGA**C**GTTAGTAAAAT

Figure 6. SNP identification for resistance and susceptible to root-knot nematode in tomatoes [26]

7. Application of SNP markers in molecular breeding

SNP is becoming to be the most useful as molecular marker in genome mapping, comparative mapping, framework mapping, varietal/line identification, map-based cloning, genome selection, association studies, diversity analysis, genetic variation, population structure, evolution analysis, bulk segregant analysis, tagging genes for economically important traits, accelerated cloning of gene/QTL of interest, marker-assisted selection (MAS) and genome-wide association studies (GWAS) in crops because of their abundance and automated high throughput genotyping. MAS has many advantages than conventional phenotypic selection because MAS is simpler than phenotypic screening which can save time, resources and effort; selection can be carried out at the seedling stage; and single plants can be accurately selected. Moreover, SNP markers can be used in molecular breeding in population genetics, such as disease association and pathogen detection.

In tomato (Solanuum lycopersicum L. syn Lycopersicon esculentum Mill.) breeding, SNP markers were used in selection Tomato Mosaic Virus (ToMV) and Tomato spotted wilt virus (TSWV) resistance genes [11]. In this research, SNP markers were used in association study, to create a genetic test that will screen for a disease in which the disease-causing gene had already been identified. Then, collect leaves samples from a group of plants affected by the disease and analyze their DNA for SNP patterns. Furthermore, compare these patterns to patterns obtained by analyzing the DNA from group of plants unaffected by the disease. This comparison can detect differences between the SNP patterns of the two groups of plants. Therefore, SNP marker has become important and useful in the selection of disease resistance genes. These markers will provide breeders with a tool in selection of Tm-2 and Tm-22

resistance genes of ToMV in tomato breeding program. SNPs can differentiate resistance and susceptible allele at Tm-2 locus. In addition, SNP markers also can differentiate Sw5-b and Sw5-a resistance genes and sw5-b susceptible gene of TSWV.

Furthermore, in molecular breeding, SNP markers also can be used in quantitative trait locus (QTL)/gene discovery. For example, in maize (Zea mays), SNP markers have facilitated the dissection of complex traits such as flowering time by using 5000 Recombinant Inbred Lines (RILs) and genotyping with 1,200 SNP markers. Next, the genetic architecture of flowering time was discovered and controlled by small additive QTL rather than a single large-effect QTL [27]. In addition, based on Poland et al. (2011) [28], by using 5000 RILs and 1.6 million SNP markers, 29 QTL were discovered and candidate genes for northern leaf blight disease were identified. Moreover, a study from Pioneer Hi-Bred International Inc. (a private breeding program) by using their proprietary SNP markers developed by them self, reported identifying a high-oil QTL (qHO6) affecting maize seed oil and oleic acid contents. This QTL encodes an acyl-CoA:diacylglycerol acyltransferase (DGAT1-2), which catalyzes the final step of oil synthesis [30]. According to Kump et al. (2011) [30], the genetic structure of northern leaf blight, southern leaf blight, and leaf architecture was studied using ~1.6 million SNPs in maize. Additionally, five SNP primers which are associated with northern leaf blight resistant gene (Ht1) were identified by Junta et al. (2020) [31]. These SNP primers are MZSNP-0055106, MZSNP-0065744, MZSNP-0070164, MZSNP0063922, and MZSNP-0073150 that located on chromosome 2.

In rice (Oryza sativa) breeding, a GWAS was performed using ~3.6 million SNPs from ~50,000 rice accessions identified genomic regions associated with 14 agronomic traits [32]. These traits including morphological characteristics (tiller number and leaf angle), yield components (grain width, grain length, grain weight and spikelet number), grain quality (gelatinization temperature and amylose content), coloration (apiculus color, pericarp color and hull color) and physiological features (heading date, drought tolerant and degree of seed shattering). Furthermore, according to McNally et al. (2009) [33], SNPs revealed the breeding history and relationships among the 20 rice varieties; some SNPs are associated with agronomic traits that used in rice improvement. These comprehensive SNP data provide a foundation for deep exploration of rice diversity and gene–trait relationships and their use for future rice improvement. Based on Ayres (2000) [34], SNP markers were used to identify Waxy gene (control amylose synthesis by coding starch synthase enzyme) and sd-1 (semi dwarfing gene). In addition, 1536 SNP markers used for identify genetic diversity on Malaysian rice varieties [35]. Meanwhile, only 932 SNPs that have high quality alleles and amplified across the rice varieties.

In wheat (Triticum aestivum), SNP markers were mapped between the known flanking markers for Fhb1 (Fusarium head blight resistance gene). These new markers would be useful for MAS and fine mapping towards cloning Fhb1 gene [36]. Moreover, SNP markers have also been used to study the evolution of genes such as WAG-2 in wheat [37]. Identification of Pin b gene (Puroindolin b) for grain hardiness [38] and Rht1 and Rht2 (semi dwarfing gene) also using SNP markers [39].

In soybean (Glycine max), SNP markers were used to increase the efficiency and cost effectiveness through MAS and enhance the resolution within the target locus. The gene Rag1 (Soybean aphid resistance gene) was mapped between two SNP markers that corresponded to a physical distance of 115kb and identified several candidate genes [40]. In another aphid resistance gene, Rag2, originally mapped to a 10cM interval, was fine mapped to a 54kb interval using SNP markers that were developed by resequencing of target intervals and sequence-tagged sites [41]. Ha et al. (2007) [42] identified SNP markers tightly linked to a QTL conferring resistance to southern root-knot nematode by developing these SNP markers from the bacterial artificial chromosome (BAC) ends and SSR-containing genomic DNA clones. According to Shi et al. (2009) [43], SNP markers were useful in identification and selection of Soybean Mosaic Virus (SMV) resistance genes. SMV is the most destructive viral disease in soybean. Then, genetic resistance is the primary method of controlling this disease. These genes are Rsv1, Rsv3, and Rsv4. Furthermore, SNP markers were developed for those genes. There are 2 SNP markers were identified from the 3gG2 gene (the resistance allele at Rsv1 locus and the gene has been cloned and sequenced) at Rsv1 locus, 4 SNPs near Rsv1, 5 near Rsv3, and 5 near Rsv4 were validated. SNP markers also were used to identify Rhg1 and Rhg4 (Soybean Cyst Nematode resistance allele) [44].

In sugar beet (Beta vulgaris), SNP markers were developed to map QTL for Beet necrotic yellow vein virus resistance genes, Rz4 [45] and Rz5 I [46]. Additionally, in cowpea (Vigna unguiculata), a

consensus genetic map was developed based on EST-derived SNPs that a very important resource for genomic and QTL mapping studies [47]. In Arabidopsis, SNP markers were utilized to clone the VTC2 gene based on the fine mapping and map based cloning approaches. Jander et al. (2002) [48] fine mapped the gene interval from ~980kb region to a 20kb interval and additional nine candidate genes were identified in that interval and subsequently the underlying mutation was discovered.

In oil palm (Elaeis guineensis), SNP markers were used for genetic diversity [49]. The genetic evaluation of oil palm germplasm collections is important to insight into the variability among populations. The information obtained is also useful for incorporating new genetic materials into current breeding programs. Furthermore, genetic diversity information among populations is important to identify selected palms with economically important traits which are being incorporated into the current breeding programs; these include high oil yield, low height increment, large kernel, long stalk, low levels of lipase, and high levels of carotene, vitamin E, iodine, and oleic acid. In this research, 219 oil palms from two natural Angolan populations are used and a total of 62 SNP markers were designed from oil palm genomic sequences and converted to cleaved amplified polymorphic sequence (CAPS). Based on cluster analysis using unweighted pair group method with arithmetic, the 219 oil palms fell into two clusters or genetic groups that do not coincide with the geographic populations. Cluster I consisted of eight palms and cluster II included 211 palms. Furthermore, molecular variance analysis revealed that only 7% of the total genetic variation was explained by the variation between populations and 93% of the total genetic variation was attributed to variation within populations. Moreover, levels of genetic variation in plants are directly associated with breeding system. This information can be applied in selecting a sampling strategy for genetic conservation. In addition, SNP markers also use for identifying fumarate hydratase 1 (FUM1) gene located on chromosome 4, which associated with nitrogen uptake in oil palm [50].

Genetic diversity study in cassava (Manihot esculenta Crantz) was done by using 5600 informative SNP markers obtained from GBS analysis [51]. The result showed three main clusters from 96 cassava genotypes, at similarity of 0.41. This genetic diversity suggests that cassava genotypes might contains alleles with high additive genetic variance that very important for the genetic improvement, conservation, and provide database for parental selection in order to develop hybrid vigor cassava. Furthermore, genetic diversity study of 70 taro (Colocasia esculenta) accessions of Hawaiian, South Pacific, Pulauan, and mainland Asian origins was done by using 2400 SNP markers [52]. The disease resistant gene in taro was revealed by using this genetic diversity study.

In pea (Pisum sativum), SNP markers were applied in fine mapping within chosen QTL confidence intervals and marker assisted breeding for important traits in pea improvement [53]. Moreover, in eggplant (Solanum melongena), the genetic diversity revealed by SNP markers [17]. In this research, 384 of the 2,201 highest quality SNPs (score > 0.6) were applied to genotype 23 eggplant germplasms that have variation in fruit shape and skin color. About 94 SNP markers in common bean (Phaseolus vulgaris L.), were used in genetic diversity. These markers were evaluated from 70 cultivated and wild accessions according to their gene pool, race and country of origin [20]. In Chickpea (Cicer arietinum), 1022 SNPs were used to make high-resolution genetic linkage map [21]. Furthermore, in Camelina sativa, 533 SNPs were applied to mapped potential candidate genes and to assess genetic variation among a collection of 175 accessions. The SNPs will provide useful tools for future crop improvement of C. sativa as an industrial oilseed [22].

SNP markers were used on GWAS and candidate gene association (CGA) studies. A GWAS approach was applied to understand the genetic architecture of complex traits, for example complex diseases of northern and southern corn leaf blights [30]. As GWAS requires large number of molecular markers, the utility of GWAS in dissection of molecular basis of traits in polyploid crops such as wheat, and cotton has been fairly limited due to the insufficient number of polymorphic markers and the absence of reference genome.

8. Conclusion

Genomic information of plants obtained from genetic diversity, association studies, evolution analysis, quantitative trait loci (QTL), marker-assisted selection (MAS) and genome-wide association studies

(GWAS) are very useful for increasing food sustainability. To achieve food sustainability in this changing climate condition, plants with superior characteristics must be developed by using genomic information. Genes associated with important traits can be identified efficiently with SNP markers. Genomic information of many tropical crops have been identified in order to improve their characteristics.

References

[1]   McCouch, S. R., Zhao, K., Wright, M., Tung, C. W., Ebana, K., Thomson, M., Reynolds, A., Wang, D., Declerck, G., Ali, M. L., McClung, A., Eizenga, G. & Bustamante, C. Development of genome-wide SNP assays for rice. Breed Sci. 60, 524–535 (2010).

[2]   Edwards, D. & Batley, J. Plant genome sequencing: applications for crop improvement. Plant Biotech J. 8, 2–9 (2010).

[3]   Thomson, M. J. High-throughput SNP genotyping to accelerate crop improvement. Plant Breed. Biotech. 2, 3, 195–212 (2014). doi:10.9787/PBB.2014.2.3.195

[4]   Borlaug, N. E. Contributions of conventional plant breeding to food production. Science. 219, 689–693 (1983).

[5]   Tester, M. & Langridge, P. Breeding technologies to increase crop production in a changing world. Science. 327, 818–822 (2010).

[6]   Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. & Blaxer, M. L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet.12, 499–510 (2011).

[7]   Kim, C., Guo, H., Kong, W., Chandnani, R. & Shuang, L. Application of genotyping by sequencingtechnology to a variety of crop breeding programs. Plant Science. 242, 14–22 (2016).

[8]   Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PlosOne3. E3376 (2008).

[9]   Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T.T., Mast, J., Sunayama-Morita, T. & Stern, D.L. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. 21, 610–617(2011).

[10]   Elshire, R. J., Glaubitz, J. C., Qun, J., Poland, J. A., Kawamoto, K., Buckler, E. S., Mitchell, S. E.& Robust, A. Simple genotyping by sequencing (GBS) approach for high diversity species. PlosOne. e19379 (2011).

[11]   Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PlosOne. 7. E37135 (2012).

[12]   Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PlosOne. 7. E32253 (2012).

[13]   Masher, M., Wu, S., Amand, P. S., Stein, N. & Poland, J. Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PlosOne. 8. E76925 (2013).

[14]   Truong, H. T., Ramos, A. M., Yalcin, F., de Ruiter, H. J., van deer Poel, et al. Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PlosOne.7. E37565 (2012).

[15]   Stolle, E. & Moritz, R. F. RESTseq-efficient benchtop population genomics with RESTriction Fragmentt SEQuencing. PlosOne. 8. E63960 (2013).

[16]   Manson-Miller, J., Sanchez-Mendez, D. C., Fass, J., Henry, I. M., Tai, T. H. & Comai, L. Reference genome-independent assessment of mutation density using restriction enzyme-phase sequencing. BMC Genomics. 13, 72 (2012).

[17]   Hamilton, J. P., Hansey, C. N., Whitty, B. R., Stoffel, K., Massa, A. N., Deynze, A. V., De Jong, W, S., Douches, D. S. & Buell, C. R. Single nucleotide polymorphism discovery in elite north

american potato germplasm. BMC Genomics. 12, 302 (2011). doi:10.1186/1471-2164-12-302

[18] Barchi, L., Lanteri, S., Portis, E., Acquadro, A., Vale, G., Toppino, L. & Rotino, G. L. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. BMC Genomics. 12, 304 (2011).

[19] Nelson, J. C., Wang, S., Wu, Y., Li, X., Antony, G., White, F. F. & Yu, J. Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. BMC Genomics. 12, 352 (2011).

[20] Cortes, A. J., Chavarro, M. C. & Blair, M. W. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). Theoreticl and Applied Genetics. 123, 827–845. (2011). doi:10.1007/s00122-011-1630-8

[21] Gaur, R., Azam, S., Jeena, G., Khan, A. W., Choudhary, S., Jain, M., Yadav, G., Tyagi, A. K., Chattopadhyay, D. & Bhatia, S. High-throughput SNP discovery and genotyping for constructing asaturated linkage map of chickpea (*Cicer arietinum* L.). DNA Research. 19, 357–373. (2012). doi:10.1093/dnares/dss018

[22] Singh, R., Bollina, V., Higgins, E. E., Clarke, W. E., Eynck, C., Sidebottom, C., Gugel, R., Snowdon, R. & Parkin, I. A. P. Single-nucleotide polymorphism identification and genotyping in *Camelina sativa*. Molecular Breeding. 35, 35 (2015).

[23] Maughan, P. J., Smith, S. M., Rojas-Beltrian, J. A., Elzinga, D., Raney, J. A., Jellen, E. N., Bonifacio, A., Udall, J. A. & Fairbanks, D. J. Single nucleotide polymorphisms identification, characterization and linkage mapping in *Chenopodium quinoa*. The Plant Genome (2012). doi:10.3835/plantgenoe2012.06.0011

[24] Yu, H., Xie, W., Li, J., Zhou, F. & Zhang, Q. A whole-genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol J. 12, 28–37 (2014).

[25] Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T. & Han, B. High-throughput genotyping by whole-genome resequencing. Genome Res. 19, 1068–1076 (2009).

[26] Shi, A., Vierling, R., Grazzini, R., Chen, P., Caton, H. & Panthee, D. Molecular Markers for *Tm-2*Alleles of Tomato Mosaic Virus Resistance in Tomato. American Journal of Plant Sciences. 2, 180–189 (2011).

[27] Buckler, E. S., Holland, J. B., Bradbury, P. J., et al. The genetic architecture of maize flowering time. *Science,* 325, 5941: 714-718. (2009).

[28] Poland, J. A., Bradbury, P. J., Buckler, E. S. & Nelson, R. J. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proceedings of the National Academy of Sciences of the United States of America. 108, 17, 6893–6898 (2011).

[29] Zheng, P., Allen, W. B., Roeslar, K., et al. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nature Genetics. 40, 3, 367–372 (2008).

[30] Kump, K. L., Bradbury, P. J., Wisser, R. J., et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nature Genetics. 43, 2, 163–168 (2011).

[31] Junta, K., Saritnum, O., Juwattanasamran, R., Puddhanon, P., Struss, D., Puttarach, J. & Songchanthiek, S. Identification of single nucleotide polymorphism markers associated with northern corn leaf blight resistance in sweet corn. Chiang Mai University (CMU) Journal of NaturalSciences. 19, 1, 155–162 (2020).

[32] Huang, X., Wei, X., Sang, T. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Genetics. 42, 11, 961–967 (2010).

[33] McNelly, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., Zeller, G., Clark, R. M., Hoen, D. R., Bureau, T. E., Stokowski, R., Ballinger, D. G., Frazer, K. A., Cox, D. R., Padhukasahasram, B., Bustamante, C. D., Weigel, D., Mackill, D. J., Bruskiewich, R. M., Ratsch, G., Buell, C. R., Leung, H. & Leach, J. E. Genomewide SNP variation reveals relationships amonglandraces and modern varieties of rice. PNAS (2009). doi: 10.1073.pnas.0900992106

[34] Ayres, N. M., Anna, M. M., Larkin, P. D., Bligh, H. F. J., Jonnes, C. et al. Microsatellites and a

single nucleotide polymorphism differentiate apparent amylase classes in an extended pedigree of US rice germplasm (2000). http://www.nal.usda.gov/ttic/tektran/data/000007/55/0000075521.html.

[35] Razak, A. S., Nor Azman, N. H. E., Kamaruzaman, R., Saidon, S. A., Mohd Yisof, M. F., Ismail, S. N., Jaafar, M. A. & Abdullah, N. Genetic diversity of released Malaysian rice varieties based onsingle nucleotide polymorphism markers. Czech J. Genet. Plant Breed. 56, 62–70 (2020).

[36] Gupta, P. K., Langridge, P. & Mir, R. R. Marker-assisted wheat breeding: present status and future possibilities. Molecular Breeding. 26, 2, 145–161 (2010).

[37] Wei, O., Peng, Z., Zhou, Y., Yang, Z., Wu, K. & Ouyang, Z. Nucleotide diversity and molecular evolution of the WAG2 gene in common wheat (*Triticum aestivum* L.) and its relatives. Genetics and Molecular Biology. 34, 4, 606–615 (2011).

[38] Huang, X. & Coder, M. S. Development of SNP assays for genotyping the Puroindoline b gene forgrain hardiness in wheat using pyrosequencing. J Agric Food Chem. 53, 2070–2075 (2005).

[39] Ellis, M., Speilmeyer, W., Gale, K., Rebetzke, G. & Richards, R. "Perfect" markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. Theor. Appl. Genet. 105, 1038–1042 (2005).

[40] Kim, K. S., Bellendir, S., Hudson, K. A., et al. Fine mapping the soybean aphid resistance gene*Rag1* in soybean. Theoretical and Applied Genetics. 120, 5, 1063–1071 (2010a).

[41] Kim, K. S., Hill, C. B., Hartman, G. L., Hyten, D. L., Hudson, M. E. & Diers, B. W. Fine mapping of the soybean aphid-resistance gene *Rag2* in soybean PI200538. Theoretical and Applied Genetics.121, 3, 599–610 (2010b).

[42] Ha, B. K., Hussey, R. S. & Boerma, H. R. Development of SNP assays for marker-assisted selection of two southern root-knot nematode resistance QTL in soybean. Crop Science. 47, 2, 73–82 (2007).

[43] Shi, A., Chen, P., Vierling, R., Cervantes, I. & Shakiba, E. Development of multiplex SNP assays for selection of soybean mosaic virus resistance genes in soybean. World Soybean Research Conference VIII, August 10-15, 2009, Beijing, China (2009).

[44] Gupta, P. K., Roy, J. K. & Prasad, M. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr. Sci. 80, 524–535 (2001).

[45] Grimmer, M. K., Trybush, S., Hanley, S., Francis, S. A., Karp, A. & Asher, M. J. C. An anchored linkage map for sugar beet based on AFLP, SNP and RAPD markers and QTL mapping of a new source of resistance to Beet necrotic yellow vein virus. Theoretical and Applied Genetics. 114, 7, 1151–1160 (2007).

[46] Grimmer, M. K., Kraft, T., Francis, S. A. & Asher, M. J. C. QTL mapping of BNYVV resistance from the WB258 source in sugar beet. Plant Breeding. 127, 6, 650–652 (2008).

[47] Muchero, W., Diop, N. N., Bhat, P. R., et al. A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proceedings of the National Academy of Sciences of the United States of America. 106, 43, 18159–18164 (2009).

[48] Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M. & Last, R. L. Arabidopsis map-based cloning in the post-genome era. Plant Physiology. 129, 2, 440–445. (2002).

[49] Ong, P. W., Maizura, I., Abdullah, N. A. P., Rafii, M. Y., Ooi, L. C. L., Low, E. T. L. & Singh, R. Development of SNP markers and their application for genetic diversity analysis in the oil palm (*Elaeis guineensis*). Genetics and Molecular Research. 14, 4, 12205–12216 (2015).

[50] Maryanto, S. D., Tanjung, Z. A., Sudania, W. M., Kusnandar, A. S., Roberdi, Pujianto, Utomo, C. & Liwang, T. Polymorphism of *Fumarate Hydratase* 1 (FUM1) gene associated with nitrogen uptake in oil palm (*Elaeis guineensis*). Biodiversitas, 21, 6, 2462–2466 (2020).

[51] Karim, K. Y., Ifie, B., Dzidzienyo, D., Danquah, E. Y., Blay, E. T., Whyte, J. B. A., Kulakow, P., Rabbi, I., Parkes, E., Omoigui, L., Norman, P. E. and Iluebbey, P. (2020). Genetic characterization of cassava (*Manihot esculenta* Crantz) genotypes using agro-morphological and single nucleotide polymorphism markers. *Physiol Mol Biol Plants,* 26, 2: 317-330.

[52] Helmkampf, M., Wolfgruber, T. K., Bellinger, M. R., Paudel, R., Kantar, M. B., Miyasaka, S. C.,

Kimball, H. L., Brown, A., Veillet, A., Read, A. and Shintaku, M. (2018). Phylogenetic relationships, breeding implications, and cultivation history of Hawaiian taro (*Colocasia esculenta*) through genome-wide SNP genotyping. *Journal of Heredity,* 272-282. doi:10.1093/jhered/esx070

[53] Boutet, G., Carvalho, S. A., Falque, M., Peterlongo, P., Lhuillier, E., Bouchez, O., Lavaud, C., Pilet-Nayel, M., Riviere, N., and Baranger, A. (2016). SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *Genomics*, 17: 121. doi:10.1186/s12864-016-2447-2