We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding

Hande Morgil, Yusuf Can Gercek and Isil Tulum

Abstract

Recent advances in genome technology revealed various single nucleotide polymorphisms (SNPs), the most common form of DNA sequence variation between alleles, in several plant species. The discovery and application of SNPs increased our knowledge about genetic diversity and a better understanding on crop improvement. Natural breeding process which takes an agelong time during collecting, cultivating, and domestication has been accelerated by detecting dozens of SNPs on various species using advanced biotechnological techniques such as next-generation sequencing. This will result in the improvement of economically important traits. Therefore, we would like to focus on the discovery, current technologies, and applications of SNPs in breeding. The chapter covers the following topics: (1) introduction, (2) application of SNPs, (3) techniques to detect SNPs, (4) importance of SNPs for crop improvement, and (5) conclusion.

Keywords: SNPs, SNP identification, plant evolution, crop improvement

1. Introduction

Understanding the distribution and diversity of plant species is increasingly important to meet the demands of the growing population. Loss and deterioration have been observed in agricultural lands due to reasons such as salinization, environmental pollution, urban growth, temperature, and global climate change [1]. Prehistoric people were able to transform crops into crops that serve food for humanity by using traditional treatment techniques. These cultivated plants, when compared to their wild relatives, differed due to such characteristics that affected the plant breeders directly such as flowering time, sizes of the reproductive organs, and seed loss.

Significant improvements occurred in the productivity of agricultural products in the last century [2]. However, there are still areas in need of improvement. Today, there is still a great need for new genotypes in agricultural lands due to various social and cultural changes. Plant producers have to comply with the market needs, consumer demands, and growing agricultural problems. While much of the progress made so far has been achieved through classical improvement techniques, future prospects are under the control of biotechnology as a basic condition for achieving greater probability of success in product development [3]. In biotechnology, the study and use of DNA markers for plant breeding are encouraging for the

future [4]. The use of DNA markers associated with crop yield is common in the development of various crops such as rice (*Oryza sativa*) [5], corn (*Zea mays*) [6], wheat (*Triticum aestivum*) [7], and tomatoes (*Lycopersicon esculentum*) [8].

If a single nucleotide change is detected by comparing the DNA of different living species, it is evaluated as there is a single nucleotide polymorphism. These changes in a single position are used as an effective genetic marker practically in both animal [9] and the plant [10, 11] species. Single nucleotide polymorphism (SNP) genotyping [12, 13] studies and the rapid progress in the development of genomic tools have led to the development of new powerful approaches in mapping complex features and identifying the causes.

In parallel with the increase in multidisciplinary studies and the development of technology, it is essential to use both traditional breeding techniques and new tools emerging in the field of molecular genetics [14]. In these tools, the two most used methods in terms of the low costs and high performance in obtaining data are microarrays and next-generation sequencing.

Since the end of the twentieth century, microarrays have been used in the first place to know the transcriptional activity of a biological sample [15]. Although other techniques have been previously used in gene expression studies such as northern blot or quantitative PCR, the ability to determine the level of less represented genes of a mixture facilitated the analysis of thousands of genes in the same reaction and increased sensitivity [16]. In the next-generation sequencing, the main goal is to parallelize DNA sequencing so that the molecules of thousands or millions of genetic materials could be read simultaneously. Regardless of the technique used, it identifies a large number of markers, allowing the development of high-density genetic maps [17]. This technology has been successfully used to detect SNPs of different genetically well-known species such as pine or corn [18, 19].

The wealth of data required to reveal evolutionary processes is based on highly efficient DNA sequencing. This technology enables nucleotide diversity studies related to a wide variety of species. The determination of the functionality of the genes of the wild species that have increased and continued in recent years and the presence of beneficial alleles for indirect plant breeding and yield improvement studies still make up an important topic for the future that is open for further improvement.

2. Applications of SNPs

Determining the DNA sequence variation in the genome is very important for plant genetics and breeding. Genetic variation can be analyzed using various molecular markers. The discovery of single nucleotide polymorphisms (SNPs), which underpin the difference between alleles, has been simplified and accelerated by recent advances in next- and third-generation sequencing technology and MALDI-TOF mass spectrophotometry compared to traditional methods [20]. Even creating machine learning models to select true SNPs directly from sequence data appears to be groundbreaking in this area [21].

The selection of SNPs enables the selection of desired lines in large-scale populations. The marker can be used to modulate the cultivation program for the determination of the relevant feature and improvement of the crop more economically using new-generation technologies than using traditional methods [22]. Today, plant breeding is dependent on SNPs and similar differences for fast and cost-effective analysis of germplasm and feature mapping. These differences improve the understanding of genetics that can change the strategy of developing new varieties.

Because the desired trait is under genetic control, phenotypic experiments can be attempted faster, and the breeder not only does the early trait selection but also can transmit the desired allele to a large number of populations [23].

As genomes of many species are fully sequenced, including human, *Arabidopsis*, and rice, the discovery of interest-specific sequence differences becomes easier [24]. These sequences enable screening of more than 1 million SNPs for each species. These polymorphisms can be used as simple genetic markers that can be identified around almost any gene. The usage of SNPs in detecting relationships between allelic forms of a gene and phenotypes, especially common diseases with multifactor genetics, high-resolution genetic map construction, linkage disequilibrium-based association mapping, genetic diagnostics, genetic diversity analysis, cultivar identification, and phylogenetic analysis, creates great potential for characterization of genetic resources [25]. SNPs are frequently found in the genome with at least one common (>20% allele frequency) SNP density per kilobase pair. It is mostly biallelic and therefore easy to analyze. More importantly, SNPs allow the combination of candidate gene approach and fusion-based fine mapping to identify the genes of interest [26].

The application of SNPs on genetic diversity is very important in terms of illuminating the relationships between varieties. This allows plant growers to improve crop plants and protect germplasm. Genetic diversity information can then be evaluated to identify new alleles in breeding programs. SNPs have been applied for several years to assess diversity in specific genes or genomic regions, and the results are estimated to extract phylogenetic relationships between species. However, the emergence of new and third-generation technologies allows the genome to assess SNP-based genetic diversity at scale and can be useful in conserving diversity in domesticated populations. Plant phylogenetic and evolutionary researches is traditionally based on sequence genes and hence the knowledge of SNPs [27]. Nuclear and chloroplast gene regions, which have been preserved for generations, are a rich source of phylogenetic information for evolutionary analysis in plants. The diversity and genotyping of the SNP sequence in these protected regions is used to explain a wide variety of phylogenetic and evolutionary relationships and inheritance extraction [28]. However, molecular phylogenetic studies only provide information about the distribution of populations, and contribution to agricultural applications is negligible compared to SNPs determined by transcriptome analyzes [29].

SNPs can also be used to discover new genes and their functions by affecting gene expression and transcriptional and translational promoter activities. Therefore, they may be responsible for phenotypic variations between individuals in improving agronomic features. It is also important to know the location of SNP in the genome, because if SNP is present in the coding region, it can greatly affect the activity and thermostability level of an enzyme or a similar product [30]. Sometimes it also depends on the substituted amino acid positions because some amino acid controls the activity of the expressed regions. Recent technological advances make it easier to identify various SNPs that can be used for product developments. It shows that SNPs in the functional parts of the gene can control the level of biotic and abiotic stress and improve the variety of abiotic and biotic stress tolerance crops by changing expressed region [31].

The integration of genomic technologies with traditional breeding can have a big impact in dealing with current and future environmental challenges more effectively [32]. In these conditions, germplasm in all plant species is imperative for rapid genetic gains in the productivity of these species using supportive approaches such as genetic and genomic sources.
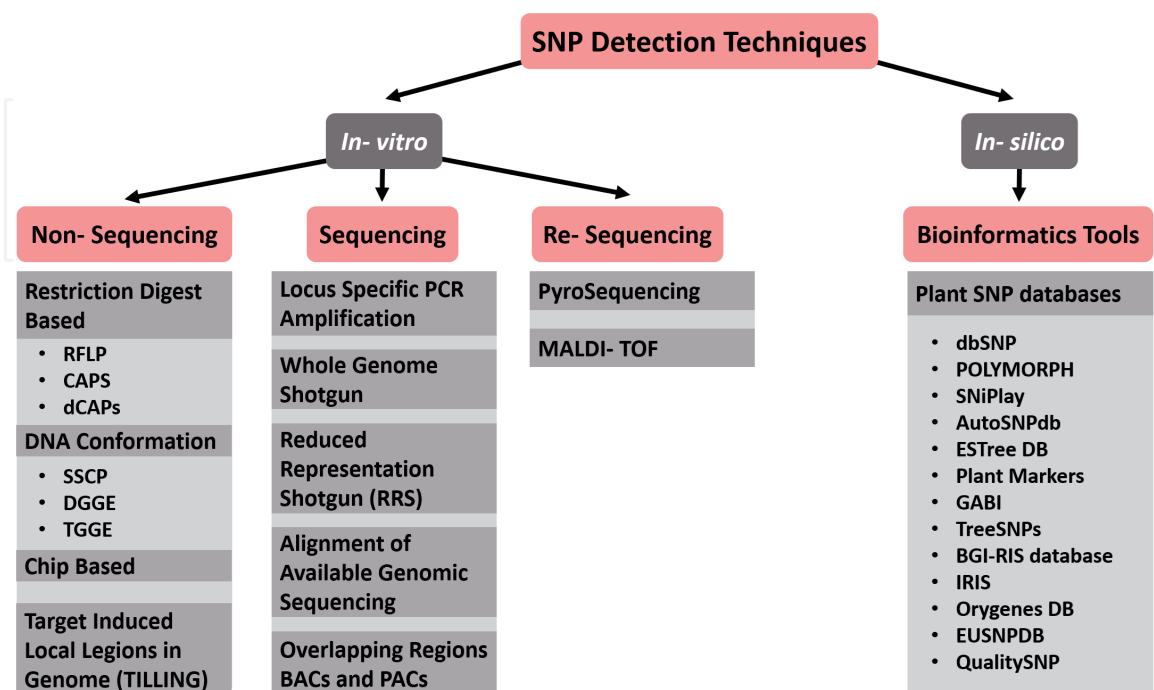
3. Techniques to detect SNPs

The choice of the methods for SNP detection is diverse. The SNP detection technologies have been evolved with the discovery of new techniques on reporter systems, fluorescent probes, development of enzymatic assays, use of highly sensitive instruments, and mostly the accelerated high-throughput sequencing technology and bioinformatic tools. In the post-genomic era, the accuracy and sensitivity of the detection methods have increased with a cost-effective manner.

The basic idea behind SNP detection is whether identifying a novel polymorphism that is previously not defined or searching for an already-known polymorphism. The techniques for detection can be divided into two main groups: (i) in vitro and (ii) in silico techniques (**Figure 1**). In vitro techniques comprise of non-sequencing, sequencing, and re-sequencing methods.

3.1 In vitro techniques

3.1.1 Non-sequencing techniques

The firstly developed non-sequencing techniques are restriction digestion-based techniques such as restriction fragment length polymorphisms (RFLPs), cleaved amplified polymorphic sequences (CAPs), and derived cleaved amplified polymorphic sequences (dCAPs). These techniques mainly aim to create or disrupt a restriction enzyme recognition site [33]. Another group of non-sequencing technique is DNA conformation techniques which comprise denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), and single-strand conformation polymorphism (SSCP). The basis of these techniques is the separation of DNA fragments of the same length with different base composition on their three-dimensional conformation [34]. The chip-based methods are based on DNA hybridization like DNA microarrays and rely on the biochemical principle of nucleotide complementation. Affymetrix and Illumina SNP Chips use

Figure 1.
Techniques to detect SNPs. In vitro techniques include non-sequencing, sequencing, and re-sequencing methods. In silico techniques basically includes bioinformatic tools with several output-/input-oriented algorithms.

the technology to hybridize fragmented single-stranded DNA to arrays containing thousands of nucleotide probe sequences that are designed to bind to a target DNA sequence [35]. Target-induced local lesions in genome (TILLING) is a reverse genetics approach that combines chemical mutagenesis with a sensitive mutation detection instrument called denaturing HPLC (DHPLC) [36]. The need of several optimization steps and "hundreds of bases"-long probes to detect only a small fraction of the region of interest made the non-sequencing methods very laborious and expensive. However, several newly developed approaches provide greater efficiencies. From all of these methodologies, direct DNA sequencing technologies are considered as the most used and benefited for SNP detection.

3.1.2 Sequencing techniques

One of the first designed sequencing-based techniques for SNP detection is locus-specific PCR amplification. In this approach, a large number of loci are targeted using locus-specific PCR primers and through conducting direct sequencing of genomic PCR products. Another sequencing-based technique is reduced representation shotgun (RRS). This method is based on the migration pattern of genomic segments of the same origin with the same size in gel electrophoresis [37]. Comparison in overlapped regions of bacterial artificial chromosome (BAC) or P1-derived artificial chromosome (PAC) clone regions is another sequencing-based approach for SNP discovery [38].

3.1.3 Re-sequencing techniques

Beside these techniques, there are re-sequencing approaches including matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF/MS) and pyrosequencing. MALDI-TOF/MS is based on differentiating genotypes by comparing the mass of DNA fragments after a single ddNTP primer extension reaction. This technique does not require labeling, and the detection depends on the mass of the ddNTP that is incorporated [39]. Pyrosequencing is a rapid re-sequencing approach in which the sequencing is performed by detecting the nucleotide incorporated by a DNA polymerase and monitored by a measure of pyrophosphate (PPi) release [40].
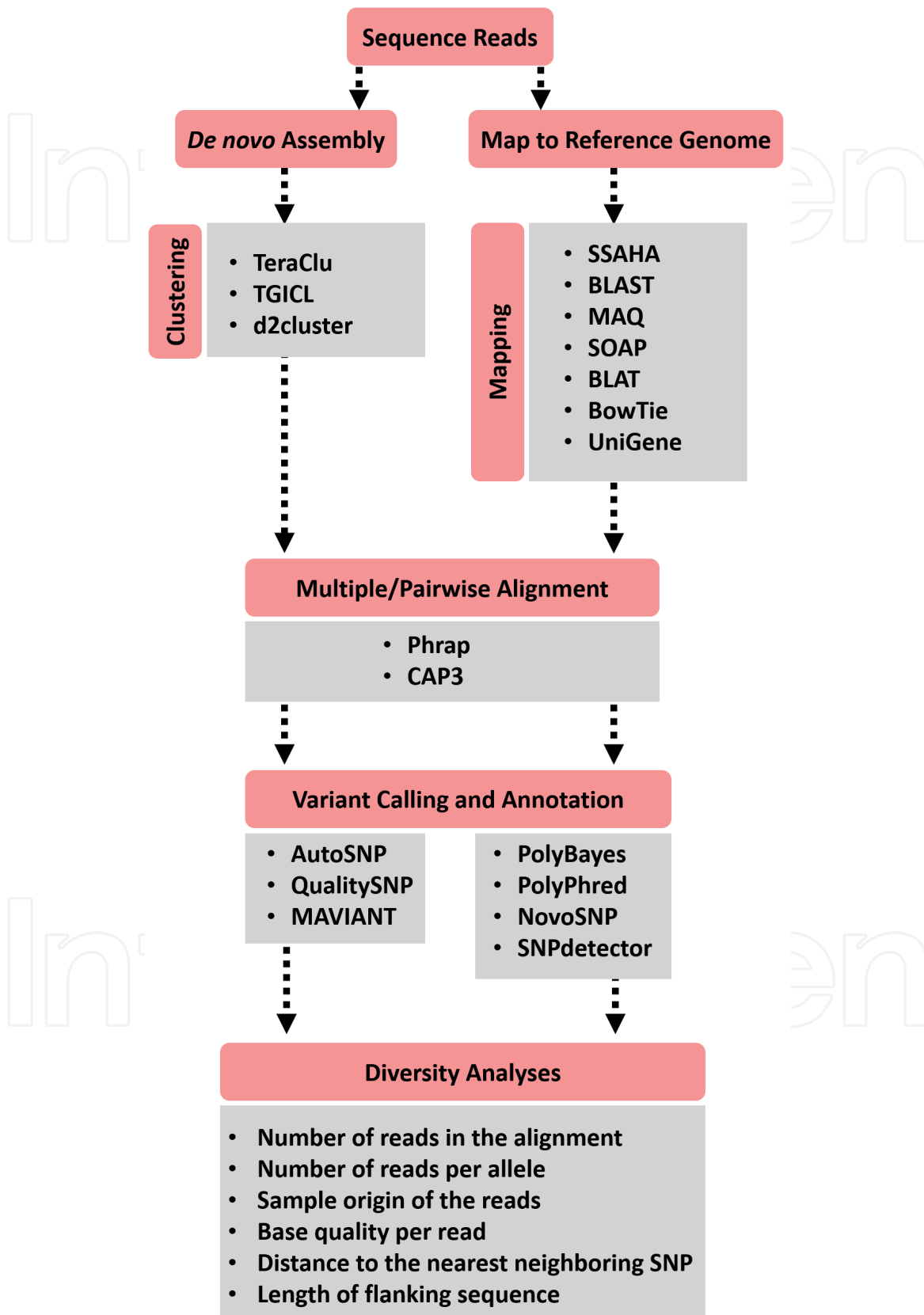
3.2 In silico techniques

As mentioned above, many experimental strategies are currently available for SNP detection. In vitro SNP detection methods are often composed of several laborious steps or require specialized instruments which makes the process high-cost and compound.

The developments in the sequencing technologies have resulted in decreasing cost along with rapid progress in next-generation sequencing (NGS) and related bioinformatic computing resources. These developments accelerated the whole-genome association studies (WGAS) and the identification of many new SNPs in model and non-model plants. In the post-genomic era, SNPs became the commonly used marker systems in many plants with several advantages such as stability, ease of use, considerably low mutation rates, and high-throughput genotyping [41]. NGS platforms generate a considerable amount of data in which results in the urge of alternative data storage methods and shorter processing time.

In silico methods are easy to apply to the SNPs that are occurring in known genomes or sequences of a species of interest. Bioinformatic research is constantly developing online and stand-alone tools, new software, and algorithms to analyze

the SNPs. The recently developed open-source and freely available bioinformatic software have speed up the SNP detection and reduced the costs. The important point is the selection of the software, sequence platform, file requirement,

```
                        ┌──────────────────────┐
                        │   Sequence Reads     │
                        └──────────────────────┘
            ┌───────────────────────┐   ┌──────────────────────────┐
            │   De novo Assembly    │   │ Map to Reference Genome  │
            └───────────────────────┘   └──────────────────────────┘
```

Sequence Reads

De novo Assembly → **Map to Reference Genome**

Clustering
- TeraClu
- TGICL
- d2cluster

Mapping
- SSAHA
- BLAST
- MAQ
- SOAP
- BLAT
- BowTie
- UniGene

Multiple/Pairwise Alignment
- Phrap
- CAP3

Variant Calling and Annotation
- AutoSNP
- QualitySNP
- MAVIANT
- PolyBayes
- PolyPhred
- NovoSNP
- SNPdetector

Diversity Analyses
- Number of reads in the alignment
- Number of reads per allele
- Sample origin of the reads
- Base quality per read
- Distance to the nearest neighboring SNP
- Length of flanking sequence

Figure 2.
A common workflow schema for in silico SNP mining. According to the input data type, the steps and the algorithms change. If the data set is de novo assembly output, clustering step is needed. If the data set is based on a reference mapped output, mapping step is required. In order to mine possible SNPs, after clustering and mapping steps, alignment, variant calling, annotation, and diversity analysis should be applied, respectively.

algorithmic background, operating systems, and organism of interest affect the choice of bioinformatic platform or pipeline to use. There are many databases and resources available today to describe SNPs in many plants.

The methods used for SNP mining will be quite similar for both database-derived and high-throughput sequencing-derived data. NGS technologies, Illumina GA/Solexa, SOLiDTM, Oxford Nanopore high-throughput sequencing, generate large amount of sequence data therefore many new SNPs. The method of choice may vary with different source data and varying approach. There are different analysis steps which apply to the two types of sequence data: reference sequence data where sequence data is acquired from species for which a reference sequence is accessible and de novo sequence data. In either case there are three main steps: (i) group the sequence reads according to their sequence similarities, and confirm identity of reads covering the same part of the genome or having the same transcript origin; (ii) align the reads; and (iii) scan for sequence variants (**Figure 2**).

If a reference sequence is available, the first step will be determining a homology search tool to map the new sequence reads to the reference. There are several tools for global or local mapping such as BLAST and SSAHA for whole-genome data. If the short reads are derived from Illumina, specially developed tools such as SOAP and MAQ are available. UniGene set is designed for mapping transcript data. The next step is to select a multiple or pairwise alignment tool to align the mapped reads to the reference sequence. Software tools like CAP3 and Phrap have been extensively used for this purpose.

In the case of de novo sequence data, an additional step called as clustering is needed to group the short sequence reads. TeraClu, TGICL, and d2cluster are the mostly used tools to fragmentate the input data and assemble them into individual contigs. After the clustering step, all nucleotides from individual reads at the identical position on the gene are aligned similarly using CAP3 and Phrap.

The final step is the SNP calling or validation. If the fragments are from a trace file or a base quality score, PolyBayes, PolyPhred, novoSNP, and SNPdetector are very well-known tools. If de novo SNP mining is performed, AutoSNP, QualitySNP, and MAVIANT can be used. There are several SNP mining tools or databases available specialized for plants such as dbSNP, ESTree DB, POLYMORPH, SNiPlay, AutoSNPdb, IRIS, etc. Although these tools are the frequently used, reliable and accurate tools, new tools, and platforms are being developed.

4. Importance of SNPs for crop improvement

Single nucleotide polymorphism (SNP) causes genetic diversity among individuals of the species and can occur at different frequencies in different species throughout the entire genome. SNPs can cause phenotypic diversity among individuals such as the color of different plants or fruits, fruit size, ripening, flowering time adaptation, crop quality, grain yield, or tolerance to various abiotic and biotic factors [42]. While SNPs can cause changes in amino acids in the exon of a gene, it can also be silent. In addition, it can occur in noncoding regions. SNPs can influence promoter activity for gene expression and finally produce a functional protein by transcription. Therefore, identifying functional SNPs in genes and determining their effects on the phenotype can lead to a better understanding of the effects on gene function for product development [43].

Conventional breeding and marker-assisted breeding are two approaches used to perform plant breeding [42, 44]. However, in plant breeding, publications on the application of molecular markers compared to conventional breeding have increased significantly over the past 15 years. Plant breeding forms and will

continue to be the basis for increasing scientific efficiency in the field of food, feed, and industry. The reason why conventional breeding is increasingly preferred is that it requires hybridization between various parents and then selection over a long time (5–15 years) generation based on phenotypic selection to obtain the advanced product [45].

Rapid progress in sequence technologies, including SNP genotyping and genome sequencing, has given new and powerful approaches to mapping complex features and then identifying genes that cause this complexity. Although these methods are first applied in human genetics, their applications in plant genetics and product development are becoming popular [46]. With these new techniques, it is an important advantage to create experimental populations of germplasm collections and homozygous individuals in plants in a short time.

The most obvious advantages of SNP markers are that they are flexible and fast and provide data management convenience. For example, biallelic SNP markers are easy to combine data between groups and create large databases of this data because there are only two alleles for each location. This will also ensure the same allele detection on different genotyping platforms after appropriate quality assessment of these data. Using bioinformatic tools to transform SNP markers from different studies into the same DNA chain can contribute to improvement efforts [47]. With the help of a high-quality reference genome, the fusion sequence and SNPs also provide a stronger analysis of the entire SNP catalog for each species. As the most common type of DNA polymorphism, SNPs can also be specific at the genome-wide locus, which can reveal the selection of SNP variants at the target locus as well as informative marker sets for specific germplasm pools.

Due to the availability of technologies that provide validation and detection of SNPs, the development of SNP markers has become a routine process, especially in products with a reference genome. Appearance of whole-genome sequencing (GWAS) and de novo sequencing of unknown genomes has emerged. However, since 2006, it has been observed that SNPs have been used in publications derived from academic researches that have resulted in the development of crops such as water-tolerant rice, rust-resistant wheat varieties "Patwin," and low phytic acid corn and, briefly, in order to correct the problems in agriculture. Although it does not normally disclose the details of reproductive methodologies to the public, it is known that SNP tokens have been applied in several articles published by companies such as Monsanto, Pioneer Hi Bred International, and Syngenta [48].

The effects on plant protein function and gene expression against a changing condition may result from SNPs occurring in coding regions and regulatory sequences, respectively. Therefore, SNPs have a great potential in genetic, reproductive, and economic importance [49].

Thus, the targeted product can be developed using only the data of the databanks before the fieldwork to efficiently cultivate a crop on logical and evolutionary studies. For example, in plant genetics, SNPs are widely used to identify the cis-regulatory variation within a species based on allele-specific analysis and to discover genes linked to complex genetic features by revealing its distribution, and it gives information about the adaptation of the species in that region. This situation also allows the investigation of the effects of changing conditions, especially by determining SNPs at the transcriptome level with RNA-seq technology. Using the RNA-seq data against two different conditions of the phenotype, you can create a SNP catalog and evaluate the effects on the protein sequence and which SNPs have a significant change in allele frequency. In addition, de novo and reference-based SNP discovery is carried out in many organisms, including many plants with little or no genetic information [50]. The availability of NGS provides a convenient approach to discover all SNPs and learn about genomic position and genotyping in

one step. There are many advantages in performing SNP analysis using RNA-seq data. First, thousands of SNPs can be discovered, and expression levels of millions of functional genes with sequence variations can be observed simultaneously. Second, the location of variations in coding regions associated with biological and agronomic properties of plants can be identified, and phenotypes can be estimated by genotypes [51]. It also provides information for relevant studies such as gene characterization, gene expression measurement, and posttranslational process analysis. In addition, emerging technologies have allowed de novo scanning of SNPs computationally even in the absence of a reference genome sequence of any plant variety. Thus, the targeted product can be developed using only the data from data banks before the fieldwork to efficiently cultivate a crop [52]. The development and advancement of SNP technology are extensive for both evolutionary and molecular geneticists, plant breeders, and industry and will be valuable for us to understand and develop crop species.

5. Conclusion

In order to understand evolutionary and genetic relationships between/within species, elucidate traits of agronomic interest in crops, and clarify prone to diseases, SNPs are the first approach. Especially in plant genetic research and breeding, identifying the genetic loci that are responsible for trait variation is fundamental. With the advantages of stability, budget friendly improvements, and high-throughput assays, SNP has become increasingly important in crop genetic studies. The development of genotyping tools for model and non-model crops allows the detection of the variations, and it has been successfully applied in plant science for many years. The shift to the high-throughput genotyping assays and development of next-generation sequencing technologies accelerated the discovery of polymorphisms. However, the error-prone fashion of the NGS analysis tools is still a big concern which can lead to false-positive SNPs. There is a need for the development of a tool for extracting bulk of data, support for the data analyses, and intelligent decision on the accuracy. To fulfill this need, instead of using binary composition of nucleotides, machine learning approaches are being developed. Integrated SNP Mining and Utilization (ISMU) Pipeline is one of the first trials to develop a machine learning approach to SNP discovery. The integrated approach alongside the recent innovations will allow an increased knowledge and application of SNPs in the future.

Author details

Hande Morgil, Yusuf Can Gercek and Isil Tulum*
Department of Biology, Istanbul University, Istanbul, Turkey

*Address all correspondence to: iciltlm@gmail.com

References

[1] Backlund P, Janetos A, Schimel D. The effects of climate change on agriculture, land resources, water resources, and biodiversity in the United States. In: Synthesis and Assessment Product 4.3. Washington, DC: US Environmental Protection Agency, Climate Change Science Program; 2008. p. 240

[2] Bohra A et al. Genomics-assisted breeding in four major pulse crops of developing countries: Present status and prospects. Theoretical and Applied Genetics. 2014;**127**(6):1263-1291

[3] Lucht JM. Public acceptance of plant biotechnology and GM crops. Viruses. 2015;**7**(8):4254-4281

[4] Lateef DD et al. DNA marker technologies in plants and applications for crop improvements. Journal of Biosciences and Medicines. 2015;**3**(05):7

[5] Mackill DJ, Nguyen HT, Zhang J. Use of molecular markers in plant improvement programs for rainfed lowland rice. Field Crops Research. 1999;**64**(1-2):177-185

[6] Ortiz R. Molecular plant breeding. Crop Science. 2010;**50**(5):2196

[7] Suwarno WB et al. Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theoretical and Applied Genetics. 2015;**128**(5):851-864

[8] Landjeva S, Korzun V, Börner A. Molecular markers: Actual and potential contributions to wheat genome characterization and breeding. Euphytica. 2007;**156**(3):271-296

[9] Illa-Berenguer E et al. Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. Theoretical and Applied Genetics. 2015;**128**(7):1329-1342

[10] Kim J-J et al. Development of SNP-based human identification system. International Journal of Legal Medicine. 2010;**124**(2):125-131

[11] Ganal MW et al. Large SNP arrays for genotyping in crop plants. Journal of Biosciences. 2012;**37**(5):821-828

[12] Ragousiss J. Genotyping technologies for genetic research. Annual Review of Genomics and Human Genetics. 2009;**10**:117-133

[13] Appleby N, Edwards D, Batley J. New technologies for ultra-high throughput genotyping in plants. Plant Genomics. Methods in Molecular Biology™ (Methods and Protocols). USA: Humana Press; 2009;**513**

[14] Perez-De-Castro M et al. Application of genomic tools in plant breeding. Current Genomics. 2012;**13**(3):179-195

[15] Slonim DK, Yanai I. Getting started in gene expression microarray analysis. PLoS Computational Biology. 2009;**5**:10

[16] Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. Journal of Computational Biology. 2000;**7**(6):819-837

[17] Rasheed A et al. Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. Molecular Plant. 2017;**10**(8):1047-1064

[18] Eckert AJ et al. High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). Tree Genetics & Genomes. 2009;**5**(1):225-234

[19] Yan J et al. High-throughput SNP genotyping with the GoldenGate assay in maize. Molecular Breeding. 2010;**25**(3):441-451

[20] Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM, et al. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell. 2016;**166**(2):481-491

[21] Bevan M, Walsh S. The Arabidopsis genome: A foundation for plant research. Genome Research. 2005;**15**(12):1632-1642

[22] Brachi B, Villoutreix R, Faure N, Hautekèete N, Piquot Y, Pauwels M, et al. Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. Molecular Ecology. 2013;**22**(16):4222-4240

[23] Brennan AC, Méndez-Vigo B, Haddioui A, Martínez-Zapater JM, Picó FX, Alonso-Blanco C. The genetic structure of *Arabidopsis thaliana* in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. BMC Plant Biology. 2014;**14**(1):17

[24] Fournier Level A, Wilczek AM, Cooper MD, Roe JL, Anderson J, Eaton D, et al. Paths to selection on life history loci in different natural environments across the native range of *Arabidopsis thaliana*. Molecular Ecology. 2013;**22**(13):3552-3566

[25] Freudenthal JA, Ankenbrand MJ, Grimm DG, Korte A. GWAS-Flow: A GPU accelerated framework for efficient permutation based genome-wide association studies. BioRxiv. 2019;**1**:783100

[26] King G, Nienhuis J, Hussey C. Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms. Theoretical and Applied Genetics. 1993;**86**(8):1028-1032

[27] Lasky JR, Des Marais DL, McKAY JK, Richards JH, Juenger TE,

Keitt TH. Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. Molecular Ecology. 2012;**21**(22):5512-5529

[28] Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, et al. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biology. 2005;**3**(7):1289-1299

[29] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007;**81**(3):559-575

[30] Somerville C, Koornneef M. A fortunate choice: The history of Arabidopsis as a model plant. Nature Reviews. Genetics. 2002;**3**(11):883-889

[31] Xavier A, Muir WM, Rainey KM. Impact of imputation methods on the amount of genetic variation captured by a single-nucleotide polymorphism panel in soybeans. BMC Bioinformatics. 2016;**17**(1):55

[32] Yu C, Arcos-Burgos M, Baune BT, Arolt V, Dannlowski U, Wong ML, et al. Low-frequency and rare variants may contribute to elucidate the genetics of major depressive disorder. Translational Psychiatry. 2018;**8**(1):1-8

[33] Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, et al. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PLoS One. 2011;**6**(3)

[34] Konstantinos KV, Panagiotis P, Antonios VT, Agelos P, Argiris NV. PCR–SSCP: A method for the molecular analysis of genetic diseases. Molecular Biotechnology. 2008;**38**(2):155-163

[35] LaFramboise T. Single nucleotide polymorphism arrays: A decade

of biological, computational and technological advances. Nucleic Acids Research. 2009;**37**(13):4181-4193

[36] McCallum CM, Comai L, Greene EA, Henikoff S. Targeting induced locallesions in genomes (TILLING) for plant functional genomics. Plant Physiology. 2000;**123**(2):439-442

[37] Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000;**407**(6803):513-516

[38] Liu Z. Single nucleotide polymorphism (SNP). Aquaculture Genome Technologies. USA: Blackwell Publishing; 2007:59-72

[39] Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: Practice, problems and promise. Human Mutation. 2001;**17**(4):296-304

[40] Storm N, Darnhofer-Patel B. MALDI-TOF mass spectrometry-based SNP genotyping. In: Single Nucleotide Polymorphisms. Totowa, NJ: Springer; 2003. pp. 241-262

[41] Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Research. 2001;**11**(1):3-11

[42] Huq MA et al. Identification of functional SNPs in genes and their effects on plant phenotypes. Journal of Plant Biotechnology. 2016;**43**(1):1-11

[43] Thomson MJ. High-throughput SNP genotyping to accelerate crop improvement. Plant Breeding and Biotechnology. 2014;**2**(3):195-212

[44] Kilian A et al. The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In: Proceedings of the International Congress in the Wake of the Double Helix: From the Green Revolution to the Gene Revolution. 2003. pp. 443-461

[45] Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, et al. Rapid transcriptome characterization for nonmodel organism using 454 pyrosequencing. Molecular Ecology. 2008;**17**:1636-1647

[46] Kwak M, Velasco D, Gepts P. Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). Journal of Heredity. 2008;**99**:283-291

[47] Shakira G, Azim A. Metagenomics and its application in rumen ecosystem: Potential biotechnological prospects. Pakistan Journal of Nutrition. 2009;**8**:1309-1315

[48] Varshney RK, Nayak SN, May GD, Jackson SA. Next generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology. 2009;**27**:522-530

[49] Xu P, Wu X, Luo J, Wang B, Liu Y, Ehlers JD, et al. Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. BMC Genomics. 2011;**12**:467-477

[50] Duran C, Appleby N, Clark T, Wood D, Imelfort M, Batley J, et al. Auto SNPdb: An annotated single nucleotide polymorphism database for crop plants. Nucleic Acids Research. 2009;**37**:951-953

[51] McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Qi SQ. Genetic properties of the maize nested association mapping population. Science. 2009;**325**:737-740

[52] Imelfort M, Duran C, Batley J, Edwards D. Discovering genetic polymorphisms in next-generation sequencing data. Journal of Plant Biotechnology. 2009;**7**:312-317