JUST AGRICULTURE
multidisciplinary e-Newsletter

Revolutionizing Fruit Crop Improvement: Harnessing SNP-Based Genotyping Arrays

Rashmi Upreti[1]* and Pushpesh Joshi[2]
[1]Department of Horticulture, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, Uttarakhand
[2]Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut, U.P

ARTICLE ID: 37

Abstract

Traditional phenotypic selection methods, reliant on morphological markers, face limitations due to environmental influences and the subjective nature of scoring. Biochemical markers, while providing more genetic insight, are also constrained by their stage specificity and instability. The introduction of Single nucleotide polymorphisms (SNPs) has revolutionized crop improvement by offering high polymorphic information content and efficient, non-gel-based detection methods. SNP arrays, developed from early genotyping techniques, facilitate high-throughput screening and allow for comprehensive genomic analysis, linking phenotypic variations to genetic differences. Despite challenges in developing SNP arrays for polyploid species, ongoing advancements in high-throughput sequencing technologies promise to improve these tools, ultimately enhancing genomic selection and crop improvement strategies.

Introduction

Earlier, selection is considered as fundamental step for improving the economically important traits through different conventional approaches in plant breeding which is based on phenotypes of the plant. These are later called as morphological markers, provide some idea about the genotype only at their appearance. However, most of the traits are controlled by many genes and highly influenced by environment, subjected to allelic interactions like epistasis or pleiotropy. And, use of morphological markers excludes analysis of non-coding sequences of genomes which accounts for more than 95% of the total genome in higher plants (O'Neill *et al*., 2003). Scoring of these markers is subjective and results may vary breeder to breeder. All these constraints make the use of phenotypic markers limited and tedious. Another set of markers called as biochemical markers (proteins, isozyme/allozyme and secondary

JUST AGRICULTURE
multidisciplinary e-Newsletter

metabolites) may overcome most of the limitations of morphological markers. Proteins/ allozymes are minimally influenced by environment. They provide closest insight into the genetic makeup. They are co-dominant markers makes detection of recessive alleles easy in heterozygotes. However, the biochemical markers are stage specific, difficult to isolate and unstable at room temperature. The use of secondary metabolites is restricted to only those plants that produce suitable range of those metabolites through complex pathways.

A new class of markers called molecular markers have been introduced in the last three decades which have revolutionised the crop improvement (Lateef *et al*., 2015). Number of markers such as first generation (Restriction fragment length polymorphism), second generation (Amplified fragment length polymorphism, Simple sequence repeats) and third generation (Expressed sequenced tags, SNPs) used as efficient tools by plants breeders in MAS (Marker Assisted Selection). A novel class of markers namely SNPs has emerged recently as an important tool in genomics and are increasingly used in various laboratories for diverse applications. They possess unique merits over others having high polymorphic information content, non-gel based and are less time consuming. The natural age long process of breeding including collection, cultivation and domestication has been accelerated for economically important traits by detecting dozens of SNPs using next-generation sequencing (Morgil *et al*., 2020). SNP array is a type of DNA microarray which is used to detect specific alleles of SNPs present on the array through labelled probe locating the SNP position upon hybridisation with fragmented DNA. Three mandatory components of the SNP arrays included an array containing immobilised allele specific oligonucleotide probes, fragmented nucleic acid sequences of target labelled with fluorescent dyes, detection system that records and interprets hybridisation signal.

History of SNP Array

The large genotyping arrays in crop plants was expanded much later than in human genetics due to polyploid nature of many crop species which makes SNP identification and SNP calling more tedious and complex work than humans. The SNP identification was expensive and complex before establishment of next generation sequencing (NGS) technologies. SNP arrays were first commercially produced by Affymetrix nearly two decades ago. Wang *et al*. (1998) initially prototyped the Hu SNP assay which was designed to genotype 1494 SNPs on one chip. Subsequent versions increased stepwise from 10,000 to nearly one

JUST AGRICULTURE
multidisciplinary e-Newsletter

million in the current release. SNP genotyping techniques include low-throughput gel-based approach, cleaved amplified polymorphic sequence (CAPS) marker approach, PCR-based fluorescently-labelled high-throughput methods, high-resolution melting (HRM) curve analysis, TaqManR and KASPTM assay, fixed array systems such as Illumina Infinium, Affymetrix Axiom (Allen *et al*., 2017), and NGS enabled approaches such as restriction enzyme-based genotyping by sequencing (GBS). With the help of NGS technologies, large number of SNPs can be easily identified in almost all plant species through comparative sequencing approaches.

SNP Array: Significance and applications in plant improvement

A progression in gathering genomic information and the development of new genomic analysis methods assists in the ability to link phenotypic variations to genomic differences in most of the crops. Two instances of high-throughput platforms for gathering genomic information include next generation sequencing and the development of DNA microarrays (Davey *et al*., 2011). The development of high-throughput sequencing technologies has allowed researchers to explore how genetic diversity across the whole genome of crops was shaped through domestication by comparing the patterns of genetic variation between wild and cultivated species. The SNP array approach is faster, high-throughput, cheaper and convenient alternative genotyping technology for germplasm screening compared to a whole genome sequencing (Yu *et al*., 2014). To investigate domestication history and population structure, higher-resolution arrays are needed for large scale genotyping in plants and ultimately provide information on genetic resources for future breeding. It has been used in Genome-wide association study (GWAS) to identify significantly related loci. SNP arrays made it possible to create saturated, supersaturated genetic maps thereby enabling genome-wide tracking, fine mapping of target regions and accelerated cloning of genes of interest. They are considered as important tool for linkage mapping, map-based cloning and genomic selection (Clarke *et al*., 2016). Unlike other genotyping approaches, SNP array has advantages in high-throughput genotyping. First, the computational analysis of the data generated through NGS-based method is relatively easy especially NGS-library construction and downstream computational processing for SNP calling. Second, the array is designed using specific alleles from the genomic regions of interest. However, construction of SNP genotyping array required prior SNP genomic locations, designing and optimisation is time-taking and its construction face an

JUST AGRICULTURE
multidisciplinary e-Newsletter

issue of ascertainment biasness. Development of array has progressed slowly in polyploid species due to their complex inheritance. Various SNP arrays have been used in diploid crops successfully, *i.e.*, in Apple 480K SNP array, Maize 600K SNP array (Unterseer *et al.*, 2014), and in Rice 700K SNP array (McCouch *et al.*, 2016). Many fruits crops are sequenced irrespective of ploidy barrier and thus, array has been developed and summarised in table 1.

Table 1: Axiom arrays developed for some fruits important crops

S.No.	Crops	Ploidy	Genome size	SNP arrays	Distributor	References
1	Cherry	2n=4x=32	344.29 Mb	6K	Illumina Infinium bead chip	Peace *et al.*, 2012
2	Apple	n=17	~750 Mb	8K		Chagne *et al.*, 2012
3	Peach	2n=2x=16	265 Mb	9K		Verde *et al.*, 2012
4	Grape	2n=4x=76	~500 Mb	18K		Laucou *et al.*, 2018
5	Avocado	2n=2x=24	~920 Mb	6K		Kuhn *et al.*, 2019
6	Apple	n=17	~750 Mb	480K	Affimetrix Axiom	Bianco *et al.*, 2016
7	Potato	2n=4x=48	844 Mb	20K	Affimetrix Axiom	Vos *et al.*, 2015
8	Rose	2n=2x=14	560 Mb	68K	Affimetrix Axiom	Koning-Boucoiran *et al.*, 2015
9	Walnut	2n=2x=32	667 Mb	700K		Marrano *et al.*, 2019
10	Pear	2n=2x=34	527 Mb	70K & 200K		Montanari *et al.*, 2019
11	Strawberry	2n=8x=56	813.4 Mb	90K	Affimetrix Axiom	Bassil *et al.*, 2015
17	Grapes	2n=4x=76	~500 Mb	18K, 9K	Illumina Infinium bead chip	Myles *et al.*, 2010
18	Lettuce	2n=2x=18	2.5 Gb	35K	Affimetrix Gene Chip	Stoffel *et al.*, 2012

JUST AGRICULTURE
multidisciplinary e-Newsletter

28	Tomato	2n=2x=24	950 Mb	10K	Illumina Infinium bead chip	Sim *et al*., 2012

Conclusion

Genotyping array has been evolved for considerable number of fruit crops in order to get insight in the genomic constitution and can be exploited for its improvement. However, many other crops have SNP array of more than 5000 SNPs are developed which can be improved by high-throughput sequencing technologies and high-quality reference genome. It is expected that these first-generation arrays would be replaced by improved array having a greater number of true allele specific SNPs and haplotypes defining SNPs. These would be extensively useful in genomic associations approaches and genomic selection in fruit breeding.

References

Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L., Wilkinson, P. A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., Griffiths, S., Bentley, A. R., Alda, M., Jack, P., Phillips, A. L., Edwards, K. J. (2017). Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant biotechnology journal*, *15*(3), 390–401.

Chagné, D., Crowhurst, R. N., Troggio, M., Davey, M. W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R. P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J. D., Iezzoni, A., Mockler, T., Wilhelm, L., Van de Weg, E., Gardiner, S. E., Bassil, N., & Peace, C. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PloS one, 7(2), e31745.

Bassil, N. V., Davis, T. M., Zhang, H., Ficklin, S., Mittmann, M., Webster, T., Mahoney, L., Wood, D., Alperin, E. S., Rosyara, U. R., Koehorst-Vanc Putten, H., Monfort, A., Sargent, D. J., Amaya, I., Denoyes, B., Bianco, L., van Dijk, T., Pirani, A., Iezzoni, A., Main, D., van de Weg, E. (2015). Development and preliminary evaluation of a 90 K Axiom® SNP array for the allooctoploid cultivated strawberry *Fragaria* × *ananassa*. *BMC genomics*, *16*(1), 155.

Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denancé, C., Théron, A., Poncet, C., Micheletti, D., Kerschbamer, E., Di Pierro, E. A., Larger, S., Pindo, M., Van de Weg,

E., Davassi, A., Laurens, F., Velasco, R., Durel, C. E., & Troggio, M. (2016). Development and validation of the Axiom(®) Apple480K SNP genotyping array. *The Plant journal: for cell and molecular biology*, *86*(1), 62–74.

Clarke, W.E., Higgins, E.E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., Batley, J., Edwards, D., Meng, J., Li, R. and Lawley, C.T., 2016. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theoretical and Applied Genetics*, *129*(10), pp.1887-1899.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), pp.499-510.

Koning-Boucoiran, C. F., Esselink, G. D., Vukosavljev, M., van 't Westende, W. P., Gitonga, V. W., Krens, F. A., Voorrips, R. E., van de Weg, W. E., Schulz, D., Debener, T., Maliepaard, C., Arens, P., & Smulders, M. J. (2015). Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (*Rosa* L.). *Frontiers in plant science*, *6*, 249.

Kuhn, D.N., Livingstone, D.S., Richards, J.H., Manosalva, P., Van den Berg, N., Chambers, A.H. 2019. Application of genomic tools to avocado (*Persea americana*) breeding: SNP discovery for genotyping and germplasm characterization. *Scientia Horticulturae*, 246, 1–11.

Lateef, D. D. (2015). DNA marker technologies in plants and applications for crop improvements. *Journal of Biosciences and Medicines*, *3*(5), 7-18.

Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A. F., Bérard, A. & Boursiquot, J. M. (2018). Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PloS one*, *13*(2), e0192540.

Marrano, A., Martínez-García, P. J., Bianco, L., Sideli, G. M., Di Pierro, E. A., Leslie, C. A. & Neale, D. B. (2019). A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density Axiom™ J. regia 700K SNP genotyping array. *Plant Biotechnology Journal*, *17*(6), 1027-1036.

McCouch, S.R., Wright, M.H., Tung, C.W., Maron, L.G., McNally, K.L., Fitzgerald, M., Singh, N., DeClerck, G., Agosto-Perez, F., Korniliev, P. and Greenberg, A.J., 2016. Open access resources for genome-wide association mapping in rice. *Nature communications*, *7*(1), pp.1-14.

Montanari, S., Bianco, L., Allen, B. J., Martínez-García, P. J., Bassil, N. V., Postman, J. & Neale, D. B. (2019). Development of a highly efficient Axiom™ 70 K SNP array for *Pyrus* and evaluation for high-density mapping and germplasm characterization. *BMC genomics*, *20*, 1-18.

Morgil, H., Gercek, Y.C. and Tulum, I., 2020. Single nucleotide polymorphisms (SNPs) in plant genetics and breeding. In *The Recent Topics in Genetic Polymorphisms*. London, UK: IntechOpen.

Myles, S., Chia, J.M., Hurwitz, B., Simon, C., Zhong, G.Y., Buckler, E. and Ware, D., 2010. Rapid genomic characterization of the genus *Vitis*. *PloS one*, *5*(1), p.e8219.

O'Neill R, Snowdon R & Köhler W, Population genetics: Aspects of biodiversity, *Prog Bot*, 64 (2003) 115-137.

Peace, C., Bassil, N., Main, D., Ficklin, S., Rosyara, U. R., Stegmeir, T. & Iezzoni, A. (2012). Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLoS One*, *7*(12), e48305.

Stoffel, K., van Leeuwen, H., Kozik, A., Caldwell, D., Ashrafi, H., Cui, X., Tan, X., Hill, T., Reyes-Chin-Wo, S., Truco, M.J. (2012). Development and application of a 6.5 million feature Affymetrix Genechip_ for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca* spp.). *BMC Genomics*, 13:185.

Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., Van Deynze, A., Hamilton, J.P., Buell, C.R., Causse, M., Wijeratne, S. and Francis, D.M., 2012. Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PloS one*, *7*(7), p.e40563.

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T.M., Fries, R., Pausch, H. and Bertani, C., 2014. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC genomics*, *15*(1), pp.1-15.

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C.T., Gasic, K., Micheletti, D., Rosyara, U.R., Cattonaro, F., Vendramin, E. (2012). Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PloS one*, 7, e35668.

Vos, P.G., Uitdewilligen, J.G., Voorrips, R.E., Visser, R.G. and van Eck, H.J., 2015. Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. *Theoretical and Applied Genetics*, *128*(12), pp.2387-2401.

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R. & Lander, E. S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, *280*(5366), 1077-1082.

Yu, H., Xie, W., Li, J., Zhou, F. & Zhang, Q. A 2014. Whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnology Journal*, 12, 28–37.