Scientific Papers. Series B, Horticulture. Vol. LXIII, No. 1, 2019
Print ISSN 2285-5653, CD-ROM ISSN 2285-5661, Online ISSN 2286-1580, ISSN-L 2285-5653

MOLECULAR MARKERS ASSOCIATED WITH SPECIFIC QUANTITATIVE TRAIT LOCI (QTL) IN PLANT RESEARCH

Anca Amalia UDRIȘTE[1], Liliana BĂDULESCU[1, 2]

[1]Research Center for Studies of Food and Agricultural Products Quality HORTINVEST, University of Agronomic Sciences and Veterinary Medicine of Bucharest, 59 Marasti Blvd., District 1, Bucharest, Romania
[2]University of Agronomic Sciences and Veterinary Medicine of Bucharest, Faculty of Horticulture, Department of Horticultural Systems Bioengineering, 59 Marasti Blvd., District 1, Bucharest, Romania

Corresponding author email: amaliaudriste@gmail.com

Abstract

The use of molecular markers, that allow to know the set of genes associated with a particular quantitative trait or QTL is widely extended, including: restriction fragment length polymorphism (RFLP), random-amplified polymorphic DNA (RAPD), cleaved amplified polymorphic sequences (CAPS), amplified fragment length polymorphism (AFLP), microsatellites, and single-nucleotide polymorphisms (SNPs). In addition to classical methodology, new approaches based on the next generation sequencing (NGS) methodology that enables rapid sequencing of the base pairs in DNA or RNA samples, are proving to be fundamental. SNPs have become the focus of a large number of studies designed to identify critical differences in DNA sequence that contribute to phenotypic variation for specific traits. Methods for the analysis of SNPs comprise two distinct steps, one for allele discrimination and one for allele detection. This review article describes the methods available for allele discrimination based on hybridization, primer extension, ligation and enzymatic cleavage, as well as, RAPD and CAPS molecular markers analysis.

Key words*: molecular markers, phenotypic variation, SNP.*

INTRODUCTION

Molecular approach in selecting plants with desired traits is called marker-assisted selection (MAS), a technology widely used in developing disease resistance varieties combined with suitable agronomic traits. Different types of DNA molecular markers have been developed and applied in the field of agricultural biotechnology or promoted for plant breeding in various fruit growing research.

The conventional breeding systems require several breeding cycles to combine many target traits in a particular line or genotype. In contrast, MAS offers the potential to raise target traits in single genotype more efficiently and precisely in fewer selection cycles and with less losses (Babu et al., 2004, Ion et al., 2016, Ion & Badulescu, 2016). Therefore, gene based markers tightly linked to the resistance genes are used for the rapid and efficient indirect selection of target genes (Ion, 2018).

Molecular markers utilized in MAS should: (i) co-segregate or be tightly linked to the trait object of selection; (ii) it should lend itself to a mass-screening for the identification of the marker genotype in breeding lines; (iii) its validity should be recognized in a laboratory-independent manner, as reliable and reproducible in different laboratories. (Mohan et al., 1997; Mandolino & Carboni, 2004). The main characteristics of the most commonly used molecular markers are shown in Table 1.

Biochemical markers

Historically, the earliest molecular markers to be extensively used in population studies and in plant breeding have been biochemical markers or isozymes (Lewontin and Hubby, 1966).

Isozyme variation have been associated to specific traits such as disease resistance (e.g. to root knot disease in tomato; Rick and Fobes, 1974). Besides, the codominant nature of isoenzyme markers made them particularly useful in the estimation of heterozygosity and

in studies of gene flow from crop species to their wild relatives (e.g. Bartsch et al.,1999).

Restriction fragment length polymorphism (RFLP)

RFPL markers rely on differences, at the genomic DNA level, in the target sequences of the restriction endonucleases. Such differences lead to variant DNA fragment length upon restriction, usually visualized by agarose gel separation, followed by hybridization of the immobilized fragments to a labeled (usually radioactively) probe and autoradiography. An RFLP marker is codominant and identifies one single locus at a time. (Beckmann and Soller, 1983; Tanksley et al., 1989).

Table 1. The main characteristics of the most commonly used molecular markers.

RFLP	RAPD	SNP	SSR	CAPS
Co-dominant	Dominant	Co-dominant	Co-dominant	Co-dominant
No PCR	PCR	PCR	PCR	PCR
Sequencing	No sequencing	Sequencing	Sequencing	Sequencing
High reproducibility	High reproducibility	High reproducibility	High reproducibility	High reproducibility
Medium polymorphism level	Very high polymorphism level	High polymorphism level	High polymorphism level	Very high polymorphism level
High cost	Less cost	Variable cost	High cost	Variable cost

Random ampified polymorphic DNA (RAPD)

The earliest PCR-based markers to be extensively applied to plant breeding and MAS were the random ampified polymorphic DNA (RAPD) markers (Williams et al., 1990). In this case, PCR amplification is mediated by short decamer primers of random sequence. Such primers find with a certain frequency annealing sites on the opposite strands of the target DNA molecule. The annealing sites of the decamer primers can be variable, and consequently some of the amplified fragments will be polymorphic in the different DNA fragments. The nature of RAPDs (and of many PCR-based markers) is dominant and multilocus (G. Mandolino & A. Carboni, 2004).

Single nucleotide polymorphism (SNP)

For germplasm analysis and plant MAS sellection in most cases is necessary the isolation of the RAPD bands found associated to a trait of interest. The relevant RAPD fragments are gel-isolated, cloned and sequenced. Specific 20-mer primers are then designed, amplifying only the sequence found by genetic analysis to be linked to the trait, resulting sequence characterized amplified region SCAR markers or single nucleotide polymorphism (SNP) (G. Mandolino & A. Carboni, 2004). SNPs are the most common DNA polymorphisms in genome sequences of plants, human and animals, and they are thought to be the major source of phenotypic variations, due to they can provide a great marker density. SNPs are typically biallelic and provide the ultimate form of molecular markers as a single nucleotide base, the smallest unit of inheritance (Lateef, 2015).

Simple sequence repeats (SSR)

The most widely used molecular markers in plant breeding are microsatellites or SSR. These are short sequences of two, three or more nucleotides that are repeated for a variable number of times in the genome. These markers are codominant, highly reliable, polymorphic and permit multistage selection. Primers are designed on the basis of the DNA sequences flanking the repeat stretch, able to amplify the target sequence by PCR. In general, one single locus is identified by each PCR reaction, but the number of alleles that can be identified is very high, as the variability in the repeated motif number is high in the plant genomes (Morgante and Olivieri, 1993) (Mandolino and Carboni, 2004). Furthermore, SSR markers are readily accessible through published linkage

maps and public databases, and allow differentiation between homozygotes and heterozygotes, used to screen, characterize and evaluate genetic diversity in many plant species.

Cleaved amplified polymorphic sequence (CAPS)

CAPS markers, are a combination of RFLP markers and PCR, and have been used in genotyping, map-based cloning and molecular identification studies. In this technique, target DNA is amplified using PCR and then its digestion is performed with specific restriction enzymes (Michaels et al., 1998). Agarose or acrylamide gel is used for the visualization of CAPS products. The primers used in this technique are developed from sequence information present in a databank of genomics or cloned RAPD bands or cDNA sequences. CAPS markers are versatile and the possibility to find DNA polymorphism can be increased by combining CAPS with single-strand conformational polymorphism, SCAR, AFLP or RAPD (Agarwal et al., 2008).

Next-generation sequencing (NGS)

NGS, relies on massively parallel sequencing and imaging techniques to yield several millions of DNA bases per run (Shendure and Ji, 2008). Several NGS platforms (Roche 454 FLX Titanium, Illumina MiSeq and HiSeq2500, Ion Torrent PGM), have been developed and used recently for ultra high sequencing operations, where as many as 500,000 sequencing-by-synthesis operations may be run in parallel (Quail et al., 2012).

Genotyping-by-sequencing (GBS), has been developed and applied in sequencing multiplexed samples that combine molecular marker discovery and genotyping. GBS is a novel application of NGS protocols for discovering and genotyping SNPs in crop genomes and populations that it generates large numbers of SNPs for use in genetic analyses and genotyping.

The GBS approach includes the digestio of genomic DNA with restriction enzymes followed by the ligation of barcode adapter, PCR amplification and sequencing of the amplified DNA pool on a single lane of flow cells. GBS is becoming increasingly important as a cost- effective and unique tool for genomics-assisted breeding in a range of plant species (Jiangfeng He et al, 2014).

Quantitative trait loci (QTL)

Quantitative trait locus (QTL) analysis is a statistical method that links phenotypic data (trait measurements) and genotypic data (usually molecular markers) in an agree to explain the genetic basis of variation in complex traits (Kearsey, 1998). QTL analysis allows researchers, in diverse fields, to link certain complex phenotypes to specific regions of chromosomes. The goal of this process is to identify the interaction and precise location of these regions (see Fig. 1).

The QTL mapping approach, has become known for the genetic studies of some of the relevant attributes in various fruit traits like fruit size, shape, firmness, netting, color and assorted metabolites involving carotenes, sugars and organic acids have been identified during recent times in a few melon maps (Baloch et al., 2016, Monforte et al., 2004).

In order to begin a QTL analysis, scientists require organisms that differ genetically with regard to a specific trait of interest and genetic markers that distinguish between these traits of interest. Then, the parental strains are crossed, resulting in heterozygous (F_1) individuals, and these individuals are then crossed using one of a number of different schemes (Darvasi, 1998). Finally, the phenotypes and genotypes of the derived (F_2) population are scored. Markers that are genetically linked to a QTL influencing the trait of interest will segregate more frequently with trait values, whereas unlinked markers will not show significant association with phenotype (Miles et al., 2008, Nadeem et al., 2018).

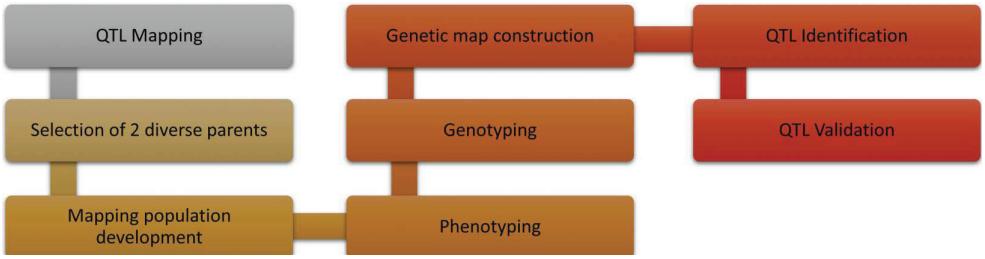

Figure 1. QTL mapping methodology (Nadeem M.,A., et al., 2018).

The forensic applications of DNA markers

Molecular markers are extremely useful in plant genetics and breeding. Markers are prerequisite for gene mapping and tagging, segregation analysis, genetic diagnosis, forensic examination, phylogenetic analysis and numerous biological applications (Lam et al., 2010).

Molecular markers can be useful for forensic purposes in order to identify illicit material types, to determine the origin of seized samples and also help to identify suspect plant cultivars. *Cannabis* is a precious plant with multiple applications; hence the possibility of engineering it genetically to produce useful compounds or raw products is highly valuable. Nowadays new perspectives for hemp as a crop that can be grown for food and non-food reasons are starting. According to Hansen (2009), as a result of its numerous nutritional benefits, many new food products containing hemp seed and its oil are in the marketplace, including pasta, tortilla chips, salad dressings, snack products and frozen desserts.

In spite of the numerous new applications nowadays possible for hemp oil and fibers, a disadvantage for the circulation of hemp cultivation is still represented by the presence of hemp psychoactive components. Hemp plants are characterized by the presence of terpenophenolic substances known as cannabinoids, which accumulate mainly in the glandular trichomes of the plant (Mechoulam, 1970), the most abundant are cannabidiol (CBD) and D-9-tetrahydrocannabinol (THC) (Piluzza et al., 2013).

The THC content of recreational *Cannabis* has drastically increased in the last 30 years, from 3% in 1980s to almost 20% nowadays, with very low level of the other cannabinoids such as CBD (Farag and Kayser, 2015). The use of molecular markers for DNA polymorphism analysis (RAPD) can be useful for the acquirement of information at an early stage of growth about the hemp fiber or drug type. The DNA polymorphisms generated by the selected primers appear to be quite effective in resolving Cannabis type determinations according to Piluzza et al. (2013). RAPD markers allow random sampling of markers overwhole genomic DNA and do not require any previous information on the genome of the organism under investigation, comparative to other molecular markers.

Rotherham and Harbison (2011), developed a single nucleotide polymorphism (SNP) assay capable of discriminating between high and low levels of THC *Cannabis* varieties based on sequence variation in the THCA synthase gene and tested this assay on drug and non-drug varieties of *Cannabis*. Non-drug plants were found to be homozygous at the four sites assayed while drug *Cannabis* plants were either homozygous or heterozygous.

However, hemp and marijuana varieties are hardly distinguishable morphologically and discrimination of drug vs. non-drug chemotypes by quantitative THC-dosage has also proven inadequate due to its dependence on environmental factors, to the strong variation during the plant's life cycle, as well as between individual plants (Welling et al., 2016).

CBD compound is the most prevalent phyto-cannabinoid in the fiber-type hemp and the second most important in the drug chemotypes. In addition, the qualitative assessment of THC:CBD ratio is also problematic for an unequivocal discrimination between fibre and drug types, due to the presence of a largely variable intermediate chemotype class and the common practice among drug breeders to produce hybrid varieties (Stagginus et al., 2014, Tipparat al. 2012).

Genetic tools may overcome these issues by genotype loci directly linked to THC synthesis in association with chemotype profiling, yet genotyping my be compromised be complex gene duplications, pseudogenes and that a only limited number of varieties of Cannabis has been validated (Weiblen et al., 2015; McKernan et al., 2015).

Dufresnes et al. (2017), provided a high-density SNP data for *Cannabis*, by genotyping 13 microsatellite loci (STRs) in 1324 samples selected specifically for fibre (24 hemp varieties) and drug (15 marijuana varieties) production and showed that these loci are sufficient to capture most of the genome-wide diversity patterns recently revealed by NGS data. This microsatellites database is the most powerful resource suitable for routine forensic analyses of *Cannabis*.

Yet, it remains limited by several aspects. First, drug vs. non-drug discrimination can be ambiguous for some samples, given the lack of differentiation and/or crossbreeding practices between few hemp and marijuana varieties (Dufresnes et al. 2017). Given the tremendous diversity of marijuana and the legal difficulty to access samples, development of molecular markers would allow unprecedented opportunities to extend forensic advances and promote the development of the industrial and therapeutic potential of this species.

However, more recent quantitative trait loci (QTL) mapping experiments were performed by Weiblen et al. (2015) where cannabinoid profiles of the same 62 individual genotypes were analyzed with linkage map using Windows QTL CARTOGRAPHER v.2.5 software. A subset of 62 plants from the F2 population were genotyped for 103 DNA markers, 16 microsatellite markers, CBDA and THCA synthase gene sequences to construct a linkage map showing a model of codominant alleles at a single locus.

As a results, the diversity of THCA and CBDA synthase gene sequences, enzyme coding loci on the map and patterns of expression, suggested multiple linked loci, in the mapping population (Weiblen et al., 2015).

All experiments suggested an evolutionary genetic basis for the differentiation of hemp and marijuana, by predicting that homozygous plants for functional CBDA synthase gene lack the capacity to yield > 0.3% THC (Weiblen et al., 2015). In conclusion, screening for the presence of nonfunctional CBDA gene could be used to verify seed sources before planting.

CONCLUSIONS

The last years have confirmed a continuous development in the molecular markers technology from earliest molecular markers to next-generation sequencing (NGS), with a diversity of array technology-based markers.

Advancements in the sequencing technologies have led to the development of NGS platforms that are low cost with high throughput.

The coming years are likely to see continued innovations in molecular marker technology to make it more precise, productive and cost effective in order to investigate the underlying biology of various traits of interest.

In the era of next-generation sequencing, there has arisen an urgent need for proper population design, advanced statistical strategies, and precision phenotyping to fully exploit high-throughput genotyping.

REFERENCES

Agarwal, M., Shrivastava, N., Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep*., 27(4), 617–631.

Babu, R., Nair, S.K., Prasanna, B.M., Gupta, H.S. (2004). Integrating marker-assisted selection in crop breeding-prospects and challenges. *Crop Sci*., 87 (5), 606–619.

Baker, P.B., Gough, T.A., Taylor, B.J. (1982). The physical and chemical-features of Cannabis plants grown in the United-Kingdom of Great-Britain and Northern-Ireland from seeds of known origin. *Bull Narc*., 34, 27–36.

Baloch, A.M., Baloch, A.W., Liu, S., Gao, P., Baloch, M.J., Wang, X., Davis, A.R., Ali, M., Luan, F., (2016). Linkage map construction and QTL analysis of fruit traits in melon (cucumis melo L.) based on caps markers. *Pak. J. Bot.,* 48 (4), 1579–1584.

Darvasi, A. (1998). Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genetics*, 18, 19–24.

Dufresnes, C., Jan, C., Bienert, F., Goudet J., Fumagalli L. (2017). Broad-Scale Genetic Diversity of Cannabis for Forensic Applications. *PLoS ONE*, 12 (1), e0170522.

Farag, S., Kayser, O. (2015). Cannabinoids production by hairy root cultures of Cannabis sativa L. *Am. J. Plant Sci.* 6, 1874–1884.

Hansen, R. (2009). Industrial hemp profile http://www.agmrc.org/commodities_products/fiber/industrial _hemp_profile.cfm

Ion, L., Moreau, F., Badulescu, L. (2016) Studies of real-time Rt-PCR detection and regular PCR for phenotyping the peach old varieties. International *Journal of Agriculture Innovations and Research*. Vol. 4(4), 796-800.

Ion, L., Badulescu, L. (2016) Studies of the old romanian apple varieties involved in new breeding programs for resistance to *Venturia inaequalis. Journal of Agricultural Science* Vol. 8(5), 1-9.

Ion, L. (2018) Research on the isolation of genomic DNA from old apple varieties. *Scientific Papers. Series B. Horticulture.* Vol. LXII, 55-58.

Jiangfeng, He, Xiaoqing, Zhao, André, Laroche, Zhen-Xiang, Lu, HongKui, Liu, Ziqin, Li. (2014). Genotyping-by-sequencing(GBS), an ultimate marker assisted selection (MAS) tool to accelerate plant breeding. *Front Plant Sci*., 5, 484.

Kearsey, M. J.(1998). The principles of QTL analysis (a minimal mathematics approach). *Journal of Experimental Botany*, 49, 1619–1623.

Lam, H.M., Xu,X., Liu,X., Chen,W., Yang,G., Wong,F.L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat.Genet*. 42.

Lateef, D.D. (2015). DNA marker technologies in plants and applications for crop improvements. *J. Biosci.Med*., 3. 7–18.

McKernan, K.J., Helbert, Y., Tadigotla, V., McLaughlin, S., Spangler, J., Zhang, L., et al. (2015). Single molecule sequencing of THCA synthase reveals copy number variation in modern drug-type *Cannabis sativa* L. BioRxiv.

Mechoulam, R. (1970). Marijuana chemistry. *Science* 168, 1159–1166

Michaels, S.D., Amasino, R. M. (1998). A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. *Plant J.,* 14(3), 381–385.

Miles, C. and Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nature Education*, 1(1), 208.

Monforte, A.J., Oliver, M., Gonzalo, M.J., Alvarez, J.M., Dolcet Sanjuan, R., Arus, P., (2004). Identification of quantitative trait loci involved in fruit quality traits in melon (*Cucumis melo* L.). *Theor. Appl. Genet*., 108, 750–758.

Nadeem, M.,A., Nawaz, M.,A., Shahid, M.,Q., Doğan, Y., Comertpay,G.,Yıldız,M., Hatipoğlu, R., Ahmad,F., Alsaleh,A., Labhane, N., Özkan, H., Chung G. & Baloch F.S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment,* 32(2), 261-285.

Piluzza, G., Delogu, G., Cabras, A., Marceddu, S., Bullitta, S. (2013). Differentiation between fiber and drug types of hemp (*Cannabis sativa* L.) from a collection of wild and domesticated accessions. *Genet Resour Crop Evol*, 60, 2331–2342.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., et al.(2012). A tale of three next generation sequencing platforms:comparison of IonTorrent, Pacific Biosciences and Illumina MiSeqsequencers. *BMC Genomics*, 13, 341.

Rotherham, D., Harbison, S.A. (2011). Differentiation of drug and non-drug Cannabis using a single nucleotide polymorphism (SNP) assay. *Forensic Science International*, 207, 193–197

Rowan, M.G., Fairbairn, J.W. (1977). Cannabinoid patterns in seedlings of Cannabis-sativa L and their use in determination of chemical race. *J Pharm Pharmacol.,* 29, 491–494.

Shendure, J., and Ji, H. (2008). Next-generation DNAsequencing. *Nat.Biotechnol.*, 26, 1135–1145.

Stagginus, C., Zorntlein, S., de Meijer E. (2014). A PCR marker linked to a THCA synthase polymorphism is a reliable tool to discriminate potentially THC-rich plants of *Cannabis sativa* L. *J Forensic Sci*., 59, 919–926.

Tipparat, P., Natakankitkul, S., Chamnivikaipong, P., Chutiwat, S. (2012). Characteristics of cannabinoids composition of Cannabis plants grown in Northern Thailand and its forensic application. *Forensic Sci Int*., 215, 164–170.

Weiblen, G.D., Wenger, J.P., Craft, K.J., ElSohly, M.A., Mehmedic, Z., Treiber, E.L., et al. (2015). Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytol*., 208, 1241–1250.

Welling, M.T., Liu, L., Shapter, T., Raymond, C.A., King, G.J. (2016). Characterization of cannabinoid composition in a diverse *Cannabis sativa* L. germplasm collection. *Euphytica*, 208, 463–475.