Review

Advance of molecular marker application in the tobacco research

X. Z. Liu and H. Y. Zhang*

Biotechnology laboratory; Southwest Forestry College, White Dragon Temple, Kunming, Yunnan Province, 650224, People’s Republic of China.

Accepted 21 November, 2008

Tobacco (Nicotiana spp.) is one of the most important commercial crops in the world. During the last two decades, molecular markers have entered the scene of genetic improvement in different fields of agricultural research. The principles and characteristics of several molecular markers such as RFLP, RAPD, AFLP, microsatellites and minisatellites applied in tobacco genetics and breeding were reviewed. The application and development of molecular marker in tobacco genetic research was presented emphatically in the following areas: evolutionary genetics, population genetics and genotyping of cultivars, mapping and tagging of genes, and diversity analysis of germplasm. Finally, the perspective of molecular marker’s application in tobacco genetic breeding in the future was discussed.

Key words: Molecular marker, tobacco, genetic and breeding.

INTRODUCTION

Tobacco (Nicotiana spp.) has been cultivated by man for thousands of years and has served as a medicinal herb, trade commodity, and crop plant in many different cultures. Nowadays, it becomes one of the most important commercial crops in the world. Within the past several decades, this plant has found yet another use, serving as a widely utilized model system in plant-cell culture and genetic-engineering research (Zhang et al., 2007). Because of its economic importance and value as a biological research tool, numerous investigations have been undertaken to examine its evolutionary origin and genome structure and organization. Morphological, karyotypical, and physiological characters have already been used to study the genetic background of tobacco (Goodspeed, 1945; Zhang, 1994; Lei et al., 1997; Lu, 1997). However, morphological characters usually vary with environments. The number of karyotypical characters is limited, and the study of genotypic diversity based on isozyme variation is restricted to a few polymorphic enzyme systems encoded by a small number of loci (Lu, 1997). With the advantages such as highly polymorphic nature, codominant inheritance, easy access, easy and fast assay, high reproducibility and easy exchange of data between laboratories, a new generation of markers has been introduced over the last two decades, which has revolutionized the entire scenario of biological sciences. DNA-based molecular markers have acted as versatile tools and have found their own position in various fields like characterization of genetic variability, genome fingerprinting, genome mapping, gene localization, analysis of genome evolution, population genetics, taxonomy, plant breeding, and diagnostics. Nowadays, these molecular markers have been employed to study the genetics and breeding of tobacco.

For simplicity, we have divided the review in three parts. The first part is a general description of most of the available DNA marker types employed in tobacco research, the second includes their application in tobacco genomics and breeding programmes, and the third is the perspective of application of molecular markers in tobacco genetic breeding.

TYPES AND DESCRIPTION OF DNA MARKERS USED IN THE TOBACCO RESEARCH

Restriction fragment length polymorphism (RFLP)

RFLP is the identification of specific restriction enzymes
that reveals a pattern difference between the DNA fragment sizes in individual organisms. To discover RFLPs, restriction enzymes (RE) are used to cut DNA at specific 4-6 bp recognition sites. Sample DNA is cut with one or more RE’s and resulting fragments are separated according to molecular size using gel electrophoresis (Botstein et al., 1980). Differences in fragment length result from base substitutions, additions, deletions or sequence rearrangements within RE recognition sequences. RFLPs are simply inherited naturally occurring Mendelian characters. RFLPs, being codominant markers, can detect coupling phase of DNA molecules, as DNA fragments from all homologous chromosomes are detected. They are very reliable markers in linkage analysis and breeding and can easily determine if a linked trait is present in a homozygous or heterozygous state in an individual, the information highly desirable for recursive traits (Winter and Kahl, 1995). However, the large amount of DNA required for restriction digestion and Southern blotting hampered the utility of RFLP. The assay is time-consuming and labour-intensive and only one out of several markers may be polymorphic, which is highly inconvenient especially for crosses between closely related species.

Randomly amplified polymorphic DNA markers (RAPD)

In 1990, Williams et al. developed a new PCR-based genetic assay namely randomly amplified polymorphic DNA (RAPD). The standard RAPD technology utilizes short synthetic oligonucleotides (10 bases long) of random sequences as primers to amplify nanogram amounts of total genomic DNA under low annealing temperatures by PCR. Amplification products are generally separated on agarose gels and stained with ethidium bromide (Williams et al., 1990). On an average, each primer directs amplification of several discrete loci in the genome, making the assay useful for efficient screening of nucleotide sequence polymorphism between individuals. Since RAPD amplification is directed with a single, arbitrary and short oligonucleotide primer, DNA virtually from all sources is amenable to amplification. However, due to the stoichiastic nature of DNA amplification with random sequence primers, it is important to optimize and maintain consistent reaction conditions for reproducible DNA amplification. RAPD markers are dominant markers and have found a wide range of applications in gene mapping, population genetics, molecular evolutionary genetics and plant and animal breeding. This is mainly due to the speed, cost and efficiency of the RAPD technique to generate large numbers of markers in a short period compared with previous methods. Therefore, RAPD technique can be performed in a moderate laboratory for most of its applications. Because the chance of comigrating bands being homologous becomes less as populations diverge, it was suggested that RAPD analysis gives more accurate estimates between closely related populations and less accurate estimates for distantly related populations (Williams et al., 1993; Smith and Williams, 1994). Despite problems such as poor reproducibility faint or fuzzy products, and difficulty in scoring bands, the RAPD method will probably be important as long as other DNA-based techniques remain unavailable in terms of cost, time and labour.

Amplified fragment length polymorphism (AFLP)

AFLP is a technique based on the detection of genomic restriction fragments by PCR amplification and can be used for DNAs of any origin or complexity (Vos et al., 1995). After final amplification, selectively amplified fragments are separated by gel electrophoresis and visualized autoradiographically. Msel-Msel fragments are excluded from the autorad because only EcoRI-directed primers are normally labeled. Using 3-bp selective primer extensions gives 128 possible linker combinations. Therefore, 128 subsets of genomic DNA can be readily amplified. Thus, thousands of markers can be generated quite rapidly. Typically, the autorad has 100-300 fingerprints with sizes ranging from 80 to 500 nucleotides. Most AFLP fragments correspond to unique positions on the genome and hence can be exploited as landmarks in genetic and physical mapping (Vos et al., 1995). AFLPs are reliable and extremely useful as tools for DNA fingerprinting and also for cloning and mapping of variety-specific genomic DNA sequences (Hongtrakul et al., 1997; Yong et al., 1996; Paglia et al., 1998). Of cause, there are still some weaknesses such as difficult in developing locus-specific markers from individual fragments and need to use different kits adapted to the size of the genome being analyze.

Microsatellites and minisatellites

The term “microsatellites” was coined by Litt and Lutty (1989), while the term minisatellites was introduced by Jeffrey et al. (1985). Both are multilocus probes creating complex banding patterns and are usually non-species specific occurring ubiquitously. They essentially belong to the repetitive DNA family. Microsatellites are repeated as of only a few bases, like two or three or five, and the whole repetitive region spans less than 150 bp. Minisatellites are larger, usually from 9 up to 80 bp. The size of a minisatellite ranges from 1 to 20 kb. Microsatellites and minisatellites form an ideal marker system creating complex banding patterns by simultaneously detecting multiple DNA loci. The number of repeats for a given microsatellites or minisatellite may differ between individuals. This feature is the basis of DNA fingerprinting. They are located in non-coding regions. Many alleles exist in a population, the level of heterozygosity is high and they follow Mendelian inheri-
tance. Some of the prominent features of these markers are that they are dominant fingerprinting markers and codominant STMS (sequence tagged microsatellites) markers.

APPLICATIONS OF MOLECULARS IN TOBACCO GENOME ANALYSIS AND BREEDING

Molecular markers have been looked upon as tools for a large number of applications ranging from localization of a gene to improvement of plant varieties by marker-assisted selection. They have also become extremely popular markers for phylogenetic analysis adding new dimensions to the evolutionary theories. If we look at the history of the development of these markers, it is evident that they have been improved over the last two decades to provide easy, fast and automated assistance to scientists and breeders. Genome analysis based on molecular markers has generated a vast amount of information and a number of databases are being generated to preserve and popularize it.

Evolutionary genetics

The advances in DNA techniques have had a great impact in addressing problems in many aspects of biology. Comparative biochemical studies and examinations of organellar (plastid and mitochondrial) genome organization support the viewpoint that *Nicotiana sylvestris* is the maternal parent and *Nicotiana tomentosiformis* the paternal parent (Shinshi et al., 1988; Sperisen et al., 1991). Analysis of molecular features, such as repetitive DNA sequences and the structure of various nuclear gene family members, has also been employed to trace the molecular evolution of tobacco and study the genetic diversity in genus *Nicotiana* (Kuhrová et al., 1991; Komarnitsky et al., 1998; Lim et al., 2000a, 2000b). In most cases, these studies also concluded that *N. tomentosiformis* was the progenitor parental donor along with *N. sylvestris*. Based on investigations of the structure and organization of mitochondrial and chloroplast genomes, Olmstead and Palmer (1991) demonstrated that a species similar to *N. sylvestris* donated the maternal genome of tobacco. Lim et al. (2000a) suggested that the majority of the genes present in *N. tabacum* were contributed by *N. sylvestris*, with the remainder being contributed by *N. tomentosiformis*. Pairwise comparisons of the AFLP profiles of wild and cultivated *Nicotiana* species show that polymorphic bands present in *N. tabacum* can be found in at least one of three proposed wild progenitor species (i.e., *N. sylvestris*, *N. tomentosiformis*, and *Nicotiana otophora*), and this observation provides additional support for these species contributing to the origin of *N. tabacum* (Ren and Timko, 2001).

Population genetics and genotyping of cultivars

Application of DNA-based approaches to population genetic studies has been limited, probably due to the need for large samples of individuals from each population to provide an accurate estimate of allele and genotype frequencies. The relatively high cost, the requirement for sophisticated equipment and well-trained personnel, and low speed are other limiting factors in population genetic studies. The RAPD technique has received a great deal of attention from population geneticists because of its simplicity and rapidity in revealing DNA-level genetic variation, and therefore has been praised as the DNA equivalent of allozyme electrophoresis (Hedrick, 1992; Skibinski, 1994). Numerous types of tobacco are grown commercially and are defined to a large extent by region and (or) area of production, method of curing, and intended use in manufacturing, as well as by some distinct morphological characters and chemical differences (Wernsman 1987). At the present time, only limited information is available on the relationship between morphological variability and genetic diversity in cultivated tobacco. Attempts have been made to examine the degree of relatedness between tobacco cultivars and types at the molecular genetic level. The repetitive and arbitrary DNA markers are markers of choice in genotyping of cultivars. Zhang et al. (2005) employed the RAPD analysis to assess the polymorphism, similarities and relationships among *N. tabacum* L. cultivars. One hundred and forty-nine bands were detected, of which 94 were polymorphic (63.1%). A primer distinguishing all of the tested cultivars was found. The result showed that RAPD assay could discriminate those flue-cure tobacco cultivars with similar genotypes. Genetic diversity and genotyping in Indian FCV and burley tobacco cultivars was also investigated by Saraka and Rao using RAPD technique (Saraka and Rao, 2008). In these studies, the markers found specific to the varieties can be used in correct identification of the carrier genotypes in trade and commerce.

Mapping and tagging of genes

The most economic and effective method to control the disease is the utilization of natural genetic resistance and breeding of resistant cultivars. And one of the most widely used applications of the molecular technique is the identification of markers linked to traits of interest without the necessity for mapping the entire genome. Using tightly linked molecular markers to target genes to screen progenies is one of the most effective methods to carry out artificial selection. Identification and utilization of new resistant genes are important for breeding new tobacco cultivars with enhanced resistance to disease. Molecular-assisted breeding of disease resistant tobacco plants can avoid most inoculation procedure of disease to
cut down on workload, can identify the resistance of plants in early stage of breeding, increasing selection veracity and efficiency, and can accelerate reasonable and rapid utilization of the resistant gene, shortening the breeding cycle. Near-isogenic lines (NILs) and bulked segregant analysis (BSA) are both efficient methods to isolate DNA segments linked to certain traits. Markers tagged and mapped with specific genes in tobacco have been identified to include a lot of important disease resistant gene (Bai et al., 1995; Bai et al., 1996; Yi et al., 1998a; Yi et al., 1998b; Kenward et al., 1999; Johnson et al., 2002; Zhang et al., 2008a; Liu et al., 2009). Julio et al. (2006b) conducted Amplified fragment length polymorphism (AFLP) on a set of 92 *N. tabacum* L. accessions from diverse types and breeding origins to identify markers associated with disease resistances. In his study, seven fragments were associated with three different resistances: two for the blue-mold (*Peronospora tabacina* Adam) resistance derived from *Nicotiana debneyi* Domin, two for the Vn gene (Potato Virus Y susceptibility), and three for the black root rot (*Chalara elegans*) resistance of *N. debneyi* origin. The finding of linked marker of disease resistant gene on the DNA level can be applied to assist selection, increase breeding efficiency and provide the base for gene cloning later.

**Diversity analysis of germplasm**

Germplasm analysis to study genetic diversity is another important area in which a lot of efforts have been put in. Characterization and quantification of genetic diversity has long been a major goal in breeding. In plant breeding programs, information on genetic diversity is essential for a rational use of genetic resources. It is particularly useful in characterizing individual accessions and cultivars, in detecting duplications of genetic materials in germplasm collections, and as a general guide in selecting parents for hybridization in breeding programs and in developing informative mapping populations for genome mapping. Studies revealed that cultured tobacco groups were divided primarily based upon geographic origin and manufacturing quality traits (Ren and Timko, 2001; Zhang et al., 2006; Zhang et al., 2008). The amount of genetic polymorphism present among cultivated tobacco lines (*N. tabacum*) was limited (Ren and Timko, 2001; Zhang et al., 2006; Siva-Raju et al., 2008; Zhang et al., 2008), as evidenced by the high degree of similarity. A greater amount of genetic polymorphism exists among wild species of *Nicotiana* than among cultivated forms (Ren and Timko, 2001). The high degree of genetic polymorphism among species that was observed by Ren and Timko (2001) using AFLP analysis is consistent with the observations of Bogani et al. (1997) following their analysis of interspecific variation using RAPD analysis. These studies indicated that the present day commonly grown tobacco germplasm has narrow genetic diversity among the cultivars, necessitating a sustained effort to preserve tobacco germplasm resources. In another study, amplified fragment length polymorphism (AFLP) analysis was used to determine genetic variation in 54 varieties of cultivated tobacco (*N. tabacum* and *N. rustica*) and three accessions of exotic germplasm.

**PERPECTIVE OF MOLECULAR MARKERS’ APPLICATION IN TOBACCO GENETICS AND BREEDING**

Narrow genetic diversity has been revealed among the commonly grown tobacco germplasm (Ren and Timko, 2001; Zhang et al., 2006; Zhang et al., 2008), and it is possible that a large proportion of valuable tobacco germplasm may already have been lost through the popularity of certain cultivars in commercial planting and the continuous artificial selection (Zhang et al., 2006). Since greater genetic variation exists among wild species of *Nicotiana* than among cultivated forms (Ren and Timko, 2001), to avoid further degradation of germplasm resources, crosses should be made with genetically distant varieties or wild species. More information of genetic variation within commercially cultivated tobacco between cultivated forms of *N. tabacum* and its wild relatives should be revealed by molecular markers. Species-specific markers identified in studies are useful in identification of the true hybrids and monitoring introgression of useful genes from the wild relatives.

Quantitative genetic studies, including the use of quantitative trait locus (QTL) mapping techniques, provide an opportunity to investigate the underlying genetic mechanisms that regulate developmental programs in plant architecture. The information of QTL has potential in strategic planning of future breeding towards tobacco sustainability. Many attempts are being made towards pyramiding different resistance genes for a specific disease or pest attack (Nishi et al., 2003). Although some tobacco quantitative trait loci (QTLs) have been investigated in a recombinant inbred line (RIL) population (Julio et al., 2006a), other important QTLs like alkaloid contents and manufacturing quality traits (e.g., wrapper quality, filler quality) should also be investigated.

Construction of genetic linkage map is an important part of genetics research. Linkage map would provide the basis for the genetic mapping of traits in tobacco and for further analyses of the tobacco genome. But reports on linkage map of tobacco are still very limited (Bindler et al., 2007). Linkage map should also be generated through different kink of molecular marker such as RAPD, AFLP, and microsatellite markers.

**ACKNOWLEDGMENTS**

The study was supported by grants from the Natural Science Foundation of Yunnan (Grant No. 2007 C 216 M), Educational Commission of Yunnan (Grant No. 0820025) and “948” Project of Ministry of Agriculture of China (Grant No. 2008-4-30).

---

**ACKNOWLEDGMENTS**

The study was supported by grants from the Natural Science Foundation of Yunnan (Grant No. 2007 C 216 M), Educational Commission of Yunnan (Grant No. 0820025) and “948” Project of Ministry of Agriculture of China (Grant No. 2008-4-30).
REFERENCES


