Int. J. Indust. Entomol. Vol. 19, No. 1, 2009, pp. 143~154

[Review]

International Journal of
Industrial Entomology

Single Nucleotide Polymorphisms (SNPs) for Advanced Genomic Research in Sericulture

Kunjupillai Vijayan*
Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan-115, ROC.

(Received 26 February 2009; Accepted 20 August 2009)

Single nucleotide polymorphisms (SNPs) are the most frequent form of variation in the genome of any organism. Owing to their greater abundance, they are considered useful for identifying cultivars, construction of higher density linkage maps, and detection of genes (QTLs) associated with complex agronomic traits and diseases. Although, SNPs have been used recently for constructing a high density genetic map in silkworm and a set of 118 SNPs have been identified in tasar silkworms, not much progress has been made in sericulture to utilize the vast potential of SNPs. Thus, this review mainly focuses on some of the important methods of SNP discovery, validation and genotyping. Emphasis has also been given to the possible uses of SNP genotyping in the improvement of silkworms and their host plants.

Key words: Mulberry, Sericulture, Silkworm, Single nucleotide polymorphism

Introduction

Single nucleotide polymorphisms are single base variations widely distributed throughout genomes of individuals with the frequency of >1% in a population (Halushka *et al.,* 1999). SNPs represent the most common type of genetic variations present in the genome and are accounting for approximately 90% of all polymorphisms (Collins *et al.,* 1997). It has been estimated that SNPs occur at a frequency of one in every thousand nucleotides in the human genome, though the precise estimates of SNP frequency are difficult to determine and often vary across species, populations and even genomic regions. For instance, in maize SNP has been reported to occur in one out of every 60~120 bp (Ching *et al.,* 2002), in sugar beat its frequency is one in every 130 bp with an average nucleotide diversity of 7.6×10^{-3} (Schneider *et al*., 2001), while in humans it is estimated that at least 1 SNP is present in every 1000 bp genome (Sachidanandam *et al.,* 2001). In practice, SNPs are predominantly bi-allelic, even though in principle any of the four nucleotides can be present at any position in a stretch of DNA. This contradiction is due to the low frequency of mutations that reaches to the level of new SNPs i.e., >1% in the population. Basically, SNPs are formed from transitions and transversions. In transitions, a purine is exchanged for the other purine (A<=>G), while on the reverse strand a pyrimidine is exchanged for the other pyrimidine (C<=>T). Transversions consist of purine-pyrimidine (and their complementary pyrimide-purine exchanges): A<=>C (T<=>G), A<=>T (T<=>A), C<=>G (G<=>C). The occurrences of transitions and transversions are not equal in the genome. Although there are three times as many options for transversions as for transitions, transitions are found twice as often as transversions. The higher level of transitions might be related to 5-methylcytosine deamination reactions which occur primarily in CpG doublets. Although a repair mechanism specifically recognizes G-T mismatches and replaces the thymine with cytosine, the 5-methylcytosine to T transition mutation occurs about 10 times more often than other transitions (Holliday and Grigg, 1993). SNPs are present in exons, introns, promoters, enhancers, and intergenic regions. Because SNPs can be analyzed using high throughput and cost effective systems, they are useful for rapid identification of crop cultivars, construction of ultra high-density genetic maps, association studies related to genetic disorders and identification of disease alleles for complex diseases (Kota *et al.,* 2003). It is expected that SNP panel may replace mic-

*To whom the correspondence addressed
Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan-115, ROC, Tel: +886-227871078;
Email:kvijayan01@yahoo.com

rosatellite panels as informative markers in linkage studies very soon (Kruglyak, 1997). Since SNPs in coding regions can lead to alteration of amino acid sequences or early termination of translation and can therefore affect gene function, these functional SNPs would be valuable markers if the altered gene function or altered phenotype is of importance to the breeders (Van *et al.,* 2005). SNPs could also be used in pharmacogenetics for the development of personalized drugs. It has been estimated that there are likely to be around 10 million common SNPs in the human genome and that these common SNPs account for around 90% of human genetic variation (Carlson *et al.,* 2003; Reich *et al.,* 2003). Worldwide public efforts have so far identified over 7 million SNPs and more than 3 million of which have already been validated (www.ncbi.nlm.nih. gov/SNP/index.html).

Present scenario of utilization of SNPs in sericulture

Since, genotyping provides a measurement of the genetic variation between members of a species that are essential for geneticists and breeders for understanding the genome architecture and to devise strategies for genetic improvements of the organism, taking advantages of the recent advances in the development of molecular markers many attempts have been made to unravel the genetic diversity present in mulberry and silkworm breeding stocks, land races and wild populations. Molecular markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), inter simple sequence repeats (ISSR), simple sequence repeats (SSR), amplified fragment length polymorphism (AFLP) have been used extensively for genotyping of *Morus* spp. (Chatterjee *et al.,* 2004a; Sharma *et al.,* 2000; Vijayan *et al.,* 2004, Vijayan, 2007; Zhao and Pan, 2004), *Bombyx mori* (Nagaraja and Nagaraju, 1995; Reddy *et al.,* 1999a, b; Sethuraman *et al.,* 2002; Chatterjee and Mohandas, 2003), *Samia cynthia ricini* (Vijayan *et al.,* 2006), and *Antherea mylitta* (Chatterjee *et al.,* 2004b; Kar *et al.,* 2005; Mahendran *et al.*, 2006). Genetic linkage maps have also been made for the silkworm, *B. mori,* (Mase *et al.,* 2007; Nagaraja *et al.,* 2005; Promboon *et al.,* 1995; Prasad *et al.,* 2005; Shi *et al.,* 1995; Tan *et al.,* 2000) and its host plant mulberry (Venkateswaralu *et al.,* 2007). However, baring the efforts made by Cheng *et al.* (2004) to identify candidate SNPs from EST data and Yamamoto *et al.* (2008) to construct a linkage map using SNPs from BAC end-sequences, use of this highly potential new generation molecular marker has not been advanced for the benefit of sericulture. Therefore, it is my feeling that it is high time for the molecular biologists working in sericulture to make sincere effort to utilize SNPs for the benefit of the industry. With this objective, I have made an overview of the most commonly used methods for discovering, validating and genotyping of SNPs. While reviewing the applications of SNPs that have already been made on silkworm genetic analysis, special emphasis has been given to point out how utilization of data from SNPs can advance our knowledge on the genetics and genomics of silkworms and their host plants.

Discovery and validation of SNPs

Development of SNP markers usually follow a three-part strategy consisting of SNP discovery based on analysis of a small set of individuals, validation in a larger set and then genotyping in a larger population. There are many methods currently available for discovering SNPs. If prior information of the genome is available then locus specific primers can be synthesized for PCR amplification from several individuals of the same species. The PCR amplicons are then sequenced and alignment of the sequence is used for identifying SNPs. If no information is available on the genome of the species, whole genome shotgun sequencing method is the option. In this method random clones from genomic library prepared from a mixture of DNA from several individuals are sequenced. The sequences can then be aligned to detect SNPs. In this way, SNPs from several loci can be obtained in a species. The reduced representation shotgun (RRS) is a modified form of genome sequencing, in which equimolecular quantities of genomic DNA from several unrelated individuals is mixed, digested with appropriate enzyme, electrophoresed, cut a narrow band of clonable size, re-extracted the DNA from the band, made partial genomic library and clones of which are sequenced. Sequences representing the same loci are aligned to discover putative SNPs (Gupta *et al.,* 2001). Overlapping of sequences from BAC clones has also been used to detect single nucleotide polymorphism among individuals of populations or species. Identification of SNPs from RFLP, STS, CAPS, RAPD and AFLP markers is also quite possible. Markers linked tightly to particular genes can be isolated and sequenced to identify putative SNPs. Examples of conversion of RFLP, CAPS and RAPD markers into SNPs have been reported extensively. For instance, tightly linked RFLP and CAPS markers were developed and studied for wheat leaf rust resistance gene *Lr19* and *Lr47* (Tyrka *et al.,* 2004). RAPD and CAPS markers were used for identification of SNPs for nucleus-cytoplasm compatibility gene *Ncc-tmp1A* of wheat and *Aegilops squarrosa* hybrids (Asakura *et al.,* 2001). Nevertheless, with the growth of high-throughput sequencing technology, large amounts of data have been submitted to various DNA databases and that could be used for SNP discovery (Taillon-Miller *et al.,* 1998). Sequences from these databases, at least if

sequences from more than two individual of the same population/species are available, can be retrieved to create multiple alignments using standard softwares like ClustalW. Once a multiple alignment is constructed, nucleotide differences among the individual sequences can be analyzed. Owing to the presence of sequencing errors, not every nucleotide position with mismatches automatically implies a polymorphic site. Although it is impossible to decide which the case is with certainty, the success of SNP detection ultimately depends on how well one is able to discriminate true polymorphisms from the likely sequencing errors. This is usually accomplished by statistical considerations that take advantage of measures of sequence accuracy accompanying the analyzed sequences (Ewing and Green, 1998; Ewing *et al.,* 1998). The result, ideally, is a set of candidate SNPs, each with an associated SNP score that indicates the confidence of the prediction. Accurate confidence values can be extremely useful for the experimentalist in selecting which SNPs to use in a study or for further characterization, and enables one to use the highest number of candidates within the bounds of an acceptable false positive rate to identify putative SNP sites. There are many base-calling algorithms like PHRED, which can identify sequencing errors (Ewing and Green 1998) and many soft wares that can identify SNPs from aligned sequences, e.g. AutoSNP (http://hornbill.cspp.latrobe.edu.au/), Forage (http://biobase.biotech.kth.se/forage/). Once the putative SNPs are identified, they can be validated using a small number of individuals with any of the following detection methods.

Genotyping of SNP markers

A number of methods, novel and high-throughput, have been developed for SNP genotyping and most of these methods predominantly involve pre-amplification of the target region from genome followed by the identification of specific alleles by restriction endonuclease digestion, hybridization probes, mismatched DNA substrates in conjunction with polymerase or ligase, or template dependent nucleotide incorporation. A brief overview of some of the important and widely applied technologies for SNP genotyping is given below.

Restriction Fragment Length Polymorphism (RFLP)-PCR analysis

Restriction fragment length polymorphism (RFLP)-PCR analysis is considered to be the simplest and earliest method to detect SNPs. It makes use of the many different restriction endonucleases and their high affinity to unique and specific restriction sites. In this method, a specific target region of genomic DNA is amplified by PCR and the PCR amplicon is then digested with individual restriction enzymes and visualized after being gel-electrophoresed. If the SNP alters a restriction site the restriction pattern is changed. A major limitation of this method is that only those SNPs which change the restriction enzyme sites are detected. Likewise, the processes of gel electrophoresis are not yet readily automated.

Single-strand confirmation polymorphism (SSCP)

It is one of the most widely used mutation detection system. The principle behind this technique is that under neutral condition, the single-stranded DNA (ssDNA) folds into a tertiary structure. The conformation is sequence dependent and most single base pair mutations will alter the shape of the structure. When applied to a gel, the tertiary shape will determine the mobility of the ssDNA, providing a mechanism to differentiate between SNP alleles. This method first involves PCR amplification of the target DNA. The double-stranded PCR products are denatured using heat and formaldehyde to produce ssDNA. The ssDNA is applied to a non-denaturing electrophoresis gel and allowed to fold into a tertiary structure. Differences in DNA sequences alter the tertiary conformation, which in turn affect the mobility of the ssDNA. Based on their mobility differences, SNPs are detected (Costabile *et al*., 2006). This method is widely used because it is technically simple, relatively inexpensive and uses commonly available equipment. However compared to other SNP genotyping methods, the sensitivity of this assay is lower. It has been found that the ssDNA conformation is highly dependent on temperature and it is not generally apparent what the ideal temperature is. Therefore, it is often required that the assay needs to be repeated under different temperatures. Another, disadvantage of this technique is the size limitation of the DNA fragment as the sensitivity of the technique reduces with higher fragment size, generally when bigger than 400 bp are used (Costabile *et al*., 2006).

Dynamic allele-specific hybridization (DASH)

Dynamic allele-specific hybridization (DASH) is essentially an advanced form of allele-specific hybridization. This technique takes advantage of the differences in the melting temperature of the DNA that results from the instability of mismatched base pairs. The analysis starts with PCR amplification of the genomic DNA. One of the primers used for PCR amplification is biotinylated so that the PCR product can be attached to a streptavidin column and washed with NaOH to remove the unbiotinylated second strand. The biotinylated single strand DNA that retained on the streptavidin column is then hybridized to an allele specific oligonucleotide probe containing the SNP site, in the presence of a low-cost fluorescent dye

that intercalates only into double-stranded DNA. The intensity of the fluorescence is then measured by increasing the temperature until the Tm can be determined. A rapid and sudden fall in fluorescence indicates the melting temperature of the DNA duplex. A PCR product with an SNP site shows a low *Tm* than those without an SNP (Howell *et al.,* 1999).

Molecular beacons

Molecular beacons are oligonucleotide probes that have two complementary DNA sequences flanking the target DNA sequence and a fluorophore/quencher dye pair at opposite ends of each probe (Abraya *et al.,* 2003; Tyagi *et al.,* 1998). When it is not hybridized to the target DNA, the probe adopts a hairpin loop conformation due to the self-annealing of the two complementary ends bringing reporter and the quencher dyes close together. This prevents the flurophore from emitting any fluorescence. However, when the probe encounters its target genomic DNA during the assay, it will anneal and hybridize. Because of the length of the probe sequence, the hairpin segment of the probe will be denatured in favor of forming a longer, more stable probe-target hybrid. This conformational change separates the donor dye from the quencher dye, which permits emission of fluorescent signals. The probe sequence is designed in such a way that if it encounters a target sequence with as little as one non-complementary nucleotide, the molecular beacon will preferentially stay in its natural hairpin state and no fluorescence will be observed, as the fluorophore remains quenched. For SNP genotyping, two molecular beacons with exact sequence matches to the wild-type and variant alleles are used in the same PCR. The use of two differentially labeled molecular beacons in the same PCR reaction allows simultaneous detection of three possible allelic combinations.

SNP microarrays

A DNA microarray, small chip, generally about a square centimeter, most commonly made of glass, plastic or silicon, has revolutionized the application of high-throughput genotyping in the last few years. In high density oligonucleotide SNP arrays, hundreds of thousands of oligonucleotides are attached to a small chip in an ordered array, allowing for many SNPs to be interrogated simultaneously (Rapley and Harbron, 2004). The DNA of interest is PCR amplified to incorporate fluorescently labeled nucleotides and then hybridized to the chip (Lipshutz *et al.,* 1999). Each oligonucleotide in the array acts as an allele-specific probe. Well-matched sequences hybridize strongly than mismatched probes. Upon stringent washing, the mismatched probes are washed away. The signals from the hybridized probes are quantified by high resolution fluorescent scanning and analyzed by sophisticated programs. Although oligonucleotide microarrays have a comparatively lower specificity and sensitivity, the scale of SNPs that can be interrogated is a major benefit. Many biotechnology companies like Affymetrix (Santa Clara CA), and Nanogen (San Diego, CA) are marketing such DNA macroarrays.

Allele-specific amplification

Allele-specific amplification, also known as the amplification refractory mutation system (ARMS), is also a commonly used method for SNP detection (Newton *et al.,* 1991). This procedure is based on the observation that PCR amplification occurs only if nucleotides near the 3' end of the PCR primer are exactly complementary to the primer-binding site on the target sequence. Therefore, one primer is designed in such a way that one of primer will have its 3' nucleotide sits over the polymorphic SNP site. In its simplest form the PCR products are gel electrophoresed and the banding pattern is visualized with ethidium bromide staining. The advantage of this method is the detection of the amplification products on agarose gel. On the other hand, the single base-pair change at the 3' primer end is often not sufficient to ensure reliable discrimination between the two SNP alleles (Ahmadian *et al*., 2001). To overcome these disadvantages, some improvement was achieved with PCR amplification with a new technique called mutagenically separated PCR (MS-PCR). In this method, two ARMS primers of different lengths for each allele are used. Thus, the resulting amplicons will be of different lengths for the two alleles. The major advantage of this method is that PCR amplification of at least one of the two allelic should happen, therefore it exclude the possibility of false-negative results resulting from non-amplification.

TaqMan assay

Taking advantages of the 5'-3' exonuclease activity of the *Taq* DNA polymerase the TaqMan assay was designed based on the principles that the fluorescence resonance energy transfer (FRET) system requires two different oligonucleotide probes carrying a donor fluorophore at the 3'-end and an acceptor fluorophore at the 5'-end. While the probe is intact, the quencher will remain in close proximity to the fluorophore, eliminating the fluorophore's signal. The assay is performed with both forward and reverse PCR primers to amplify a region that includes the SNP polymorphic site. During the PCR amplification step, if the allele-specific probe is perfectly complementary to the SNP allele, it will bind to the target DNA strand and then get degraded by 5'-nuclease activity of the *Taq* polymerase as it

extends the DNA from the PCR primers. The degradation of the probe results in the separation of the fluorophore from the quencher molecule, generating a detectable signal. If the allele-specific probe is not perfectly complementary, it will have lower melting temperature and not bind as efficiently. This prevents the nuclease from acting on the probe (McGuigan and Ralston, 2002). Since the Taqman assay is based on PCR, it is relatively simple to implement. The Taqman assay can be multiplexed by combining the detection of up to seven SNPs in one reaction. The assay is usually carried out in a 96-well format, but availability of a 384-well version of Applied Biosystems' AB7900HT real-time thermocycler with a capacity of 100,000 SNP typings per day enables very high throughput. The specificity of the approach is based on differential hybridization of the two probes to their corresponding target sequences. As with all systems based on hybridization, it can sometimes be difficult to achieve very specific binding of probes to one or other allele.

Invader assay

The Invader assay is based on the unique capability of a class of natural enzymes called flap-endonucleases and engineered enzymes termed cleavases (Olivier, 2005). DNA molecules are cleaved at specific structures produced by the addition of sequence specific oligonucleotides to genomic DNA. The oligonucleotide that binds on the upstream of the SNP site is called invader, while the other one that binds to the downstream is called probe. The resulting adjacent duplexes have to overlap by at least one nucleotide to create an efficient substrate. The last base of the invader oligonucleotide is a non-matching base that overlaps the SNP nucleotide in the target DNA. The 5'-end of the probe oligonucleotide contains an unpaired region called the "flap". If the target DNA contains the desired allele, the Invader and allele-specific probes will bind to the target DNA forming the tripartite structure. This structure is recognized by cleavase, which will cleave and release the flap from the 3' end of the allele-specific probe. If the SNP nucleotide in the target DNA is not complementary allele-specific probe, the correct tripartite structure is not formed and no cleavage occurs. The Invader assay is usually coupled with fluorescence resonance energy transfer (FRET) system to detect the cleavage event. In this setup, a quencher molecule is attached to the 3' end and a fluorophore is attached to the 5' end of the allele-specific probe. If cleavage occurs, the fluorophore will be separated from the quencher molecule generating a detectable signal (Olivier, 2005). An advantage of the Invader assay is that it does not require thermocycling because of its isothermic, balanced equilibrium. Each target-specific product enables the cleavage of many FRET probes. Under standard conditions ca. 10^6~10^7 labeled cleaved flaps are produced per hour. Another advantage of this assay is that it works with genomic DNA and does not require PCR amplification thereby avoiding potential contamination problems.

MALDI-TOF mass spectrometry

Matrix assisted laser desorption ionization time of flight (MALDI–TOF) mass spectrometry (MS) was first introduced by Karas and Hillenkamp (1998) as a revolutionary method for ionizing and mass analyzing large biomolecules. They discovered that irradiation of crystals formed by suitable small organic molecules with a short laser pulse at a wavelength close to a resonant absorption band of the matrix molecules caused an energy transfer and desorption process, producing gas phase matrix ions. More importantly, they also found that if a low concentration of a non-absorbing analyte, such as a protein or nucleic acid molecule, was added to the matrix in solution and embedded in the solid matrix crystals formed by drying of the mixture, the nonabsorbing, intact analyte molecules were also desorbed into the gas-phase and ionized upon irradiation with the laser, facilitating their mass analysis. Typically, predominantly singly charged molecular ions (both negative and positive) are detected by MALDI–TOF MS, with these ions being created in a gas-phase proton transfer reaction with the matrix molecules. Analyzing nucleic acids with MALDI–TOF MS has several advantages. Firstly, it is fast; ionization, separation by size and detection of nucleic acids takes milliseconds to complete. Because signals from multiple laser pulses (~20–100 pulses) are usually averaged to obtain a final mass spectrum, the total analysis time can take as little as ten seconds. Secondly, the results are absolute, being based on the intrinsic property of the mass-to-charge ratio (m/z). This is inherently more accurate than electrophoresis-based or hybridization-array-based methods, which are both susceptible to complications from secondary-structure formation in nucleic acids (Griffin and Smith, 2000). Furthermore, enzymatic DNA sequencing coupled with MALDI–TOF mass spectrometric analysis has been shown to be effective at discovering previously unknown single-nucleotide substitution mutations (Kirpekar *et al.*, 1998). The greatest promise of MALDI-TOF MS for SNP analysis arises from its ability to genotype many SNPs rapidly, accurately and simultaneously. However, MALDI-TOF MS suffers from the draw back of signal intensity loss with increasing DNA size, owing to the size dependent tendency of phosphodiester backbone of the DNA to fragment during MALDI process. Secondly it requires a larger investment in capital expenditure and technical expertise, but has a lower reagent cost.

Pyrosequencing

Pyrosequencing is a rapid-sequencing technique, particularly suitable for sequencing short segments of PCR-amplified DNA, based on detection of released pyrophosphate (PPi) during DNA synthesis. In a cascade of enzymatic reactions, visible light is generated that is proportional to the number of incorporated nucleotides. The cascade starts with a nucleic acid polymerization reaction in which inorganic PPi is released as a result of polymerase-mediated incorporation of nucleotides. The released PPi is subsequently converted to ATP by ATP sulfurylase, which provide the energy to luciferase to oxidize luciferin and produce light. Since the Results are automatically assigned by pattern recognition. Pyrosequencing offers many advantages as the quality of the results is high with very clear distinction between genotypes, the approach does not require a high level of technical expertise, and the quantitative nature of the signals produced makes it suitable for gene-dosage studies, including estimation of gene frequencies on pooled DNA. Some of these features make pyrosequencing highly amenable for numerous clinical applications. However, its major disadvantage is the cost of the reagents and a limited capacity for multiplexing (Ronaghi *et al.,* 1998).

Single-base extension or minisequencing

In single-base extension or minisequencing an oligonucleotide primer is annealed to the single strand of a PCR amplicon, with the 3' end of the primer binding one base short of the SNP site. A mixture of all the four dNTPs and one of the four possible ddNTPs that corresponds to the SNP locus is used for primer extension. DNA polymerase then catalyzes extension of the primer and the ddNTP is incorporated to the base at the SNP site. This incorporated base is detected and determines the SNP allele (Syvanen, 2001). Because, primer extension is based on the highly accurate DNA polymerase enzyme, the method is generally very reliable. The flexibility and specificity of primer extension make it amenable to high throughput analysis. Primer extension probes can be arrayed on slides allowing for many SNPs to be genotyped at once.

Oligonucleotide ligase assay

DNA ligase catalyzes the ligation of the 3' end of a DNA fragment to the 5' end of a directly adjacent DNA fragment. This mechanism can be used to interrogate a SNP by hybridizing two probes directly over the SNP polymorphic site, whereby ligation can occur if the probes are identical to the target DNA. In the oligonucleotide ligase assay, two probes are designed; an allele-specific probe which hybridizes to the target DNA so that it's 3' base is situated directly over the SNP nucleotide and a second probe that hybridizes the template upstream (downstream according to the complementary strand) of the SNP polymorphic site providing a 5' end for the ligation reaction. If the allele-specific probe matches the target DNA, it will fully hybridize to the target DNA and ligation can occur. Ligation does not generally occur in the presence of a mismatched 3' base. Ligated or unligated products can be detected by gel electrophoresis, MALDI-TOF mass spectrometry or by capillary electrophoresis for large-scale applications (Rapley and Harbron, 2004).

Constant denaturant capillary electrophoresis

Constant Denaturant Capillary Electrophoresis (SDCE) uses the principle that electrophoretic separation of samples inside narrow-bore capillaries with electrophoretic media is capable of separating small molecules with high resolution (Jorgenson and Lukacs, 1981) along with high fidelity PCR amplification. The denaturing conditions in CDCE are achieved by heating a section of capillary in a temperature-controlled water jacket. Using CDCE, it is possible to measure point mutations at frequencies as low as 10^{-6}. The CDCE instruments are constantly under upgradation and the present version is with two wavelength detectors, which have the ability to use samples labeled with two different fluorescent dyes, thus permitting comparison of two separate channels. Separation of PCR products is generally conducted in capillaries with an internal diameter of 75 μm at a constant current of 9 μA. Further improvement with multiple capillary arrays and automation systems should increase the speed and the scale of this technique.

These are some of the important techniques being used for SNP genotypes, though it is not a complete list of all the methods being used for SNP genotyping. Since, different methods have different advantages and limitations; selection of the most appropriate genotyping platform requires numerous considerations, including scale, cost, compatibility with other instruments in the laboratory, requirements for liquid-handling systems, technical support from manufacturers of the instruments and chemicals, availability of the technical expertise in the laboratory, throughput, and the nature and objective of the research.

Application of SNPs in sericulture

SNPs for silkworm, *B. mori* genome analysis.

The mulberry silkworm, *B. mori*, is one of the most important insect species that has been used extensively for silk production for more than 5000 years. Over the years of research, a large number of silkworm races have been

evolved through systematic breeding. These silkworm races, inbred lines, and their hybrids show substantial variations for a large number of quantitative traits. Most of these silkworm races are still differentiated based on their phenotypic traits such as color and other marking of mature silkworms, larval duration, color and shape of the cocoon, and the color and other properties of the silk fiber. Since several of the silkworm race that are currently used in breeding and for commercial seed production were developed from crosses involving many races or varieties, it is often difficult to distinguish them unambiguously with these phenotypic characters. Here arises the need for reliable yardsticks to identify them easily and reliably. DNA markers are stable, reliable and efficient, though not as easily obtainable as the phenotypic markers. A variety of DNA markers such as RAPD, ISSR RFLP, SSR, EST and SNP markers are available now and each of these marker has its on strength and weakness. Since a number of articles have already published on the merits and demerits of generally used molecular markers (Vijayan *et al.,* 2007), in the present paper those points are not discussed. Nonetheless, it should be noted here that as the SNPs are the most abundant molecular marker available, they can be used for race identification in IPR registrations, parental identifications of hybrids in commercial seed production, determination of duplicates in germplasm resources etc.

Genetic mapping of *B. mori*

For the silkworm, *B. mori*, molecular linkage maps composed of various markers including about 1000 RAPDs (Promboon *et al.,* 1995; Yasukochi, 1998), 250 RFLPs (Shi *et al.,* 1995; Kadono-Okuda *et al.,* 2002; Nguu *et al.,* 2005), 545 AFLPs (Tan *et al.,* 2005), 500 SSRs (Miao *et al.,* 2006; Prasad *et al.,* 2005) and nearly 400 ESTs (Yasukochi *et al.,* 2006) have been generated. Saturation of these maps with more number of reproducible molecular markers can be achieved with SNPs as demonstrated by Cheng et al. (2004), who identified 101 putative SNPs with an average density of 1.395 SNPs per 1000 bp by aligning ESTs developed from an inbred silkworm strain "Dazao". Out of these 62.4% were of transverse types while the remaining 37.5% were transition types. The nucleotide diversity estimated was 1.3×10^{-3}. Based on this rate of nucleotide diversity and the genome size of 530-Mb (Gage, 1974), it was estimated that *B. mori* genome may have about 673000 SNPs (Cheng *et al.,* 2004). However, due to the continued inbreeding the possibility of a reduction in SNP density in silkworm genome can not be ruled out. Subsequent efforts on identification and utilization of SNPs in silkworm have resulted in the identification of more than 1755 SNPs using bacterial artificial chromosomes (BAC) end sequences (Yamamoto *et al.,* 2008). These SNPs were found segregating into 28 linkage groups, with a total recombination length of 1,413 cM. They successfully assigned 26 of the SNP linkage groups into the classical silkworm chromosomes 1-26. Although, the average distance between markers was 0.81 cM, the markers were not evenly distributed through out the linkage map. Some of these SNPs are available now at http://silkworm.genomics.org.cn. Additionally, it is well known that cSNPs (SNPs present in cDNA) are of major interest in the study of phenotypic variation and functional gene mapping as cSNPs provide two main advantages for genetic analysis. One is that, it is used to construct genomic markers associated with genes. Another is the identification of sequence variations associated with EST sequences from individuals in a population (Picoult-Newberg *et al.,* 1999). Thus, it is sure that in the coming years, SNPs are going to play more prominent roles in genotyping, genetic mapping, QTLs identification, and in marker assisted selection breeding.

SNPs for silkworm pathological studies

There are a number of important potential advantages to using SNPs to investigate the genetic determinants of complex diseases. Research on population genetics demonstrated that evolutionary potential of a pathogen is reflected in its genetic diversity and its genetic structure. Pathogen populations with higher evolutionary potential are more likely to overcome host resistance than pathogen populations with a lower evolutionary potential (Mac Donald and Linde, 2002). The resulting changes in population structure or virulence can lead to host resistance breakdown. Therefore understanding the pathogen genetic diversity is critical in developing an effective disease control strategy. The mulberry silkworm, *B. mori*, is infested with many pathogens including virus, protozoa, bacteria, and fungi. For instances, pebrine is one of the deadly diseases that cause wide spread destruction in most of the silkworm rearing countries. This disease is caused by a protozoan (*Nosema bombycis*). Based on the pathogenicity, mode of transmission, tissue specificity of infection and differences in the gene sequences of the small subunit at least four microsporidian isolates have been identified in India (Rao *et al.,* 2004). This clearly indicates the high evolutionary potential of this microsporidian parasite. Although attempts have been made to characterize these isolated through molecular markers such as ISSR markers (Rao *et al.,* 2005), proper genotyping of these isolates require large number of markers. Thus, SNPs can be of much use to genotype these isolates, due to their abundance in the genome, and relating them with the nature and severity of the disease. The same principle is also

applicable to other pathogens like fungi, bacteria and virus. Similarly, results from other organisms including human revealed that single nucleotide polymorphism (SNP) mapping has accelerated complex disease gene localization by providing a tool to narrow the linkage region by the detection of multiple SNPs associated with the disease in a relatively small linkage disequilibrium (LD) region. Identification of susceptibility genes will enable a better understanding of the mechanisms of the disease processes and will facilitate the discovery of new and more efficacious medicines. Whole genome SNP maps will also allow abbreviated SNP profiles to be developed for pharmacogenetic applications, enabling physicians to tailor therapeutic regimens (i.e., identify patients likely to receive therapeutic benefit and not suffer adverse reactions). Further, because of their abundance SNPs have high potential for detailed haplotype analysis and application in association studies. In addition, SNPs are more suitable for association studies because they are almost exclusively identical by descent because of their relatively low mutation rate. Following identification of a SNP or SNP-based haplotype that is closely associated with a disease or associated trait, it might be possible to use this information to develop diagnostic tests. The ability to determine risk of disease prior to the onset of symptoms would be potentially of great benefit in many conditions. Therefore, SNPs can be of much use in genetic diversity analysis of the pathogens and silkworms, and their associations.

SNPs in the genomics of wild silkworms

In addition to the domesticated mulberry silkworm, *B. mori*, a few more silkworm species are used for commercial utilization. Important among them are the tropical tasar silkworm, *A. mylitta*, the eri silkworm S*amia cynthia ricini*, and the muga silkworm, *Antheraea assamica*. These silkworms are endemic to India and the morphologically divergent populations distributed in different geographical regions are termed as "ecological races" or "ecoraces". Some of the important ecoraces of *A. mylitta*, are 'Daba', 'Raily', 'Sukinda', 'Sarihan', 'Laria', 'Andhra local', 'Bhandara local' and 'Modal'. Like wise the ecoraces (~16 numbers) of *S. cynthia ricini* are widely distributed in the Brahmaputra river valley in North-Eastern India. Populations of the muga silk worm, *A. assamica,* are mostly confined to the forests of Assam. Owing to the continued deforestation activities, population fragmentation and inbreeding most of these ecoraces are under the threat of genetic diversity erosion and decline or complete elimination of the populations. This drastic loss of genetic diversity coupled with loss of populations has raised concern not only among the silkworm farmers and sericulturists but also among the ecologists and conservationists. Thus, appropriate strategies are to be chalked out to conserve these precious wild silkworms. Generally, these 'ecoraces' are distinguished by a few phenotypic characters like fecundity, voltinism, cocoon weight, silk ratio and also in their host plant preference. However, using the phenotypic traits, it is often found difficult to identify them properly as the phenotypic traits changes with environmental conditions and developmental stages. Another problem pertaining to these wild silkworm populations is the lack of adequate information on the genetic structuring among different populations of the same ecorace and among sympatric and allopatric ecoraces. Recently several new methods have been developed for estimating the comparative evolutionary changes that take place in populations (Myers *et al*., 2005). Similarly, methods for quantifying recombination rates and identifying recombination hotspots have also been formulated (McVean, 2002). One of the fundamental parameter being used in population genetic analyses is $\theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per generation. There are several estimators of θ, including the commonly used estimator by Watterson (1975), based on the number of segregating sites. One reason for the interest in this parameter is that it is informative regarding both demographic processes and natural selection (Hudson *et al*., 1987). For example, a reduction in θ in a region with normal or elevated between-species divergence suggests the action of recent natural selection acting in the region. Therefore, estimates of θ can be used to identify candidate regions of recent selection. In addition, the relationship between recombination rates and θ is highly informative regarding the relative importance of genetic drift and natural selection in shaping diversity in the genome. In *Drosophila*, it is well established that θ varies with the local recombination rate (Begun and Aquadro, 1992). This has been interpreted as evidence for the action of selection in the genome. Both positive and negative selection can lead to a reduction in population genetic variability, and in both cases the effect is stronger in regions of low recombination. In order to understand these genetic changes a large number of molecular markers need to be developed cost effectively. Although, a few molecular markers such as RAPD, ISSR and RFLP have been used to investigate the inter-populational (Chatterjee *et al.,* 2004b), inter ecoracial (Kar *et al.,* 2005; Mahendran *et al.,* 2006) diversity, the most potential SNPs have not been tested so far. An important characteristic of SNPs is that they are thought to have very low mutation rates, ~10^{-8}–10^{-9} in humans. The population genetical parameter $N_e\mu$ (μ = mutation rate per generation, N_e = effective population size) was estimated to be 10^{-4} by Wang *et al.* (1998).

This implies that the probability of two mutations occurring in the same locus is very low and consequently, the data are essentially diallelic. Another important property of SNPs is that they are abundant and distributed widely across the genome and easy to reproduce in different laboratories. Thus, SNPs are best suited for the comparative evolutionary analyses of these ecoraces. The vast EST data base (http://www.cdfd.org.in/wildsilkbase/home.php) along with the 118 cSNPs developed by Dr. J. Nagaraju and his school of scientists at CDFD, Hyderabad, India (Arunkumar *et al.,* 2008a-b) are valuable resources for genomic analysis of the wild silkworms.

Another potential use of SNPs in wild silkworms is in association mapping or linkage disequilibrium (LD) mapping, though yet to be initiated. LD is defined as the non-random association of alleles at different loci (Flint-Garcia *et al*., 2003). If there is linkage disequilibrium between a marker and a trait, then that marker can be associated with phenotypic values using appropriate statistical analysis (Cardon and Bell, 2001). Because LD mapping exploits genomic diversity present in natural populations, the requirement of developing specific mapping population can be avoided. This makes LD mapping the most suitable mapping method for the wild silkworms. Association mapping make use of all the recombination events that have taken place since the founder population. Based on the scale and focus of the study, association mapping can be of candidate-gene association mapping, which relates polymorphisms in selected candidate genes that have purported roles in controlling phenotypic variation for specific traits and genome-wide association mapping, or genome scan, which surveys genetic variation in the whole genome to find signals of association for various complex traits (Risch and Merikangas, 1996). LD has already been extensively used in mapping complex disease genes in humans (Jorde 2000), and for agronomic traits in plants (Remington *et al*., 2001; Aranzana, 2005).

Use of SNP in host plants

Unfortunately, no SNPs have been discovered in mulberry or any other host plants of the silkworms. This is mainly due to the paucity of adequate information on the genome of these host plants. However, the growing number of ESTs in mulberry (Lal *et al.,* 2008) may help to identify a number of SNPs if the same from two or more populations/cultivars/genotype/species are available. Likewise, SNPs can also be developed through locus specific amplification (LSA) and comparative re-sequencing from multiple individuals (Rieder *et al.,* 1998) by utilizing the information available from the genomic sequences deposited from markers like ISSR and RAPD that are associated with valuable phenotypic traits. In LSA procedure, PCR products amplified from genomic DNA from different individuals are directly sequenced and compared. This strategy has the advantage that the required coverage of several reads per locus can be reached immediately and the generated sequence traces can be compared directly. The speed-limiting factor in the LSA strategy is the design of primers for amplification. Many SNP genotyping assays, however, require a step of amplification of the region flanking the SNP in which the amplification primers can be employed again. The SNPs, thus, discovered can be validated with a few species and those identified as useful SNPs can be widely used for genotyping, identification of cultivars, construction of highly saturated genetic maps, development of association maps (LD maps) for both agronomic traits and disease resistance. Similarly, it is reported that mulberry and other host plants are susceptible to a number of pests and diseases (Teotia and Sen, 1994). Genotypic differences have also been observed for the disease susceptibility. It is also noted that many plant pathogens differ from their closest relatives by only a few bases on different genes. Thus, using SNPs it is possible to unravel the genetic diversity among both the host plants and the pathogens. Such studies will be of much use in understanding the mechanism of host plant resistance, epidemiology and for effective disease management. Further, all the advantages of SNP genotyping associated with disease managements mentioned in the pathological portions of silkworm diseases are equally applicable to the host plant disease managements.

Conclusions

Thus, it can be concluded that, although an impressive degree of progress has been made in the area of SNP genotyping in many agriculturally important crop plants and animals, not much progress has been made in sericulture to tap the tremendous potential attached with this third generation marker system. Therefore, it is high time to make concerted efforts to utilize the advantages of SNP genotyping in cultivar identification, QTL mapping, marker assisted selection and in association mapping to tackle complex traits like disease resistance and biochemical contents of the leaf.

References

Abravaya K, Huff J, Marshall R, Merchant B, Mullen CS, Chneider G, Robinson J (2003) Molecular beacons as diagnostic tools: technology and applications. Clin Chem Lab Med 41, 468-474.

Ahmadian A, Gharizadeh B, O'Meara D, Odeberg J, Lundeberg J (2001) Genotyping by apyrase-mediated allele-specific extension. Nucl Acids Res 29, 1-21.

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genetics 1, 531-539.

Arunkumar KP, Tomar A, Daimon T, Shimada T, Nagaraju J (2008a) WildSilkbase: An EST database of wild silkmoths. BMC Genomics 9, 338.

Arunkumar KP, Kifayathullah L, Nagaraju J (2008b) Microsatellite markers for the Indian golden silkmoth, *Antheraea assama* (Saturniidae: Lepidoptera). Mol Ecol Resour (doi: 10.1111/j.1755-0998.2008.02414.x).

Asakura N (2001) Single nucleotide polymorphism in an STS region linked to the *Ncc-Tmp1A* locusare informative for characterizing the differentiation of Chromosome 1A in wheat. Genes Genet Syst 76, 295-304.

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356, 519-520.

Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nature Genet 33, 518-521.

Chatterjee SN, Mohandas TP (2003) Identification of ISSR markers associated with productivity traits in silkworm, *Bombyx mori* L. Genome 46, 438-447.

Chatterjee SN, Nagaraja GM, Srivastava PP, Naik G (2004a) Morphological and molecular variation of *Morus laevigata* in India. Genetica 39, 1612-1624.

Chatterjee SN, Vijayan K, Roy GC, Nair CV (2004b) ISSR profiling of genetic variability in the ecotypes of *Antheraea mylitta* Drury, the tropical tasar silkworm. Russ J Genetics 40, 21-217.

Cheng TC, Xia QY, Qian JF, Liu Y, Zha XF, Xiang ZH (2004) Mining single nucleotide polymorphism from EST data of silkworm, *Bombyx mori*, inbred strain Dazo. Insect Biochem Mol Biol 34, 523-530.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3, 19.

Collins FS, Guyer MS, Chravarti A (1997) Variations on a theme: Cataloging human DNA sequence variation. Science 128, 1580-1581.

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nat Rev Genet 2, 91-99.

Costabile M, Quach A, Ferrante A (2006) Molecular approaches in the diagnosis of primary immunodeficiency diseases. Hum Mutat 27, 1163-73.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8, 186-194.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8, 175-185.

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54, 357-374.

Gage LP (1974). The *Bombyx mori* genome: analysis by DNA reassociation kinentics. Chromosoma 45, 27-42.

Griffin TJ, Smith LM (2000) Single-nucleotide polymorphism analysis by MALDI–TOF mass spectrometry. TIBTECH 18, 77-84.

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphism: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80, 524-535.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22, 239-247.

Howell W, Jobs M, Gyllensten U, Brookes A (1999) Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. Nat Biotechnol. 17, 87-88.

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153-159.

Hudson RR (2001) Two-locus sampling distributions and their application. Genetics 159, 1805-1817.

Jorde JB (2000) Linkage disequilibrium and the search for complex disease gene. Genome Res 10, 1435-1444.

Kar PK, Vijayan K, Nair CV, Mohandas TP, Saratchandra B, Thangavalu K (2005) Genetic variability and genetic structure of wild and semi-domestic populations of tasar silkworm (*Antheraea mylitta*) ecorace Daba as revealed through ISSR markers. Genetica 125, 173-183.

Kadono-Okuda K, Kosegawa E, Mase K, Hara W (2002) Linkage analysis of maternal EST cDNA clones covering all twenty-eight chromosomes in the silkworm, *Bombyx mori* . Insect Mol Biol 11, 443-451.

Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Anal Chem 60, 2299-2303.

Kimiko Y, Junko N, Keiko KO, Junko N, Motoe S, Yoshitaka S, Yutaka B, Hiroshi F, Goldsmith MR, Kazuei M (2006) Construction of a single nucleotide polymorphism linkage map for the silkworm , *Bombyx mori,* based on BAC end-sequences. Genetics 173, 151-161.

Kirpekar F, Nordhoff E, Larsen LK, Kristainsen K, Roepstroff P, Hillemkamp F (1998) DNA sequence analysis by MALDI mass spectrometry. Nucl Acids Res 26, 2554-2559.

Kota R, Rudd S, Facius A, Kolesov G, Theil T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphism from large EST collections in barley (*Hordeum vulagre* L.). Mol Gen Genomics 270, 24-33.

Kruglyak L (1997) The use of genetic map of biallelic markers in linkage studies. Nat Genet 17, 21-24.

Lal S, Ravi V, Khurana JP, Khurana P (2008) Repertoire of leaf expressed sequence tags (ESTs) and partial characterization of stress-related and membrane transporter genes from mulberry (*Morus indica* L.). Tree Genet Genome (In press).

Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. Nat Genet 21, 20-24.

MacDonald B, Linde C (2002) Pathogen population genetics, evolutionary potential and durable resistance. Ann Rev Phytopathol 40, 349-379.

Mahendran B, Padhi BK, Ghosh SK, Kundu SC (2006) Genetic variation in ecoraces of tropical tasar silkworm, *Antherae mylitta* D. using RFLF technique. Curr Sci 90, 100-103.

Mase K, Iizuka T, Yamamoto T, Okada E, Hara WH (2007) Genetic mapping of a food preference gene in the silkworm, *Bombyx mori*, using restriction fragment length polymorphism (RFLPs). Genes Genet Syst 82, 249-256.

McGuigan FE, Ralston SH (2002) Single nucleotide polymorphism detection: allelic discrimination using TaqMan. Psychiatr Genet 12, 133-136.

McVean GA (2002) A genealogical interpretation of linkage disequilibrium. Genetics 162, 987-991.

Miao X *et al.* (2006) Simple sequence repeat-based consensus linkage map of *Bombyx mori*. Proc Natl Acad Sci USA 102, 16303-16308.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. Science 310, 321-324.

Nagaraja GM, Nagaraju J (1995) Genome fingerprinting in silkworm *Bombyx mori* using random arbitrary primers. Electrophoresis 16, 1633-1638.

Nagaraja GM, Mahesh G, Satish V, Madhu M, Muthulakshmi M, Nagaraju J (2005) Genetic mapping of Z chromosome and identification of W chromosome-specific markers in the silkworm, *Bombyx mori*. Heredity 95, 148-157.

Newton CR (1991) Genetic analysis in cystic fibrosis using the amplification refractory mutation system (ARMS): the J3.11 MspI polymorphism. J Med Genet 28, 248-251.

Nguu EK, Kadono-Okuda K, Mase K, Kosegawa E, Hara W (2005) Molecular linkage map for the silkworm, *Bombyx mori*, based on restriction fragment length polymorphism of cDNA clones. J Insect Biotechnol Sericol 74, 5-13.

Olivier M (2005) The Invader assay for SNP genotyping. Mutat Res 573, 103-110.

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. Genome Res 9, 167-174.

Prasad MD *et al.* (2005) Survey and analysis of microsatellites in the silkworm, *Bombyx mori*: frequency, distribution, mutations, marker potential and their conservation in heterologous species. Genetics 169, 197-214.

Promboon A, Shimada T, Fujisawa H, Kobayashi M (1995) Linkage map of random amplified polymorphic DNAs (RAPDs) in the silkworm, *Bombyx mori*. Genet Res Camb 66, 1-7.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98, 11479-11484.

Rao SN, Muthulakshmi M, Kanginakudru S, Nagaraju J (2004) Phylogenetic relationships of three new microsporidian isolates from the silkworm, *Bombyx mori*. J Invertebr Pathol 86, 87-95.

Rao SN, Nath BS, Saratchandra B (2005) Characterization and phylogenetic relationships among microsporidia infecting silkworm, *Bombyx mori*, using inter simple sequence repeat (ISSR) and small subunit rRNA (SSU-rRNA) sequence analysis. Genome 48, 355-366.

Rapley R, Harbron S (2004) Molecular Analysis and Genome Discovery. Chichester. John Wiley & Sons Ltd.

Reddy KD, Nagaraju J, Abraham EG (1999a) Genetic characterization of silkworm *Bombyx mori* by simple sequence repeats (SSR) – anchored PCR. Heredity 83, 681-687.

Reddy KD, Abraham EG, Nagaraju J (1999b) Microsatellites of the silkworm, *Bombyx mori*: abundance, polymorphism and strain characterization. Genome 42, 1057-1065.

Reich DE, Gabriel SB, Altshuler D (2003) Quality and completeness of SNP databases. Nat Genet 33, 457-458.

Rieder MJ, Taylor SL, Tobe VO, Nickerson DA (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucl Acids Res 26, 967-973.

Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. Science 281, 363-365.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928-933.

Schneider K, Weisshar B, Borchardt DC, Salamini F (2001) SNP frequency and allelic haplotype structure of *Beta vulagris* expressed genes. Mol Breed 8, 63-74.

Sethuraman BN, Mohandas TP, Chatterjee SN (2002) DNA fingerprinting with homologous multilocus probes and search for DNA markers associated with yield attributes in silkworm, *Bombyx mori*. Eur J Entomol 99, 267-276.

Shi J, Heckel DG, Goldsmith MR (1995) A genetic linkage map for the domesticated silkworm, *Bombyx mori*, based on restriction fragment length polymorphisms. Genet Res 66, 109-126.

Sharma A, Sharma R, Machii H (2000) Assessment of genetic diversity in a *Morus* germplasm collection using fluorescence-based AFLP markers. Theor Appl Genet 101, 1049-1055.

Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2, 930-942.

Taillon-Miller P, Gu ZJ, Li Q, Hillier L, Kwok PY (1998) Overlapping genomic sequences: a treasure trove of single-

nucleotide polymorphisms. Genome Res 8, 748-754.
Tan YD, Wan C, Zhu Y, Lu C, Xiang Z, Deng HW (2001) An amplified fragment length polymorphism map of the silkworm. Genetics 157, 1277-1284.
Teotia RS, Sen SK (1994) Mulberry diseases in India and their control. Sericologia 34, 1-18.
Tyrka M *et al*. (2004) Developement of single nucleotide polymorphism marker of the wheat Lr1 leaf rust resistance gene. Cell Mol Biol Lett 9, 879-889.
Tyagi S, Bratu DP, Kramer FR (1998) Multicolor molecular beacons for allele discrimination. Nat Biotechnol 16, 49-53.
Van K, Hwang EY, Kim MY, Park HJ, Lee SH, Cregan PB (2005) Discovery of SNPs in soybean phenotypes frequently used as the parents of mapping populations in the United States and Korea. J Heridity 96, 529-535.
Venkateswarlu M, Raje Urs S, Nath BS, Shashidhar HE, Maheswaran M, Veeraiah TM, Sabitha MG (2006) A first genetic linkage map of mulberry (*Morus* spp.) using RAPD, ISSR, and SSR markers and pseudotestcross mapping strategy. Tree Genet Genomes 3, 15-24.
Vijayan K, Kar PK, Tikader A, Srivastava PP, Awasthi AK, Thangavelu K, Saratchandra B (2004) Molecular evaluation of genetic variability in wild populations of mulberry (*Morus serrata* Roxb.). Plant Breed 123, 568-572.
Vijayan K, Chatterjee SN, Nair CV (2005) Molecular characterization of mulberry genetic resources indigenous to India. Genet Resour Crop Evol 52, 77-86.
Vijayan K, Anuradha HJ, Nair CV, Pradeep AR, Awasthi AK, Saratchandra B, Rahman SAS, Singh KC, Chakraborti R, Raje Urs S (2006) Genetic diversity and differentiation among populations of the Indian eri silkworm, *Samia cynthia ricini*, revealed by ISSR markers. J Insect Sci 6, 30.
Vijayan K (2007) Molecular markers and their application in mulberry breeding. Int J Indust Entomol 15, 145-156.
Wang DG (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280, 1077-1082.
Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7, 256-276.
Yamamoto K, Narukawa J, Kadono-Okuda K, Nohata J, Sasanuma YM, Suetsugu Y, Banno Y, Fujii H, Goldsmith MR, Mita K (2008) A BAC-based integrated linkage map of the silkworm, *Bombyx mori*. Genome Biol 9, R21.
Yasukochi Y, Ashakumary LA, Baba K, Yoshido A, Sahara K (2006) A second generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. Genetics 173, 1319-1328.
Zhao W, Pan Y (2004) Genetic diversity of genus *Morus* revealed by RAPD markers in China. Int J Agric Biol 6, 950-954.