

INVITED REVIEW

Molecular marker systems in insects: current trends and future avenues

SUSANTA K. BEHURA

349 Morrill Hall, Department of Entomology, 505 S Goodwin Avenue, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Abstract

Insects comprise the largest species composition in the entire animal kingdom and possess a vast undiscovered genetic diversity and gene pool that can be better explored using molecular marker techniques. Current trends of application of DNA marker techniques in diverse domains of insect ecological studies show that mitochondrial DNA (mtDNA), microsatellites, random amplified polymorphic DNA (RAPD), expressed sequence tags (EST) and amplified fragment length polymorphism (AFLP) markers have contributed significantly for progresses towards understanding genetic basis of insect diversity and for mapping medically and agriculturally important genes and quantitative trait loci in insect pests. Apart from these popular marker systems, other novel approaches including transposon display, sequence-specific amplification polymorphism (S-SAP), repeat-associated polymerase chain reaction (PCR) markers have been identified as alternate marker systems in insect studies. Besides, whole genome microarray and single nucleotide polymorphism (SNP) assays are becoming more popular to screen genome-wide polymorphisms in fast and cost effective manner. However, use of such methodologies has not gained widespread popularity in entomological studies. The current study highlights the recent trends of applications of molecular markers in insect studies and explores the technological advancements in molecular marker tools and modern high throughput genotyping methodologies that may be applied in entomological researches for better understanding of insect ecology at molecular level.

Keywords: insect, microarray, molecular marker, SNP, SFP, transposon display

Received 14 February 2006; revision accepted 2 May 2006

Introduction

Insects represent a major life form on earth. Nearly 900 000 insect species have been discovered by now, comprising 75% of all the recorded animal species. Insects can be found in almost all ecosystems; in deserts as well as in the Antarctic. Even the oceans have not been spared. Some insect species (*Halobates*) are also known to inhabit on ocean surfaces (Ikawa *et al.* 2002). This truly demonstrates the immense ecological adaptation of insects and their extraordinarily adaptable life forms in diverse environments where other animals can't survive. This biological success is attributed to the enormous diversity of their size and body structure, mating strategies, and the remarkable adaptive abilities in

feeding and behaviour. Such diversity makes insects impose a delicate relationship to human life. Insects are beneficial as they pollinate crops, act as natural enemies of damaging pests, and produce useful products such as honey, silk and wax for humans. At the same time, insects are major pests of our food crops, act as vectors for transmitting deadly diseases, cause damage to our urban infrastructure, environment, forest and natural resources, and sometimes interfere in our international trade, commerce and economic affairs. Thus, very often one wonders what makes insects so diverged and how the genes and genetic make-up of insects contribute to their adaptable life forms, and most importantly, how they may affect human life either directly or indirectly. Insect populations even within a species vary in their behaviour and morphology that attributes to their complex interaction with the environment (Dempster & McLean 1999). Thus, the study of insect ecology is important to understand their

Correspondence: Susanta K. Behura, Fax: 217-244-3499; E-mail: susanta@life.uiuc.edu

evolution and diversification, and their influence on the functional and trophic links between different components of associated habitats (Speight *et al.* 2005).

Significant progress has been made in understanding insect diversity and ecology by using classical genetic principles. Common visible markers including eye colour, body spots or bands and hairs or spines were used as phenotypic markers in studying pattern of dispersal, mating behaviour and inheritance of genetic traits in insects (Bartlett *et al.* 1968; Fay & Craig 1969; Bond *et al.* 1970; Bartlett & Butler 1975). Although the phenotypic markers are found at all time of life span of the organism and can be readily used for studies in field conditions, they suffer from many practical limitations. The major drawback is that these visible phenotypes are relatively infrequent and often hard to score. Also, it is difficult and time-consuming to induce genetic mutations in laboratory populations to develop new phenotypic markers, and sometimes they interfere with the overall fitness of the organism. Furthermore, identification of phenotype markers must be accompanied with the information as to how the trait is inherited to the offspring before they are used as faithful markers. Because the phenotype markers are rare, use of these markers in mapping a trait is difficult. For all such difficulties and with the concurrent advancement in biochemical methodologies, protein markers then became more popular. Protein markers made a significant contribution in the early periods when DNA technologies were not so much advanced as it is now (Loxdale & Lushai 1998). The electrophoretic pattern of allozymes as detected in polyacrylamide or starch gels were used to identify different alleles of a given gene. Based on banding patterns, it was possible to determine the extent of heterogeneity of a locus in a population based on the variant allozymes produced by the individuals. Banding pattern of isozymes (products of different loci, but with similar function) could also be used for studying genetic variation within and between populations (Steiner & Joslyn 1979; Bartlett 1981; Loxdale *et al.* 1985). Protein markers have also been successfully exploited in studying insecticide resistance (Maa & Terriere 1983), pathogen identification (Wilding *et al.* 1993), chromosome mapping (Loukas *et al.* 1979) or detection of prey in insect predators (Solomon *et al.* 1996) to name a few. However, with the development of DNA-based marker systems, it was found that greater level of polymorphism could be obtained by using DNA markers than by using protein markers in many situations (Richardson *et al.* 1986). This is because mutations in introns or even in the codons of a gene can potentially provide more variation at the DNA level than at the protein level. Moreover, DNA samples are more stable than proteins and are unchanged for detection at all time and tissue of the organism unlike proteins. Thus, DNA markers became the most common yardsticks for measuring genetic differences between individuals or within and between related species or

populations. The unprecedented advancements in modern molecular biology, particularly in those of DNA marker technology, have already created a wealth of technical know-how that finds useful applications of these markers especially in molecular ecology research in insects (Hoy 2003).

Over the last past 15 years or so, DNA makers have made a significant contribution to rapid rise of molecular studies of genetic relatedness, phylogeny, population dynamics or gene and genome mapping in insects (Loxdale & Lushai 1998; Avise 2000, 2004; Severson *et al.* 2001; Heckel 2003). Many improvements have been made to enhance reproducibility, power of resolution (ability to reveal more informative polymorphisms from less number of loci) and more importantly, the cost and time consumption in developing and scoring the marker loci. Since then, application of DNA markers in entomology has gone through and is still undergoing a noticeable change in continuously accommodating new technologies for robust and less expensive genotyping methods. Entomologists are getting more accustomed with the refinement of marker systems and are applying the new techniques to study insect genomes more efficiently. Traditionally, mitochondrial DNA (mtDNA) has been a choice of marker for studying genetic variations in insect species. Mitochondrial gene sequences have been used for phylogenetic and population-genetic studies to construct evolutionary history of related insect species. More precisely, these markers have provided invaluable insights into the history and genetic basis of speciation and phenotypic evolution of recently diverged species. Moreover, mtDNA sequences are often transferred to the nucleus giving rise to the so-called nuclear mtDNA (Numt). Variations in copy number and size of Numts are also used to assess the inter-specific diversity of these loci in insects (Richly & Leister 2004). Microsatellites are also used as popular markers in insect studies because of high abundance and highly variable nature of their loci in genome. Despite their rising popularity in insect research, microsatellites have not yet become a major marker system in entomology. Recent studies have shown that microsatellite markers can be faithfully extended beyond population genetics and can be used for studying phylogenetic relationships of closely related species with fewer loci than previously assumed (Schlotterer 2001). However, with the introduction of random amplified polymorphic DNA (RAPD) technique (Williams *et al.* 1990), use of (polymerase chain reaction) PCR-based fingerprinting assays gained popularity mainly because of the easy-to-perform and easy-to-score procedures for these marker loci. But, because RAPD markers suffer from poor reproducibility, use of these markers in insect ecological studies was limited (Black 1993). Then, the development of amplified fragment length polymorphism (AFLP) (Vos *et al.* 1995) technology was adopted as a better alternative to generate more numbers of multiloci reproducible markers more reliable than RAPD markers.

Today, molecular marker technology has reached a new height with the power and the precision of modern genomic tools. High throughput genotyping methods are now available that can be used for genome-wide mutation screening in hundreds or even thousands of individuals (as much as 300 000 genotypes) as quickly as in a day! In principle, if the loci sequence is known in one individual, genotyping them in thousands of individuals in that population is not a technical bottleneck any more. Innovative assay methods and implementation of advance instrumentations have been devised to accurately genotype large numbers of markers in parallel (Lyamichev *et al.* 1999; Kwok 2001; Tsuchihashi & Dracopoli 2002). These large-scale genotyping techniques are reliable enough to discriminate allelic variations even at single nucleotide level. Also, these methods are less time-consuming and are less prone to sample-to-sample variations associated with single tube genotyping methods. Moreover, the assay costs are also gradually decreasing by continuous instrumental innovations and often by commercialization. This review is intended to study the current status of our knowledge on application of molecular markers, and the sciences and techniques of these systems, to study genetic diversity in insects that shapes their evolution and ecology. One of the major objectives of this work is to explore the potential utility of alternate novel marker systems and modern high throughput genotyping technologies applicable in these studies for a better understanding of insect ecology.

Conventional marker systems

DNA markers such as mtDNA, RAPD, AFLP, microsatellites and ESTs have been used as popular marker systems in insect genetics research. Although there are inherent advantages and disadvantages associated with each marker systems, the choice of applying them depends upon the objectives of a study. Mitochondrial DNA is used for marker analyses largely because of their maternal inheritance, haploid status, and high rate of evolution. In insects, a further advantage of using mitochondrial markers is that many of these loci can be readily amplified by using universal primers designed from highly conserved mitochondrial genes (Roehrdanz 1993; Kambhampati & Smith 1995; Lunt *et al.* 1996; Zhang & Hewitt 1997; Lanave *et al.* 2002). Moreover, upon amplification, these loci can be used for genotyping by restriction fragment length polymorphisms (RFLP) by easy and simple restriction digestions and gel electrophoresis (Behura *et al.* 2001b). Because of higher evolutionary rate, these marker loci are preferred to use in constructing phylogeny and inferring evolutionary history for ecological studies. However, often in cases, particularly in insects, these markers are difficult to use in evolutionary and phylogenetic studies as mtDNA undergoes selective sweep, paternal leakage or even nuclear integration (Zhang & Hewitt 1996; Behura *et al.* 2001b; Hurst & Jiggins 2005). Another class of markers generated by

arbitrarily primed PCR-based DNA fingerprinting methods, such as RAPD, DNA amplification polymorphisms (DAF) and arbitrarily primed-PCR (AP-PCR), are easy to perform and relatively easy to score (Black 1993). However, these markers, because of poor reliability and reproducibility, are not suitable for population studies (Black 1993). For example, in gypsy moth (*Lymantria dispar* Linnaeus), it was observed that a fragment was present in the RAPD amplification products of F1 progeny although it was not amplifiable from either parent (Reineke *et al.* 1999). This is particularly a problem when dealing with pooled DNA as template in RAPD-PCR, as in the case of bulk segregant analysis (BSA) or while comparing individual fingerprints of a population where genetic admixture is suspected. However, RAPD loci can be used for diagnostic purposes after converting these loci into reliable STS (sequence tagged site; Hunt & Page 1994; Laayouni *et al.* 2005), SCAR (sequence characterized amplified region; Behura *et al.* 1999, 2000) or STAR (sequence-tagged RAPD; Fulton *et al.* 2001) by sequencing the RAPD fragments and designing sequence specific PCR primers. On the other hand, AFLP technique (Vos *et al.* 1995) is a better marker system as compared to RAPD and RFLP methods. It combines the easiness of RAPD and the reliability of RFLP. The genomic DNA is first digested with two restriction enzymes, one 4-cutter and another 6-cutter. The digested DNA is then ligated with adapters compatible to the two restriction sites. The ligated material is then used as template for selective amplification with primers extended at the 3' end with selective nucleotides. The technique is relatively time-consuming and often utilizes radiolabelling of primers. Also, scoring of the loci requires electrophoresis of the amplified products in sequencing gels or in automated sequencing machine. Sometimes, AFLP loci may contain repeats such as microsatellite in their sequences and may pose difficulty in scoring the alleles (Wong *et al.* 2000). However, the major advantage of AFLP system is that the selective PCR generates more numbers of marker loci, on average 50–100 bands, per primer pairs per sample. Also, AFLP loci are highly reproducible and are generally codominant in nature, although sometimes dominant AFLP loci are also amplified (Wong *et al.* 2000). AFLP markers are suitable for mapping of genes and quantitative trait loci and to generate linkage maps of genomes (Heckel *et al.* 1999; Behura *et al.* 2004; Ruppell *et al.* 2004). Another feature of the AFLP technique is that it can be applied to cDNA and the results can be used to detect differentially expressed genes in insects (Reineke *et al.* 2003).

The use of microsatellite markers in ecological studies is preferred as these are highly polymorphic in nature and are also abundant to generate large numbers of markers compared to the AFLP system. Large numbers of microsatellites can be isolated by generating a genomic library of small fragments either amplified by PCR or generated by partial digestion of genomic DNA (Lunt *et al.* 1999) and

enriched with simple sequences repeats (Ostrander *et al.* 1992; Kandpal *et al.* 1994). Also, microsatellite markers are amenable for high throughput genotyping by nonradioactive labelling and scoring by automated sequencing machines (Nagaraju *et al.* 2002). Although the cost involved in the initial setup is higher than that of the AFLP, this is often compromised in population and ecological studies where more number of genotypes are required for meaningful statistical analyses. However, recent findings have shown that in some cases microsatellites are not neutral markers as they are found involved in gene regulation, genetic hitchhiking, and sex-specific differential selection (Schlotterer 2000; Li *et al.* 2002), and hence, those loci may not be useful for evolutionary and phylogenetic studies. Another class of popular markers, expressed sequence tags (ESTs) has recently been used as major resources for insect genetic studies. Use of commercial kits for preparation of cDNA libraries and automated sequencing methods has made it possible to generate large sets of ESTs in a relatively fast and efficient manner. Although, the major applications of ESTs in insects have been for transcriptome analyses (Whitfield *et al.* 2002; Pedra *et al.* 2003; Nakabachi *et al.* 2005), they have also been used for integrated linkage mapping of insect genomes (Fulton *et al.* 2001; Severson *et al.* 2002; Graham *et al.* 2004). Development of these molecular markers in insects represents a valuable resource in further molecular ecological studies in those species. A list of different molecular markers generated from selected insect species is compiled in Table S1, Supplementary material.

Major applications of molecular markers in studying insect ecology

Ecological research on insect species provide invaluable information on population structure, speciation, gene flow and genetic diversity, and provide explanation on insect diversity based on their interaction with environmental factors, either biotic (including other biological species) or abiotic. In many a time, molecular marker data help to distinguish between different species, when there is no other comprehensive way available to do so. The traditional approach of discriminating two species could be testing if members of the two populations cannot produce fertile offspring. However, some ecologists reject this notion. There are cases where individuals of one population can produce fertile offspring with a second population, and individuals of the second population can produce fertile offspring with members of a third population, but members of the first and third population cannot produce fertile offspring. Though, the hierarchical process from individuals to species evolution has conceived no less than 20 species concepts (Mayden 1997), only half of those recognize biological processes such as reproduction and competition to contribute to the process of evolution within species

(Hey 2001). These concepts are primarily based upon variations in biological, ecological, evolutionary, genetic and phylogenetic observations as opposed to other indirect parameters such as cohesion, recognition, phenetism, polythetism and succession. However, the 'species problem' is still an open question as to what should be called a species and what should not. In *'The Blind Watchmaker'*, Dawkins (1986) argues that two organisms are conspecific if and only if they have the same number of chromosomes, and for each chromosome, both organisms have the same number of nucleotides. At this juncture, DNA sequence is considered as basic biomarker for drawing ecological lines between different hierarchies of speciation.

In insects, DNA markers are used to provide raw information based on which an ecologist makes estimates of genetic diversity and gene flow between species, identifies haplotypes and lineages or predicts migration and colonization history (Hale & Singh 1991; Yoon & Aquadro 1994; Behura *et al.* 2001b; Salvato *et al.* 2002; Llewellyn *et al.* 2003; Bosio *et al.* 2005). Also, molecular data provide the means to differentiate sympatric species from allopatric species and parapatric species (Ballinger-Crabtree *et al.* 1992; Wilkerson *et al.* 1993; Favia *et al.* 1994; Apostol *et al.* 1996; Banuls *et al.* 1999; Ayres *et al.* 2003; Margonari *et al.* 2004). Gene flow and genetic variations within and between insect species, measured from marker data, are critical to establish meaningful explanation for population structure and dynamics (Cervera *et al.* 2000; Wong *et al.* 2000; Takami *et al.* 2004; Mendelson & Shaw 2005). Similarly, natural selection is a major factor that is regarded as the driving engine for population diversity. In this regard, molecular markers are used to infer phylogeny and biogeography of insect populations and to understand modes of evolution and evolutionary trajectories (Luque *et al.* 2002; Chatterjee & Mohandas 2003; Chatterjee & Tanushree 2004; Prasad *et al.* 2005). Also, diagnostic molecular markers, based on linkage to certain traits or genes, are used for diagnostic purposes of individual insects (Hunt & Page 1994; Behura *et al.* 1999; Behura *et al.* 1999, 2000; Kengne *et al.* 2001; Manguin *et al.* 2002; Kethidi *et al.* 2003; Ullmann *et al.* 2003). Often in cases where two parapatric populations coexist but they do not cross-breed, it is assumed that there is behavioural mechanism that is preventing gene flow between them. Without reproductive isolation, population differences cannot develop, and given reproductive isolation, gene flow between the populations cannot merge the differences. Thus, mating behaviour of insects, their behavioural plasticity and their interaction with hosts and parasites, out of other unknown factors, play important roles in contributing to the ecological diversity of insect species. Applications of molecular markers, with representative examples, in major domains of insect ecological research are discussed below to show the effectiveness of DNA marker systems to unravel the underlying genetic factors that shape ecological diversity in insect species.

Mating, parentage and kinship

DNA markers can unravel information to determine parentage and kinship relations in insects. One of the innovative works done in this regard was to use RAPD markers to determine paternity in two odonate species of Anisopteran dragonflies, *Anax parthenope* (Julius Brauer) and *Orthetrum coerulescens* (Keeled Skimmer) (Hadrys *et al.* 1993). RAPD banding patterns were used to access paternity of 'synthetic offsprings' generated by quantitative mixing of genomic DNA from putative parents. This approach has been helpful to establish the paternity of guarding males in species where no information on mating histories of both males and females are known. Using molecular markers, it has been shown how the females in some nonparthenogenetic insect species such as white pine weevil, an important forest pest, carry sperms of more than one male from one season to the next (Lewis *et al.* 2002). The results of such studies provide explanation on how the offspring are produced in new habitats where no males are available for mating and as the basis of their colonization in new geographical regions. Also, DNA markers are quite helpful to study sperm competition in doubly mated female individuals. Simmons & Achmann (2000) used microsatellite markers in the bush cricket *Requena verticalis* Walker (Tettigoniidae: Listroselidinae), and by manipulating the numbers of sperm transferred by the first and by the second male, they were able to find out that the ejaculate quantity of the first male was the determining factor to establish the paternity identity of the offspring. These studies provide information on the role of male competition for selective fitness of offspring. Identification of egg paternity by molecular markers is also used to validate the theoretical models that predict influence of paternal care on reproductive success in some insect species such as golden egg bugs, *Phyllomorpha laciniata* Villers (Heteroptera, Coreidae) (Garcia-Gonzalez *et al.* 2003). In these insects, males somehow recognize the eggs that will produce their true genetic offspring and then take care of those eggs. But as more than one male mate a female in the field condition, a male individual can neither discriminate their own eggs, nor can they predict if their own eggs will be fertilized. Marker-assisted selection of paternity, in these situations, has been extremely helpful to determine the mating history of female individuals (Garcia-Gonzalez *et al.* 2003). The information gained from these studies is crucial to understand remating propensity of females and to estimate the success of sterile insect technique (SIT) often used to control pest populations (Bonizzoni *et al.* 2002).

Some insects such as cricket, *Gryllus bimaculatus* de Geer (Ensifera, Gryllidae), use polyandry as a mating strategy to minimize genetic inbreeding of their brood. In these species, polyandrous females avoid fertilizing eggs with sperm from genetically incompatible males. Using microsatellite markers, this theory was tested by examining history of

paternity where a female was mated with related and unrelated males (Bretman *et al.* 2004). From these molecular data, it was found that unrelated males were more successful in gaining paternity than related males. Similarly, molecular markers are used to understand parthenogenetic mating behaviour in aphid species where cyclical parthenogenesis is a mating strategy by which the insects can populate themselves quickly in a new environment without undergoing the sexual mating (Delmotte *et al.* 2002). Microsatellite markers were used to measure the genetic diversity between sexual and asexual aphid populations. The data indicated that sexual populations have high allelic polymorphism but with little geographical differences, possibly because of population subdivision, inbreeding, selection and large dispersal ability (Delmotte *et al.* 2002). In contrast, asexual populations showed less allelic polymorphism but high heterozygosity at most loci. These results implicate that there are major genetic differences between sexual and asexual aphids. To further understand this, sexual and asexual aphid populations coexisting in the same habitat (such as *Rhopalosiphum padi* L.) were further analysed. RAPD markers associated with life cycle variation and breeding traits in these insects have been converted into codominant sequence-characterized amplified region (SCAR) markers (Simon *et al.* 1999). Using these SCAR markers in segregating and natural populations of known breeding systems, complete linkage was found in segregating populations. Apparently the association in field populations was on an average of 94%. Such information has potential use for studying the role of genetic mechanism of sexual and asexual mating behaviour in dispersal and colonization of aphid populations in geographical regions.

Insect plant interaction

One of the most appealing applications of molecular markers in insect studies is probably on those for insect-plant interaction. DNA markers provide utility in tagging and mapping genes in important crop plants that provide resistance to damaging insect pests, and are also useful in characterizing avirulence genes in insects interacting with the host plants (Harris *et al.* 2003). The molecular genetic information generated by marker data is used to characterize phenotypic ability of insect to attack specific plant varieties. Specific examples of marker usage in these aspects can be evident from molecular studies on gall midges (*Cecidomyiidae* family of Diptera), major insect pests of rice and wheat. Rice gall midge (*Orseolia oryzae* Wood-Mason) is a major problem in rice cropping in many South and Southeast Asian countries because of emergence of different biotypes in their population that breaks down the resistance of host plants. This causes unexpected damage of rice production worth of nearly half a billion dollars every year. Similarly, a closely related species of gall midge, known as Hessian fly (*Mayetiola*

destructor Say) causes heavy economic loss in wheat crops. In spite of employing more than 32 resistance genes in wheat fields to protect Hessian fly damages in North America, different virulent strains have evolved in response to the resistance deployment. This has led to cause occasional outbreaks and irregular damages to wheat crops accounting to an average of \$100 million each year. In such situations, where no effective chemical or cultural control methods are available, one of the strategies has been to exploit molecular and genetic tools to understand genetic interaction between insect and host plant. This will help identify specific genes involved in the interaction that can further be exploited to develop transgenic-based management strategies. One of the important aspects of these researches has been to characterize the emerging virulent strains of gall midge populations and to understand how they interact with host plants. Using RAPD-PCR with pooled DNA from different strains (or biotypes) of Asian rice gall midge, distinct loci specific to individual strains were identified (Behura *et al.* 1999). By confirming by Southern blotting and sequencing, codominant SCAR markers were developed, and allele-specific amplification using these diagnostic markers was useful to establish good correlation between genotypes to the observed phenotypes of these biotypes (ability/inability to attack host plants). An AFLP marker was further discovered using bulk segregant methods, and the locus showed linkage with *vGm2* avirulence gene that interacted with the corresponding resistance gene (*Gm2*) in rice (Behura *et al.* 2000). Similarly, in Hessian fly, RAPD and AFLP markers have been employed in combination with bulk segregant analysis to identify major avirulence genes that condition the resistance mechanism in wheat varieties cultivated in the United States (Stuart *et al.* 1998; Schulte *et al.* 1999; Rider *et al.* 2002; Behura *et al.* 2004). Although avirulence genes of some plant pathogens have been cloned, no insect avirulence gene has been cloned yet. Thus, these markers have potential to exploit map-based cloning strategy to isolate avirulence genes from these economical important gall midges.

Molecular research on aphids represents another clear example on marker-based study on insect–plant interaction. More than 250 aphid species are pests on horticultural crops throughout the world. Most aphid species are relatively host-plant specific, although they may occasionally feed on alternate hosts. If no resistance is triggered from the host, aphids can stunt plant growth, and cause galls to form on leaves, stems and roots. Some aphid species undergo cyclical parthenogenesis (the alternation of sexual reproduction and a phase of parthenogenetic reproduction). They use one host plant as the ‘primary host’ for sexual reproduction and another as a ‘secondary host’ for parthenogenetic reproduction. In *Aphidinae*, migrants returning to the primary host are winged males and winged females (parthenogenic). Later on, winged parthenogenic females return to the secondary

host (herbaceous plant). In these situations, DNA markers provide useful information to understand the genetic basis of the adult polypheny in such aphid clones. Using RAPD-PCR, it was found that major genetic differences existed between winged phenotypes and the wingless phenotypes of the asexual adult aphids (Lushai *et al.* 1997). Similarly, in natural populations of grain aphids, *Sitobion avenae* (Fabricius), feeding on different hosts of grasses and cereals, it was clearly demonstrated that the RAPD banding pattern could be correlated to host adaptation (Lushai *et al.* 2002). These profiles could identify ‘specialist’ genotypes found on specific grasses from the ‘generalist’ genotypes colonizing on multiple host types including cultivated cereals or native grasses. Genetic basis of the host plant association has also been studied in lettuce root aphid (*Pemphigus bursarius* Linnaeus) using microsatellite markers (Miller *et al.* 2005). The degree of virulence (aggressiveness) of individual clones in pea aphids, in response to natural resistance in alfalfa, has been evaluated using RAPD markers (Bournoville *et al.* 2000). These works clearly demonstrate that application of molecular markers are helpful in better understanding the mechanistic and evolutionary basis for the genetic interaction between insect pests and their host plants.

Insect pathogen interaction

Molecular markers are useful in understanding genetic interaction between disease-causing pathogens and the insect vectors that spread them (Crampton *et al.* 1997). *Triatoma infestans* (Klug) (*Reduviidae*) is a major insect vector of Chagas’ disease in many South American countries. It transmits the *Trypanosoma cruzi*, the agent of Chagas’ disease. Mixed and pure clones of *Triatoma cruzi* in the gut of *Triatoma infestans* have been studied by using RAPD profiles to provide information on the vectorial ability of the insects (Pinto *et al.* 1998). Similarly, molecular markers were applied to determine the vectorial ability of mosquitoes (Severson *et al.* 2001) by means of mapping quantitative trait loci (QTL) that determined if a species could transmit the malaria parasite (Severson *et al.* 1995; Bosio *et al.* 2000). Expressed sequence tag (EST) markers have also been used to identify the potential target genes that are involved in immunity to specific pathogens causing malaria (Bartholomay *et al.* 2004). *Metarhizium anisopliae* is a causative agent of green muscardine disease, and is transmitted by a broad range of insects. To explore the molecular basis of this process, EST markers were used to study the pattern of gene expression in responses to diverse insect cuticles and were found that an array of secreted protein had potential activity in the hosts in these responsive actions (Freimoser *et al.* 2003).

Apart from medically important insects, DNA markers have also been used in insect-borne viruses that are important for agriculture. These viruses sometimes act as damaging pathogens for agricultural crops and forest

resources. The brown citrus aphid, *Toxoptera citricida* (Kirkaldy), is considered the primary vector of citrus tristeza virus (CTV), a damaging pathogen that causes heavy losses to citrus industries worldwide. The alate form of these host aphids is an important stage of their life cycle where they spread themselves and spread the viruses in citrus-growing area. ESTs as molecular markers have been developed to study the gene expression pattern at this life cycle of the host aphid to better understand the genetics of its vectorial ability to transmit the citrus tristeza virus (Hunter *et al.* 2003). Similarly, entomopathogenic fungi also play crucial role in the host pathogen interaction in many agriculturally important insects that is analysed by the use of various molecular markers (Fegan *et al.* 1993; Tigano-Milani *et al.* 1995). Some entomopathogens are beneficial for humans as they act as natural biocontrol of damaging insect pests. Understanding genetic interactions between insects and these entomopathogens is a major research focus towards developing natural biocontrols for pests. Identification of these pathogens in diverse insect species is an important screening step that often requires application of molecular markers (Hodge *et al.* 1995; Castrillo *et al.* 2004). Also, DNA markers are used to characterize new and better cell lines for production of viable pathogens as biopesticides (Sudeep *et al.* 2005).

Insecticide resistance

Insecticide resistance is another important focus in entomological research and bears medical and agricultural importance. Molecular markers are used for identification and mapping of resistance genes in insects against insecticides. In malaria control programs, difficulties arise because of emerging resistance in the mosquito vectors to DDT. DDT resistance in the major malaria vector in Africa, *Anopheles gambiae*, is associated with an increased metabolism of the insecticide. Use of microsatellite markers in mapping experiments has identified QTL in *Anopheles gambiae* that determines the DDT resistance phenotypes (Ranson *et al.* 2000). Using RFLP markers, it was discovered that DDT resistance in houseflies, *Musca domestica* was associated with the 'knockdown' *kdr* trait (Knipple *et al.* 1994). 'Knockdown' *kdr* trait is associated with the reduced neuronal sensitivity to pyrethroids insecticides in turn related to a mutation of the sodium channel gating system. Gene-specific PCR markers have also been used to examine expression pattern of cytochrome P450 genes near a DDT resistance gene in *Drosophila* (Brandt *et al.* 2002). Similarly, using random amplified DNA markers, genetic loci have been mapped in lesser grain borer, *Rhyzopertha dominica* (Fabricius) that determines high-level resistance to phosphine (Schlipalius *et al.* 2002). Use of AFLP markers has facilitated the identification of resistance loci in Colorado potato beetle to pyrethroid (Hawthorne 2001) and in diamondback moth

to *Bacillus thuringiensis* toxins (Heckel *et al.* 1999). Also, biochemical markers have been successfully used for molecular diagnostic purposes for screening resistance to methyl-parathion in western corn rootworms (Zhou *et al.* 2002) and neonicotinoid cross-resistance in *Aleyrodoidea* whiteflies (*Bemisia tabaci*) (Rauch & Nauen 2003).

Prey, predator and parasites

Molecular markers provide information to understand the prey–predator–parasite trophic interactions in insects. Particularly, they are helpful for identification and quantification of prey in the diets of insect predators. Identification of more than one type of prey in the gut of a single insect is performed by genetic fingerprinting of its gut contents and then by developing diagnostic marker systems. Diagnostic PCR-based markers such as SCARs derived from specific RAPD loci were used to identify *Trialeurodes vaporariorum* and *Helicoverpa armigera* prey in the gut of *Dicyphus tamaninii* (Wagner) (Agusti *et al.* 1999, 2000). Mitochondrial DNA markers have also been used to identify *Collembolan* species that comprise a major source of alternative prey to linyphiid spiders (Agusti *et al.* 2003). Apart from identification, marker data are also used to quantify the parasitism rate in insects. Identification of immature parasitoids of *Lysiphlebus testaceipes* and green bugs in different species of small grain cereal aphids have been performed by using DNA markers and the information has been used for estimation of parasitism rates in aphids (Walton *et al.* 1990; Jones *et al.* 2005). Correlation of parasitoid frequencies with molecular data suggested that application of molecular markers could provide accurate estimates of parasitism rates. A comparison of parasite load based on immunological techniques and that based on PCR showed that the PCR technique was more efficient in such studies (Chen *et al.* 2000). Multiplex PCRs using more than one diagnostic marker have been used to develop fast and reliable methods to analyse gut contents in predators. Use of fluorescent-labelled multiplex PCR is a novel approach in this regard (Harper *et al.* 2005). Using this method, more than 10 species have been identified simultaneously. Similarly, microsatellite markers have been used to study the association between host races of *Acyrtosiphon pisum* and their symbionts (Simon *et al.* 2003). Similar work has been performed to identify the bacterial endosymbiont *Wolbachia* in Asian rice gall midge populations (Behura *et al.* 2001b). It was found that adult midges are infested by *Wolbachia* in different frequencies in different biotypes. The results obtained from PCR-RFLP data using mtDNA showed that *Wolbachia* had a profound role in genetic hitchhiking in female midges. These results bear important implications of bacterial parasites causing sex-ratio distortion and population hitchhiking in rice gall midge biotypes.

Behavioural studies

Behavioural plasticity in social insects represents a complex biological phenomenon that is getting attention of molecular biologists. Honeybee (*Apis mellifera*) is an emerging model organism that is being studied for social behaviour at molecular level (Robinson *et al.* 2005). ESTs have been used as expression markers in microarray formats to predict the nursing and foraging behaviour in individual bees (Whitfield *et al.* 2003). In honeybee, these social behaviours are polygenic traits and are influenced by more than one gene referred to as QTL. The two major QTLs that determine the foraging behaviour in honeybee have been identified by employing RAPD markers in backcross populations between bees collecting nectar and those collecting pollen (Hunt & Page 1995). Exploiting similar procedures with molecular markers in honeybee, colony level behaviours such as stinging behaviour, body size, pheromone alarm level, traits for reversal learning and hygienic behaviour have also been dissected at the level of specific genomic regions (Breed *et al.* 2004). AFLP markers and microsatellites have been used in dissecting the guarding and stinging behaviours in honeybee (Arechavaleta-Velasco *et al.* 2003). In bumble bees, AFLP markers have been employed to study genetic basis of ecological implications of foraging range and nest density behaviours (Knight *et al.* 2005). In Australian meat ants, *Iridomyrmex purpureus* (Smith), the movement of females between colonies has been successfully traced by using mtDNA markers. The data help understand the kin selection as the explanation for multiqueen colonies in this species (Crozier *et al.* 1997). Also, in sugar ants (*Camponotus consobrinus* Erichson), microsatellites and mtDNA have been used as markers to study their colony behaviour, and it was found that unexpected genetic diversity and complexity existed in colony structure and behaviour in these ants (Crozier *et al.* 1997).

Gene, genome and QTL mapping

Genetic linkage mapping, based on recombination frequency, uses molecular markers for tagging and mapping of specific genes and QTL in insects. Construction of genetic map uses segregating populations (generally F2 backcross progenies) to genotype them at hundreds of marker loci. Based on the inheritance pattern of these loci, multipoint linkage analysis is performed to generate different linkage groups. In insects, RAPD and AFLP markers have been extensively used to generate genetic maps. RAPD-based linkage maps have been constructed for genomes of honeybee (Hunt & Page 1995), silkworm (Yasukochi 1998), beetle (Beeman & Brown 1999; Yezerski *et al.* 2003) and sawfly (Nishimori *et al.* 2000), whereas AFLP-based linkage maps are available for genomes of silkworm (Tan *et al.* 2001), Colorado potato beetle (Hawthorne 2001), red flour beetle (Zhong *et al.* 2004), Hawaiian cricket (Parsons & Shaw 2002), European corn borer (Dopman *et al.*

2004), Hessian fly (Behura *et al.* 2004) and butterfly (Wang & Porter 2004; Tobler *et al.* 2005). Microsatellite-based maps have also been constructed for the genomes of *Drosophila mojavensis* (Staten *et al.* 2004) and honeybee (Solignac *et al.* 2004).

Apart from genome mapping, DNA markers are also used to tag and map specific gene(s) responsible for a phenotypic trait in insects. For example, to tag an avirulence gene in an insect pest that is responsible for triggering resistance mechanism in host plant (a monogenic trait), backcross F2 progenies are generated from a cross between a virulent male and avirulent female. Mating strategy is designed depending on the objective of the experiment and is also based on the sex determination mechanism of the organism. The segregating progenies are then scored for the parental phenotypes, virulent type or avirulent type as in this case, and then are subjected to multiloci DNA fingerprinting to score individual genotypes. These marker data are then used for performing two-point (or three-point) linkage analysis to determine the genetic distance of the marker(s) segregating in association with the phenotype. Using such approaches, AFLP markers have been used in gall midges to identify and map major avirulence genes in different biotypes of Asian rice gall midge (*Orseolia oryzae*) (Behura *et al.* 2000), and of Hessian fly (*Mayetiola destructor*) (Rider *et al.* 2002; Behura *et al.* 2004). However, when more than one gene govern the trait (polygenic trait), markers are identified those link to the chromosomal region containing the genes, often known as QTL. QTL mapping is often difficult unlike mapping monogenic traits. Unfortunately, many economical and behavioural traits in insects are polygenic in nature. QTL mapping is primarily based on identifying segregation of a single gene marker and estimating the effect of its linkage to the polygenic trait. Statistical techniques, for example, analysis of variance (ANOVA), are used to check if trait means for one marker genotype at any given locus differ significantly from the trait means for alternative marker genotypes. This will indicate if there is an association between the segregating markers and the character of interest. Generally, three major methods are employed for QTL mapping: single marker analysis, multiple regressions (interval mapping), and marker regression approach. The choice of a marker system for the genotyping in these mapping is a crucial factor as one needs dense markers near the QTL for QTLs are very low in heritability by nature. AFLP markers have been successfully used to map QTL governing ageing in *Drosophila melanogaster* (Luckinbill & Golenberg 2002), foraging behaviour in honeybee (Ruppell *et al.* 2004), susceptibility in beetles to tapeworm parasites (Zhong *et al.* 2005), variation in pheromone composition in *Heliothis* species (Groot *et al.* 2004) and Bt (*Bacillus thuringiensis*) resistance in diamondback moths (Heckel *et al.* 1999). AFLP markers have also been used to trace the phenotype traits of wing size in pea aphids (*Acyrthosiphon pisum*) (Braendle *et al.* 2005), to study genetics

of male dispersal in moth species (*Thaumetopoea pityocampa* and *Th. wilkinsoni*) (Salvato *et al.* 2002), and to investigate the genetic basis of haplodiploidy in *Venturia canescens* wasps (Reineke & Lobmann 2005). Apart from AFLP markers, uses of microsatellite markers have been very successful specially in honeybees, bumblebees and ants to identify genes responsible for diversity in foraging range and mating behaviour (Pearcy *et al.* 2004; Knight *et al.* 2005), host parasitization (Solignac *et al.* 2005), colonization (Jensen *et al.* 2005), and kinship relation (Trontti *et al.* 2005).

Comparative genomics and cytogenetics

Molecular markers have important applications in modern cytogenetics and comparative genomic studies of insects. Satellite DNA as marker can provide useful information on gene structure, order and polymorphisms of chromosomal DNA. Fluorescence in-situ hybridization (FISH) on chromosome preparations with (TTAGG)_n satellite markers as probes showed that these repeats were extremely conserved in telomeres of chromosomes in many insect species and also in hexapods, crustaceans, myriapods and pycnogonids (Vitkova *et al.* 2005). Ribosomal DNA as probes is also useful in FISH for comparative cytogenetic studies (Cabrero *et al.* 2003). Transposable elements, because of their repetitiveness in the chromosome, are suitable markers to physically localize them onto chromosomal DNA and to compare heterochromatin and euchromatin regions of closely related species (Berloco *et al.* 2005; Torti *et al.* 2005). FISH localization of BAC clones genetically mapped onto the genome of Hessian fly (*Mayetiola destructor*), an insect pest of wheat, demonstrated severe crossover suppression in an autosomal segment containing two major avirulence genes required for host resistance (Behura *et al.* 2004). Also, a ~1 megabase region of the sex chromosome X2 of Hessian fly has been assembled into a single contig using bacterial artificial chromosomes (BACs) with the help of genetically mapped molecular markers. The BAC clones have been used in fibre-FISH experiments to physically map them onto Hessian fly DNA to confirm their order of location on the chromosomes. These BAC ends have then been used for comparative sequence analyses with the genome sequences of *Anopheles gambiae* and *D. melanogaster* to understand the order and synteny of these loci in these three dipteran insects (Lobo *et al.* 2006). Similar studies have been performed in *Anopheles* species, using extensive *in situ* hybridization experiments, to understand inversion and gene order shuffling in *Anopheles funestus* as compared to *A. gambiae*, (Sharakhov *et al.* 2002). Apart from comparative studies, molecular markers are also used to study cytogenetics of unusual sex determination mechanisms in insects, such as unisexual progeny production in gall midges, or production of only daughters by the virgin females in the wasp *Trichogramma cacoeciae* (thelytoky) (Harris *et al.* 2003; Vavre *et al.* 2004).

Study of fossils and museum samples

DNA analyses of fossil samples are important tools in studies of archaeology, conservation biology, forensic science and entomology. Information obtained from ancient DNA can contribute to the interpretation of history of extinct populations or species (Brown & Brown 1994). Recovery of good quality DNA from museum samples is the first step towards successful analyses of their sequences. Though in some cases, isolation of quality DNA from amber- and copal-preserved insects is problematic (Austin *et al.* 1997; Walden & Robertson 1997), proper optimization of the DNA extraction procedure is required for individual cases (Junqueira *et al.* 2002). Successful recovery of ancient DNA from specimens of Dominican amber fossil termites (DeSalle *et al.* 1992) and wood gnats (DeSalle 1994) has been reported. DeSalle *et al.* (1992) were able to amplify gene fragments of mitochondrial 16S ribosomal gene from the ancient DNA samples by PCR, and based on phylogenetic analyses, it was shown that the fossil termites shared several sequence attributes with *Mastotermes darwiniensis* (Froggatt). Based on sequence polymorphisms of DNA isolated from museum specimens from the late 19th and early 20th centuries, Goldstein & Desalle (2003) were able to identify a relatively recent fixation event of anthropogenic character associated with the extinction and fragmentation of populations in tiger beetle *Cicindela d. dorsalis* (Coleoptera: Carabidae). Santolamazza *et al.* (2004) used PCR-RFLP of DNA from museum samples to identify *Anopheles arabiensis*, *A. gambiae* and their two molecular forms. Similarly, Cano & Borucki (1995) used 16S ribosomal DNA markers to identify an ancient bacterium closely related to extant *Bacillus sphaericus* from fossil sample of 25–40 million years old of extinct bees. DNA bar-coding of museum specimens is successfully used in conjunction with morphological study to show that *Astraptes fulgerator*, a Neotropical skipper butterfly (Lepidoptera: Hesperidae) is a complex of at least 10 species in northwestern Costa Rica (Hebert *et al.* 2004). Widmer *et al.* (2000) collected insects from nectar-rich plants flowering near natural orchid populations and from museum samples to the study of orchid–pollinator relationships. Sequence analysis of the nuclear ribosomal ITS region, in this study, allowed the identification of the orchid species or species-group from which the pollinaria originated. Such DNA method is extremely helpful in understanding orchid–pollinator relationships and especially the degree of pollinator-specificity as direct observations of pollinator visits to orchids are often difficult and time-consuming.

Novel marker systems

Apart from the above popular markers, novel marker systems are recognized to have potential applications in insect studies.

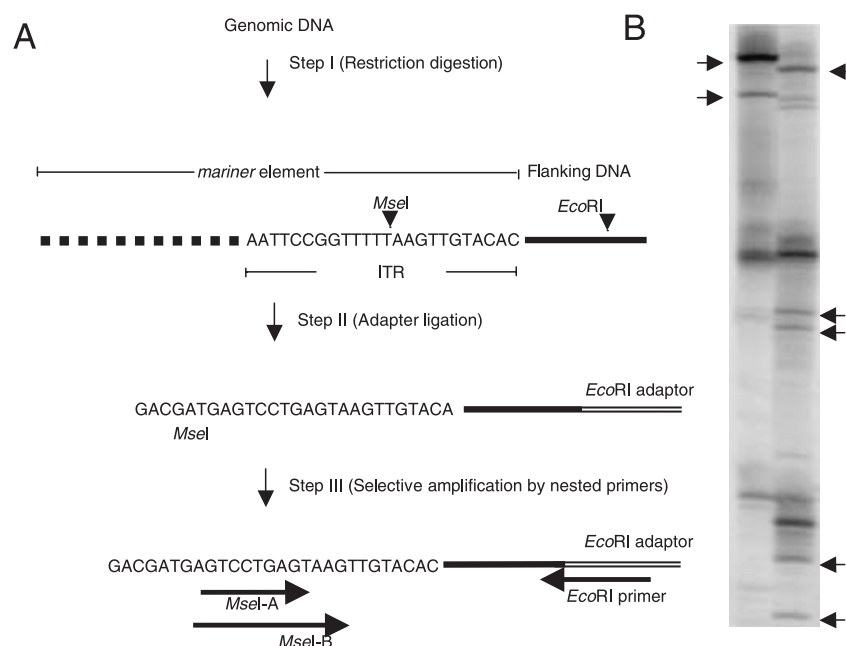


Fig. 1 (A) Principle of 'transposon display' technique. Genomic DNA is digested with a restriction enzyme that has recognition sites in the conserved sites such as inverted terminal region of the transposon. Another enzyme is employed in the second digestion that digests the DNA in the flanking sites of the transposons. Sequences of multiple copies of that element should be checked prior to selection of the second enzyme to make sure that its recognition site is not present in the internal sequences of the transposon. After the digestion, specific adapters compatible to the restriction sites are used to ligate to the digested molecules. The ligated molecules are used as templates to perform two rounds of PCR. In the first PCR, adapter-specific primers are used. The product of this PCR is diluted and used as template for the second PCR where a nested primer (radiolabelled) specific to the transposon is used. The 3' end of this nested primer is extended arbitrarily up to three nucleotides (if inverted terminal repeat sequences are used to make primers in the first PCR). When internal regions of the element (such as conserved motifs of transposase) is used as anchor primer, the 3' end extension can be made according to the flanking sequences beyond that motif in the known sequence of the transposable element. The amplified products are then resolved through a sequencing gel and autoradiography is performed to score the banding pattern. (B) An example of banding pattern generated by the above method using genomic DNA of two laboratory strains of Hessian fly, *Mayetiola destructor* (this was done in the laboratory of Dr Jeff Stuart at Purdue University, USA in the course of standardization of this technique in Hessian fly). The primers employed in this display used the sequences of inverted terminal repeats of *mariner*-like elements. Arrows show the differences in the banding patterns.

Most of these markers are developed by PCR-based techniques such as 'transposon display' (Waugh *et al.* 1997) or 'anchored polymerase chain reaction' (Ayyadevara *et al.* 2000). Numerous kinds of repeat elements such as DNA transposable elements, retro-elements, satellite repeats are present in insect genomes (Box S1, Supplementary material). Some of them are present in diverse insect species and also show sequence conservation (to some degree) within or between the species. Using PCR, it is possible to anchor to these conserved regions and amplify the flanking polymorphisms to generate hundreds of repeat-associated genetic markers in an inexpensive way. Transposon display (TD) and sequence-specific amplification polymorphism (S-SAP) are two such techniques which have been very well applied in marker development projects in plants (Waugh *et al.* 1997; Casa *et al.* 2000; Kentner *et al.* 2003; Lee *et al.* 2005) and also have recently been applied in entomology (Behura *et al.* 2001a; Guimond *et al.* 2003; Zampicini *et al.* 2004).

Transposon display

Transposon display technique is applied to identify polymorphisms associated, generally in the insertion sites of transposons, by using PCR (Waugh *et al.* 1997). The technique is very similar to the AFLP method (Vos *et al.* 1995) except that it employs one primer anchoring to a conserved region of the transposon (commonly, the terminal repeats or internal transposase coding sequences) and another primer to the adapter attached to flanking sites generated by restriction digestion (Fig. 1). Then, nested PCRs are performed to selectively amplify the insertion site polymorphisms.

The DNA mediated transposon '*mariner*' is widespread in insects (Robertson 1993). It is a part of the IS630-Tc1-*mariner* superfamily where the DD34D domains of the elements are highly conserved. The *mariner* elements are 1.2–1.4 kb in length and have two ~12-bp perfect inverted terminal repeats (ITRs) required for transposition. They code for a protein (transposase) that catalyses the excision and

insertion events during transposition. Robertson (1993) designed degenerate primer pairs based on the highly conserved amino acid motifs in the *mariner* transposase coding regions. These primers have been extensively used to identify *mariner*-like elements in different insect species (Box S1, Supplementary material). Because of widespread presence in insects and presence of conserved sequence motifs, *mariner* elements are used as anchors in the genome to develop transposon display methods. *Mariner*-like elements present in the genome of Asian rice gall midge (*Orseolia oryzae*), an economically important pest of rice, have been exploited using this approach to characterize the insertion site polymorphisms of these elements in different biotype populations in India (Behura *et al.* 2001a). *Mariner*-specific primers, along with different combinations of adapter-specific primers, were employed to identify the polymorphisms within and between five biotypes of this pest. Southern hybridization of genomic DNA with a *mariner* probe suggested that these elements were highly repetitive elements in rice gall midge genome. Moreover, the transposon display results showed that specific insertion site polymorphisms were present in *Orseolia oryzae* genome in a biotype specific manner. Southern hybridization of the gel blots (of transposon display products) with a *mariner* probe confirmed these results. Such observation implicated that insertion of *mariner* elements in rice gall midge was probably associated with some 'fixed' mutations in the flanking regions. Transposon display technique has also been used (Behura *et al.* unpublished data) to identify and map three *mariner*-flanking polymorphisms as additional markers to the linkage map of Hessian fly (*Mayetiola destructor*), a major insect pest of wheat (Behura *et al.* 2004).

Hermes is another DNA-mediated transposon that has potential utility in transposon display in insects. It is a member of the *hAT* (*hobo*, *Ac*, *Tam3*) family and is present in at least 15 insect species including fruit fly, mosquitoes, beetles and moths. The element is ~3 kb in length with 17-bp ITRs. Guimond *et al.* (2003) used *hermes* element, cloned in a plasmid, and injected to *D. melanogaster* embryos to study the insertion patterns of these elements in the transgenic flies. Using transposon display method, Guimond *et al.* (2003) were able to investigate the insertion patterns of *hermes* elements in the transgenic flies and showed that *hermes* can be used as a potential vector for insect genetic transformation experiments. Similarly, *Wukong* is another highly repetitive transposable element (with inverted terminal repeats) discovered in the genome of yellow fever mosquito *Aedes aegypti* (Tu 1997). Structurally, these are very similar to the miniature inverted-repeat transposable elements (MITEs) found in plants. The sequences of these MITE-like elements show high sequence similarities between copies, which are highly iterative (2100–3000 per haploid genome) (Tu 1997). This shows its potential utility in transposon display of yellow fever mosquitoes to generate large

numbers of linked markers. Similarly, RNA-mediated mobile elements can also be used for transposon display methods. Such an approach has recently been incorporated in population studies of the midge *Chironomus riparius* Meigen (Zampicini *et al.* 2004). The retrotransposon, NLRCth1 was used to generate fingerprints of the samples and to assess genetic distances between the midge populations.

Apart from transposable elements, many repetitive satellite elements present in insect genomes can also be used in anchored PCR-based assays (Ayyadevara *et al.* 2000) to develop genetic markers. For example, *Bkm* element (*Banded krait minor*) discovered in *D. melanogaster* (Singh *et al.* 1984; Accession no. K01664) is a moderately repetitive satellite sequence in fly genome and has been successfully used in fingerprinting of ecotypes of silkworm (Nagaraju *et al.* 1995). Similarly, an *AluI* repeat element has been identified in the genome of honeybee (*Apis mellifera*) which is estimated to be in 23 000 copies in bee genome (Tares *et al.* 1993). The element is 176 bp long (Accession no. X57427) and is extremely conserved among the copies. This has been used as probe in RFLP fingerprinting of different geographical populations of honeybee (Tares *et al.* 1993). There are many other such minisatellite elements (see Box S1, Supplementary material), such as *cla* element (X78085) in Anisopodidae *Chironomus thummi*; *A4* element (AF465798) in phytophagous grasshopper *Dichroplus elongates* (Acrididae); *pRa-43* (AY141125) in *Rhynchosciara Americana* (Diptera: Sciaridae); *bm1* repeat locus (AB063493) in *Bombyx mori* or *Dromini* (M62837) in *Drosophila mauritiana*; those may also be used to develop anchored PCR markers. Using similar PCR-based approach, STS markers can be developed from direct amplification of BAC ends either by sequencing them or by direct PCR amplification of the ends of large insert clones such as BACs (Lobo *et al.* 2006), YACs (yeast artificial chromosomes) and even cosmids. A method of direct PCR amplification of BAC ends is provided in Box S2, Supplementary material.

The major advantage of these novel marker systems is that they do not require establishing the sequences of each locus to generate these markers. Because the technique uses just one degenerate primer along with adapter specific primers, the cost of generating and screening novel markers is less than that of the conventional markers. The methods are relatively simple, can be performed by either radiolabelling of anchor primer, or by silver staining of the PCR products. In some species, DNA transposons are present in high abundance, thus making these methods more suitable for construction of linkage maps for their genome. Moreover, many different restriction enzymes and/or linker specific primers can potentially be used to generate the flanking site polymorphisms at both sites of the elements. *In situ* hybridization with the transposon as the probe can be used to detect the physical locations of these loci in the chromosomes. This can be compared with the genetic map positions

of the loci by transposon display to identify any discrepancy between genetic and physical positions of the markers. These approaches may help identify marker recombination or crossover suppression in certain chromosomal regions (Behura *et al.* 2004; Yoshido *et al.* 2005). However, in spite of these advantages, the novel markers have not gained widespread popularity. The major drawback in these techniques is that if nested PCR conditions are not properly optimized, nontarget amplifications are often generated. However, Southern hybridization may be employed on the gel blots of the PCR products in order to exclude the nontarget amplifications (Behura *et al.* 2001a).

Some concerns on marker applications in insect research

Increasing evidences are now known for indirect selection on mtDNA and microsatellite loci that goes against the so-called 'neutral' nature of these popular markers. Thus, their application as molecular markers, particularly in the studies of species evolution and population dynamics, can potentially mislead the interpretation based on the data generated from such marker assays. Thus, these issues should be understood for meaningful interpretation of data derived from mtDNA and microsatellites markers in some situations.

Integration of mitochondrial sequences in nuclear DNA (referred to as Numts) has been discovered in a variety of organisms including insects (Zhang & Hewitt 1996). Thus, PCR amplification using mtDNA marker loci using total genomic DNA can potentially amplify these nuclear copies. These sequences complicate the employment of mtDNA as a molecular marker in ecological studies. In insects, because of the relative small genome size, high copy number of Numt sequences may interfere in effective separation of mtDNA from its nuclear paralogs. Numt sequences do not retain their function after their departure from the mitochondrial genome; hence, they diverge from each other without any selective constraints (Bensasson *et al.* 2000). This has been evident in brown mountain grasshopper, *Podisma pedestris* Linneus (Orthoptera: Acrididae) where the Numt sequences were found to have different mitochondrial origin (Bensasson *et al.* 2000). Thus, population studies using mitochondrial markers derived from these loci can potentially mislead the results.

Another problematic issue of using mtDNA markers has been identified in cases where the host insect harbours maternally inherited microorganisms such as *Wolbachia*. It is a gram-negative endosymbiotic bacterium that causes many developmental defects such as cytoplasmic incompatibility, feminization and sex ratio distortion (Werren 1997). *Wolbachia* is widespread in insects and more and more insects are now being tested by molecular methods to have these bacteria (Jeyaprakash & Hoy 2000). As the *Wolbachia* infection sweeps through an insect species, the frequency

of mitochondria from infected individuals also increases in the population due to the similar mode of transmission used by *Wolbachia* and the mitochondria. As a result, the spread of the mtDNA from infected individuals reaches high prevalence in these populations (Turelli *et al.* 1992; Rigaud *et al.* 1999). This phenomenon is commonly referred to as 'genetic hitchhiking'. Thus, indirect selection of host mtDNA by such microorganisms has been a major concern (Hurst & Jiggins 2005). Apart from selective sweep, presence of more than one strain of *Wolbachia*, and genetic recombination between the strains have also been recognized as alternative routes for reduction or elevation of mtDNA polymorphism in insect populations (Rousset & Solignac 1995; Hurst *et al.* 1999; James & Ballard 2000; Jiggins 2003). Coupled blockage of maternal inheritance of host mtDNA and that of *Wolbachia* is another unusual confounding interference of *Wolbachia* on host mtDNA observed in *Orseolia oryzae* (Behura *et al.* 2001b). A detailed list of similar but commonly observed cases of indirect selection of endosymbiont microorganisms on host mtDNA has been documented in a recent review (Hurst & Jiggins 2005). Thus, inferring evolutionary history of populations solely based on use of mtDNA markers in insect species harbouring such maternally inherited microorganisms may be misleading.

Similarly, natural selection on microsatellite loci has been observed in many cases that raise doubts to use them as faithful genetic markers (Schlotterer 2000). Differences in microsatellite variations in sex chromosomes compared to autosomes represent an example of such differential selection (Bachtrog *et al.* 1999). The rate of mutation in microsatellite loci is also dependent on repeat units, repeat length, flanking sites and recombination rates of the loci (Schlotterer 2000). Also, genetic hitchhiking (by means of selective sweep) and background selection of microsatellite loci in *Drosophila* have been demonstrated (Kauer *et al.* 2003). Analysing microsatellites loci in ancestral (African) and colonizing (non-African) populations of *D. melanogaster*, it was found that while background selection was the contributing factor for neutral variability in the ancestral populations, variation in colonizing populations was predominantly determined by genetic hitchhiking. Thus, positive selection of microsatellite loci could cause local reduction in genetic variability in such microsatellite loci (Harrift *et al.* 2003). Significant loss of mutations in X-linked microsatellite loci in non-African populations of *Drosophila simulans* compared to the African counterparts caused 'fixation' tendency of these loci in the colonizing populations (Schofl & Schlotterer 2004). Similar bias in rate of genetic variations in microsatellite loci associated with transposable elements has also been documented in *D. melanogaster* (Souames *et al.* 2003). Apart from that, microsatellites have also been suspected to influence gene spacing and secondary folding of DNA in the chromosomes. They are also known to affect chromatin structures in insects (Li *et al.* 2004). In *D. melanogaster*,

association of AAGAC satellite repeats with the rolled (*ri*) gene implicates its possible role in rendering the *ri* gene to undergo extensive polytenization in the salivary gland chromosomes (Berghella & Dimitri 1996). Regulatory roles of simple sequence repeats in gene expression also support the non-neutral nature of those loci (Li *et al.* 2002). Comparison of transcription factors between *Tribolium castaneum* and *D. melanogaster* revealed that many trinucleotides repeat found in *Drosophila* were almost absent in the corresponding homologous genes in the *Tribolium* (Schmid & Tautz 1999). These lines of evidences suggest that microsatellites, once considered as junk DNA, may have potential regulatory roles and are potentially non-neutral (see Li *et al.* 2002 for review).

Null alleles of DNA marker loci also pose difficulties in many genetic analyses. The major problematic issues have been identified in microsatellite applications in studies of parentage, forensic testing or in estimation of allele frequencies (Dakin & Avise 2004). These are alleles of a locus in which the microsatellite flanking primers do not produce any product in PCR. Scoring for these loci should be done with care in order to avoid false inclusion of the presence of a homozygote. This can lead to a false interpretation in the genetic analyses. The conventional way of avoiding this potential difficulty is to treat all homozygotes as heterozygotes with detectable null alleles so that a true heterozygote is not excluded on the basis of homozygosity. Several computer programs, such as PROBMAX, PAPA, CERVUS, PARENTE, FAMOZ, NEWPAT, are also available to tackle the difficulties of null alleles (Jones & Ardren 2003).

Large-scale genotyping

Single nucleotide polymorphisms (SNPs) are highly abundant markers found in the genome of all organisms. These markers have been extensively used in human genome projects (Kwok 2001), and researchers are showing much interest in these markers to explore and exploit in other animals and plants, as well. Some progress have been made in insects too, though mostly limited to *Drosophila* and mosquitoes. In *D. melanogaster*, Teeter *et al.* (2000) used a resequencing approach (of a number of gene fragments and sequence tagged sites) and developed a low resolution SNP map of *Drosophila melanogaster* for the first time. Subsequently, larger sets of STSs were used for sequencing to develop a large SNP data set for *D. melanogaster* (Hoskins *et al.* 2001). Based on this study, at least 52 SNPs in X chromosome, 138 in chromosome 2 and 89 in chromosome 3 have been reported. An SNP map of chromosome 3 of *D. melanogaster* has also been developed (Martin *et al.* 2001). A major contribution to *Drosophila* SNPs was made by Berger *et al.* (2001) by identifying 7223 SNP markers. SNP mapping in *Drosophila nicastrin* has been established by use of denaturing high performance liquid chromatography (DHPLC) (Nairz *et al.* 2002). In *Anopheles gambiae*, single

nucleotide polymorphisms in coding regions of three strains have been identified (Morlais *et al.* 2004). In *Aedes aegypti*, single strand conformation polymorphism (SSCP) analyses using cDNA revealed single nucleotide polymorphisms that were used for genome mapping (Fulton *et al.* 2001; Severson *et al.* 2002). In *Bombyx mori*, EST sequences and genomic contig sequences were aligned to identify 101 candidate SNPs in its genome (Cheng *et al.* 2004). Because SNPs are abundant, they are suitable for large-scale genotyping projects. Some recent advancement in this marker system should be discussed to know how SNPs are identified and how they can be used for genotyping.

In principle, SNPs can be identified by generating nucleotide sequences of a particular DNA region or a particular gene isolated from several individuals in a population. Because the nucleotide sequences from different individuals may vary at many loci in that DNA region, these allelic variations may be manifested as potential SNPs in the multiple alignments of the individual sequences. The approach seems very primitive when dealing with a small piece of DNA fragment or a single gene. But when it comes to SNP-hunting in a chromosome or in the whole genome, the difficulties are manifold. Multiple alignments become increasingly difficult to find the coordinates in the reference sequence due to false alignments of paralogs, repeats, gene families or gene duplications. In such cases, finding true alignment depends upon the correct anchoring to the actual reference sequences. Three main aspects that need to be taken care of are (i) correct anchoring of the multiple alignments, (ii) filtering out the paralogs, and (iii) discriminating true SNPs from sequencing errors. By using the alignment depth, the base call rate, value for quality base, rigorous statistical methods are available that can be used in estimating the probability that a mutation is a potential SNP and not a sequencing error. Several computer algorithms that can be used to identify potential SNPs in DNA trace sequences are listed in Box 1.

Selection of a suitable method for SNP genotyping is mostly dependent on the objective of a project. Commonly, there have been two broad aspects; one is SNP genotyping in particular gene or set of genes, and the other objective is the genome-wide SNP screening. Simple methods of restriction digestion or SSCP or other low to medium throughput techniques can be used to genotype a small set SNPs in a known gene(s). Many of the reaction formats in these methods are homogenous in nature and can be performed in 96-well plates without sophisticated automation and specialized instrumentation. Entomologists and ecologists who don't have access to high throughput genotyping facilities may rely on these methodologies. Specific examples are described on how old techniques that was once performed in single tubes can be formatted into 96-well plates by suitable scaling up procedures. One example is PCR-RFLP method that can be scaled up to for multiplex

Box 1 Computer software to predict single nucleotide polymorphism in DNA sequences

POLYBAYES (Marth *et al.* 1999) program is scripted to perform anchored alignment, paralog filtering and SNP detection using Bayesian methods and reports the SNP locations and the associated probabilities in a text formatted output file.

AUTOSNP program (Barker *et al.* 2003) uses a gapped FASTA-formatted multiple alignment file in cap3 assemble program to identify the potential SNPs. AUTOSNP has recently been used in a web-based SNP Server for online services of SNP identification (<http://hornbill.cspp.latrobe.edu.au/snpdiscovery.html>).

FORAGE program (Unneberg *et al.* 2005) uses two neural networks and a set of training data set of SNPs to identify a base discrepancy in the multiple alignments. First, the query data set (for example EST sequences) are used to perform the basic filtrations. If a site in the alignment exists in at least two sequences with a discrepancy, it uses it as a vector with component features such as the major allele frequency, smallest error probabilities for each minor allele, and SNP probability. The vector is then

submitted to both neural networks and if a unanimous quorum is reached, it is recorded as a SNP candidate.

INSNP program (Manaster *et al.* 2005) uses first stretch of six consecutive bases where there are no differences between the reference and any one the trace sequences. It uses this as the coordinate to anchor the sequences, and checks for base differences till the half distance of each sequences. Only those base differences are marked as SNPs.

NOVOSNP software (Weckx *et al.* 2005) uses a cumulative scoring scheme to score each position in the alignment, which is performed by addition of three subcores (feature score, difference score and peak shift score) determined independently from forward and reverse reads of the aligned sequences with an extra 'type' score. The cumulative scoring scheme shows how well both the forward and the reverse reads match at the particular base which forms the basis of detecting a single nucleotide polymorphism.

SNPCEQER (Flood *et al.* 2002) is used in Beckman CEQ2000 (Beckman) genotype system. It reports discrepancies between individual sequences and the consensus sequence and detects SNPs based on individual base call values. It is available free at <http://innovation.swmed.edu>.

genotyping of SNPs. There are simple computer software such as SNPKIT (Hao *et al.* 2002), SNICKER (Niu & Hu 2004) and SNP CUTTER (Zhang *et al.* 2005) that can be used to identify restriction enzyme recognition sites overlapping the polymorphic bases (after the candidate SNPs in the sequences are recognized). These then can be used in designing pilot scale restriction reactions in 96-well plates to process large number of SNPs in parallel. McSNP (Akey *et al.* 2001) is such a novel SNP genotyping method that refers to as melting curve analysis of SNPs (McSNP). McSNP combines restriction enzyme digestion with DNA melting curve analysis. Melting curve analysis is performed by slowly heating DNA fragments in the presence of the dsDNA-specific fluorescent dye, SyberGreen. As the sample is heated, fluorescence rapidly decreases when the melting temperature of a particular fragment is reached. Using this method, it is possible to determine the allele composition of DNA samples. McSNP is well suited for medium-throughput genotyping because 96 samples can be analysed and scored in just 20 min (Akey *et al.* 2001). Restriction digestion followed by single base extension is also another genotyping approach based on PCR-RFLP (Che & Chen 2004). In this method, a restriction site (*FokI*) is incorporated in the forward primer so that the PCR product can be digested with that enzyme. Upon digestion, the digested molecules carry 5'-overhang structure that can be used for primer extension with labelled ddNTPs. The labelled products are then separated in an automated sequencer to

discriminate the alleles. A similar method, but carried out on solid support, is also used for multiplex genotyping (Shapero *et al.* 2001). PCR amplifications are carried out on microspheres using primers (one primer contains recognition site for restriction enzyme *BbvI*) designed for each polymorphic locus. Then *BbvI* is used to digest the PCR products. The digested DNA is then used for primer extensions with multi colour fluorescent probes in mini-sequencing reaction formats. Apart from the PCR-RFLP methods, other PCR-based techniques such as allele-specific PCR, AFLP, *TaqMan* PCR can also be used for genotyping SNPs and the principles of some of these techniques are briefly outlined in Box 2.

The genotyping techniques based on DNA confirmation are based on differentiation of the DNA molecules due to conformational changes arising from difference(s) in their nucleotide sequences. When a single-stranded DNA molecule is electrophoresed in a nondenaturing gel medium, intrastrand-base pairing creates specific conformation for each molecule. Based on this principle, SSCP is used for genotyping purposes (Orita *et al.* 1989). Traditionally, the PCR product is denatured by heating and then snap-cooled on ice and run in a nondenaturing polyacrylamide gel. The bands are then detected either by ethidium bromide staining or by autoradiography. Genotyping of large number of samples by SSCP is preferably performed either by silver staining or by using fluorescently labelled primers (Schmalenberger & Tebbe 2003; Dong & Zhu 2005). High

Box 2 Low to medium throughput genotyping approaches

Allele-specific PCR: Allele-specific PCR-based assay (ASP) (Saiki *et al.* 1986) uses conventional PCR with two primers designed from SNP-flanking sequences. A common primer matching to both the allele along with two specific primers, with 3' ends matching to the polymorphic bases in the two alleles, are used so that allele specific amplification occurs. Simple electrophoretic gels can detect the products.

TaqMan assay: The TaqMan technique (Livak 1999) utilizes the exonuclease activity (5'–3') of *Taq* DNA polymerase. The oligonucleotide probe contains a fluorescent reporter dye at the 5' end and a universal quencher fluorescent dye at the 3' end. As long as they are close to each other, the fluorescence is quenched. When the probe and the template DNA match, the exonuclease activity of the polymerase cleaves the probe and the fluorophore is released. The emitted fluorescence is then monitored to discriminate the alleles.

On-off switch assay: Use of DNA polymerase with proof reading activity in combination of allele-specific modified primers (generally phosphorothiorate, PTO or locked nucleic acid, LNA to render resistance to the exo-activity of the polymerase) has also been successfully exploited for SNP genotyping (Lin-Ling *et al.* 2005; Zhang *et al.* 2005). Termed as 'on/off switch assay' it works on the simple principle that if the modified primer matches to the template SNP, the proof reading activity does not come into action, and hence, the primer extension takes place. But if the modified primer is mismatched, the DNA polymerase then attempts to employ its exo activity, but fails (because of primer modification). Thus, no primer extension takes place.

RCA technology: SNP specific ligation using padlock oligos followed by rolling cycle amplification (RCA) of the alleles is also an attractive method for SNP genotyping (Qi *et al.* 2001). The method uses cyber-gold staining of the reaction product to identify the polymorphic DNA molecules. Only ~50 ng starting DNA is required per assay, and hybridization-based bi-allelic assay of the rolling cycle PCR products with fluorescence dye-labelled probes makes it ideal for cost effective genotyping method. Commercial format of RCA format is available from Amersham Pharmacia Biotech Inc.

RNaseCut method: Similarly, RNaseCut is an innovative method that is based on digestion of RNA transcripts from PCR fragments (Krebs *et al.* 2003). These are then digested with guanosine-specific ribonuclease T1 to generate sequence specific fingerprints that are then analysed via MALDI-spectrometry.

Phosphodiesterase: Phosphodiesterase enzyme is also used for SNP discrimination. Allele specific PCR product is generated by terminally modified (methylphosphonate) oligo in presence of ddNTPs (Sauer *et al.* 2000). When the product is digested by phosphodiesterase, it digests the amplicon leaving back the modified backbone of the primer portion that is linked to the negatively charged phosphate group of dideoxynucleotide matching to the SNP. This negatively charged tag could then be monitored in MALDI to assign its genotype.

Pyrosequencing: In a pyrosequencing reaction, the DNA polymerase incorporates a nucleotide to the template DNA and it releases pyrophosphate (PPi) that is converted to ATP (by ATP sulphurylase). This releases energy that is converted to light (through oxidation of luciferin), which is detected by a photon detector in order to identify the nucleotide. Template for pyrosequencing is generated by using two or more primers (a common labelled primer at one end and allele specific primers on the other end of the SNPs) in the same reaction to first amplify the DNA containing multiple SNPs. After cleaning the products, they are then passed through a column of sepharose beads (coated with streptavidin) so that after washing single stranded DNAs can be isolated. This is then used to carry out pyrosequencing reactions using the allele specific primers to determine the genotypes in the individual samples. Recently, a multiplexed format of SNP assay based on pyrosequencing method has been developed for high throughput genotyping (Zhou *et al.* 2005).

SNPwave: This is an AFLP-based method that utilizes padlock probes followed by selective amplification (van Eijk *et al.* 2004). This method utilizes the ligation-mediated circularization of the padlock probes. Upon hybridization to the denatured genomic DNA, the 3'-OH end of the probe makes a phosphodiester bond with the adjacent 5'-PO₄ end of a common sequence in the probe resulting in a closed circularization of the molecule. This happens only if the 3' end of the probe matches to the polymorphic base, but when there is mismatch, no close circularization takes place. In the second step, these closed circular molecules are amplified selectively with AFLP primers with two selective nucleotides at their 3' end. This way only those closed ligation probes are amplified which contain perfectly base-paired nucleotides adjacent to common primer sequences. The bands are separated in a capillary (MegaBASE using 1 × LPA matrix) sequencing machine based on their size differences induced by stuffer nucleotides.

SNaPIT Technology: SNaPIT technology (Curran *et al.* 2002) uses the high specificity of Uracil-DNA

Box 2 *Continued*

Glycosylase (UDG), a DNA repair enzyme to digest the template DNA in allele specific manner. At the first step, a PCR is performed to replace the thymine residues with uracils in the template DNA. Then UDG is used to excise the uracil bases and generates apyrimidinic sites, which are cleaved under alkaline conditions. This generates fragments of different lengths based on the alleles, which can be readily detected in denaturing conditions by PAGE or using an ABI-PRISM Genetic Analyser.

ARMS assay: The amplification refractory mutation system (ARMS)-PCR (Newton *et al.* 1989) is similar to allele-specific PCR assay. The PCR assay uses oligonucleotides with 3' ends matching to the two alleles. Two SNP specific primers designed from SNP-flanking sequences at one end and another common primer matching to both the allele on the other side are used to discriminate the two alleles, and simple electrophoretic gels can detect the products. Alternatively, SyberGreen can be used as dsDNA binding dye to efficiently detect large-scale.

throughput approaches for parallel SSCP scoring can be achieved by using automated capillary sequencers, with optimized electrophoretic conditions (Baba *et al.* 2003). Cleavage fragment length polymorphism (CFLP) is another conformation-based genotyping method. The method relies on the use of cleavage I enzyme that cleaves the DNA strands at upstream junctions of any duplex structure (Lyamichev *et al.* 1999). Using this approach, heteroduplexes can be generated by mixing wild type and mutant type DNAs, and then cleavage I is used to digest the duplex DNAs. This can reveal any sequence polymorphisms between the wild type and the mutants. Temperature and assay time are important in the reproducibility of the CFLP pattern, and thus, optimization in the parameters are needed to run all sample in parallel conditions (Oldenburg & Siebert 2000). Nucleotides labelled with different fluorescent dyes may also be used in this method.

Denaturing high-performance liquid chromatography (DHPLC) is also used for SNP genotyping. It uses the fact that under partially denaturing conditions, heteroduplex DNA molecules are denatured faster than the homoduplex molecules. Thus, when DNA samples are subjected in an acetonitrile gradient in a liquid chromatogram, the heteroduplex molecules show more retention time than the homoduplex molecules (Underhill *et al.* 1997). The advantages of DHPLC over SSCP or CFLP are that no PCR amplification is required prior to genotyping and also that the triethylammonium acetate media reduce the range of melting temperature of the amplicons (Nairz *et al.* 2002). Denaturing gradient gel electrophoresis (DGGE) (Fischer & Lerman 1983) employs the differential melting behaviour of DNA molecules based on its sequence composition. A single base difference between two DNA molecules can alter the way they may denature. DNA molecule shows differential electrophoretic mobility based on the concentration of the denaturant or also temperature. DGGE methods are specifically suitable when GC rich sequences in the DNA pose difficulty for mutation screening.

Genome-wide SNP mapping is undertaken to construct haplotype maps or linkage blocks, to develop high reso-

lution linkage maps and map unknown genes, or to study population structure and genetic diversity. It involves the screening of many thousands or even millions of SNPs present in the genome in large number (hundreds or even thousands) of individuals. This is where the need of high throughput reaction formats and detection systems are necessary to achieve the goal in a cost effective way. There are several high throughput methods now available that can be used for genotyping thousands of loci in relatively short time. Large-scale genotyping is generally carried out using two basic steps: allele discrimination, and allele detection. There are three basic biochemical principles generally used for allele discrimination reactions. These principles are allele specific hybridization, primer extension or ligation (Fig. 2). Allele specific hybridization is a straightforward approach where the probe is hybridized to the DNA if the allele sequence is complementary to that of the probe. There will be no hybridization if the DNA contains mismatching sequence. In primer extension method, an oligonucleotide probe is used that has a 3' end base complementary to the polymorphic base in the template DNA. When chain elongation takes place, only the template matching to the oligo is allowed to extend. On the other hand, in primer ligation method, two oligos are used, one specific to each allele and another common to both the alleles. The first step of the assay is performed at low annealing temperature to anneal the common oligo to the template DNA. The second step is performed at high temperature to ligate the allele-specific oligos. The ligation will take place only when the allele-specific oligo matches to the SNP site. There will be no ligation product in the case of mismatching oligo. Using these principles, either homogenous or solid phase reaction formats can be adopted for SNP genotyping. In homogenous formats, the whole assay is performed in solution from beginning to the end without any need of separation or purification of reaction products. On the other hand, solid phase reactions are performed on the solid surfaces of glass slides or array beads or silicon chips, and these methods are preferred for large-scale genotyping assays. In these cases, the allele specific oligo probes are covalently attached to

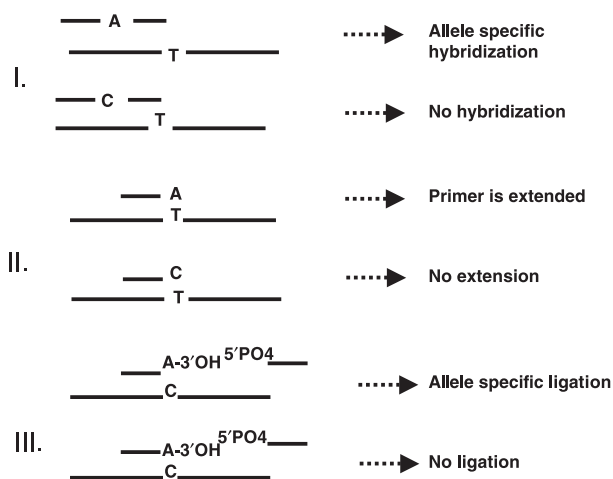


Fig. 2 Principles of different methods used for allele discrimination. I. Allele-specific hybridization: When two alleles having a single nucleotide polymorphism (A to C) hybridized to the oligonucleotide probe with 'T' at the position of mutation, only the allele with 'A' as the polymorphic base, hybridizes. The other allele with 'C' does not hybridize. II. Primer extension method: It uses an oligonucleotide probe with 3'-end base complementary to the polymorphic base in the template DNA. When chain elongation takes place in PCR, only that template matching to the oligo is extended. The primer for the other allele is not extended. III. Allele specific ligation: It uses two oligos. One primer is allele specific and fluorescently labelled. The second primer (unlabelled) is common to both the alleles. The first step of the assay is performed at low annealing temperature to anneal the common oligo to the template DNA. The second step is performed at high temperature to ligate the allele specific oligos. The ligation will take place only when the allele specific oligo matches to the SNP and there will be no ligation product in the case of mismatching oligo. Monitoring the fluorescence of the product allows discrimination of the two alleles.

these solid surfaces, and sample DNA (to be genotyped) is used as template in the allele discrimination reactions.

Common detection techniques in SNP assays are based on fluorescence, luminescence, and fluorescence polarization or fluorescence resonance energy transfer (FRET). Fluorescent labelling of a nucleotide or an oligo probe is the most common approach. Incorporation of the fluorogenic material into the reaction product identifies the allele matching or mismatching to the polymorphic base. Luminescence is mostly used in pyrosequencing methods, and is a direct measure of amount of template DNA incorporated in the ATP-dependent luciferase reactions. Pyrosequencing uses four enzymatic reactions, namely: (i) polymerase reaction for extending DNA strands, (ii) ppi conversion for producing ATP by the catalysis of ATP sulphurylase, (iii) light production by a luciferase-luciferin reaction, and (iv) a dNTP degradation reaction by apyrase. If the added dNTP is complementary to the base at SNP site and is incorporated into the hybridized sequencing primer, a peak is observed in the sequencing spectrum identifying the allele. Similarly, the

principle of fluorescence polarization is based on physical properties of plane polarized light. Upon excitement, the emitted light from the fluorescently labelled molecule is polarized to a degree that depends directly on the molecular weight of the DNA. Thus, monitoring the fluorescence polarization of the emitted light, the differences in the molecular weight of two DNA molecules differing in a single nucleotide can be determined. On the other hand, the principle of FRET detection is based on close proximity of the donor and the acceptor dye whose emission spectra overlap with each other. If this proximity is destroyed or created in the allele discrimination reaction, transfer of fluorescence resonance energy takes place. It is a preferred method of detection for high throughput SNP assays using primer extension or ligation (where proximity is created) or 5' nuclease reactions (where proximity is destroyed). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOF-MS) is another attractive method for allele detection. It is based on the principle that when a temporally and spatially well-defined group of ions of differing mass/charge ratios are allowed to drift in a region of constant electric field, they will traverse with time that depends upon their mass/charge ratios. The method can discriminate single base differences in nucleotides in two DNA alleles based on their mass/charge ratio (Wada *et al.* 1999). Based on the above genotyping and detection methods, several high throughput technologies are available for assaying large number of SNPs in parallel. Some of the array-based SNP genotyping methodologies, including Affymetrix Tag Arrays (Fan *et al.* 2000), Illumina's bead-array system (Gunderson *et al.* 2005) and arrayed primer extension (APEX) (Kurg *et al.* 2000), are very promising for exceptionally high throughput and multiplexing assays. Principles of some of these techniques are briefly described in Box 3.

Like SNP, 'single feature polymorphism' (SFP) is an emerging marker system to screen genome-wide DNA polymorphisms. Whole genome microarray is used as a genotyping platform for this purpose to identify these polymorphisms. Single features are fabricated by using oligos (typically 25 bp) from the genome sequences. The basic principle of identification of single feature polymorphism using this method is illustrated in Fig. 3. This method has been used successfully to identify genomic regions and QTLs governing various phenotypic traits in plants and also in yeast (Borevitz *et al.* 2003; Winzeler *et al.* 2003). The method can be used to detect SFPs in insects to understand genetic diversity in closely related species or between different biotypes or strains. It has been recently used for genomic differentiation of two sympatric forms M and S of African malaria mosquito *A. gambiae* (Turner *et al.* 2005). Using Affymetrix GENECHIP containing 142 065 probes (25 bp) covering the entire genome of *A. gambiae*; Turner *et al.* (2005) performed microarray hybridization with genomic DNA from the two strains. Using this approach, they identified

Box 3 High throughput genotyping technologies

Invader assay: It uses flap endonucleases that specifically recognize the overlapping nucleotides in a DNA structure formed by hybridizing two overlapping oligos to a target template molecule (Lyamichev *et al.* 1999). It then cleaves the downstream oligonucleotide probe to remove the 5' end flap nucleotides of the perfect overlapping structure. Two signal probes (one for each allele) are designed so that they overlap the polymorphic base in the template DNA. When an upstream oligo probe (invader) is used for hybridization, the downstream oligo is cleaved if only the signal probe has a match to the polymorphic base. No cleavage takes place if the oligo is not specific to the allele. Then the cleaved allele-specific probe is used in a (fluorescence resonance energy transfer) FRET-based reaction to amplify the signal so that the oligo cleavage can be easily monitored (Ledford *et al.* 2000).

APEX: This is basically a primer extension-based allele discrimination technique coupled with microarray-based detection. The oligonucleotide probes are immobilized on a glass surface. The target genomic DNA is digested with specific restriction enzymes and is hybridized to the immobilized primers. Then, labelled ddNTPs are used for primer extension that can detect a match or mismatch in the template DNA based on monitoring the labelled ddNTP incorporation in the reaction products (Kurg *et al.* 2000).

Bead array technology: The technology (Gunderson *et al.* 2005) uses beads with two 75-mer oligos as capture probes. The capture sequences in both oligos are the same; only the 3' end base differs from each other. Pre-amplified genomic DNA is used for hybridization onto

the bead array. Then, an array-based primer extension reaction is performed in presence of biotin-labelled dNTPs. If the captured genomic DNA molecule had a match to the 3' end of the bead probe, biotin label gets incorporated, and if that is a mismatch, no label is incorporated. The incorporated biotin signal is then amplified using a sandwich protocol and then after washing the array, the signal is recorded by a suitable imaging system to score the SNPs.

Tag Array: The technology (Fan *et al.* 2000) uses generic oligonucleotide arrays containing thousands of pre-selected 20-mer oligonucleotide tags. The genomic DNA is amplified by SNP specific primers. The PCR product is then used for single base extension (SBE) reactions using chimeric primers with 3' end complementary to the specific SNP sites and 5' end complementary to specific probe tags on the array. Multi-coloured dyes are used to label ddNTPs that is used in the SBE reaction. The extension primers, terminating one base before the SNP sites are thus extended with different label for the two SNP alleles. This is then hybridized to the tag array. The genotypes are deduced from the fluorescence intensity ratio of the colours.

MassCode: The technology (Kokoris *et al.* 2000) uses small tags covalently attached to the primers used for allele-specific amplifications. The tag is attached to the primer via a photo-cleavable linker molecule. After the amplifications steps, the tags are cleaved from the primers by photolysis. SNP alleles in the product are then determined by the relative ratio of the paired tags (coming from matched and mismatched alleles). Multiplexed formats using 30 different tags can be done using this approach with a potential to genotype 40 000 SNPs in a day.

three specific genomic regions differentiating the M and S strains. Mutations in these three regions (~2.8Mb that contained 67 genes) on chromosome 2 L, 2R and X are fixed unlike those in the nearby loci. In spite of gene flow between the two strains, these three specific genomic regions remained differentiated between M and S strains suggesting that these are speciation islands containing genes responsible for developing reproductive isolation between them (Turner *et al.* 2005). Particularly, this method is regarded as a very powerful emerging technology for genomic analysis of populations undergoing recent genetic differentiation (Butlin & Roper 2005).

Concluding remarks

Because of the simple nature of DNA itself, we tend to assume that the technologies to detect DNA polymor-

phisms are not infinite. If this statement is correct, then we must be getting near the asymptote of the sigmoidal graph of DNA techniques vs. time (let's say since 1953), and that it is unlikely (although not impossible) that many more techniques will be developed in the next 20 years or so. However, at this point, the field of molecular marker technology is fast progressing by adopting new forms and innovative approaches of the existing genetic principles in detecting DNA polymorphism.

One example of this scenario is the use of novel approaches for generating new class of genetic markers. Insect chromosomes are home to numerous DNA and RNA transposons, some of which are highly abundant in the genome. Hence, comparative sequence studies of the repeat elements in diverse insect species can provide useful information on how to make use of them for developing abundant markers that can be used in those species. Another example is the

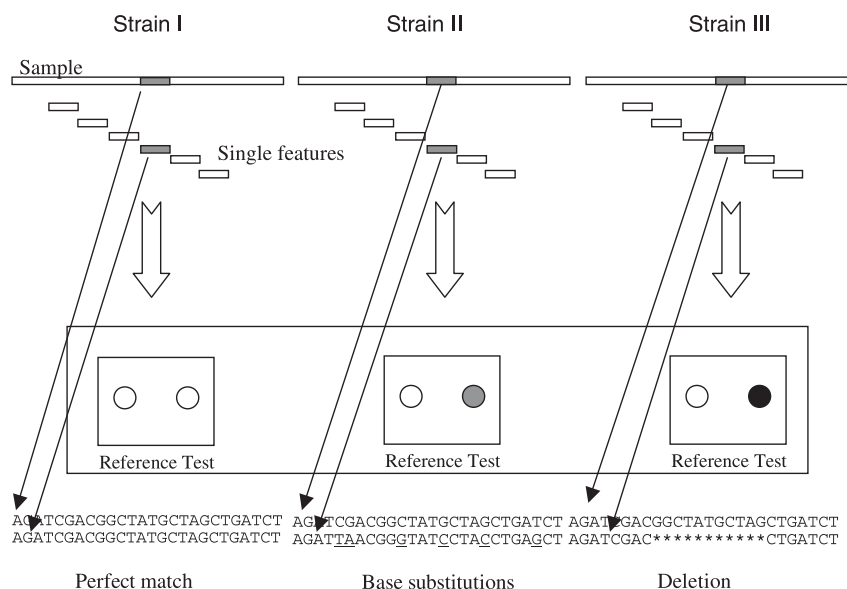


Fig. 3 Microarray methods used for detection of single feature polymorphisms (SFPs). Genomic DNA from three strains of a species shows differential intensity in the signal when hybridized to a single feature to a particular gene compared to the reference gene signal. Mutations (either base substitution or deletion) in the corresponding locus in the related strains give rise to differential hybridization strength, when hybridized to a particular 'feature'. That generates differential signal intensity in the microarray hybridization. A reference feature is used as control to compare the signal intensities in samples. Arbitrary sequences are provided to explain the principle. The common sequence (top) for each strain represents a hypothetical feature. The DNA sequences of the loci corresponding to the feature in the three species are shown below (indicated by arrows). The bases underlined shows sequence differences of the targets (in strain II). The deletion mutations of the target compared to the feature in strain III are shown by 'asterisks'.

possible application of genomic tools to identify EST markers and other cDNA markers that may be derived from microRNA targets. Global analysis of EST using microarray technology has shown unseen power in revealing expression polymorphism in insects (Whitfield *et al.* 2002, 2003). The expression and mRNA abundance of many protein-coding genes are regulated by microRNAs. The microRNAs are small (~22 nt) regulatory RNA molecules that bind to the small stretches in the 3'-untranslated regions (UTR) of target mRNAs and cause post-transcriptional regulation of the target genes. A target mRNA often bears conserved sequences at several locations in the 3-UTR (for example, K box, Brd box, GY box in those of mRNA of Notch genes) where microRNAs bind. Thus, in principle, primers designed based on these conserved regions can be used in PCR amplification of the cDNA fragments corresponding to the microRNA binding regions (Boutla *et al.* 2003). Also, as the 5' end of the microRNA (2–8 bases) makes perfect complementary base pairing with the mRNA, it should be possible to use PCR primers, with 3' end (8 bases) complementary to the 5' ends of microRNAs, to fingerprint cDNA by cDNA-PCR approaches (Behura *et al.* unpublished observation). Because, microRNA has hundreds of targets in the genome and often each target has binding regions for multiple microRNAs, these abundant microRNA-binding sites in messenger RNA can be used to generate thousands of cDNA markers.

The most awaiting aspect, at least to me, is the widespread use of high throughput SNP genotyping in insect studies. These markers are highly abundant, and hence, can be the best marker system to perform genetic analyses for ecological studies. Candidate SNPs can be identified by sequencing a set of loci from a small set of individuals, and the sequence information then can be used for genotyping a large number of individuals in that population. Where genomic trace sequences and overlapping EST sequences are available, those can also be exploited computationally to identify potential SNP markers for genotyping individuals. Although SNP mapping have mostly been limited to *Drosophila* research (Berger *et al.* 2001; Hoskins *et al.* 2001; Kaminker *et al.* 2002) at the moment, these markers have huge potential in other insects to develop SNP genotyping projects, where genome sequences are available. The other system that holds huge promises is the analyses of DNA polymorphisms in microarrays. Also, promising methods are now available to isolate microsatellite sequences (Zane *et al.* 2002) and to genotype them using fluorescent-ISSR-PCR (Nagaraju *et al.* 2002). Application of whole genome microarray represents another new approach to perform genomic DNA analyses in closely related insect species (Butlin & Roper 2005; Turner *et al.* 2005).

After all, DNA sequencing is the ultimate polymorphism detection system and may be that in 20 years' time, the

technology will be so routine that an insect will be placed in a small tube with buffer and homogenized and then some relatively short time later, the entire genome sequence will emerge as data from a sequencing machine, or certainly, large sections of it. I realize that this is at present science fiction, but the fact may not be that far ahead. At the moment, a total of eight insect species are in genome assembly stages and another 35 are in progress for genome sequencing (NCBI genome resources; <http://www.ncbi.nlm.nih.gov/genomes>). These projects will help us to perform genome-wide mutation analyses and to apply high throughput genotyping technologies in insects by using the forthcoming genome sequences. Comparative genomic approaches can be applied to identify highly conserved gene family, transposons, repeat elements, and noncoding conserved regulatory elements in insects that can be exploited to develop new generation of marker systems. Application of such comparative genomic approaches in insects has just begun (Kaufman *et al.* 2002) and genome sequences of *D. melanogaster* and *Anopheles gambiae* along with the forthcoming genome sequences of other important insect species such as *Apis mellifera*, *Bombyx mori*, *Aedes aegypti*, *Drosophila yakuba*, *Drosophila pseudoobscura* will bring into light the many aspects of comparative genomics of these insects very soon. This information can potentially be exploited to develop new generation marker systems based on highly conserved regions in insect genomes and may also provide useful information towards developing generic marker systems such as universal microarrays (Belosludtsev *et al.* 2004) that can probably be readily applied to many related insect species. These technologies in the field of entomology are expected to provide new direction to study insect genomes in an unprecedented way in the years to come.

Acknowledgements

I am grateful to Dr Charles Whitfield of University of Illinois at Urbana-Champaign for his encouragement and support. I am also thankful to Dr Jeffery J. Stuart of Purdue University, West Lafayette, Indiana, for his encouragement and to Drs H. D. Loxdale and Louis Bernatchez for their helpful suggestion in improving the manuscript. I apologize for not being able to include other works on insect molecular markers due to space limitation.

Supplementary material

The supplementary material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC3014/MEC3014sm.htm>

Table S1. A list of insect species and number of available molecular markers

Box S1. A list of transposons and repeat elements in insects

Box S2. Protocol for development of BAC end specific markers

References

- Agusti N, de Vicente MC, Gabarra R (1999) Developing SCAR markers to study predation on *Trialeurodes vaporariorum*. *Insect Molecular Biology*, **9**, 263–268.
- Agusti N, de Vicente MC, Gabarra R (2000) Development of sequence amplified characterized region (SCAR) markers of *Helicoverpa armigera*: a new polymerase chain reaction-based technique for predator gut analysis. *Molecular Ecology*, **8**, 1467–1474.
- Agusti N, Shayler SP, Harwood JD *et al.* (2003) Collembola as alternative prey sustaining spiders in arable ecosystems: prey detection within predators using molecular markers. *Molecular Ecology*, **10**, 3467–3475.
- Akey JM, Sosnoski D, Parra E *et al.* (2001) Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *BioTechniques*, **30**, 358–362.
- Apostol BL, Black WC 4th, Reiter P *et al.* (1996) Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. *Heredity*, **76**, 325–334.
- Archavaleta-Velasco ME, Hunt GJ, Emore C (2003) Quantitative trait loci that influence the expression of guarding and stinging behaviors of individual honeybees. *Behavioral Genetics*, **33**, 357–364.
- Austin JJ, Ross AJ, Smith AB *et al.* (1997) Problems of reproducibility — does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **264**, 467–474.
- Avis JC (2000) *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.
- Avis JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn, pp. 684. Sinauer Associates, Sunderland, Massachusetts.
- Ayres CF, Melo-Santos MA, Sole-Cava AM *et al.* (2003) Genetic differentiation of *Aedes aegypti* (Diptera: Culicidae), the major dengue vector in Brazil. *Journal of Medical Entomology*, **40**, 430–435.
- Ayyadevara S, Thaden JJ, Shmookler RJ (2000) Anchor polymerase chain reaction display: a high-throughput method to resolve, score, and isolate dimorphic genetic markers based on interspersed repetitive DNA elements. *Analytical Biochemistry*, **284**, 19–28.
- Baba S, Kukita Y, Higasa K *et al.* (2003) Single-stranded conformational polymorphism analysis using automated capillary array electrophoresis apparatuses. *Biotechniques*, **34**, 746–750.
- Bachtrog D, Weiss S, Zangerl B *et al.* (1999) Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Molecular Biology and Evolution*, **16**, 602–610.
- Ballinger-Crabtree ME, Black WC 4th, Miller BR (1992) Use of genetic polymorphisms detected by the random-amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) for differentiation and identification of *Aedes aegypti* subspecies and populations. *American Journal of Tropical Medicine and Hygiene*, **47**, 893–901.
- Banuls AL, Jonquieres R, Guerrini F *et al.* (1999) Genetic analysis of leishmania parasites in Ecuador: are *Leishmania* (Viannia) *panamensis* and *Leishmania* (V.) *guyanensis* distinct taxa? *American Journal of Tropical Medicine and Hygiene*, **61**, 838–845.
- Barker G, Batley J, O'Sullivan H *et al.* (2003) Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics*, **19**, 421–422.
- Bartholomay LC, Cho WL, Rocheleau TA *et al.* (2004) Description of the transcriptomes of immune response-activated hemocytes

- from the mosquito vectors *Aedes aegypti* and *Armigeres subalbatus*. *Infection Immunity*, **72**, 4114–4126.
- Bartlett AC (1981) Isozyme polymorphism in populations of the pink bollworm. *Annals of Entomological Society of America*, **74**, 9–13.
- Bartlett AC, Butler GD Jr (1975) Genetic control of the cabbage looper by a recessive lethal mutation. *Journal of Economic Entomology*, **68**, 331–335.
- Bartlett AC, Wilson NM, Mattix EB (1968) The fate of genetic markers in populations of boll weevils. *Journal of Economic Entomology*, **61**, 808–812.
- Beeman RW, Brown SJ (1999) RAPD-based genetic linkage maps of *Tribolium castaneum*. *Genetics*, **153**, 333–338.
- Behura SK, Nair S, Sahu SC *et al.* (2000) An AFLP marker that differentiates biotypes of the Asian rice gall midge (*Orseolia oryzae*, Wood-Mason) is sex-linked and also linked to avirulence. *Molecular and General Genetics*, **263**, 328–334.
- Behura SK, Nair S, Mohan M (2001a) Polymorphisms flanking the *mariner* integration sites in the rice gall midge (*Orseolia oryzae* Wood-Mason) genome are biotype-specific. *Genome*, **44**, 947–954.
- Behura SK, Sahu SC, Mohan M *et al.* (2001b) *Wolbachia* in the Asian rice gall midge, *Orseolia oryzae* (Wood-Mason): correlation between host mitotypes and infection status. *Insect Molecular Biology*, **10**, 163–171.
- Behura SK, Sahu SC, Rajamni S *et al.* (1999) Differentiation of Asian rice gall midge, *Orseolia oryzae* (Wood-Mason), biotypes by sequence characterized amplified regions (SCARs). *Insect Molecular Biology*, **8**, 391–397.
- Behura SK, Valicente FH, Rider SD Jr *et al.* (2004) A physically anchored genetic map and linkage to avirulence reveals recombination suppression over the proximal region of Hessian fly chromosome A2. *Genetics*, **167**, 343–355.
- Belosludtsev YY, Bowerman D, Weil R *et al.* (2004) Organism identification using a genome sequence-independent microarray probe set. *BioTechniques*, **37**, 654–660.
- Bensasson D, Zhang DX, Hewitt GM (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution*, **17**, 406–415.
- Berger J, Suzuki T, Senti KA *et al.* (2001) Genetic mapping with SNP markers in *Drosophila*. *Nature Genetics*, **29**, 475–481.
- Berghella L, Dimitri P (1996) The heterochromatic rolled gene of *Drosophila melanogaster* is extensively polytenized and transcriptionally active in the salivary gland chromocenter. *Genetics*, **144**, 117–125.
- Berlolo M, Fanti L, Sheen F *et al.* (2005) Heterochromatic distribution of HeT-A- and TART-like sequences in several *Drosophila* species. *Cytogenetic and Genome Research*, **110**, 124–133.
- Black WC (1993) PCR with arbitrary primers: approach with care. *Insect Molecular Biology*, **2**, 1–6.
- Bond HA, Craig GB Jr, Fay RW (1970) Field mating and movement of *Aedes aegypti*. *Mosquito News*, **30**, 394–402.
- Bonizzoni M, Katsoyannos BI, Marguerie R *et al.* (2002) Microsatellite analysis reveals remating by wild Mediterranean fruit fly females, *Ceratitis capitata*. *Molecular Ecology*, **11**, 1915–1921.
- Borevitz JO, Liang D, Plouffe D *et al.* (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research*, **13**, 513–523.
- Bosio CF, Fulton RE, Salasek ML *et al.* (2000) Quantitative trait loci that control vector competence for dengue-2 virus in the mosquito *Aedes aegypti*. *Genetics*, **156**, 687–698.
- Bosio CF, Harrington LC, Jones JW *et al.* (2005) Genetic structure of *Aedes aegypti* populations in Thailand using mitochondrial DNA. *American Journal of Tropical Medicine and Hygiene*, **72**, 434–442.
- Bournoville R, Simon JC, Badenhausser I *et al.* (2000) Clones of pea aphid, *Acyrtosiphon pisum* (Hemiptera: aphididae) distinguished using genetic markers, differ in their damaging effect on a resistant alfalfa cultivar. *Bulletin of Entomological Research*, **90**, 33–39.
- Boutla A, Delidakis C, Tabler M (2003) Developmental defects by antisense-mediated inactivation of micro-RNAs 2 and 13 in *Drosophila* and the identification of putative target genes. *Nucleic Acids Research*, **31**, 4973–4980.
- Braendle C, Caillaud MC, Stern DL (2005) Genetic mapping of *aphicarus* – a sex-linked locus controlling a wing polymorphism in the pea aphid (*Acyrtosiphon pisum*). *Heredity*, **94**, 435–442.
- Brandt A, Scharf M, Pedra JH *et al.* (2002) Differential expression and induction of two *Drosophila* cytochrome P450 genes near the *Rst* (2) DDT locus. *Insect Molecular Biology*, **11**, 337–341.
- Breed MD, Guzman-Novoa E, Hunt GJ (2004) Defensive behavior of honeybees: organization, genetics, and comparisons with other bees. *Annual Review of Entomology*, **49**, 271–298.
- Bretman A, Wedell N, Tregenza T (2004) Molecular evidence of post-copulatory inbreeding avoidance in the field cricket *Gryllus bimaculatus*. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **271**, 159–164.
- Brown TA, Brown KA (1994) Ancient DNA: using molecular biology to explore the past. *Bioessays*, **16**, 719–726.
- Butlin R, Roper C (2005) Evolutionary genetics: microarrays and species origins. *Nature*, **437**, 199–201.
- Cabrero J, Bugrov A, Warchalowska-Sliwa E *et al.* (2003) Comparative FISH analysis in five species of *Eyprepocnemidinae* grasshoppers. *Heredity*, **90**, 377–381.
- Cano RJ, Borucki MK (1995) Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. *Science*, **268**, 1060–1064.
- Casa AM, Brouwer C, Nagel A *et al.* (2000) Inaugural article: the MITE family *heartbreaker* (Hbr): molecular markers in maize. *Proceedings of National Academy of Sciences, USA*, **97**, 10083–10089.
- Castrillo LA, Griggs MH, Vandenberg JD (2004) Vegetative compatibility groups in indigenous and mass-released strains of the entomopathogenic fungus *Beauveria bassiana*: likelihood of recombination in the field. *Journal of Invertebrate Pathology*, **86**, 26–37.
- Cervera MT, Cabezas JA, Simon B *et al.* (2000) Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on AFLP analysis. *Bulletin of Entomological Research*, **90**, 391–396.
- Chatterjee SN, Mohandas TP (2003) Identification of ISSR markers associated with productivity traits in silkworm, *Bombyx mori* L. *Genome*, **46**, 438–447.
- Chatterjee SN, Tanushree T (2004) Molecular profiling of silkworm biodiversity in India. *Genetika*, **40**, 1618–1627.
- Che Y, Chen X (2004) A multiplexing single nucleotide polymorphism typing method based on restriction-enzyme-mediated single-base extension and capillary electrophoresis. *Analytical Biochemistry*, **329**, 220–229.
- Chen Y, Giles KL, Payton ME *et al.* (2000) Identifying key cereal aphid predators by molecular gut analysis. *Molecular Ecology*, **9**, 1887–1898.
- Cheng TC, Xia QY, Qian JF *et al.* (2004) Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain Dazao. *Insect Biochemistry and Molecular Biology*, **34**, 523–530.

- Crampton JM, Beard CB, Louis C (1997) *The Molecular Biology of Insect Disease Vectors: A Methods Manual*. Kluwer Academic Publishers, Boston, Massachusetts.
- Crozier RH, Oldroyd BP, Tay WT *et al.* (1997) Molecular advances in understanding social insect population structure. *Electrophoresis*, **18**, 1672–1675.
- Curran S, Hill L, O'Grady G *et al.* (2002) Validation of single nucleotide polymorphism quantification in pooled DNA samples with SNaPIT: a glycosylase-mediated methods for polymorphism detection method. *Molecular Biotechnology*, **22**, 253–262.
- Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504–509.
- Dawkins R (1986) *The Blind Watchmaker*. W.W. Norton & Company, New York.
- Delmotte F, Leterme N, Gauthier JP *et al.* (2002) Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, **11**, 711–723.
- Dempster J, McLean I (1999) *Insect Populations: In theory and in Practice*. Kluwer Academic Publishers, Boston, Massachusetts.
- DeSalle R (1994) Implications of ancient DNA for phylogenetic studies. *Experientia*, **50**, 543–550.
- DeSalle R, Gatesy J, Wheeler W *et al.* (1992) DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science*, **257**, 1933–1936.
- Dong Y, Zhu H (2005) Single-strand conformational polymorphism analysis: basic principles and routine practice. *Methods in Molecular Medicine*, **108**, 149–157.
- Dopman EB, Bogdanowicz SM, Harrison RG (2004) Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). *Genetics*, **167**, 301–309.
- van Eijk MJ, Broekhof JL, van der Poel HJ *et al.* (2004) SNPWave: a flexible multiplexed SNP genotyping technology. *Nucleic Acids Research*, **32**, 47.
- Fan JB, Chen X, Halushka MK *et al.* (2000) Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Research*, **10**, 853–860.
- Favia G, Dimopoulos G, Louis C (1994) Analysis of the *Anopheles gambiae* genome using RAPD markers. *Insect Molecular Biology*, **3**, 149–157.
- Fay RW, Craig GB (1969) Genetically marked *Aedes aegypti* in studies of field populations. *Mosquito News*, **29**, 121–127.
- Fegan M, Manners JM, Maclean DJ *et al.* (1993) Random amplified polymorphic DNA markers reveal a high degree of genetic diversity in the entomopathogenic fungus *Metarhizium anisopliae* var. *anisopliae*. *Journal of General Microbiology*, **139**, 2075–2081.
- Fischer SG, Lerman LS (1983) DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proceedings of National Academy of Sciences, USA*, **80**, 1579–1583.
- Flood EM, Tang F, Horvath MM *et al.* (2002) SNPCEQer: detecting SNPs in sequences generated by the Beckman CEQ2000 DNA Analysis System. *BioTechniques*, **33**, 814–816.
- Freimoser FM, Screen S, Bagga S *et al.* (2003) Expressed sequence tag (EST) analysis of two subspecies of *Metarhizium anisopliae* reveals a plethora of secreted proteins with potential activity in insect hosts. *Microbiology*, **49**, 239–247.
- Fulton RE, Salasek ML, DuTeau NM *et al.* (2001) SSCP analysis of cDNA markers provides a dense linkage map of the *Aedes aegypti* genome. *Genetics*, **158**, 715–726.
- Garcia-Gonzalez F, Nunez Y, Ponz F *et al.* (2003) Sperm competition mechanisms, confidence of paternity, and the evolution of paternal care in the golden egg bug (*Phyllomorpha laciniata*). *International Journal of Organic Evolution*, **57**, 1078–1088.
- Goldstein PZ, Desalle R (2003) Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. *Molecular Ecology*, **12**, 1993–1998.
- Graham DH, Holmes JL, Black WC 4th (2004) Identification of quantitative trait loci affecting sex determination in the Eastern treehole mosquito (*Ochlerotatus triseriatus*). *Journal of Heredity*, **95**, 35–45.
- Groot AT, Ward C, Wang J *et al.* (2004) Introgressing pheromone QTL between species: towards an evolutionary understanding of differentiation in sexual communication. *Journal of Chemical Ecology*, **30**, 2495–2514.
- Guimond N, Bideshi DK, Pinkerton AC *et al.* (2003) Patterns of *Hermes* transposition in *Drosophila melanogaster*. *Molecular Genetics and Genomics*, **268**, 779–790.
- Gunderson KL, Steemers FJ, Lee G *et al.* (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genetics*, **37**, 549–554.
- Hadrys H, Schierwater B, Dellaporta SL *et al.* (1993) Determination of paternity in dragonflies by random amplified polymorphic DNA fingerprinting. *Molecular Ecology*, **2**, 79–87.
- Hale LR, Singh RS (1991) A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history vs. selection in the genetic structure of geographic populations. *Genetics*, **129**, 103–117.
- Hao K, Niu T, Sangokoya C *et al.* (2002) SNPKIT: an efficient approach to systematic evaluation of candidate single nucleotide polymorphisms in public databases. *Biotechniques*, **33**, 822.
- Harper GL, King RA, Dodd CS *et al.* (2005) Rapid screening of invertebrate predators for multiple prey DNA targets. *Molecular Ecology*, **14**, 819–827.
- Harriett B, Kauer M, Schlotterer C (2003) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proceedings of National Academy of Sciences, USA*, **99**, 12949–12954.
- Harris MO, Stuart JJ, Mohan M *et al.* (2003) Grasses and gall midges: plant defense and insect adaptation. *Annual Review of Entomology*, **48**, 549–577.
- Hawthorne DJ (2001) AFLP-based genetic linkage map of the Colorado potato beetle *Leptinotarsa decemlineata*: sex chromosomes and a pyrethroid-resistance candidate gene. *Genetics*, **158**, 695–700.
- Hebert PD, Penton EH, Burns JM *et al.* (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.
- Heckel DG (2003) Genomics in pure and applied entomology. *Annual Review of Entomology*, **48**, 235–260.
- Heckel DG, Gahan LJ, Liu YB *et al.* (1999) Genetic mapping of resistance to *Bacillus thuringiensis* toxins in diamondback moth using biphasic linkage analysis. *Proceedings of National Academy of Sciences, USA*, **96**, 8373–8377.
- Hey J (2001) The mind of the species problem. *Trends in Ecology & Evolution*, **16**, 326–329.
- Hodge KT, Sawyer AJ, Humber RA (1995) RAPD-PCR for identification of *Zoopthora radicans* isolates in biological control of the potato leafhopper. *Journal of Invertebrate Pathology*, **65**, 1–9.
- Hoskins RA, Phan AC, Naeemuddin M *et al.* (2001) SNP markers

- for genetic mapping in *Drosophila melanogaster*. *Genome Research*, **11**, 1100–1113.
- Hoy MA (2003) *Insect Molecular Genetics*, 2nd edn, Academic Press/Elsevier, San Diego, California.
- Hunt GJ, Page RE (1994) Linkage analysis of sex determination in honeybee (*Apis mellifera*). *Molecular and General Genetics*, **244**, 512–518.
- Hunt GJ, Page RE (1995) Linkage map of the honeybee, *Apis mellifera*, based on RAPD markers. *Genetics*, **139**, 1371–1382.
- Hunter WB, Dang PM, Bausher MG *et al.* (2003) Aphid biology: expressed genes from alate *Toxoptera citricida*, the brown citrus aphid. *Journal of Insect Science*, **3**, 23.
- Hurst GD, Graf von der Schulenburg JH, Majerus TM *et al.* (1999) Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. *Insect Molecular Biology*, **8**, 133–139.
- Hurst GD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceeding of Royal Society of London. Series B, Biological Sciences*, **272**, 1525–1534.
- Ikawa T, Okabe H, Hoshizaki S *et al.* (2002) Species composition and distribution of ocean skaters Halobates (Hemiptera: Gerridae) in the western Pacific Ocean. *Entomological Science*, **5**, 1–6.
- James AC, Ballard JW (2000) Expression of cytoplasmic incompatibility in *Drosophila simulans* and its impact on infection frequencies and distribution of *Wolbachia pipientis*. *International Journal of Organic Evolution*, **54**, 1661–1672.
- Jensen AB, Palmer KA, Boomsma JJ *et al.* (2005) Varying degrees of *Apis mellifera* ligustica introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, **14**, 93–106.
- Jeyaprasanth A, Hoy MA (2000) Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. *Insect Molecular Biology*, **9**, 393–405.
- Jiggins FM (2003) Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, **164**, 5–12.
- Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–2523.
- Jones DB, Giles KL, Chen Y *et al.* (2005) Estimation of Hymenopteran parasitism in cereal aphids by using molecular markers. *Journal of Economic Entomology*, **98**, 217–221.
- Junqueira AC, Lessinger AC, Azeredo-Espin AM (2002) Methods for the recovery of mitochondrial DNA sequences from museum specimens of myiasis-causing flies. *Medical and Veterinary Entomology*, **16**, 39–45.
- Kambhampati S, Smith PT (1995) PCR primers for the amplification of four insect mitochondrial gene fragments. *Insect Molecular Biology*, **4**, 233–236.
- Kaminker JS, Bergman CM, Kronmiller B *et al.* (2002) The transposable elements of *Drosophila melanogaster* – a genomics perspective. *Genome Biology*, **3**, 0084.1–0084.0084.20.
- Kandpal RP, Kandpal G, Weissman SM (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proceedings of National Academy of Sciences, USA*, **91**, 88–92.
- Kauer MO, Dieringer D, Schlotterer C (2003) A microsatellite variability screen for positive selection associated with the 'out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics*, **165**, 1137–1148.
- Kaufman TC, Severson DW, Robinson GE (2002) The *Anopheles* genome and comparative insect genomics. *Science*, **298**, 97–98.
- Kengne P, Trung HD, Baimai V *et al.* (2001) A multiplex PCR-based method derived from random amplified polymorphic DNA (RAPD) markers for the identification of species of the *Anopheles minimus* group in Southeast Asia. *Insect Molecular Biology*, **10**, 427–435.
- Kentner EK, Arnold ML, Wessler SR (2003) Characterization of high-copy-number retrotransposons from the large genomes of the *Louisiana iris* species and their use as molecular markers. *Genetics*, **164**, 685–697.
- Kethidi DR, Roden DB, Ladd TR *et al.* (2003) Development of SCAR markers for the DNA-based detection of the Asian long-horned beetle, *Anoplophora glabripennis* (Motschulsky). *Archives of Insect Biochemistry and Physiology*, **52**, 193–204.
- Knight ME, Martin AP, Bishop S *et al.* (2005) An interspecific comparison of foraging range and nest density of four bumblebee (*Bombus*) species. *Molecular Ecology*, **14**, 1811–1820.
- Knipple DC, Doyle KE, Marsella-Herrick PA *et al.* (1994) Tight genetic linkage between the *kdr* insecticide resistance trait and a voltage-sensitive sodium channel gene in the house fly. *Proceedings of the National Academy of Sciences, USA*, **91**, 2483–2487.
- Kokoris M, Dix K, Moynihan K *et al.* (2000) High-throughput SNP genotyping with the Masscode system. *Molecular Diagnostics*, **5**, 329–340.
- Krebs S, Medugorac I, Seichter D *et al.* (2003) RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. *Nucleic Acids Research*, **31**, e37.
- Kurg A, Tonisson N, Georgiou I *et al.* (2000) Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genetical Testing*, **4**, 1–7.
- Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. *Annual Review of Genomics and Human Genetics*, **2**, 235–258.
- Laayouni H, Santos M, Fontdevila A (2005) Toward a physical map of *Drosophila buzzatii*: Use of randomly amplified polymorphic DNA polymorphisms and sequence-tagged site landmarks. *Genetics*, **156**, 1797–1816.
- Lanave C, Licciulli F, De Robertis M *et al.* (2002) Update of AMmtDB: a database of multi-aligned Metazoa mitochondrial DNA sequences. *Nucleic Acids Research*, **30**, 174–175.
- Ledford M, Friedman KD, Hessner MJ *et al.* (2000) A multi-site study for detection of the factor V (Leiden) mutation from genomic DNA using a homogeneous invader microtiter plate fluorescence resonance energy transfer (FRET) assay. *Journal of Molecular Diagnostics*, **2**, 97–104.
- Lee JK, Kwon SJ, Park KC *et al.* (2005) Isaac-CACTA transposons: new genetic markers in maize and sorghum. *Genome*, **48**, 455–460.
- Lewis KG, Liewlaksaneeyanawin C, Alfaro RI *et al.* (2002) Sexual reproduction in the white pine weevil (*Pissodes strobi* [Peck] [Coleoptera: Curculionidae]): implications for population genetic diversity. *Journal of Heredity*, **93**, 165–169.
- Li YC, Korol AB, Fahima T *et al.* (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, **11**, 2453–2465.
- Li YC, Korol AB, Fahima T *et al.* (2004) Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution*, **21**, 991–1007.
- Lin-Ling C, Zhang J, Sommer SS *et al.* (2005) Single-base discrimination mediated by proofreading inert allele specific primers. *Journal of Biochemical and Molecular Biology*, **38**, 24–27.
- Livak KJ (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genetical Analysis*, **14**, 143–149.

- Llewellyn KS, Loxdale HD, Harrington R *et al.* (2003) Migration and genetic structure of the grain aphid (*Sitobion avenae*) in Britain related to climate and clonal fluctuation as revealed using microsatellites. *Molecular Ecology*, **12**, 21–34.
- Lobo NF, Behura SK, Aggarwal R *et al.* (2006) Genomic analysis of a 1 Mb region near the telomere of Hessian fly chromosome X2 and avirulence gene *vH13*. *BMC Genomics*, **7**, 7.
- Loukas M, Krimbas CB, Mavragani-Tsipidou P *et al.* (1979) Genetics of *Drosophila subobscura* populations. VIII. Allozyme loci and their chromosome maps. *Journal of Heredity*, **70**, 17–26.
- Loxdale HD, Lushai G (1998) Molecular markers in entomology. *Bulletin of Entomological Research*, **88**, 577–600.
- Loxdale HD, Tarr IJ, Weber CP *et al.* (1985) Electrophoretic study of enzymes from cereal aphid populations. III. Spatial and temporal genetic variation of populations of *Sitobion avenae* (F.) (Hemiptera: Aphididae). *Bulletin of Entomological Research*, **75**, 121–141.
- Luckinbill LS, Golenberg EM (2002) Genes affecting aging: mapping quantitative trait loci in *Drosophila melanogaster* using amplified fragment length polymorphisms (AFLPs). *Genetica*, **114**, 147–156.
- Lunt DH, Hutchinson WF, Carvalho GR (1999) An efficient method for PCR-based identification of microsatellite arrays (PIMA). *Molecular Ecology*, **8**, 893–894.
- Lunt DH, Zhang DX, Szymura JM *et al.* (1996) The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology*, **5**, 153–165.
- Luque C, Legal L, Staudter H *et al.* (2002) ISSR (inter simple sequence repeats) as genetic markers in *noctuids* (Lepidoptera). *Heredity*, **136**, 251–253.
- Lushai G, Loxdale HD, Brookes CP *et al.* (1997) Genotypic variation among different phenotypes within aphid clones. *Proceedings of Royal Society of London. Series B, Biological Sciences*, **264**, 725–730.
- Lushai G, Markovitch O, Loxdale HD (2002) Host-based genotype variation in insects revisited. *Bulletin of Entomological Research*, **92**, 159–164.
- Lyamichev V, Mast AL, Hall JG *et al.* (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology*, **17**, 292–296.
- Maa WCJ, Terriere LC (1983) Age-dependent variation in enzymatic and electrophoretic properties of house fly (*M. domestica*) carboxylesterases. *Comparative Biochemistry and Physiology*, **74C**, 461–467.
- Manaster C, Zheng W, Teuber M *et al.* (2005) INSNP: a tool for automated detection and visualization of SNPs and InDels. *Human Mutation*, **26**, 11–19.
- Manguin S, Kengne P, Sonnier L *et al.* (2002) SCAR markers and multiplex PCR-based identification of isomorphic species in the *Anopheles dirus* complex in Southeast Asia. *Medical and Veterinary Entomology*, **16**, 46–54.
- Margonari CS, Fortes-Dias CL, Dias ES (2004) Genetic variability in geographical populations of *Lutzomyia whitmani* elucidated by RAPD-PCR. *Journal of Medical Entomology*, **41**, 187–192.
- Marth GT, Korf I, Yandell MD *et al.* (1999) A general approach to single-nucleotide polymorphism discovery. *Nature Genetics*, **23**, 452–456.
- Martin SG, Dobi KC, St Johnston D (2001) A rapid method to map mutations in *Drosophila*. *Genome Biology*, **2**, research0036.1–research0036.12.
- Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: *Species: the Units of Biodiversity* (eds Claridge MF, Dawah HA, Wilson MR), pp. 381–424. Chapman & Hall, London.
- Mendelson TC, Shaw KL (2005) Use of AFLP markers in surveys of arthropod diversity. *Methods in Enzymology*, **395**, 161–177.
- Miller NJ, Kift NB, Tatchell GM (2005) Host-associated populations in the lettuce root aphid, *Pemphigus bursarius* (L.). *Heredity*, **94**, 556–564.
- Morlais I, Poncon N, Simard F *et al.* (2004) Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. *American Journal of Tropical Medicine and Hygiene*, **71**, 795–802.
- Nagaraju J, Kathirvel M, Subbaiah EV *et al.* (2002) FISSR-PCR: a simple and sensitive assay for high throughput genotyping and genetic mapping. *Molecular Cellular Probes*, **16**, 67–72.
- Nagaraju J, Sharma A, Sethuraman BN *et al.* (1995) DNA fingerprinting in silkworm *Bombyx mori* using banded krait minor satellite DNA-derived probe. *Electrophoresis*, **16**, 1639–1642.
- Nairz K, Stocker H, Schindelholz B *et al.* (2002) High-resolution SNP mapping by denaturing HPLC. *Proceedings of National Academy of Sciences, USA*, **99**, 10575–10580.
- Nakabachi A, Shigenobu S, Sakazume N *et al.* (2005) Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera. *Proceedings of National Academy of Sciences, USA*, **102**, 5477–5482.
- Newton CR, Heptinstall LE, Summers C *et al.* (1989) Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. *Lancet*, **2**, 1481–1483.
- Nishimori Y, Lee JM, Sumitani M *et al.* (2000) A linkage map of the turnip sawfly *Athalia rosae* (Hymenoptera: Symphyta) based on random amplified polymorphic DNAs. *Genes and Genetics Systems*, **75**, 159–166.
- Niu T, Hu Z (2004) SNPICKER: a graphical tool for primer picking in designing mutagenic endonuclease restriction assays. *Bioinformatics*, **20**, 3263–3265.
- Oldenburg MC, Siebert M (2000) New cleavage fragment length polymorphism method improves the mutation detection assay. *BioTechniques*, **28**, 351–357.
- Orita M, Iwahana H, Kanazawa H *et al.* (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences, USA*, **8**, 2766–2770.
- Ostrander EA, Jong PM, Rine J *et al.* (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of National Academy of Sciences, USA*, **89**, 3419–3423.
- Parsons YM, Shaw KL (2002) Mapping unexplored genomes: a genetic linkage map of the Hawaiian cricket *Laupala*. *Genetics*, **162**, 1275–1282.
- Pearcy M, Aron S, Doums C *et al.* (2004) Conditional use of sex and parthenogenesis for worker and queen production in ants. *Science*, **306**, 1780–1783.
- Pedra JH, Brandt A, Westerman R *et al.* (2003) Transcriptome analysis of the cowpea weevil bruchid: identification of putative proteinases and alpha-amylases associated with food breakdown. *Insect Molecular Biology*, **12**, 405–412.
- Pinto AS, de Lana M, Bastrenta B *et al.* (1998) Compared vectorial transmissibility of pure and mixed clonal genotypes of *Trypanosoma cruzi* in *Triatoma Infestans*. *Parasitology Research*, **84**, 348–353.

- Prasad MD, Muthulakshmi M, Madhu M *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.
- Qi X, Bakht S, Devos KM *et al.* (2001) L-RCA (ligation-rolling circle amplification): a general method for genotyping of single nucleotide polymorphisms (SNPs). *Nucleic Acids Research*, **29**, E116.
- Ranson H, Jensen B, Wang X *et al.* (2000) Genetic mapping of two loci affecting DDT resistance in the malaria vector *Anopheles gambiae*. *Insect Molecular Biology*, **9**, 499–507.
- Rauch N, Nauen R (2003) Identification of biochemical markers linked to neonicotinoid cross resistance in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Archives of Insect Biochemistry and Physiology*, **54**, 165–176.
- Reineke A, Karlovsky P, Zebitz CP (1999) Suppression of randomly primed polymerase chain reaction products (random amplified polymorphic DNA) in heterozygous diploids. *Molecular Ecology*, **8**, 1449–1455.
- Reineke A, Lobmann S (2005) Gene expression changes in *Ephesia kuehniella* caterpillars after parasitization by the endoparasitic wasp *Venturia canescens* analyzed through cDNA-AFLPs. *Journal of Insect Physiology*, **51**, 923–932.
- Reineke A, Schmidt O, Zebitz CP (2003) Differential gene expression in two strains of the endoparasitic wasp *Venturia canescens* identified by cDNA-amplified fragment length polymorphism analysis. *Molecular Ecology*, **12**, 3485–3492.
- Richardson BJ, Baverstock PR, Adams M (1986) Allozyme electrophoresis: a handbook for animal systematics and population studies. London Academic Press.
- Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution*, **21**, 1081–1084.
- Rider SD Jr, Sun W, Ratcliffe RH *et al.* (2002) Chromosome landing near avirulence gene *vH13* in the Hessian fly. *Genome*, **45**, 812–822.
- Rigaud T, Bouchon D, Souty-Grosset C *et al.* (1999) Mitochondrial DNA polymorphism, sex ratio distorters and population genetics in the isopod *Armadillidium vulgare*. *Genetics*, **152**, 1669–1677.
- Robertson HM (1993) The *mariner* transposable element is widespread in insects. *Nature*, **362**, 241–245.
- Robinson GE, Grozinger CM, Whitfield CW (2005) Sociogenomics: social life in molecular terms. *Nature Reviews Genetics*, **6**, 257–270.
- Roehrdanz RL (1993) An improved primer for PCR amplification of mitochondrial DNA in a variety of insect species. *Insect Molecular Biology*, **2**, 89–91.
- Rousset F, Solignac M (1995) Evolution of single and double *Wolbachia* symbioses during speciation in the *Drosophila simulans* complex. *Proceedings of the National Academy of Sciences, USA*, **92**, 6389–6393.
- Ruppell O, Pankiw T, Page RE (2004) Pleiotropy, epistasis and new QTL: the genetic architecture of honeybee foraging behavior. *Journal of Heredity*, **95**, 481–491.
- Saiki RK, Bugawan TL, Horn GT *et al.* (1986) Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature*, **324**, 163–166.
- Salvato P, Battisti A, Concato S *et al.* (2002) Genetic differentiation in the winter pine processionary moth (*Thaumetopoea pityocampa-wilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers. *Molecular Ecology*, **11**, 2435–2444.
- Santolamazza F, Della TA, Caccone A (2004) A new polymerase chain reaction–restriction fragment length polymorphism method to identify *Anopheles arabiensis* from *An. gambiae* and its two molecular forms from degraded DNA templates or museum samples. *American Journal of Tropical Medicine and Hygiene*, **70**, 604–606.
- Sauer S, Gelfand DH, Boussicault F *et al.* (2000) Facile method for automated genotyping of single nucleotide polymorphisms by mass spectrometry. *Nucleic Acids Research*, **30**, e22.
- Schluppius DI, Cheng Q, Reilly PE *et al.* (2002) Genetic linkage analysis of the lesser grain borer *Rhyzopertha dominica* identifies two loci that confer high-level resistance to the fumigant phosphine. *Genetics*, **161**, 773–782.
- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**, 365–371.
- Schlötterer C (2001) Genealogical inference of closely related species based on microsatellites. *Genetical Research*, **78**, 209–212.
- Schmalenberger A, Tebbe CC (2003) Bacterial diversity in maize rhizospheres: conclusions on the use of genetic profiles based on PCR-amplified partial small subunit rRNA genes in ecological studies. *Molecular Ecology*, **12**, 251–262.
- Schmid KJ, Tautz D (1999) A comparison of homologous developmental genes from *Drosophila* and *Tribolium* reveals major differences in length and trinucleotide repeat content. *Journal of Molecular Evolution*, **49**, 558–566.
- Schofl G, Schlötterer C (2004) Patterns of microsatellite variability among X chromosomes and autosomes indicate a high frequency of beneficial mutations in non-African *D. simulans*. *Molecular Biology and Evolution*, **21**, 1384–1390.
- Schulte SJ, Rider SD Jr, Hatchett JH *et al.* (1999) Molecular genetic mapping of three X-linked avirulence genes, *vH6*, *vH9* and *vH13*, in the Hessian fly. *Genome*, **42**, 821–828.
- Severson DW, Brown SE, Knudson DL (2001) Genetic and physical mapping in mosquitoes: molecular approaches. *Annual Review of Entomology*, **46**, 183–219.
- Severson DW, Meece JK, Lovin DD *et al.* (2002) Linkage map organization of expressed sequence tags and sequence tagged sites in the mosquito, *Aedes aegypti*. *Insect Molecular Biology*, **11**, 371–378.
- Severson DW, Thathy V, Mori A *et al.* (1995) Restriction fragment length polymorphism mapping of quantitative trait loci for malaria parasite susceptibility in the mosquito *Aedes aegypti*. *Genetics*, **139**, 1711–1717.
- Shapero MH, Leuther KK, Nguyen A *et al.* (2001) SNP genotyping by multiplexed solid-phase amplification and fluorescent minisequencing. *Genome Research*, **11**, 1926–1934.
- Sharakhov IV, Serazin AC, Grushko OG *et al.* (2002) Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science*, **298**, 182–185.
- Simmons LW, Achmann R (2000) Microsatellite analysis of sperm-use patterns in the bush cricket *Requena verticalis*. *International Journal of Organic Evolution*, **54**, 942–952.
- Simon JC, Carre S, Boutin M *et al.* (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 1703–1712.
- Simon JC, Leterme N, Latorre A (1999) Molecular markers linked to breeding system differences in segregating and natural populations of the cereal aphid *Rhopalosiphum padi* L. *Molecular Ecology*, **8**, 965–973.
- Singh L, Phillips C, Jones KW (1984) The conserved nucleotide sequences of *Bkm*, which define *Sxr* in the mouse, are transcribed. *Cell*, **36**, 111–120.
- Solignac M, Cornuet JM, Vautrin D *et al.* (2005) The invasive Korea and Japan types of *Varroa destructor*, ectoparasitic mites of the western honeybee (*Apis mellifera*), are two partly isolated clones.

- Proceedings of the Royal Society of London. Series B, Biological Sciences*, **272**, 411–419.
- Solignac M, Vautrin D, Baudry E *et al.* (2004) A microsatellite-based linkage map of the honeybee, *Apis mellifera* L. *Genetics*, **167**, 253–262.
- Solomon MG, Fitzgerald JD, Murray RA (1996) Electrophoretic approaches to predator–prey interactions. In: *The Ecology of Agricultural Pests: Biochemical Approaches* (eds Symondson WOC, Liddell JE), pp. 457–468. Chapman & Hall, London.
- Souames S, Bonnivard E, Bazin C *et al.* (2003) High mutation rate of TPE repeats: a microsatellite in the putative transposase of the *hobo* element in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **20**, 1826–1832.
- Speight MR, Watt A, Hunter M (2005) *Ecology of Insects: Concepts and Applications*, 2nd edn, Blackwell Science, London.
- Staten R, Schully SD, Noor MA (2004) A microsatellite linkage map of *Drosophila mojavensis*. *BMC Genetics*, **5**, 12.
- Steiner WWM, Joslyn DJ (1979) Electrophoretic techniques for the genetic study of mosquitoes. *Mosquito News*, **39**, 35–54.
- Stuart JJ, Schulte SJ, Hall PS *et al.* (1998) Genetic mapping of Hessian fly avirulence gene *vH6* using bulked segregant analysis. *Genome*, **41**, 702–708.
- Sudeep AB, Khushiramani R, Athawale SS *et al.* (2005) Characterization of a newly established potato tuber moth (*Plutheirinaea operculella* Zeller) cell line. *Indian Journal of Medical Research*, **121**, 159–163.
- Takami Y, Koshio C, Ishii M *et al.* (2004) Genetic diversity and structure of urban populations of *Pieris* butterflies assessed using amplified fragment length polymorphism. *Molecular Ecology*, **13**, 245–258.
- Tan YD, Wan C, Zhu Y *et al.* (2001) An amplified fragment length polymorphism map of the silkworm. *Genetics*, **157**, 1277–1284.
- Tares S, Cornuet JM, Abad P (1993) Characterization of an unusually conserved *AluI* highly reiterated DNA sequence family from the honeybee, *Apis mellifera*. *Genetics*, **134**, 1195–1204.
- Teeter K, Naeemuddin M, Gasperini R *et al.* (2000) Haplotype dimorphism in a SNP collection from *Drosophila melanogaster*. *Journal of Experimental Zoology*, **288**, 63–75.
- Tigano-Milani MS, Honeycutt RJ, Lacey LA *et al.* (1995) Genetic variability of *Paecilomyces fumosoroseus* isolates revealed by molecular markers. *Journal of Invertebrate Pathology*, **65**, 274–282.
- Tobler A, Kapan D, Flanagan NS *et al.* (2005) First-generation linkage map of the warningly colored butterfly *Heliconius erato*. *Heredity*, **94**, 408–417.
- Torti C, Gomulski LM, Bonizzoni M *et al.* (2005) *Cchobo*, a hobo-related sequence in *Ceratitidis capitata*. *Genetica*, **123**, 313–325.
- Trontti K, Aron S, Sundstrom L (2005) Inbreeding and kinship in the ant *Plagiolepis pygmaea*. *Molecular Ecology*, **14**, 2007–2015.
- Tsuchihashi Z, Dracopoli NC (2002) Progress in high throughput SNP genotyping methods. *Pharmacogenomics Journal*, **2**, 103–110.
- Tu Z (1997) Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proceedings of National Academy of Sciences, USA*, **94**, 7475–7480.
- Turelli M, Hoffmann AA, McKechnie SW (1992) Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics*, **132**, 713–723.
- Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.
- Ullmann AJ, Piesman J, Dolan MC *et al.* (2003) A preliminary linkage map of the hard tick, *Ixodes scapularis*. *Insect Molecular Biology*, **12**, 201–210.
- Underhill PA, Jin L, Lin AA *et al.* (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research*, **7**, 996–1005.
- Unneberg P, Stromberg M, Sterky F (2005) SNP discovery using advanced algorithms and neural networks. *Bioinformatics*, **21**, 2528–2530.
- Vavre F, de Jong JH, Stouthamer R (2004) Cytogenetic mechanism and genetic consequences of thelytoky in the wasp *Trichogramma cacoeciae*. *Heredity*, **93**, 592–596.
- Vitkova M, Kral J, Traut W *et al.* (2005) The evolutionary origin of insect telomeric repeats, (TTAGG)_n. *Chromosome Research*, **13**, 145–156.
- Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.
- Wada Y, Mitsumori K, Terachi T *et al.* (1999) Measurement of polymorphic trinucleotide repeats in the androgen receptor gene by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, **34**, 885–888.
- Walden KK, Robertson HM (1997) Ancient DNA from amber fossil bees? *Molecular Biology and Evolution*, **14**, 1075–1077.
- Walton MP, Powell W, Loxdale HD *et al.* (1990) Electrophoresis as a tool for estimating levels of hymenopterous parasitism in field populations of the cereal aphid, *Sitobion avenae*. *Entomologia Experimentalis et Applicata*, **54**, 271–279.
- Wang B, Porter AH (2004) An AFLP-based interspecific linkage map of sympatric, hybridizing *Colias* butterflies. *Genetics*, **168**, 215–225.
- Waugh R, McLean K, Flavell AJ *et al.* (1997) Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (SSAP). *Molecular and General Genetics*, **253**, 687–694.
- Weckx S, Del-Favero J, Rademakers R *et al.* (2005) NOVO SNP, a novel computational tool for sequence variation discovery. *Genome Research*, **15**, 436–442.
- Werren JH (1997) Biology of *Wolbachia*. *Annual Review of Entomology*, **42**, 587–609.
- Whitfield CW, Band MR, Bonaldo MF *et al.* (2002) Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honeybee. *Genome Research*, **12**, 555–566.
- Whitfield CW, Cziko AM, Robinson GE (2003) Gene expression profiles in the brain predict behavior in individual honeybees. *Science*, **302**, 296–299.
- Widmer A, Cozzolino S, Pellegrino G *et al.* (2000) Molecular analysis of orchid pollinaria and pollinaria-remains found on insects. *Molecular Ecology*, **9**, 1911–1914.
- Wilding N, Mardell SK, Brookes CP *et al.* (1993) The use of polyacrylamide gel electrophoresis of enzymes to identify entomophthorean fungi in aphid hosts. *Journal of Invertebrate Pathology*, **62**, 268–272.
- Wilkerson RC, Parsons TJ, Albright DG *et al.* (1993) Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: Anopheles). *Insect Molecular Biology*, **1**, 205–211.
- Williams JG, Kubelik AR, Livak KJ *et al.* (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**, 6531–6535.
- Winzler EA, Castillo-Davis CI, Oshiro G *et al.* (2003) Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics*, **163**, 79–89.

- Wong A, Forbes MR, Smith ML (2000) Characterization of AFLP markers in damselflies: prevalence of codominant markers and implications for population genetic applications. *Genome*, **44**, 677–684.
- Yasukochi Y (1998) A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. *Genetics*, **150**, 1513–1525.
- Yezerski A, Stevens L, Ametrano J (2003) A genetic linkage map for *Tribolium confusum* based on random amplified polymorphic DNAs and recombinant inbred lines. *Insect Molecular Biology*, **12**, 517–526.
- Yoon CK, Aquadro CF (1994) Mitochondrial DNA variation among the *Drosophila athabasca* semispecies and *Drosophila affinis*. *Journal of Heredity*, **85**, 421–426.
- Yoshido A, Bando H, Yasukochi Y *et al.* (2005) The *Bombyx mori* karyotype and the assignment of linkage groups. *Genetics*, **170**, 675–685.
- Zampicini G, Blinov A, Cervella P *et al.* (2004) Insertional polymorphism of a non-LTR mobile element (NLRCth1) in European populations of *Chironomus riparius* (Diptera, Chironomidae) as detected by transposon insertion display. *Genome*, **47**, 1154–1163.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**, 1–16.
- Zhang DX, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, **11**, 247–251.
- Zhang DX, Hewitt GM (1997) Assessment of the universality and utility of a set of conserved mitochondrial COI primers in insects. *Insect Molecular Biology*, **6**, 143–150.
- Zhang J, Li K, Pardinas JR *et al.* (2005) Proofreading genotyping assays mediated by high fidelity exo+ DNA polymerases. *Trends in Biotechnology*, **23**, 92–96.
- Zhang R, Zhu Z, Zhu H *et al.* (2005) SNP CUTTER: a comprehensive tool for SNP PCR-RFLP assay design. *Nucleic Acids Research*, **1**, 33.
- Zhong D, Pai A, Yan G (2004) AFLP-based genetic linkage map for the red flour beetle (*Tribolium castaneum*). *Journal of Heredity*, **95**, 53–61.
- Zhong D, Pai A, Yan G (2005) Costly resistance to parasitism: evidence from simultaneous quantitative trait loci mapping for resistance and fitness in *Tribolium castaneum*. *Genetics*, **169**, 2127–2135.
- Zhou GH, Gotou M, Kajiyama T *et al.* (2005) Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATI). *Nucleic Acids Research*, **33**, 133.
- Zhou X, Scharf ME, Parimi S *et al.* (2002) Diagnostic assays based on esterase-mediated resistance mechanisms in western corn rootworms (Coleoptera: Chrysomelidae). *Journal of Economic Entomology*, **95**, 1261–1266.

Susanta K. Behura works on molecular tagging and mapping of avirulence genes in agricultural insect pests such as rice gall midge and Hessian fly. Currently, he is involved in SNP mapping projects of honeybee genome towards understanding population structure and the spread of killer bees in the United States. He is employing large scale SNP genotyping methods along with microarray expressions of brain genes for eQTL mapping of behavioral phenotypes. He is also generating expression profiles of microRNAs in developing bee brain and to study their possible role in age dependent behavioral plasticity (interested in foraging behavior).
