Microsatellites in plants: A new class of molecular markers

P. K. Gupta, H. S. Balyan, P. C. Sharma and B. Ramesh

Department of Agricultural Botany, Ch. Charan Singh University, Meerut 250 004, India

The present status of research on microsatellites, also called simple sequence repeats (SSRs) or short tandem repeats (STRs) in higher plants has been reviewed. Results on the occurrence, distribution and the level of polymorphism of different microsatellites examined or searched from DNA sequence databases for different plant materials have been tabulated and briefly described. Information on number of loci for each SSR and the number of alleles at each locus, wherever available, has also been included. Methods used for developing microsatellite markers have been briefly discussed. Achievements made and the future possibilities for the use of microsatellites in areas like selection and diagnostics in segregating populations, genome selection during gene introgression (in backcross programmes), genome mapping, gene tagging, cultivar identification, germplasm characterization, estimation of genetic relatedness, etc. have also been reviewed and critically discussed. The addition of microsatellite markers to the repertoire of other available molecular markers should prove very useful in a variety of research areas involving plant systems.

MICROSATELLITES are tandem repeats of DNA sequences of only a few base pairs (1–6 bp) in length, the most abundant being the dinucleotide repeats. The term microsatellite was introduced by Litt and Luty[1] to characterize the simple sequence stretches amplified by polymerase chain reaction (PCR). These are also known as short tandem repeats (STRs)[2] or simple sequence repeats (SSRs)[3] and differ from minisatellites (often called VNTRs[4]), which are repeated sequences having repeat units ranging from 11 to 60 bp in length. The minisatellites were first reported by Jeffreys *et al.*[5], though their utility through PCR was suggested later[6]. The microsatellites are randomly and more evenly dispersed in the genome[7] than minisatellites, which are generally confined to telomeres. A dinucleotide like $(CA)_n$ occurs in human genome, as many as 50,000 times, with n ranging from 10 to 60 (ref. 8). The tri- and tetra-nucleotide repeats are also common in human genome.

The DNA sequences flanking SSRs are known to be conserved in the same manner as those flanking minisatellites (VNTRs). These conserved sequences have been used for designing suitable primers for amplification of the SSR loci using PCR. Any such primer or a pair of primers, when used to amplify a particular SSR locus in a number of genotypes, will reveal SSR polymorphism, in the form of differences in length of the amplified product, each length representing an allele at that locus. The length differences are attributed to the variation in the number of repeat units at a particular SSR locus, possibly caused by slippage during replication[9]. The initial experiments with microsatellites revealed their following attributes: (i) these markers are codominant, like most RFLPs, (ii) many alleles exist in a population and the level of heterozygosity is extremely high, and (iii) the markers are inherited in Mendelian fashion and thus can be used for linkage analysis. The microsatellites or SSRs are being used and will certainly have increased use in future in our efforts to map genomes, to quantify genetic diversity and to characterize accessions in plant germplasm collections[10].

Occurrence and distribution

Microsatellites are an important class of DNA markers because of their abundance and length hypervariability. They occur frequently and randomly in all eukaryotic DNAs examined[1,11–13] and represent a vast source of highly informative markers[7,8,11,12,14,15]. Microsatellites have been found and used for genetic analysis in many a mammalian species[16–19] and to a lesser extent in other eukaryotes, e.g. insects[20,21], birds[19], fish[22], mouse[23], cattle[24] and plants[14,15,25–41]. The data-base search revealed that the relative abundance of different microsatellite motifs in plants and animals differs considerably[28,29]. For example, $(CA)_n$ repeat is one of the most frequently occurring microsatellites (several tens of thousands of copies) in human and many mammals[10,12,18], but is comparatively less frequent in plants[28]. In contrast, $(AT)_n$ microsatellites are the most abundant dinucleotide repeats in plants. Further, greater abundance of $(GA)_n$ repeats than $(CA)_n$ repeats appears to be a consistent feature of plant genomes[26,42]. Trinucleotides and tetranucleotide repeats are also found in plant genomes, the most frequent of them being $(AAG)_n$ and $(AAT)_n$ (refs 26,28,29,43,44). Search was also made in GenBank™ by Sarkar *et al.*[45] for purine/pyrimidine repeats greater than 13 units in length. Their calculations suggested

that for every 100 kb of DNA, on an average there were 2–3 such SSR loci in primates and 1.8 loci in yeast. Condit and Hubbell[25] screened DNA libraries of five tropical tree species and *Zea mays* for the presence of $(AC)_n$ and $(AG)_n$, which together ranged from 5×10^3 to 3×10^5 blocks per genome among the six species examined. Search was also made by Cregan[46] for di-, tri- and tetrameric repeats and on an average a frequency of 0.224 microsatellites per 100 kb of DNA was found (Table 1) in ten plant species including yeast. The potential of microsatellite sequences as genetic markers in hexaploid wheat (*Triticum aestivum*) with respect to their abundance, variation, chromosomal location and usefulness in related species was investigated by Roder *et al.*[40]. The total number of $(GA)_n$ blocks was estimated to be 3.6×10^4 and the number of $(GT)_n$ blocks was estimated to be 2.3×10^4 per haploid wheat genome. Plaschke *et al.*[41] utilized 23 wheat microsatellite markers for the detection of genetic diversity among 40 wheat cultivars/lines and found that these 23 microsatellites, represented by a total of 142 alleles, were located on 15 different chromosomes. The number of alleles at one locus ranged from 3 to 16 with an average of 6.2 alleles, while the average dinucleotide repeat number ranged from 13 to 41. It was shown that the distribution of SSR loci over the different chromosomes and chromosome arms was random and that the highest proportion of microsatellites occurred on the B genome[41]. Table 2 gives information on the distribution of microsatellites in some of the plant materials. Distribution of $(GGC)_n$ microsatellite in a wide variety of eukaryotic genomes was also reported by Zhao and Kochert[15]. This microsatellite is more abundant in monocots (rice, maize, bamboo, wild grasses) than in dicots (peanuts, alfalfa).

Recently, Rubinsztein *et al.*[47], on the basis of a study of allele length distributions for 42 microsatellites in humans and their related primate species, reported a highly significant trend for the loci to be longer in humans than in other primates suggesting that microsatellites can evolve directionally and at different rates in closely related species.

Terminology

Weber[7] categorized microsatellite arrays mainly as 'perfect repeats' (without interruptions), 'imperfect repeats' (interrupted by non-repeat bases) and 'compound repeats' (two or more repeat runs present adjacent to each other) (Table 3). Smith and Devey[37] also observed in *Pinus radiata*, simple sequence repeats pertaining to these three categories and those having other constitutions like perfect + compound (perfect), and compound (perfect) + imperfect.

Level of polymorphism

Variation in the number of tandemly repeated core sequence of nucleotides at a SSR locus among different genotypes provides the basis for polymorphism that can be used in plant genetic studies[25,46]. Recent reports indicate that SSR loci for a number of core repeat units are highly polymorphic between species, and more importantly, between individuals within species and populations[26,36]. There is also a good linear relationship between the number of alleles detected at a locus and the length of the microsatellite array. Thus the larger the repeat number in a microsatellite, the larger is the number of alleles detected[33]. Microsatellites are abundant and highly polymorphic in plant species including Triticeae such as wheat[40] and barley[36]. In wheat, microsatellites are relatively long containing up to 40 nucleotide repeats[40].

In humans also, informativeness of microsatellites tend to increase with the increase in number of repeats, so

Table 1. The frequencies and average distances between all possible dimeric, trimeric and tetrameric simple sequence repeat DNA sequences in plant species determined from a search of GenBank™ (from Cregan[46])

Plant species	Kilobases searched (kbp)	Dimeric repeats*		Trimeric repeats**		Tetrameric repeats***	
		No.	Distance between repeats (kbp)	No.	Distance between repeats (kbp)	No.	Distance between repeats (kbp)
Saccharomyces cervisiae	2288	40	57	29	79	2	1144
Nicotiana tabacum	118	4	29	0	—	0	—
Glycine max	212	6	35	1	212	4	53
Lycopersicon esculentum	135	2	68	0	—	1	135
Triticum aestivum	151	1	151	43	4	0	—
Medicago sativa	30	0	—	1	30	0	—
Pisum sativum	129	3	43	3	43	2	64
Zea mays	368	3	123	4	91	6	61
Arabidopsis thaliana	247	4	62	1	247	0	—
Oryza sativa	137	2	68	2	68	6	23

*Dimeric repeats such as $(AT)_n$ with $n > 9$.
**Trimeric repeats such as $(ATT)_n$ with $n > 7$.
***Tetrameric repeats such as $(AGAT)_n$ with $n > 4$.

Table 2. Occurrence and polymorphism of microsatellites in plants

Crop and species	No. of genotypes examined	Repeat unit	No. of loci	No. of alleles at one or more loci	Ref.
Soybean (*Glycine max*; *G. soja*)	43	$(AT)_n$	2	14	26
		$(ATT)_n$	1	7	
		$(CA/GT)_n$	3		
		$(AT)_4$	—	1	
		$(ATT)_7$	—	1	
Glycine max	10	$(AT)_n$	1	8	29
	07	$(TAT)_n$	1	7	
Glycine max; *Glycine soja*	12–61	$(AT)_n$	3	23	85*
		$(CT)_{16}$	1	4	
		$(AT)_9(AAT)_6$	1	4	
		$(TA)_{10}$	1	7	
		$(TAT)_{20}$	1	7	
Glycine max	—		1	23	86
	96	$(AT)_n$	3	21–26	79
		$(ATT)_n$	4	11–19	
Phaseolus (18 species)	90	$(CA)_8$	Occurs in all species		83
		$(CAC)_5$			
		$(GATA)_4$			
		$(GACA)_4$	Occurs in 13 species		
Grapevine (*Vitis* sp.)	26	Imperfect $(A)_n$	1	8	30
		Imperfect $(TAA)_n$			
		Perfect $(GA)_n$	1	13	
		$(GTAT)_n(GT)_n$	1	4	
		Perfect $(AG)_n$	1	9	
		Perfect $(GT)_n$	1	12	
Arabidopsis thaliana	6	$(AT)_n$	2	9	42*
		$(A)_n$	2	9	
		$(CA)_n$	1	3	
		$(AG)_n$	6	23	
		$(GA)_n$	14	57	
		$(CT)_n$	4	19	
		$(TC)_n$	1	3	
Rapeseed (*Brassica napus* L.)	8	$(CA)_8$, $(CT)_8$, $(GTG)_5$, $(GACA)_4$, $(GGAT)_4$, $(GATA)_4$	—	—	81
		$(GA)_n$	1 SSR/100 kb DNA		87
		$(CA)_n$	1 SSR/400 kb DNA		
		$(GATA)_n$	1 SSR/560 kb DNA		
	4	$(AAG)_n$	10	2	88
		$(ATG)_n$, $(GTG)_n$	2	3	
Maize (*Zea mays*)	8	$(CT)_n$	4		31*
		$(AG)_n$	1		
		(GCT)	1		
Wild and cultivated barley (*Hordeum vulgare*)			4	33, 28, 37	36*
Wheat (*Triticum aestivum*)	18	$(GA)_n$	15	4.6 (2–7)	41
		$(GT)_n$			
	40		23	6.2 (3–16)	39
		$(GA)_n$	8	3–16	40
		$(GT)_n$	4	3–6	
		Compound	11	3–9	

Table 2. (*Contd*)

Crop and species	No. of genotypes examined	Repeat unit	No. of loci	No. of alleles at one or more loci	Ref.
Rice (*Oryza* sp.)	20	$TAA(GA)_7A(GA)_2A(GA)_{11}TTGC$	1	6	32
		$AAC(GA)_{20}AGTAA$	1	11	
		$CTCTAT(GT)_{12}TTT$	1	5	
		$(GT)_{16}TT(GT)_4GAC$	1	9	
		$GGAA(GA)_{16}GGGG$	1	8	
		$CGG(T)_{15}(GT)_{14}TTC$	1	8	
		$CT(GA)_{16}GG$	1	8	
		$GT(GA)_9TAG(ATC)_4$	1	6	
	16	$(GGC)_n$	1	6	14*, 15*
Oryza sativa	238	$CT(GA)_9GGT$, $GCT(GGT)_{10}GCA$	10	2–14	33*
Oryza sativa and 8 other *Oryza* spp.	> 15	$(TG)_{10}$, $(GT)_n$, $(GAA)_6$, $(CAC)_5$, $(GATA)_4$, $(GGAT)_4$, $(GACA)_4$	—	—	82, 89, 90
Seashore paspalum (*Paspalum vaginatum*)	46	$(GA)_n$, $(CA)_n$	5	14 (6–16)	91
Wild yam (*Dioscorea tokoro*)	19–23	$(CT)_{17}$	1	6	35
		$(CT)_{16}T(CT)_3T(CT)_3T(CT)_3$	1	9	
		$(CT)_{12}TT(CT)_4$	1	4	
		$(AT)_8$	1	3	
		$(GT)_9GAG(GA)_4TAATACAGT(AAT)_8$	1	7	
		$(TA)_{12}(CA)_{19}$	1	8	
Tomato (*Lycopersicon esculentum* × *L. pennellii* – F_2)	84	GATA, GACA	32		73*
Six angiosperm tree species	6	1 per 15–200 kb			25
Bur oak (*Quercus macrocarpa*)		$(CA)_n$, $(GA)_n$	3	11–20	92
Pinus radiata	96	$(GA)_{10}$, $(CA)_{10}$	2	6	37

*In these studies different loci have also been mapped.

Table 3. Terminology of microsatellites

Category (Weber[7])	Examples (Smith and Devey[37])
Perfect	$(CA)_{14}$
Compound (perfect)	$(TA)_{31}$ $(CA)_{42}$
Imperfect	TAAT $(TAA)_5$ AATATAATA

that the number of repeats below 12 usually showed very low levels of polymorphism. However, the number of repeats, below which the degree of polymorphism would dramatically drop off may be lower in rice than in humans[33].

Development of microsatellite markers

Microsatellite arrays being shorter in length than minisatellites are easy to clone, sequence and amplify through PCR. New microsatellites can be cloned directly from total genomic DNA libraries or libraries enriched for specific microsatellites[48,49]. For this purpose, genomic libraries with relatively small clones may also be generated in sequencing vectors by ligating the sequencing vector with genomic DNA restricted with frequent cutting enzymes. The library can subsequently be screened for microsatellites by hybridization with repetitive oligonucleotide probes, and the positive clones identified on screening can then be sequenced. A more efficient way of microsatellite enrichment of genomic library has recently been worked out using the bead-enrichment method[50]. For libraries of larger insert sizes, either the positive clones may be subcloned first or a pair of degenerate sequencing primers may be directly used, which will anneal directly to the sequences in the clone, and thus facilitate determination of the flanking sequences[51]. Alternatively, known DNA sequences (microsatellites) may be searched from the data bases like EMBL and GenBank™, and flanking sequences noted (Table 4). Once the flanking sequences are known, primers may be designed either by manual inspection or with the help of computer programs[52–54]. While making a choice for primer, one has to take into consideration

Table 4. Simple sequence repeats in gene sequences available in the database*

SSR	Species	No. of genes	Important gene(s)
$(AT)_{10}$	*Antirrhinum majus*	1	*tap2* gene
	Arabidopsis thaliana	4	*atpC1* (ATP synthase gamma subunit); chitinase; *CRA* (storage protein); *ats1A* (rbcS)
	Hordeum vulgare	1	Rubisco activase
	Catharanthus roseus	1	*cyc02* mRNA
	Daucus carota	1	Carrot V-type H + ATPase gene
	Lilium henryi	1	*del* transposon gene
	Zea mays	2	En-1 mosaic protein; ZC2 (Zein)
	Pisum sativum	2	*gpal* (GAPDH); *P4* (organ specific)
	Solanum tuberosum	1	Patatin
	Oryza sativa	2	Oryzacystatin; *phy18*
	Glycine max	4	*Gy2* (Glycinin); *lbc* (leghemoglobin gene); *Gmhsp17.6-L* (heat shock protein gene); *sc514* (gene for lipoxygenase)
	Nicotiana tabacum	3	*t3-ars*; glycine-rich protein gene; *nia-1* gene for nitrate reductase
	N. plumbaginifolia	1	Cab-E gene 5′-flanking region
	Lycopersicon esculentum	2	*Td* (threonine deaminase gene); fruit ripening specific mRNA
	Triticum aestivum	1	Alpha-amylase gene
	Candida albicans	1	*TEF-2* (elongation factor gene)
$(CT)_{10}$	*Lemna gibba*	5	rbcS genes-*SSU26, SSU40A, SSU40B, SSU5A, SSU5B*
	Zea mays	1	*Gpc1*
	Ricinus communis	1	*R. communis* mRNA
	Spinacia oleracea	1	*ACP1* (acyl carrier protein I) mRNA
$(GA)_{10}$	*Lemna gibba*	1	chlorophyll *a/b* apoprotein gene
	Solanum tuberosum	1	*ST-LS1* (Light-inducible tissue-specific gene)
$(GT)_{10}$	*Petunia hybrida*	1	*CHI-Bgene* (Chalcone flavanone isomerase)
$(AAT)_7$	*Medicago sativa*	1	Leghemoglobin gene
	Petunia hybrida	1	rbcS gene-*SSU491*
	Lycopersicon esculentum	1	Gene for an elongation factor
$(ATT)_7$	*Dolichos biflorus*	1	Lectin *DB58* gene
	Psium sativum	1	*P4* (organ specific)
	Glycine max	1	*SbPRP1* (soybean proline-rich protein gene)
	Spinacia oleracea	1	mRNA, photosystem I subunit V
	Nicotiana plumbaginifolia	1	*ATP2-1* (mt ATP synthase gene)
$(AGC)_7$	*Zea mays*	2	GSH glutathione S-transferase III mRNA; spontaneous deletion sequence from waxy (*wx-B*) gene
	Oryza sativa	1	Cab2R
	Triticum aestivum	5	Gliadin genes (1 gene; 4 cds clones)
$(ACC)_7$	*Candida maltosa*	1	Phosphoribosyl- amidoimidazole-succino carboxamide-synthetase (complete cds)
$(TTC)_7$	*Mesembryanthemum crystallinum*	2	*fnr-A* (ferredoxin-NADP⁺ reductase); *ppc1* (phosphoenol-pyruvate carboxylase)
	Catharanthus roseus	1	mRNA for strictosidine synthase
	Cucurbita maxima	1	Nitrate reductase mRNA
	Oryza sativa	1	*rbcS* gene
$(ATG)_7$	*Arabidopsis thaliana*	1	*atpC2* (ATP synthase gamma subunit)
$(TAG)_7$	*Hordeum vulgare*	1	5SrRNA gene
$(ATC)_7$	*Pisum sativum*	1	mRNA for cytoplasmic lipoxygenase
	Sinapis alba	1	mRNA for chalcone synthase

Table 4. (*Contd*)

SSR	Species	No. of genes	Important gene(s)
$(AAC)_7$	*Dianthus caryophyllus*	1	1-amino-cyclopropane-1- carboxylate synthase (CARAcc)
	Sorghum bicolor	2	pGK1 (Kafirin DNA); pSK8 (Kafirin mRNA)
	Triticum aestivum	~20	Gliadins (mainly cds clones)
	T. urartu	2	Gliadins (complete cds)
	Candida tropicalis	2	*cat* gene (catalase); *POX9* (peroxisomal catalase)
$(AAG)_7$	*Pisum sativum*	1	*leg*J (legumin)
$(GGT)_7$	*Cucurbita maxima*	1	Nitrate reductase mRNA
	Candida tropicalis	1	*POX2* (complete cds)
$(GCT)_7$	*Hordeum vulgare*	1	mRNA for *ADH1*
	Triticum aestivum	1	Alpha-Amy 12/34 gene 5'- region
$(GCA)_7$	*Zea mays*	1	Spontaneous deletion sequence from waxy (*wx-B*) gene
	Triticum aestivum	~8	Gliadin genes (mainly cds clones)
$(CCG)_7$	*Sorghum vulgare*	1	Hydroxyproline-rich glycoprotein gene
$(GGC)_7$	*Zea mays*	1	Albumin b-32 mRNA
$(ACAT)_7$	*Lemna gibba*	1	Chorophyll *a/b* apoprotein gene (cds)
$(AATT)_4$	*Pisum sativum*	1	*S2* (organ specific)
	Solanum tuberosum	3	Patatin genes (three clones PS3, PS27, 5B6B)
	Oryza sativa	1	Waxy gene for glycogen synthetase
	Lycopersicon esculentum	1	Dispersed repeat CR1 associated with *Cab1* gene
$(TTTA)_4$	*Glycine max*	2	*Lbc1*; nodulin-35
$(AAAT)_4$	*Hordeum vulgare*	1	Thiol protease aleurone gene
	Oryza sativa	5	Glutelins (DNA and cds clones)
$(GGGC)_4$	*Zea mays*	1	Chloroplast GADPH gene
$(ACGG)_4$	*Zea mays*	1	*Gpc1* (GADPH subunit C)
$(CATT)_4$	*Pisum sativum*	1	*Fed-1* (ferredoxin I)
$(ATAC)_4$	*Sorghum vulgare*	1	Hydroxyproline-rich glycoprotein gene
$(TATG)_4$	*Glycine max*	1	*Gy3* (glycinin subunit G3)
$(ATAG)_4$	*Lemna gibba*	1	Chlorophyl *a/b* apoprotein gene (complete cds)
$(GTGA)_4$	*Glycine max*	1	*Tgm1* (seed lectin gene transposable element)
$(CGCT)_4$	*Zea mays*	1	Chloroplast GADPH gene (cds)

*Compiled from data supplied by Susan McCarthy, Co-ordinator, Plant Genome Data and Information Centre, National Agricultural Library, Beltsville, USA.

firstly, the melting temperature of primer DNA (based on GC content) permitting optimum PCR amplification conditions and *secondly*, the nucleotide sequence of the primer, that should avoid self-annealing.

After PCR amplification, products may be separated by electrophoresis on agarose gels and visualized by ethidium bromide staining. For better resolution and detection of smaller differences in amplified products, polyacrylamide gels and silver staining are preferred. Microsatellite products can also be labelled by ^{32}P or ^{35}S to allow their visualization by autoradiography. Reducing the number of reaction cycles may also improve the resolution. Denaturation gels containing formamide and urea may be used to eliminate heteroduplex formation or spurious conformations. Resolution may also be improved by the fluorescent labelling of PCR primers, so that the different alleles can be distinguished by quantitative differences in the relative intensities of fluorescence in the corresponding bands. Overlapping products may also be sometimes obtained due to amplification of alleles of same length at different loci through the use of different sets of primers. These may be resolved if different fluorescent dyes are used with different sets of primers, so that computer analysis of the output signal from scanning laser enables distinction between different overlapping primer products, due to different characteristic wavelengths of different dyes. Simultaneous amplification of different loci by multiplex PCR has also been facilitated by the use of different fluorescent dyes for different primers associated with different loci.

Applications

Marker-assisted selection in plant breeding

One of the basic requirements in any plant breeding exercise is the ability to screen large segregating population for a desirable trait. Since several of these desirable traits, particularly the quantitative traits are influenced by environment, the visual selection or selection based on phenotypic estimations may not be effective. In other cases, a phenotypic trait like disease susceptibility may not express, if suitable environment is not available, and during pyramiding of resistance genes, selection for additional resistance genes may not be feasible in the presence of an already existing resistance gene. In still other cases, the contrasting forms may not be distinguishable at the seeding stage, making it necessary to grow population up to the adult stage before exercising selection. Raising large populations up to the adult stage for mere selection needs effort and expense, and therefore, a plant breeder would like to exercise selection at the seedling stage itself, if feasible. These difficulties can be largely overcome, if molecular marker-assisted plant breeding is exercised.

During the last more than ten years, it has been emphasized that DNA-based molecular markers can actually be used to facilitate plant breeding[55,56], which is sometimes described as 'molecular breeding'. For this purpose, initially restriction fragment length polymorphisms (RFLPs) were proposed and used for genetic mapping in a number of crop plants. Subsequently, several other PCR-based molecular markers including random amplified polymorphic DNAs (RAPDs) and microsatellites or SSRs were suggested and utilized. Table 5 gives advantages and disadvantages of these different molecular markers. The use of these molecular markers is, however, determined by several factors, a major factor being whether their use is cost effective.

Since the use of RFLPs and RAPDs in plant breeding has been discussed in many reviews[57,58], no attempt will be made here to discuss their use. It has been emphasized in recent years that RFLP markers are not particularly suitable for plant breeding, which requires screening of large populations, although Southern blotting has now been automated. Instead, increasing interest in the use of microsatellites has been witnessed (particularly the dinucleotide repeats), in view of their ubiquitous presence, abundance and hypervariability. Even though the cost of *Taq* polymerase used in PCR amplification of microsatellites may limit its use at present, but the price of this enzyme is expected to fall in due course[59].

For genetic diagnostic in plant breeding, even though a complete automation may take time, some automation has already been achieved for DNA extraction at the rate of 5–10 μg of DNA per tissue sample per minute. In several laboratories, gel loading, visualization and image recording have also been automated. Documentation and analysis of results can also be automated. Computer softwares have been developed to handle data generated from molecular marker-assisted assays on large populations. However, no reliable and completely automatic system (involving steps from DNA extraction up to data analysis) is available so far, but should be developed in due course of time[59].

Genome selection during gene introgression in plant breeding

The use of minisatellites has also been recommended for facilitating gene introgression in plant breeding programmes[60]. In such a breeding programme, F_1 hybrid is backcrossed repeatedly to the recipient parent with

Table 5. Comparison of microsatellites with RFLP and RAPD markers

Characteristic	RFLP	RAPD	Microsatellites
Principle involved	DNA blot hybridization	PCR amplification with random primers	PCR amplification of simple sequence repeat loci
Type of polymorphism	Single base changes; insertions; deletions	Single base substitutions; insertions; deletions	Variation in number of repeat motifs
Genomic abundance	High	Very high	Medium
Level of polymorphism	Medium	Medium	High
Inheritance	Codominant	Dominant	Codominant
Amount of DNA required	5–10 μg	10–25 ng	50–100 ng
Sequence information required	No	No	Yes
Radioactive detection required	Yes/no	No	No
Development costs	Medium	Medium	High
Start up costs	Medium/high	Low	High
Detection	Autoradiography; biotin labelling	Ethidium bromide; silver staining; fluorescence	Ethidium bromide; silver staining; fluorescence

an objective to restore the genome of the recipient parent with introgressed trait included in it. The number of backcrosses in this programme can be reduced through the use of polymorphic markers scattered in the genome[57,61,62]. For this purpose the use of highly polymorphic minisatellite loci consisting of variable number of tandem repeats (VNTRs) of relatively short sequences mentioned earlier has been recommended. The probes into these minisatellites reveal DNA patterns which can be used as 'DNA fingerprints' as done in rice[63]. However, the complex hybridization patterns and preferential localization (telomeres) makes their use limited as molecular markers. DNA fingerprinting using synthetic oligonucleotides (representing di-, tri- and tetranucleotide repeats[64]) can, therefore, be used for this purpose.

In view of the above, we feel that since microsatellites can be used for 'DNA fingerprinting', they can also be used for genome selection during the introgression breeding programme. The genome selection is based on the assumption that the genomes of recipient and donor parents can be tagged with the help of DNA fingerprints. Therefore, the efficiency of genome selection is proportional to the faithfulness of genome tagging, if genome tagging is used for selecting individuals in the tails of the distributions in backcross generations.

Genome mapping

During the last more than ten years, molecular maps based on DNA markers have been prepared in several crops, some of them being available now in a fairly saturated state. Genes for many phenotypic traits have also been added to these maps, providing composite maps having molecular markers as well as genes for specific traits. In some cases, this has also led to the establishment of close linkage between specific genes and specific molecular markers, so that the molecular markers can be used for diagnostics during screening of segregating populations or diverse germplasm. Availability of a saturated genetic map for a crop also allows map-based cloning and subsequent isolation of desirable genes. Since many related genomes (for example, human and other primates; chickpea, lentil and pea; Triticeae members) show synteny, markers developed for one genome can be used for other related genomes. The most commonly used markers for this purpose are RFLPs, although in recent years RAPDs have been added and RFLPs have been converted into STSs (sequence tagged sites). Only very recently a beginning has been made to use microsatellites for genetic/physical mapping (Table 2). In wheat, recently Devos *et al.*[65] have shown the potential of microsatellite sequences as a PCR-based alternative to RFLP markers. The study on the application of two microsatellite sequences (one in a γ-gliadin pseudogene and the other one in LMW-glutenin gene) in wheat storage proteins as molecular markers revealed that these sequences are genome specific and displayed high levels of variation. Further, the study revealed that even within multigene families, minor sequence variation in the microsatellite flanking regions can be exploited to construct highly specific primer sets, as was demonstrated with the γ-gliadin gene family[65].

An evaluation of the sequence tagged site markers for genetic analysis has been done in *Citrus* and related species[66]. The study gives details of the properties of two STMS loci that were isolated using bead enrichment method[50]. In addition, the extent of primer sequence conservation across the range of *Citrus* species and species of other related genera was examined to determine the utility of STMS markers for mapping through intergeneric crosses and for comparative genome analysis. During this study, most of the major species of *Citrus* likely to be used in breeding programmes, were examined. Since the markers were present and were polymorphic in all species, they appear to be ideal to assist in tracing both monogenic and polygenic traits in future.

Various computer programs are available that calculate recombination values and map distances which are then used to create genetic maps. LINKAGE-1 (ref. 67) is based on chi-square analysis and only allows the evaluation of pairwise (two-point) analyses of recombination values. GMENDEL[68] uses a log-likelihood method or G statistics. JOINMAP[69] accepts data with different expected segregation ratios and integrates data from different populations. MAPMAKER[70] performs multipoint analyses using maximum likelihood in F_2 and backcross generations. In some recent studies, a high degree of reproducibility and accuracy in allele size determination was achieved using the GENESCAN automated software[66,71]. Recently, DRAWMAP[72] has also been developed for drawing genetic linkage maps.

The first attempt to use microsatellites for genetic mapping was made by Zhao and Kochert[14,15] in rice, using $(GGC)_n$ microsatellite. During the last two years several other reports of genetic mapping using microsatellites have become available (Table 2). A prerequisite for genetic mapping is the identification of microsatellite loci, an objective that can be achieved by screening either the available DNA sequences from the data bases, or the genomic libraries. Both the above approaches have actually been used for designing primers used for microsatellite amplification, although several of these studies were meant for a study of polymorphism at microsatellite loci, rather than for genetic mapping. In tomato, however, GATA-containing microsatellites were detected with the oligonucleotide probe $(GATA)_4$ by Vosman *et al.*[27]. These microsatellite loci were used for mapping on the existing map having 51 RFLP markers

Table 6. Gene tagging using microsatellites

Crop	Gene tagged	Ref.
Glycine max (soybean)	Soybean mosaic virus resistance	34
Oryza sativa (rice)	Yield (QTLs)	38

covering all chromosomes. No PCR amplification was involved, so that no primers were designed. Only 44 plants out of the 84 plants drawn from F_2 population derived from *Lycopersicon esculentum* (allround) × *L. pennellii* (LA 716) and earlier used for the above RFLP map, were used for mapping 32 microsatellite loci containing GATA and GACA repeats[73].

Gene tagging

Although there are numerous examples of tagging important genes in crop plants employing RFLP and RAPD markers, only two examples are available till date where microsatellite markers have been used for gene tagging[34,38] (Table 6). Efforts are under way in this direction and many more genes will be tagged using microsatellite markers in the near future.

To find a molecular marker linked to a particular monogenic trait, one need not necessarily require a saturated genetic map. One can make use of near isogenic lines[74,75] or follow the approach of bulked segregant analysis[76,77].

Cultivar identification, estimation of genetic relatedness and germplasm conservation

In the past, hybridization based (involving minisatellite and SSR probes) and PCR-based (RAPDs) fingerprinting methods[78] have been used for identification of plant cultivars as well as in areas like plant propriety rights protection, etc. Moreover, assessment of genetic diversity among cultivars and their wild relatives has recently attracted increased attention in efforts to cope with the commonly encountered reduction of diversity due to the practice of growing monocultures. Therefore, efforts to characterize existing germplasm and available genetic diversity are warranted. Considering advantages available with microsatellite markers, these markers are certain to find increased usage in efforts to quantify genetic diversity and to characterize accessions in plant germplasm collections. The usefulness of these markers for germplasm characterization has been demonstrated for sweet potato[79], soybean[80], rapeseed[81], rice[82] and phaseolus[83]. Primers representing a combination of two tetranucleotides or compound microsatellites, are equally effective for characterization of germplasm. The polymorphism obtained may be used to distinguish individual plant varieties[84]. Although the development and use of SSR markers require an investment in funds and facilities, the utility of these PCR-detectable genetic markers clearly warrants it.

1. Litt, M. and Luty, J. A., *Am. J. Hum. Genet.*, 1989, **44**, 397–401.
2. Edwards, A., Civitello, A., Hammond, H. A. and Caskey, C. T., *Am. J. Hum. Genet.*, 1991, **49**, 746–756.
3. Jacob, H. J., Lindapaintner, K., Lincoln, S. E., Kusumi, K., Bunker, R. K., Mao, Yi-Pei, Ganten, D., Dzau, V. J. and Lander, E. S., *Cell*, 1991, **67**, 213–224.
4. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlim, E. and White, R., *Science*, 1987, **235**, 1616–1622.
5. Jeffreys, A. J., Wilson, V. and Thein, S. L., *Nature*, 1985, **314**, 67–73.
6. Jeffreys, A. J., Wilson, V., Neumann, R. and Keyte, J., *Nucleic Acids Res.*, 1988, **16**, 10953–10971.
7. Weber, J. L., *Genomics*, 1990, **7**, 524–530.
8. Weber, J. L. and May, P. E., *Am. J. Hum. Genet.*, 1989, **44**, 388–396.
9. Schlotterer, C. and Tautz, D., *Nucleic Acids Res.*, 1993, **20**, 211–215.
10. Hamada, H., Petrino, M. G. and Kakunaga, T., *Proc. Natl. Acad. Sci. USA*, 1982, **79**, 6465–6469.
11. Tautz. D., *Nucleic Acids Res.*, 1989, **17**, 6463–6471.
12. Stallings, R. L., Torney, D. C., Hildebrand, C. E., Longmire, J. L., Deaven, L. L., Jett, J. H., Doggett, N. A. and Moyzis, R. K., *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 6218–6222.
13. Beckmann, J. S. and Soller, M., *Bio/Technology*, 1990, **8**, 930–932.
14. Zhao, X. and Kochert, G., *Mol. Gen. Genet.*, 1992, **231**, 353–359.
15. Zhao, X. and Kochert, G., *Plant Mol. Biol.*, 1993, **21**, 607–614.
16. Moore, S. S., Sargeant, L. L., King, T. J., Mattick, J. S., Georges, M. and Hetzel, D. J. S., *Genomics*, 1991, **10**, 654–660.
17. Rassmann, K., Schlotterer, C. and Tautz, D., *Electrophoresis*, 1991, **12**, 113–118.
18. Beckmann, J. S. and Weber, J. L., *Genomics*, 1992, **12**, 627–631.
19. Moran, C., *J. Hered.*, 1993, **84**, 274–280.
20. Estoup, A., Solignac, M., Harry, M. and Cornuet, J. M., *Nucleic Acids Res.*, 1993, **21**, 1427–1431.
21. Hughes, C. R. and Queller, D. C., *Mol. Ecol.*, 1993, **2**, 131–137.
22. Rico, C., Zadworny, D., Kuhnlein, U. and Fitzgerald, G. J., *Mol. Ecol.*, 1993, **2**, 271–272.
23. Kondo, Y., Mori, M., Kuramoto, Y., Yamada, J., Beckman, J. S., Simon-Chazottes, D., Montagutelli, X., Guenet, J. and Serikawa, T., *Mamm. Genome*, 1993, **4**, 571–576.
24. Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sunden, S. L. F., Hawkins, G. A., Toldo, S. S., Fries, R., Grosz, M. D., Yoo, J. Y. and Beattie, C. W., *Genetics*, 1994, **136**, 619–639.
25. Condit, R. and Hubbell, S. P., *Genome*, 1991, **34**, 66–71.
26. Akkaya, M. S., Bhagwat, A. A. and Cregan, P. B., *Genetics*, 1992, **132**, 1131–1139.
27. Vosman, B., Arens, P., Rus-Kortekaas, W. and Smulders, M. J. M., *Theor. Appl. Genet.*, 1992, **85**, 239–244.
28. Legercrantz, U., Ellegren, H. and Anderson, L., *Nucleic Acids Res.*, 1993, **21**, 1111–1115.
29. Morgante, M. and Olivieri, A. M., *Plant J.*, 1993, **3**, 175–182.
30. Thomas, M. R. and Scott, N. S., *Theor. Appl. Genet.*, 1993, **86**, 985–990.
31. Senior, M. L. and Heun, M., *Genome*, 1993, **36**, 884–889.
32. Wu, K.-S. and Tanksley S. D., *Mol. Gen. Genet.*, 1993, **241**, 225–235.
33. Yang, G. P., Saghai-Maroof, M. A., Xu, C. G., Zhang, Q. and Biyashev, R. M., *Mol. Gen. Genet.*, 1994, **245**, 187–194.
34. Yu, Y. G., Saghai-Maroof, M. A., Buss, G. R., Maughan, P. J. and Tolin, S. A., *Phytopathology*, 1994, **84**, 60–64.

35. Terauchi, R. and Konuma, A., *Genome*, 1994, **37**, 794–801.
36. Saghai-Maroof, M. A., Biyashev, R. M., Yang, G. P., Zhang, Q. and Allard, R. W., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 5466–5470.
37. Smith, D. N. and Devey, M. E., *Genome*, 1994, **37**, 977–983.
38. Zhang, Q., Gao, Y. J., Yang, S. H., Ragals, R., Saghai-Maroof, M. A. and Li, Z. B., *Theor. Appl. Genet.*, 1995 (in press).
39. Plaschke, J., Borner, A., Wendehake, K., Ganal, M. W., and Roder, M. S., *Euphytica*, 1995 (in press).
40. Roder, M. S., Plaschke, J., Konig, S. U., Borner, A., Sorrels, M. E., Tanksley, S. D. and Ganal, M. W., *Mol. Gen. Genet.*, 1995, **246**, 327–333.
41. Plaschke, J., Ganal, M. and Roder, M., *Theor. Appl. Genet.*, 1995 (in press).
42. Bell, C. J. and Ecker, J. R., *Genomics*, 1994, **19**, 137–144.
43. Weising, K., Weigand, F., Driesel, A. J., Kahl, G., Zischler, H. and Epplen, J. T., *Nucleic Acids Res.*, 1989, **17**, 10128.
44. Wang, Z., Weber, J. L., Zhong, G. and Tanksley, S. D., *Theor. Appl. Genet.*, 1994, **88**, 1–6.
45. Sarkar, G., Paynton, C. and Sommer, S. S., *Nucleic Acids Res.*, 1991, **19**, 631–636.
46. Cregan, P. B., *Probe*, 1992, **2**, 18–22.
47. Rubinsztein, D. C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S., Margolis, R. L., Ross, C. A. and Ferguson-Smith, M. A., *Nature Genetics*, 1995, **10**, 337–343.
48. Ostrander, E. A., Jong, P. M., Rine, J. and Duyk, G., *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 3419–3423.
49. Grist, S. A., Firgaira, F. A. and Morley, A. A., *Biotechniques*, 1993, **15**, 304–309.
50. Kijas, J. M. H., Fowler, J. C. S., Garbett, C. A. and Thomas, M. R., *Biotechniques*, 1994, **16**, 656–662.
51. Browne, D. L. and Litt, M., *Nucleic Acids Res.*, 1992, **20**, 141.
52. Rychlik, W. and Rhoads, R. E., *Nucleic Acids Res.*, 1989, **17**, 8543–8551.
53. Lowe, T., Sharefkin, J., Yang, S. Q. and Dieffenbach, C. W., *Nucleic Acids Res.*, 1990, **18**, 1757–1761.
54. Eberhardt, N. L., *Biotechniques*, 1992, **13**, 914–917.
55. Beckmann, J. S. and Soller, M., *Theor. Appl. Genet.*, 1983, **67**, 35–43.
56. Tanksley, S. D., Young, N. D., Paterson, A. H. and Bonierbale, M. W., *Bio/Technology*, 1989, **7**, 257–264.
57. Beckmann, J. S. and Soller, M., *Euphytica*, 1986, **35**, 111–124.
58. Williams, J. G. K., Hanafey, M. K., Rafalski, J. A. and Tingey, S. V., *Methods Enzymol.*, 1993, **218**, 704–740.
59. Rafalski, J. A. and Tingey, S. V., *Trends Genet.*, 1993, **9**, 275–280.
60. Hillel, J., Schaap, T., Haberfeld, A., Jeffreys, A. J., Plotzky, Y., Cahaner, A. and Lavi, U., *Genetics*, 1990, **124**, 783–789.
61. Tanksley, S. D. and Rick, C. M., *Theor. Appl. Genet.*, 1980, **57**, 161–170.
62. Tanksley, S. D., Medina-Filho, H. and Rick, C. M., *Theor. Appl. Genet.*, 1981, **60**, 291–296.
63. Dallas, J. F., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 6831–6835.
64. Sharma, P. C., Winter, P., Bunger, T., Huttel, B., Weigand, F., Weising, K. and Kahl, G., *Theor. Appl. Genet.*, 1995, **90**, 90–96.
65. Devos, K. M., Bryan, G. J., Collins, A. J., Stephenson, P. and Gale, M. D., *Theor. Appl. Genet.*, 1995, **90**, 247–252.
66. Kijas, J. M. H., Fowler, J. C. S. and Thomas, M. R., *Genome*, 1995, **38**, 349–355.
67. Suiter, K. A., Wendel, J. F. and Case, J. S., *J. Hered.*, 1983, **74**, 203–204.
68. Liu, B.-H. and Knap, S. J., *J. Hered.*, 1990, **81**, 407.
69. Stam, P., *Plant J.*, 1993, **3**, 739–744.
70. Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, A., Lincoln, S. E. and Newburg, L., *Genomics*, 1987, **1**, 174–181.
71. Kimpton, C. P., Gill, P., Walton, A., Urquhart, A., Millican, E. S. and Adams, M., *PCR Meth. Appl.*, 1993, **3**, 13–22.
72. Van Ooijen, J. W., *J. Hered.*, 1994, **85**, 66.
73. Arens, P., Odinot, P., Van Heusden, A. W., Lindhout, P. and Vosman, B., *Genome*, 1995, **38**, 84–90.
74. Martin, G. B., Williams, J. G. K. and Tanksley, S. D., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 2336–2340.
75. Paran, I., Kesseli, R. V. and Michelmore, R. W., *Genome*, 1991, **34**, 1021–1027.
76. Michelmore, R. W., Paran, I. and Kesseli, R. V., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 9828–9832.
77. Kesseli, R. V., Paran, I. and Michelmore, R. W., in *Application of RAPD Technology to Plant Breeding* (ed. Neff, M.), ASHS Publishers, St. Paul, MN, 1993.
78. Weising, K., Nybom, H., Wolff, K. and Meyer, W., *DNA Fingerprinting in Plants and Fungi*, CRC Press, Florida, 1995.
79. Jarret, R. L. and Bowen, N., *Plant Genetic Resources Newsletter*, 1994, **100**, 9–11.
80. Rongwen, J., Akkaya, M. S., Bhagwat, A. A., Lavi, U. and Cregan, P. B., *Theor. Appl. Genet.*, 1995, **90**, 43–48.
81. Poulsen, G. B., Kahl, G. and Weising, K., *Theor. Appl. Genet.*, 1993, **85**, 994–1000.
82. Ramakrishna, W., Chowdari, K. V., Lagu, M. D., Gupta, V. S. and Ranjekar, P. K., *Theor. Appl. Genet.*, 1995, **90**, 1000–1006.
83. Hamann, A., Zink, D. and Nagl, W., *Genome*, 1995, **38**, 507–515.
84. Gupta, M., Chyi, Y. S., Romero-Severson, J. and Owen, J. L., *Theor. Appl. Genet.*, 1994, **89**, 998–1006.
85. Morgante, M., Rafalski, A., Biddle, P., Tingey, S. and Olivieri, A. N., *Genome*, 1994, **37**, 763–769.
86. Cregan, P. B., Akkaya, M. S., Bhagwat, A. A., Lavi, U., Rongwen, J., in *Plant Genome Analysis* (ed. Gresshoff, P. M.), CRC Press, pp. 43–49.
87. Kresovich, S., Szewc-McFadden, A. K., Bliek, S. M. and McFerson, J. R., *Theor. Appl. Genet.*, 1995, **91**, 206–211.
88. Depeiges, A., Goubely, G., Lenoir, A., Cocherel, S., Picard, G., Raynal, M., Grellet, F., Delseny, M., *Theor. Appl. Genet.*, 1995, **91**, 160–168.
89. Gupta, V. S., Ramakrishna, W., Rawat, S. R. and Ranjekar, P. K., *Biochem. Genet.*, 1994, **32**, 1–8.
90. Ramakrishna, W., Lagu, M. D., Gupta, V. S. and Ranjekar, P. K., *Theor. Appl. Genet*, 1994, **88**, 402–406.
91. Liu, Z.-W., Jarret, R. L., Kresovich, S. and Duncan, R. R., *Theor. Appl. Genet.*, 1995, **91**, 47–52.
92. Dow, B. D., Ashley, M. V. and Howe, H. F., *Theor. Appl. Genet.*, 1995, **91**, 137–141.

ACKNOWLEDGEMENT. We thank the Department of Biotechnology, New Delhi, for financial assistance.

Received 16 August 1995; accepted 18 November 1995