REVIEW OF LITERATURE
Chapter 2

Review of Literature

Purity of hybrid seed lots is assayed conventionally by a grow out test (GOT) on a representative sample of the seed that is to be marketed. The GOT involves growing plants to maturity and assessing several morphological and floral characteristics that distinguish the hybrid (Yashitola et al., 2002). With increasing numbers of cultivars and a finite number of conventional characters, it has become apparent that such traits will not suffice to establish uniqueness (Rongwen et al., 1995). In addition to this, GOT can be affected by adverse environmental conditions. Thus there is a need for an assay to assess genetic purity of hybrid seeds that is both accurate and faster than the GOT.

To solve this problem, molecular markers like proteins, isoenzymes, and DNA markers can be used. Chemical or biochemical analyses have been developed and have provided excellent (although frequently limited) cultivar identification procedures in a number of species. Traits such as the presence or absence of isozymes, specific protein bands in PAGE (Polyacrylamide gel electrophoresis) and SDS-PAGE (Sodium-dodecyl-sulphate Polyacrylamide gel electrophoresis) respectively or the presence or the level of storage compound such as specific oils have been used for cultivar identification (Preston et al., 1999). Initially proteins and isoenzymes were used to test hybrid seed purity and for genotype identification (Arus, 1983; Nijenhuis, 1971).
2.1 Study of soluble seed protein profiles

2.1.1 Gossypium species

Castleberry and Coleman (1972) classified four storm proof and non-storm proof varieties of cotton into four groups on the basis of their globulin patterns.

In the studies on water, salt and alkali soluble seed proteins of various cotton cultivars Ibragimov et al. (1973) found differences in albumin fractions in respect of the amount of electrophoretic zones and their mobility.

The electrophoretic spectrum of soluble seed proteins of 30 G. hirsutum cultivars showed marked differences between fuzzy seeds and genetically fuzzyless seeds and also between long fibred and short fibred cultivars (Zapruder et al., 1980).

In an electrophoretic analysis of the soluble protein fractions in the seed of a mutant variety 59041 of G. barbadense, Yunuskhanov et al. (1982) found the protein markers H 0.13 and B 0.18. To this they opined that the gene for protein B 0.18 has been replaced in the mutant by the gene for H 0.13. This was further substantiated by the presence of fuzz on the chalazal section and lateral surface of the mutant seed like that in G. hirsutum seed controlled by the gene FiFi, which evidently was linked with the gene for protein marker H 0.13.

Kapse and Nerkar (1985) studied disc-electrophoretic patterns of soluble seed proteins in four intra-G. hirsutum hybrids, two G. hirsutum x G. barbadense hybrids, parents and G. arboreum. They observed unique differences in the banding patterns of the cultivars. They opined that this technique could be useful to identify cotton cultivars and can be complementary, if not a substitute, to field-test for the determination of genetic purity of seed lots.
2.1.2 Wheat and related species

Based on the electrophoretic pattern, Chen and Bushuk (1970) reported that the proteins of triticale were inherited from its parents — wheat and rye. Further they observed the differences in the pattern of gliadins and glutenins in a *durum* cultivar and a cultivar of hard red spring wheat and opined that these qualitative difference might account for the difference in their bread making quality and provide additional evidence in the support of the hypothesis that genes for bread making quality factors are in the D genome.

Zillman and Bushuk (1979) observed that seed source had no effect on gliadin electrophoregrams of 5 Canadian wheat cultivars grown at 10 locations in Manitoba and Saskatchewan nor the 5 Australian cultivars grown in Canada and Australia. Their results confirmed that the gliadin electrophoregram is independent of environment and substantiated the suitability of the electrophoregram for wheat cultivar identification.

Ellis (1971) reported that SGE [Single gel electrophoresis] of grain proteins made it possible to place a number of wheat cultivars in groups and 17 cultivars could be identified by a combination of their electrophoretic pattern, phenol reaction, coleoptile anthocyanin and grain hardness.

Three lab procedures namely, phenol reaction, hardness testing and electrophoresis were examined by Wrigley and Shepherd (1974) for identification of about 50 wheat cultivars grown in Australia. They reported that the most discriminating of these methods was SGE of gliadin proteins extracted from single grains.
Chapter 2: Review of Literature

Dhaliwal (1977) electrophoresed albumins, globulins, gliadins and glutenins of several species of *Triticum* and *Aegilops* on SDS-PAGE. He observed that electrophoretic patterns differed from one species to another only in a few minor components or density of certain components.

Shewry *et al.* (1978b) reported that the SDS-PAGE technique is more sensitive in distinguishing gliadin of different wheat varieties than SGE.

Du Cros and Wrigkey (1979) compared several electrophoretic methods in terms of their reliability, convenience and ability to distinguish between varieties of wheat, barley, rye and triticale using protein extracted from the endosperm. Gradient gel electrophoresis and isoelectric focusing (IEF) offered improvement over the use of SGE and uniform pore polyacrylamide gels.

Zillman and Bushuk (1979c) obtained 85 distinct gliadin electrophoregram formulae for the 88 Canadian wheat cultivars studied and observed only 3 pairs of formulae essentially identical, which they attributed to very close genetic relationships among these 3 cultivars.

Fullington *et al.* (1980) observed that each of 5 wheat varieties studied exhibited at least one outstanding feature in its densitometric pattern when the total protein extract from single seeds was studied by SDS-PAGE.

Jones *et al.* (1982) determined and catalogued 88 United States wheat cultivars based on PAGE patterns of their gliadins. Most of the cultivars were readily differentiated by their electrophoregrams. However, some closely related cultivars gave them identical patterns and were thus not uniquely identifiable by PAGE.
Damania et al. (1983) examined prolams in 3168 single seeds of 64 landraces of wheat and barley from Nepal and Yemen Arab Republic using the vertical slab PAGE method. They reported that this technique was easy and rapid in assessing the extent of variation present within the population of landraces.

Mansur-Vergara et al. (1984) used the technique of SDS-PAGE of total protein extraction to estimate the storage protein genetic variability among 841 accessions of wild emmer wheat collected from various ecological regions in the Middle East. The computerized densitometer scanning of protein bands indicated that the greatest genetic variability occurred for proteins in the high molecular weight (MW) region (above 70,000 MW) followed by those in the medium range (70,000 to 30,300 MW). Comparatively little variability was revealed for protein sub-units of molecular weight below 33,300 MW.

Studies of Virdi and Larter (1984) revealed that polymorphism could exist in the protein banding patterns of any given population such as the triticale amphidiploids and their parents when prolams were extracted and analysed by PAGE. The degree of polymorphism was greatest in triticale populations synthesized from open-pollinated heterozygous rye strains.

Burgoon et al. (1985) devised a method to detect non-durum wheat gliadin proteins in pasta using aluminium lactate-PAGE at pH 3.1. Pastas were prepared from various mixtures of durum and hard red winter wheat flours and gliadins were extracted. Zymograms of pasta extracts matched those of the corresponding flour extracts. Durum wheat extracts lack certain slow moving gliadin protein bands present in other flours. Pasta adulterated with as little as 5% hard red winter flour could be detected with this method.
Chapter 2: Review of Literature

Damania (1985) analysed gliadins of tetraploid and hexaploid wheats from Nepal using the Al-lactate PAGE method. The tetraploids lacked slow moving bands in the ‘omega’ region found in hexaploid wheats. These bands also aided in identifying different hexaploid wheat cultivars.

Khan et al. (1985) investigated the PAGE procedure to determine its suitability for wheat variety identification by analysis of gliadin proteins. Gels of varying firmness were obtained by varying relative amounts of catalysts. Gel firmness and pore size affected resolution of gliadin proteins.

Sapirstein and Bushuk (1985) described a computer-aided methodology that facilitates identification and comparison of wheat cultivars based on gliadin relative mobility and band density data. The data was used in an equation to quantify electrophoretic pattern homology that determines the order of cultivar ranking. A separate computer programme evaluates the uniqueness of the unknown electrophoregram and identifies diverse genotypes. Their results (Sapirstein and Bushuk, 1986) provided strong evidence of the influence of gliadin composition to discriminate wheat cultivars according to functional type.

Cooke et al. (1986) described two methods of electrophoresis that enabled seeds of bread wheat, durum, wheat, rye and triticale to be characterized and distinguished from each other. The variation in prolamine and esterases composition both within and between different cultivars of the species was examined and generalized protein for each species was presented. They concluded that different electrophoretic techniques can provide a very rapid, definite way of characterizing and distinguishing between seeds of various cereal species and opined that this is another example of the increasing utility of electrophoresis in the areas of plant genetics, breeding and testing.
Zawistowska et al. (1986) observed that PAGE and SDS-PAGE patterns were similar for both Neepawa and Chile varieties of wheat but reversed phase-high performance liquid chromatography (RP-HPLC) patterns were quite different for low molecular weight proteins.

2.1.3 *Hordeum* species

McDaniel (1970) characterized individual seed proteins of selected *Hordeum* cultivars by a disc gel electrophoresis procedure and reported that cultivars related by descent, breeding or geography were more similar in protein phenotype than relatively unrelated cultivars. Variations in proteins between cultivars of *H. vulgare* were greater in some cases than protein variation between *H. vulgare*, *H. spotaneum* and *H. irregular*. Further he opined that this procedure might be particularly important as an initial screening tool in the effective production of heterotic barley hybrids.

McCausland and Wrigley (1977) examined endosperm proteins of 19 barley cultivars. They obtained 7 different patterns for hordeins by SGE and 13 different patterns by flat gel IEF of water-soluble and hordein proteins, which aided in cultivar identification in barley.

Shewry et al. (1978) compared the polypeptide composition of the hordein fractions from 29 varieties of barley on SDS-PAGE and IEF. A total of 11 different patterns were recognized using the former system while further small differences were revealed by IEF and 2-dimensional analysis. They reported that if these techniques were considered along with other grain characters it would be possible to uniquely identify all 29 varieties.
Shewry et al. (1978) compared the polypeptide composition of hordein polypeptides of single seeds of 88 varieties of barley on SDS-PAGE. They reported that this technique is of great potential value both for the commercial identification of grains samples and as an aid to existing techniques for establishment of varietal distinctness.

Burbidge et al. (1986) established methods of barley identification such as visual examination and gel electrophoresis. However, they reported that high performance liquid chromatography (HPLC) offers speed, automatic sorting of results and distinction between some similar varieties.

Gebre et al. (1986) used PAGE of hordein polypeptides to identify and catalog 40 barley cultivars. An electrophoregram formula was prepared for each cultivar using relative band mobility. Only 10 cultivars showed unique hordein band patterns and rest were separated into 17 groups, each group consisting of several cultivars. Some differences, mainly due to presence or absence of faint bands, were observed among cultivars within groups. In general, they reported that single kernels gave better resolution than meal samples.

Nielsen and Johansen (1986) opined that the hordein PAGE pattern was clearly the most powerful single system for identifying barley cultivars.

2.1.4 *Avena* species

Murray et al. (1970) studied seed proteins of 3 synthetic ampidiploids, 1 autotetraploid and 8 species that include representatives of all ploidy levels and karyotypes of *Avena*. They reported that A and C genome diploids were readily separated by band differences. The A genome protein patterns were essentially identical while the C genome patterns differed primarily in band
number. The tetraploid protein patterns differed but the hexaploid had high homology. However, intra-specific polymorphism was detected in the hexaploid cultivars.

Cooke and Draper (1986) examined the soluble protein of single seeds of different species of *Avena* and cultivars of *A. sativa* by PAGE using lactate buffer. Considerable polymorphism was observed by them in the avenin composition of the different species. It was possible to identify the species of wild oats unequivocally and to distinguish them from cultivars of *A. sativa*. The method was applicable to both whole and broken caryopses. Hence they opined that this technique provides a definite reference for the identification of *Avena* species in seed testing laboratory.

### 2.1.5 Rice

Studies of Siddiq *et al.* (1972) revealed that there were marked variations in the seed protein-banding pattern within and between three sub-species of *Oryza sativa*. Variability for number and intensity of protein bands was wider in *indica* than in *japonica* and *javanica* sub-species.

Iwasaki *et al.* (1982) studied proteins extracted from milled rice of long, medium and slender grain varieties by SGE. Twenty bands were resolved for both albumin and globulins, which supported the idea of classifying rice varieties into groups on the basis of electrophoregrams.

Chauhan and Nanda (1984) tested the efficacy of certain physio-chemical characters of grain and electrophoretic methods in rice varietal identification. They observed a wide variation in the pattern of protein bands, their electrophoretic mobility and intensity. However, they reported that the electrophoretic spectrum of salt soluble seed proteins exhibited an
ambiguous picture and resulted in greater controversy in classification compared to that of physio-chemical characters. Hence they suggested that instead of total salt soluble seed proteins, the variations in the specific proteins (enzymes) might be screened to explore their possible use in varietal identification.

On the contrary, Sarkar and Bose (1984) observed both qualitative and quantitative differences in electrophoreogram tracings in salt soluble protein fractions in a number of rice varieties. Analysis of variance was found to be useful in estimating the quantitative differences. The tracings of pattern appeared to be unique for each of the varieties investigated and remained constant under different environmental conditions. They suggested that this technique could be conveniently used for variety identification in rice.

Similarly Guo et al. (1986) examined the prolamin fractions by electro-focusing and categorized 25 varieties of Chinese rice into 4 major groups. Four major protein bands allowed differentiation of indica, japonica and glutinous rices and their hybrids. Minor bands allowed further differentiation within each group.

2.1.6 Sorghum

Chavan et al. (1980) investigated the soluble proteins of high and low tannin sorghum cultivars by disc gel electrophoresis. They observed that while in the low tannin cultivars IS 4129 and IS 3441, 17 and 16 bands were resolved, respectively, not a single band was resolved in the high tannin cultivars. Further they opined that tannin forms complexes with proteins during extraction and alters their properties, i.e. binding of tannins will increase the molecular weight of protein molecules, which may block the entry of such complexes into the gel.
Tripathi et al. (1981, 1983) made comparison of protein pattern in anthers and seeds of 6 different male-sterile lines, maintainer lines and fertility restorer lines of sorghum by disc gel electrophoresis. They observed both qualitative and quantitative differences in soluble protein patterns from male-sterile and maintainer lines.

Using vertical slab PAGE and IEF, Tripathi et al. (1982) studied the soluble protein patterns from seeds and anthers of 24 different male-sterile lines, maintainer lines and fertility restorer lines of sorghum. They observed consistent qualitative and quantitative differences between the lines and opined that these techniques are useful in characterizing male-sterile lines.

Sastry et al. (1986) investigated the genetic variability of storage proteins in the grains of several inbreds, hybrids and varieties representing different races of sorghum by IEF and HPLC. It was observed that grains of different inbreds usually contained different kafirin and alcohol soluble proteins. The hybrids contained proteins inherited from both parents with those from the female predominating. Representatives of the different sorghum races varied in their component alcohol soluble proteins, those most widely removed from the geographical center of origin showing greater differences.

### 2.1.7 Maize

Wall et al. (1984) observed that the electrophoretic patterns of zeins from extracts of different inbreds have different components and hybrids contain polypeptides from both parents, but the female parent’s contribution was quantitatively greater. They also recorded that inbreds that were derived from crosses usually contained some zeins from both parents. They concluded that electrophoretic patterns can characterize zeins in corn inbreds and can aid in relating corn hybrids and inbreds to parental lines.
Wilson (1985) studied endosperm zeins of different corn inbreds by IEF on agarose gels, which were resolved into 30 components. Each inbred exhibited 7 to 12 bands, which were useful for comparing inbreds, varieties and hybrids. A nomenclature based on IEF patterns was described. Closely related inbreds were often very similar but some differences occurred.

Smith and Wych (1986) used electrophoretic and morphological traits to estimate the percentage of female selfed plants in seed lots of four hybrids maize cultivars at contamination levels of 0, 3, 6 and 12%. Field determination of contaminants were carried out at seedling, anthesis and maturity stages while electrophoretic tests were performed on coleoptile tissue from 5-days old seedlings. Results based upon the morphological comparison at seedling and anthesis stages were highly inaccurate. Field comparisons at maturity were generally accurate. However, electrophoresis results were consistently more accurate with added advantages of speed and ability to determine the percentage of female sibs prior to harvest.

2.1.8 Soybean

Seed proteins of 61 soybean varieties were analyzed by Larsen (1967) on disc gel electrophoresis. The studies revealed two components of stained proteins that separated the varieties into two major groups. Component A was present in 13 varieties and component B in 48 varieties. In no instances were A and B observed in a single variety. Further Larsen and Caldwell (1968) reported that when a variety with A seed protein was crossed with a variety with B seed protein, the seed from the cross contained both proteins but at lower concentrations of each than when occurring alone. Selfed F₁ plants produced F₂ seed of A, AB and B types in ratio 1:2:1 respectively. The data indicated that these proteins are controlled by a pair of co-dominant alleles at a single locus.
Chapter 2: Review of Literature

By electrophoresis of seed proteins on gradient polyacrylamide slab gels, Lowry et al. (1974) detected differences between Lee and Pickett varieties of soybean and mutants.

Orf et al. (1980) studied PAGE of seed protein extracts from 3338 soybean accessions. They reported that the technique was useful in classifying soybean cultivars into broad categories.

Chauhan et al. (1985) investigated electrophoretic variations in proteins and enzymes in dry and germinated soybean and barley seeds by PAGE. Their results indicated that electrophoretic profiles of proteins and enzymes are species specific and can act as markers in assessing seed quality.

2.1.9 Grain legumes

Characterisation of 32 cultivars of Phaseolus vulgaris was done by densitograms of water soluble proteins separated by SGE (Adriaanse et al., 1969). This study also indicated that the protein patterns are unlikely to be affected by external conditions such as nitrogenous fertilization, climatic conditions and soil properties.

Ziegenfus and Clarkson (1971) reported that PAGE of salt soluble proteins was more sensitive than immuno-electrophoresis and double diffusion in detecting differences among the 7 taxa investigated in the genus Acer.

Blagrove and Gillepse (1978) examined the extracts from 80 pure lines of winged bean (Psophocarpus tetragonolobus) by PAGE and observed differences in the proportion of one seed globulin polypeptide among the lines studied.
Barratt (1980) examined the globulin fraction from *Vicia faba* seeds of different cultivars and different lines of the same cultivar by SDS-PAGE. The technique separated the globulin polypeptides into distinct groups termed as A, B, and C. Distinct differences between cultivars and the different lines of the same cultivar were observed in group A and B polypeptides with little variation in group C. Variation in polypeptide composition of different cultivars was also detected in respect of virilin and legumin by SDS-PAGE.

Wolf (1980) investigated the seed albumins of several members of the family *Papilionaceae* by gel electrophoresis. He found relationships between mutants, varieties and sub-species. More distant related ones did not show similarities in banding pattern. He also observed that the environment had no influence on the albumin pattern if genetically identical material of *Pisum sativum* was grown in different places and different vegetation periods.

Przybylska (1986) reported that the electrophoretic patterns of the legumin fraction obtained by means of urea-PAGE and IEF, and isoenzyme phenotypes were useful in discriminating between and within collection samples of *Pisum*.

Hussain *et al.* (1986) studied seed proteins by SDS-PAGE to discriminate 7 different cultivars of field beans. Their results indicated that sufficient variation was present among the 7 cultivars to afford their unambiguous identification.

### 2.1.10 Groundnut

Dawson and McIntosh (1973) analysed arachin and conarachin fractions extracted from 8 samples of groundnut varieties grown in different locations. They observed marked variations in the electrophoretic patterns, especially of the conarachin fractions. Amino acid composition of the fractions showed highly significant differences.
2.1.11 Forages

In their studies on 30 alfalfa varieties and 26 species accessions of *Medicago sativa* and *M. falcate*, Bingham and Yeh (1971) observed that protein patterns, in general, were basically similar in cultivated and wild materials and at diploid and tetraploid levels. One major band was present in *M. falcate* and absent in pure *M. sativa*. Further they separated the cultivated varieties into 4 major groups on the basis of the absences or presences and density of one main band.

Wilkinson and Beard (1972) could distinguish 6 creeping bent grass (*Agrotis palustris*) cultivars by their leaf protein banding patterns. Out of the 10 Kentucky blue grass (*Poa pratensis*) cultivars, 6 could be placed into two groups and 2 could be identified singly, while 2 showed no characteristic banding pattern. Hence they suggested that cultivar identification need not rely on electrophoretic pattern alone.

Dalling *et al.* (1979) separated seed globulin proteins of 17 subterranean clover (*Trifolium subterraneum*) cultivars using PAGE. They observed that each cultivar had a unique banding pattern and suggested that the technique could be useful to the plant breeder for cultivar identification and determination of genetic purity of a seed sample.

On the basis of the disc gel electrophoretic analysis of seed proteins and enzymes of diploid and tetraploid species of Italian rye grass (*Lolium multiflorum*) and perennial rye grass (*Lolium perenne*), Nakamura (1979) recommended esterase analysis for the separation of rye grass species.
De Prins and Van de Weghe (1983) characterized some Italian rye grass (*Lolium multiflorum*) and perennial rye grass (*Lolium perenne*) cultivars by seed protein and esterase banding patterns on pH gradient polyacrylamide gels. They also suggested that the technique could be used for rapid varietal identification. Further, they opined that densitometric scanning of the gels may make identification much easier and possibly allow detection of mixtures of cultivars.

### 2.1.12 Poinsettia

Soluble leaf proteins from 18 cultivars of poinsettia were analyzed by Werner and Sink (1977) using disc gel electrophoresis. They reported that the soluble protein banding pattern was the same for all of these cultivars.

### 2.2 Study of esterase profiles

#### 2.2.1 *Gossypium* species

In PAGE analysis of esterase (EST) isozymes in dormant seeds of 29 species and 29 varieties in genus *Gossypium*, Cherry and Katterman (1971a) observed that species and/or varieties within each genome had banding patterns more similar to one another than to members in other genome groups. Only minor differences distinguished A and B genome species whereas band variations were greater between more distantly related than closely related species in the C, D and E genome groups. *G. longicalyx* (formerly considered an E genome species) showed an overall banding pattern unique to itself. These results further supported the classification of the diploid species into the presently recognized six-genome groups A to F. Variation was observed also for leucine amino peptidases (LAP) and catalases (CAT) between species of different genome groups (Cherry et al., 1972), though within species polymorphism noted for EST was not observed for LAT and CAT.
Cherry and Katterman (1971b) observed six different EST isozymograms for single seed samples of four natural populations of G. thurberi and suggested that intra-specific variation should be evaluated before any inter-specific analyses were undertaken.

### 2.2.2 Barley

Kahler and Allard (1970) studied SGE of 30 parental lines of barley. They studied F1 and F2 progenies from crosses between selected parents and progenies obtained by selfing individuals from advanced generations and concluded that esterase isozymes were useful research tools for determining allelic variability in barley populations.

Fedak and Rajhathy (1972a) analyzed 55 Canadian barley cultivars and some parental genotypes by SGE for EST isozyme patterns. Each cultivar was assigned to one of 9 different patterns observed. They reported that polymorphism was a varietal characteristic and apparently not associated with age of cultivar or area of adaptability.

Fedak and Rajhathy (1972b) studied isozyme patterns in hybrid barley and found that hybrid plants contain the sum of their parental bands for PER [peroxidase], EST and AMY [amylase] isozymes. Genetic studies of EST and AMY isozymes revealed that the additive type of gene action governs band variation between parents, while combination of epitasis and over-dominance type of gene action were observed in the hybrids.

Out of three isozyme systems (viz. EST, PER and ACP) studied in shoot extracts of barley by Bassiri (1976), EST alone could be used to differentiate among most of the cultivars.

24
Andersen (1982) examined 47 barley cultivars for AMY isozymes in sprouted grains and EST in the first developed leaf. He concluded usefulness of the AMY and EST isozymes as a valuable supplement in identification of barley cultivars in the laboratory.

2.2.3 Oats
Singh et al. (1973) investigated 10 cultivars of oat for the isozymes of EST, LAP and PER. They observed that each variety had a stable and characteristic isozyme pattern, which makes this technique useful for varietal identification.

2.2.4 Wheat
Cubadda et al. (1975) studied the EST isozyme banding patterns in caryopses of *Triticum aestivum*, *T. durum* and *T. timopheevi* by gel electrophoresis. In *T. aestivum* 17 bands were ascertained in pH 5-8 range. Eleven were of high intensity, four were weak and two were very weak. Using Chinese Spring nullitetrasomic lines, it was possible to locate the genes controlling several isozymes on the chromosomes of homologous group III.

2.2.5 Rice
Isozyme polymorphism for EST was studied by Nakagahra et al. (1975) in 776 native varieties of *Orzyza sativa* collected from Asian countries. EST isozymes occurred in altogether 14 bands of which 9 were easily distinguishable. This indicated the presence of substantially complex variation. Each isozyme differed in its frequency of occurrence in each of 8 areas extending from Sri Lanka to Japan.

Second (1982) studied isozyme patterns of 13 enzymes including EST, MDH, ADH, CAT and ACP in the leaf blade extracts of 1948 strains of *O. sativa*, *O. glaberrima* and *O. breviligulata* by SGE. The polymorphism of the banding patterns was highest in *O. sativa*. 

25
2.2.6 Grain legumes

Based on the study of EST and PER isozyme patterns by SGE, Bassiri and Rouhani (1977) could identify 40 broad bean cultivars and reported that the differences in banding patterns among cultivars of the same region were as great as those of the cultivars from unrelated regions.

Using SGE, Bassiri and Adams (1978) studied EST, ACP and PER isozyme patterns in primary leaves, stems and roots of 13 species of the genus *Phaseolus*. Most species were observed to exhibit unique banding patterns in different tissues for each isozyme system.

Among the isozyme systems studied in primary leaf, stem and root tissue from seedlings of 34 United States common bean cultivars by Bassiri and Adams (1978), PER and EST were found suitable for cultivar identification while ACP could not be considered as a good system for this purpose.

Mahmoud *et al.* (1984) examined three isozyme systems namely AMY, EST and GOT in seeds of peas. Clear variation was observed in banding patterns between different lines. Their investigations revealed that each of the isozyme systems studied is genetically controlled by co-dominant alleles at a single locus.

Janardhanan *et al.* (1986) studied changes in isozyme profiles of 5 enzymes that occur during seed development and germination in wild pulse *Mucuna utilis* by PAGE. The isozyme patterns towards the end of seed maturation and during early germination were more or less identical and could be used effectively in cultivar identification.
Chapter 2: Review of Literature

Mittal et al. (1987) used isozymic variation for estimating genetic diversity in green gram. Eight varieties were characterized for EST, AMY, ACP, CAT and PER isozymes. ACP and EST exhibited more polymorphism than the other isozymes. Differences in these systems were used to calculate the diversity index (DI).

2.2.7 Soybean

Payne and Koszykowski (1978) analysed seed extracts of 44 soybean cultivars by PAGE for EST activity. They consistently detected five EST isozymes in all cultivars. Quantitative differences were observed among cultivars in the activity of two iso-esterases, which could serve as an aid in identifying soybean cultivars.

2.2.8 Forages

Spoor and Hay (1979) investigated the use of PAGE for identification of cultivars of Poa pratensis. Among 16 lines and 14 ecotypes studied by them, considerable variation was observed in the banding pattern of the seed EST and PER isozymes.

Quiros (1980) studied the applicability of SGE in the identification of alfalfa mother plants. He observed that the zymograms of PER, EST and ACP from leaf tissues were useful in ‘fingerprinting’ each of the 21 mother plants tested. He concluded that SGE is a simple, inexpensive and reliable technique for the identification of alfalfa plants.

2.2.9 Maize

Hunter and Kannenberg (1971) used isozyme variability as an indicator of genetic diversity in maize, characterizing 15 inbred lines for EST, AMY, LAP, CAT, ADH [alcohol dehydrogenase] and PER isozymes. Differences in these isozyme systems were used to calculate diversity index.
Goodman and Stuber (1980) studied the fingerprinting of maize inbred lines surveyed by them in USA and Canada by analyzing 13 isozyme systems including EST, MDH [malate dehydrogenase], ADH, CAT and ACP in coleoptiles of 5 days old seedlings. They opined that laboratory genotyping or isozyme fingerprinting can be used for establishing cultivar identity or uniqueness and can complement, if not substitute, field testing of various seed lots.

2.2.10 Sorghum

To compare different male-sterile, maintainer and restorer lines of sorghum, Tripathi et al. (1981, 1983) studied EST, PER, ADH and GDH isozyme patterns by PAGE of anther and seed extracts. Based on EST isozyme patterns and densitographs, the diverse male-sterile lines with different cytoplasms were classified into three groups. Each group was further differentiated on the basis of minor differences in EST isozyme pattern. PER patterns showed characteristic, but not clear-cut, differences. ADH and GDH isozyme patterns in general were similar in both male-sterile and maintainer lines.

Tripathi et al. (1982) studied isozyme pattern of seed and anther extracts of 24 cytoplasmic male sterile sorghum lines using PAGE and IEF. EST isozyme patterns of milo-based male-sterile lines differed from male-sterile lines having cytoplasms of diverse Indian origin and thus served as aid in characterizing male-sterile sorghum stocks.

2.2.11 Brassica species

Arus et al. (1985) analyzed six isozyme genes in seed samples of 65 commercial F₁ hybrids of four horticultural groups of B. oleracea. Their
Chapter 2: Review of Literature

results indicated that it was possible to determine purity in 91% of hybrids analyzed and 74% of hybrids were individually distinguishable by their isozyme phenotypes.

Kumar and Gupta (1985) made a study of seed PER and EST isozyme patterns by SGE to characterize 12 genotypes of Indian mustard. They ascertained that these patterns, especially the staining intensity of bands, could be used successfully to characterize different Indian mustard genotypes.

Schenck and Wolf (1986) obtained three somatic hybrids by fusion of protoplasts of *B. oleraceae* and *B. campestris*. Gel electrophoretic patterns of the fusion products were compared with their respective parental species. By comparing multiple forms of EST and phosphorylases, they demonstrated that in all the cases the hybrid plants contained one or more enzymes from each parent.

### 2.2.12 Cassava

Hussain *et al.* (1987) developed a methodology for determining EST isozyme patterns by PAGE in viable root tissue extracts of 20 cassava cultivars. Each of 20 cultivars was characterized by a unique band pattern. The pattern was observed to be highly reproducible in band mobility and intensity. Further they reported that this method could be used to identify unknown cassava cultivars by comparing their EST isozyme pattern with those of known cultivars.

Ramirez *et al.* (1987) developed a methodology based on SGE and PAGE for determining isozyme electrophoregrams of 16 enzymes of cassava varieties as potential genotypic markers. Extracts of 5 different tissues (root, stem, leaf, petiole and bud) were examined. They observed that the nodal
portion of the shoots gave isozyme patterns with the largest number of bands. The limited number of varieties that were examined could be distinguished by sequential classification on the basis of isozyme patterns of ACP, EST, GOT and PGI. EST, which is a complex group of isozymes, showed a large variation in the fast migrating zone with little variation among different tissues.

2.2.13 Potato

Le (1994) analyzed esterase and peroxidase isozymes by means of isoelectric focusing in a number of varieties of potato cultured in vitro. Electrophoretic profiles obtained in the study were used for varietal identification by various countries.

2.2.14 Radish

Park et al. (1995) used the isozyme electrophoretic technique for cultivar identification and seed purity testing in radish, Chinese cabbage and their F₁ hybrids.

2.2.15 Nicotiana species

Smith et al. (1970) observed unique banding patterns of seedling root PER and seed EST in 61 and 55 species of Nicotiana, respectively. They reported that no single specific band was common to all Nicotiana species. Comparison of PER and EST zymograms of 17 artificially synthesized amphidiploids revealed that most of the bands had mobility similar to one or both parents, only 25% being new (possible hybrid) bands. They concluded that this method of band assessment is a reliable measure of genetic similarity.
Wilkinson et al. (1985) performed isozyme separation using PAGE for EST, CAT, PER and MDH from leaf samples of 10 tobacco cultivars. Leaf material from the earliest maturity dates exhibited the highest number of isozymes and the greenhouse environments produced fewer bands per enzymes than the field environment. No cultivar differences were observed for EST; however, cultivar groupings were observed for other three enzymes investigated.

### 2.3 Use in Plant Breeder’s Rights (PBR)

Electrophoretic characters have been accepted as a special test of distinctness in at least three cases of PBR. The one International Union for Protection of New Plant Varieties (UPOV) member state to officially grant protection to the three varieties is Sweden.

Red fescue variety ‘Satin’ is morphologically indistinguishable from two other varieties, ‘Dawson’ and ‘Polar’. Using the technique of IEF, seed EST isozymes were separated and the zymograms of the three varieties compared. ‘Satin’ zymograms lacked a prominent high-pH band present in the zymograms of both ‘Dawson’ and ‘Polar’. This isozymic difference was accepted as an evidence of varietal novelty and a grant of PBR was issued for ‘Satin’ in 1978 (Clapham and Almgard, 1978).

Barley variety ‘Pernilla’ was distinguished from morphologically similar ‘Gunilla’ by differences in gliadin banding patterns and rights were then granted for this cultivar in 1979. Red clover variety ‘Kora’ could not be clearly distinguished from its parent, ‘Disa’ and the cultivar ‘Reko’. However, general seed protein patterns of the cultivars were qualitatively dissimilar, allowing PBR to be granted for ‘Kora’ on 1980 (UPOV, 1980).
2.4 Study of DNA profiles

Limited genetic variation of a limited number of potential isozyme and protein markers limit their use in genotype identification and hybrid seed purity testing (Burrow and Blake, 1998). Isozymes and proteins were widely used as molecular markers up to the late 1980s. However, as the extent of knowledge about plant genomes and the technologies for unlocking the information contained within those genomes have increased rapidly, DNA methods have become an increasingly interesting future technology for cultivar identification.

As the methods for evaluating variation directly at the DNA level became widely available during the mid 1980s, DNA-based markers replaced isozymes and proteins as the markers of choice. Since then molecular biology has ushered in a new era with techniques that directly assay DNA sequence variation and overcome many of the problems that have previously limited the applied use of biochemical markers.

DNA analysis has a number of advantages as a source of genetic markers for cultivar identification:

- The vast genomes of plants provide an essentially infinite range of possibilities for the design of DNA markers that differentiate among varieties.

- The DNA of an organism is identical in all parts of the plant (with some exceptions like triploid endosperm of a cereal seed, which has a different genetic constitution than either the diploid maternal tissue of the pericarp or the diploid zygotic tissue of the embryo of that same seed).

- There is no effect of environment on DNA sequence.
DNA analysis can be highly automated and be objectively scored (at least in theory).

It is possible to build up databases of DNA data so that the methodology is highly transportable between analytical laboratories across the world.

It is relatively simple to add information on new varieties or markers to an existing database.

DNA analysis can be applied early in the growth of the crop whereas some morphological characters may only be expressed in mature plants (Preston et al., 1999).

### 2.5 DNA markers

Many DNA-based molecular markers (DNA markers) reveal neutral sites of variation at the DNA sequence level. By 'neutral' is meant that, unlike morphological markers, these variations do not show them themselves in phenotype and each might be nothing more than a single nucleotide difference in gene or a piece of repetitive DNA. They have the big advantage that they are much more numerous than morphological markers and individually they typically do not disturb the physiology of the organism (Jones et al., 1997). Several types of DNA markers have been used widely for molecular marker studies. Recently Mohan et al. (1997) and Kumar (1999) had extensively reviewed the details of these markers (Table 2.1).

### 2.6 DNA markers in crops other than sorghum

#### 2.6.1 Maize

Stubber et al. (1987) demonstrated that in two maize F2 populations, COTX and CMT, isozyme marker loci could be effective in identifying and locating many QTLs affecting the expression of grain yield and 24 yield-related traits.
Smith et al. (1990) calculated genetic distances among 37 inbred lines representing a wide range of related and unrelated elite Corn Belt germplasm of maize using 257 probe-restriction enzyme combinations. This RFLP-based genetic distance had a high correlation with single-cross grain yield performance and grain yield heterosis. They suggested that the measures of similarity calculated from RFLP data, coupled with pedigree knowledge and using molecular markers to locate QTLs, could allow maize breeders to predict combinations of lines that result in high-yielding, single-cross hybrids.

### Table 2.1. Molecular marker techniques

<table>
<thead>
<tr>
<th>Techniques</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFLP</td>
<td>Amplified fragment length polymorphism</td>
</tr>
<tr>
<td>ALP</td>
<td>Amplicon length polymorphism</td>
</tr>
<tr>
<td>AP-PCR</td>
<td>Arbitrarily primed PCR</td>
</tr>
<tr>
<td>AS-PCR</td>
<td>Allele-specific PCR</td>
</tr>
<tr>
<td>CAPS</td>
<td>Cleaved amplified polymorphic sequence</td>
</tr>
<tr>
<td>DAF</td>
<td>DNA amplification fingerprinting</td>
</tr>
<tr>
<td>ISA</td>
<td>Inter-SSR amplification</td>
</tr>
<tr>
<td>RAPD</td>
<td>Random-amplified polymorphic DNA</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>SAP</td>
<td>Specific amplicon polymorphism</td>
</tr>
<tr>
<td>SCAR</td>
<td>Sequence characterized amplified region</td>
</tr>
<tr>
<td>SSCP</td>
<td>Single-strand conformation polymorphism</td>
</tr>
<tr>
<td>SSLP</td>
<td>Microsatellite simple sequence length polymorphism</td>
</tr>
<tr>
<td>SSLP</td>
<td>Minisatellite simple sequence length polymorphism</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple sequence repeat</td>
</tr>
<tr>
<td>STS</td>
<td>Sequence tagged site</td>
</tr>
</tbody>
</table>

### 2.6.2 Wheat

He et al. (1992) detected DNA sequence polymorphism among different genotypes of wheat using RAPD markers. Of the 65 primer combinations used for PCR amplification, over 38% of them produced readily detectable
and reproducible DNA polymorphism between ‘spring’ wheat line ‘SO852’ and ‘winter’ wheat variety ‘Clark’. A high level of polymorphism was observed among a number of commercial varieties and breeding lines of wheat.

2.6.3 Rice

Virk et al. (1995) subjected a set of accessions of *Oryza sativa* from IRRI (Philippines), that included known and suspected duplicates as well as closely related germplasm, to RAPD analysis. The results indicated the accurate discrimination of true and suspected duplicate accessions could be made using this method.

Temnykh et al. (2000) reported that 312 microsatellite markers provide whole genome coverage with an average density of one SSLP per 6 cM in rice (*Oryza sativa* L.). Twenty-six SSLP markers were identified in published sequences of known genes, 65 were developed based on parental cDNA sequences available in GenBank, and 97 were isolated from genomic libraries.

Raghunathachari et al. (2000) applied 10 RAPD primers to 18 accessions of Indian scented rice (*Oryza sativa* L.). A total of 144 different marker bands were produced, of which 95% were polymorphic. Thus, with these ten selected RAPD primers sufficient polymorphism could be detected to allow identification of each of the individual accessions.

Gupta et al. (2001) analyzed 16 isolates of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* representing different geographical locations in India along with 2 isolates from Philippines using polymorphic RAPDs. The data using RAPD-PCR and IS- (Insertion Sequence)-PCR approaches revealed their potential in rapid identification of isolates, in assessment of genetic variation in the Indian pathogen population and in generating unique DNA fragments specific to 8 isolates of *X. oryzae* pv. *oryzae*.
2.6.4 Lentil
Sharma et al. (1995) used 24 RAPD markers to distinguished between 6 different Lens taxa representing cultivated lentil and its wild relatives. These generated a total of 88 polymorphic bands in 54 accessions and were used to partition variation within and among Lens taxa.

2.6.5 Strawberry
Graham et al. (1996) used RAPD markers to examine the genetic relatedness of 8 strawberry cultivars released from four breeding programmes around the world. Ten RAPD primers successfully amplified DNA fragments from each cultivar and specific fingerprints were generated from the molecular marker data.

2.6.6 Olive
Belaj et al. (2003) used RAPD markers for the study of 19 Albanian olive cultivars and two wild olives (Oleasters). A total of 76 polymorphic bands out of 107 reproducible bands were obtained using 16 primers. The number of polymorphic bands detected by each of the 16 primers ranged from 1 to 9. All accessions could be identified by the combination of four primers: OPA-19, OPA-02, OPK-16 and OPP-19.

2.6.7 Mango
Karihaloo et al. (2003) carried out RAPD analysis in 29 Indian mango cultivars comprising popular landraces and some advanced cultivars. PCR amplification with 24 primers generated 314 bands, 94% of which were polymorphic.

2.6.8 Orange
Abkenar and Isshiki (2003) used RAPD markers to evaluate genetic similarity and inter-relationships among 31 acid citrus species and cultivars.
including sour oranges (6 accessions), Yuzu (4 accessions) and its relatives (21 accessions). Out of the 60 decamer primers screened, 27 were selected that produced 108 bands; 76 of which were polymorphic. Species-specific and cultivar-specific RAPD markers were also found.

2.7 DNA markers in sorghum

Binelli et al. (1992) utilized cDNA and genomic clones from maize to initiate the construction of a RFLP-based linkage map in *Sorghum bicolor*. A total of 159 maize clones were hybridized to genomic DNA of the two sorghum parents in order to detect polymorphism: 154 probes hybridized to sorghum and 58 out of these were polymorphic. In almost all cases hybridization patterns were similar between maize and sorghum.

Berhan et al. (1993) used cloned maize genes and random maize genomic fragments to construct a RFLP-based genetic map of sorghum and to compare the structure of the maize and sorghum genomes. Most (266/280) of the maize DNA fragments hybridized to sorghum DNA and 145 of them detected polymorphism.

Tao et al. (1993) used RAPD and AFLP markers to determine the frequency of DNA polymorphism in cultivated grain sorghum (*Sorghum bicolor*). Similar frequencies of polymorphism to that obtained with RAPDs were obtained with RFLPs. Results from these experiments indicated that a high level of genetic uniformity exists within the range *S. bicolor* cultivars evaluated.

Chittenden et al. (1994) described a detailed RFLP map of a small (56 individuals) *S. bicolor × S. propinquum* F₂ population, linked to an estimated 93% of the genome, with markers at average spacing of 5.2 cM. This map
Chapter 2: Review of Literature

provided evidence supporting ancestral duplication of chromosomes or chromosome segments in the evolution of modern-day “diploid” (n = 10) sorghum.

Pammi et al. (1994) used primers to amplify RAPDs in 32 genotypes of sorghum. Under standard conditions, approximately 80% of the primers tested amplified DNA and most revealed 1-5 polymorphisms between BTx623 and IS 3620C.

Xu et al. (1994) constructed a RFLP-based linkage map of sorghum composed principally of markers detected with low-copy-number nuclear DNA clones from sorghum. The map spanned 1789 cM and consisted of 190 loci arranged into 14 linkage groups. Frequency of RFLP was found to be 51%. A minimal estimate of the number of clones that detect duplicate sequences was 11% and null alleles occurred at 13% of the mapped RFLP loci.

Deu et al. (1994) performed RFLP analysis on 94 sorghum varieties, chosen to represent the main cross combination (race × geographic origin). A total of 50 polymorphic probe-enzyme combinations yield 158 polymorphic bands. The bicolour race appeared highly variable (as is expected of this race, which is essentially comprised of all small-seeded variants in the cultivated species that have substantial glume coverage). The apparent geographic differentiation was related to the contrasting distribution of these races and to a greater similarity between races localized in Southern Africa.

Pereira and Lee (1995) used RFLPs to determine the genetic location and effects of genomic regions controlling plant height in sorghum using 152 F2 plants from the cross CK 60 × PI 229828. Tallness was dominant and
Chapter 2: Review of Literature

conferred by alleles from PI 229828 for three QTLs. At the 4\textsuperscript{th} QTL, the allele from PI 229828 contributed to increased plant height but short stature was partially dominant.

Dufour \textit{et al.} (1997) constructed a sorghum composite linkage map with two recombinant inbred line populations using heterologous RFLP probes. This map includes 199 loci revealed by 188 probes and distributed on 13 linkage groups.

Menkir \textit{et al.} (1997) used RAPD markers to assess genetic diversity and taxonomic relationships in 190 accessions sampled to represent the cultivated races of sorghum [\textit{Sorghum bicolor} (L.) Moench]. A high level of variation was detected among the cultivated genotypes. Partitioning the genetic variation in cultivated sorghum with Shannon's diversity index revealed that 86\% of the total genetic variation occurred among accessions and 14\% among races.

Tao \textit{et al.} (1998) constructed a genetic map using 120 F\textsubscript{5} sorghum RILs developed from a cross between two elite Australian sorghum inbred lines, QL 39 and QL 41. A total of 155 RFLP loci and 8 SSR loci were mapped on 21 linkage groups covering a map distance of approximately 1400 cM.

Boivin \textit{et al.} (1999) prepared a near-saturated sorghum genetic linkage map using RFLP and AFLP markers. The RFLP-based map includes 343 loci distributed across 11 linkage groups spanning 1352 cM. The AFLP map includes 1378 loci distributed across 11 linkage groups spanning 849 cM. Taking into account the different data sets, they constructed a combined genetic linkage map including 443 loci spanning 1899 cM.
Peng et al. (1999) constructed a RFLP linkage map of Sorghum bicolor (L.) Moench in a population of 137 F$_6$ RILs using sorghum, maize, oat, barley, and rice DNA clones. The resulting map consisted of 10 linkage group and 323 marker loci.

Subudhi et al. (2000) reported the consistency of QTLs controlling stay-green in sorghum. They also identified the QTLs regions and marker loci showing significant epistatic interaction in the expression of the stay-green trait.

Haussmann et al. (2002) constructed a combined sorghum linkage map from two RIL populations using AFLP, SSR, RFLP and RAPD markers and compared this with other sorghum maps. It was in good agreement with the other sorghum linkage maps, from which it deviated by a few apparent inversions or deletions and additional distal regions.

Menz et al. (2002) constructed a high-density genetic map of the sorghum genome using AFLP technology and a RIL population derived from the cross BTx623 × IS 3620C. The 1713 cM map encompassed 2926 loci distributed across 10 linkage groups; 2454 of those loci are AFLP products, 203 are RFLP markers detected by cDNA and genomic clones from rice, barley, oat, and maize, and 136 are SSRs previously mapped in sorghum.

Bowers et al. (2003) reported a genetic recombination map for sorghum of 2512 loci spaced at average 0.4 cM intervals based on 2050 RFLP probes, including 865 heterologous probes that foster comparative genomics of sugarcane, maize, rice, pearl millet, buffel grass, the Triticae
(wheat, barley and oats), rice and *Arabidopsis*. Mapped loci identify 61.5% of recombination events in the small mapping progeny set of 65 F$_2$ individuals derived from the interspecific cross of BTx623 × *S. propinquum* and reveal strong positive cross-over interference acting across intervals of ≤50 cM.

### 2.8 SSR markers

Although known by many names and acronyms, including simple tandem repeats (STRs), variable tandem repeats (VTRs), microsatellites, and simple sequence repeats (SSRs), SSRs have received considerable attention and are probably the current marker system of choice for marker-based genetic analysis and marker-assisted plant breeding (Akkaya *et al.* 1992; Chin *et al.* 1996). SSRs are co-dominant, occur in high frequency and appear to be distributed throughout the genomes of most if not all higher plants and animals. They also display high levels of polymorphism even among closely related accessions and are amenable to simple and inexpensive PCR-based assays (Brown *et al*., 1996).

The repeat regions are generally composed of perfectly repeated di-, tri-, tetra-, and sometimes greater length, nucleotide sequences (Tautz and Ranz, 1984) that exhibit a high degree of polymorphism (Weber and May, 1989). Variability in the number of repeat units is the typical basis of observed polymorphism. The high degree of polymorphism is thought to be due to increased rates of sequence mutation affecting the number of repeat motifs present at an SSR locus with the observed variation likely due to replication slippage or unequal crossing over (Edward *et al*., 1992).
In plant genomes, the overall frequency of microsatellite repeats appears to be lower than animal genomes (Morgante and Oliveri, 1993; Wu and Tanksley, 1993); although the incidence of closely spaced repeats has been borne out experimentally (Gupta et al., 1994; Zietkiewicz et al., 1994). In humans, AC or TC are very common repeat units but in plants AT is more common followed by AG or TC. In general, plants have about 10 times less SSRs than humans (Mohan et al., 1997).

Unique sequences that flank the tandem repeats can be used for making PCR primers. There are well-established methods of finding microsatellites by screening phage libraries with oligonucleotide probes. But a quicker, if limited, approach is to examine DNA sequence data banks for their presence (Burr, 2001). SSR-based primers representing tri-, tetra-, and penta- nucleotide repeats have been used successfully to generate distinct banding patterns that are resolvable on low-resolution agarose gels using ethidium bromide staining (Gupta et al., 1994; Weising et al., 1995), through primer radio-labeling followed by auto-radiography (Gupta et al., 1994), or through primer labeling with fluorescent dyes and automated high-resolution visualization of PCR products separated by PAGE or capillary electrophoresis. As would be predicted, the best product size discrimination is obtained with capillary electrophoresis or polyacrylamide-based gel analysis, although agarose gels are sufficient for many applications (Vogel and Scolink, 1997). A good visualization and size discrimination of PCR-products is possible by silver staining polyacrylamide gels (Tegelstrom, 1992).
Chapter 2: Review of Literature

In any case SSRs are generally among the most reliable and highly reproducible of molecular markers. SSRs have several advantages over the other DNA markers such as RAPDs, RFLPs, or AFLPs. The advantages include the uniform genome coverage, high levels of polymorphism, codominance and locus-specific PCR-assays. As a result, SSR diversity has been successfully used as tool in genotyping and studying genotypic diversity of many plant species. In addition, they are useful in pedigree analysis because they represent single loci and can uniquely define genotypes. SSR loci give good discrimination between closely related individuals, in some cases even when only a few loci were employed (Ghebru et al., 2002). One of the advantages of microsatellite markers is that they can be shared between laboratories by simply exchanging primer DNA sequences. Thus microsatellite markers should provide for better information exchange (Rafalski et al., 1994). Because SSR markers typically detect only single genetic loci that are co-dominantly inherited, and are highly polymorphic, they can be extremely informative in pedigree tracing studies, in the analysis of progeny from multiparent matings, in a wide range of mapping applications and genotype identification (Akkaya et al., 1995). These markers can require a considerable investment to generate but are then inexpensive to use in any molecular analysis. The large start up costs for this technique should be justifiable for crops where large-scale mapping and MAS (marker assisted selection) are a practical necessity (Hash and Bramel-Cox, 2000).

Markers can be used to prove the uniqueness of a cultivar, a characteristic that is essential in protecting the intellectual property that it
Chapter 2: Review of Literature

represents with PVP or patents. Currently, most breeders are using simple sequence repeats (SSR) markers because they are easy to use, plentiful for most crop species in which the necessary DNA sequencing investments have been made, and are quite polymorphic and as informative as older RFLP markers and perhaps one tenth as expensive (Sneller, 2002). A combination of SSR and morphological descriptors showed the best compromise regarding genetic relationships and the needs of clear classification for PVP and may help to establish minimum genetic distances for distinctness with PVP Office definition (Giancola et al., 2002).

Inter simple sequence repeat (ISSR)-PCR is a simple and quick method that combines most of the advantages of SSRs and AFLPs with the universality of RAPD markers. ISSR markers are highly polymorphic and are useful in studies on genetic diversity, phylogeny, gene tagging, genome mapping and evolutionary biology. ISSR will have an important role in securing plant variety rights by virtue of its unique efficiency in distinguishing even closely related germplasm (Reddy et al., 2002).

2.9 SSR markers in crops other than sorghum

2.9.1 Soybean

Akkaya et al. (1992) studied presence and degree of simple sequence repeats (SSRs) DNA length polymorphism in soybean.

Rongwen et al. (1995) assessed the ability of SSR length polymorphism at a small number of loci to distinguish individuals in a group of 96 diverse soybean genotypes. The objective of the work was to provide an initial evaluation of SSR markers to develop unique DNA profiles of soybean. The set of microsatellite markers used provides a positive
assessment of the ability to produce unique DNA profiles of soybean genotypes. They also concluded that with careful selection of microsatellite loci for high levels of polymorphism it is very likely that 10 or 15 loci may be adequate to distinguish closely related cultivars arising from similar pedigrees.

Cregan et al. (1999) examined two sets of soybean BAC clones to uncover new SSRs associated with previously characterized RFLP, AFLP or STS markers. They found five or more \((\text{ATT})_n\) and \((\text{AT})_n\) microsatellite sequences for each set of BAC clones and developed 3 new polymorphic SSR markers for each linkage group.

Bommi and Ferguson (2005) used eleven SSR markers to distinguish between four elite soybean cultivars using high-resolution agarose gel electrophoresis. Of the 11 SSR markers evaluated, five markers exhibited visible polymorphism between the four cultivars. Two SSR markers were found that could distinguish between all four cultivars.

### 2.9.2 Maize

Chin et al. (1996) made a database survey of 576 maize sequences from the GenBank and EMBL databases to determine the abundance of maize microsatellites. They identified 200 potential microsatellites. Using high-resolution agarose gels, they showed that 69 of 200 potential microsatellites were polymorphic and yielded 2-4 alleles each.

Despite the economic advantages of AFLPs, the French association of maize breeders (SEPROMA) recently recommended a set of uniformly
distributed SSR markers for essentially derived varieties (EDVs) identification. SSRs provide a higher degree of transparency for legal issues than AFLPs due to their co-dominant inheritance, their known map positions and their public availability (Heckenberger et al., 2003).

### 2.9.3 Rice

Xiao et al. (1996) examined the relationship of genetic diversity with hybrid performance and heterosis to assess whether such PCR-based markers (SSR and RAPD) were useful for evaluating germplasm and predicting F$_1$ performance and heterosis in rice. Results indicated that genetic distance measures based on RAPDs and SSRs may be useful for predicting yield potential and heterosis of intra-subspecies but not inter-subspecies hybrids.

Cho et al. (2000) derived 194 microsatellite markers from genomic library screening and 129 microsatellite markers from analysis of rice-expressed sequence tags (ESTs) available in public DNA databases for comparing polymorphism. They found that microsatellites derived from genome libraries detected a higher level of polymorphism than those derived from ESTs contained in GenBank databases (83.8% versus 54.0%).

Yashitola et al. (2002) had screened several microsatellite and STS markers to identify polymorphisms that distinguish certain CMS and restorer lines and their hybrids in rice. They also studied usefulness of these polymorphisms for determining hybrid seed purity.

Ni et al. (2002) used 111 microsatellite markers distributed over the entire rice nuclear genome for evaluating genetic diversity and determining
cultivar identity of 38 rice cultivars of particular interest to U.S. breeding programs and two wild species accessions. The results suggested that a relatively small number of microsatellite markers could be used for the estimation of genetic diversity and identification of rice cultivars.

2.9.4 Wheat

Microsatellites, also termed simple sequence repeats (SSRs) have been proposed as one of the most suitable marker systems for assessment of genetic variation and diversity among wheat varieties/lines, because they are multiallelic, chromosome specific and evenly distributed along chromosomes (Roder et al., 1998a,b).

Using a set of only 12 primer pairs of SSR markers, Prasad et al. (2000) distinguished 48 of the 55 wheat genotypes evaluated. The results demonstrated the utility of microsatellite markers for detecting polymorphism leading to genotype identification and for estimating genetic diversity.

Manifesto et al. (2001) used a selected subset of 10 highly informative SSR markers to construct an identification matrix that allowed the discrimination of 105 wheat cultivars. The identification matrix based on these SSRs provides a rapid and reliable method for cultivar identification that might be used for quality control in certified seed production programmes, to identify sources of seed contamination, and to maintain pure and clean germplasm collections.
Chapter 2: Review of Literature

Huang et al. (2002) used a set of 24 wheat microsatellite markers for characterization and assessment of genetic diversity in 998 accessions of hexaploid bread wheat. Comparative analysis of microsatellite diversity among eight geographical regions revealed that the accessions from the Near East and the Middle East exhibited more genetic diversity than those from the other regions. Huang also reported that SSR markers permit the fast and high throughput fingerprinting of large numbers of accessions from a germplasm collection in order to assess genetic diversity.

2.9.5 Grape

Moreno et al. (1998) carried out ISSR-PCR on 12 clones of grapevine (Vitis vinifera L.) variety ‘Garnacha’ and one accession of a related variety attempting to detect polymorphism within a variety that they had been unable to reveal previously with RAPDs, and to quantify the intra-laboratory reproducibility of the assay. They couldn’t detect any intra-varietal differences with ISSR-PCR. However, the ISSR approach revealed highly reproducible polymorphism between the two varieties tested. The reproducibility of the banding patterns ranged from 86% to 94%.

2.9.6 Picea

Scotti et al. (2000) demonstrated that microsatellite repeats also occur in expressed regions in Picea.

Yazdani et al. (2003) found many polymorphic co-dominant highly-repeatable SSR microsatellite markers in Picea abies and concluded that they can be used for population studies, seed certification, quality control of controlled crosses, paternity identification, pollen contamination assessment, and mapping of QTLs in related families.
A'Hara and Cottrell (2004) developed microsatellite markers in Sitka spruce using over 10,000 sequences of expressed regions of DNA from *P. glauca* in GenBank. He found nine polymorphic loci that demonstrated Mendelian inheritance in Sitka spruce.

### 2.9.7 Barley

DNA-based varietal analysis has now moved from being a research tool to become a routine analytical method (Henry, 2001). Henry et al. (2001), have developed microsatellite markers for barley that successfully distinguish cultivars—Stirling, Fitzgerald and Gairdner from Harrington.

The Variety Identification Unit of the Canadian Grain Commission (CGC) have developed and implemented DNA-based methods to enhance barley variety identification capabilities. For six-rowed barley varieties, CGC use six microsatellite markers that are multiplexed into two PCR sets. For two-rowed barley varieties, CGC currently use 12 STS markers that are multiplexed into 3 sets (Demeke et al., 2003).

### 2.9.8 Cotton

Hybrid cotton ‘H 6’ and its parents G.Cot.10 (male) and G.Cot.100 (female) were studied for identification and to test the purity of commercial hybrid seed samples with three PCR-based molecular marker systems (RAPD, ISSR, and microsatellite markers) by Dongre and Parkhi (2005). Results indicated that using all three marker systems — RAPD, ISSR and SSR in combination is faster and more reliable than using the three in isolation, for the identification of cotton hybrids.
2.9.9 Miscellaneous

Gupta et al. (1994) evaluated the potential of SSR-based primers in amplifying DNA markers from plant and vertebrate DNAs. They used DNA of pine, grapes, lettuce, tomato, maize, salmon, chicken, cow and human to demonstrate that such primers produce polymorphic band patterns from complex eukaryotic genomes. These polymorphic bands were heritable and can be used as DNA markers.

Wang et al. (1994) searched for all possible STRs ranging from mononucleotide repeats up to tetra-nucleotide repeats on EMBL and GenBank DNA sequence databases of 3026 kb nuclear DNA and 1268 kb organelle DNA in 54 and 28 plant species (plus algae), respectively. An extreme rareness of STRs (4 STRs in 1268 kb DNA) was detected in organellar DNA compared with nuclear DNA sequences. In nuclear DNA sequences, (AT)$_n$ sequence repeats were most abundant.

Varshney et al. (2002) investigated available expressed sequence tags (ESTs) of cereal species like barley, maize, oat, sorghum, rice, rye and wheat for a study of abundance, frequency and distribution of various types of microsatellites. SSRs were present in about 7% to 10% of the total ESTs in the investigated cereal genomes.

2.10 Sorghum SSR markers

SSR-containing clones isolated from bacterial artificial chromosomes (BACs) and enriched genomic DNA (gDNA) libraries, and database sequences that contain SSRs were the sources for the primer sequences of sorghum SSRs mapped by Bhattramakki et al. (2000). Targeted isolation of SSR loci using
BAC clones as proposed by Cregan et al. (1999) is likely to be an efficient method for placing SSR loci in a particular segment of the genome. BTx623 (Frederiksen and Miller, 1972) is the reference genotype used for sorghum molecular marker genotyping and it was the source of the DNA used to construct the enriched libraries and the two sorghum BAC libraries that are currently available (Bhattramakki et al., 2000). PCR primers for the amplification of DNA fragments containing SSRs from sorghum were successfully developed through three different approaches by Brown et al. (1996) and they reported that polymorphic sorghum fragments can be amplified using at least some maize SSR primers.

Map locations of 46 previously reported SSRs (Taramino et al., 1997; Tao et al., 1998; Kong et al., 2000) and 113 novel SSR loci, including four SSR-containing gene loci (Bhattramakki et al., 2000), have been reported to date for sorghum. Map locations of additional sorghum SSR loci have been reported by Klein et al. (2001); Agrama et al. (2002); Schloss et al. (2002; via Chittenden et al., 1994 and Bowers et al., 2003); Tao et al. (2003); Haussmann et al. (2004); Casa et al. (2005); Folkertsma et al. (2005). SSR markers have been incorporated into existing RFLP-based sorghum genetic linkage maps of Xu et al. (1994) by Kong et al. (1997) and of Peng et al. (1999) by Bhattramakki et al. (2000). The number of polymorphic SSR loci available per sorghum linkage group ranges from 14 to 53 and many more are in pipeline (CT Hash, pers. comm.). Eight SSR loci that, although monomorphic among 18 survey accessions, have high degree of homology to known genes (Bhattramakki et al., 2000), have yet to be mapped. The
average number of alleles detected per locus at the polymorphic loci was 3.88. \((\text{AG/TC})_n\) and \((\text{AC/TG})_n\) repeats comprise the majority of these sorghum SSRs (52%) and 91% of the dinucleotide SSRs at these loci (Bhattramakki et al. 2000). It was found that as much as 57% of SSRs containing triplets rich in G-C base pairs were located in the gene coding regions of the total genomic DNA (Wang et al., 1994).

The estimated average probability that two accessions in a sorghum working group (sub-race) will have different alleles at a locus ranges from 0.88 to 0.67 depending upon the working group to which the accessions belong (Kong et al., 2000). In addition, the number of alleles per locus is positively correlated \((r = 0.68\) which is significant at 1% level) with the numbers of repeated units at the locus in BTx623, the strain from which the SSRs were originally isolated (Kong et al., 2000). This confirms that most \textit{Sorghum bicolor} SSR loci are sufficiently polymorphic to be useful in marker-assisted selection programmes (Kong et al., 2000).

Taramino et al. (1997) tested 13 SSR primers on 9 inbred lines of \textit{Sorghum bicolor} of different geographic origin to explore the degree of polymorphism. PCR analysis on acrylamide gels revealed a high degree of polymorphism (0.80). One locus in particular, \textit{SbAGE01}, allowed the identification of all 9 inbred lines used.

Dean et al. (1999) demonstrated that 15 SSR markers from the set developed by Brown et al. (1996) provided substantial genetic resolution among 19 'Orange' sorghum accessions. The SSRs employed in this analysis
had diversity indices ranging from 0.18 to 0.77. The marker set of 15 SSR loci detected two genetically redundant groups among the 19 Orange accessions evaluated.

Dje *et al.* (1999) investigated the level of genetic diversity and the population genetic structure of sorghum landraces from North-western Morocco based on direct field-sampling, using both allozyme and microsatellite markers. Microsatellite markers showed a much higher degree of polymorphism than allozymes, but relative measures of genetic structure such as Wright's inbreeding coefficient and Nei's coefficient of genetic differentiation were similar for the two sets of markers.

Dje *et al.* (2000) evaluated the use of microsatellite markers to quantify the genetic diversity within as well as among 25 accessions sampled from the world germplasm collection of sorghum. They found considerable variation at five SSR loci analyzed, with an average number of alleles per locus equal to 2.4 within accessions and 19.2 across the sample of 25 accessions. They also showed that microsatellite data are useful in identifying individual accessions with high relative contribution to the overall allelic diversity of the collection.

On the basis of polymorphism at 15 microsatellite loci, Grenier *et al.* (2000) assessed genetic diversity in 3 subsets constituted from the ICRISAT sorghum collection using random vs. non-random sampling procedures. The global molecular diversity retained in each subset was not affected by a sampling procedure based upon phenotypic characters.
Chapter 2: Review of Literature

Smith et al. (2000) evaluated potential utility of SSR technology for applications in research, product development, seed production and genetic resource conservation management for sorghum. Fifty genetically diverse elite sorghum inbreds were used to compare the discrimination abilities of 15 SSR primers with 104 RFLPs and to compare the association among lines revealed by these molecular data and by pedigrees. Results indicated that this set of SSRs could be used to help genetic conservation management and to support Intellectual Property Protection (IPP).

Ghebru et al. (2002) used a precise high-throughput approach to characterize diversity in 28 Eritrean sorghum landraces and to compare this diversity to representative samples of the world sorghum collection. Pools of simple sequence repeat (SSR) markers were sized and scored on automated DNA-sizing gels. An exceptionally high level of diversity was observed among the 28 Eritrean sorghum landraces, compared to other sorghum germplasm. Individual landraces were found to carry a high level of within-population diversity and heterozygosity and between-population diversity was equally high.

Schloss et al. (2002) evaluated sorghum RFLP probe sequences for presence of SSRs. Sixty SSRs were developed and assayed in an array of sorghum germplasm comprising inbred, landraces, and wild relatives. SSRs located in coding regions were less polymorphic \( (D = 0.07, \text{ averaged over 3 loci}) \) than those from gene-flanking regions, UTRs, and introns \( (D = 0.49, \text{ averaged over 3 loci}) \).
averaged over 16 loci. Map locations of these 60 RFLP-SSR loci, having \textit{Xcup} locus name prefixes, are expected to be the same as those detected by the RFLP probes from which they are derived, and these are reported by Chittenden et al. (1994) and Bowers et al. (2003).

Agrama and Tuinstra (2003) assayed 22 sorghum genotypes for polymorphism using 32 RAPD primers and 28 pairs of sorghum SSR primers. The results indicated that SSR markers are highly polymorphic with an average of 4.5 alleles per primer. The RAPD primers were less polymorphic with nearly 40% of the repeatable fragments detected being monomorphic. Genetic distances calculated from SSR data were highly correlated with the distances based on the geographic origin and race classification of the 22 diverse sorghum accessions used in this study.

Uptmoor et al. (2003) compared genetic relatedness of 46 sorghum accessions derived from Southern Africa using RAPD, AFLP and SSR markers. Mean genetic similarity was estimated at 0.88 based on RAPDs, 0.85 using AFLPs and 0.31 based on SSRs. Results on genetic relatedness and genetic diversity revealed a clear separation between landraces and improved varieties but a similar level of genetic diversity within both subgroups.

Grenier et al. (2004) assessed phenotypic diversity and compared pattern of distribution among Sudanese sorghum landraces collected from different geographical regions. Phenotypic diversity among landraces was high, as expressed by the large range of variation for mean quantitative traits and the high (0.81) Shannon-Weaver diversity index.
Menz et al. (2004) utilized a set of 100 SSRs, 1318 EcoRI/Msel AFLP and 496 PstI/Msel AFLP markers with known map positions to assess genetic similarity in a group of sterility-maintaining (B-line), fertility-restoring (R-line) and IS public inbreds. Cluster analysis of genetic similarity estimates revealed that the classification of sorghum inbreds is based on the sorghum working groups, Zera-zera, Kafir, Kafir-Milo, Durra and Feterita. Cluster analysis failed to give a clear differentiation between B-lines and R-lines, suggesting that R-lines and B-lines do not represent well-defined heterotic groups in this set of public inbred lines.

Several of the SSR loci evaluated by Casa et al. (2005) were monomorphic among a set of accessions assembled to represent the existent variability in cultivated sorghum.

To elucidate the genetic diversity and differentiation among Guinea-race sorghum landraces, Folkertsma et al. (2005) selected 100 accessions from ICRISAT sorghum Guinea-race core collection and genotyped these using 21 SSR markers. Stratification of the accessions into 11 countries and 5 eco-regional groups confirmed earlier reports on the spread of Guinea-race sorghum across Africa and South Asia: most of the variation was found among the accessions from semi-arid and Sahelian Africa and the least among accessions from South Asia. In addition, accessions from South Asia most closely resembled those from Southern and Eastern Africa. Stratification of accessions according to their Snowden classification indicated clear genetic variation between Margeritiferum, Conspicuum and Roxburghii accessions whereas the Gambicum and Guineense accessions were genetically similar.
2.11 Other non-mapping applications of markers in sorghum

Jordan et al. (2003) investigated the value of molecular marker-based distance information to identify high yielding grain sorghum hybrids in Australia. One hundred and sixty-two hybrid combinations derived from 70 inbred lines were screened with 113 mapped RFLP markers. They suggested that using data from just two linkage groups 38% of the variation in hybrid performance for grain yield could be explained. A model combining phenotypic trait data and parental diversity on particular linkage groups explained 71% of the variation in grain yield and has potential for use in the selection of heterotic hybrids.