Genetic diversity and detection of Kunitz protein in local soybean varieties

by

Prevashinee Padayachee

Submitted in partial fulfillment of

the requirements for the degree of

Master of Technology

in the

Department of Biotechnology

Durban Institute of Technology

Durban

South Africa

2004

PREFACE

The experiments in this thesis were conducted under the direct supervision of Professor B Odhav and Mrs T. Erasmus in the Research Department at Pannar, Greytown, during the period of January 2001 to June 2003. No work represented in this dissertation has been submitted to any other tertiary institution, either in part or full. The opinions and views expressed in this dissertation are those of the author, and do not necessarily reflect those of the Durban Institute of Technology.

P. Padayachee

ACKNOWLEDGEMENTS

The author wishes to express her sincere gratitude to the following individuals for their assistance in the preparation of this dissertation:

Professor B. Odhav, supervisor, Research Professor, Department of Technology, Durban Institute of Technology, for her expert guidance and supervision of this project as well as her valuable suggestions.

Mrs T. Erasmus, Researcher, Pannar, Greytown, for her help, encouragement and invaluable input during the preparation of this work.

Mr F. Pienaar, Biotechnology Technician, Pannar, Greytown. Thank you for your friendship and encouragement during the research.

Mrs Y. Babooram, Technical Staff, Pannar, Greytown, for your help in the processing of the samples and protein extractions.

Professor A. Oberholzer, University of Potchefstroom, for the kind use of your software for the preparation of the data.

Mr A.Jarvie, Plant Breeder, Pannar, Greytown, for providing the samples that were utilized in this study.

Mr Nimmesh Ramkisson for all your love and understanding during the endless days I had to dedicate towards my work. Thank you for being so supportive. I could not have done it without you.

DEDICATION

I dedicate this work to my family and friends, who have always showered me with love and support. Thank you for your strength and encouragement that I am so fortunate to be able to share in.

TABLE OF CONTENTS **page**

CHAPTER THREE: MATERIAL AND METHODS

CHAPTER FOUR: RESULTS

CHAPTER FIVE: DISCUSSION

LIST OF FIGURES **page**

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES **page**

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF ABBREVIATIONS

AFLP	-	amplified fragment length polymorphism
ELISA	-	enzyme linked immunoassay
GM	-	genetically modified
GMO	-	genetically modified organism
mt	-	mitochondrial
PCR	-	polymerase chain reaction
RAPD	-	random amplified polymorphic DNA
RFLPs -	-	restriction fragment length polymorphisms
SNP	-	single nucleotide polymorphisms
SSRs	-	single sequence repeats
ti	-	trypsin inhibitor
UPGMA	-	umweighted pair group method with arithmetic mean

ABSTRACT

South Africa produces 190 000 tonnes of soybean per annum. Seed producing companies require knowledge of the diversity of the germplasm to produce hybrids that will be competitive in local and overseas markets. Furthermore, they need to ascertain the presence/absence of the anti-nutritional factor, Kunitz trypsin inhibitor protein. Currently, seed producing companies plant the seed and wait for the grow-out in order to select desirable traits. This process is time-consuming, tedious and does not necessarily ensure the selection of the best genetic stability as it is based on phenotypic expression alone. This study was undertaken to evaluate a molecular method to determine the genetic diversity among soybean parent lines and optimize a method which can be used to evaluate seeds for the Kunitz trypsin inhibitor protein. Experiments were set up to: a) determine the genetic diversity in South-African bred soybean using amplified fragment length polymorphisms (AFLP) analysis, and b) determine the absence or presence of the Kunitz trypsin inhibitor protein using iso-electric focussing. Prior to application, the methods were tested using various parameters, and were than optimised for reproducibility, repeatablity, ease of operation and costs.

The genetic diversity of 40 soybean parent lines were evaluated using two established AFLP protocols. An isoelectric focussing technique using agarose gels was optimised to screen 30 soybean cultivars for Kunitz trypsin inhibitor protein. One method of AFLP analysis (Method A) showed too few polymorphisms whilst an alternative method that screens small genomes such soybean genome (Method B) was suitable. Method B identified a total of 187 fragments of which 22.4% were polymorphic. Primer pairs detected different levels of polymorphism, ranging from 2.6% to greater than 8%.

A screening technique for the detection of Kunitz trypsin inhibitor protein using a wide range (pH 3-10) agarose gel and R-250 brilliant blue stain was developed. Of the 30 hybrids analyzed, 83.4% tested positive for Kunitz protein and 16.6% tested negative.

The results of this study have led to the development of an AFLP technique, which will provide knowledge on the genetic distances between lines, prior to planting. This will result in a more efficient and effective selection procedure for genetic diversity amongst soybean collections maintained by individual seed companies. AFLP analysis can be used to determine homozygosity and heterozygosity of parents. The optimised screening technique for the Kunitz trypsin inhibitor protein provides a technique that can seed users, within a day of testing, of the absence/presence of Kunitz trypsin inhibitor protein. This data is essentail in informing those concerned of the suitability of soybean varieties for monogastric animal feed. It will also assists in determining the processing necessary if the seed is intended for food purposes. Thus, these techniques empower soybean users with a powerful, effective and viable tool in a multi-million rand industry.

CHAPTER ONE: THE PROBLEM AND ITS SETTING

1.1. Importance of the Study

For plant breeders a major challenge is selecting the traits to emphasize in a cultivar development programme. Soybean is a self-pollinating crop and therefore has a very low genetic diversity. To develop cultivars, most soybean breeding programmes focus on development of improved varieties combining enhanced nutritional quality with high yields and other agronomic traits. The current commercial market farmers require a soybean variety that has a high vigour, high yield, pest resistance, and is also lacking in an anti-nutritional factor namely, the Kunitz Trypsin inhibitor that causes pancreatic hypertrophy in monogastric animals. It is therefore of importance for the seed supplier and user to know the genetic diversity of soybean, and if the seeds contain the Kunitz trypsin inhibitor protein. This requires fast, accurate, cost effective and reproducible techniques.

1.2. Statement of the Problem

Plant breeders seek to improve varieties and genetic stocks for grain quality; yielding ability and resistance to diseases, insects and pests. They also have to maintain and evaluate diversity of genetic stocks and develop special-use varieties of species such as soybean with improved characteristics for food use. Soybean breeders evaluate and use the diversity of the germplasm collection to increase knowledge of soybean genetics, develop soybean for specific food and

feed use and develop improved disease resistance. Progress in this area improves soybean varieties which survive environmental stress and are competitive in the domestic and international markets.

Knowledge of the occurrence of the Kunitz trypsin inhibitor protein is important since it directs soybean's usage and processing requirements. In order to obtain this knowledge, a fast, accurate, cost-effective and reproducible method is necessary. Production of soybean with have a high yield and vigour also require knowledge regarding the genetic diversity of the parent lines used in crosses.

Pannar Seed (Pty) Ltd., a South-African seed producing company, produces maize, sunflower, dry beans and vegetable seeds for the local and overseas markets. Plant breeders produce the seed for commercial use. In order to assist the plant breeder in selecting suitable parent material for his crosses this study was undertaken to determine the genetic diversity among soybean parent lines and develop a rapid method to screening for the Kunitz trypsin inhibitor protein. Experiments were set up to: a) determine the genetic diversity in South-African bred soybean using amplified fragment length polymorphisms (AFLP) analysis, and b) to optimize a procedure for the detection of Kunitz trypsin inhibitor protein using iso-electric focusing.

1.3. Limitations

The screening technique for Kunitz trypsin inhibitor protein had to be developed around the current infrastructure in the Pannar laboratory. This was essential so as to avoid the costly start-up costs usually associated with the implementation of a new technique.

1.4. Thesis Organisation

This thesis is divided into five chapters. Chapter One covers a brief introduction to the problem and its significance. Chapter Two contains a summary of the literature associated with our study, while Chapter Three contains the details of the materials and methodology adopted for this study. Chapter Four handles the results of experiments carried out for our study and Chapter Five focuses on the discussion of results and conclusions reached.

CHAPTER TWO: LITERATURE REVIEW

2.1. History of Soybean

Soybean (*Glycine max)* is thought to be derived from *G. ussuriensis* Regel & Maacc, from a legume which is found wild throughout eastern Asia, possibly in hybridization with *G. tomentosa* Benth, which grows in southern China. *G. racilis* Skvortzor is considered an intermediate semi-cultivated species between the wild *G. ussuriensis* and *G. max.* (J W Purseglove, 2003). The soybean was cultivated in China before 3000 B.C., and was classified as one of the five sacred crops. The first written record is a 2200 B.C. farming manual advising Chinese farmers how to get the best from their crop. The western world only discovered soybean as a source of oil and protein in the 18th century, with the commercial cultivation starting for the first time in the United States in 1804. In the past thirty years, world production has risen to 180 million tonnes, the United States producing 44%, Brazil 19%, Argentina 17% and China 9%. The balance of the world production of 11% is made up, by Europe, India and South Africa (Specialised Protein Products, 2003).

2.2. Soybean Production in South Africa

The first record of soybean cultivation in South Africa appears in the Cedara Memoirs of 1903. The seed was imported from China but did not germinate properly. Successful soybean cultivation was initiated at a Research Centre in

Potchefstroom. In the early fifties, the center produced a well-adapted line which was named "Geduld" (meaning 'Patience'). From 10 000 tonnes produced in the seventies, national production had risen to 190 000 tonnes by 2001. Initially soybean was restricted to the Bapsfontein area and the Northern Lowveld region (Specialised Protein Products, 2003). Soybean is now commercially grown in the following areas as indicated in Figure 2-1:

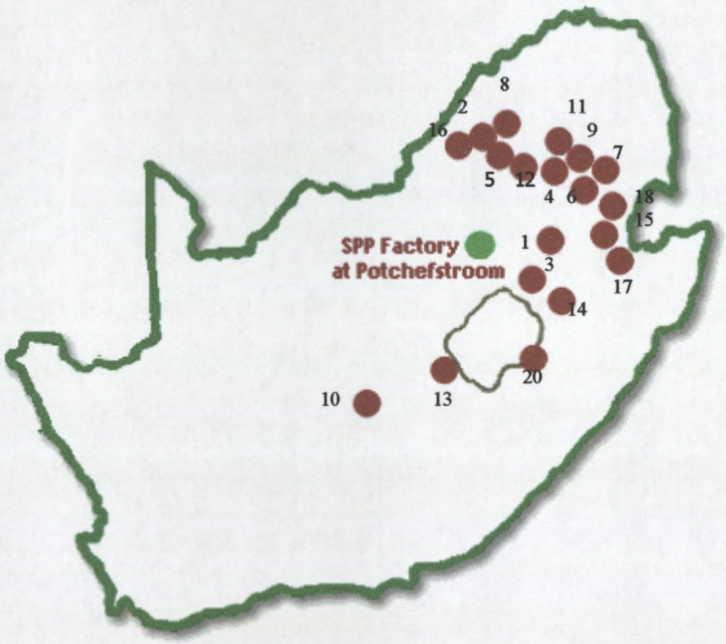

Figure 2-1 Map of the main regions in SA for soybean production (Specialised Protein Products, 2003)

1. Amersfoort	6. Carolina	11. Koedoeskop	16. Thabazimbi
2. Beestekraal	7. Ermelo	12. Middelsburg	17. Vrede
3. Bergville	8. Frankfort	13. Morgenson	18. Vryheid
4. Bethal	9. Hendrina	14. Newcastle	19. Winterton
5. Brits	10.Hopetown	15. Paulpietersburg	

Currently South Africa produces approximately 190 000 tonnes of commercial soybean per annum. Between 60 000 - 65 000 tonnes are processed for human consumption and the remainder is used mainly for animal feed. In December 2001, the Executive Council of Genetically Modified Organisms permitted soybeans to be grown commercially in South Africa from genetically modified (GM) seed. This is the first time that GM soy crop has been approved (other GM products grown commercially in South Africa are cotton and maize). The variety of soybean is resistant to a herbicide: Roundup, which kills other plants. The herbicide is widely used to kill weeds, which compete with crops for water and nutrients (Specialised Protein Products, 2003).

The aim of the GMO Act of 1997, which was implemented in December 1999, is to control the import of live GM products. The Act protects the consumer as well as the environment through its well-structured regulations. Although currently no mandatory labeling is required in South Africa for foods containing GMOs, The South African Department of Health is in the process of legislating for the labeling of products containing GM ingredients. In addition, the government is also planning to communicate a better awareness and understanding of biotechnology to the public (Specialised Protein Products, 2003).

2.3 Cultivation

Soybean is a frost-sensitive summer annual, and it takes about 75-80 d(days) for the beans to fully mature (IFR Communications, 2001). It is an erect, bushy

herbaceous annual (Figure 2.2) that can reach a height of 1.5 metres. The leaves are unifoliate, opposite and ovate, the secondary leaves are trifoliolate and alternate, and compound leaves with four or more leaflets are occasionally present. The nodulated root system consists of a taproot from which emerges a lateral root system. The plants of most cultivars are covered with fine trichomes. The papilionaceous flower consists of a tubular calyx of five sepals, a corolla of five petals, one pistil and nine fused stamens with a single separate posterior stamen. The stamens form a ring at the base of the stigma and elongate one day before pollination. Seeds are borne in hairy pods which grow in clusters of three to five; each pod contains two or three seeds, which resembles peas. When the seeds are mature, the upright vine and foliage begin to shrivel and the leaves fall away. Harvesting by machine must be completed before the pods shatter. Roundup and other non-selective herbicides are used extensively for weed control in soya cultivation, but they cannot be applied to weeds within growing crops because they will kill the crop as well as the weeds. Using biotechnology, plants are being developed that are tolerant to Roundup herbicide; farmers will be thus able to spray soybean crops during the growing season (IFR Communications, 2001).

roasted soybean used as confectioneries, soy nut butter, soy coffee, and other soy derivatives of oriental foods (Whigham, 2001).

Soybean uses can be further divided into two groups: products derived from soy oil and Soybean protein products. These are outlined in Figures 2-3 and 2-4 respectively.

There occur three categories of oil products (Figure 2-3), glycerol products, Refined Soy oil products and Soybean lecithin:

- Glycerol products are broken down into, fatty acids and sterols.
- Refined oil products can be either edible or technical products. Examples of edible refined oil products are coffee creamers, cooking oil, margarine, mayonnaise, pharmaceuticals, salad dressings, shortenings, etc. Technical uses of refined soybean oil products include anti-static agents, caulking compounds, disinfectants, fungicides, inks, paints, protective coatings, etc.
- Soybean lecithin products are used for nutritional, medicinal and technical uses. These technical applications include use for dispersing, stabilizing and wetting agents.

Soybean protein products (Figure 2-4) can be divided into soy flour concentrates & isolates which have technical and edible uses and soybean meal.

- The edible uses of Soybean concentrates and isolates include baby food, candy products, cereals, food drinks, noodles, yeast, beer, ale, etc., and the technical products range from adhesives to antibiotics, binders, cosmetics, inks, paints, plastics, textiles, etc.

- Soybean meal and there are feed uses such as aquaculture feeds, bee foods, calf feed replacers, fish food, livestock feeds, poultry feeds and pet foods. The hulls that are removed from the soybean can be used as dairy feed.

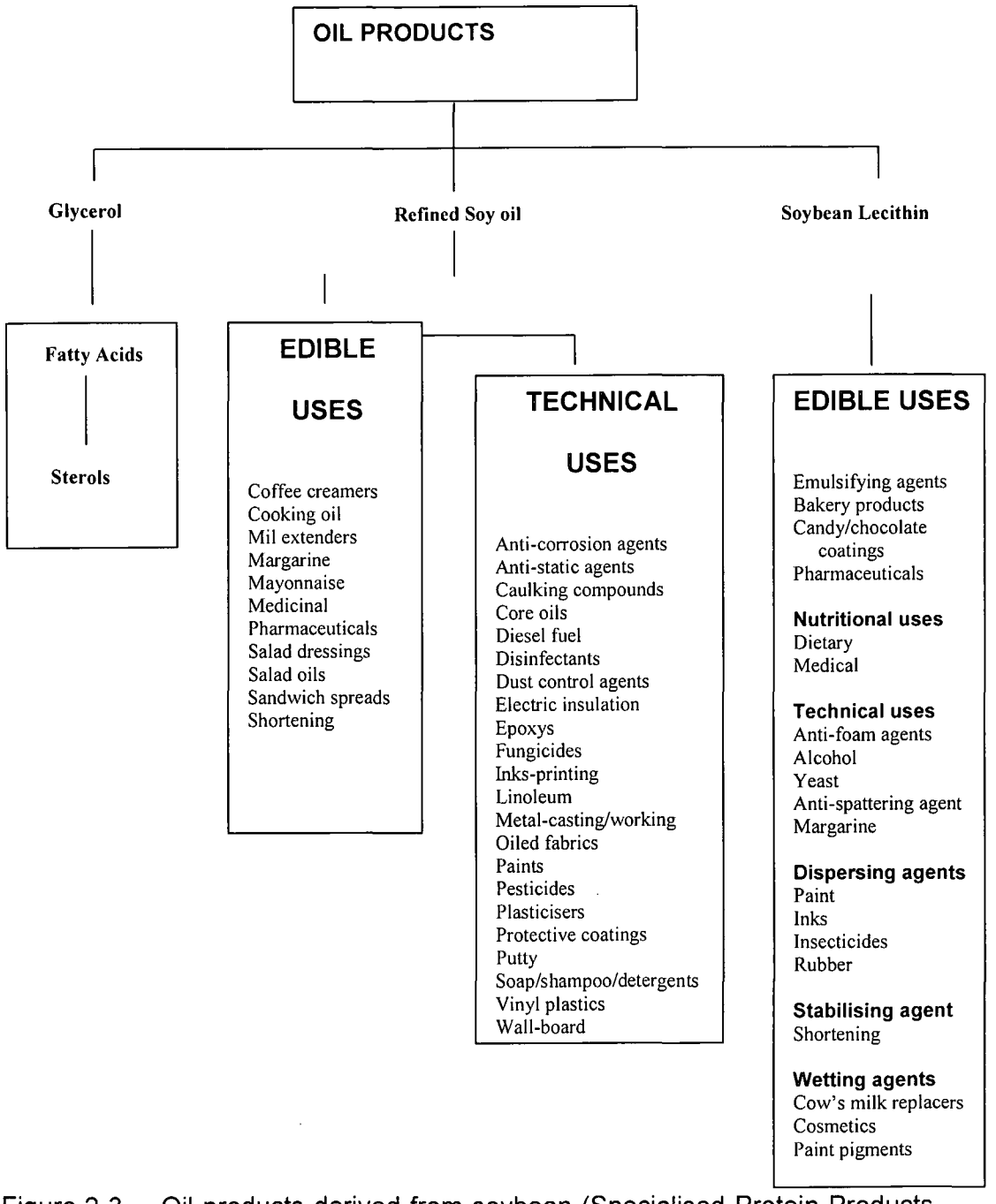

Figure 2-3 Oil products derived from soybean (Specialised Protein Products, 2003)

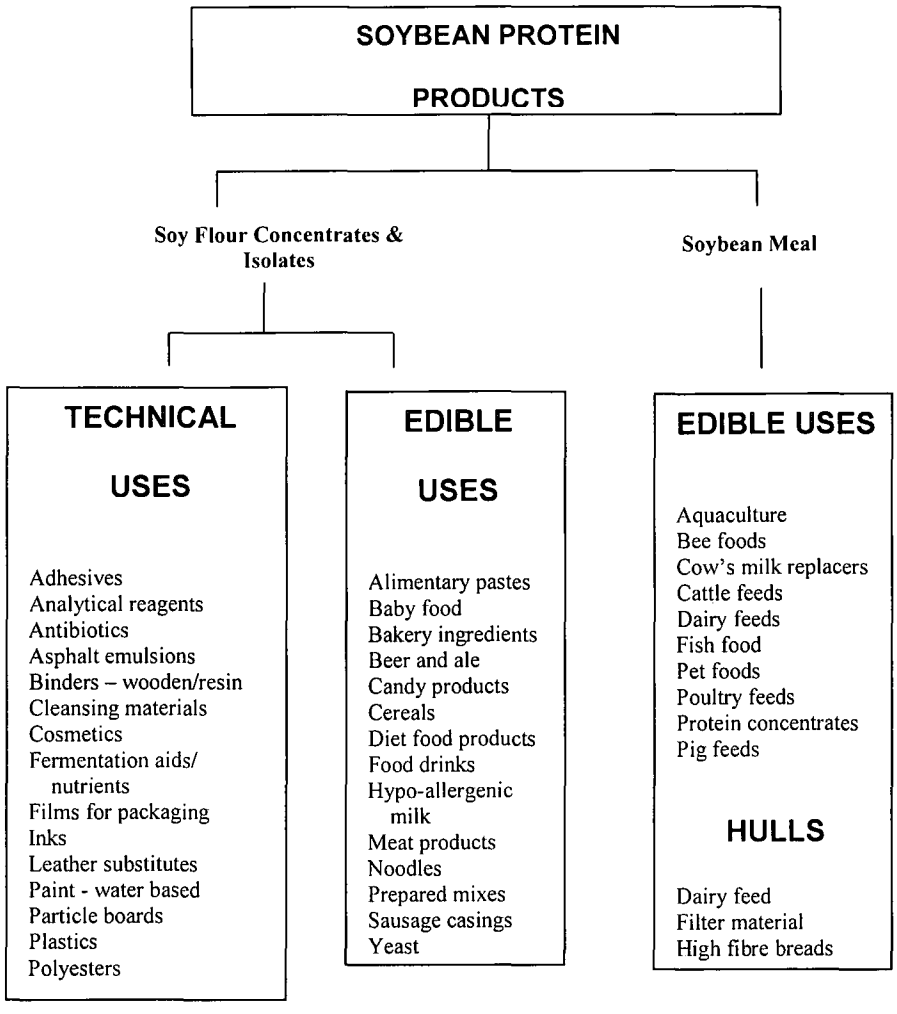

Figure 2-4 Soybean protein products (Specialised Protein Products, 2003)

2.5 Genetics of Soybean

2.5.1 Traditional breeding versus marker assisted breeding

Farmers have noticed for centuries that some individual plants in a given species manage to survive diseases or epidemics of insects relatively unscathed, while their neighbours succumb to infection or insect predation. In 1905, Sir Roland Biffen of Cambridge, England, wondered whether healthy plants inherited pest resistance, just as they might inherit the tendency to be tall

or short. His experiments on two varieties of wheat showed that the ability to resist infection by a rust fungus was indeed inherited, a discovery that intensified attempts by farmers and plant breeders to produce varieties of pest-resistant crop plants. Similar attempts continue today and primarily involve screening a large number of plant varieties to identify those that are resistant to particular pests. Resistant varieties are then crossed with those desirable for other reasons, for example, because they produce more grain per acre. Careful selection and repetitive crossing of progeny can eventually generate varieties that are high yielding and resistant to particular pests. But the process is extremely time consuming and it can take more than fifteen years to bring a new variety to market (National Academy of Sciences, 2001).

One challenge encountered in traditional breeding is that generally only closely related species of plants can be cross-bred. If no varieties are naturally resistant to a particular fungus or insect, traditional breeders have no way to create resistance to that fungus or insect besides population breeding. Furthermore, breeders frequently face a situation in which a resistance gene is closely linked to a gene that adversely affects the quality of a crop, that is, where the two traits are always inherited together. For example, insect resistance in lettuce might be inherited along with a tendency for the lettuce to taste bitter. In the early 1990s, despite plant breeder's best attempts to improve the pest and disease resistance of cotton, corn, rice, and other crops, farmers worldwide lost about one-fourth of their crops to pests and diseases (National Academy of Sciences, 2001).

2.5.2 Developments in soybean breeding

One of the major challenges for soybean breeders is selecting the traits to emphasize in a cultivar development programme. Through biotechnology research, the list of new traits that are available to breeders has expanded and will is becoming even longer (Fehr, 2001).

Herbicide resistance is one such trait. The success of Roundup Ready Soybean indicates that tolerance to selected herbicides is important to farmers. It seems likely that new herbicides will be developed to which genetic resistance will be incorporated into the soybean. The breeder must therefore decide if the new product will be successful against herbicides and if the investment in breeding cultivars for resistance will be profitable (Fehr, 2001).

The problem of insect resistance has been solved for Coleoptera by the incorporation of the Bt gene into soybean. It is not likely that the gene will be important for cultivars developed for the Midwest USA where these insects are not a consistent problem in soybeans. The gene may be more important for cultivars grown in the southern US where these insects occur(Fehr, 2001).

Disease resistance is another important aspect in soybean breeding. No novel genes for disease resistance in soybean have been developed through biotechnology, but research is underway. Biotechnology is used to more efficiently select for the resistance genes that are commonly used by soybean breeders in cultivar development. There is research in progress to increase the

protein quality, namely the lysine and methionine content of soybean. A gene was transferred from the Brazil nut plant to soybean that altered the amino acid composition of the protein. Unfortunately, people who were allergic to Brazil nut were also allergic to the soybeans with the Brazil nut gene. As a result, that modification will not become a commercial product. Other strategies for modifying amino acid content are under investigation. Research into soybean oil quality are also being carried out. Major genetic changes have been made in the composition of soybean oil. Cultivars have been grown on a limited acreage that have either reduced saturate content, elevated oleic acid content, or reduced linolenic acid content. The challenge is to determine if there is an end use for the oil that will justify the cost of producing and marketing it (Fehr, 2001).

The carbohydrate quality of soybean has also been challenged. Cultivars were grown for market testing that have reduced stachyose content. It is believed that reduction of the sugar will improve the energy available from soybean for poultry and swine. Also the beany flavour of soybean protein has been removed genetically by eliminating the three lipoxygenase enzymes that are responsible for the development of the flavour. Cultivars have been grown in the Midwest USA to determine the market potential of lipoxygenase-free soybeans, particularly in food products made from whole soybeans, such as soymilk, or products made from soybean protein, such as bakery products (Fehr, 2001).

2.5.3 The soybean genome

The soybean has three fewer pairs of chromosomes and about 1.7 billion fewer DNA units than humans. Perry B, Cregan, a plant geneticist used single sequence repeats (SSRs) to construct a map of the soybean genome. Cregan (1999) used SSRs to find unique variations to serve as markers for nearby genes. Cregan (1999) found that the soybean's 20 pairs of chromosomes contain a total of 1.3 billion bases. Thus far they have found more than 700 SSR markers from within these 20 chromosomes pairs. Now they are filling in the details by use of single nucleotide polymorphism (SNP). There is an average of 65 million bases per chromosome. It would be ideal to have a SNP marker every 100,000 to 250,000 bases (ARS News and Information, 2001).

2.5.4 The Kunitz trypsin inhibitor protein

In 1917, T.B. Osborne and L.B. Mendel, researchers at the Connecticut Agricultural Experiment Station, demonstrated that unheated, raw soybean meal is inferior in nutritional quality to steam-heated soybean meal. Their research become the foundation for the development of the lucrative soybean processing industry and expanded the potential use of soybean meal as a high-protein feed. Physiologically the ingestion of raw soybean meal by monogastric animals (for example, poultry and swine) causes pancreatic hypertrophy, a condition in which the pancreas enlarges and ultimately ceases to function. Moses Kunitz, professor of biochemistry at New York's Rockefeller University, purified and characterized the predominant antinutritional factor in soybean seed in 1945 and called this factor: trypsin inhibitor (Hymowitz, 2001).

Soybean contain two major proteinase inhibitor classes- the Kunitz trypsin inhibitor and the Bowman-Birk proteinase inhibitor. These inhibitors are found in other legumes, are localized within protein bodies and are specific for serine proteases. The Kunitz trypsin inhibitor protein class is the most prevalent soybean protease inhibitor, is represented by a 21-kD protein and is specific for trypsin. By contrast the Bowman-Birk class consists of several related 8-kD proteins, and inhibits trypsin, chymotrypsin and elastase. Proteinase inhibitors have been proposed to function as an endogenous insecticide. Because soybean null lines exist lacking either the Kunitz trypsin inhibitor or the Bowman-Birk protease inhibitor, it is suggested that these proteins are not essential for normal growth and development (Jofuku and Goldberg, 1989).

The soybean Kunitz trypsin inhibitor is a beta-sheet protein with unusual stability when exposed to chemical and thermal denaturation. Thermal denaturation of the Kunitz trypsin inhibitor is readily reversible to the native form upon cooling (Roychaudhuri *et al.,* 2003). Human trypsin is more resistant to inhibition than is the trypsin of other mammalian species. The effect on human trypsin of soybean trypsin inhibition in soy protein does not appear to be a potential hazard to man. In animal diets, however, pancreatic toxicity must be considered whenever soybean protein is utilized (Flavin, 1982).

In 1968 the Illinois Agricultural Experiment Station began to study the genetic variations of the Kunitz trypsin inhibitor in soybean. Four different trypsin inhibitor (*Ti*) forms were discovered. Three of the forms identified as Ti^a, Ti^b, and

Ti^c are distinguishable from each other by banding patterns produced on gels. The differences are due to amino acid substitutions in the protein. Genetic studies have revealed that the three forms are controlled by a co-dominant multiple allelic system at a single locus. The fourth form, found in soybean lacks the Kunitz trypsin inhibitor. The absence of the protein is inherited as a recessive allele to Ti^a, Ti^b and Ti^c, and has been designated ti.

'Kunitz soybean', which lacks the Kunitz inhibitor protein, has the following advantages over other commercial varieties. These are:

- Farm-grown soybeans can be fed directly to swine in their finishing rations.
- Kunitz can replace as much as 50 percent of the heated soybean meal in chick rations.
- Processors can save about 25 percent in energy costs in processing Kunitz soybeans.
- Kunitz can be exported to those countries lacking soybean processing facilities.
- Kunitz is an excellent raw product for companies making soy-based foods for babies (Hymowitz, 2001).

2.5.5 Current methods for the detection of Kunitz protein

The variants having the ti allele lack Kunitz trypsin inhibitor protein, which forms the basis for differentiating hybrids for this allele. Polyacrylamide gel electrophoresis has been frequently used to differentiate between the three electrophoretic forms of Kunitz trypsin inhibitor. This has been used to determine genetic diversity by Singh *et al.* (1969) and Orf and Hymowitz (1979).

In 1980 Hildebrand and Hymowitz (1980) developed a rapid test for Kunitz *ti* activity in soybean by use of Azocoll (a proteolytic substrate consisting of a dye bound to a protein), which is insoluble. When proteases such as trypsin react with Azocoll, the red dye is released into the solution. Trypsin inhibitors, however, prevent the tryptic digestion of Azocoll.

The current method of screening for *ti* containing soybean is polyacrylamide disc gel electrophoresis, which is slow and requires specialized equipment. Soybean breeders and crop testing laboratories therefore need a simple and rapid technique for determining the occurrence of the Kunitz ti allele in soybean seeds for use. Recently Duranti *et al,* (2003) used a one step method for the purification of Kunitz trypsin inhibitor. The method consisted of heating the soybean at 60 degrees C for 90 min. This was followed by loading of the aqueous medium onto an affinity chromatography column with immobilized trypsin. The retained fraction was eluted with 0.01 N HCl. Brandon *et al.* (1991) used another method, an antibody detection method based on Elisa (enzyme-linked immunosorbent assay) to measure the inhibitor content of soy concentrates & isolates and flours, both heated and unheated, as well as a soy infant formula. They concluded that the immunoassays complement the established enzymatic assays of trypsin inhibitors and have advantages in measuring low levels of inhibitors in processed foods and differentiating between the Kunitz and Bowman-Birk inhibitors. Their findings have significant impact on food safety methods.

2.5.6. Methods of DNA fingerprinting for soybean classification

The fast growth of DNA technology has aided in the use of genetic markers as a way to characterize populations. Generally, DNA fingerprinting is characterized as classical, hybridization-based or polymerase chain reaction (PCR) based.

Classical fingerprinting involves cutting the DNA with restriction endonucleases followed by electrophoretic separation. This strategy was used to develop a DNA-based marker identification system called Restriction Fragment Length Polymorphisms (RFLPs). An RFLP may be the result of length mutation, and/or point mutation at a restriction enzyme cleavage site at a given chromosomal location. RFLPs can be detected by analyzing restriction digests of genomic DNA through Southern hybridization. The probes used in RFLP analysis can be generated from cloned genomic, cDNA or mtDNA fragments, or from specific DNA segments, amplified using PCR. Thus, depending on the probe used, RFLPs can be used to analyze mtDNA variation, ribosomal (r)DNA region variation, repetitive and single-copy sequence variations. RFLPs are co-dominant markers. This makes them suitable for studying population genetics as well as for linkage map construction. By employing probes that detect multiple loci and dispersed repetitive sequences, the sensitivity of the RFLP method can be enhanced to fingerprinting. In addition, synthetic simple repeat oligonucleotides can be used as fingerprinting probes (Wang and Szmidt, 2001).

PCR-RFLPs can also be applied for fingerprinting. DNA hybridization-based RFLP analysis requires the isolation of large amounts of purified DNA. With PCR it becomes possible to analyze specific sequences from small amounts of tissue. The advantages of PCR-RFLP lie in its speed, sensitivity and specificity. PCR can be applied on crude DNA extracts with a pair of region-specific primers. Variation of the amplified fragment can be further analyzed by restriction enzyme digestion and electrophoretic separation. The regions most commonly examined by PCR-RFLP are the rDNA sequences (Wang and Szmitz, 2001).

Variation within a species can also be assayed using the random amplified polymorphic DNA (RAPD) method, in which arbitrary short oligonucleotide primers, targeting unknown sequences in the genome, are used to generate amplification products that often show size polymorphisms within species. RAPD analysis offers the possibility of creating polymorphisms without any prior knowledge of the DNA sequences of the organisms investigated. The patterns produced are highly polymorphic, allowing discrimination between isolates of a species if sufficient numbers of primers are screened (Welsh and McClelland, 1990).

RAPD analysis is fast and economical for screening large numbers of samples. However, some researchers are critical of the sometimes-poor reproducibility of RAPD patterns. The reproducibility within a laboratory is usually satisfactory (Tommerup *et al.*, 1995). However, inter-laboratory comparison of RAPD patterns may not always be applicable since the RAPD patterns can be

influenced by many technical factors (Penner *et al.,* 1993). This implies that diagnostic RAPD markers specified purely by mobility/ size may not be confirmed by another laboratory, thus posing uncertainty for data cross-referencing in race/ type identifications. For any RAPD marker to be used as a diagnostic tool by a wider group of researchers, it is necessary to characterize it more thoroughly, through isolation, cloning and sequencing to generate either probes or specific primers for future applications. The main limitation of RAPD analysis in population studies is the dominant character of RAPD markers. In the study of diploid organisms, homozyote AA can not be distinguished from heterozyote Aa, since both will give a RAPD pattern with a band corresponding to an A (Wang and Szmidt, 2001).

The amplified fragment length polymorphism (AFLP) assay is actually a combination of the classical, hybridization-based approach and the PCR-based approach. In the AFLP procedure (Fig. 2-3), DNA is digested with restriction endonucleases, and DNA adaptors are ligated to the ends of the DNA fragments to create template DNA for amplification by PCR. The amplified fragments are then analyzed through polyacrylamide sequencing gel electrophoresis, thereby creating a fingerprint. The AFLP assay promises to be an excellent tool in characterizing population variance and genetic diversity in soybean (Cheung, 2001).

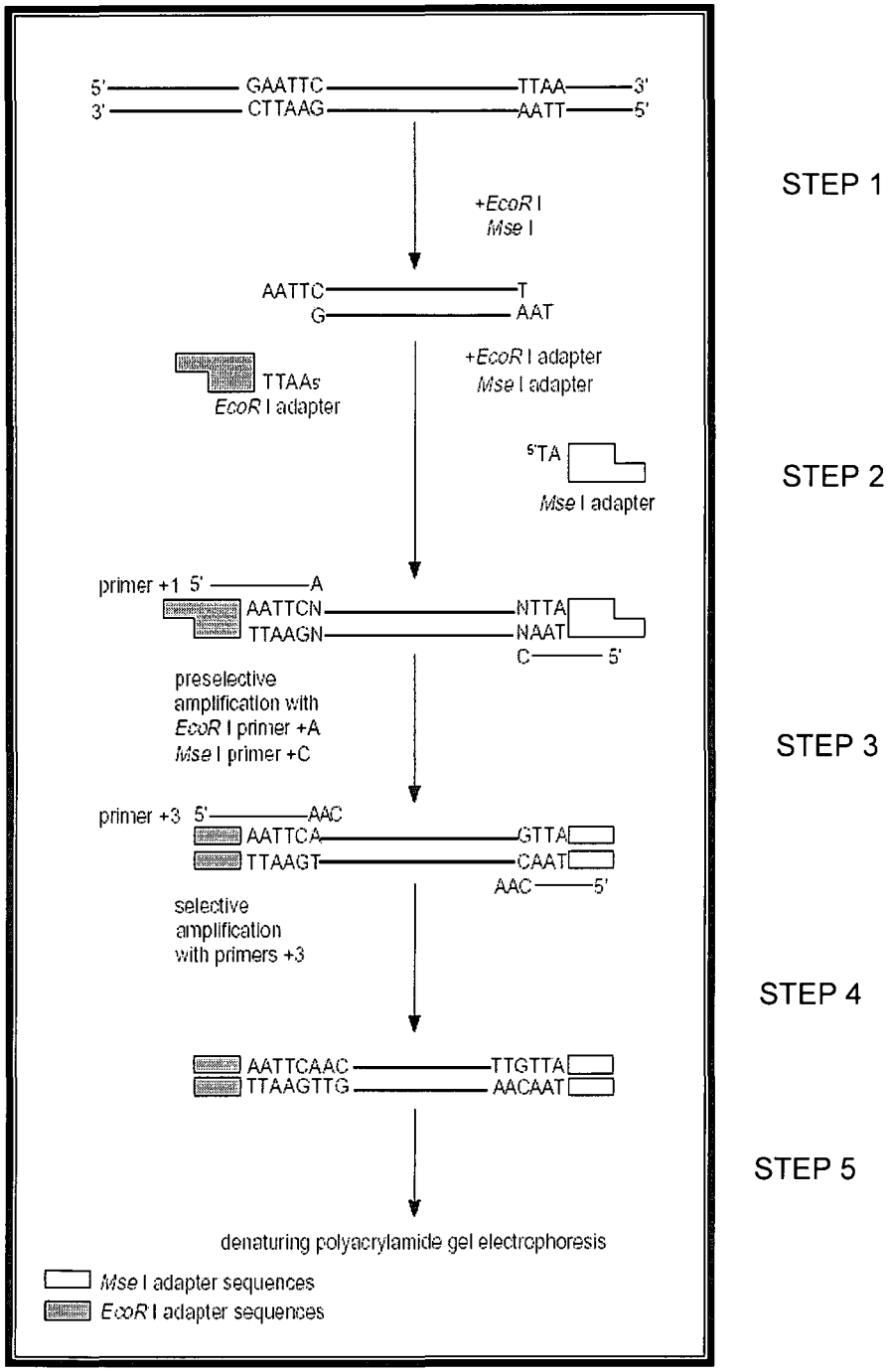

Figure 2-5. A schematic diagram showing the basic steps involved in the AFLP procedure (Applied Biosystems, 1997)

2.5.7. Comparison of DNA fingerprinting methods used for soybean typing

Kiem *et al.* (1989) used 17 RFLP markers to evaluate 58 *G.soja* and *G.max*. accessions that included ancestral lines and cultivars. Molecular diversity, based on RFLP polymorphisms was 0.16 for the cultivars and 0.26 for the ancestors. Kiem *et al.* (1992) surveyed 38 soybean genotypes with 132 RFLP probes. Molecular diversity was 0.30, indicating that diversity in soybean was greater than previously reported. The authors cautioned that diversity estimates in their report were probably inflated because the genotypes included in the study were selected to represent the greatest amount of diversity and not to represent the typical material found in breeding programmes.

Skorupska *et al.* (1993) identified 29 RFLP probes with gene diversity scores ≥ 0.30 in elite and ancestral lines of those cultivars. Lorenzen *et al.* (1995) identified 67 polymorphic RFLP loci with gene diversity scores of ≥ 0.30 in elite and ancestral cultivars in North American breeding programmes.

Maughan *et al.* (1996) used AFLP to test for species diversity, inheritance and near-isogenic line analysis. AFLP primer pairs detected a total of 759 AFLP fragments in a sample of 23 accessions of wild and cultivated soybean, with an average of 51 fragments produced per primer pair per accession. Two-hundred and seventy four fragments (36% of the total observed) were polymorphic, among which 127 (17%) were polymorphic in *G.max*. and 237 (31%) were

polymorphic in *G.soja*. The number of polymorphic loci detected per AFLP primer pair in a sample of 23 accessions ranged from 9 to 27.

Doldi *et al.* (1997) employed RAPD and microsatellite techniques to evaluate the genetic diversity among 18 soybean genotypes. Out of the 33 random primers used in RAPD reactions, only 12 showed polymorphisms useful for characterization of these genotypes. In contrast, all 12 microsatellite primer pairs used in the study detected polymorphism with 2-6 allelles per locus.

Not all RAPD fragments are equally informative, only 34% of the markers used by Thompson *et al.* (1998) to evaluate diversity in a group of 18 North American soybean ancestors and 17 PIs from the USDA Soybean Germplasm Collection were polymorphic. Skorupska *et al.* (1993) identified 29 RFLP probes with gene diversity scores ≥ 0.30 in elite and ancestral lines of those cultivars. Lorenzen *et al.* (1995) identified 67 polymorphic RFLP loci with gene diversity scores of ≥ 0.30 in elite and ancestral cultivars in North American breeding programmes.

CHAPTER THREE: MATERIAL AND METHODS

This chapter is divided into two parts (3.1 and 3.2). The first part deals with analyzing the genetic diversity in soybean using AFLP and the second part deals with developing a technique to screen for Kunitz trypsin inhibitor protein. In the first part leaf samples from 160 soybean parent lines from Pannar Seed (Pty) Ltd. were used for DNA extractions and the AFLP profiles of these samples were analyzed. In the second part, the detection of the Kunitz Trypsin inhibitor protein using isoelectric focusing is discussed.

3.1 AFLP Analysis of Soybean

Two different methods were employed for the AFLP analysis, Method A (K. Edwards, pers. Comm.) and Method B (Life Technologies Corporation). The DNA used in both AFLP methods was extracted using the standard protocol provided in the GenElute plant genomic DNA extraction kit which was purchased from Sigma Aldrich. The DNA samples were then analyzed by AFLP method using two different protocols. In the first method (Method A) the DNA was digested with restriction endonucleases, ligated to adapters and amplified using known primers obtained from GibcoBRL. The method involved digesting DNA with Mlu I and Mse I restriction enzymes and the ligation of double stranded adapters to the restriction fragments. This was followed by selection using Dyna beads and the selective amplification of DNA using PCR with known primers. One of the primers was radiolabelled and the PCR products were visualized by first running them on a sequencing gel and thereafter exposing the gel to X-ray film.

The second method(Method B) is generally recommended for small genotypes to ascertain genetic diversity. In this method a similar protocol was followed but the DNA was amplified using primers whose sequences were unknown. This method was obtained from Life Technologies Co. in the form of an AFLP kit and involved the digestion of DNA with restriction endonucleases, the ligation of double-stranded DNA adapters to the restriction fragments, preselective amplification, selective amplification and denaturing polyacrylamide gel electrophoresis. All reagents used in this method were obtained from Invitrogen with the exception of Taq DNA polymerase (Amplitaq-PerkinElmer Life Sciences) and gamma ^{33}P (Nen- Separation Scientific).

Figure 3.1 gives a schematic representation of the protocol followed.

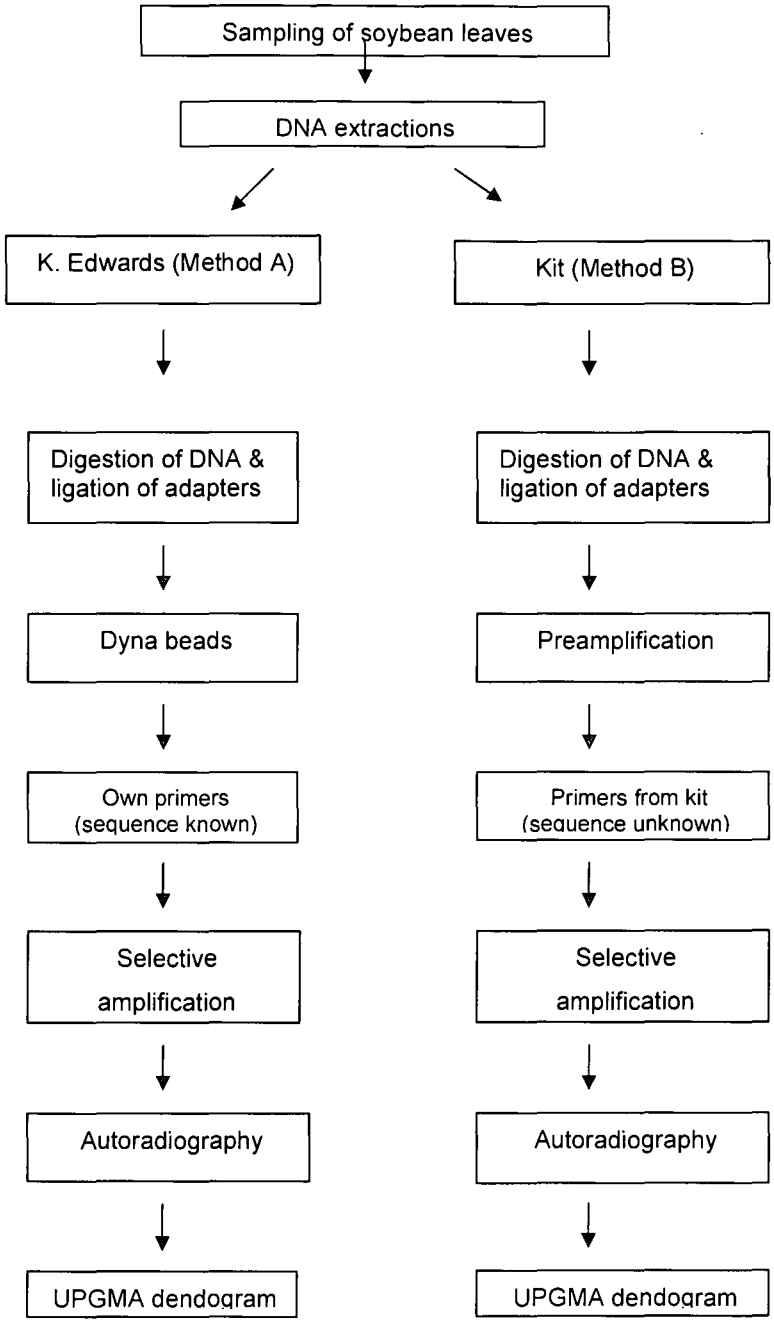

Figure 3-1. Schematic representation of the methodology used to determine genetic diversity in *G. max*

3.1.1 Sampling

Soybean (*G. max)* leaf samples used in this aspect of the study were obtained from Pannar Seed (PO Box 32, Greytown, 3250). The samples (160) were made up of 40 parent lines which formed part of Mr A. Jarvie's (Plant Breeder, Pannar Research, Greytown) soybean breeding programme. This was made up of bi-parental crosses of soybean plants.

Bi-Parental crosses of soybean plants lead to a collection of soybean parent lines which were cultivated by Mr A Jarvie. These samples were used in the study for DNA fingerprinting. The youngest leaves (4 to 6 leaves approximately 4 cm long) of each line were selected, harvested and bulked when the plants were six weeks old. Four replicates of each of forty parent lines were sampled in 15 ml eppendorf tubes and stored on ice during transit to the laboratory.

3.1.2 Extraction and purification of genomic DNA

The extraction and purification of genomic DNA from the leaf of the 40 soybean lines was carried out using the GenElute Plant Genomic DNA Kit (SIGMA-G2N350). The method involved the disruption of the plant cells by grinding the leaf tissue in liquid nitrogen using a mortar and pestle, followed by lysis of plant tissue at 65°C using the Lysis buffers provided in the kit. RNase A (50 units) (Sigma-R6513) was added to each sample to remove contaminating RNA prior to incubation at 65°C. The precipitation solution, also part of the kit, was added and samples were incubated on ice for 5 min. Following incubation the samples

were centrifuged at *12 000 x g* to pellet the debris. The binding solution from the kit was added to the supernatent and centrifuged through a binding column. The samples were washed using the wash solution from the kit and eluted into a clean 1.5 ml microcentrifuge tube with pre-warmed (65°C) elution solution which is also part of the kit. The DNA samples were stored at -20°C. Details of the methodology can be found in Appendix 1.

3.1.3. Determination of genomic DNA concentration

DNA concentrations were determined using agarose gel electrophoresis (Plant Molecular Biology, 1997). Quantification of the DNA in the extract was achieved by analysis of the agarose gel with ethidium bromide staining. The extracted DNA (3 µl) was analyzed on a low percentage gel (0.7% $^{w}/_{v}$) agarose (Amresco-9012366) gel alongside 30 ng of Lambda DNA (Promega) standard. The gel was visualized in an ultraviolet light box (UVP) and DNA concentrations relative to the Lambda DNA reference was calculated using Genetools software by SYNGENE (UK). The DNA was subsequently diluted using sterile distilled water (SABAX) to a working concentration of 150 ng/µl in 1.5 ml eppendorf tubes. All the DNA samples were stored at –20°C. The details of reagents and method is outlined in Appendix 1B.

3.1.4. AFLP Method A

DNA digestion and adaptor ligation

Purified DNA (150 ng/µl) was added to 5 µl of 10X One-phor-all buffer (AEC Amersham-270901), 10 units of *Mlu* I restriction enzyme (Amersham-E1071Y), 10 units of *Mse* I (New England Biolab-525L) and the solution was brought up to a final volume of 50 µl (per reaction) with sterile distilled water. The digestion mixture was incubated at 37°C in a waterbath for 1 h.

One microlitre *Mse* I (50 pmol) adaptor (Gibco BRL), 1 µl *Mlu* I (5pmol)adaptor (Gibco BRL), 1 µl 10X All-Phor-One buffer (Amersham), 1.2 µl of 10 mM ATP (Amersham-272056), 0.2 µl of T4 ligase (Roche-70042X) and sterile distilled water were added to a final volume of 10 µl. This mixture was added to the digestion mixture, vortexed and incubated in a waterbath at 37°C for a further 3 h.

Selection of biotinylated DNA fragments

Twenty microlitres Streptavidin (M-280) Dyna Beads (Dynal-Norway) was used per sample. The beads were washed four times with 20 µl of Tris-EDTA wash buffer (Appendix 1B). Twenty microlitres of beads was added to each 60 µl digestion-ligation mixture and incubated on ice for 30 min with gentle manual agitation every 5 to 10 min. One hundred and twenty microlitres of wash buffer

was added to each sample. The samples were then washed 3 times with 200 µl wash buffer. After the third wash all the liquid was aspirated and 100 µl TE buffer (Appendix 1B) was added. Samples were stored at -20°C.

Labelling of Mse I primer

One tenth microlitre *Mse*I primer (300 ng/µl) ,(Gibco BRL), was added to 0.1 µl 10X One-phor-all buffer, 0.1 µl gamma-P^{33} (Life Technologies Co.), 0.05 µl kinase (USB-E70031Z) and 0.65 µl sterile distilled water. The mixture was incubated for 1 hour at 37°C in a waterbath and then for 10 min at 65°C on a heating block (Thermolyne).

Selective amplification by PCR

One microlitre of DNA (selected by Dyna beads) was added to 2 µl 10X Taq buffer (Roche), 1 µl *Mlu* I primer (Gibco BRL), 1 µl $MgCl_2$ (50 mM) (Roche), 12.9 µl distilled water, 0.1 µl Taq (0.5 units) (Roche-N8080161) 1 µl dNTP's (Roche-1969064) (100 µM), and 1 µl labeled *Mse* I primer. The final reaction volume was 20 µl. These were placed in the PCR (GeneAmp PCR 9700) machine. The cycle profile used for PCR is outlined in Table 3-1.

Table 3-1. The PCR profile used for selective AFLP amplification of samples

	Number of cycles	Temperature (°C)	Time
Initial denaturation	1	72	1 min
		94	2 min
Denaturation	12	94	20 s
Annealing		65	30 s
Elongation		72	2 min
Denaturation	25	94	20 s
Annealing		56	30 s
Elongation		72	2 min
Final Elongation	1	72	30 s

The PCR machine was programmed with a holding step to keep the reactions at 4°C after the PCR was complete.

3.1.5. AFLP METHOD B

Restriction digestion of genomic DNA

Purified DNA (1.67 μl) which was extracted using the GenElute plant genomic DNA kit (Sigma) was added to a 1.5 ml microcentrifuge tube. Five microlitres of reaction buffer (50 mM Tris-HCl pH 7.5/ 50 mM magnesium acetate/ 250 mM potassium acetate) was added to 2.5 units of *Mse*I restriction enzymes (*Eco*RI/ *Mse* I -1.25 units/μl each in 10 mM Tris-HCl pH 7.4/ 50 mM NaCl/ 0.1mM EDTA/ 1 mM DTT/ 0.1 mg/ ml BSA/ 50% glycerol ($^{v}/_{v}$)/ 0.1% Triton X- 100) in

reaction buffer. part no. 51115) was added to 2.5 units of *Eco*RI and 2.5 units of *Mse*I restriction enzymes (part no. 51114) in reaction buffer. This was made up to a final volume of 25 µl with distilled water. The mixture was vortexed briefly and incubated for 2 h at 37°C. Thereafter it was incubated for a further 15 min at 75°C to inactivate the restriction endonucleases. The samples were placed on ice for 5 min.

Ligation of adapters

Adapter ligation solution (part no. 51116) (24 µl) and 1 µl T4 DNA ligase (part no. yo1301) were added to the digested DNA from the above step. This mixture was mixed gently at room temperature, vortexed briefly and incubated at 20°C for 2 h. Following incubation, a ten-fold dilution of this ligation reaction was performed by transferring 10 µl of the reaction mixture to a 1.5 ml microcentrifuge tube and adding 90 µl of TE buffer (part no. 50282). These samples were mixed by vortexing. This formed the template DNA for the PCR in the next step (Preamplification). The diluted DNA as well as the unused DNA portions were stored at -20°C.

Preamplification

In this step genomic DNA was amplified with AFLP primers each having one selective nucleotide. Diluted template DNA (5 µl) from the above step was added to 40 µl pre-amp primer mix (part no. 50142), 5 µl 10X PCR buffer with

Mg^{2+} (part no. 50284) and 1 μl Amplitaq DNA polymerase (Roche-N8080161). Samples were mixed by vortexing briefly.

The PCR cycle was performed for 20 cycles each at:

94°C for 30 s

56°C for 60 s

72°C for 60 s

Once the PCR cycle was completed the amplified DNA was kept at 4°C.

A 1:50 dilution of the PCR products was performed as follows: 3 μl of the PCR products from the previous step was added to 147 μl of TE buffer (part no. 51147). This was sufficient for 30 selective amplifications. Both unused diluted and undiluted reactions were stored at -20°C.

Primer labelling

Primer labelling was performed by phosphorylating the 5' end of the *Eco*RI primers with ATP and T4 kinase (gamma ^{33}P-Separation Scientific,SA). The following components were added to a 1.5 ml microcentrifuge tube: 18 μl *Eco*RI primer (27.8 ng/μl), 10 μl distilled water, 10 μl (gamma ^{33}P) ATP and 10 μl of T4 kinase (part no. 50286). The mixture was vortexed briefly and incubated at 37°C for 1 h. The kinase enzyme was inactivated at 70°C for 10 min on a heating block.

Selective AFLP amplification

Primer combinations from the Invitrogen kit (Life Technologies Co.- USA) were used for selective PCR.

The primer definitions were:

E-AAC (Cat no. 50141) and M-CAA (Cat no. 50133)

E-ACA (Cat no. 50136) and M-CTT (Cat no. 50192)

E-ACA (Cat no. 50136) and M-CAA (Cat no. 50133)

The PCR reactions were assembled by making two mixes (Mix 1 and Mix 2) for each primer combination. Mix 1 was made up of 5 μl of the labeled primer from the previous step and 45 μl of the *Mse* I primer (6.7 ng/μl), containing dNTPs. The total volume of Mix 1 was 50 μl and this was sufficient for 10 samples or reactions.

Mix 2 was made up of 79 μl distilled water, 20 μl 10X PCR buffer plus Mg^{2+} (Part no. 50284), and 1 μl Amplitaq DNA polymerase (5 units/μl). The total volume of Mix 2 was 100 μl and this was sufficient for 10 samples or reactions.

Each AFLP reaction was assembled by combining 5 μl of diluted template DNA from the preamplification step with 5 μl of Mix 1 (primers and dNTPs) and 10 μl of Mix 2 (Taq DNA polymerase and PCR buffer). This gave a final reaction volume of 20 μl for PCR. Details of the PCR cycle used for selective amplification are outlined in Table 3-2.

Table 3-2. PCR profile used for selective amplification of *G max*. samples

	No of cycles	Temperature (°C)	Time
Denaturation	1	94	30s
Annealing		65	30 s
Elongation		72	60 s
Denaturation	[1]12	94	30s
Annealing		65	30 s
Elongation		72	60 s
Denaturation	23	94	30 s
Annealing		56	30 s
Elongation		72	60 s

[1] The annealing temperature was lowered by 0.7°C for 12 cycles. This gave a touch down phase of 13 cycles.

3.1.6. Analysis of PCR products on polyacrylamide sequencing gel

The PCR products were analyzed on a 4% denaturing polyacrylamide sequencing gel. A 4% denaturing polyacrylamide sequencing gel was prepared as described in Appendix 1. Electrophoresis was carried out using gel sequencing apparatus by Life Technologies (Model S2001). The gel was pre-electrophoresed at constant power (55 W) for 60 min. Prior to loading, 10 μl formamide loading buffer (Appendix 1) was added to each reaction. Samples were denatured (GeneAmp PCR 9700) at 95°C for 2 min. Denatured PCR products (4.5 μl) were loaded onto the gel with ^{33}P labelled Molecular Weight Marker XIII (Roche-1721925) as a reference (Appendix 1). The running buffer was 1X TBE buffer (Appendix 1). The gel was run at 55 W, 2000 V and 135 mAMP for approximately 2 h. The gel was then transferred to Whatman paper

and dried on a gel drier (Hoefer GD 2000) at 80°C for 2 h and then exposed to Kodak X-ray film for 12 to 24 h. The X-ray film was developed in a darkroom using Kodak fixing and developing solutions (Appendix 1) and allowed to dry. The X-ray was captured by UV-Grabit software and analyzed first by visual rating of the loci and thereafter using Popgene Version 1.31 to obtain dendograms.

The total number of loci was calculated by visually examining the gels and counting the loci. Polymorphic bands were determined by scoring the loci that showed differences between the individual samples. This was also scored visually. The percentage polymorphism was calculated by dividing the polymorphic loci yielded by each primer pair by the total number of loci produced by that primer pair. Fragment size was determined by comparing the loci with the radiolabelled molecular weight marker that was loaded on each gel as a reference for size determination. The validity of the tests was ascertained by running a control, in our case tomato DNA. The banding pattern of the tomato DNA was checked for consistency in each run so as to determine the validity of each run and our consistency in assembling each set of PCR reactions. Repeatability and reproducibility of the method was ascertained by running duplicate and replicate assays. The data obtained from the visual scoring of the gels was then processed using Popgene software. Popgene is a user-friendly Microsoft Window-based computer package for the analysis of genetic variation among and within natural populations using co-dominant and dominant markers and quantitative traits. UPGMA is the simplest method of tree construction.

3.2. Screening Technique for Kunitz Trypsin Inhibitor Protein

This involved the extraction of Kunitz protein. Prior to analyzing the Kunitz protein the isoelectric focusing (IEF) technique for detection of this protein using polyacrylamide gels was optimized for pH range and staining technique. The optimization of the screening technique is illustrated if Figure 3-2. Ten protein extractions were done per parent line and were run on both wide pH and narrow pH range gels to determine the most suitable gel type. In order to ascertain the optimal staining procedure, gels were stained with both silver stain and R-250 brilliant blue. A trypsin inhibitor marker was run as a standard reference on all gels.

This involved the extraction of Kunitz protein as shown below in Fig 3.2

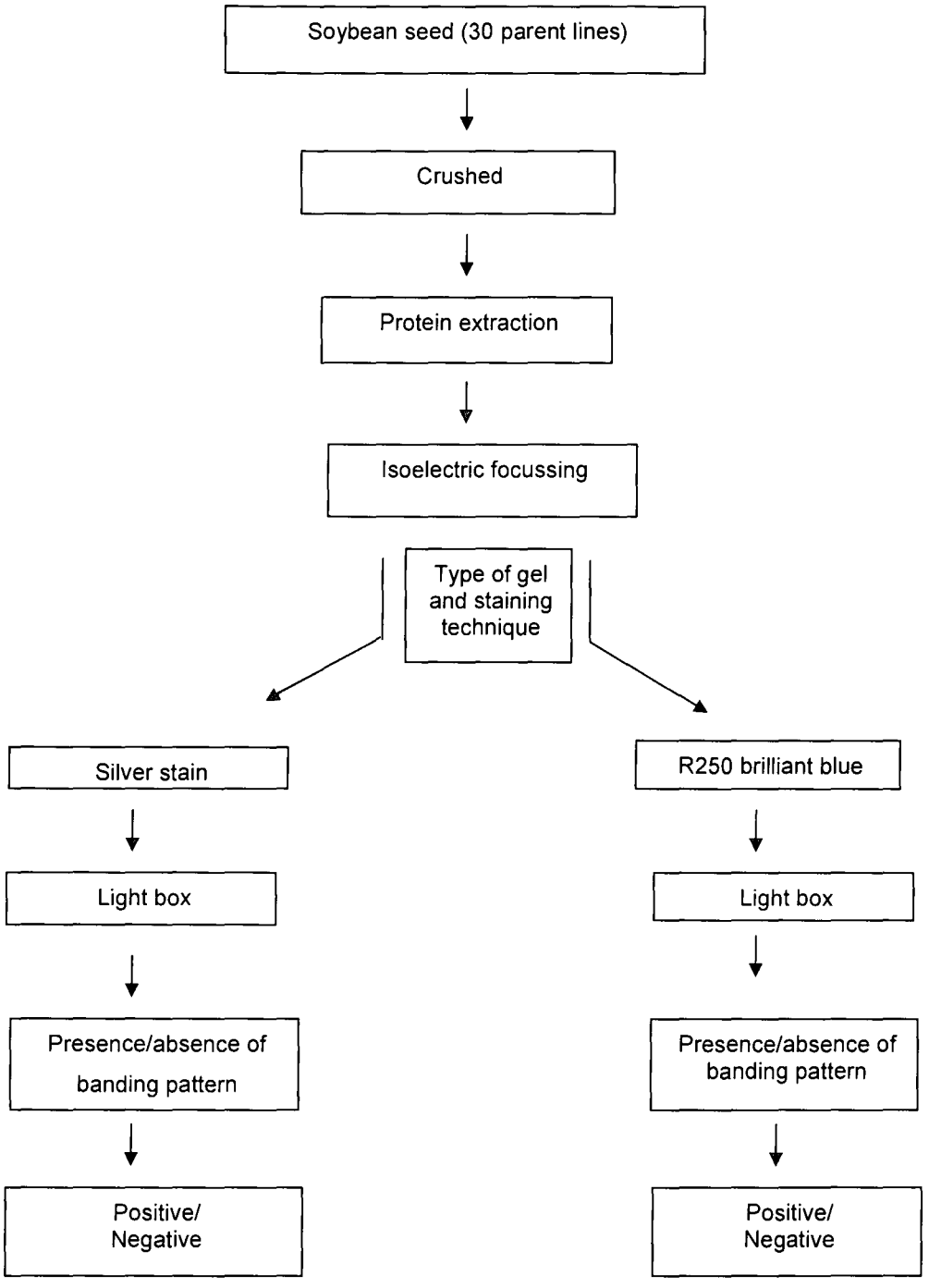

Figure 3-2. Optimisation of screening technique for the detection of Kunitz trypsin inhibitor protein

3.2.1. Sampling

Thirty soybean cultivars for the extraction of Kunitz trypsin inhibitor protein were obtained from Mr. A. Jarvie (Plant Breeder-Pannar Research, Greytown). Among the thirty cultivars were four control samples (sample numbers 8, 20, 21 and 24). The control samples were known controls used by the breeder in his soybean breeding programme. Samples numbers 8 and 21 did not contain the trypsin inhibitor and were thus the negative controls while Samples 20 and 24 were positive controls. From each of 30 different plants samples, 30 individual seed were randomly selected and extracted for the Kunitz trypsin inhibitor protein. This gave a total of 900 protein extractions. Each sample was run in triplicate to ensure reproducibility of the technique. Prior to extraction, the soybean samples were stored at RT (21°C) away from moisture.

3.2.2. Protein extraction from soybean seeds

The extraction procedure used in this method is an adaptation of the method of Hymowitz and Hadley (1972). Individual soybean seed were crushed using a mortar and pestle to a fine meal. A 1.5 ml aliquot of extraction buffer solution containing 1 ml of buffer solution (Appendix 1) and 0.5 ml of a 40% sucrose solution (Appendix 1) were added to the soybean meal. The debris was removed by centrifugation at 2000 x g for 10 min at 4°C (Harrier 18/80 centrifuge). The supernatent containing the protein was decanted into a storage vial and stored at –20°C.

3.2.3 Isoelectric focusing to screen for Kunitz trypsin inhibitor protein

The method prescribed with the flat bed focusing apparatus (Pharmacia Biotech Multiphor II) from Perkin-Elmer corporation was used. Precast agarose gels (FS 5080 and 5480, PerkinElmer, UK) in the pH range 4-5 (narrow range) and 3-10 (wide range) respectively were utilized for isoelectric focusing. The flat bed IEF apparatus was connected to a circulating waterbath and the temperature was set at 4°C. The gel was removed from its package and a reference corner was cut (top right). Water (2 ml) was placed on the cooling plate prior to placing the gel on the plate. The gel was centered on the cooling plate and excess water was blotted from the periphery of gel using paper towel. Three pre-cut IEF wicks (Perkin-Elmer, UK) were prepared. Two anode wicks, with their rough side down on paper towel (Kimberly Clark) were evenly saturated with 4 to 5 ml of anolyte solution (Appendix 1). Excess fluid was gently blotted off from wicks. The wicks were then placed on the gel and a finger was run along edge of wick to ensure even contact with gel. New gloves were put on and hands washed prior to preparing the third wick. A wick, rough side down, was saturated with 4 to 5 ml catholyte solution (Appendix 1). The wick was gently blotted with paper towel. Pre-focusing, was carried out by connecting to a power- supply (Major Science, MP-2000P) and running at constant power for 10 min at 30 W (voltage limit=1500). Following this period, the power supply was turned off and the wicks and electrodes blotted to remove excess buffer. Templates (52 well) (Perkin-Elmer, UK) were positioned 2 cm from central cathode, for the narrow

range gel (FS5080) and 4 cm from the cathode for the wide range gel (FS5480). 12 μl supernatent was applied to each well on the template. The reference Trypsin inhibitor protein marker (Sigma) was loaded in the middle. The gel was connected to a power supply and run at 40 W min for 30 min, followed by 40 min at 40 W, stopping midway through to blot and remove templates. At the end of the run, the power supply was turned off and the wicks removed. The gel was immediately stained with either silver or Coomassie stain. All gels were repeated three times in order to ensure reproducibility of the technique and this was achieved in 100% of the samples tested.

3.2.4 Comparison of staining techniques

The positive and negative control samples were run on two narrow range (pH 4-5) and two wide range (pH 3-10) agarose IEF gels. These were stained using two techniques, the silver staining technique (Perkin-Elmer) and the R250 Brilliant blue stain (Perkin-Elmer, UK).

3.2.4.1. Silver staining technique

The silver staining method for agarose gels (PerkinElmer, UK) for total protein detection was tested for reliability, ease of application, economic feasibility and time constraints. The gel was placed in a staining dish (Pyrex) and 200 ml of fixative solution (Appendix 1) was poured onto the gel. This gel was rocked on an orbital shaker (Vivid Air) for 20 min. The fixative solution was then discarded. The gel was washed while rocking for 30 min in 500 ml distilled water. The

distilled water was discarded and this step was repeated. The water was discarded and two pieces of Whatman filter paper saturated with distilled water was placed over the gel's surface. Several layers of paper towel was placed over the Whatman paper. A 10 kg weight was placed over the paper towel. The gel was press blotted for 10 min.

The gel was allowed to dry completely on a flat surface overnight at RT. All glassware used for the staining procedure was washed in Farmers Reducing Wash Solution (Appendix 1). The front of the gel was briefly rinsed in distilled water and subsequently placed in a clean staining tray (Pyrex), agarose side up. Distilled water (108 ml) was dispensed into a clean beaker with a magnetic stirrer bar. Solution A (23 ml) (Appendix 1), Solution B (23 ml) (Appendix 1) and Solution C (23 ml) (Appendix 1) was added to the beaker. Solution D (3.2 ml) (Appendix 1) was added and this was designated Solution #1.

Into another clean beaker, distilled water (108 ml) was added together with a magnetic stirrer bar and 5.5 g of sodium carbonate. The solution was stirred until dissolved. This was designated Solution # 2.

While stirring 90 ml of Solution #1 was added to 90 ml of Solution #2. The stain was immediately poured over the gel surface and the gel was rocked on an orbital shaker until the first bands began to appear. The staining solution was discarded and the gel was briefly rinsed in distilled water for 2 to 3 s. Both sides of the gel were gently wiped with wet paper towel (Kimwipes).

The gel was placed back into the glass staining dish and a second application of the stain was applied. While stirring, 90 ml of Solution #1 and 90 ml of Solution #2 was added to a beaker. The mixture was poured onto the gel and allowed to stain until the bands reached the desired intensity. The staining solution was discarded and the gel was briefly rinsed in distilled water for 2 to 3 seconds. Both sides of the gel were gently wiped with wet paper towel and stop solution (Appendix 1) was poured over the gel surface. The stop solution was discarded after 10 min and the gel was placed in 200 ml of distilled water for 10 min. After discarding the water the gel was allowed to air-dry overnight at RT.

3.2.4.2. R250 Brilliant Blue Staining

After electrophoresis the gel was placed into a staining dish (Pyrex) and 200 ml of fixative solution (Appendix 1) was added. The gel was rocked for 20 min on an orbital shaker (Vivid Air). The fixative solution was discarded and the gel was washed twice in 500 ml distilled water while shaking on an orbital shaker. The distilled water was discarded. Two pieces of Whatman filter paper saturated with distilled water were placed over the gel's surface. Several layers of paper towel were placed over the Whatman paper. A 10 kg weight was placed over the paper towel. The gel was press blotted for 10 min. The gel was allowed to dry completely on a flat surface overnight at RT.

The gel was placed in a staining dish, agarose side up and covered in 200 ml of Brilliant Blue R-250 stain (Appendix 1). The gel was rocked until the bands reached the desired intensity. The stain was discarded and 200 ml of destain

solution (Appendix 1) was applied. This procedure was repeated until the background was clear. The gel was then immersed in 500 ml of distilled water for 10 min. The gel was removed from the distilled water and air-dried overnight at RT.

3.3 Screening for the Kunitz Trypsin Inhibitor Protein

The wide range gel and Brilliant Blue staining procedure was used to analyse the thirty cultivars (30 extractions per cultivar) using the optimized conditions. The results of these were analysed visually using a light box. The banding patterns were visually scored against the Trypsin inhibitor protein standard (Sigma). The presence of loci produced by the protein standard was used to determine the presence/absence of the Kunitz trypsin inhibitor protein. Samples containing the trypsin inhibitor protein band were scored as positive and samples lacking the trypsin inhibitor band were scored as being negative. The number of samples that were positive or negative were merely counted from visual examination of the gels on a light box with reference to the trypsin marker that was loaded on each gel. Within each sample there were 30 individual protein extractions that were analyzed. The percentage positive was determined by dividing the number of samples with the presence of the trypsin band by the total number of extracts analyzed for the sample. Similarly, the percentage negative was determined by dividing the number of samples lacking the trypsin band by the total number of extracts analyzed for that sample.

CHAPTER FOUR: RESULTS

4.1 Genetic Diversity in *G. max*

The genetic diversity (polymorphism) among the forty *G. max.* parent lines using primer combinations (Mse 4-1/Mlu 1-5, E-AAC/ M-CAA, E-ACA/ M-CTT and E-ACA/ M-CAA) and comparing these using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) showed that Method A (K. Edwards, pers comm.) yielded too few polymorphisms (Figure 4-1) compared to Method B which yielded a larger number of polymorphisms for screening the genetic diversity (Figure 4-2) and therefore all subsequent experiments were carried out using Method B.

4.1.1 Repeatability, reproducibility and validity of AFLP technique

In order to facilitate scoring of the AFLP gels and to ensure that the method employed was repeatable as well as reproducible, replicates and duplicates of samples were loaded on gels (Figure 4-3). In order to ensure validity of the results obtained, an internal check namely, tomato DNA was loaded on AFLP gels (Figure 4-3). The internal check was examined on each duplicate gel for consistency in the banding patterns.

Using Methods A and B to analyse forty *G.max.* samples with primer combinations, Mse4-1/ Mlu1-5, E-AAC/ M-CAA and E-ACA/ M-CTT and E-ACA/ M-CAA, identified a total of 187 fragments of which 42 (22.4%) were

polymorphic. Figure 4-3 is a partial representation of an AFLP pattern obtained for Samples 1 to 17.

Based on the percentage of polymorphic fragments (polymorphic fragments divided by total number of fragments observed) primer pairs also detected different levels of polymorphism, ranging from 2.6% (Mse4-1/ Mlu1-5) to greater than 8% (E-ACA/ M-CAA). The frequency of individual polymorphic fragments observed among the forty lines varied greatly from 0.03 (polymorphism present in 1 line) to 0.9 (polymorphism present in 39 lines).

The major portion (32.5%) of the polymorphic AFLP fragments were detected in one plant, (17.5%) in two plants, (5%) in three plants and (2.5%) in four out of 40 plants. The remaining AFLP fragments were detected in five to 30 of the 40 plants.

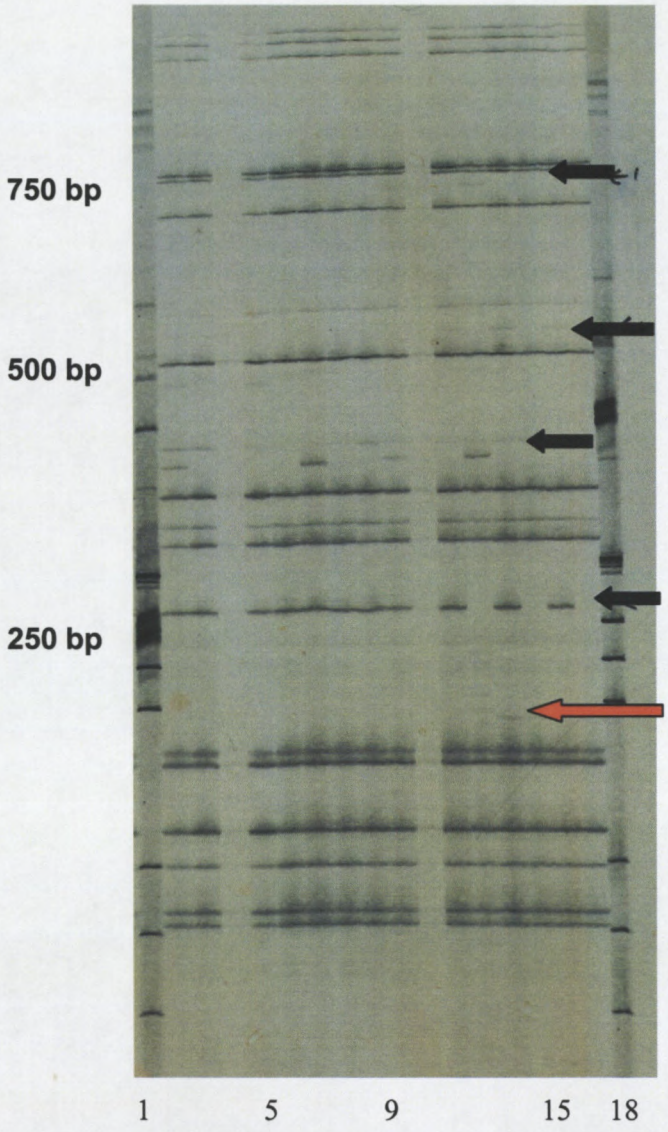

Figure 4-1. AFLP (Method A) using primer combination Mse 4 and Mlu 5. Lanes 1 and 18 contain the labeled marker. Lanes 2 to 17 contain Samples 1 to 16. Four polymorphic loci are detected using this primer combination as indicated by arrows.

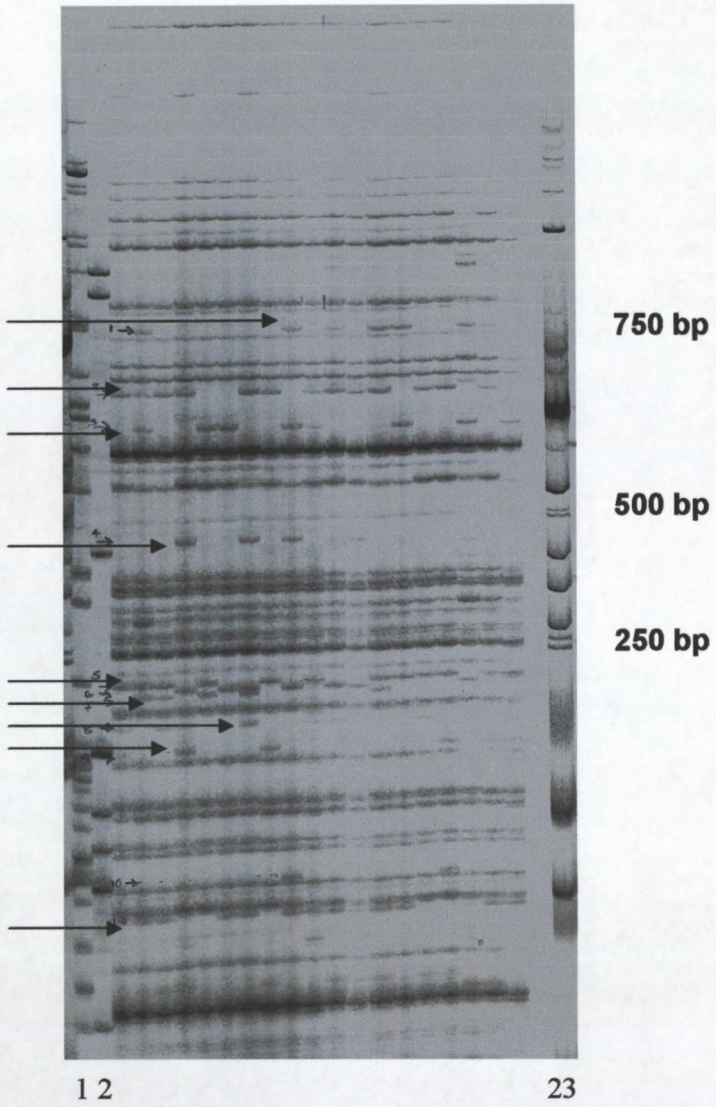

Figure 4-2 AFLP (Method B) using primer pair M-CAT/ E-ACA. Lane 1 was loaded with the control tomato DNA while lanes 2 and 23 contain the marker and lanes 3 to 22 contain Samples 1 to 20. Arrows mark polymorphic PCR fragments that represent potential AFLP markers in the 250-500 base pair region.

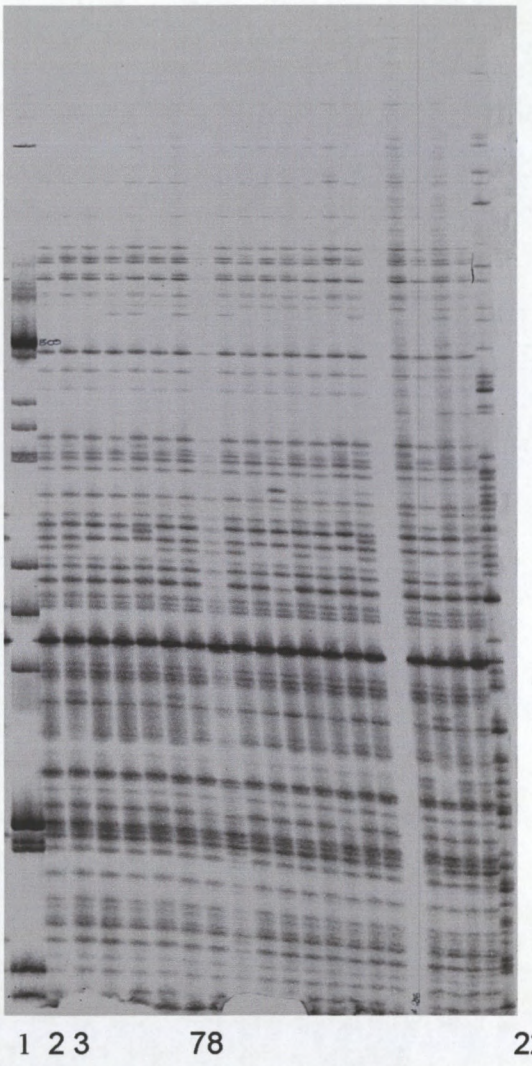

Figure 4-3 AFLP (Method B) using the primer combination M-CAA/ E-ACA. Lanes 2 and 3 were loaded with duplicates of the same sample while lanes 7 and 8 were loaded with replicates of another sample. Lane 22 contains the internal check (tomato DNA). The marker is loaded in lane 1 and Samples 2 to 17 are loaded in the remaining wells

Table 4-1 Levels of polymorphisms exhibited by the 40 soybean samples

Primer combination	Total number of fragments	Number of polymorphic bands	% Polymorphic bands
M-CAA/ E-AAC	24	5	20.8
Mse4-1/ Mlu1-5	49	11	22.5
M-CAT/ E-ACA	48	11	23
M-CAA/ E-ACA	66	15	22.7
TOTAL	187	42	
AVERAGE	46.75	10.5	
% Polymorphism		22.5	

4.1.2. AFLP fragment size

The size of AFLP fragments was determined by comparing sequencing ladder of Molecular Weight Marker XIII (Roche) to AFLP patterns (Figure 4-2). AFLP fragment sizes ranged from approximately 50 to 1500 base pairs. Polymorphic fragments were distributed across the entire size range. The major portion (72%) of polymorphic fragments was between 500 and 750 base pairs. The remaining (18%) of the polymorphic fragments were equally shared among the remaining size ranges of 750 to 1500 base pairs.

4.1.3. UPGMA analysis of data

The cluster analysis of soybean parents (40) generated from the AFLP data showed five clusters (Figure 4-4). Parental lines 1, 21, 13 and 15 were least related and therefore the furthest distance apart on the dendogram (Figure 4-4).

Samples 2 and 5 as well as Samples 14 and 19 were the most closely related parents. Samples 3, 4, 6, 7, 8, 9, 10, 11, 12, 16, 17, 18, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 and 40 are all an equal distance apart from each other in the dendogram.

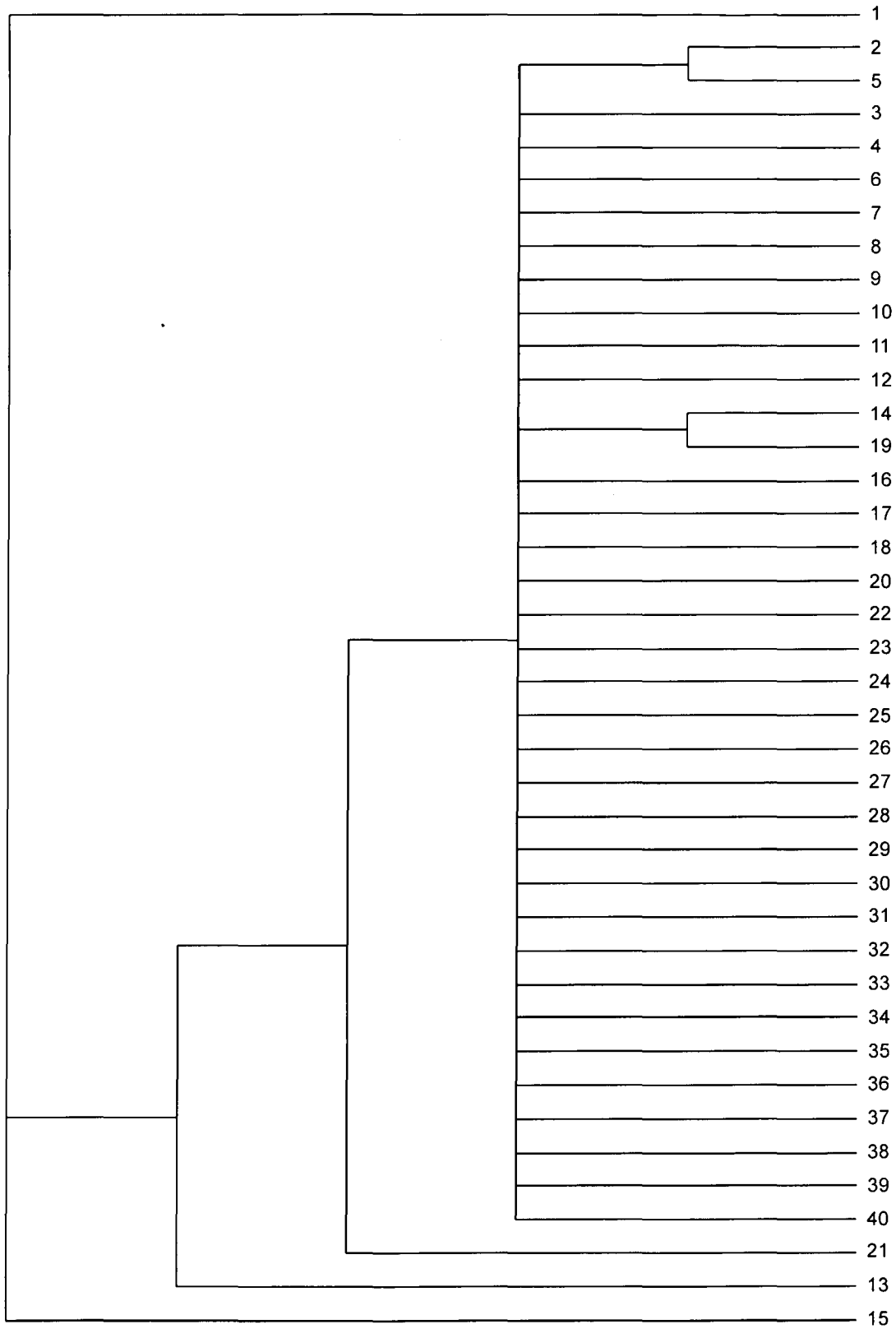

Figure 4-4. Composite dendogram of forty soybean lines obtained using four primer pairs for AFLP analysis

In order to develop a screening technique for Kunitz protein, optimization was carried out whereby the pH range of agarose gels and different staining techniques were investigated for the detection of Kunitz protein.

4.2. Optimization on Wide and Narrow Range Agarose Gels

A screening technique for Kunitz protein was optimized with five individual seeds chosen from each of the four control Samples 8, 20, 21 and 24. Samples 8 and 21 were known to lack the Kunitz protein while samples 20 and 24 were known to have it. Protein extractions were carried out and the extracts were run on both a wide range (pH 3-10) and narrow range (4-6) agarose IEF gels. Two protein stains viz. R-250 Brilliant blue stain and silver staining were used to visualize the Kunitz protein. When stained with R-250 Brilliant blue the narrow range gel (pH 4-6) did not stain with ease and the banding pattern observed was not well focused (Figure 4-5). The banding pattern observed was fuzzy and very few protein bands were visible. However, as expected the positive control samples (20 and 24) exhibited the Kunitz trypsin inhibitor protein band whilst the negative control samples (8 and 21) lacked this band (Figure 4-5). The same extracts loaded on the wide range gel (pH3-10) and stained with R-250 brilliant blue, showed a more defined banding pattern with a sharper and more defined banding profile. All the positive control samples showed the presence of the trypsin inhibitor protein band while all the negative control samples showed an absence of the protein (Figure 4-6).

A comparison of the silver stain and the R250 brilliant blue staining technique on the narrow range gel (Figure 4-7) showed less clarity in the banding profile as compared to the wide range gel (Figure 4-8). The narrow range gel showed a great deal more background staining than the wide range gel although the staining protocol was standardized. As expected Samples 20 and 24 (positive controls) showed the presence of the trypsin inhibitor protein and Samples 8 and 21 (negative controls) lacked the protein band.

The wide range gel (Figure 4-8) showed a more focused and clear banding profile with good separation of proteins. There was less background staining and the bands were more intense than the narrow range gel. However, the control samples did not show the expected banding patterns. Both the negative controls (Samples 8 and 21) showed a lack of the protein. Sample 20 and 24 are known positive samples. However, only Sample 24 exhibited the trypsin inhibitor protein and Sample 20 did not.

The results from the 900 soybean proteins extracts that were run on wide range agarose gels (Hypure FS-5480) and stained with R-250 brilliant blue stain are shown in Table 4-1. In order to ensure reproducibility of the technique, each of the 30 samples, (made up of 30 protein extractions each) were run on three separate occasions, on separate gels.

Figures 4-9, 4-10 and 4-11 illustrate the reproducibility of the technique by showing the three separate runs of the thirty individual seed extract from Sample number 5. Figure 4-9 shows the first run of the thirty protein extracts

from Sample number 5. The reproducibility of the technique is indicated by the consistent banding patterns obtained (Figure 4-10 and Figure 4-11) when the thirty individual protein extracts were run using our optimized technique on different occasions.

The results of 900 soybean protein extracts from 30 samples showed that 46.6% (n=14) of the samples tested were found to contain all extracts that were positive for the trypsin inhibitor protein, 16.6% (n=5) were found to contain all extracts that were negative for the protein, 36.6% (n=11) were found to contain both positive and negative extracts within a specific sample.

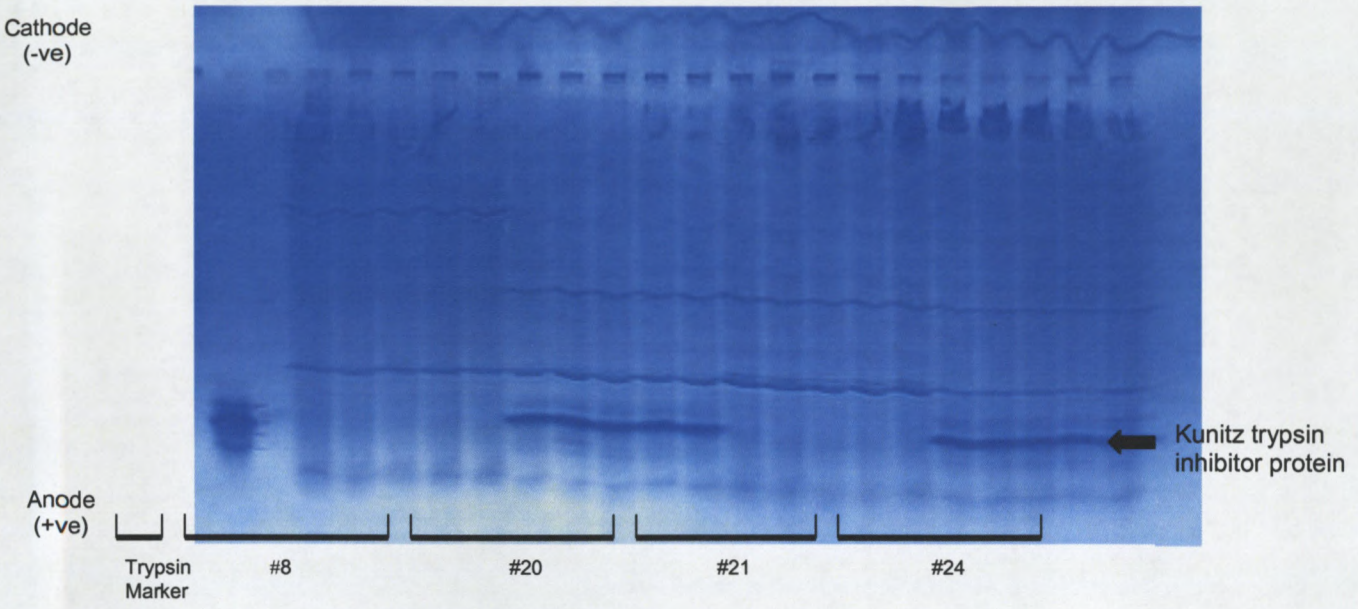

Figure 4-5. The protein extracts of 5 individual soybean seeds of Sample 8, 20, 21, 21 and 24 were loaded along with the Kunitz trypsin inhibitor standard on a narrow range (pH 4-6) agarose IEF gel. The gel was stained with R-250 brilliant blue. Samples 8 and 21 showed an absence of the trypsin inhibitor protein and Samples 20 and 24 showed the presence of the protein.

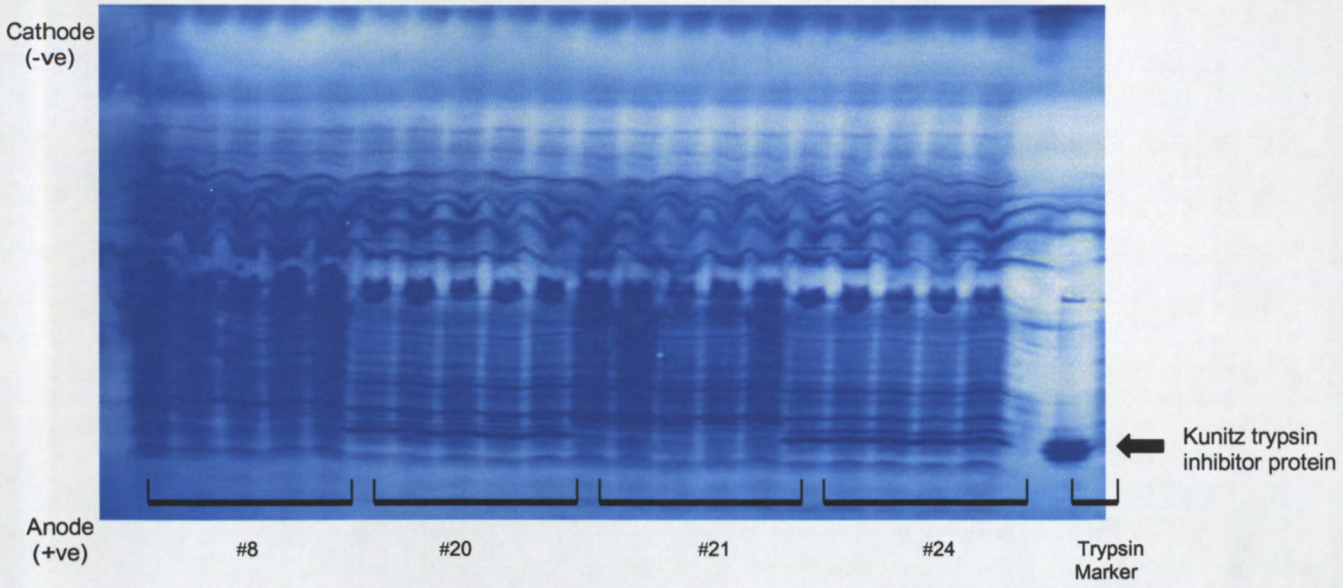

Figure 4-6 .The protein extracts of 5 individual soybean seeds of Samples 8, 20, 21, 21 and 24 were loaded along with the Kunitz trypsin inhibitor standard on a wide range (pH 3-10) agarose IEF gel. The gel was stained with R-250 brilliant blue. Samples 8 and 21 showed an absence of the trypsin inhibitor protein and Samples 20 and 24 showed the presence of the protein.

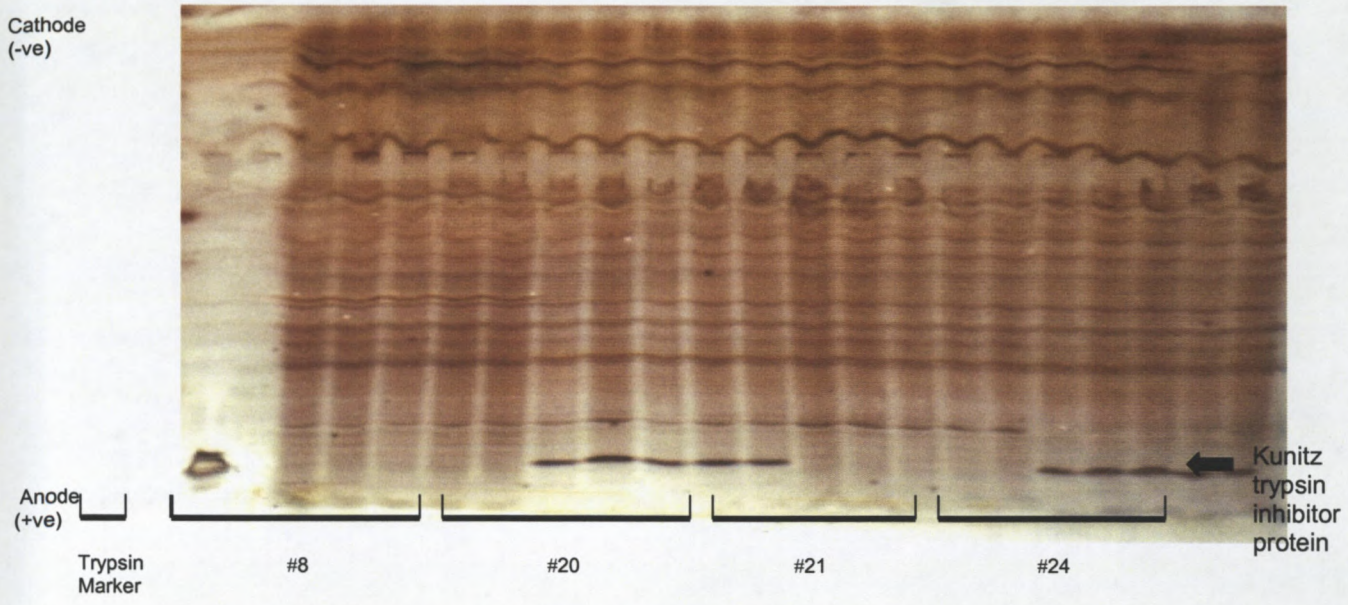

Figure 4-7. The protein extracts of 5 individual soybean seeds of Samples 8, 20, 21, 21 and 24 were loaded along with the Kunitz trypsin inhibitor standard on a narrow range (pH 4-6) agarose IEF gel. The gel was stained with silver stain. Samples 8 and 21 showed an absence of the trypsin inhibitor protein and Samples 20 and 24 showed the presence of the protein.

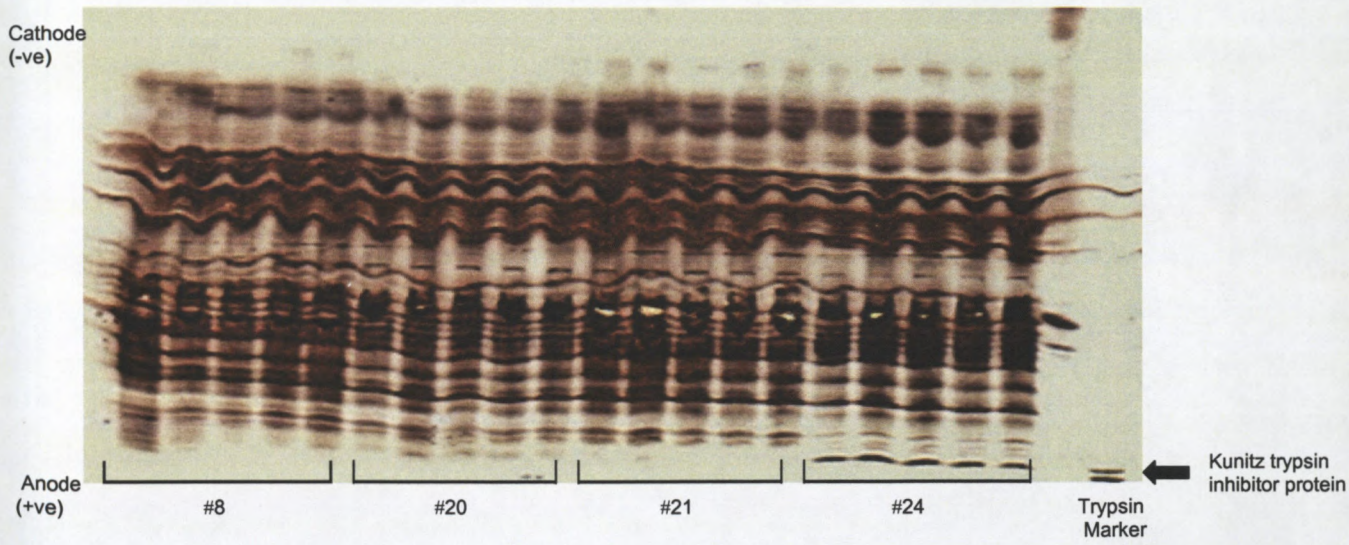

Figure 4-8. The protein extracts of 5 individual soybean seeds of Samples 8, 20, 21, 21 and 24 were loaded along with the Kunitz trypsin inhibitor standard on a wide range (pH 3-10) agarose IEF gel.The gel was stained with silver stain. Samples 8, 20 and 21 showed an absence of the trypsin inhibitor protein and Sample 24 showed the presence of the protein

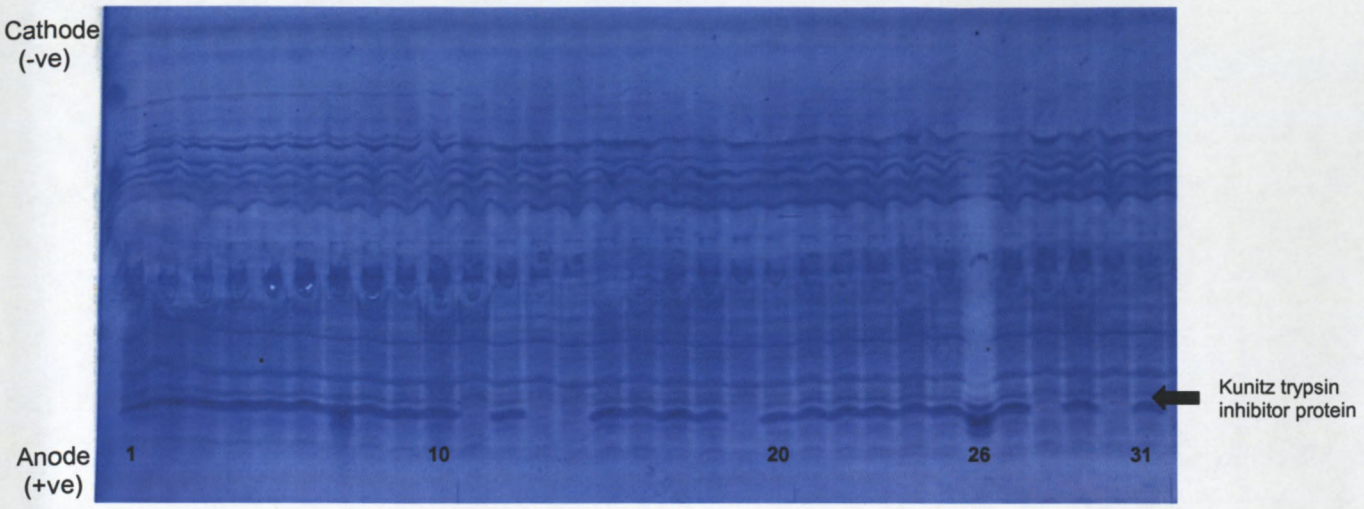

Fig. 4-9 Hypure 5480 gel loaded with Sample 5 to show reproducibility. Lanes 1 to 25 and lanes 27 to 31 are loaded with the 30 protein extractions from sample number 5 that were extracted from individual seeds, while lane 26 contains the Kunitz trypsin inhibitor marker. The arrow on the right of the gel indicates the Kunitz Trypsin inhibitor marker band.

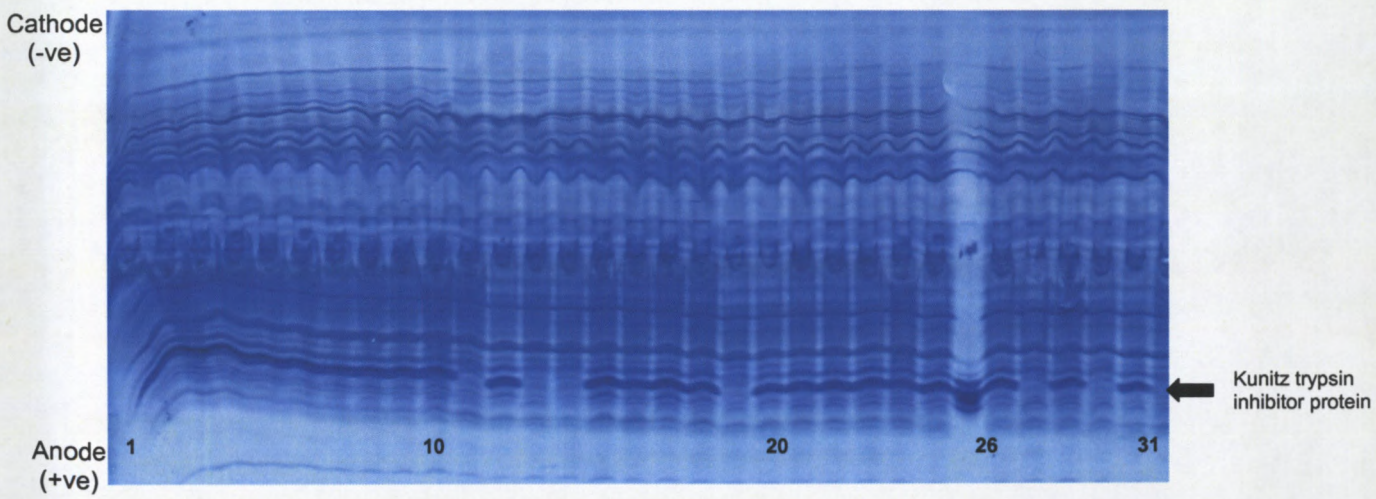

Fig. 4-10 A Hypure 5480 gel loaded with duplicates of Sample 5 (lanes 1 to 25 and lanes 27 to 31) and run on a separate occasion to ascertain the reproducibility of the technique. Lane 26 contains the Kunitz trypsin inhibitor marker. The arrow on the right of the gel indicates the Kunitz trypsin inhibitor marker band.

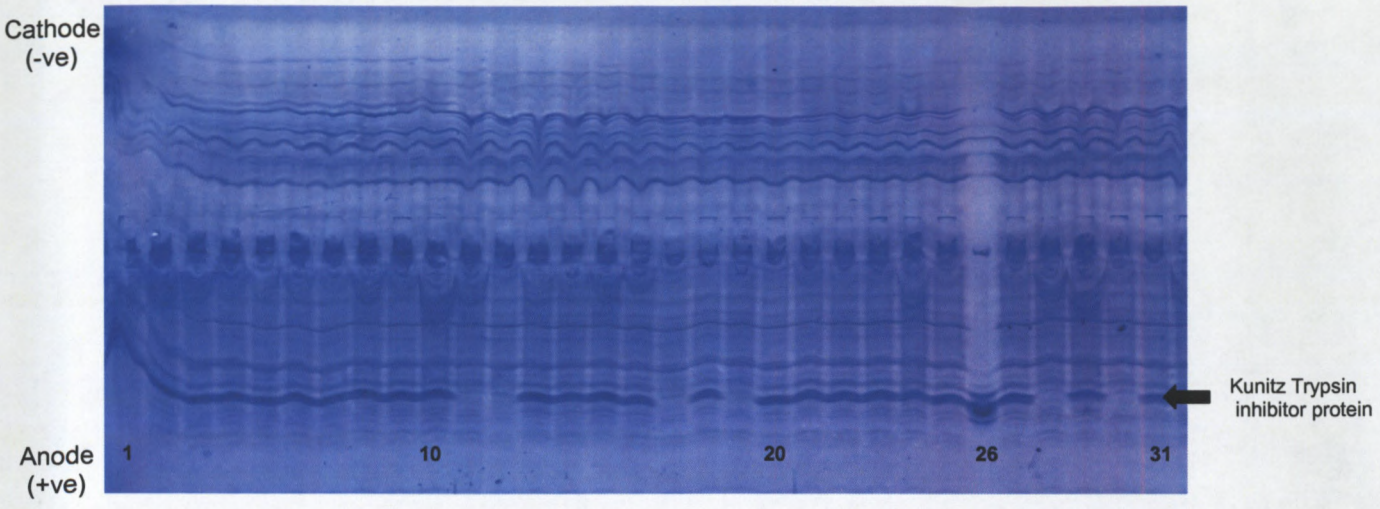

Fig. 4-11 A Hypure 5480 gel loaded with duplicates of Sample 5 (lanes 1 to 25 and lanes 27 to 31) and run on a separate occasion to ascertain the reproducibility of the technique. Lane 26 contains the Kunitz trypsin inhibitor marker. The arrow on the right of the gel indicates the Kunitz trypsin inhibitor marker band.

Table. 4-2. Summary of results of the 900* soybean protein extract samples that were run on Hypure (5080) agarose gels

Sample Number	Number Positive	Number Negative	% Positive For Kunitz Protein	% Negative for Kunitz Protein
1	30	-	100	0
2	30	-	100	0
3	30	-	100	0
4	-	30	-	100
5	24	6	80	20
6	30	-	100	0
7	21	9	70	30
8	-	30	-	100
9	30	-	100	-
10	30	-	100	-
11	24	6	80	20
12	30	-	100	-
13	18	12	60	40
14	27	3	90	10
15	30	-	100	-
16	24	6	80	20
17	30	-	100	-
18	21	9	70	30
19	15	15	50	50
20	30	-	100	-
21	-	30	-	100
22	-	30	-	100
23	21	9	70	30
24	30	-	100	-
25	30	-	100	-
26	30	-	100	-
27	18	12	60	40
28	24	6	80	20
29	-	30	-	100
30	30	-	100	-

* Thirty individual seeds were randomly selected from each of 30 plant samples Hence each Sample contains 30 seeds giving total of 900 extractions.

CHAPTER FIVE: DISCUSSION

In our study the AFLP assay, which is combination of the classical, hybridization-based approach and the PCR-based approach was used. This technique is reported to be an excellent tool in characterizing population variance and genetic diversity in soybean (Cheung, 2001).

Our studies show that the AFLP procedure is highly reproducible and efficient. The technique can be performed with ease and is able to produce polymorphisms even among small genotypes. The method has the ability to differentiate between closely related plants belonging to a small genotype, for example soybean.

5.1. Repeatability

It was important to assess the consistency of the AFLP banding patterns to examine the amount and possible sources of error involved in the procedure. In order to do this, duplicate assays and replicate assays were run on gels. Firstly, duplicate assays (one assay loaded twice on a gel) were run on different points of the same gel. Also replicate assays (the same DNA used to perform PCR on different occasions run on different gels) were run. The results were interesting in that, as anticipated, there was very little error for duplicate assays while there appeared to be a much greater amount of error in replicate assays. It is therefore useful according to Zeid (2003) that scoring

across gels should be backed by replicates and several appropriate check entries.

5.2 Reproducibility

During the replicate and duplicate assay trials it was found that there are several factors that affect the reproducibility of the assays. We found that DNA used for AFLP must be of the highest purity in order to ensure that complete digestion by restriction enzymes is accomplished. Incomplete digestion of DNA generates partial fragments, predominantly of high molecular weight as is indicated by the red arrow in Figure 4-1 in lane 14. Subsequently amplification of fragments that are not fully digested generates an altered banding pattern and this may be misinterpreted as false polymorphisms. Similar findings were reported by Blears *et al.* (1998). Blears *et al.* (1998) also stated that although the AFLP procedure is insensitive to the template DNA concentration, at very high template dilutions (picogram quantities), the nucleotide sequences flanking the restriction site will no longer be random for a small pool of restriction fragments and variations in the banding patterns may be observed. Typically, 0.05 μg of genomic DNA is required for small genomes ranging from 50 to 500 megabases and 0.5 μg for genomes of 500 to 5000 megabases. This study used 0.2 ng of genomic DNA. There was an observance of 'phantom' bands, however these bands were not scored. We were able to differentiate between the common banding patterns and 'phantom' banding patterns in the samples by examining the replicates and especially the duplicates of each sample.

5.3. Ease of operation

AFLP markers were chosen because of their advantages over other markers available for this type of research. A successful AFLP assay of a genotype results in amplification products that are separated by gel electrophoresis. The presence of the PCR products is detected as banding pattern seen on X-ray film. Scoring of the banding pattern was performed visually. The bands were rated in a 1/0 matrix, indicating presence or absence of a band respectively.

5.4 Genetic Diversity of Local Soybean

Our studies using 40 parent lines and four primer pairs detected a total of 187 fragments of which, 22.4% were polymorphic. The number of polymorphic loci per primer pair ranged from 5 to 15 in the 40 samples. Each primer pair detected an average of 46 loci. These results reflect the ability of the AFLP method to be highly informative in genetic diversity studies. The level of polymorphism was 22% and this value is among the highest that can be attained in a self-pollinating crop such as soybean with the use of classical, hybridization and PCR-based molecular methods.

In another study by Maughan *et al.,* (1996) that used AFLP for species diversity, inheritance and near-isogenic line analysis AFLP primer pairs detected a total of 759 AFLP fragments in a sample of 23 accessions of wild and cultivated soybean, with an average of 51 fragments produced per primer pair per accession. Two-hundred and seventy four fragments (36%) were

polymorphic, among which 127 (17%) were polymorphic in *G.max.* and 237 (31%) were polymorphic in *G.soja.* The number of polymorphic loci detected per AFLP primer pair in a sample of 23 accessions ranged from 9 to 27.

Welsh and McClelland. (1990) used RAPD to detect variation in species. RAPD is capable of creating polymorphisms without any prior knowledge of the DNA sequences of the species. The patterns produced are highly polymorphic, allowing discrimination between isolates of a species if sufficient numbers of primers are screened. The method is fast and economic for screening large numbers of samples. However, there seems to be poor reproducibility of RAPD patterns (Tommerup *et al.*, 1995). This implies that diagnostic RAPD markers may not be confirmed by another laboratory, thus posing uncertainty for data cross-referencing in race/ type identifications. The main limitation of RAPD analysis in population studies is the dominant character of RAPD markers. In the study of diploid organisms, homozyote AA can not be distinguished from heterozyote Aa, since both will give a RAPD pattern with a band corresponding to a A (Wang and Szmidt, 2001). Not all RAPD fragments are equally informative. Only 34% of the markers used by Thompson *et al.* (1998) to evaluate diversity of 18 North American soybean ancestors and 17 PIs from the USDA Soybean Germplasm Collection were polymorphic.

A comparison (Table 5-1) of the use of different genetic methods for soybean fingerprinting i.e. RFLP, PCR, RAPD and our study AFLP showed that this technique can differentiate among different soybean parents and is able to

provide the breeder with useful information on the relationship between the parent lines and their potential use in cultivar development.

Table 5-1 A comparison of the different genetic techniques used to fingerprint soybean

	Principle of technique	Description of samples tested	Findings
RFLP	Co-dominant markers that can be detected by analyzing restriction digests of genomic DNA through Southern hybridization	58 *G.soja* and *G.max* accessions including ancestral lines and cultivars	Molecular diversity was 0.16 for the cultivars and 0.26 for the ancestors (Kiem *et al.*,1989)
		38 soybean genotypes were surveyed with 132 RFLP probes	Molecular diversity was 0.30 (Kiem *et al.*, 1992)
		Elite and ancestral lines	Identified 29 RFLP probes with gene diversity scores ≥0.30 (Skorupska *et al.*,1993)
		Elite and ancestral cultivars in North America	Identified 67 polymorphic RFLP markers with gene diversity scores of ≥0.30 (Lorenzen *et al.*, 1995)
RAPD	Arbitrary short oligonucleotide primers targeting unknown sequences in the genome are used to generate amplification products that show size polymorphisms within species	18 soybean genotypes	Out of the 33 random primers used, only 12 showed polymorphisms useful for characterization of genotypes (Doldi *et al.*, 1997)
		18 North American soybean ancestors and 17 PI's from the USDA Soybean Germplasm Collection	Only 34% of the markers used were polymorphic
AFLP	DNA is digested, adapters ligated to the ends and amplified by PCR. Amplified fragments are analyzed through PAG electrophoresis after pre-selection of fragments.	23 accessions of wild and cultivated soybean	AFLP primers detected a total of 759 AFLP fragments, with an average of 51 fragments produced per primer pair. 36% of the fragments were polymorphic. The number of polymorphic loci per primer pair ranged

			from 9 to 27 (Maughan *et al.*, 1996)
Our Study (AFLP)	Used AFLP analysis	40 soybean parent lines	Primers detected a total of 187 fragments of which 42% were polymorphic.

5.5 Screening for Kunitz Trypsin Inhibitor Protein

5.5.1. Method development

In this method development stage both a narrow range (4-5) and a broad range (3-10) IEF gel were investigated since it was known that the pI of the kunitz trypsin inhibitor protein was 4.6. From the literature it was ascertained that agarose gels can be stained for proteins using two types of staining techniques R-250 brilliant blue stain and silver stain. The silver stain is considered to be to be a more sensitive stain. It was therefore decided that the control samples would be run on both wide and narrow range gels under standardized conditions and to stain the gels with both types of protein stains in order to determine which would be the optimal staining technique. All gels used for the optimization were loaded with the same samples and loading sequence. All parameters in terms of protein extraction, running conditions and amount of extract loaded were standardized. There were three basic differences in the running of the two gel types. The first was in the stains that were applied i.e. one wide gel was stained with silver while the other was stained with R-250 brilliant blue and one narrow gel was stained with silver while the other was stained with R-250 brilliant blue. The second difference

was that the point of application of the sample loading templates differed for the wide and narrow range gels. For the narrow pH range gel the sample templates were placed 2 cm from the central cathode while for the wide range gel the sample templates were placed 4 cm from the central cathode. The third difference was that the catholyte solution for the two gel types were different. A 0.5 M NaOH cathode solution was used for the narrow range gel and a 0.25M NaOH cathode solution was used for the wide range gel.

5.2.2 Ease of operation

The Kunitz trypsin inhibitor protein marker has a pI of 4.6 and was visible on both gel types when stained with R-250 brilliant blue. Figures 4-5 and 4-6 differences in resolution of the banding patterns produced by the two gel types. The Kunitz protein was visible on both the wide and narrow range gels when stained with R-250 brilliant blue as the isoelectric point of the protein falls within the window of both gel types. However, there were differences in terms of the resolution, number of bands and field of separation on the two gels. The wide pH range gel showed better separation of bands and a higher resolution, while the narrow range gel exhibited fuzziness and more background staining. The banding patterns observed on both gels correlated with the expected banding patterns of the control samples.

The silver staining technique applied to the two gel types (Figures 4-7 and 4-8) was unsuccessful in the case of the wide range gel. Positive control sample twenty showed a negative banding pattern. However, the controls run on the

narrow range gel showed the expected banding pattern. Silver staining is a more sensitive staining technique for proteins on polyacrylamide gels however agarose gels are not ideally suited for silver staining (Westermeier, 2001). Once again the wide range gel showed a greater resolution when silver stained than the narrow gel. This could be a result of the difference in the catholyte buffers used on the gels and the presence of different ampholytes. The point of application of the extracts could perhaps have also played a role in the difference in resolution observed in the gels.

At the end of the optimization process it was clear that the R-250 brilliant blue staining technique was superior over the silver stain for the detection of the Kunitz protein on agarose gels. It was also concluded that the wide pH range gels were more suited for this screening method due to its greater resolving power.

5.5.3. Method application

Nine hundred soybean protein extracts were run on wide range gels and stained using R-250 Brilliant blue stain. Table 4-2 shows a summary of the results obtained on each of the three runs. All protein extracts were run in triplicate on three separate occasions in order to ensure reproducibility of the technique. There was a 100% correlation of the results obtained for each extract when run in triplicate on IEF. This indicates that the technique is highly reproducible.

5.6. Conclusions

Plant breeding programmes were more traditional previously. Selection of parental material for breeding was done by planting out the seed and waiting for the adult plants to grow, before selecting for desirable traits that could be incorporated into the parental material for future crosses. This process was time-consuming and did not necessarily ensure the selection of the best genetic stability as it was based on phenotypic expression alone. However breeding programmes are enhanced by DNA technology, for example AFLP analysis, which in conjunction with traditional breeding methods, can speed up the development of soybean varieties. This is accomplished by being able to select parent material for the development of lines by AFLP analysis and gaining knowledge on the genetic distances between lines before seeds are planted. This may result in a speedier and more accurate selection procedure. Also AFLP analysis can be used to determine homozygosity and heterozygosity of parents. The AFLP analysis has therefore made a tremendous impact on plant breeding programmes.

The objectives of the study were fulfilled in that the screening technique optimized for the detection of the Kunitz protein on IEF was economical. Start-up costs would be minimal as the equipment is readily available. The cost of analysis per soybean seed tested was in the region of R4-50. The technique proved to be highly accurate and reproducible. When extracts were run in triplicate, there was 100% correlation for each run. From a safety point of view, it is more desirable to use agarose gels as opposed to polyacrylamide

gels. This method proved efficient as we are able to run 100 individual protein extracts on each gel. The method was not as time-consuming as other electrophoresis techniques e.g. vertical PAG electrophoresis. The total time required, from the sample preparation stage to the final destaining of the gel is 5 h. Results can therefore be available within a day of receiving samples. All of the above reasons contribute to making this screening technique desirable for use in a commercial seed testing laboratory where time is of the essence and the volume of seed to be tested is large.

All soybean varieties sold for commercial use by Pannar can now be tested at the protein level for the Kunitz trypsin inhibitor protein using the optimized protocol. The knowledge of its Kunitz status is important to the farmers as it will determine the use of the soybean. For example, the presence of the protein will mean that the soybean will be unsuitable as feed for monogastric animals. It will also determine the processing necessary if the seed is intended for food purposes. It is therefore vital that farmers purchasing soybean seed know the characteristics of the seed in order for them to determine its use. We have managed to optimize a protocol for the determination of the Kunitz protein without resorting to the DNA level, which is more time-consuming and expensive. This protocol also ensures results within a day of receipt of a sample.

REFERENCES

Applied Biosystems. 1997. AFLP Plant Mapping. USA. Perkin-Elmer.

ARS News and Information. 2001. *There's Treasure in the Soybean Genome Map* [online]. Available from: http://www.ars.usda.gov/is/AR/archive/feb99/map0299.html [Accessed 3 September 2002].

Blears, M.J., De Grandis, S.A., Lee, H. and Trevors, J.T. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology and Biotechnology* **21:** 99-114

Cheung, W.D. 2001. *AFLP DNA Fingerprinting of Colletotrichum spp* [online]. Available from: http://www.wpi.edu/~wcheung/rppaper/rp.html [Accessed 23 March 2001].

Cregan, P.B. 1999. There's a treasure in the soybean genome map. *Agricultural Research Magazine* **12:** 56-62

Doldi, M.L., Vollmann, J. and Lelley, T. 1997. Genetic diversity in soybean as determined by RAPD and microsatellite analysis. *Plant Breeding* **116:** 331-335.

Duranti, M., Barbiroli, A., Scarafoni, A., Tedeschi, G. and Morazzoni, P. 2003. One step purification of Kunitz soybean trypsin protein. *Protein Expression and Purification* **30:** 167-170.

Fehr, W.R. 2001. *Soybean Breeding* [online]. Available from: http://www.agron.iastate.edu/soybean/soybreed.html [Accessed 29 March 2001]

Flavin, D.F. 1982.The effects of soybean trypsin inhibitors on the pancreas of animals and man: a review. *Veterinary and Human Toxicology* **24:** 8-25.

Harper, D. 1997. In the field with herbicide resistant crops: RoundUp Ready soybeans. *Proceedings of the Western Society of Weed Science*, p8.

Hildebrand, D.F. and Hymowitz, T. 1980. Rapid test for Kunitz Trypsin inhibitor activity in soybean seeds. *Crop Science* **20:** 818-819.

Hymowitz, T. and Hadley, H.H. 1972. Inheritance of a trypsin inhibitor variant in seed protein of soybeans. *Crop Science* **12:**197-198.

Hymowitz, T. 2001. *The Kunitz Soybean Variety* [online]. Available from: http://www.ag.uiuc.edu/~vista/html_pubs/irspsm91/kunitz.html [Accessed 21 March 2001].

IFR Communuications. 2001.*Soya* [online]. Available from: http://www.ifrn.bbsrc.ac.uk/public/FoodInfoSheets/soya.html [Accessed 6 March 2001].

Jofuku, K.D. and Goldberg, R.B. 1989. Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. *The Plant Cell* **1:** 1079-1093.

Jones, P., Sutton, J.M. and Sutton, M. 1997. Plant molecular biology: essential techniques. New York: John Wiley and Sons.

Kiem, P., Shoemaker, R.C. and Palmer, R.G. 1989. Restriction fragment length polymorphism diversity in soybean. *Theoretical and Applied Genetics* **77:** 786-792.

Kiem, P., Beavis, W., Schupp, J. and Freestone, R. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. *Theoretical and Applied Genetics* **85:** 205-212.

Life Technologies. 2002. AFLP Analysis System, Foster City, California.

Lorenzen, L.L., Boutin, S., Young, N., Specht, J.E., and Shoemaker, R.C. 1995. Soybean pedigree analysis using map-based markers: tracking RFLP markers in cultivars. *Crop Science* **35:** 1326-1336.

Maughan, P.J., Maroof, M.A.S., Buss, G.R., and Huestis, G.M. 1996. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. *Theoretical and Applied Genetics* **93:** 392-401.

National Academy of Sciences. 2001. *Limits of Traditional Breeding* [online]. Available from: http://www.4.nas.edu/beyond/beyonddiscovery.nsf/web/seeds.html [Accessed 21 March 2001].

Orf, J.H. and Hymowitz, T. 1979. Inheritance of the absence of the Kunitz trypsin inhibitor in seed protein of soybeans. *Crop Science* **19:** 107-109

Penner, G.A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., Laroche, A., Scoles, G. and Fedak, G. 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods and Applications* **2:** 341-345.

Purseglove, J.W. 2003. Trpoical crops – dicotyledons. John Wiley and Sons Inc., New York.

Roychaudhuri, R., Sarath, G., Zeece, M. and Markwell, J. 2003. Reversible denaturation of the soybean Kunitz trypsin inhibitor. *Biochemistry* **412**: 20-26.

Skorupska, H.T., Shoemaker, R.C., Warner, A., Shipe, E.R., and Bridges, W.C. 1993. Restriction fragment length polymorphism in soybean germplasm of the southern USA. *Crop Science* **33:** 1169-1176.

Singh, L., Wilson, C.M., and Hadley, H.H. 1969. Genetic differences in soybean trypsin inhibitors separated by disc electrophoresis. *Crop Science* **9:** 489-491.

Specialised Protein Products.2003. *About Soy* [online]. Available from: http://www.Specialised Protein Products/frame_aboutsoy.html [Accessed 20 June 2003]

Thompson, J.A., Nelson, R.L. and Vodkin, L.O. 1998. Identification of diverse soybean germplasm using RAPD markers. *Crop Science* **38:** 1348-1355.

Tommerup, I.C., Barton, J.E., and O'Brien, P.A. 1995. Reliability of RAPD fingerprinting of three basidiomycete fungi, *Laccaria, Hydnangium* and *Rhizoctonia. Mycology Research* **99:** 179-186.

Wang, X., and Szmidt, A.E. 2001. *Population Genetic Analysis of Pathogenic Fungi* [online]. Available from: http://133.5.173.66/pgen/abstract/xrwbook98.html [Accessed 24 October 2001].

Welsh, J., and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**: 7213-7218.

Westermeier, R. 2001. Electrophoresis in Practice. Third Edition. New York: Wiley-VCH.

Whigham, K. 2001. *Soybean Usage* [online]. Available from: http://www.agron.iastate.edu/soybean/usage.html [Accessed 29 March 2001]

Winter, P. 2003. Soy it isn't so (online). Available from: http://www.handpen.com/Bio/medical/html [Accessed on 9 July 2002].

Zeid, M.M. 2003. Analysis of genetic diversity based on molecular markers (AFLP) and of heterosis in faba bean (*Vicia faba L*). Doctoral Dissertation, Geor-August-University Gottingen, Germany.

APPENDIX 1

A. Procedure for GenElute Plant Genomic DNA kit (Sigma-G2N350)

1. In order to release DNA from plant tissue, the plant material was ground in liquid nitrogen, using a mortar and pestle. Up to 100 mg of ground plant tissue was lysed with 350 μl of lysis solution (Part A) and 50 ul of lysis solution (Part B). The mixture was vortexed and inverted to mix thoroughly. The samples were incubated at 65°C for 10 min.
2. The debris was removed by adding 130 μl of precipitation solution to the samples and inverting to mix. The samples were then incubated on ice for 5 min. The debris was pelleted by centrifugation at 12,000 x g for 5 min.
3. The DNA was then bound to a column. 700 μl of binding solution was added to the filtrate and mixed thoroughly by inversion. 700 μl of the mixture was transferred to a binding column and allowed to centrifuge at 12,000 x g for 1 min. The flow through was discarded. This step was repeated with the remainder of the mixture. The column was then transferred to a new collection tube.
4. The sample was then washed to remove contaminants. 500 μl of wash solution was added to the column and this was centrifuged for a min. A second 500 μl wash solution was added to the column and the step was repeated.

5. The final step was the elution of the purified DNA. The column was transferred to a new collection tube and 100 µl of elution solution (pre-warmed to 65°C) was added to the column. This was placed in a centrifuge for 2 minutes. The column was discarded and samples were stored at –20 °C.

B. Procedure for determination of DNA concentration by agarose gel electrophoresis.

1. A 0.7% agarose gel was made by adding 10 ml of 10X TBE buffer to 90 ml of water and 0.7 g of agarose. The mixture was heated in a microwave for 2 min until the agarose had melted.
2. The gel was allowed to cool to approximately 40°C and 10 µl of ethidium bromide (10 mg/ml) was added to the gel.
3. The gel was poured into a gel cast (OWL Scientific) fitted with combs in order to make wells.
4.

C. Reagents for determination of genomic DNA concentration

- 0.7 % ($^{w}/_{v}$) agarose gel (Amresco-9012366)
- 10X Tris-borate (TBE) buffer (pH 8.3)- Dissolve 108 g Tris (Sigma-T1378), 55 g Boric acid (ACE- analytical grade) and 9.3 g Ethylene Diamine Tetra-aceticacid Di-sodium salt (EDTA) (Saarchem-Analytical grade) in 1000 ml distilled water.
- 1 X TBE -Dilute 1 part of 10X TBE buffer with 9 parts distilled water
- Lamda DNA (Promega-6638601)
- 1% ($^{w}/_{v}$) Ethidium Bromide ((Sigma-E4391) - Dissolve one tablet in 1 ml of distilled water

D. Reagents for selection of biotinylated DNA fragments (Method A)

- Wash buffer (10 mM Tris-HCl / 0.1 mM EDTA / 100 mM NaCl)
- TE buffer (10 mM Tris-HCl / 0.1 mM EDTA)

E. Reagents for Method B

Restriction digestion of genomic DNA

- 5X reaction buffer (50 mM Tris-HCl pH 7.5/ 50 mM magnesium acetate/ 250 mM potassium acetate)
- *Eco*RI/ *Mse* I (1.25 units/ μl each in 10 mM Tris-HCl pH 7.4/ 50 mM NaCl/ 0.1mM EDTA/ 1 mM DTT/ 0.1 mg/ ml BSA/ 50% glycerol ($^{v}/_{v}$)/ 0.1% Triton X- 100)

Ligation of Adapters

- Adapter ligation solution (*EcoR*I/ *Mse*I adapters/ 0.4 mM ATP/ 10 mM Tris-HCl [pH 7.5] / 10 mM magnesium acetate/ 50 mM Potassium acetate)
- T4 DNA ligase (1 unit/ μl in 10 mM Tris-HCl [pH 7.5] / 1 mM DTT/ 50 mM potassium chloride/ 50% ($^{v}/_{v}$) glycerol)
- TE buffer (10 mM Tris-HCl [pH 8]/ 0.1 mM EDTA)

Preamplification

- Pre-amp primer mix (Invitrogen-50142)
- 10X PCR buffer (200 mM Tris-HCl [pH 8.4]/ 15 mM magnesium chloride/ 500 mM potassium chloride)
- Amplitaq DNA polymerase –1 unit/µl (Roche-N8080161)

Primer labeling

- *Eco*RI primer (27.8 ng/ µl)
- 5X kinase buffer (350 mM Tris-HCl [pH 7.6]/ 50 mM magnesium chloride/ 500 mM potassium chloride/ 5 mM 2-mercaptoethanol)
- [Gamma ^{33}P] ATP (3000 Ci/mmol)
- T4 kinase (10 units/ µl in 50 mM Tris-HCl [pH 7.6]/ 25 mM potassium chloride/ 1 mM 2-mercaptoethanol/ 0.1 µM ATP/ 50% ($^{v}/_{v}$) glycerol

Selective AFLP amplification

- *Mse*I primer (6.7 ng/µl and dNTPs)
- 10X PCR buffer plus Mg (200 mM Tris-HCl [pH 8.4]/ 15 mM magnesium chloride/ 500 mM potassium chloride)
- Amplitaq DNA polymerase- 5 units/ µl (Roche-N8080161)

Analysis of PCR products on polyacrylamide sequencing gel

- 4 % denaturing polyacrylamide gel – dissolve 30 g of urea (AEC Amersham) in 7.5 ml of 19:1 acrylamide-bis gel stock (Gibco BRL), 7.5 ml 10X TBE and 30 ml distilled water. Make up to 75 ml with distilled water. Add 40 µl N tetramethylethylenediamine (TEMED) (BDH) and 400 ul of 10 % ammonium persulphate (Stratagene). Shark-tooth combs were inserted and the gel was allowed to polymerize overnight.
- 10X Tris-borate (TBE) buffer (pH 8.3)- Dissolve 108 g Tris (Sigma-T1378), 55 g Boric acid (ACE- analytical grade) and 9.3 g Ethylene Diamine Tetra-aceticacid Di-sodium salt (EDTA) (Saarchem-Analytical grade) in 1000 ml distilled water.
- Formamide loading buffer (98 % formamide/ 10 mM EDTA/ 0.01% bromophenol blue/ 0.01% xylene cyanol)
- Labelled MWM XIII (Roche-721925) – Add 0.5 µl of One-phor-all buffer (AEC Amersham) to 2 µl of the marker and 0.3 µl of (10 U/ul) kinase (USB). Add 1 µl of gamma ^{33}P. Incubate the mixture at 37°C for 15 min followed by a further incubation at 65°C for 15 min. Add 3.8 µl of TE buffer and 23 µl of formamide loading buffer.
- Kodak fixing replenisher- add 473ml of distilled water to 103 ml fixing solution
- Kodak developing replenisher- add 473 ml distilled water to 103 ml developing solution.

F. Reagents for Kunitz protein extraction from soybean samples

- Buffer solution (0.092 M Tris-HCl/ 0.023 M $CaCl_2.2H_2O$ brought to pH 8.1 with HCL)
- 40 % sucrose solution (dissolve 40 g of sucrose (Saarchem) in 100 ml water.

G. Reagents for isoelectric focusing

- Anolyte solution (0.5 M acetic acid) (BDH- analytical grade)
- Catholyte solution (0.5 M NaOH (Saarchem- analytical grade)/ 30 mM Lysine (Sigma-L5501)/ 30mM L-Arginine (Sigma- A5006)
- Trypsin inhibitor protein marker (Sigma-T102)

H. Reagents for Silver staining procedure

- Fixative solution- Add 40 g trichloroacetic acid (ACE- analytical grade) to 200 ml of distilled water. Stir until dissolved.
- Farmers Reducing Solution- dissolve 2.5 g potassium ferricyanide (Saarchem- Analytical grade) and 4 g sodium thiosulfate (Saarchem- Analytical grade) in 200 ml distilled water.
- Solution A- Dissolve 2 g of ammonium nitrate (Saarchem- Analytical grade) in 100 ml of distilled water.
- Solution B- Dissolve 2 g of silver nitrate (Sigma-S6506) in 100 ml distilled water.
- Solution C- Dissolve 10 g of tungstosilicic acid (Sigma-T95395) in 100 ml distilled water.

- Solution D- 37% Formaldehyde (Sigma- F1268)
- Sodium Carbonate (Saarchem- Analytical grade)
- Stop Solution- Add 5 ml of glacial acetic acid (BDH-analytical grade) to 95 ml of distilled water.

I. Reagents for Coomassie Staining procedure

- Fixative solution- Add 40 g of trichloroacetic acid (ACE-analytical grade) to 200 ml of distilled water.
- Stain- Dissolve 1 g of Brilliant Blue R250 (Sigma-B0149) in 350 ml of 2-propanol (BDH- Analytical grade) and 80 ml of glacial acetic acid (BDH-analytical grade).
- Destain- Add 80 ml of glacial acetic (BDH-analytical grade) to 240 ml ethanol (BDH- analytical grade) and 680 ml of distilled water. Stir gently.

APPENDIX 2

AFLP: Amplified fragment length polymorphism. A sensitive method for detecting DNA polymorphism. Following restriction enzyme digestion of DNA, a subset of the DNA fragments are selected for PCR amplification and visualization.

Allelle: a variant segment of the genetic material. Diploid organisms have two potential alleles for any particular stretch (gene) of DNA. If the alleles are the same (or indistinguishable) on both chromosomes, the individual is a homozygote. If the alleles differ, it is a heterozygote.

Amplicon: amplified DNA product derived from PCR.

Blunt ends: Restriction enzyme cut that produces even 5' or 3' ends. Blunt ends are useful when no specificity of ligation is possible, but higher concentrations of DNA ligase are required.

Codominant: expression of heterozyote phenotypes that differ from either homozygote phenotype. Microsatellites are codominant genetic markers, because one can distinguish a heterozygote (two bands) from each of the homozygotes (single bands).

dNTP: A deoxyribonucleotide (A,G, C, or T)

Electrophoresis: polarized acetate, agarose or acrylamide gel through which one runs proteins or DNA. The material then separates by weight or polarity and allows one to distinguish variants (e.g. alleles or enzyme variants).

Gene diversity: A measure of genetic variation in a population.

Genetic markers: any trait used as a marker of genetic variation within and among individuals and taxa. Traits used include phenotypic traits (eye colour), protein products (allozymes, albumin) and segments of the DNA.

Genome size: The genome is the collective term for all the complement of hereditary material found in an organism (e.g., all the DNA in the set of chromosomes in eukaryotes). Genome size range from approximately 10^4 base pairs (bp) in some viruses to approximately 10^{10} in many angiosperm plants, to $<10^{10}$ in some salamanders and fishes. Mammals have approximately 2-3 X 10^9 bp.

Genotype: The set of DNA variants found at one or more loci in an individual. The information from which genotypes are developed could include allozyme alleles, sequence information or RFLP variants.

Introgression: Movement of genes (or traits) between species or between well-differentiated populations.

Ladder: A series of known-size fragments run in a gel to allow sizing of fragments of target DNA run in other lanes.

Lamda: lamda (λ) phage DNA is a useful tool in molecular biology. Because its entire sequence is known, it is often used to create a ladder of known-size fragments for sizing bands on gels. It is also a useful cloning vector.

Ligation: Enzyme-mediated procedure for joining segments of DNA. Variants include blunt end ligation (both strands of DNA end at the same point, so that any other blunt end can be ligated onto it) or sticky end ligation (one strand overhangs by a few base pairs; this requires a specific enzyme to recognize and initiate ligation/ synthesis).

Oligonucleotide: short chain of nucleotides. Synthesized in the lab as a starting point for developing primers or for use as a probe.

PAGE: Polyacrylamide gel electrophoresis. A technique for separating DNA fragments based on differential mobility in a gel.

PCR: polymerase chain reaction. Technique for amplifying nucleic acids in a thermal cycler. Involves the use of forward and reverse primer pairs that start off the reaction. End yield is many orders of magnitude more DNA of the target sequence than one started with. The resulting amplified DNA can then be visualized with stains or radioactive labeling or sized with fluorescent markers in a sequencer.

Primer: Short, pre-existing single-stranded polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase. The primer anneals to a nucleic acid template and promote copying of the template, starting from the primer site.

Restriction enzyme/ endonuclease: DNA-cutting enzymes found in bacteria and harvested from them for use. Because they cut within the molecule, they are often called restriction endnucleases. A restriction enzyme recognizes and cuts DNA at a particular sequence of nucleotides.

Taq polymerase: A thermostable DNA polymerase from *Thermus aquaticus,* a hot springs bacterium. It is used in PCR because it does not degrade during the high heat cycles generated by a thermal cycler.

Thermal cycler: the 'engine' or PCR machine in which the PCR is performed.