GENETIC IDENTIFICATION OF SOYBEAN [GLYCINE MAX (L.) MERR.] GROWING IN TURKEY FOR MOLECULAR BREEDING USING MOLECULAR MARKERS

H.C. Vural
Selcuk University, Department of Biology, Molecular Biology, Selcuklu, Konya, Turkey
Correspondence to: Hasibe Cingilli Vural
E-mail: hcingilli@selcuk.edu.tr

ABSTRACT
The Soybean (Glycine max) is a species of legume which is a plant in the family Fabaceae. It is an annual plant that has been used in China and in other countries for 5,000 years as a food and a component of drugs. Soy contains significant amounts of all the essential amino acids for humans, and therefore is a good source of protein. Soybeans are the primary ingredient in many processed foods, including dairy product substitutes. The feasibility of detecting yields quality in soybean genotypes by a polymerase chain reaction (PCR) method is determined. PCR is a sensitive method for analyzing DNA and it is considered the most important method for detection of soybeans in processed or raw foods. Namely, PCR is a commonly applied nucleic amplification method which is specific and sensitive enough to detect even tiny amounts of organism-specific DNA sequences. This study focuses on the PCR detection in food, describes rapid and reliable DNA extraction methods which can be applied to a variety of food samples and details of PCR amplification protocols for sensitive and specific detection of soybean genotypes growing in Turkey. All soybean samples were evidenced by all PCR primers as soybean products.


Keywords: genomic DNA extraction, soybean (Glycine max. L.), fingerprinting, molecular marker

Abbreviations: CTAB: hexadecyltrimethylammonium bromide

Introduction
Legumes have begun to draw much attention through recent genomic and phylogenetic studies (11). The crop legumes, such as Lotus, Medicago, Pisum, Glycine, Phaselous, and Vigna, also receive attention from researchers because they are economically important products (8). Legumes play an important role in the traditional diets of many regions throughout the world. In contrast, in Western countries beans tend to play only a minor dietary role despite the fact that they are low in fat and are excellent sources of protein, dietary fibers, and a variety of micronutrients and phytochemicals. Soybeans are unique among the legumes because they are a concentrated source of isoflavones. Isoflavones have weak estrogenic properties. Furthermore, Genistein is one of several known isoflavones. Isoflavones, such as genistein and daidzein, are found in a number of plants with soybeans and soy products like tofu and textured vegetable protein being the primary food source. Soy isoflavones are a group of compounds found in and isolated from the soybean. Besides functioning as antioxidants, many isoflavones have been shown to interact with animal and human estrogen receptors, causing effects in the body similar to those caused by the estrogen hormone (9). Soy isoflavones also produce non-hormonal effects that influence the signal transduction. Soyfoods and isoflavones have received considerable attention for their potential role in preventing and treating cancer and osteoporosis. Vegetable protein sources like soybeans, canola and maize gluten are good alternatives to fish meal.

This study investigates soy DNA fragments (120 and 195 bp) and a 180-bp fragment of the lectin gene of soybean (Glycine max). Soybean lectin (SBL) specifically binds to terminal N-acetyl-D-galactosamine with greatest affinity and to a lesser extent with D-galactose (2). Lectins are carbohydrate-binding proteins or glycoproteins that occur widely in plants, animals and microorganisms. A lectin was first found in the seeds of Ricinus communis (Euphorbiaceae) about one hundred years ago. Since then, numerous plant lectins have been isolated, for the most part, from the seeds of leguminous and gramineous plants. Although lectins have been extensively studied with respect to carbohydrate binding specificity and potential utility for the isolation and characterization of glycoconjugates, the physiological role of lectins in plant is not yet well understood (5). Lectins have a potential use in cancer treatment strategies due to the fact that lectins which are present on the surface of tumor cells are capable of binding exogenous carbohydrate-containing molecules and internalize them by endocytosis (4).

Fresh or frozen tissue is usually used as a source of DNA for PCR and RAPD analysis. We have found that leaves can be allowed to dry at room temperature before extraction of DNA. Heating the leaves or microwave drying resulted in poor recovery of DNA. Storage of fresh leaves in paper envelopes in the laboratory was the most successful approach. This allowed the tissue to dry out over a period of several days and DNA could be extracted at any time, providing a convenient method for the collection and analysis of field material. As DNA is a rather stable molecule it is the preferred analyte for almost
any kind of sample (raw materials, ingredients, processed foods). In addition, PCR, which is the most common DNA detection method, is very sensitive and therefore, very small aliquots of vegetal material are required for the analysis. The development and use of molecular markers for the detection and exploitation of DNA polymorphisms is one of the most significant developments in the field of molecular genetics. The presence of various types of molecular markers, and differences in their principles, methodologies, and applications require careful consideration in choosing one or more of such methods. Analysis of plant and animal genomes using DNA markers are proving valuable in breeding programs to rapidly develop improved crop and livestock strains with enhanced productivity. Genome studies are enabling scientists to gain valuable insights into how the crop and livestock genomes are organized (10). They are also providing a multitude of practical applications such as variety identification through DNA fingerprinting and development of genetic maps, which facilitate indirect selection of economic traits, such as disease resistance. The four principal approaches for identifying polymorphic DNA markers have been so far Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), microsatellites or short sequence repeats (SSR) and Amplified Fragment Length Polymorphism (AFLP). All four approaches have some merits but also a few inherent disadvantages. The RFLP is a laborious technique that relies on Southern blotting and results in the detection of fewer alleles than other methods. The RAPD approach is instead sensitive to reaction conditions and presents problems of reproducibility. The microsatellites strategy on its turn provides high throughput data on polymorphism but requires lengthy studies involving cloning and sequencing to obtain information on flanking nucleotide sequences.

Materials and Methods

Plant material and DNA isolation

Fresh, seed, plantlets as well as herbarium specimens were used in this study for DNA extraction. Herbarium samples were collected by the author in the field and gathered in herbaria. Soybean (Glycine max. L.) seeds or grain samples were provided by Karadeniz Agricultural Research Institute. The analysed samples included dried soybean seeds that were used directly for DNA extraction. They were found to yield DNAs comparable in quality and quantity when using EZ1 Nucleic acid isolation analyser (QIAGEN, 2007), CTAB method, Plant Genomic DNA Purification Kit method, and DNA extraction with phenol purification and liquid nitrogen method (6). A soybean seed bulk sample was ground to fine powder. This experiment was repeated twice under the same conditions for all DNA samples. Furthermore, soybean seeds were germinated then these plantlets were ground to powder by liquid nitrogen treatment (Fig. 1).

Fig. 1. Agarose gel electrophoresis of genomic DNA isolated from soybean genotypes growing in Turkey. Genomic DNAs were loaded in a 0.7% agarose gel and separated by electrophoresis for 90 min at 50 V, then visualised by ethidium bromide staining with transillumination. Respectively: Lane 1, 1 kb ladder size standard; Lane 2, negative control with water blank; Lane 3, 4 and 5, genomic DNA isolated from Soybean with Bio Robot EZ1

Molecular material

DNA was extracted from Soybean [Glycine max (L.) Merr.] growing in Turkey and isolated genomic DNA was amplified with four primer pairs, obtained from Operon Technologies Inc. (Almeda CA, USA) (Table 1).

Primer for PCR amplification of Soybean [Glycine max (L.) Merr.] used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Anneling Tm</th>
<th>Primer bp</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA</td>
<td>54°C</td>
<td>320</td>
<td>5'-CCACTATCCTCTCAGCAAGACCCCTTCC-3'</td>
</tr>
<tr>
<td>SPB</td>
<td>54°C</td>
<td>320</td>
<td>5'-CTTCTGTGCTGTAGCCACTGTGTC-3'</td>
</tr>
<tr>
<td>LE5</td>
<td>53°C</td>
<td>180</td>
<td>5'-TCAACGAAAACGAGTCTGGTG-3'</td>
</tr>
<tr>
<td>LE6</td>
<td>53°C</td>
<td>180</td>
<td>5'-GGTGGAGGCATCATAGGTAT-3'</td>
</tr>
<tr>
<td>CP 03-5&quot;</td>
<td>59°C</td>
<td>123</td>
<td>5'-CGGACGAGAATAAAGATAAAT-3'</td>
</tr>
<tr>
<td>CP 03-3&quot;</td>
<td>59°C</td>
<td>123</td>
<td>5'-TCTCAAGAAGCGAGAAGAAAGG-3'</td>
</tr>
</tbody>
</table>

TABLE 1

BIOTECHNOL. & BIOTECHNOL. EQ. 24/2010/3 2005
DNA digestion
Each restriction enzyme recognises a unique, specific sequence of, usually, 4-6 base pairs (bp) in length, termed a restriction site, where the enzyme cuts (or restricts) the DNA. In general, restriction sites will occur throughout the genome and, consequently, application of the enzyme to total genomic DNA (restriction of the DNA) results in its conversion into millions of fragments. The frequency of restriction sites will vary depending on both the restriction enzyme and on the genome.

Polymerase chain reaction amplifications
In this study, four primer pairs were used for PCR amplification. We performed different PCR optimisations of the PCR reaction mix and conditions for each primer pair.

SPA/SPB primer pair
PCR was carried out in two tubes of a total volume of 25 μl. The tubes contained 2 μl of genomic DNA, 0.2 μl of each primer, 2.5 μl 10x PCR Buffer, 3.5 μl of 50 mmol/l MgCl$_2$, 2.5 μl of 10 mmol/l dNTPs, and 0.3 μl of Taq DNA polymerase. The program was initiated on Bio-Rad PCR amplifier with 5 min of denaturation, followed by 10 cycles of amplification with denaturation for 15 sec at 94°C, first annealing for 20 sec at 55°C and an extension at 72°C for 15 sec; followed by 40 cycles of amplification with denaturation for 1.5 min, second annealing for 1.5 min at 61°C, an extension at 72°C for 1 min, and a final elongation at 72°C for 5 min.

LE5/LE6 primer pair
PCR was carried out in two tubes of a total volume of 25 μl. The tubes contained 2 μl of genomic DNA, 0.2 μl of each primer, 2.5 μl 10x PCR Buffer, 3.5 μl of 50 mmol/l MgCl$_2$, 2.5 μl of 10 mmol/l dNTPs and 0.3 μl of Taq DNA polymerase. The program was initiated on Bio-Rad PCR amplifier with 5 min of denaturation, followed by 10 cycles of amplification with denaturation for 15 sec at 94°C, first annealing for 20 sec at 54°C and an extension at 72°C for 15 sec; followed by 40 cycles of amplification with denaturation for 1.5 min, second annealing for 1.5 min at 61°C, an extension at 72°C for 1 min, and a final elongation at 72°C for 5 min.

CP03-F/CP03-R primer pair
PCR was carried out in two tubes of a total volume of 25 μl. The tubes contained 2 μl of genomic DNA, 0.2 μl of each primer, 2.5 μl 10x PCR Buffer, 3.5 μl of 50 mmol/l MgCl$_2$, 2.5 μl of 10 mmol/l dNTPs and 0.3 μl of Taq DNA polymerase. The program was initiated on Bio-Rad PCR amplifier with 5 min of denaturation, followed by 10 cycles of amplification with denaturation for 15 sec at 94°C, first annealing for 20 sec at 59°C and an extension at 72°C for 15 sec; followed by 40 cycles of amplification with denaturation for 1.5 min, second annealing for 1.5 min at 61°C, an extension at 72°C for 1 min, and a final elongation at 72°C for 5 min.

Gym 81/Gym 82 primer pair
PCR was carried out in two tubes of a total volume of 25 μl. The tubes contained 2 μl of genomic DNA, 0.2 μl of each primer, 2.5 μl 10x PCR Buffer, 2.5 μl of 50 mmol/l MgCl$_2$, 2.5 μl of 10 mmol/l dNTPs, and 0.3 μl of Taq DNA polymerase. The program was initiated on Bio-Rad PCR amplifier with 5 min of denaturation, followed by 10 cycles of amplification with denaturation for 15 sec at 94°C, first annealing for 20 sec at 55°C and an extension at 72°C for 15 sec; followed by 40 cycles of amplification with denaturation for 1.5 min, second annealing for 1.5 min at 61°C, an extension at 72°C for 1 min, and a final elongation at 72°C for 5 min.

All amplification products (10 μl) were electrophoresed in 1.5% (w/v) TAE agarose gels containing 1 g/ml ethidium bromide. The electrophoresis was run in TBE at 100V for 40min and DNA bands were observed under ultraviolet light and photographed by Image System (Fig. 2, Fig. 3).

PCR-RFLP analysis
RFLP analysis was performed as described by Shiraishi et al. (7). High molecular weight DNA was digested with an appropriate restriction endonuclease and the digest was fractionated by electrophoresis in a 0.7% agarose gel. The diversity in soybeans has previously been assayed using restriction fragment length polymorphism (RFLP) and sequencing analyses of genomic DNA (nDNA). Here we describe a method based on PCR-RFLP for identification as new primers were designed to artificially create restriction sites that spanned them. The PCR-RFLP method enabled us to identify nucleotides at each of the mutation sites easily and reliably. However, mutations were not observed in this study. Digestive reactions were incubated at optimal assay temperature (37°C) for 2 h. The results from RFLP analysis after DNA electrophoresis were read and

BIOTECHNOL. & BIOTECHNOL. EQ. 24/2010/3
photographed as described above, except that the concentration of the agarose gel was 2%.

![Agarose gel electrophoresis of PCR products amplified from genomic DNA of soybean genotypes. Lane 1 or M, 50 bp ladder size standard; Lane 2, 100 bp ladder size Standard; Lane 3, PCR products are amplified with Gym81/Gym82 primer pairs](image.png)

**Fig. 3.** Agarose gel electrophoresis of PCR products amplified from genomic DNA of soybean genotypes. Lane 1 or M, 50 bp ladder size standard; Lane 2, 100 bp ladder size Standard; Lane 3, PCR products are amplified with Gym81/Gym82 primer pairs

### Results and Discussion

This study was performed to investigate the quality of the extracted DNA samples from Soybean *Glycine max* (L.) Merr. genotypes. Integrity of DNAs from all extractions, as well as, isolation conditions were checked by agarose gel electrophoresis and a typical result is illustrated on **Fig. 1**. Genomic DNA solutions from soybean obtained by both extraction methods (EZ1 and CTAB) were of adequate purity ($A_{260}/A_{280}=1.7-2.0$) and yield for applying PCR, resulting in amplification of the expected 100 bp fragment using soybean specific primer pairs.

DNA quality is determined by its fragment length and its degree of damage due to exposure to heat, low pH and/or nucleases that cause hydrolysis, depurination and/or enzymatic degradation. Therefore, DNA quality varies according to the material under examination, the degree of processing of the sample and the applied DNA extraction method. It is important to keep in mind that DNA isolated from processed foods and certain agricultural materials such as cured tobacco leaf is of low quality, with available target sequences being rather short, e.g. 100-400 bp for soybean protein preparations and processed tomato products (3).

A sensitive qualitative detection method for soybeans in foods by using the polymerase chain reaction was developed. A sample of lecithin was processed and submitted to PCR with primer pairs LE5/LE6. In all reactions the expected product was produced by the control primer pair LE5/LE6. Once BIOTECHNOL. & BIOTECHNOL. EQ. 24/2010/3 again, this indicated that the quantity and quality of lecithin extracted DNA was suitable for PCR. In this study, four pairs of primers, LE5/LE6, CP03-F/CP03-R, SPA/SPB, and Gym81/Gym82 were used for PCR detection of Soybean genotype sequences growing in Turkey. These primers were designed for the present study and can be used as a specific test for the presence of Soybean. When used in amplification reactions the SPA/SPB primer pair produces a 320 bp product which is diagnostic for the species specific to Turkey. Additionally, the other primers LE5/LE6 were designed to amplify a 180 bp portion of the lectin gene of *Glycine max*, and CP03-F/CP03-R for universal detection of the DNA derived from plants to verify the presence of extracted DNAs. This primer pair was designed to amplify a 123 bp portion of the chloroplast DNA and to detect the noncoding region of the chloroplast. For specific detection of soybeans with high specificity, the primer pair- Gym81/Gym82 was designed on the gene encoding the *Glycine max* repetitive sequence. The primer pair amplified a 118 bp fragment of the repetitive sequence.

In summary, a rapid PCR detection method was developed for the specific detection of soybeans growing in Turkey. The protocols reported in this study are simple, sensitive, and reliable for identifying trace amounts of soybean in processed foods. We believe that the PCR method will be used to complement the protein-based detection in the future. The presence of soybean genotypes was investigated by amplifying the soy-specific Lectin gene. Lecithin primer pairs were designed to detect soybean genotypes and lectin sequences. This primers generate smaller amplicons and therefore have a greater opportunity of mediating amplification when degraded DNA templates such as those obtained from processed material are used. LE5/LE6 amplifies a 180 bp fragment of the lectin gene. We repeated all reactions at least two times and obtained the same DNA fingerprints, indicating high reproducibility obtained with these reactions. For specific detection of soybeans with high specificity, the primer pair of Gym81/Gym82 was designed on the gene encoding the *Glycine max* repetitive sequence.

### Confirmation test of PCR products

As described above, PCR products can be confirmed using the following methods: nucleic acid sequencing; nested and touchdown PCR followed by cleavage by a proper endonuclease in order to confirm the final products of PCR-RFLP method. In this study, soybean samples were amplified with different PCR-primers: LE5/LE6, CP03-F/CP03-R, SPA/SPB, and Gym81/Gym82 and after the restriction specific products were produced.

### Conclusions

The present study describes the development of protocol for isolation of high purity DNA, optimization of the conditions for the primer pairs and is the first report in the species soybean. This will form a strong base for future molecular characterization and genetic improvement works in this promising food and industrial plant. The existence of soybean
genotypes in the sample may reflect the narrow genetic base of the gene pool of Turkey soybean germplasm. Furthermore, the use of these primers or marker loci may be a feasible alternative in identifying and evaluating the soybean to be protected. In addition, these primers may be promising for marker-assisted selection in plant improvement and for filling in gaps of pre-existing marker-based map.

Acknowledgments
This study was partially supported by Selcuk University Archeometry Laboratory and Scientific Research Foundation of Selcuk University (BAP) for providing foundation.

REFERENCES