Sequence-tagged sites of cDNA clones in *Solanum tuberosum* and their evaluation as molecular markers in Solanaceae species

Junita Liebenberg

Thesis presented in partial fulfilment of the requirements for the degree of Master of Natural Science at the University of Stellenbosch.

December 1999

Study leader: Prof. L. Warnich
Co-study leader: Dr. A.V. Peeters

DECLARATION

I the undersigned hereby declare that the work contained in this thesis is my own work and has not previously in its entirety or in part submitted at any university for a degree.

Date: 24/11/1999

ABSTRACT

Potato is an important food crop that is constantly threatened by pests and diseases. To cope with the food shortages in the water-poor subcontinent, Africa will have to increase its yield through the use of disease- and pest-resistant varieties, and increase the area planted by developing drought-tolerant varieties. In order to introduce resistant or stress-tolerant genes, either by classical breeding or genetic engineering, genome analyses with molecular markers are needed to identify and track such genes. Partially sequenced cDNA clones or expressed sequenced tags (ESTs), play an important role in genome analysis. ESTs can be used to tag genes and are a common source for the development of sequence-tagged site (STS) markers. In this study cDNA-based STS marker technology is combined with single-stranded conformation polymorphism (SSCP) analysis to detect polymorphisms.

An oligo(dT)-primed cDNA library was constructed from young potato leaves. About 3000 clones were produced, using a plasmid vector. The quality of the library was determined by means of PCR and sequencing. Forty-eight ESTs were produced and based upon similarities to characterised sequences in electronic databases, 37 cDNA sequences could be assigned putative identities. STS oligonucleotide primers were developed from the ESTs, and were subsequently used to screen a panel of 10 potato cultivars. Twelve of the 13 primer pairs could detect polymorphisms in SSCP analyses and between two and six unique profiles were amplified per assay. Five of the polymorphic STSs were characterised by sequencing. It was found that single nucleotide polymorphisms gave rise to the variation for three markers and the variations for the remaining two markers were due to insertions and deletions.

Since the STS primers were derived from partially sequenced cDNA clones, they have the advantage of providing not only a marker, but also a tag for a gene. In addition it was shown that most of these markers could also be used in other Solanaceae species. Nine markers could detect variations between six different species by standard agarose gel electrophoresis.

In summary, useful PCR-based markers, which are quick and easy to analyse, have been developed for potato and can find application in the genome analysis of potato as well as other Solanaceae species.

OPSOMMING

Die aartappel is 'n belangrike voedselgewas wat gedurig deur peste en siektes bedreig word. Om die voedseltekorte in die droogtegeteisterde subkontinent te oorkom, moet Afrika deur die gebruik van weerstandbiedende variëteite die oes verhoog, asook die area onder verbouing vergroot deur droogtetolerante variëteite te ontwikkel. Om die weerstands- en droogtebestande gene oor te dra, deur klasieke planteteelt of genetiese ingenieurswese, is genoomanalises met merkers nodig om gene te identifiseer en oordrag daarvan te volg. Die volgordes van cDNA klone of ESTs ("expressed sequenced tags"), speel 'n belangrike rol in genoomanalise. ESTs kan gebruik word om gene te merk en hulle dien as bron vir die ontwikkeling van STS ("sequence–tagged site") merkers. In hierdie studie is cDNA-gebaseerde STS merkertegnologie met SSCP ("single–strand conformation polymorphism") analise gekombineer om polimorfismes waar te neem

'n cDNA-biblioteek van jong aartappelblare is vervaardig met oligo(dT)-inleiers. Ongeveer 3000 klone is geproduseer deur van 'n plasmied vektor gebruik te maak. Die kwaliteit van die biblioteek is bepaal met behulp van PKR en volgordebepaling. Ag-en-veertig ESTs is geproduseer en volgens ooreenkomste met gekarakteriseerde volgordes in elektroniese databasisse, is 37 cDNA volgordes voorlopige geïdentifiseer. STS inleiers is ontwerp vanaf die ESTs en is gebruik om 'n paneel van 10 aartappelkultivars te amplifiseer. Twaalf van die 13 inleierpare kon polimorfismes waarneem in SSCP analises en tussen twee en ses unieke profiele is per merker geamplifiseer. Vyf van die polimorfiese STS'e is deur volgordebepaling gekarakteriseer. Daar is gevind dat enkel nukleotied polimorfismes verantwoordelik was vir die variasie van drie van die merkers en die verskille van die oorblywende twee merkers was as gevolg van insersies en delesies.

Aangesien die merkers van cDNA-kloonvolgordes afgelei is, het dit die voordeel dat dit nie slegs 'n merker nie, maar ook 'n geen vir 'n kaart verskaf. Daar is gevind dat die merkers ook in ander Solanaceae spesies gebruik kan word. Met nege merkers is variasie tussen ses verskillende spesies met agarose jelelektroforese waargeneem.

Opsommend kan gesê word dat nuttige PKR–gebaseerde merkers, wat maklik en goedkoop ge-analiseer kan word, vir die aartappel ontwikkel is en toegepassing in die genoomanalise van aartappel en ander Solanaceae kan vind.

ACKNOWLEDGEMENTS

My thanks to the following people and institutions:

The National Research Foundation (former Foundation for Research Development) for financial support.

The University of Stellenbosch for supplying the facilities.

Prof. L. Warnich and Dr. A.V. Peeters for their support and guidance.

Ms. C.E. McGregor and Mr. A. Visser, Vegetable and Ornamental Plant Institute (ARC), Roodeplaat, for supplying potato seedlings and DNA.

Mr. C.J. van Heerden, Central DNA Sequencing Facility, U.S., for DNA sequence analysis.

My friends and colleagues for their encouragement and patience.

My family for their love and support.

Frans for his love, endless patience, understanding and encouragement.

TABLE OF CONTENTS

List of tables i

List of figures ii

List of abbreviations iv

CHAPTER 1

INTRODUCTION **1**

1.1 THE POTATO **1**

- 1.1.1 Potato as a food source 1
- 1.1.2 The potato genome and the Solanaceae family 3
- 1.1.3 The potato and biotechnology 3

1.2 Characterisation of a genome **5**

- 1.2.1 DNA markers for genetic maps 6
- *1.2.1.i Single nucleotide polymorphisms* *7*
- *1.2.1.ii Single-stranded conformation polymorphism (SSCP) analysis* *7*
- *1.2.1.iii Development of new markers* *8*
- 1.2.2 Physical mapping techniques 8
- *1.2.2.i STS markers* *9*
- *1.2.2.ii Expressed sequence tags* *10*
- *1.2.2.iii ESTs as a source for the development of STSs* *11*

1.3 cDNA libraries **12**

- 1.3.1 PCR-based cDNA libraries 12
- 1.3.2 Modified cDNA libraries 13
- *1.3.2.i Subtracted cDNA libraries* *13*
- *1.3.2.ii Normalised cDNA libraries* *15*

1.4 Aims of this study **16**

CHAPTER 2

MATERIALS AND METHODS **17**

2.1 Plant material **17**

2.2 Measures to prevent RNA degradation by RNases **18**

2.3 RNA extraction **18**

2.4 Library construction **19**

- 2.4.1 cDNA synthesis 19

2.4.2 Ligation of amplification adaptors 21
2.4.3 Amplification of cDNA 21
2.4.4 Cloning of cDNA 22
2.4.4.i Ligation 22
2.4.4.ii Transformation of competent cells 22
2.4.5 Library organisation 23
2.5 Characterisation of inserts 23
2.5.1 Size determination of inserts 23
2.5.2 Sequencing 24
2.5.3 Data analysis 24
2.6 Primer design 24
2.7 Screening of genomic DNA 25
2.7.1 Genomic DNA isolation 25
2.7.2 PCR conditions 25
2.7.3 Agarose gel analysis 26
2.7.4 Single-strand conformation polymorphism (SSCP) analysis 26
2.8 Summary of procedures followed 28

CHAPTER 3

RESULTS AND DISCUSSION 29
3.1 Isolation of RNA 29
3.1.1 Isolation of total RNA 30
3.1.1.i *Determination of quantity of total RNA 32*
3.1.1.ii Determination of quality of total RNA 32
3.1.2 Isolation of mRNA 33
3.2 Library construction 34
3.2.1 cDNA synthesis 36
3.2.2 Cloning of cDNA inserts 37
3.3 Characterization of cDNA inserts 38
3.3.1 Presence and size of inserts 38
3.3.2 DNA sequencing 39
3.3.2.i Insert orientation and presence of poly(A) tails 41
3.3.2.ii Redundancy 41
3.3.2.iii Long poly(A) tails cause problems in automated sequencing 42
3.3.2.iv Sequencing errors 43
3.3.2.v Chimerism 46

3.3.3 EST analysis 46
3.4 STS analysis **54**
3.4.1 Agarose gel analysis of potato cultivars 55
3.4.2 SSCP analysis of potato cultivars 57
3.4.3 Characterisation of SSCPs 64
3.4.4 STS agarose gel analysis of the Solanaceae family 66

CHAPTER 4
CONCLUSIONS **70**

LIST OF REFERENCES **72**

LIST OF TABLES

Chapter 2

Table 2.1 The potato cultivars and members of the Solanaceae family used in this study. 17
Table 2.2 PCR conditions for different ESTs. 26

Chapter 3

Table 3.1 Potato ESTs with similarity to known genes in protein databases. 49
Table 3.2 Potato ESTs with similarity to known genes in nucleotide databases. 50
Table 3.3 STSs developed in this study: the primer sequences, the length of the cDNA product and the estimated length of the genomic amplification products in bp. 55
Table 3.4 A summary of the variation detected by sequencing a selection of SSCP fragments, generated by five STS markers. 65
Table 3.5 A summary of the unique profiles appointed to the ten cultivars for the five selected markers. 65
Table 3.6 Amplification trials for 13 potato-derived STS primer pairs in six solanaceous species. 69

LIST OF FIGURES

Chapter 1

Figure 1.1 Global potato production from 1961 to 1997. 2

Figure 1.2 Acceleration in production growth. 2

Chapter 2

Figure 2.1 cDNA synthesis and amplification based on the method described by Jepson *et al.* (1991). 20

Chapter 3

Figure 3.1 (a) Total RNA isolated with the Qiagen and the Pharmacia kit. 31

Figure 3.1 (b) Total RNA isolated with the TRIZOL method. 31

Figure 3.2 Amplified double-stranded cDNA run on a 1% agarose gel and stained with ethidium bromide. 37

Figure 3.3 Inserts amplified with M13 forward and reverse primers by means of PCR. 38

Figure 3.4 Clone 6e3 sequenced with the M13 reverse primer. 40

Figure 3.5 (a) A poly(A) tail causing errors in the adjacent sequence. 43

Figure 3.5 (b) A poly(T) tail causing errors in the adjacent sequence. 43

Figure 3.6 The alignment of the sequences of clones 10f12 and 10g2. 45

Figure 3.7 Alignment of clones 10e3 and 10h2. 52

Figure 3.8 Chromatograms of 10e3 (forward and reverse sequences) and 10h2. 53

Figure 3.9 Genomic amplification products of primer pair 6b10 on an ethidium bromide stained 1.5% agarose gel. 57

Figure 3.10 (a) SSCPs of 2g4. 59

Figure 3.10 (b) SSCPs of 3c10. 59

Figure 3.10 (c) SSCPs of 3d6. 59

Figure 3.10 (d) SSCPs of 3d11. 60

Figure 3.10 (e) SSCPs of 3h5. 60

Figure 3.10 (f) SSCPs of 3h7. 60

Figure 3.10 (g) SSCPs of 6b10. 61

Figure 3.10 (h) SSCPs of 6b11. 61

Figure 3.10 (i) SSCPs of 6d6. 61

Figure 3.10 (j) SSCPs of 6h9. 62

Figure 3.10 (k) SSCPs of 7a5. 62

Figure 3.10 (l) SSCPs of 7b4. 62

Figure 3.10 (m) SSCPs of 7e9. 63

Figure 3.11 (a) Amplification products of different species of the Solanaceae family with primer pairs 2g4, 3d6 and 6h9 on a 1.5% agarose gel. 67

Figure 3.11 (b) Amplification products of different species of the Solanaceae family with primer pairs 3d11, 3h7 and 3c10 on a 1.5% agarose gel. 67

Figure 3.11 (c) Amplification products of different species of the Solanaceae family with primer pairs 6b10, 6b11, 6d6 and 7e9 on a 1.5% agarose gel. 68

Figure 3.11 (d) Amplification products of different species of the Solanaceae family with primer pairs 7a5, 7b4 and 3h5 on a 1.5% agarose gel. 68

LIST OF ABBREVIATIONS

λ	lambda
μg	microgram
μl	microlitre
μM	micromolar
°C	degrees centigrade
A	adenine
AFLP	amplified fragment length polymorphisms
AMP	Ampicillin
AP-PCR	arbitrary primed PCR
ARC	Agricultural Research Council
ARMS	amplification refractory mutation system
ASO	Allele-specific oligonucleotide
BLAST	Basic Local Alignment Search Tool
bp	base pair
C	cytosine
cDNA	complementary DNA
cM	centiMorgan
CTAB	cetyltrimethylammonium bromide
DAF	DNA amplified fingerprints
dCTP	deoxycytosine triphosphate
ddH_2O	double-distilled deionised water
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ds	double-stranded
dT	deoxythymine
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
EST	expressed sequence tag
EtBr	ethidium bromide
FAO	Food and Agriculture Organization of the United Nations
FISH	fluorescence in situ hybridization
g	gram
G	Guanine

gDNA	genomic deoxyribonucleic acid
GITC	guanidinium isothiocyanate
h	hour(s)
href	hypertext reference
HSP	high-scoring segment pair
IPTG	isopropylthiogalactoside
kb	kilobase
KCl	potassium chloride
kg	kilogram
LB	"Luria Bertani"
LiCl	lithium chloride
min	minute(s)
M	molar
mg	milligram
$MgCl_2$	magnesium chloride
ml	millilitre
mM	millimolar
mRNA	messenger RNA
NaCl	sodium chloride
NAD	nicotinamide adenine dinucleotide
NaOH	sodium hydroxide
NCBI	National Centre for Biotechnology Information
ng	nanogram
$NH_4(SO_4)$	ammonium sulphate
nm	nanometer
ORF	open reading frame
%	percentage
PCR	polymerase chain reaction
pmol	picomole
pSK	pBluescript SK(+)
QTL	quantitative trait loci
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
rpm	revolutions per minute

RT	reverse transcriptase
SDS	sodium dodecyl sulphate
sec	second(s)
SNP	single nucleotide polymorphism
SSCP	single strand conformation polymorphism
SSR	simple sequence repeat
STE buffer	sodium chloride tris EDTA buffer
STMS	sequence-tagged microsatellite site
STS	sequence-tagged site
t	ton
T	thymine
Taq	*Thermus aquaticus*
TBE	tris borate EDTA
TE	tris EDTA
tRNA	transfer RNA
Tris	tris[hydroxymethyl]aminomethane
Tris-HCl	tris hydrochloric acid
Tm	melting temperature
U.S.	University of Stellenbosch
UTR	untranslated regions
UV	ultra-violet
V	volt
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CHAPTER 1

INTRODUCTION

1.1 THE POTATO

1.1.1 Potato as food source

With its high nutritional value and great yield potential, the potato (*Solanum tuberosum* L.) is the world's fourth most important food crop, after wheat, rice and maize. It is rich in minerals, vitamins, kilojoules and protein, but is virtually fat-free. Potato fields yield more tons per hectare than any other form of cultivation, averaging 16.1 tons per hectare worldwide. Potatoes are grown in more countries than any other crop, except maize. Futhermore, potato provides roughly half of the world's annual output of all roots and tubers, making it the largest non-cereal crop.

According to the latest FAO (Food and Agriculture Organization of the United Nations) data (FAOSTAT, September 1998), potato production worldwide stands at 295 million tons per annum and covers more than 18 million hectares. About 35% (about 100 million t) of the world production is in developing countries where potato is a staple in the diet of half a billion consumers. Although the annual production in Africa is only 7.7 million tons in contrast to the 154 million tons produced in Europe, potatoes are becoming increasingly important as a source of food, employment and income in Africa. Egypt (2.7 million t), South Africa (1.5 million t), Algeria (1.1 million t) and Morocco (1.1 million t) produce more than 80% of all the potatoes on this continent. (href1)

The potato plays an increasingly important role in the daily nutrition of poor subsistence farmers, and is often a main source of cash income. Potato production in developing countries entered a new, rapidly expanding phase in the 1990s (Figure 1.1). For the last 10 years, potato production has increased at an annual average rate of 4.5%, and the area planted by 2.4%. As potato output continues to expand, the growth rate for area planted and production continues to accelerate. As a result, the growth rate in potato production has nearly doubled over the last 20 years. While growth in the production of maize, wheat, and rice slowed down in the last decade, potato output surged ahead, thereby increasing potato's relative importance (Figure 1.2). (href1)

Figure 1.1 Global potato production from 1961 to 1997. The production in developing compared to that in developed countries. (href1)

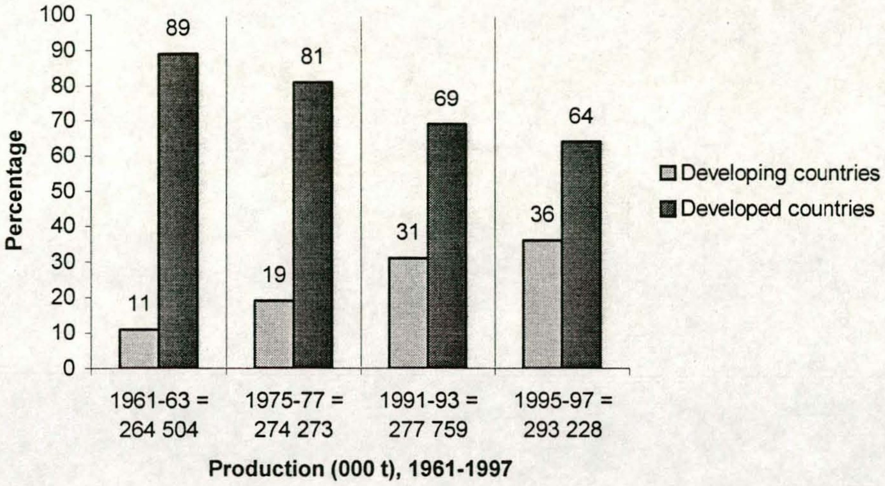
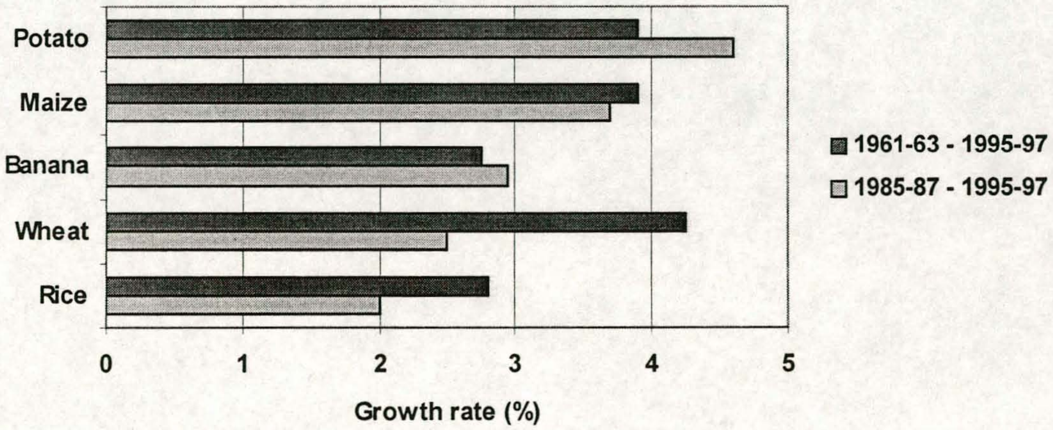


Figure 1.2 Acceleration in production growth. Average annual growth rates in production of selected crops in developing countries: 1961-63 to 1995-97 and 1985-87 to 1995-97. (href1)



1.1.2 The potato genome and the Solanaceae family

Commercial potato cultivars are highly heterozygous tetraploids ($2n = 4x = 48$). The segregation of traits in potato is complex and often masked by the heterozygosity and tetrasomic inheritance. The haploid nuclear genome size of the potato is 1.579×10^9 bp (Arumuganathan 1991).

The genus *Solanum* includes about 900 species, of which 235 form tubers (Hawkes 1990). Wild *Solanum* species related to the cultivated potato represent an important reservoir of genetic diversity for potato breeders. During 8000 years of cultivation farmers have developed thousands of varieties.

The Solanaceae family includes other economically important crops, such as tomato, tobacco, pepper, and aubergine. Great homology exists between the different species of the Solanaceae family (Provan *et al.* 1996). The genetic contents of the potato and tomato chromosomes are quite similar, differing only by a few inversions (Bonierbale *et al.* 1988). This was shown by mapping of restriction fragment length polymorphism (RFLP) loci in an interspecific cross involving *S. phureja* and a hybrid between *S. tuberosum* ssp. *tuberosum* and *S. chacoense* with tomato probes. Bonierbale *et al.* reported that virtually all RFLP probes mapped in the tomato genome, cross-hybridized with potato DNA and can therefore be used for potato mapping. Tanksley *et al.* (1992) constructed molecular linkage maps for the potato and tomato genomes based on a common set of probes. These maps made it possible to determine the breakpoints corresponding to five chromosomal inversions that differentiate the two genomes. The total of 1400 markers on these maps placed tomato and potato amongst the most thoroughly mapped plant species (Tanksley *et al.* 1992).

1.1.3 The potato and biotechnology

The European cultivated potato was derived from a narrow genetic base and was reproduced by clonal reproduction. As a consequence it lacked genes for adequate levels of resistance to pathogens and pests such as late blight (Nelson 1995) and potato cyst nematodes (Marks & Brodie 1998). During the twentieth century, attempts have been made to remedy these deficiencies by introgression of resistance genes into ssp. *tuberosum* from wild and cultivated Solanum species of Central and South America (Hermsen 1994). However, this failed to give durable resistance and production is thus maintained by the application of substantial quantities of chemicals. Of the world food crops, potato is the largest single recipient of agricultural pesticides (href1). Dependence on these chemicals has increased production costs and exposed the environment, farm workers, and consumers to associated health risks. Pesticides are often handled inappropriately and in general their use is poorly regulated in developing countries. At the same time lack of access to pesticides reduces potato production. Biotechnology can be employed to lessen the dependence on pesticides by incorporating resistance directly into the crops. This can reduce the risks to the environment, avert unnecessary exposure of farm workers and consumers to toxic pesticides, and contribute to increased food security and profitability.

By introgression of disease resistance genes from the potato's many primitive relatives and wild species, classical breeders have improved the resistance spectrum of modern cultivars (Bradshaw & Mackay 1994). However, the introgression of such resistance takes much time and effort. The hybridisation of the cultivated potato with even its closest relatives usually requires several generations of back crossing and selection to restore the yield and quality of a modern cultivar. It also risks the introgression of undesirable traits, such as high levels of toxic glycoalkaloids, in the tubers (Van Gelder *et al.* 1998). Though the potato is capable of self-fertilisation, it suffers from marked inbreeding depression. Moreover, it is extremely difficult to unravel the genetics of complex traits in a heterozygous tetraploid, such as the potato. With the potential for four different alleles at each locus, in crosses made between parents with complementary traits, the probability of recovering recombinants in their progeny combining the best features of both parents is low (Bradshaw 1994). Despite successful research into increasing the efficiency of selection in breeding programs (Bradshaw & Mackay 1994; Bradshaw *et al.* 1995), it remains unlikely that classical breeding will be able to produce cultivars that are resistant to all possible diseases and pests.

Biotechnology can provide the solution to the dilemma. The transfer of specific genes that confer resistance, as opposed to the use of entire genomes, would be much more efficient. For the past 15 years, plant reseachers have improved crop yields and pest resistance through a combination of breeding and transformation. By using bacteria to transfer foreign DNA into plant cells, the potato was the first commercial crop to be protected against insect pests through biotechnology (href2). The NewLeaf® Potato, introduced in 1995 by Monsanto, is resistant to the Colorado potato beetle, as a result of the introduction of a gene from the *Bacillus thuringiensis* bacteria (href3). Monsanto has also developed the following three products which are expected to be on the market within six years. By introducing a starch-producing gene from a soil bacterium into a potato plant, a higher solids potato was produced. With the reduction in the percentage of water in the genetically improved potato, less oil is absorbed during processing, resulting in a reduction in cooking time and costs. The NewLeaf® Plus Potato, which protects itself against the Colorado potato beetle and potato leaf roll virus, and the NewLeaf® Y Insect- and Virus-Protected Potato, which is resistant to the Colorado potato beetle and potato virus Y, will also be commercialised.

1.2 CHARACTERISATION OF A GENOME

Plants provide a unique biological frontier to study many processes not found in other systems and are one of the major targets of useful genetic engineering. For molecular interpretation of plant cellular processes and plant genetic engineering, development of gene resources, in the form of genetic and physical maps, is needed. A linkage map provides a unifying framework for the genetic description of a species. Detection of close linkage between molecular markers and loci controlling traits of economic interest provides a basis for marker-assisted selection and for map-based positional cloning. Comparative mapping can be used to detect linkage groups and, therefore, chromosome rearrangement between taxa (Tanksley *et al.* 1992; Moore *et al.* 1993). Due to the similarities between plant species in genes and their arangement, now referred to as conserved DNA sequences, a gene identified in one species can sometimes be used to identify the same gene in related species. To facilitate the movement from the genetic locus to the cloned gene, physical maps are constructed (Hauge & Goodmann 1992).

Despite large investment in plant molecular research in recent years, plant gene resources are still limited and more markers are needed to fill the gaps on current linkage maps (Harushima *et al.* 1998). Once compiled, the genetic map is used to develop research strategies. When the source of a particular desired trait has been identified, the gene or genes can be cloned and directly transferred, or incorporated indirectly into other varieties using associated markers to facilitate the process. Within this context, conserved DNA sequences are of major interest.

Currently, several molecular marker technologies are available to detect variation and conservation in nuclear genomes. Variation can be detected at the level of the protein product of a gene, as in the case of isozymes, or directly at the DNA level. Although isozymes have been widely used, their limited number prevents them from providing complete genome coverage (Smith & Smith 1990). In recent years, PCR-based molecular markers have gained popularity in population genetics and genome mapping, because they are potentially unlimited in number and their detection requires mere nanogram quantities of DNA.

1.2.1 DNA markers for genetic maps

Genetic maps for potato now exist for several species, including wild potatoes, primitive cultivated potatoes from the Andes, and modern cultivated potatoes. Currently several maps are available on the World Wide Web (href4). These include a RFLP map consisting of markers from 211 cloned potato sequences, 26 tomato genomic sequences and several cDNA clones for potato and maize genes. High density molecular linkage maps, comprised of more than 1000 markers that have been constructed for the potato and tomato genome (Tanksley *et al.* 1992), are accessible at the same Internet location. Additionally, an amplified fragment length polymorphism (AFLP) marker catalogue is available for gene mapping within the cultivated potato (href5). The catalogue consists of of 733 AFLP markers which are mapped relative to 217 RFLP, isozyme and morphological trait loci (Rouppe van der Voort *et al.* 1998). A map consisting of 65 microsatellite markers was constructed by Milbourne *et al.* (1998). Still more markers are needed to locate genes and quantitative trait loci (Bradshaw *et al.* 1998).

Numerous genetic markers are available and are being developed. At present, RFLPs are predominantly used as anchor points for map construction. However, RFLPs are tedious to handle and the number of genotypes that can be screened in a short time-span is limited. Alternatively, conversion of RFLP probes into PCR-based markers provides a rapid, safe and efficient method for screening large populations. Microsatellites or simple sequence repeats (SSRs) (Tautz 1989), random amplified polymorphic DNAs (RAPDs) (Welsh and McClelland 1990), DNA amplification fingerprinting (DAF) (Caetona-Anollès *et al.* 1991), arbitrarily-primed PCR (AP-PCR) (Williams *et al.* 1990), and the recently–developed AFLP marker technique (Vos *et al.* 1995) contribute an almost unlimited number of markers.

Recently there has been great interest in single nucleotide polymorphisms (SNPs) since these are the most common variation in a genome (href6). In an analysis of random genome fragments from the horse and human genomes, Reynolds *et al.* (1998) found that SNPs are present in every 250 - 400 nucleotides. cSNPs, SNPs occuring in coding sequences (Collins *et al.* 1997), were specifically targeted in this study.

1.2.1.i Single nucleotide polymorphisms

Methods to genotype SNPs are increasingly important in gene mapping. A common technique is single-stranded conformational polymorphism analysis (SSCP) (Orita *et al.* 1989), which depends on electrophoretic resolution of DNA sequence conformation differences. The method was favoured in this study and will be discussed in section 1.2.1.ii.

Allele-specific oligonucleotide hybridisation (ASO) (Higuchi *et al.* 1988) is used to differentiate between alleles when relatively short synthetic oligonucleotide probes hybridise with target allele sequences but not with others, based on different thermal stabilities. However, control of hybridisation and wash temperatures is critical. In another technique, *Taq*Man ASO (Holland *et al.* 1991), the ability of *Taq* DNA polymerase to degrade oligonucleotide probes hybridised to target DNA during PCR is utilised but the method has been found to be impractical for genotyping of diverse SNPs.

Other techniques commonly used to detect SNPs include Southern blot analysis, denaturing gradient gel electrophoresis (Myers *et al.* 1987), the amplification refractory mutation system (ARMS) (Newton *et al.* 1989), and chemical cleavage methodologies (Cotton *et al.* 1988). These methods, however, are cumbersome and not amenable to automation (Reynolds *et al.* 1998). Direct fluorescence–based DNA sequencing (Hattori *et al.* 1993) can detect the exact nucleotide variation in a sample, but the technique is expensive.

DNA fingerprinting techniques such as AFLP, RAPD, DAF and AP–PCR are useful in studying complex genomes, but not so for examining specific nucleotide changes in the genome, as is needed for specific mutation or polymorphism detection (Reynolds *et al.* 1998).

1.2.1.ii Single-stranded conformation polymorphism (SSCP) analysis

Single-stranded DNA has a tendency to fold up and form complex structures stabilised by weak intramolecular bonds. These will be mostly base-pairing hydrogen bonds, therefore the precise structure formed will depend on the nucleotide sequence. The electrophoretic mobility of such structures through nondenaturing gels will depend not only on their chain lengths, but also on their conformations, which are dictated by the DNA sequence. For

SSCP, amplified DNA samples are denatured and then run on a non-denaturing polyacrylamide gel (Orita *et al.* 1989).

Although SSCP does not reveal the nature or position of any mutation detected, the technique is adequately sensitive, simple and cost effective (Sheffield *et al.* 1993). The precise pattern of bands seen is very dependent on details of the conditions, but provided that the temperature and length of electrophoresis are controlled, most single base differences in PCR products up to 400 bp in size will be distinguished (Hayashi 1996).

1.2.1.iii Development of new genetic markers

Traditionally genomic DNA libraries were the main source for the development of new markers. However, cDNA libraries are becoming the source of choice since a marker that originates from a cDNA sequence identifies a gene itself and is not merely linked to a gene (Hussey & Hunsperger 1996). Either cDNA or genomic DNA libraries can be constructed from a whole organism, a certain tissue or even a particular cell type (Kwak *et al.* 1997). To increase the chances of isolating a particular type of marker, for example a microsatellite marker, an enriched library would be constructed (Kijas *et al.* 1995). Isolation of a marker requires screening of the library by either hybridisation or sequencing of selected clones.

1.2.2 Physical mapping techniques

Physical maps consist of a linearly ordered set of DNA fragments encompassing the genome or region of interest. Physical mapping provides distinct benefits. Firstly the map provides access to any region of the genome which can be genetically identified. In other words, the physical map serves as a cloning tool by facilitating the movement from the genetic locus to the cloned gene. Secondly, physical maps provide a starting point for studying global genomic organisation. As an increasing number of genes are cloned and molecular biological information is accumulated, one can begin to investigate the physical linkage of cloned genes, study the organisation and distribution of repetitive elements, and analyse the correlation between physical and genetic distance.

The physical mapping techniques include restriction mapping, which gives the relative positions of restriction enzyme cutting sites on a DNA molecule, fluorescent *in situ* hybridisation (FISH), which positions marker locations by hybridising the marker to intact chromosomes, and sequence-tagged site (STS) mapping. This technique locates the positions of short sequences by PCR and/or hybridisation analysis of genome fragments. The physical map with the highest possible resolution is the genomic sequence itself but, well in advance of assembling the final genomic reference sequence, localized regions can be brought into focus as they are completed (Schuler 1998). At present the most powerful physical mapping technique is STS mapping (Brown 1999). In the human genome mapping project, an attempt to convert the genetic map to a physical map has been greatly simplified by using STSs (Schuler 1998).

1.2.2.i STS markers

"*An STS is any piece of DNA whose sequence is known and for which a specific PCR assay has been designed*" (Strachan & Read 1996: 595)

The concept of STSs was developed by Olson *et al.* (1989) in an attempt to systematise landmarking of the human genome. Two problems were experienced in the physical mapping of man. One was the difficulty of merging mapping data gathered by diverse methods in different laboratories into a consensus physical map. The second was the logistics and expense of managing the huge collections of cloned segments on which the mapping data would depend almost exclusively. Olson *et al.* (1989) proposed the use of short tracts of single–copy DNA sequences that can be easily recovered by PCR as landmarks that define positions on the physical map. An STS is fully portable once the two sequences of the primers are known and these can be obtained from databanks. STSs can also be developed for unique regions along the genome that vary in length from one individual to another.

To qualify as an STS a sequence must satisfy two criteria. Its sequence must be known so that a PCR assay can be set up, and the STS sequence must have a unique location in the genome. These are easy criteria to meet and STSs can be obtained in many ways. The most common source is expressed sequence tags (ESTs), which are short sequences obtained by analysis of cDNA clones (Mara *et al.* 1998). An EST can be used as an STS, assuming that it originates from a unique gene and not from a member of a gene family.

UNIVERSITEITSBIBLIOTEEK U. S.

Another class of STS markers amplifies microsatellite simple sequence repeat regions (Kijas 1995). These are referred to as sequence-tagged microsatellite sites (STMSs) (Beckmann & Soller 1990). For human (Weissenbach 1993), mouse (Kondo *et al.* 1993) and many other mapping initiatives, STMSs appear to be the method of choice. These markers are codominant and have a high level of polymorphism. However, within plant genomes information regarding microsatellite polymorphism, genomic distribution and relative abundance is scarce (Kijas 1995). The lengthy cloning and screening procedures required to identify candidate markers are more laborious in plants, which may have lower microsatellite copy numbers compared to mammals (Lagercrantz *et al.* 1993). For this reason ESTs seemed a more viable source for the development of STS markers.

Other DNA markers, such as RFLPs, can also be converted to STSs in physical mapping. Like STMSs they also provide a direct connection between the physical and genetic maps. Random genomic sequences obtained by sequencing random pieces of cloned genomic DNA, or sequences that have been deposited in databases, are also used as sources for STSs.

A database of STSs (dbSTS) is accessible at NCBI (href7). Most of the entries in the database originate from human (57 700). Currently there are no potato entries in the latest issue (September 1999) of the database, but other plants, such as rice (339) and maize (202), are accounted for.

1.2.2.ii Expressed Sequence Tags

Partial cDNA sequences are referred to as expressed sequence tags or ESTs (Adams *et al.* 1991)

The expressed component of complex genomes may constitute only a few percent of the total genome. One way of building transcript maps is to identify genes in defined clone contigs. An alternative and more general method is to obtain partial sequences of numerous randomly selected cDNA clones and then place these on physical maps. Even short sequences from a cDNA clone permit sequence-specific primers to be designed so that a PCR assay can be developed that is specific for that sequence.

A single sequencing reaction gives all the information necessary for extensive analysis of a partial gene sequence. If one is working with a representative library, every clone sequenced has the potential to be an interesting new gene. Single-pass sequencing, while less accurate than highly redundant contigs of overlapping sequences, is accurate enough for very sensitive similarity searches (Adams *et al.* 1991). Without redundancy, a large number of independent clones can be analysed, rather than tens to hundreds of overlapping clones to determine the sequence of a single gene.

ESTs have also served as molecular genetic markers in genomic mapping (Kurata *et al.* 1994; Shen *et al.* 1994). Since the number of ESTs from various species has increased rapidly, it is now possible to compare the large number of genes and proteins they encode between animals and plants. Genes expressed in different tissues within an organism have also been randomly sequenced. Kwak *et al.* (1997) examined the expression patterns of ESTs, derived from a guard cell library of *Brassica campestris*, in several plant organisms by means of RNA blot analysis. Together with the functional categorisation of ESTs, the gel-blot data indicated that guard cells have transcriptional activities distinctive from leaf cells as a whole, consistent with their unique roles. Comparison of ESTs between different tissues yields information on the dynamics of genomic expression patterns.

The first extensive analysis of a population of expressed genes has been reported for human brain tissue (Adams *et al.* 1991, 1992). The EST technique has been applied to several plant species, including *Arabidopsis* (Höfte *et al.* 1993; Newman *et al.* 1994; Cooke *et al.* 1996), *Brassica* (Park *et al.* 1993; Kwak *et al.* 1997), rice (Uchimiya *et al.* 1992; Kurata *et al.* 1994; Sasaki *et al.* 1994), maize (Keith *et al.* 1993; Shen *et al.* 1994), and citrus fruit (Hisada *et al.* 1997). However, there are as yet no reports on potato EST projects in the literature. The latest release (August 1999) of the database of ESTs (dbEST) (Boguski *et al.* 1993) reports only 85 EST entries for potato in contrast with 46 418 rice ESTs, 37 776 *Arabidobsis* ESTs and 30 472 entries for maize (href8).

1.2.2.iii ESTs as a source for the development of STSs

Tsumura *et al.* (1997) presented the use of STSs derived from cDNAs as molecular markers in *Cryptomeria japonica*. They used these STS primers to amplify panels of genomic DNA of *C. japonica* and other conifer species. The amplification products were subsequently digested with restriction enzymes and analysed on agarose gels. Perry and Bousquet

(1998a) have developed STS markers from black spruce (*Picea mariana*) cDNA. These polymorphisms may be observed by standard agarose gel electrophoresis without the manipulation of amplified products. Such markers combine the technical simplicity of RAPDs with the specificity of SSRs and they may often be codominant.

1.3 cDNA LIBRARIES

The ability to generate cDNA from mRNA has had a considerable impact on our current understanding of genome organisation and gene expression. In most cases the cDNA is cloned before the corresponding genomic copy of the gene. This is because cloning genes directly from genomic DNA can be time-consuming, and the process is accelerated if the respective cDNA clones are used as probes. A cDNA library is a representation of the genes expressed in the tissue from which the mRNA was isolated. Therefore cDNA libraries have been instrumental in the identification of tissue-specific and developmental-stage-specific genes.

Traditionally, cDNA libraries have been constructed in order to find one gene or a few genes of interest. However, with the introduction of the EST studies related to the Human Genome Project, the interest in cDNA libraries has been expanded to finding all the genes which are expressed in a particular cell or tissue (Adams *et al.* 1991, 1992; McCombie *et al.* 1992). cDNA libraries have also formed the basis of many of the current large-scale shotgun sequencing projects. Different approaches for the construction of cDNA libraries are discussed below.

1.3.1 PCR-based cDNA libraries

Despite the existence of spesialised approaches for cloning cDNA the procedures still remain highly inefficient. Only a minor fraction of the cDNA clones give rise to recombinant clones, of which some 10^6 may have to be screened to find clones corresponding to rare mRNAs (Kimmel and Berger 1987). In general, after cDNA synthesis, only 10% of the input mRNA is actually converted to cloneable double-stranded cDNA. In addition, with the use of *in vitro* phage packaging systems or transformation of competent bacteria with plasmid vectors, the amount of ligated cDNA that is introduced into cells is even further reduced. Therefore, to have a statistically significant chance of finding the desired sequence, a technique that generates large numbers of clones from small amounts of template mRNA is essential.

This limitation of the cloning process was overcome by the subsequent amplification of cDNA *in vitro*. In 1989 Belyavsky *et al.* described a novel modification of the polymerase chain reaction for the amplification of only cDNA. In contrast to standard PCR, whereby amplification of only one sequence or a group of related sequences is achieved (Saiki *et al.* 1985), their scheme allowed the amplification of virtually all cDNA species present in the reaction mixture. Moreover, even long cDNAs, in the size range of 2.4 - 4.4 kb, were found to be amplified with reasonably high efficiency.

For amplification-based cDNA library construction, the ability to amplify all cDNA species independently of their sequence is of central importance. The cDNAs must possess a known DNA sequence (~20 bp) at each end. These ends can be generated by homopolymer tailing with terminal transferase (Gurr *et al.* 1991), by ligating adapters/linkers to the cDNAs (Ko *et al.* 1990), or by using primers that anneal by means of random hexamers at their 3'-ends (Froussard 1992). The 3'-end of the majority of eukaryotic mRNA already contains a poly(A) tail, which can be utilised in cDNA synthesis.

1.3.2 Modified cDNA libraries

1.3.2.i Subtracted cDNA libraries

Subtractive hybridisation has been applied successfully to a wide range of biological problems to identify differentially expressed genes of unknown sequence. The conventional approach has consisted of isolating clones with a labelled subtractive cDNA probe designed to detect cDNA clones which are present in one cell type, but absent or expressed at significantly lower levels in another cell type (Sargent 1987). Subtractive hybridisation typically involves hybridising tens of micrograms of poly(A)+ RNA (driver) at 10-fold excess with cDNA (target) prepared from mRNA from another cell type. The double-stranded RNA-DNA hybrids represent sequences which are present in both cell types and can be removed (subtracted) from the unhybridised single-stranded preparation using hydroxyapatite chromatography (Sargent 1987). The remaining unhybridised cDNA can be used as a probe to screen libraries or used to generate a subtractive library if enough product is available.

Fargnoli *et al.* (1990) recently developed another protocol dependent upon low-ratio hybridisation subtraction between driver RNA and target cDNA that was found to enrich for cDNAs representing low-abundance transcripts, which are induced only several-fold over the base level. The hybridisation was carried out in a phenol emulsion that altered the kinetics such that the rates of association between driver RNA and target cDNA were increased markedly. Therefore, this subtraction technique allows for the isolation of clones that are quantitatively as well as qualitatively different between cell types.

One of the most common problems with subtraction techniques is that specific mRNAs or cDNAs from the target tissue are subtracted by common sequences or repetitive elements yielding very low amounts of cloneable material. Another major disadvantage of these procedures is that they require large amounts of driver mRNA which may be difficult to obtain for some studies in which tissues are limited.

Various methods have been developed recently which combine PCR and solid-phase capture technologies to generate subtractive libraries. Several investigators have published techniques which base the subtractive hybridisation on ds cDNA which has been prepared from small amounts of mRNA and amplified with primers complementary to oligonucleotide linkers (Wang & Brown 1991). These procedures employed the photobiotinylation of the driver DNA in order to improve the efficiency of removal of hybridised DNA sequences.

Hara *et al.* (1991) introduced a novel strategy for preparing a subtracted library using an oligo(dT) primer covalently linked to latex particles. Driver cDNA is synthesised on latex beads using mRNA template from one tissue source. After removal of the mRNA template by heat denaturation and centrifugation, the first strand remains bound to the solid support where it can serve as a subtractive probe for several rounds of subtraction. The unhybridised subtracted mRNA can be recovered and converted to cDNA. Following a typical subtraction, only a few nanograms of target sequence remain and, therefore, the ds cDNA is first ligated to linkers, amplified with PCR and then subcloned into an appropriate vector.

Another technique that generates tissue-specific or stage-specific subtracted probes using magnetic oligo(dT) beads has been described (Rodriguez 1992). In this method cDNA is synthesised on the magnetic beads from total RNA (target) as well as from driver mRNA. A randomly primed radiolabelled probe is synthesised from the target cDNA and subtraction hybridisation is performed with an excess of driver cDNA bound to the solid support. The radiolabelled probe that remains can then be used to screen a library.

Many techniques have been developed for the selection of cDNAs by hybridisation with fragments of genomic DNA immobilised on a solid support (Parimoo *et al.* 1991). Low-abundance cDNAs encoded by large genomic clones can be isolated, amplified and subcloned. In the simplest of these techniques, biotinylated cloned genomic DNA is hybridised in solution with amplifiable cDNAs. The genomic clones and hybridised cDNAs are captured on streptavidin-coated magnetic beads, the cDNAs eluted, amplified with PCR and subcloned. All of the selected cDNAs that were initially present at very low abundance were found to be increased as much as 100 000-fold after two rounds of enrichment.

1.3.2.ii Normalised cDNA libraries

The abundance of different mRNA varies characteristically within a cellular population, as well as during different stages of development. An individual class of transcript can represent from one in 10^6 to >10% of the total message population (Galau *et al.* 1977). The variation in abundant classes in given cell types implies that several hundred thousand clones have to be screened to have a reasonable chance of finding a particularly rare transcript (Sambrook *et al.* 1989). The situation is even more complex with a tissue which comprises numerous cell types (Ohlsson 1989).

For the efficient cloning of cDNA belonging to a rare mRNA class it is important to establish a library in which the frequency of the occurrence of every cDNA clone is the same. Normalisation is the process by which the more abundant messages are reduced in number and rare messages increased relative to other clones within the library. Two approaches have been proposed. One approach depends on hybridisation with genomic DNA in order that the relative frequencies of cDNAs is made proportional to the frequency of genes complementary to those cDNAs in genomic DNA. The other approach is based on the reannealing of double-stranded DNA in solution following reassociation kinetics (Galau *et al.* 1977), assuming cDNA reannealing follows second-order kinetics. This implies that abundant species will anneal faster than rare species, leaving a pool of unhybridised single-stranded cDNA rich in rare species.

Only a limited number of attempts, however, have been made to construct a normalised library. Among them Ko (1990) and Patanjali *et al.* (1991) independently reported that the use of the second-order kinetics approach resulted in normalisation of the clones in cDNA libraries, containing an approximately equal representation of all sequences present in the initial preparation of poly(A)+ RNA. Sasaki *et al.* (1994) presented a semi-solid system to achieve self-hybridisation using cDNA immobilised on latex beads.

1.4 AIMS OF THIS STUDY

Plant breeders and molecular biologists are developing new varieties and cultivars by introducing favourable genes, such as disease-resistance genes, into their crops. Molecular markers are necessary to identify genes and to track their transfer by marker assisted selection. Markers derived from cDNA sequences have the advantage of pointing directly to a gene and cDNA libraries are therefore an important tool in genome analysis studies. ESTs developed from these libraries are used as tags for genes, while STS markers can also be derived from them. STS markers used as landmarks in genome mapping provide a number of advantages. STSs are free of context within a particular cloned segment, which obviates the need to maintain and distribute any biological material, and STSs are relatively simple to generate. As with any PCR-based markers that target specific sequences, considerable cost is incurred during the development of cDNA-based STS markers. Sequence data have to be obtained, followed by primer synthesis and testing. If primers generated for one species could be used in others, the time and resources required for the launching of studies involving related species would be significantly reduced.

The aims of my study were as follows:

1) to construct a cDNA library from potato leaf material;
2) to characterise selected inserts according to size;
3) to develop ESTs by sequencing selected inserts;
4) to develop STS markers by designing primers from these sequences;
5) to screen a panel of potato genomic DNA with these primers and detect polymorphisms;
6) to investigate the possibility of using these markers in other species of the Solanaceae family.

CHAPTER 2

MATERIALS AND METHODS

2.1 PLANT MATERIAL

Ten potato cultivars and five other Solanaceae species were used in this study (Table 2.1). The Vegetable and Ornamental Plant Research Institute (ARC, Roodeplaat) provided the potato genomic DNA.

The potato cultivar Aviva was used for RNA isolation. The plant was received from the Vegetable and Ornamental Plant Research Institute (ARC, Roodeplaat) and grown in a temperature-controlled greenhouse with supplementary lighting. Leaf material was collected, immediately frozen in liquid nitrogen and stored at -80°C until needed.

Table 2.1 The potato cultivars and members of the Solanaceae family used in this study. ID indicates the number or letter used to identify each entry on gel images.

Potato cultivars		**Solanaceae family**	
ID	**Cultivar**	**ID**	**Species**
1	ALTO	GP	Green pepper (*Capsicum annuum*)
2	Aviva	P	*Petunia*
3	Bravo	Po	Potato (*Solanum tuberosum*)
4	Baroc	Sn	*Solanum nigrum*
5	Charlie	Tm	Tomato (*Lycopersicon esculentum*)
6	Dawn	To	Tobacco (*Nicotiana tabacum*)
7	Devlin		
8	Darius		
9	Calibra		
10	Caren		

2.2 MEASURES TO PREVENT RNA DEGRADATION BY RNASES

Frozen leaves were kept in liquid nitrogen prior to the lysis step to prevent thawing. All solutions were kept cold (4°C) and procedures were performed as quickly as possible. Eluted RNA was kept on ice at all times and an RNase inhibitor was added if storage was required.

Sterile, disposable plastic ware was used wherever possible. Disposable gloves were worn and changed frequently. Surfaces and pipettes were treated with RNase AWAY^{TM} (MβP). Glassware was soaked overnight in 0,1% diethyl pyrocarbonate (DEPC) and autoclaved to remove any traces of DEPC.

Water and salt solutions were treated with 0.1% DEPC for at least an hour at 37°C and autoclaved. Buffers containing Tris was not treated in this way as Tris reacts with DEPC. To prepare a Tris buffer, the water was DEPC-treated and autoclaved before adding Tris. The solution was then autoclaved again. Solutions for RNA manipulations were kept separate to avoid possible contamination.

2.3 RNA EXTRACTION

Three methods were used and evaluated for the isolation of total RNA. These were the RNeasy Plant Mini kit (Qiagen), TRIzoL (GibcoBRL) and the QuickPrep Total RNA Extraction kit (Pharmacia). The procedures were performed as specified by the manufacturers. A hundred micrograms of frozen leaf tissue were used as starting material for the first two methods and method 500 mg of leaf tissue for the third.

Total RNA was quantified by means of a Pharmacia GeneQuant Spectrophotometer. The quality of the RNA was evaluated by loading 5 µl of the sample, 10 µl Ficoll Orange G and 1 µl ethidium bromide onto a 1% agarose (LE, FMC) gel. Electrophoresis was carried out in 1 X TBE running buffer at 80 V for 60 minutes. Gels were analysed under UV light.

mRNA was isolated using the Dynabeads mRNA Purification kit (Dynal) according to specifications of the manufacturers. As starting material the total yield of a total RNA preparation was used. The mRNA was eluted in 10 µl Elution solution provided with the kit. The concentration of the mRNA was determined on a Pharmacia GeneQuant Spectrophotometer.

2.4 LIBRARY CONSTRUCTION

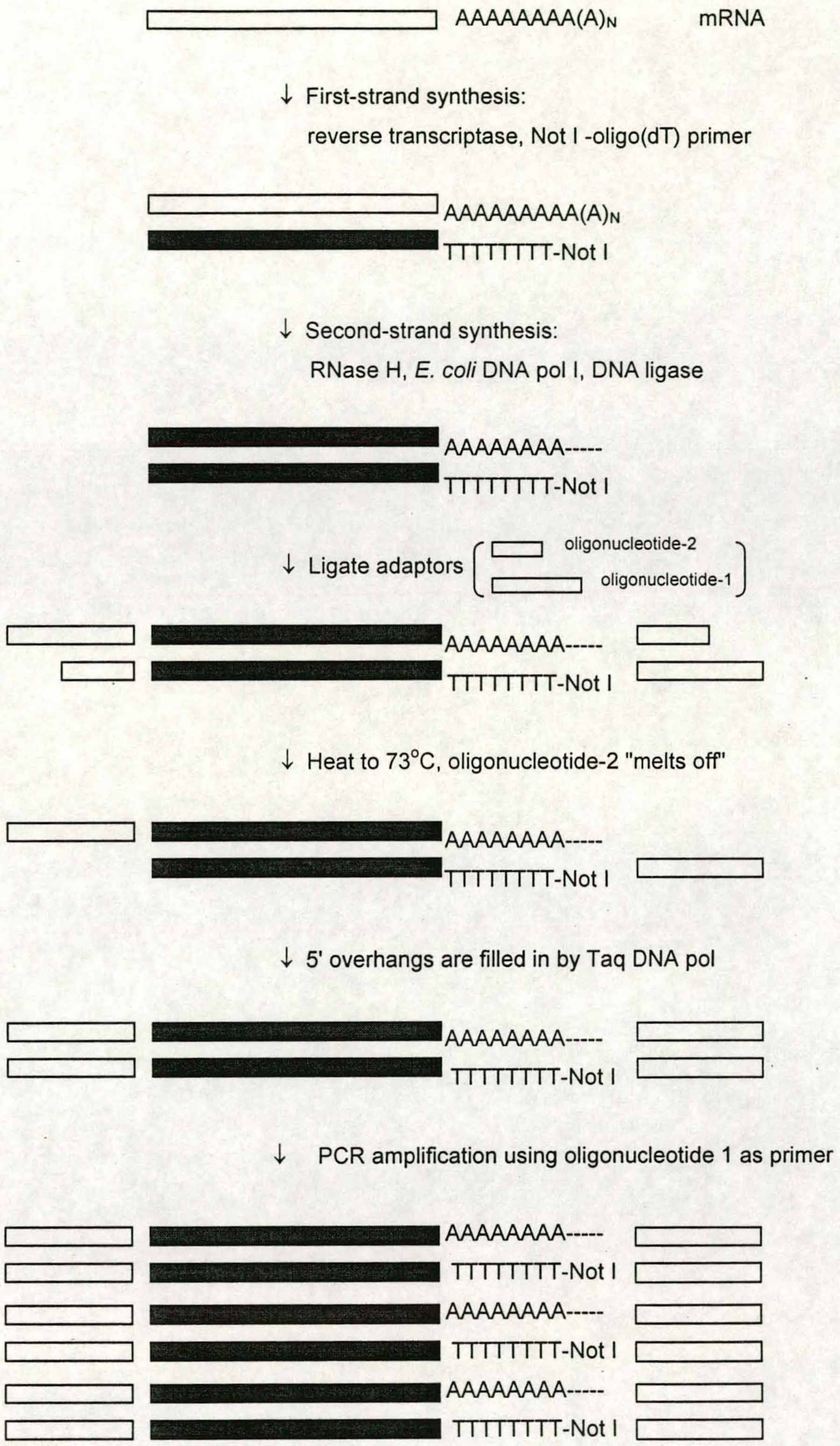
2.4.1 cDNA synthesis

The procedure followed for cDNA synthesis is outlined in Figure 2.1. The first-strand was synthesised as follows. Not I -oligo(dT) primer was added to 500 ng of mRNA at a final concentration of 2.5 µM. The reaction was incubated at 65°C for 10 minutes and immediately cooled on ice. Subsequently the following components were added to the reaction: 5 X RT buffer (provided with the enzyme, GibcoBRL), 10 mM DTT, 1 mM dNTP mix, 20 units RNase inhibitor (Boehringer Mannheim), 100 units Superscript RNase H reverse transcriptase (GibcoBRL) and sterile water up to a final reaction volume of 20 µl. The reaction was placed on ice after incubation at 42°C for 45 minutes.

Directly after the first-strand synthesis the following components were added to the same tube: 1 X buffer (50 mM Tris-HCl, pH 7.6; 100 mM KCl; 5 mM $MgCl_2$), 5 mM DTT, 0.1 mM NAD, 10 mM $NH_4(SO_4)$, 0.8 units RNase H (Boehringer Mannheim), 23 units *E. coli* DNA polymerase 1 (Boehringer Mannheim), 0.5 units DNA ligase (Epicentre Technologies), 0.05 mg/ml BSA and sterile water to a final volume of 100 µl. This was incubated for two hours at 14°C. The reaction was inactivated by incubation at 70°C for 10 minutes, centrifuged briefly and placed on ice.

Two units of T4 DNA polymerase were added to the reaction to ensure that the cDNA was blunt-ended. The reaction was incubated for 10 minutes at 37°C. To stop the reaction 10 mM EDTA was added. A phenol:chloroform:isoamyl alcohol extraction followed by a chloroform:isoamyl alcohol extraction were performed. The cDNA product was purified through a QIAquick PCR Purification column (Qiagen) and eluted in 40 µl elution buffer provided with the kit.

Figure 2.1 cDNA synthesis and amplification based on the method described by Jepson *et al.* (1991).


AAAAAAAA(A)N
mRNA
↓ First-strand synthesis:
reverse transcriptase, Not I -oligo(dT) primer
AAAAAAAAA(A)N
TTTTTTTT-Not I
↓ Second-strand synthesis:
RNase H, E. coli DNA pol I, DNA ligase
AAAAAAAA-----
TTTTTTTT-Not I
↓ Ligate adaptors
oligonucleotide-2
oligonucleotide-1
AAAAAAAA-----
TTTTTTTT-Not I
↓ Heat to 73°C, oligonucleotide-2 "melts off"
AAAAAAAA-----
TTTTTTTT-Not I
↓ 5' overhangs are filled in by Taq DNA pol
AAAAAAAA-----
TTTTTTTT-Not I
↓ PCR amplification using oligonucleotide 1 as primer
AAAAAAAA-----
TTTTTTTT-Not I
AAAAAAAA-----
TTTTTTTT-Not I
AAAAAAAA-----
TTTTTTTT-Not I

2.4.2 Ligation of amplification adaptors

Forty micrograms of oligonucleotide-1 (5' - ATGCTTAGGAATTCCGATTTAGCCTCATA - 3') and 100 μg of oligonucleotide-2 (5' - TATGAGGCTAAA - 3') were dissolved in 50 μl of STE buffer. To anneal the oligonucleotides, the reaction was heated in a water bath to 70°C for 2 minutes. The water bath was turned off and left to cool to 30°C.

An aliquot (2.5 μl) of the annealed oligonucleotide adaptor set was ligated to 10 μl cDNA by means of the Rapid DNA Ligation kit (Boehringer Mannheim) in a final reaction volume of 20 μl according to the instructions of the manufacturers. The reaction was inactivated by phenol:chloroform extraction.

2.4.3 Amplification of cDNA

The ligated cDNA product was size-fractionated using Quick SpinTM (TE) Linkers 6 columns supplied by Boehringer Mannheim, which have an exclusion limit of 194 bp. The volume of the sample obtained from the column was approximately 35 μl.

The ligated, size-fractionated cDNA was amplified, using oligonucleotide-1 as primer. A 50 μl reaction contained 1 X PCR buffer, 2 mM $MgCl_2$, 200 mM of each dNTP, 600 ng oligonucleotide-1, 5 μl cDNA template and 2.5 units Taq DNA polymerase (Bioline).

The following PCR thermal profile was performed:

73°C for 2 minutes,
followed by 45 cycles of
94°C for 48 seconds,
68°C for 68 seconds and
73°C for 3 minutes.

Ten microliters of the PCR product were run on a 1.5% agarose gel to confirm that the amplification had been successful. Electrophoresis was carried out at 80 V for 90 minutes in 1 X TBE running buffer containing ethidium bromide. Gels were analysed under UV light. A 1 kb ladder (GibcoBRL) was used as control molecular marker.

2.4.4 Cloning of cDNA

2.4.4.i Ligation

The remaining 40 μl of amplified cDNA was digested with Eco RI and Not I, using 20 units of each enzyme, 1 X SuRE/Cut™ buffer (Boehringer Mannheim) and dH_2O to a final volume of 100 μl. The reaction was incubated at 37°C for 1 hour after which another 10 units of each enzyme was added and the reaction incubated for a further 90 minutes. To inactivate the enzymes, the reaction was heated to 70°C for 15 minutes. The cDNA was purified with the QIAquick PCR Purification kit (Qiagen) and eluted in 30 μl elution buffer provided with the kit.

The same procedure was used to digest the plasmid vector, pBluescript® SK+ (pSK) (Stratagene). Subsequently, the reaction was run on a 2% Nusieve® GTG® agarose (FMC) gel for 2 hours at 80 V. The digested vector was extracted from the gel using the QIAquick Gel Extraction kit (Qiagen) in 30 μl elution buffer. The concentration of the digested cDNA and vector was determined on a 1% agarose gel by means of concentration standards and the Gel Doc 1000 (Bio Rad) gel documentation system.

The ligation was performed according to the Rapid DNA Ligation kit (Boehringer Mannheim). Different ratios of insert:vector DNA were tested to determine optimum ligation conditions. Typically, 10-100 ng of cDNA was used in a reaction at an insert:vector ratio of 4:1. The reaction was incubated for 2 hours at room temperature.

2.4.4.ii Transformation of competent cells

E. coli DH5α (Stratagene) competent cells were prepared according to the Calcium Chloride method described by Nakata *et al.* (1997). A hundred microliters of competent cells were transformed with half of the ligation mix (5 μl) as described by the suppliers (Stratagene). Non-recombinant pSK was also transformed as a control reaction. The cells were centrifuged at 1000 rpm for 10 minutes and plated onto LB plates containing ampicillin (Amp) at a final concentration of 50 μg/ml, 40 μl 100 mM IPTG solution and 40 μl 50 mg/ml X-gal solution. The plates were incubated at 37°C overnight.

2.4.5 Library organisation

After overnight incubation, blue and white colonies were visible on the LB/Amp/IPTG/X-gal agar plates. White colonies identify recombinant plasmids and blue colonies identify non-recombinants. The white colonies were individually picked with sterile tooth picks and deposited into the wells of a multi-well plate containing 200 µl LB medium with Amp (50 µg/ml) per well. Each individual clone was given a specific ID in the following format: plate - numeral, row - letter, and column - numeral. The multi-well plates were incubated overnight at 37°C with shaking. Glycerol was then added to a final concentration of 15% and the plates stored at –80°C.

When the colonies were picked, they were also duplicated on fresh LB/Amp agar plates with ID corresponding to that of the multi-well plate. These plates were used for further manipulation to prevent unnecessary thawing of the frozen cultures.

2.5 CHARACTERISATION OF INSERTS

2.5.1 Size determination of inserts

The insert size was determined by means of PCR. The amplification was done directly on the colonies to avoid preparing liquid cultures and performing plasmid extractions.

M13 forward and reverse primers were used to amplify the total insert. The 12.5 µl PCRs contained the following: 1 X PCR buffer, 2 mM $MgCl_2$, 0.378 pmol of each primer, 200 mM of each dNTP and 0.25 units Taq DNA polymerase (Bioline). A colony was then picked and mixed with the reaction mixture.

The temperature profile for the reactions was:

94°C for 7 minutes,
followed by 40 cycles of
94°C for 45 seconds,
60°C for 1 minute, and
72°C for 2 minutes.

The PCR products were analysed on a 1% agarose gel as described in paragraph 2.4.3.

2.5.2 Sequencing

Individual cDNA clones were candidates for sequencing if a single product of size ≥ 500 bp was present in the insert amplification. PCR products were purified with the QIAquick PCR Purification kit (Qiagen). Sequencing was performed using the BigDye™ Terminator Cycle Sequencing kit (PE Applied Biosystems) and M13 forward and reverse primers. Each reaction contained 3.3 pmol primer and 10 to 30 ng template cDNA. The sequencing products were analysed on an ABI PRISM™ 377 automated DNA sequencer (PE Applied Biosystems) at the U.S. core DNA sequencing facility.

2.5.3 Data analysis

Sequences were edited manually by removing vector and ambiguous sequences. Comparative sequence analysis was conducted with the BLAST (Basic Local Alignment Search Tool) program (Altschul *et al.* 1990) against the National Center for Biotechnological Information (NCBI) non-redundant protein (BLASTX) and nucleic acid (BLASTN) databases (href9). Clones with significant alignments were assigned a putative identity. In comparing the sequences with sequences in the database, a Poisson *P*-value of less than 0.01 was considered to indicate significant similarity (Altschul *et al.* 1994).

The ORF (Open Reading Frame) Finder program (href10) was used to detect open reading frames in the ESTs. The ORF Finder is a graphical analysis tool that detects all open reading frames of a selectable minimum size in a user's sequence, using the standard or alternative genetic codes.

2.6 PRIMER DESIGN

Primer sequences were selected using the computer program, Primer Design V1.01. Typically, 20-mers with a GC content between 45 and 55%, minimal secondary structure and no significant inter-primer complimentarity were selected. The complete insert sequence was taken into consideration in the design of the primer. Primer sets were ordered from either Perkin Elmer or GibcoBRL.

2.7 SCREENING OF GENOMIC DNA

2.7.1 Genomic DNA isolation

Genomic DNA was isolated using a modified version of the method described by Edwards *et al.* (1991). Two leaf discs were collected in a microfuge tube using the lid as a punch. The discs were ground in the tubes in the presence of Carborundum (400 grit). The ground material was incubated in 400 µl prewarmed Supaquick buffer (200 mM Tris-HCl, pH 7.5; 250 mM NaCl; 25 mM EDTA; 0.5% SDS) at 60°C for 30 minutes. The DNA was extracted with chloroform:isoamyl alcohol and precipitated with isopropanol. Finally the DNA was resuspended in 40 µl of TE buffer and stored at –20°C.

2.7.2 PCR conditions

Genomic amplification was conducted, using 40 ng of DNA. The $MgCl_2$ concentrations and annealing temperatures were varied until optimal results were obtained. Table 2.2 summarises the optimal conditions used for each of the EST primer sets.
The temperature profile for the reactions was:

94°C for 5 minutes,
followed by 35 cycles of
94°C for 30 seconds,
52 - 55°C 30 seconds, and
72°C for 1 minute, and
a final elongation at 72°C for 7 minutes.

Table 2.2 PCR conditions for different ESTs. The primer pair designed for EST 6f12 did not produce a clear PCR product under any of the conditions tested and was excluded from the screening.

EST	Annealing temp (°C)	[$MgCl_2$] (mM)
2g4	52	1.5
3c10	55	1
3d6	52	1.5
3d11	54	1.5
3h5	44 (10x), 45 (25x)	0.75
3h7	54	1
6b10	55	1
6b11	55	1
6d6	55	1
6h9	55	1.5
7a5	55	1.5
7b4	55	1.5
7e9	55	1.5

2.7.3 Agarose gel analysis

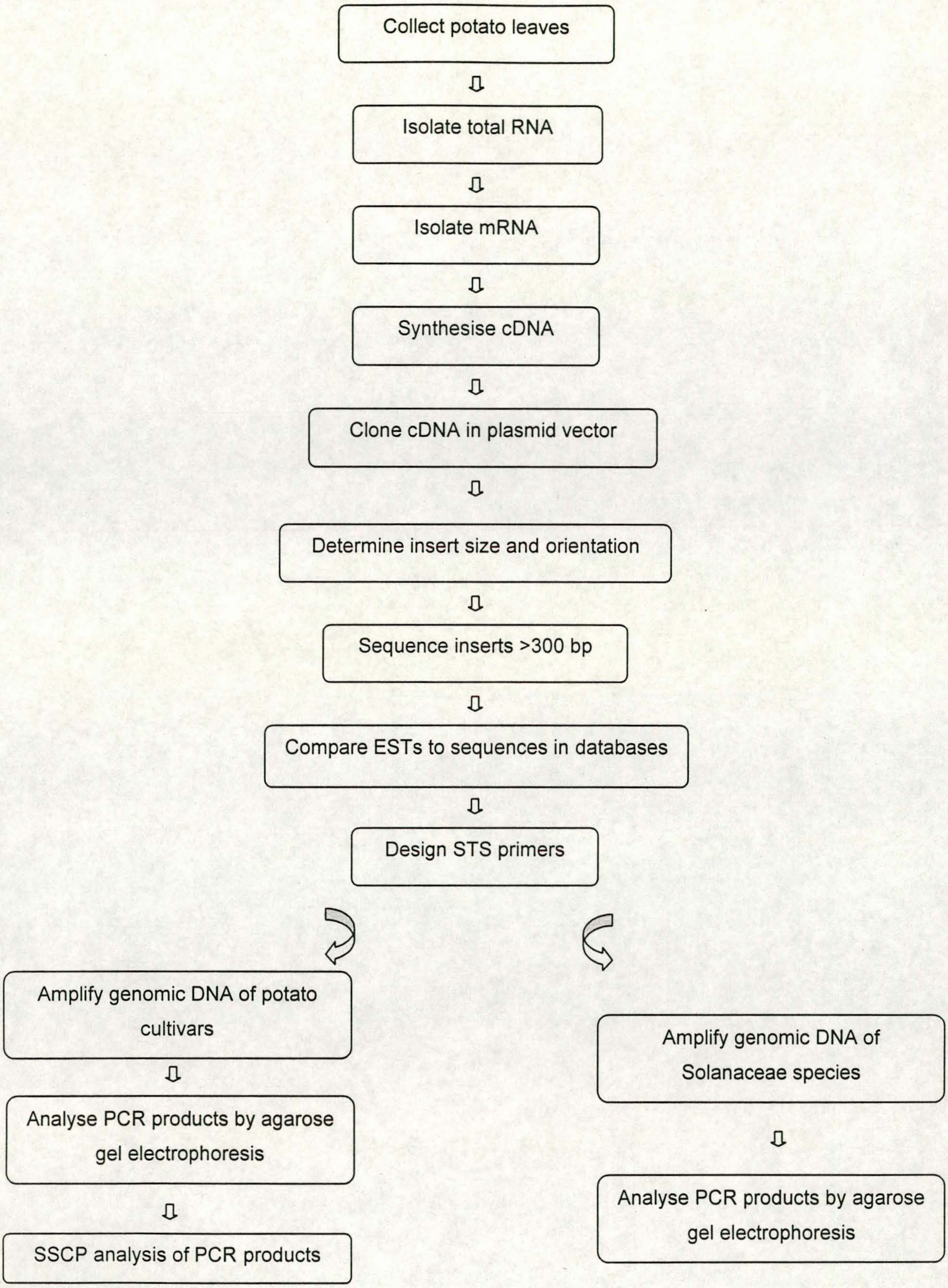
PCR analyses were performed with each primer set on the 10 potato cultivars and five other species of Solanaceae described in paragraph 2.1. PCR products were electrophoresed on a 1.5% agarose gel as described in section 2.4.3.

2.7.4 Single-stranded conformation polymorphism (SSCP) analysis

PCR products of the potato cultivars were analysed on 20% nondenaturing polyacrylamide gels with 1% cross-linking. Twenty microliters of the PCR product were mixed with 8 μl of loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) and denatured in a heating block at 100°C for 3 minutes. The electrophoresis was performed using a Hoefer Mighty Small II apparatus at 200 V for 4 to 6 hours at 15°C. The gels were stained with ethidium bromide and DNA was visualised under UV light.

Five SSCPs were characterised by automated sequencing. Polymorphic bands of primer pairs 3d11, 6d6, 7a5, 7b4 and 7e9 were reamplified and sequenced in both directions with the corresponding forward and reverse primers, as described in 2.5.2. The fragments selected for reamplification were those illustrated in Chapter (Figures 3.10 d, i, k, l and m). Alignment of the sequences was done using the Sequence Navigator program (PE Applied Biosystems).

2.8 SUMMARY OF THE PROCEDURES FOLLOWED


Collect potato leaves
Isolate total RNA
Isolate mRNA
Synthesise cDNA
Clone cDNA in plasmid vector
Determine insert size and orientation
Sequence inserts >300 bp
Compare ESTs to sequences in databases
Design STS primers
Amplify genomic DNA of potato cultivars
Analyse PCR products by agarose gel electrophoresis
SSCP analysis of PCR products
Amplify genomic DNA of Solanaceae species
Analyse PCR products by agarose gel electrophoresis

CHAPTER 3

RESULTS AND DISCUSSION

A cDNA library is a versatile tool in genome analysis. cDNA libraries are often used to isolate known genes, for example, a transferase (Schmidt *et al.* 1999) and a glucosidase gene (Taylor *et al.* 1998) were isolated from potato cDNA libraries. The various EST[•] studies have revealed that single-pass automated sequencing of randomly chosen cDNA clones is an efficient strategy for identifying unknown genes. Existing gene resources may also serve as a valuable source for identifying genetic variation. Buetow *et al.* (1999), for instance, identified more than 3 000 candidate SNPs in the human genome by examining EST data. cDNA libraries are also a source for other types of markers such as STSs[•]. Until recently, the great majority of STSs have been derived from anonymous genomic sequences. This has been highly succesful, efficient and elegant in its simplicity. However, as Wilcox and co-workers proposed in 1991, the development of STSs from the 3' untranslated regions (3' UTRs) of mRNAs has the advantage of supplying not only a marker, but also a gene for the map. In this study a potato cDNA library was used to develop STSs from the 3' ends of mRNAs. These STS primers were used to identify polymorphisms in potato cultivars.

3.1 ISOLATION OF RNA

Each organism, organ and tissue have a unique population of mRNA molecules. These mRNA populations are difficult to maintain, clone and amplify, due to the ubiquitous and persistent nature of highly active RNases in plant tissues. Therefore, they must be converted into more stable DNA or complementary DNA (cDNA) molecules. Pure, undegraded mRNA is essential for the construction of large, representative cDNA libraries and successful cDNA synthesis should yield full-length copies of the original mRNA. Hence, the quality of a cDNA library can only be as good as the quality of the mRNA. All RNA extraction techniques are therefore based on initial inactivation or inhibition of RNases by chemical or physical means, such as suboptimal pH and temperature. This is followed by separation of nucleic acids from proteins, thereby separating the RNA from RNases.

[•] See p. 9 (STS) and p. 11 (ESTs) for definitions.

3.1.1 Isolation of total RNA

The isolation of undegraded ribonucleic acid from cells and tissues involves three steps:

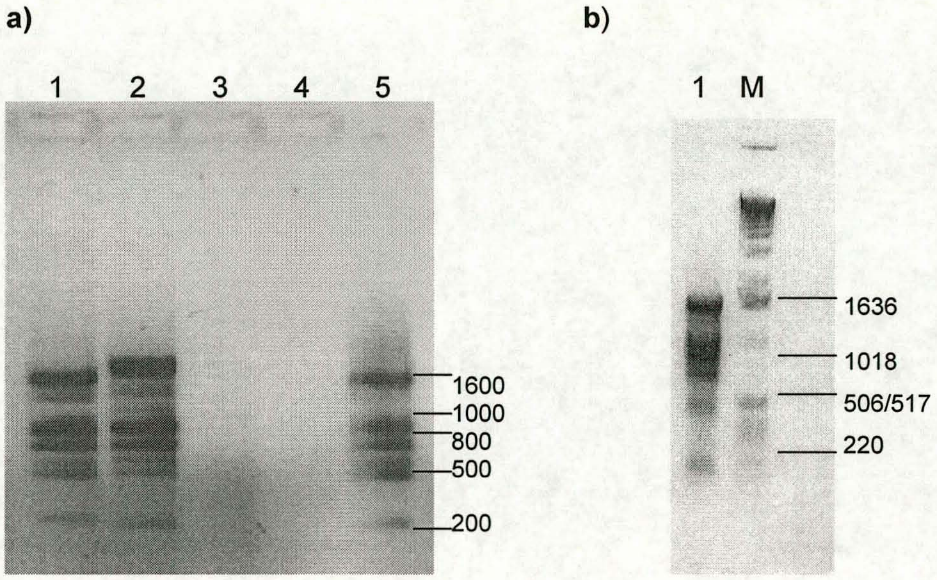
1) inhibition of endogenous nucleases,
2) deproteinisation of the RNA, and
3) physical separation of the RNA from the other components of the homogenate.

Several different protocols exist for the isolation of total RNA and mRNA. In most cases they involve either the use of detergents like cetyltrimethylammonium bromide (CTAB) or SDS, phenol extraction or density gradient centrifugation. Due to the nature of degradative nucleases, procedures that utilise high concentrations of chaotropic agents during cell lysis tend to be more reliable due to the simultaneous inactivation of RNases. These procedures also tend to generate more intact RNA. Inactivation of nucleases parallels the kinetic efficiency of protein denaturation, so that the relative potencies are guanidinium thiocyanate > guanidinium hydrochloride > urea. Including a reductant, such as 2-mercaptoethanol or dithiothreitol, to break intramolecular protein disulphide bonds that are essential for RNase activity (Sela *et al.* 1956), enhances denaturation. The addition of either competitive inhibitors of ribonucleases or chemicals such as diethyl pyrocarbonate (DEPC) does not enhance denaturation, and the latter's reactivity toward nucleic acids renders it undesirable (Ehrenberg *et al.* 1974). Guanidinium thiocyanate and chloride are among the most effective protein denaturants (Gordon 1972). Cox (1968) first introduced guanidinium chloride, which is both a strong inhibitor of ribonucleases and a deproteinisation agent, for isolation of RNA. Since then guanidinium extraction has become the method of choice for RNA purification, replacing phenol extraction (Logemann *et al.* 1987). For those samples with very high RNase content or for irreplaceable samples with unknown RNase content, guanidinium thiocyanate becomes the deproteinising agent of choice. Chirgwin *et al.* (1979) maximised the rate of denaturation by the combined use of guanidinium thiocyanate and 2-mercaptoethanol. The RNA is separated from the proteins, DNA, and polysaccharides by several ethanol precipitation steps or by sedimentation through caesium chloride. Several commercially available kits allow for the rapid isolation of RNA. In addition they provide the convenience of guaranteed RNase-free components.

In this study three methods were used to isolate total RNA: the Qiagen RNeasy Plant Mini kit, TRIzOL from GibcoBRL and Pharmacia's QuickPrep Total RNA Extraction kit. Total RNA of high quality was isolated with the first two methods mentioned (Figure 3.1(a) and (b)). Both these methods utilise guanidinium-based extraction buffers. TRIzOL is a monophasic solution of phenol and guanidinium isothiocyanate (GITC). The extraction buffer of the Qiagen kit also contains GITC, but no phenol, which makes it safer to use.

The Pharmacia kit, using a guanidinium thiocyanate buffer as well, yielded no or very degraded RNA (See Figure 3.1(a), lanes 3 and 4). This was possibly due to the lack of proper equipment. A vacuum aspiration system is required to remove the liquid phase and protein coat, which contain RNases, from the RNA pellet during the isolation procedure. As an alternative a pipette was used which could have led to insufficient removal of contaminants and subsequent degradation of RNA. In general this method is much more laborious and time consuming than the other methods used and requires five times more starting material.

Figure 3.1 (a) Total RNA isolated with the Qiagen kit (lane 1, 2 and 5) and the Pharmacia kit (lane 3 and 4). Estimated sizes are given in bp (marker not shown). **(b)** Total RNA isolated with the TRIzol method (lane 1). M = 1 kb ladder (GibcoBRL), the sizes are given in bp.



3.1.1.i Determination of quantity of total RNA

The concentration of RNA can be determined by measuring the absorbance at 260 nm (A_{260}) in a spectrophotometer. Absorbance readings should be greater than 0.05 to ensure significance. An absorbance of one unit at 260 nm corresponds to 40 µg of RNA per ml ($A_{260} = 1 = 40$ µg/ml). Absorbance readings were done using a Pharmacia GeneQuant Spectrophotometer. Approximately 20 µg total RNA per reaction was isolated with the TRIzoL method. The Qiagen kit yielded between 10 and 20 µg per reaction. Dudareva *et al.* (1996) found similar results with the Qiagen kit, isolating total RNA from *Clarkia breweri* flowers. The only report on total RNA yield from potato leaves was by Logeman *et al.* (1987), who isolated between 30 and 40 µg total RNA/100 mg tissue using a guanidinium hydrochloride method.

3.1.1.ii Determination of quality of total RNA

The quality of total RNA can be measured in terms of purity and integrity. The ratio between the readings at 260 nm and 280 nm (A_{260}/A_{280}) provides an estimate of the purity of RNA. Pure RNA preparations have A_{260}/A_{280} values of 2.0. Since the absorbance of any particular molecule is a function of its total base composition and secondary structure, RNA samples displaying A_{260}/A_{280} values in the range of 1.8 to 2.0 can be regarded as highly pure. A_{260}/A_{280} values of 1.85 to 2.1 were calculated for both extraction methods used, which indicated that pure RNA was isolated.

Agarose gel electrophoresis and ethidium bromide staining can be used to check the integrity and size distribution. Ribosomal RNA can be seen as bands of different sizes on the gel. Lanes 1, 2 and 5 in Figure 3.1(a) shows results of the Qiagen extraction method and Figure 3.1(b) shows an example of RNA isolated with TRIzoL. The RNA was found to be intact as was indicated by the clear bands, from about 200 to 1600 bp, on the gels.

Both the Qiagen and TRIzoL methods yielded high quality total RNA. Although a higher yield was found using TRIzoL, the convenience and ease of the spin columns of the Qiagen kit made it the method of choice. All the reagents required for isolation are provided by the kit and are guaranteed to be RNase-free. It is a very rapid procedure, thereby reducing the chances of degradation by RNases. The only advantage the TRIzoL procedure has over the

Qiagen method is its flexibility. The method can be scaled up to accommodate larger amounts of starting material, whereas the spin columns are developed for a specific amount of material.

3.1.2 Isolation of mRNA

Approximately three-quarters of eukaryotic RNA consists of ribosomal RNA (rRNA) in the form of 28S, 18S and the 5S rRNA. Two other types of RNA molecules, transfer RNA (tRNA) and messenger RNA (mRNA), are also present in eukaryotic cells. mRNA is the least abundant, accounting for only 1-5% of the total. It is also the most heterogeneous in size.

Putative mRNAs isolated from eukaryotic cells differ from other RNAs in that they contains relatively long stretches of adenylic acid residues. The precise role of these poly(A)-rich regions in the metabolism of mRNA is not known, but Darnell *et al.* (1971) suggested that they are involved in the transport of mRNA from the cell nucleus to the cytoplasm where protein synthesis occurs.

These properties have made it possible to purify mRNA with one of the following techniques:

1) oligo(dT)-cellulose (Aviv and Leder 1972),
2) poly(U)-sepharose (Sheiness & Darnell 1973),
3) nitrocellulose filters (Darnell *et al.* 1971),
4) paper affinity chromatography (Werner *et al.* 1984),
5) oligo(dT) covalently linked to latex particles (Hara *et al.* 1991), or
6) oligo(dT)-coated magnetic beads (Hornes & Korsnes 1990; Jakobsen *et al.* 1990).

These methods are relying on base pairing between the poly(A)+ residues on the 3' -end of mRNAs and the oligo(dT) residues coupled to a solid support. Non-poly(A)+ RNAs are washed off, and bound mRNA is subsequently eluted with a low salt buffer. New commercially available kits and protocols even allow the direct isolation of poly(A)+ RNA from tissues and cells using an oligo(dT) cellulose matrix (Fastrack, Invitrogen) or oligo(dT)-coated magnetic beads (Dynal). Purification of poly(A)+ mRNA, using oligo(dT) probes attached to paramagnetic beads, is a rapid method not requiring centrifugation or filtration steps, thereby reducing the risk of physical or enzymatic degradation. In addition, it is possible to isolate mRNA (both polyadenylated and nonpolyadenylated) with benzoylated cellulose chromatography for the production of random hexamer-primed cDNA libraries (Van Ness *et al.* 1979).

In this study mRNA was purified from total RNA with oligo(dT)-coated magnetic beads (Dynabeads, Dynal). The procedure is flexible and may be scaled up or down as required. A strong RNase-inhibiting agent (LiCl) together with stringent hybridisation and washing conditions ensure the isolation of highly purified, intact mRNA. This method eliminates time-consuming precipitation or filtration and the use of hazardous organic reagents.

The concentration of mRNA was determined by measuring the absorbance at 260 nm. Typically the mRNA yield using Dynabeads was about 1% of the total RNA yield. This was to be expected, as mRNA makes up 1 to 5% of total RNA. A_{260}/A_{280} absorbance ratios of 2.0 were repeatedly detected, thus pure mRNA was obtained.

The quality of the isolated mRNA can be checked by gel electrophoresis. However, a rather large amount, about 1 μg, of mRNA is necessary to obtain a visible result. This was done only once and a faint smear was seen as expected (results not shown). The mRNA seemed to be pure, as no rRNA bands were present. This was, however, not conclusive as there was no way to tell whether the smear represented mRNA or degraded total RNA. Northern blot determination of mRNA quality might be more suitable for this purpose.

3.2 LIBRARY CONSTRUCTION

The utility of data generated in an EST study is greatly dependent upon the quality of the cDNA library. Ideally, a cDNA library dedicated to a study of this nature should be:

1) representative, i.e., containing all sequences present in the initial poly(A)+ population at the same relative frequencies;
2) unidirectionally cloned so that orientation of each cDNA is known to facilitate subsequent sequence analysis;
3) composed of a high proportion of long or full-length inserts;
4) uncontaminated with genomic, mitochondrial or ribosomal RNA inserts; and
5) composed of a large proportion of inserts with short poly(A) tails.

During the construction of a cDNA library, polyadenylated mRNA is converted to clonable double-stranded cDNA by using a series of enzyme-catalysed reactions. The oligonucleotides used to prime the synthesis of a complementary DNA with reverse transcriptase can be:

1) oligo(dT) primers which bind to the 3' poly(A) tail thereby potentially generating a full-length copy (Krug and Berger 1987),
2) random hexamer primers (Sargent 1987), or
3) an oligo(dT) linker ligated to a plasmid vector to create a "vector–primer" (Okayama & Berg 1982).

The yield of cDNAs containing the full coding sequence depends upon the quality of the mRNA and the degree of secondary structure present in mRNA transcripts that can interfere with complete extension of the primer. Random primers, usually DNA hexamers, ensure equal representation of all mRNA sequences in the cDNA library and increase the chances of obtaining coding and 5'-untranslated sequences. However, random primers allow neither unidirectional cloning nor synthesis of full-length cDNA copies.

Earlier methods used the transient hairpin loop at the 3' end of the newly synthesised first strand as a primer to generate the second strand after the RNA fragments are removed with an alkaline buffer (Land *et al.* 1981). This method suffers from significant loss of coding information corresponding to the 5' end of the mRNA. The method used by Okayama & Berg (1982) provided a major improvement in the quality of cDNA libraries. They employed a "vector–primer" with an oligo(dT) at one end to prime poly(A)+ mRNA for first strand synthesis. The vector is digested with a restriction enzyme to form blunt ends to which poly(A) tails are added by means of terminal deoxynucleotidyl transferase. To produce the second strand, they used RNase H and DNA polymerase I mediated nick translation. Gubler & Hoffman (1983) adapted the second strand synthesis method used by Okayama & Berg into their oligo(dT) primed procedure.

There are two general approaches for unidirectional cDNA cloning. In one poly(A)+ RNA is primed from an adaptor primer which generates a unique restriction site at one end of the cDNA. After second-strand synthesis, adaptor oligonucleotides with a cohesive end different from the restriction site within the adaptor primer are ligated onto both ends of the cDNA (Han *et al.* 1987; Meissner *et al.* 1987). The second established method for unidirectional cDNA synthesis is to use the "vector–primer" (Okayama & Berg 1982; Okayama *et al.* 1987).

Although multiple types of vector may be used to construct a cDNA library, most of the large-scale cDNA sequencing studies are based on λ or phagemid libraries. Most of the current automated sequencing technology has been developed using M13 priming sites. Any vector, phage or plasmid, with these sites and the ability to be amplified in prokaryotes should be suitable for this purpose. It is important to have a highly representative library and, therefore, the cloning efficiency of the vector system chosen is of central importance.

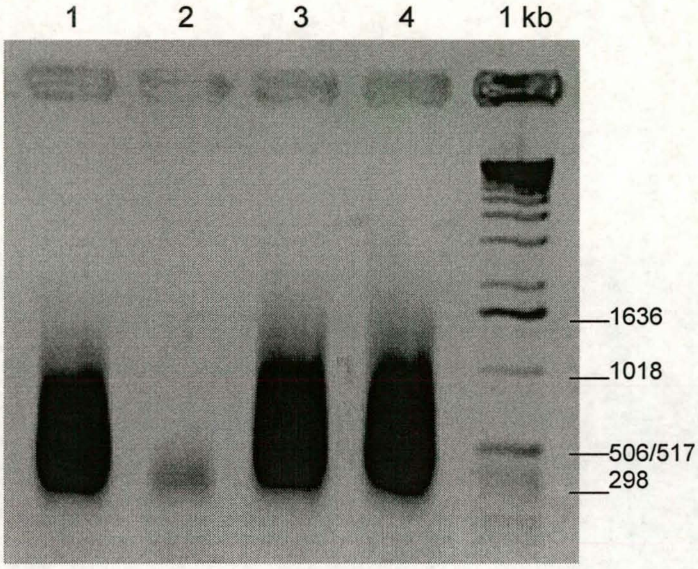
3.2.1 cDNA synthesis

In this study cDNA was synthesised and amplified according to the procedure described by Jepson *et al.* (1991), based on the method of Okayama & Berg (1982). The first strand was synthesised using a Not I -oligo(dT) primer. Second-strand synthesis was done by employing RNase H, DNA polymerase I and DNA ligase. Adaptors, containing Eco RI cutting sites, were ligated to both ends of the double-stranded cDNA. The adaptor sequences served as priming sites for the amplification of the cDNA.

Figure 3.2 shows the PCR products produced. Smears of fragments seen in lanes 1, 3 and 4 have sizes ranging from 300 - 1500 bp and are compatible with the average size of 600 bp previously found in two maize libraries by Shen *et al.* (1994). Approximately 200 ng/μl of cDNA were routinely amplified per reaction and is compatible with results reported by Jepson *et al.* (1991) for a library constructed from *Beta vulgaris*. The poor yield in lane 2 could be due to insufficient amplification, a problem in either first- or second-strand synthesis, the adaptor ligation, or the mRNA could have been degraded.

To reduce the likelihood of smaller cDNA sequences being amplified preferentially, careful consideration was paid to the choice of extension time at the PCR stage. According to Domec *et al.* (1990), a three minute extension time largely overcomes the problem of molecule size heterogeneity. In addition, size fractionation was used to eliminate very small (<194) cDNAs. This protocol provides a simple and rapid procedure to generate cDNA and can also be used for differential and subtractive screening. This procedure is straightforward when compared to other systems, in that precipitation steps are kept to a minimum.

Figure 3.2 Amplified double-stranded cDNA run on a 1% agarose gel and stained with ethidium bromide. Lanes 1 to 4 represent PCR products of cDNA synthesised on different occasions. A smear of fragments ranging in size from 300 bp to 1500 bp was found. The 1 kb ladder was used as a size marker.



3.2.2 Cloning of cDNA fragments

The amplified cDNA was digested with both Eco RI and Not I, and cloned into a plasmid vector, pBluescript® SK+. It was found that the most successful vector to insert ratio was 1:4, using about 100 ng of cDNA. Although much effort was made to optimise the cloning procedure, a large percentage of non-recombinants were still found.

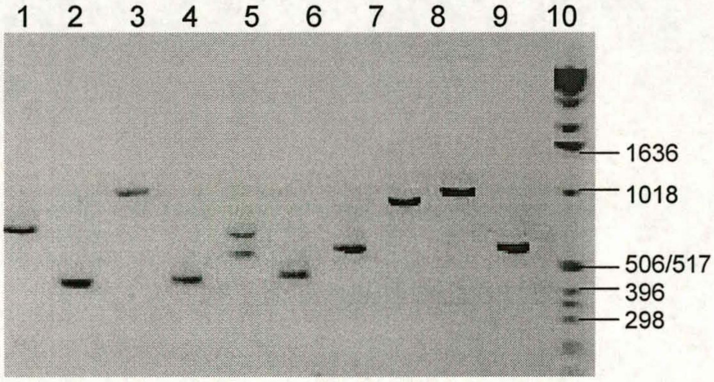
For a library to be fully representative one may need $>10^6$ clones (Kimmel & Berger, 1987). Phage vectors are often the choice for large libraries of >100 000 recombinants whereas libraries of 10 000 to 50 000 recombinants are often prepared in plasmid vectors. However, more versatile cloning strategies are available using plasmids as vectors, and plasmid libraries offer additional possibilities such as screening for expression in eukaryotic cells (Kieffer, 1991). About 3000 clones were produced in this study. The recombinants were conveniently stored as liquid cultures in 96–well plates at –80°C. Although this is not a completely representative library, it was adequate for the development of markers. Frozen cDNA samples are available for cloning and transformation when more clones are needed for future applications. Two reported plasmid cDNA libraries of potato in the literature were relatively small as well. Logemann *et al.* (1988) reported 4 000 recombinants and Taylor *et al.* (1990) constructed a library of 20 000 clones. However, larger libraries using plasmid vectors have been constructed in other crops. For example, by using techniques such as electroporation, a *Brassica napus* plasmid library with 1,5 x 10^5 recombinants has been obtained by Park *et al.* (1993).

3.3 CHARACTERISATION OF cDNA INSERTS

3.3.1 Presence and size of inserts

Randomly selected recombinant plasmids were amplified by means of PCR, using M13 forward and reverse oligonucleotide primers. Colonies were picked randomly and added directly to the PCR reaction mixture. Much time was saved using this method instead of preparing liquid cultures and performing plasmid extractions. Figure 3.3 is a typical example of PCR products obtained. The size range varied considerably, typically from 200 bp to 1600 bp.

Figure 3.3 Inserts amplified with M13 forward and reverse primers by means of PCR. On this gel the sizes of the inserts range from about 400 bp to 1000 bp. The presence of two products in lane 5 could be due to the presence of two individual plasmids carrying different cDNA inserts or a single plasmid containing two cDNA inserts. M = 1 kb ladder, sizes in bp.



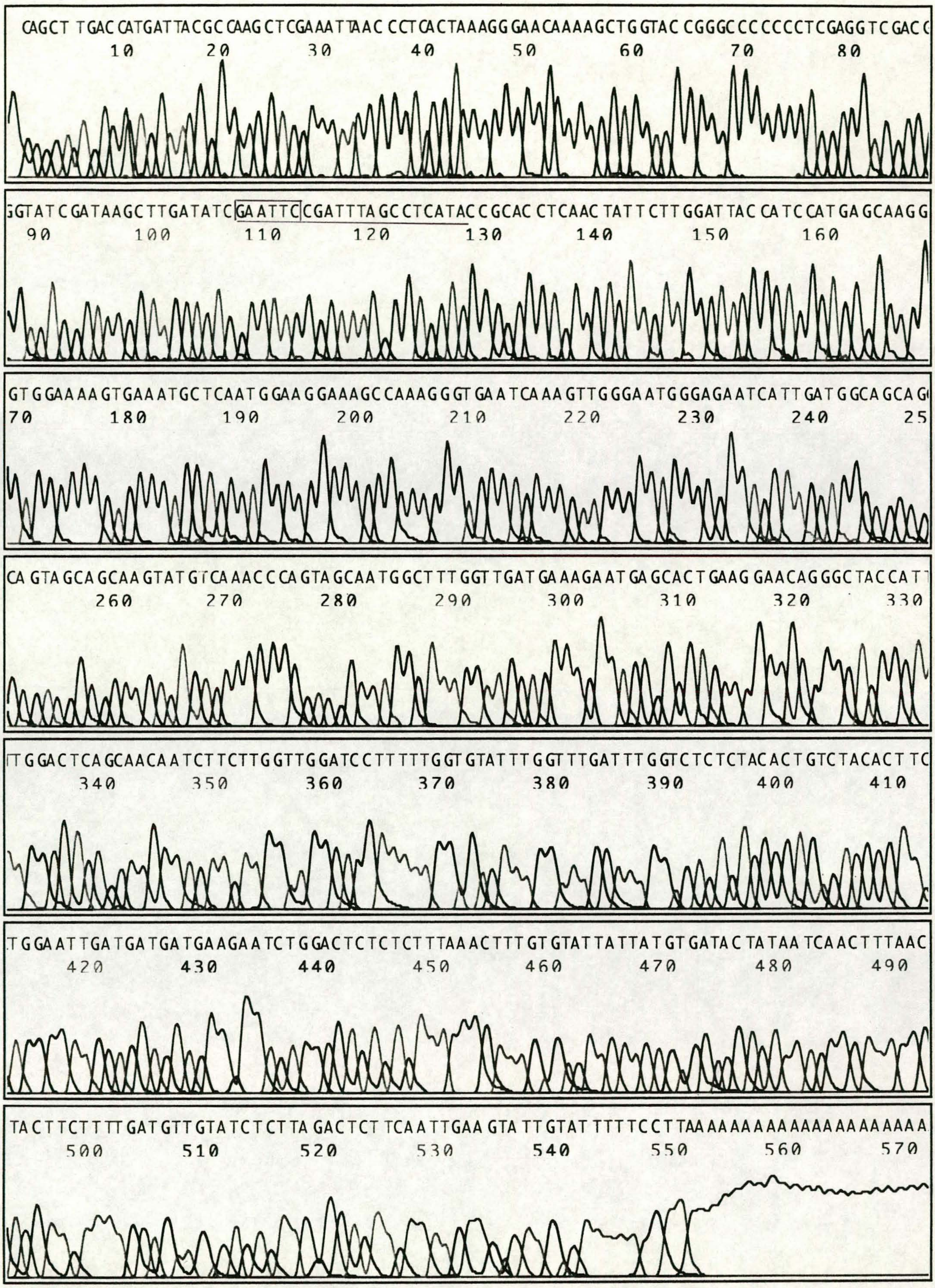
Of 250 clones amplified, 50 percent yielded no or small (<250 bp) inserts, 22 percent of the inserts were in the size range 250 to 500 bp and 24 percent were between 500 and 1000 bp. Only four percent of the inserts were larger than 1 kb. Lim *et al.* (1996) reported insert sizes ranging from 0.5 kb to 4 kb from a Chinese cabbage library and Kwak *et al.* (1997) found the average insert size of their *Brassica campestris* library to be 0.85 kb. Although the size range of the present library was comparable to the results of these studies, the number of large inserts (>1 kb) was much smaller.

The large number of clones with no or small inserts were unexpected and disappointing, as one would expect that all the clones picked would have an insert. Blue/white selection was used to identify recombinant clones. After transformation clones could be placed into three groups according to colour: blue, white and different shades of light blue. To examine the effectiveness of the selection method, colonies of the different colours were picked and analysed (results not shown). As was anticipated, the blue colonies contained no inserts and most of the white colonies did contain inserts, as was also expected. However, the light blue colonies could not be placed in either of the two categories. Some contained inserts; others did not. As it was difficult to distinguish between white and light blue colonies, clones without inserts could have been mistaken for recombinants.

3.3.2 DNA Sequencing

Clones with cDNA insert sizes larger than 300 bp were selected for sequencing. A total of 50 sequences were determined by automated sequencing. Two clones were excluded from further analysis due to poor quality sequencing, which could have been the result of impure preparations. Figure 3.4 depicts a typical cDNA insert sequenced from the 3' end with M13 reverse primer. The first part of the sequence represents the vector plus the EcoRI cutting site (GAATTC) at base 108 (boxed), followed by the adaptor sequence (CGATTTAGCCTCATA) (underlined). The rest of the sequence portrays the cDNA insert up to the poly(A), starting at base 551.

Figure 3.4 Clone 6e3 sequenced with the M13 reverse primer.



3.3.2.i Insert orientation and presence of poly(A) tails

To confirm the directionality of this library, clones were sequenced from both ends. Of the 48 clones, 43 had poly(A) tails. Eight inserts were inverted, thus this library is 83% unidirectional. According to Burglin & Barnes (1992) potential contamination of a cDNA library with genomic sequences is a matter of concern, but the presence of a poly(A) tail on most sequences confirms that this library is essentially free of genomic sequences.

Internal restriction enzyme cutting sites were responsible for the lack of poly(A) tails, as well as inverse orientation of inserts. To protect the cDNA from restriction enzymes, some protocols use 5' -methyl dCTP instead of dCTP in the nucleotide mixture for the first strand synthesis. The presence of 5' -methyl dCTP will create hemimethylated cDNA. This is not feasible for PCR-based libraries. For all the inserts to be hemimethylated, 5' -methyl dCTP should be used in the PCR reaction mixture. However, the cutting sites created at the ends of the inserts will also become methylated. Other protocols protect EcoRI internal sites by treating the double-stranded DNA with EcoRI methylase (Wu *et al.* 1987). Still, this will not solve the problem of internal cutting sites for this library entirely, as both EcoRI and NotI are used to clone the cDNA. Although Not I is a rare cutter, two clones lacking poly(A) tails contained internal Not I cutting sites.

Open reading frames were detected for all the ESTs, using the ORF Finder program. Although this is not conclusive evidence that full-length cDNA was synthesised, the presence of open reading frames together with the presence of poly(A) tails suggest that most of the cDNAs might be full-length.

3.3.2.ii Redundancy

Out of 48 ESTs, 38 were unique or nonredundant cDNA clones. Thus, 21% redundancy was found in this library. These statistics might not be significant, since only a small number of clones were sequenced. However, the redundancy observed in this library is comparable to the 25% redundancy found by Lim *et al.* (1996) in a Chinese cabbage cDNA library. Redundant clones could be transcripts of the same gene or cognate genes.

The use of differentially selected clones (Hoog *et al.* 1991) as opposed to randomly selected clones, as well as the preparation of 'sorted' libraries (Waterston *et al.* 1992), has been shown to reduce insert duplication. In addition, the type and developmental stage of the tissue used for mRNA isolation will also affect the redundancy. For example, Nelson *et al.* (1984) found that a maize library constructed from expanding leaves, will have a relatively high percentage of abundant mRNAs. Keith *et al.* (1993) chose to construct their library from mature maize leaves and found no duplicates in the 130 clones characterised. For the present library young leaves from a growing potato plant were collected and this could have contributed to the high level of redundancy.

The elimination of these uninformative clones is a priority for constructing ideal cDNA libraries. Techniques to reduce repeated sequencing of clones would become increasingly important as large numbers of cDNAs are sequenced. Subtraction, which preferentially reduces the population of certain sequences in a library, and normalisation, which results in all sequences being represented in approximately equal numbers in a library, should reduce repeated sequencing of abundantly expressed sequences. This should subsequently maximise the chance of identifying rare transcripts from specific cell populations (Cooke *et al.* 1996). However, techniques such as normalisation or subtraction were not considered for this study as they are technically more demanding.

3.3.2.iii Long Poly(A) tails cause problems in automated sequencing

Cloning of poly(A) tails is a consequence of oligo(dT) priming. Although long poly(A) tails do not ordinarily present a problem in conventional cloning applications, they have proven to be a difficulty in automated sequencing. It is difficult to sequence a length of poly(A) exceeding 30 to 40 bases, because the nucleotides in the reaction mix get depleted. This creates an imbalance, introducing errors in the adjacent sequence (Figure 3.5). Sequencing from the 3' end of the cDNA did not create serious problems when the insert was in the correct orientation, as the poly(A) tail is only at the end of the sequence (Figure 3.5 (a)). However, it did make it virtually impossible to sequence the cDNA from the 5' end, because the tail was present at the 5 prime end (Figure 3.5 (b)). As a consequence, most sequences could only be determined from one end and therefore could not be aligned and checked for errors.

A library containing a high proportion of short poly(A) tails can be achieved by using an excess of oligo(dT) primer during the first strand cDNA synthesis, thereby saturating the poly(A) tails of the mRNAs (Moreno-Palanques & Fuldner 1994). Since the reverse transcriptase enzyme cannot strand-displace (Kornberg & Baker 1992), the primer located at the most 5' -position of the poly(A) tail will serve as the origin of the cDNA, resulting in clones containing short poly(A) tails. Extension of any other primer is limited by its distance to the next downstream primer, thus generating very small fragments that can be easily eliminated by an efficient size-selection procedure (Soares 1994).

Figure 3.5 (a) A poly(A) tail causing errors in the adjacent sequence. Clone 7a5 sequenced with M13 reverse primer.

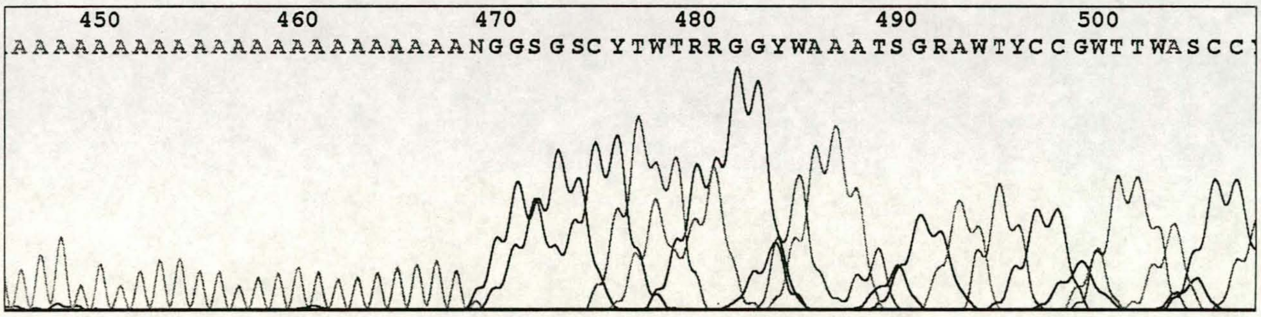
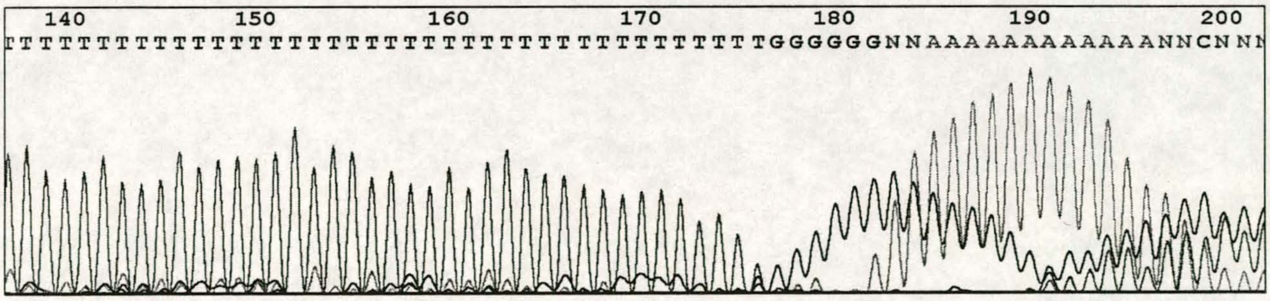


Figure 3.5 (b) A poly(T) tail causing errors in the adjacent sequence. Clone 6b11 sequenced with M13 forward primer.



3.3.2.iv Sequencing errors

Single-strand sequencing is not particularly accurate. Although it is not possible to determine the exact rate for a specific single-strand sequence without having the double-strand sequence as comparison, accuracy rate estimates of 98% (Adams *et al.* 1991) and 99.5% (Hoog 1991) have been made for entire projects. In their study, Keith *et al.* (1993) estimated the project error rate by comparing sequences obtained to sequence data from cDNA that had previously been determined, as well as data from four clones that did not contain inserts. For a total of 1897 bases they found five deletions, three mismatches, and four ambiguous base calls. This calculates to an error rate of less than 1%. According to States & Botstein (1991) an error rate of this magnitude poses no problems with the analysis of DNA similarities and should not impede amino acid alignments. PE Biosystems, supplier of the ABI 377 Sequencer and chemistries used in this study, reported a 98.5% base-calling accuracy over 800 bases on sequencing of Long Read DNA Cycle Sequencing Standard with the M13 forward primer (href11). The accuracy rate for this study is estimated to be similar. This is based on the calculated accuracy rate for the sequencing of standards in the sequencing facility used for this study (personal communication with C.J. van Heerden), the alignment of sequences determined in both directions (see Figure 3.7, 10e3), and database matches with a 100% similarity (Table 3.1, clones 2g4, 6g8 and 7f7).

The type of errors found in the EST sequences can affect the database matches. Most sequence errors are miscallings or ambiguous base calls, which did not affect the level of database matching. A miscalled base is simply regarded as a change of a single amino acid residue, or it might even not change the amino acid residue, depending on the location of the miscalled base and on the frame used to search the database. However, deletions or insertions may cause a problem, because these errors can cause frame shifts in the translated amino acid.

In this study, two clones which were found to be identical, were aligned using the program Sequence Navigator (PE Applied Biosystems) (Figure 3.6). Excluding the first 10 bases, where the sequences are compressed and errors are bound to occur, no differences were detected. This demonstrates that basecalling was consistent. Nevertheless, for the short sequence data generated by partial sequencing, frameshift errors are a concern. Special care has to be taken when STS primers are to be designed. A good practice would be to check the chromatograms for possible miscallings and eliminate areas that contain ambiguities.

Figure 3.6 The alignment of the sequences of clones 10f12 and 10h2.

```
                10         20         30         40         50         60         70         80         90        100
10f12   GAT-CGCCAA GCTCGAAATT AACCCTCACT AAAGGGAACA AAAGCTGGTA CCGGGCCCCC CCTCGAGGTC GACGGTATCG ATAAGCTTGA TATCGAATTC
10h2    GAAACGCCAA GCTCGAAATT AACCCTCACT AAAGGGAACA AAAGCTGGTA CCGGGCCCCC CCTCGAGGTC GACGGTATCG ATAAGCTTGA TATCGAATTC
        --**------ ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

               110        120        130        140        150        160        170        180        190        200
10f12   CGATTTAGCC TCATAGGCAA ATTATGGCTT CTACTTGTGC AGTTTCCATG GCAATGCCAC TGACTTACAC AAGTCACAAG AGACAACCAA GAGTTGAGGC
10h2    CGATTTAGCC TCATAGGCAA ATTATGGCTT CTACTTGTGC AGTTTCCATG GCAATGCCAC TGACTTACAC AAGTCACAAG AGACAACCAA GAGTTGAGGC
        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

               210        220        230        240        250        260        270        280        290        300
10f12   TTTCTTGAAG CCATTGCCAT TAAGGCCCTC AAGGACTACT TCAGCATCAA AACCAGTAGC AAAGTTTCAA GTGAATGCTT CCCTTAAGGA GAAAGCTTTG
10h2    TTTCTTGAAG CCATTGCCAT TAAGGCCCTC AAGGACTACT TCAGCATCAA AACCAGTAGC AAAGTTTCAA GTGAATGCTT CCCTTAAGGA GAAAGCTTTG
        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

               310        320        330        340        350        360        370        380        390        400
10f12   ACAGGATTGA CAGCAGCTGC ACTCACTGCT TCCATGGTCA TTCCCGACGT AGCCGAAGCA GCTGAGGGTG TTTCACCGTC CCTCAAGAAC TTTTTGCTCA
10h2    ACAGGATTGA CAGCAGCTGC ACTCACTGCT TCCATGGTCA TTCCCGACGT AGCCGAAGCA GCTGAGGGTG TTTCACCGTC CCTCAAGAAC TTTTTGCTCA
        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

               410        420        430        440        450        460        470        480        490        500
10f12   GCATTTCTGC AGGTGGAGTT GTGCTTGCTG CAATTCTTGG TGCTATAATT GGTGTTTCCA ATTTTGATCC TGTCAAACGG ACTTAACAAG CATTATTCTC
10h2    GCATTTCTGC AGGTGGAGTT GTGCTTGCTG CAATTCTTGG TGCTATAATT GGTGTTTCCA ATTTTGATCC TGTCAAACGG ACTTAACAAG CATTATTCTC
        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

               510        520        530        540        550        560        570        580        590
10f12   AACTCTCTTC TTGNATCTAT CTCATGTTTC CTGTCAAGAT CAATAAAATG TTTACATAAC TATCATTTCT TAAAAAAAAA AAAAAAGCGG
10h2    AACTCTCTTC TTGNATCTAT CTCATGTTTC CTGTCAAGAT CAATAAAATG TTTACATAAC TATCATTTCT TAAAAAAAAA AAAAAAGCGG
        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

3.3.2.v Chimerism

Chimeric clones often result from blunt-end ligation of cDNA molecules during the reaction in which adaptors are ligated to the cDNAs. To prevent formation of these cloning artifacts, adaptor molecules must be present in vast excess over cDNAs in the ligation reaction. Such conditions can be easily satisfied only if the cDNAs are efficiently size-selected prior to ligation. An excess of adaptors was used in this study and there was no evidence that chimeric clones were present.

Some precautions are necessary to avoid non-specific priming at GC-rich regions of the mRNAs when using large amounts of the Not I-$(dT)_{18}$ primer for first strand cDNA synthesis. Most importantly, the reaction mixture should be preincubated at 37°C before the addition of reverse transcriptase. Soares (1994) observed that if the enzyme was added to the reaction mixture at room temperature, an appreciable number of clones lacking a tail can be formed. Presumably, if a GC-rich cluster is flanked by a few A's located upstream on the mRNA, the Not I sequence of the primer (GCGGCCGC) can anneal to it while most of the oligo(dT) tail loops out. Short first-strand cDNA fragments are formed as a result.

3.3.3 EST analysis

Forty-eight random ESTs were generated with an average length of 450 bp. Sequences were examined for similarities to sequences in databases using the BLAST 2.0 programs (Altschul *et al.* 1997). BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore available sequence databases regardless of whether the query is protein or DNA. BLASTX compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against protein sequence databases. BLASTN was used to compare a nucleotide query sequence against nucleotide sequence databases.

The BLAST algorithm will find all segment pairs with scores that distingiush real matches from random backround hits. These are called high-scoring segment pairs or HSPs. Several statistics are used to describe each HSP:

- the raw alignment score;
- the raw score converted to bits of information by multiplying with Lambda ('bit' score);
- the number of times one might expect to see such a match, or a better one, merely by chance (E-value);
- the Poisson *P*-value (probability in the range of zero to one) of observing such a match; and
- the number and fraction of total residues in the HSP which are identical (Identities).

A protein match is considered to be statistically significant with a Poison *P*-value <0.1 (Altschul *et al.* 1994). BLAST 2.0 reports E-values rather than P-values because it is easier to understand the difference between, for example, an E-value of five and 10 than P-values of 0.993 and 0.99995. However, when E<0.01, P-values and E-value are nearly identical. (Altschul 1999) (href12).

In this study 32 out of 48 clones (66.7%) showed significant similarity to entries from the protein public databases. The results of the amino acid searches using BLASTX are summarised in Table 3.1. This percentage is comparable to that observed in similar studies of pea (72% of 39 ESTs; Gilpin *et al.* 1997) and black spruce (78% of 40 ESTs; Perry & Bousquet 1998a). It is much higher than similarities observed in major EST projects where numerous partial sequences are produced. For example, 31% of 1517 ESTs from *Caenorhabditis elegans* (Waterston *et al.* 1992), 32% of 1152 *A. thaliana* ESTs (Höfte *et al.* 1993), 8% of 830 rice ESTs (Uchimiya *et al.* 1992) and 17% of 2375 ESTs from human (Adams *et al.* 1991, 1992) showed significant matches with sequences in databases.

Several factors might explain the apparent differences between these studies. Most of the ESTs produced in this study match proteins which are expressed abundantly in leaves, for example the chlorophyll binding proteins. Nelson *et al.* (1984) also found a relatively high percentage of abundant mRNAs, analysing a maize cDNA library, constructed from expanding leaves. Shen *et al.* (1994) reported that 88% of ESTs produced from maize leaves showed high similarity to proteins in the abundantly expressed group. Similarly in mice, 90% of sequences detecting a significant similarity are abundantly expressed (Hoog 1991). This can be explained by the fact that abundantly expressed genes are generally easier to isolate and more likely to have been studied previously. Therefore one might expect a decrease in the fraction of identified sequences once all of the abundantly expressed group of sequences have been identified. For example, in 1993 Höfte *et al.*

reported that 32% of *Arabidopsis* EST have by that time been described, while Cooke *et al.* (1996), studying the same organism, found that only 15% of sequences correspond to previously characterised genes. Another factor could simply be the standard for declaring a "match". Standards adopted by each group vary as no universal values exist. A match could be declared either as the accurance of the homologous gene requiring very high scores, or as a much lower score indicating a similarity of function.

Only five (a12, 7a5, 17a6, 17b2 and 18c12) of the 16 sequences that did not share homology with sequences in the protein databases shared significant sequence similarity to known nucleotide sequences in the databases. The significant matches found with BLASTN are listed in Table 3.2. Most of the matches from the two searches were similar, but some differences did occur. To avoid confusion, nucleotide matches were only taken into consideration when a protein similarity was not found, as in these five cases. Among the putatively identified ESTs, 9 (24%) matched genes of potato, 13 (35%) were similar to genes of other Solanaceae, and 14 (38%) showed similarity to other plant genes. Only one clone matched a non-plant protein.

Statistical significance of sequence similarities does not necessarily imply functional similarity. Some of the matches reported may indicate the presence of a conserved domain or motif, or simply a common protein structure pattern. Precautions were taken to eliminate non-significant similarities and those that are cited should still be considered with some caution. They are only indications of the possible function of the corresponding protein, as putative functions are based on partial sequence data. For instance, Cooke *et al.* (1996) noted that a previously unknown clone showed significant similarity to *Umbrella californica* C12:0 thioesterase (Grellet *et al.* 1993) of which the corresponding recombinant protein sequence had been previously reported (Voelker *et al.* 1992). However, when a full-length clone was isolated using the EST an over-expression of the corresponding recombinant protein in *E. coli* showed that the gene encodes a thioesterase having specificity in the range C14-C18 with a marked preference for C16 (Doermann *et al.* 1995). These results and others emphasise that only biochemical and genetic approaches can demonstrate the true function when a putative function is deduced from sequence similarity.

Some clones matched the same database entry. For example, 17a3 and 17a9 both showed homology to cytochrome P-450 in the protein (Table 3.1) and nucleotide (Table 3.2) searches. After further investigation their sequences were found to be similar as was to be expected.

Table 3.1 Potato ESTs with similarity to known genes in international protein databases using the BLAST program. The best match of each EST was selected.

Clone[1]	Length (bp)[2]	Putative identification/ function[3]	Organism[4]	Score (bits)	E-value	Identities[5]	Accession no.[6]
a6	548	Chlorophyll a/b binding protein 3C precursor	Tomato	298	2e-80	144 (99%)	sp P07369
2g4	324	Chlorophyll a/b binding protein LHCII type 1 precursor	Tomato	153	5e-37	73 (100%)	sp P14274
3c10	393	Rubisco small subunit 2C	Potato	181	2e-45	88 (98%)	sp P26577
3d11	473	Photosystem II protein X precursor	Arabidopsis thaliana	84	6e-16	120 (45%)	gb AAD25151.1
3h7	381	Translation initiation factor GOS2	Maize	83	8e-16	41 (97%)	sp P56330
6b10	495	Light inducible tissue-specific ST-LS1 protein	Potato	233	7e-61	131 (87%)	sp P06183
6b11	497	Photosystem II reaction centre W protein precursor	Spinach	73	2e-26	38 (84%)	sp Q41387
6d5	233	Putative photosystem II 5 kd protein	Arabidopsis thaliana	51	1e-06	30 (73%)	gb AAD30649.1
6d6	600	Plastocyanin precursor	Tomato	193	1e-48	169 (62%)	sp P17340
6d7	506	Light inducible tissue-specific ST-LS2 protein	Potato	232	2e-60	131 (87%)	emb CAA27989
6e3	434	Photosystem II reaction centre W protein precursor	Spinach	128	2e-29	106 (64%)	sp Q41387
6f10	317	Hypothetical protein	Arabidopsis thaliana	48	2e-05	26 (69%)	gi 4115357
6f12	527	Conserved protein	Methanobacterium thermoautotrophicum	41	0.006	40 (42%)	gi 2622823
6g8	257	Chlorophyll binding protein 1C	Tomato	106	5e-23	49 (100%)	prf 1204205C
6h9	577	Plastidic aldolase NPALDP1	Nicotiana paniculata	365	1e-100	190 (93%)	dbj BAA776041
7b4	415	Unknown protein	Arabidopsis thaliana	94	5e-19	98 (52%)	gi 3128175
7e9	617	Ubiquitin extension protein	Potato	215	4e-63	133 (83%)	emb CAA71132
7f7	419	Carboxylase RBP	Nicotiana sylvestris	176	7e-44	79 (100%)	gi 170320
9g5	241	Putative photosystem II 5 kd protein	Arabidopsis thaliana	66	6e-11	41 (73%)	gb AAD30649.1
10e3	243	Rubisco small subunit 24	Potato	168	9e-42	80 (100%)	sp P26575
10f12	484	Photosystem II protein X precursor	Arabidopsis thaliana	86	2e-16	120 (45%)	gb AAD25151.1
10g2	476	Photosystem II protein X precursor	Arabidopsis thaliana	86	2e-16	120 (45%)	gb AAD25151.1
10g9	725	Photosystem I reaction centre subunit XI precursor	Barley	132	7e-45	107 (66%)	sp P23993
10h2	650	Rubisco small subunit 1	Potato	325	2e-88	157 (95%)	sp P26574
10h3	601	Thioredoxin-m	Brassica napus	151	2e-37	108 (62%)	gi 1943720
17a3	372	Putative cytochrome P-450	Nicotiana plumbaginifolia	48	3e-10	32 (65%)	gb AAB05376.3
17a9	367	Putative cytochrome P-450	Nicotiana plumbaginifolia	115	2e-25	64 (78%)	gb AAB05376.3
17b10	396	Chlorophyll a/b binding protein CP29	Vigna radiata	160	5e-39	79 (94%)	gb AAD27878
18c7	524	Chlorophyll a/b binding protein type 1 precursor	Tomato	249	9e-66	140 (87%)	pir S16294
18c9	695	Initiation factor 5A-4	Potato	115	4e-25	117 (64%)	sp P56336
19h10	735	Putative protein	Arabidopsis thaliana	298	3e-80	187 (74%)	emb CAB36531.1
20d11	674	Plastidic aldolase	Nicotiana paniculata	270	2e-75	136 (95%)	dbj BAA77603.1

(1) Clone name, (2) the length of the sequence used to search the database, (3) the protein for which similarity has been detected, (4) the organism for which the similarity was found, (5) the length and percent identity of the match, and (6) the database accession number of the clone for which similarity was found. The length listed in the table is not necessarily the length of the cDNA insert, but the length in bp of good quality sequence that could be used for the searches. The lengths of the match listed in Identities are given in amino acids.

Table 3.2 Potato ESTs identified by similarity to sequences in international nucleotide databases using the BLASTN program. The best match of each EST was selected.

Clone[1]	Putative identification/ function[2]	Organism[3]	Score (bits)	E-value	Identities[5]	Accession no.[6]
a6	Chlorophyll a/b binding protein (Lhcb1-2) gene	Potato	835	0.0	433 (99%)	gb U20983
a12	MAL3P1	Plasmodium falciparum	46.1	0.003	27 (96%)	emb Z97348.1
2g4	Chlorophyll a/b binding protein (Lhcb1-1) gene	Potato	446	1e-124	224 (100%)	gb U21111
3c10	RbcS2c gene for ribulose-1,5-bisphosphate carboxylase/oxygenase	Potato	351	6e-95	177 (100%)	emb X69762
3h7	Translation initiation factor (GOS2) mRNA	Rice	97.6	1e-18	105 (86%)	gb AF094774
6b10	mRNA for light inducible tissue-specific ST-LS1 gene	Potato	848	0.0	468 (98%)	emb X04401
6b11	mRNA for 6.1 kDa polypeptide of photosystem II	Solanum oleracea	101	1e-19	163 (83%)	emb X85038
6d6	mRNA for pre-plastocyanin	Tomato	803	0.0	505 (95%)	emb X13934
6d7	mRNA for light inducible tissue-specific ST-LS1 gene	Potato	856	0.0	492 (97%)	emb X0401
6e3	mRNA for 6.1 kDa polypeptide of photosystem II	Solanum oleracea	125	7e-27	163 (84%)	emb X85038
6f10	Aryl sulfotransferase IV gene	Rattus norvegicus	46.2	0.004	23 (100%)	gb L16241
6g8	Chlorophyll a/b binding protein (Lhcb1-1) gene	Potato	289	1e-76	150 (99%)	gb U21111
6h9	Aldolase gene	Pisum sativum	234	1e-59	250 (86%)	gb M97476
7a5	Rubisco activase mRNA	L. pennellii	383	1e-104	229 (96%)	gb AF037361
7e9	mRNA for ubiquitin extension protein	Potato	764	0.0	503 (94%)	emb Y10024
7f7	rbcS2 gene for ribulose-1,5-bisphosphate carboxylase/oxygenase	Potato	551	1e-155	286 (99%)	emb X69760
10e3	RuBP carboxylase small subunit	Tomato	297	6e-79	242 (90%)	gb M15236
10g9	mRNA for PsaL	Cucumber	117	3e-24	175 (83%)	dbj D50456
10h2	RuBP carboxylase small subunit	Tomato	870	0.0	606 (93%)	gb M15236
10h3	Thioredoxin m gene	Pisum sativum	46.4	0.007	35 (91%)	gb U35831
17a3	Putative cytochrome P-450 mRNA	Nicotiana plumbaginifolia	115	5e-24	182 (84%)	gb U35226.3
17a6	16s rRNA gene	Nicotiana plumbaginifolia	431	e-119	308 (95%)	gb M82900
17a9	Putative cytochrome P-450 mRNA	Nicotiana plumbaginifolia	157	2e-36	195 (85%)	gb U35226.3
17b2	Sulfite reductase	Tobacco	83.8	2e-14	76 (89%)	dbj AB010717
17b10	Chlorophyll a/b binding protein CP29	Vigna radiata	157	2e-36	171 (86%)	gb AF139466.1
18c7	Cab9 gene for type I (26 kD) CP29 polypeptide	Tomato	524	e-147	344 (95%)	emb X61287
18c9	mRNA for eukaryotic initiation factor 5A4	Potato	805	0.0	667 (92%)	dbj AB004825
18c12	RbcS2 gene for ribulose-(1,5)-bisphosphate carboxylase/oxygenase	Potato	189	1e-46	131 (95%)	emb X69760
19h10	Chromosome 4, BAC clone F10M23	Arabidopsis thaliana	70	8e-10	67 (88%)	emb AL035440.2
20d11	mRNA for protein homologous to plastidic aldolase	Potato	902	0.0	584 (95%)	emb Y10380

(1) Clone name, (2) the length of the sequence used to search the database, (3) the protein for which similarity has been detected, (4) the organism for which the similarity was found, (5) the length and percent identity of the match, and (6) the database accession number of the clone for which similarity was found. Match lengths are given in nucleotides.

Two clones matching the same protein in a database do not necessarily indicate duplication of the same clone. For example, clone 10e3 and 10h2 both significantly matched a Rubisco small subunit from potato in the protein search. In the nucleotide search they matched the same database entry (accession number gb M15236), a Rubisco small subunit from tomato. However, alignment of the two sequences revealed that this is not a duplication of the same cDNA insert (Figure 3.7). Parts of the sequence are similar, but they are not identical. The chromatograms of some differences, highlighted in the alignment, are shown in Figure 3.8 and verify that the discrepancies are not a result of poor quality sequencing or miscalling. Similarly, PCR artifacts could not account for so many discrepancies.

Fourteen ESTs did not match any current database entries and therefore represent new, uncharacterised genes. However, with the current rapid rate of increase in submissions to the databases, it is likely that the numbers of identifiable genes will also increase. It was found that routine resubmission of "unknown" sequences usually results in more identifications due to new additions to the databases in the interim. Development of a better computer algorithm for identifying distantly related genes showing weak amino acid identity but containing structurally or evolutionary related sequences, will also increase the level of identification of EST sequences (Park *et al.* 1993).

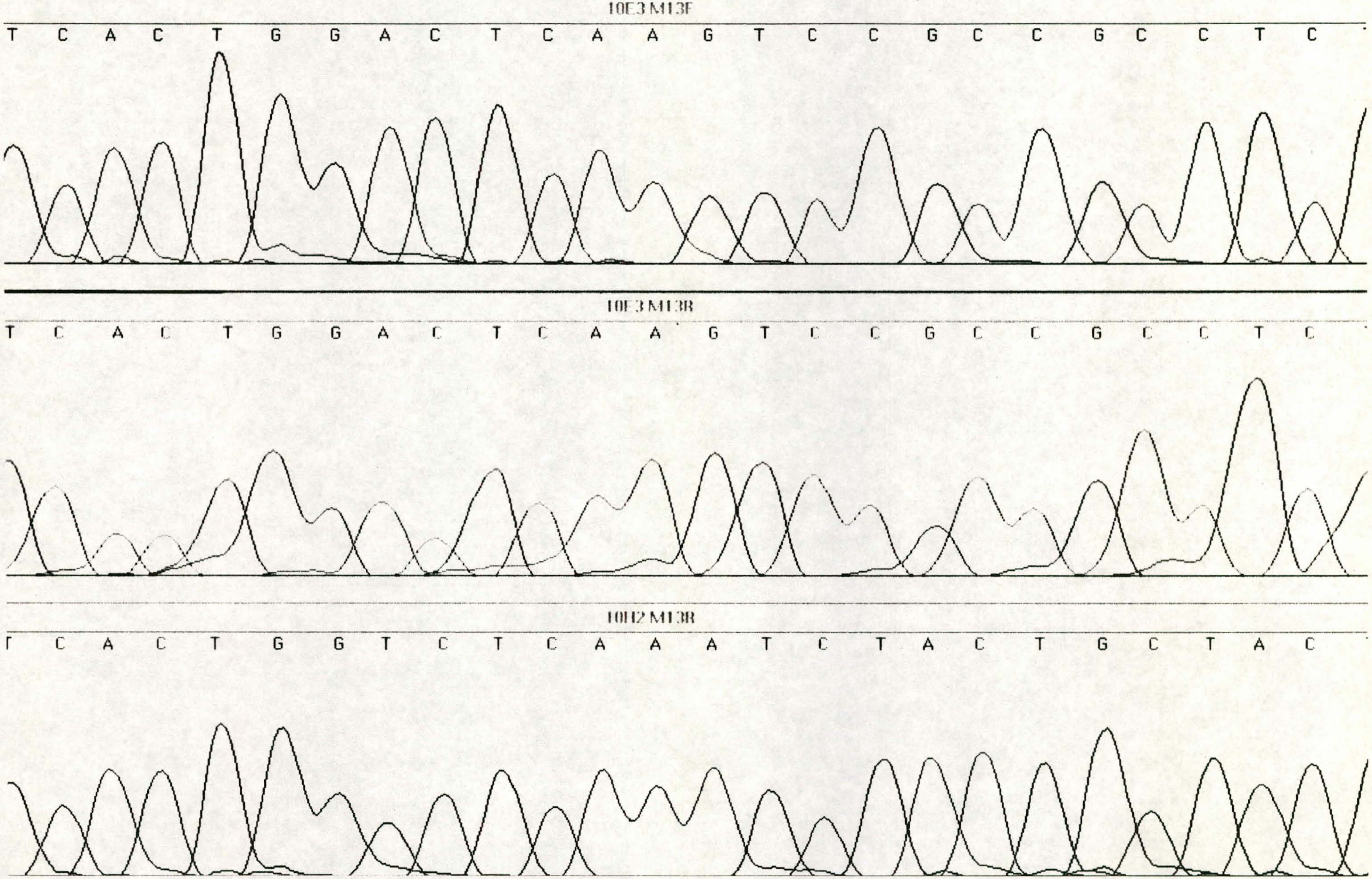
A multitude of approaches for classifying new genes now exist, including complete sequencing and expression, chromosome mapping, tissue distribution, and immunological characterisation (Adams *et al.* 1991). New genes can be identified merely by the simple nucleic acid hybridisation technique using known plant genes as probes, but some genes cannot be isolated by hybridisation, especially genes that have been identified by similarities with genes from other kingdoms. For example, Park *et al.* (1993) reported that full sequencing of a clone that showed 48% amino acid identity with the yeast RNA polymerase II subunit RPB10 is likely to encode a plant RNA polymerase subunit. However, cloning by simple hybridisation would not have been possible because of a low degree of nucleic acid homology. Thus, it appears that partial cDNA sequencing is a useful approach to identify genes that have not previously been found in plants, but that have a low degree of protein similarity with known plant or non-plant genes. Although considerable effort is required in the identification of unknown sequences, these novel unidentified plant genes could provide a powerful gene resource for the analysis of many cellular processes.

The sequencing and database analysis of random cDNA clones provide some insight into the type of genes that are functioning in plant organs. Although whole leaves that are composed of several tissues and cells were used to construct a cDNA library in this study, a better EST

Figure 3.7 Alignment of clones 10e3 and 10h2.

```
                   10         20         30         40         50         60         70         80         90        100        110
                    |          |          |          |          |          |          |          |          |          |          |
10E3 M13F  ---------- ---------- ---------- ---------- --ACGGCcAG TGAATTGTAA TACGAC-TC- ACTATAGGGC GAAT-TGGAG CTCCA---CC GCGGTGGCGG
10E3 M13R  NNNNNCCG-- CAGGGTTTTC CCAGTCACGA CGTTGTAAAA CGACGGCCAG TGAATTGTAA TACGAC-TC- ACTATAGGGC GAAT-TGGAG CTCCA---CC GCGGTGGCGG
10H2 M13R  GANACGCMAR CTCGAAATTW ACCCTCACTA AAGGGAACAA AAGCTGGTAC CGGGCCCCCC CTCGAGGTCG ACGGTATCGA TAAGCTTGAT ATCGAATTCC GATTTAGCCT
           **-******* ********** ********** ********** ***-*-**-* *-******** **---**--* --**--**-* *--**-*--* *--*-***-- -***-*--**

                  120        130        140        150        160        170        180        190        200        210        220
                    |          |          |          |          |          |          |          |          |          |          |
10E3 M13F  CCGCT--CTA GAACTAGTGG ATCCCCCGG- GCTGCAGGAA TTC----CGA T-TTAGCCT- -CATACAGCA TGGTTGCCCC CTTCACTGGA CTCAAGTCCG CCGCCTCTTT
10E3 M13R  CCGCT--CTA GAACTAGTGG ATCCCCCGG- GCTGCAGGAA TTC----CGA T-TTAGCCT- -CATACAGCA TGGTTGCCCC CTTCACTGGA CTCAAGTCCG CCGCCTCTTT
10H2 M13R  CATATGGCTT CCTCTGTTAT TTCCTCTGCA GCTGTTGCTA CACGCACCAA TGTTACACAA GCAAGCAGCA TGGTTGCACC TTTCACTGGT CTCAAATCTA CTGCTACTTT
           -***-**--* ***--**-** *---*-*-** ----**-**- **-****-*- -*---**-** *--**----- -------*-- *--------* -----*--** -*--**----

                  230        240        250        260        270        280        290        300        310        320        330
                    |          |          |          |          |          |          |          |          |          |          |
10E3 M13F  CCCTGTTTCA AGGAAGCAAA ACCTTGACAT TACCTCCATT GCTAGCAATG GTGGAAGAGT CAGATGCATG CAGGTGTGGC CACCAATTAA CATGAAGAAG TACGAGACTC
10E3 M13R  CCCTGTTTCA AGGAAGCAAA ACCTTGACAT TACCTCCATT GCTAGCAATG GTGGAAGAGT CAGATGCATG CAGGTGTGGC CACCAATTAA CATGAAGAAG TACGAGACTC
10H2 M13R  CCCTGTTTCA AGGAAGCAAA ACCTTGACAT CACTTCCATT GCTAGCAATG GTGGAAGAGT TAGGTGCATG CAGGTATGGC CACCAATTAA CATGAAGAAG TACGAGACAC
           ---------- ---------- ---------- *--*------ ---------- ---------- *--*------ -----*---- ---------- ---------- --------*-

                  340        350        360        370        380        390        400        410        420        430        440
                    |          |          |          |          |          |          |          |          |          |          |
10E3 M13F  TATCGTACCT TCCTGATTTG TCCGACGAGC AATTGCTCAA GGAAGTTGAG TACCTATTGA AAAATGGATG GGTTCCTTGC TTGGAATTCG ATATCAAGC- -----TTATC
10E3 M13R  TATCGTACCT TCCTGATTTG TCCGACGAGC AATTGCTCAA GGAAGTTGAG TACCTATTGA AAAATGGATG GGTTCCTTGC TTGGAATTCG ATATCAAGC- -----TTATC
10H2 M13R  TCTCATACCT TCCTGATTTG ACTGACGAGC AATTGCTTAA GGAAGTTGAG TACCTATTGA AAAATGGATG GGTTCCTTGC TTGGAATTTG AGACTGAGCA CGGATTTGTC
           -*--*----- ---------- *-*------- -------*-- ---------- ---------- ---------- ---------- --------*- -*-***---* *****--*--

                  450        460        470        480        490        500        510        520        530        540        550
                    |          |          |          |          |          |          |          |          |          |          |
10E3 M13F  ------GATA CC-----GTC ---------- ---------- -------GAC CTCGAGG--- ---------- --GGGGGCCC GG-------- ---------- ----------
10E3 M13R  ------GATA CC-----GTC ---------- ---------- -------GAC CTCGAGG--- ---------- --GGGGGCCC GG-------- ---------- ----------
10H2 M13R  TACCGTGAGA ACAACAAGTC ACCAGGATAC TATGATGGAA GGTACTGGAC CATGTGGAAG TTGCCCATGT TTGGGTGCAC TGATGCAACC CAAGTGTTGG CTGAGGTTCA
           ******--*- *-*****--- ********** ********** *******--- -**-*--*** ********** **---*--*- *-******** ********** **********

                  560        570        580        590        600        610        620        630        640        650        660
                    |          |          |          |          |          |          |          |          |          |          |
10E3 M13F  ---------- ------TACC ----AGCTT- ---------- -TTG--TTC- --------CC TTTAGTG--A GGGTT---AA TTTCG----- ----AGCTT- --GGCG--TA
10E3 M13R  ---------- ------TACC ----AGCTT- ---------- -TKG--TTC- --------CC TTTAGTG--A GGGTW---AW TTTCG----- ----AGCTK- --KGCG--TA
10H2 M13R  AGAGTGTAAG AACGCATACC CACAAGCATG GATCAAAATC ATTGGATTCG ACAATGTGCG TCAAGTGCAA GTGTATCAAG TTTCATTGNC TACAAGCCAN AAGGCTACTA
           ********** ******---- ****---*-* ********** *-*-**---* ********-* -**----**- -*--****-* ----****-* ****---**- ***--***--
```

Figure 3.8 Chromatograms of 10e3 (forward and reverse sequences) and 10h2.



cataloguing procedure should be achieved by constructing the library from specific tissues or cell types. In some plants it is now possible to isolate particular single cell types, such as guard cells of *Brassica campestris* (Kwak *et al.* 1997).

3.4 STS ANALYSIS

Regardless of whether the ESTs are identifiable in the databases, they can be utilised as STS markers. STS markers represent uniquely identified map positions. ESTs can serve the same purpose as the random DNA STSs and provide the additional feature of identifying an expressed gene.

Although STSs are not as yet widely used in the genome mapping of plant species, they have been successfully used for the physical mapping of human chromosomes (Srivastava *et al.* 1999). STS mapping of the ESTs in the potato genome may provide genomic information such as the distribution of expressed genes in the genome and the relationship between the expression patterns or functions of genes and their location in the genome.

The 5' -untranslated regions of many plant genes are relatively short compared to the 3' -untranslated regions (Joshi 1987) and both of the 3' - and 5' -untranslated regions are much less conserved than the coding regions. In contrast to the results from *Arabidopsis* presented by Höfte *et al.* (1993), which concerned essentially 5' sequences, most studies are focussed on systematically sequencing the 3' end of cDNA clones. While 5' sequences allow identification of putative protein products, 3' sequencing allows the development of gene-specific probes and oligonucleotide primers (Cooke *et al.* 1996; Perry & Bousquet 1998a).

In this study amplification primers were selected for 14 cDNA clones. The aim was to select a reverse primer in the presumed 3' -untranslated region and a forward primer at least 200 bp upstream within the coding region. PCR conditions were determined for all but one set of primers. These primer pairs were used to screen a panel of genomic DNA (See Table 2.1).

The primer sequences and the cDNA and genomic PCR product lengths are listed in Table 3.3. cDNA lengths were derived from the clone sequences from which the primers were designed. Genomic PCR product lengths were determined by comparing the amplified products to a size marker on agarose gels. The estimated length of the genomic PCR product was often comparable to that of the cDNA product. For example, the cDNA product using primer pair 2g4, is 166 bp and the estimated genomic product is of similar length. However, the size of the genomic product differed from that of the cDNA product in some cases. The estimated genomic amplification product of primer pair 6b11, for example, was 1 kb, in contrast to a cDNA product of 376 bp, perhaps ascribable to inclusion of one or more introns.

Table 3.3 STSs developed in this study: the primer sequences, the length of the cDNA product and the estimated length of the genomic amplification products in bp.

EST	**Forward primer**	**Reverse primer**	**Product size (bp)**	
			cDNA	**Genomic**
2g4	AAGGTGGTCAGCAAGGTTCTC	TAGCAGCTTTGACCCATTAGG	166	~160
3c10	GCTTCTTCAGTAATGTCCTC	ATCAGGAAGGTACGATAGAG	227	~300
3d6	TTTCCCACTCAGCCTCAGTTA	ACGCTCAAGTTGCTACCGTTA	309	~300
3d11	ATGGCTTCTACTTGTGCAGT	CTTGTTAAGTCCGCTTGAC	366	~350
3h5	AATGTGAGAAAGAACCAACG	GCAAATTCTGGAAAGGGAA	165	~160
3h7	CTGAACTAGGCCAGGTTATT	ACACGACTCATATTCAGACG	273	~700
6b10	CTTCCATCACTTGCTAGATC	AGGATAACATCTACTGTGCC	354	~1500
6b11	CTTGCACCACAACTTCTTCT	GAGAGAGTCCAGATTCTTCA	376	~1000
6d6	ACTGTTACCTCTGCTGCTGT	CTGCATTCAGAAGATCCTCT	402	~400
6h9	CTAATGGTCCTTCTGCACTT	TAGAGTTGCCTCAACTTCAG	379	~800
7a5	CCTTCACCAATCTCAAAGTC	CCTTAACAGTGGAACACAAC	197	~300
7b4	ATTCACCAGAGGCTCAAGTA	AGGACTACACCAACAATTGC	192	~500
7e9	AGTCATCGGACACCATTG	CTTACGGAGCCTCTGAAC	304	~300

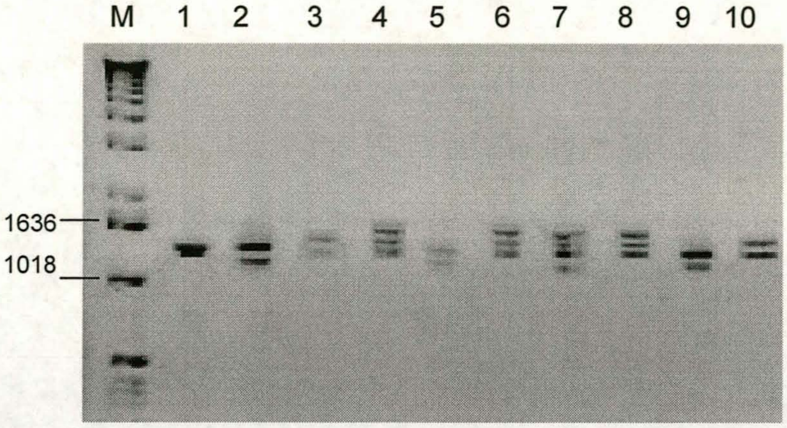
3.4.1 Agarose gel analysis of potato cultivars

An initial objective of this analysis was to detect length polymorphisms between potato cultivars by standard agarose gel electrophoresis as described by Perry & Bousquet (1998a). Only one set, 6b10, revealed polymorphisms using this procedure (Figure 3.9). A typical pattern usually associated with repetitive sequences was found. A single fragment was amplified for the cultivar ALTO (lane 1) and two to four fragments, differing in size by approximately 100 bp, were amplified for the other cultivars. Whether these represent actual alleles or PCR artifacts could not be decided. Similar results were obtained when the PCRs were repeated. No product was found when PCR conditions were changed. Fragments of diverse lengths were reamplified and sequenced in an attempt to determine the origin of length differences. However, several attempts yielded only poor quality sequences. The reaction with the forward primer sometimes produced a good quality sequence for the first 330 bp after which the sequence became impossible to read. This sequence was compared to sequences in the databases with BLASTN and it matched the same gene (light inducible tissue-specific protein) as was found using the cDNA sequence. The gene sequence was retrieved from the database and examined for repeat sequences. A short GA-repeat was found but this does not explain the difference of ± 100 bp between the fragments of the genomic PCR product. There was no sign of secondary binding sites detected for either of the two primers.

Analysis of codominant cDNA-based STS markers of three Adh loci in jack pine revealed alleles differing in size caused by the presence or absence of large (>38 bp) repeats (Perry & Furnier 1996). In black spruce, STS markers were developed from ESTs with alleles differing in size by 70 bp due to the duplication or triplication of a repeat sequence (Perry & Bousquet 1998a). These authors found that additional repeats may lead to inconsistent results.

In general, the proportion of directly observable length polymorphisms has been low, although digestion of amplification products with restriction enzymes (PCR-RFLP) has been used routinely. However, there are some reports of allelic length polymorphisms of plant genes by Tragoonrung *et al.* (1992), Bradshaw *et al.* (1994), Perry & Furnier (1996), Davis & Yu (1997), Perry & Bousquet (1998a,b) and Hernandez *et al.* (1999). The low frequency of directly observable length polymorphism may be due to the type of panel used for screening. Small panels or panels with a restricted genetic base, may include only a fraction of the total genetic diversity. The panel used in this study was omposed of only 10 closely related potato cultivars developed in South Africa. A more diverse panel of 40 rice varieties screened with 15 pairs of STS primers was found to yield 40% length variation Ghareyazie *et al.* (1995) and Perry & Bousquet (1998b) report similar results (38%) in black spruce. These results suggest that the potential success rate may be reasonably high for a wide range of plant species.

Figure 3.9 Genomic amplification products of primer pair 6b10 on an ethidium bromide stained 1.5% agarose gel. The size of the fragment in lane 1 is estimated at 1.4 kb and the fragments differ with approximately 100 bp. Lanes 1 - 10 correspond to cultivars 1 - 10 (Table 2.1). M = 1 kb ladder, the sizes are given in bp.



3.4.2 SSCP analysis of potato cultivars

Since polymorphisms could not be visualised directly on agarose gels, the application of the SSCP technique to detect point mutations or length variations of only a few base pairs was evaluated. This technique has been widely used for the detection of mutations in the human genome (Liu *et al.* 1999). Detection of SSCPs in the 3' UTRs of bovine (Lowden & Drinkwater 1999) and fish species identification with PCR-SSCP (Rehbein *et al.* 1999) have also been reported recently. SSCP markers have also been developed from cDNA clones for plant species such as blackcurrant (Lemmetty *et al.* 1999), *Piceae* (Germano & Klein 1999) and *Pinus* (Plomion *et al.* 1999).

PCR products were analysed on 20% acrylamide gels with 1 % cross-linking, using the same conditions for all the STSs. All reactions were repeated and the results were found to be reproducible. Results of the preliminary study revealed polymorphisms in 12 pairs (92%) as depicted in Figures 3.10 a - m. Unique band patterns were identified for each marker, referred to as profiles in discussions which follow. Between two and six unique profiles were amplified for each marker (See Table 3.5) using 10 potato cultivars (Table 2.1).

Two profiles were identified for each of 2g4 (Figure 3.10 a), 3d6 (Figure 3.10 c) and 7e9 (Figure 3.10 m), even though the polymorphisms were not very clear. Only a faint band for 2g4 distinguished cultivars 2 and 4 from the rest and the polymorphic bands of marker 3d6 and 7e9 were very close together. For better resolution the duration of electrophoresis could be adjusted, or a different electrophoretic system using gels of different lengths could be employed.

Simple band patterns were observed and two profiles were detected for 3c10 (Figure 3.10 b), 3d11 (Figure 3.10 d) and 3h5 (Figure 3.10 e). However, the variation for 3c10 and 3h5 were not as clear as that of 3d11.

Using primer pair 3h7 (Figure 3.10 f) a simple band pattern revealing three profiles was resolved. An extra band was found in lane 3 and one band was absent in lane 1. However, electrophoretic differences may cause samples in the lanes nearest to the edges of the gel to migrate differently and therefore the difference in lane 1 should only be regarded as a true polymorphism once the samples have been electrophoresed again with this sample loaded in a different lane.

The SSCPs of 6b10 (Figure 3.10 g) and 6b11 (Figure 3.10 h) formed very complex patterns. At least six profiles could be resolved for 6b10. This was not unexpected as comlex patterns had already been detected with agarose gel electrophoresis (Figure 3.9). Although one could see that there are variations in the faint bands in Figure 3.10h, it was not possible to identify profiles for 6b11. The addition of denaturants to the gels, such as formamide or urea, might sharpen bands and clear the background (Yip *et al.* 1999) but this strategy was not examined in this study.

Clear and simple patterns were visible for 6d6 (Figure 3.10 i), 7a5 (Figure 3.10 k), and 7b4 (Figure 3.10 l). Five profiles were recognised for 6d6 and three each for 7a5 and 7b4. No polymorphisms were detected using primer pair 6h9 (Figure 3.10 j).

Figure 3.10 (a) to (m) depicts the genomic amplification products of the STS primer pairs electrophoresed on 20% non-denaturing acrylamide gels. Lanes 1 - 10 correspond to cultivars 1 - 10 as specified in Table 2.1. Arrows indicate polymorphic bands, except for Figure 3.10 g and h. Bands framed in Figure 3.10 d, i, k, l and m are discussed in section 3.4.7.

Figure 3.10 (a) SSCPs of 2g4.

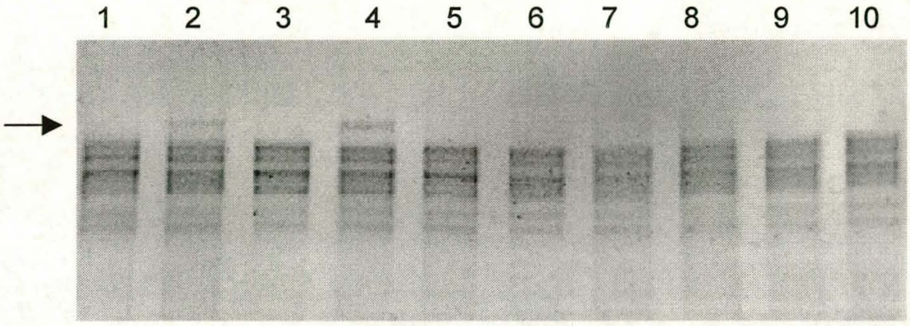

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (b) SSCPs of 3c10.

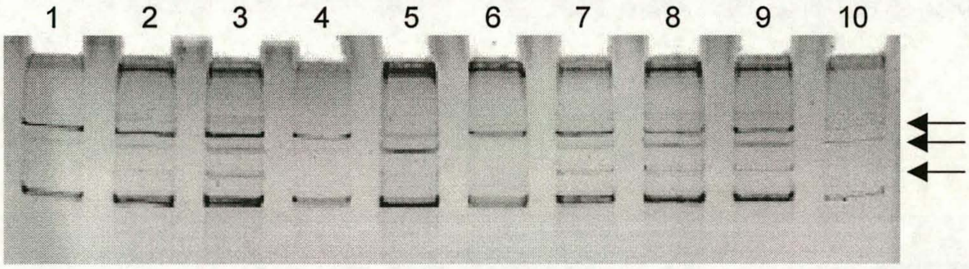

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (c) SSCPs of 3d6.

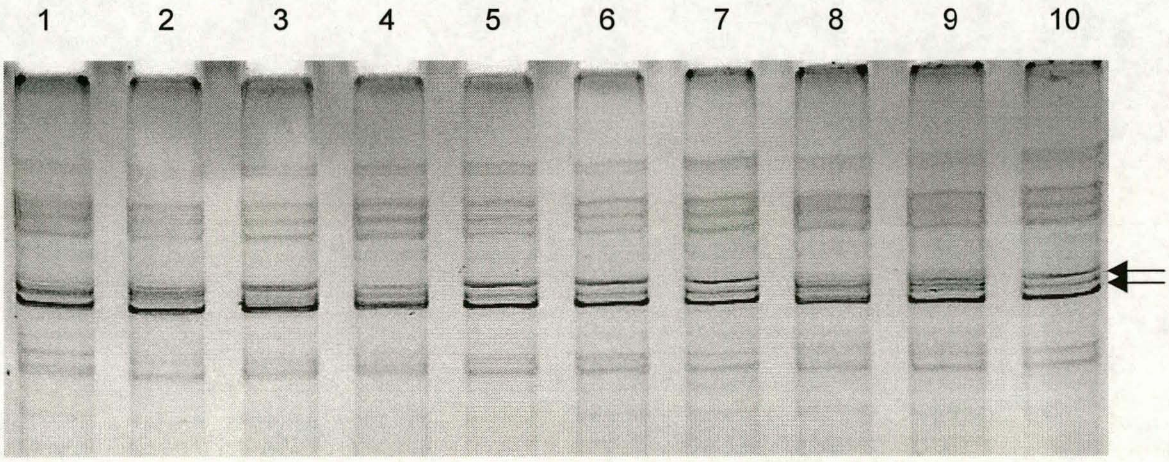

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (d) SSCPs of 3d11.

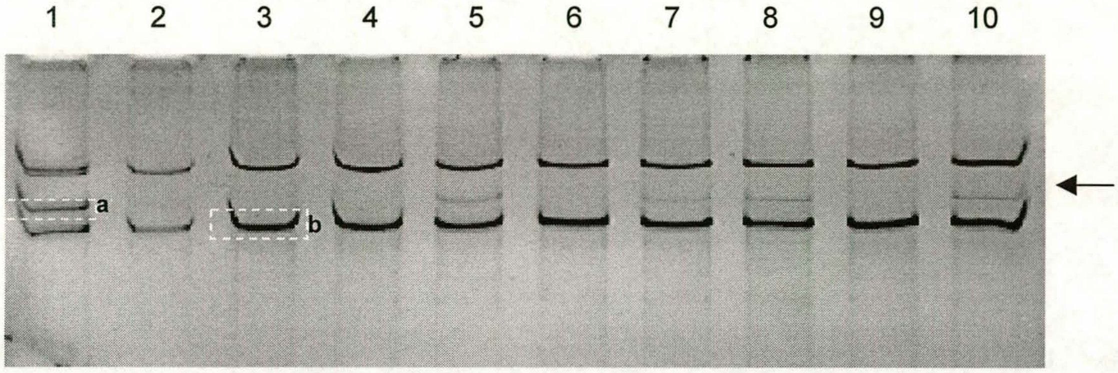

1
2
3
4
5
6
7
8
9
10
a
b

Figure 3.10 (e) SSCPs of 3h5.

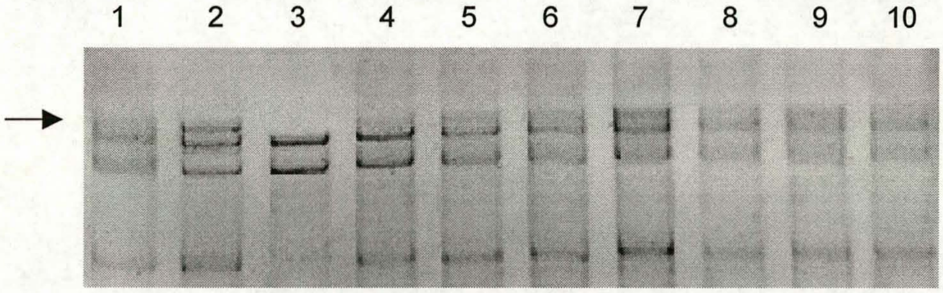

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (f) SSCPs of 3h7.

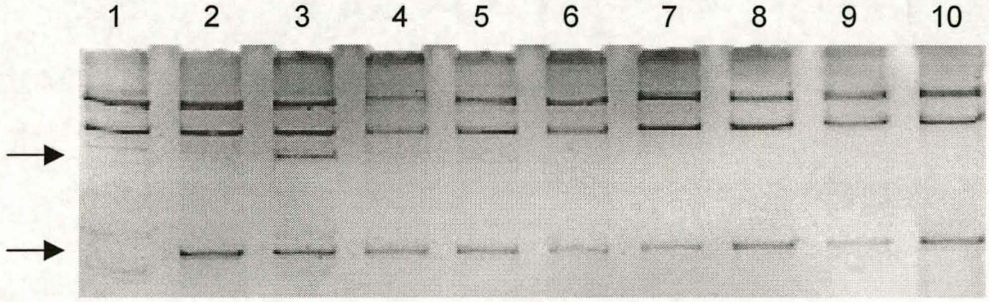

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (g) SSCPs of 6b10.

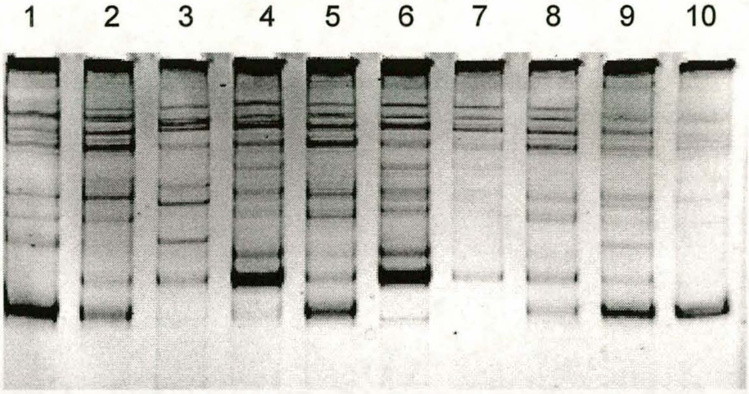

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (h) SSCPs of 6b11.

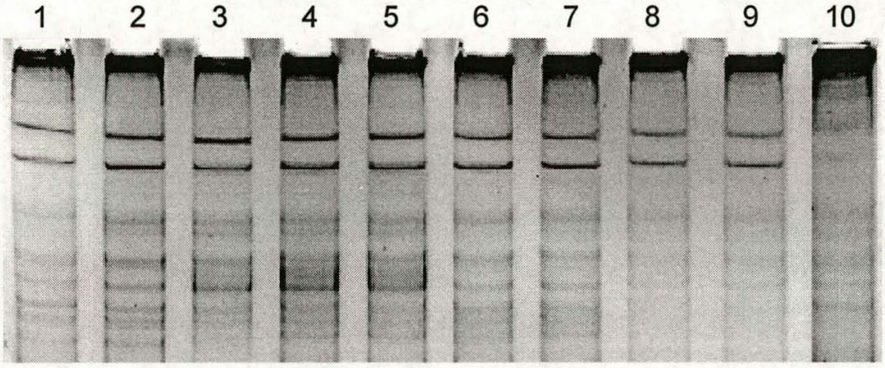

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (i) SSCPs of 6d6.

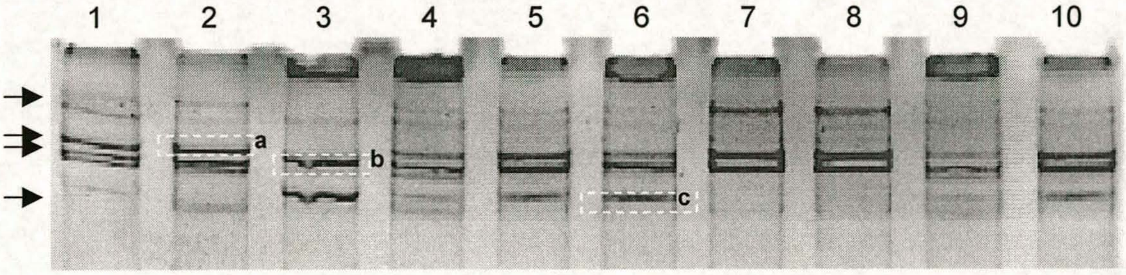

1
2
3
4
5
6
7
8
9
10
a
b
c

Figure 3.10 (j) SSCPs of 6h9.

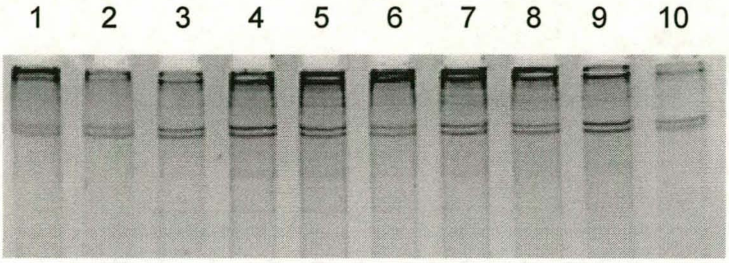

1
2
3
4
5
6
7
8
9
10

Figure 3.10 (k) SSCPs of 7a5.

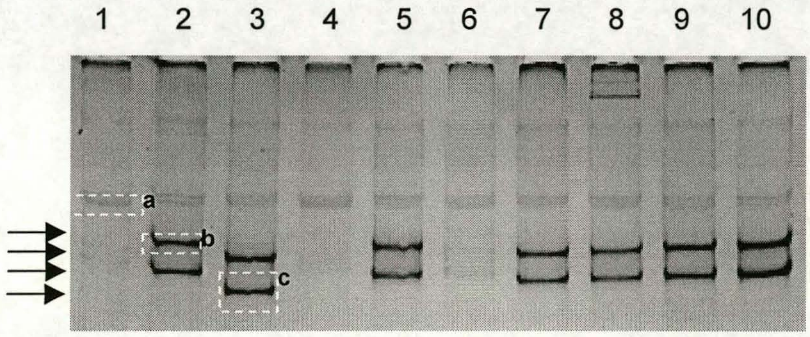

1
2
3
4
5
6
7
8
9
10
a
b
c

Figure 3.10 (l) SSCPs of 7b4.

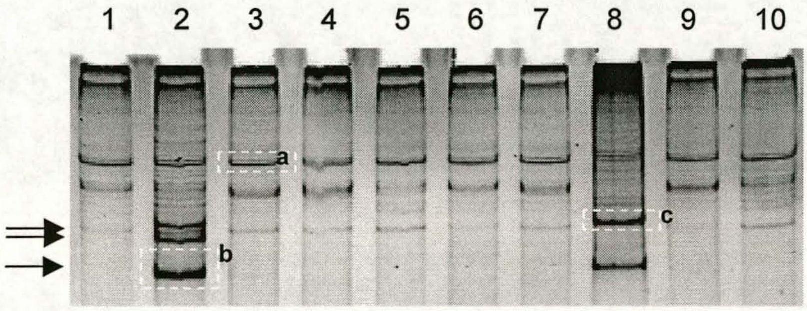
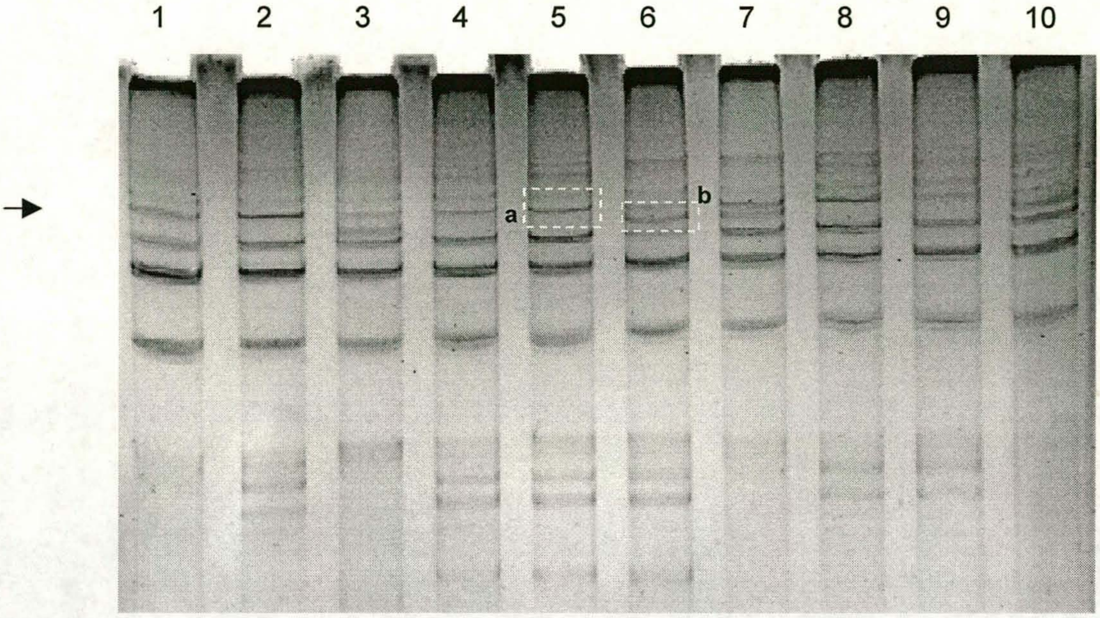

1
2
3
4
5
6
7
8
9
10
a
b
c

Figure 3.10 (m) SSCPs of 7e9.



Since introns are more variable than coding regions, one would expect products that include introns to reveal more polymorphisms. However, this was not found for the SSCP results of this study. For instance, primer pair 6d6 produced genomic and cDNA products of the same size (402 bp), but revealed five unique profiles. In contrast, primer pair 6h9 spans one or more introns (genomic fragment >800 bp; cDNA 379 bp), but no polymorphisms were detected. Although the absence of polymorphisms in the latter product cannot be ruled out, it should be noted that the optimum fragment size for SSCP analysis is between 100 and 400 bp (Hayashi 1996). The size of the product generated by primer pair 6h9 is thus beyond the resolution of the SSCP gels used in this study.

The SSCP technique provides an economic and simple way to detect polymorphisms. After the initial high costs of sequencing and primer development, no further manipulation is required. Variation is detected without the need of restriction-digestion or the use of hazardous and costly radioisotopes. Costs are kept to a minimum by running small acrylamide gels on the Mighty Small (Hoefer) or a similar apparatus.

3.4.3 Characterisation of SSCPs

Five SSCPs were selected for characterisation by sequencing (Table 3.4). Different bands (framed in Figures 3.10 d, i, k, l and m) were stabbed and reamplified. The PCR products were electrophoresed on the acrylamide gels once more to confirm that the correct band was amplified. However, the same multiple band patterns, from which a single band was selected, were again found. Therefore the original PCR product was used for sequencing. Unique profiles were identified for each marker (Table 3.5) of which most could be elucidated by the sequencing analysis.

From the sequencing results it was deduced that single nucleotide polymorphisms gave rise to the variation for markers 3d11, 6d6 and 7e9. Profile i with 3d11 (Figure 3.10 d, lane 1) corresponds to sequence $G_{25}A_{93}A_{131}$ and Profile ii (Figure 3.10 d, lane 3) to sequence $A_{25}T_{93}T_{131}$. Only three variant bands were selected for 6d6 (Figure 3.10 i), although five profiles had been detected. Profile i (Figure 3.10 i, lane 2) was identified as $G_{59}C_{60}G_{159}$, Profile ii (Figure 3.10 i, lane 3) as $C_{59}T_{60}A_{159}$ and Profile iii (Figure 3.10 i, lane 6) as $G_{59}T_{60}G_{159}$. The other two profiles (Figure 3.10 i, lane 4 and 5) could be the result of different combinations of these variant bands but since the potato genome is tetraploid, it was very difficult to assign true alleles. For 7e9, two polymorphic bands were resolved, $G_{76}A_{115}$ (Profile i - Figure 3.10 m, lane 5) or $A_{76}G_{115}$ (Profile ii - Figure 3.10 m, lane 6). Differences in restriction enzyme cutting sites in the polymorphic areas were examined by means of the Webcutter 2.0 program (href13). Restriction enzymes that could differentiate between different profiles were found for 3d11, 6d6 and 7e9 (See Table 3.4). These enzymes could be used for cultivar identification instead of the SSCP analysis. However, this would be a more laborious and expensive method, especially since the enzymes identified are not frequently used in most laboratories.

Differences for 7a5 (Figure 3.10 k) originated from an insertion and single nucleotide polymorphisms in the forward primer binding site. (Table 3.4). Whether these reflected PCR/sequencing artifacts or true polymorphisms, is not clear. The contribution of these variations to the three profiles detected is also not known at this stage.

Polymorphisms detected for 7b4 (Figure 3.10 l) are due to insertions and deletions. At nucleotide 346, either 8, 9 or 10 A's were present in Profiles i, ii and iii respectively. In Profile iii an insertion was also found after base 71.

Table 3.4 A summary of the variation detected by sequencing a selection of SSCP fragments, generated by five STS markers. Base pair numbers are given in the forward direction of the sequence starting from the first base of the forward primer (FP = forward primer; RP = reverse primer). Brackets are used to indicate insertions. RE = restriction enzymes which would detect the variation. The superscript numbers assigned to the different enzymes refer to the profile (Table 3.5) to be distinguished from the rest using the specific enzyme.

Marker	**Variation in sequence**	**RE**
3d11	FP$_{20}$-----C$_{24}$ G/A$_{25}$ C$_{26}$-----C$_{92}$ A/T$_{93}$ A$_{94}$-----A$_{130}$ A/T$_{131}$ T$_{132}$-----RP	BsaMI^{i}, BsmI^{i}, Mva12691^{i}
6d6	FP$_{20}$-----G$_{58}$ G/C$_{59}$ C/T$_{60}$ G$_{61}$-----C$_{158}$ A/G$_{159}$ A$_{160}$-----RP	Bst71I^{i}, ItaI^{i}, MspA1I^{i}, MwoI^{i}, AgeI^{ii}, AccII^{ii,iii}
7a5	T$_{3}$ T/C$_{4}$ C/A$_{5}$ A/C$_{6}$ C$_{7}$-----C$_{12}$ T/C$_{13}$ (TAGTCCAATCC) C-----RP	
7b4	FP$_{20}$-----A$_{71}$ (ACTA) C$_{72/76}$-----A$_{346/350}$ (8/9/10 T's)-----RP	
7e9	FP$_{20}$-----G$_{75}$ G/A$_{76}$ T$_{77}$-----C$_{114}$ A/G$_{115}$ C$_{116}$-----RP .	AciI^{i}, MspA1^{i}, MnlI^{i}, NheI^{ii}, PstNHI^{ii}

Table 3.5 A summary of the unique profiles assigned to the ten cultivars for the five selected markers. Each unique profile was designated by a Roman numeral (See Figure 3.10 d, i, k, l and m). Some conspicuous bands were not taken into consideration.

Marker	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3d11	i	ii	ii	ii	i	ii	i	i	ii	i
6d6	i	i	ii	v	iii	iii	iv	iv	v	iii
7a5	i	ii	iii	i	ii	i	ii	ii	ii	ii
7b4	i	ii	i	i	i	i	i	iii	i	i
7e9	i	i	ii	i	i	ii	ii	i	i	ii

3.4.4 STS agarose gel analysis of the Solanaceae family

There are few reports on the transferability of cDNA-based STS primers among plant species. Tsumura *et al.* (1997) recently reported the use of primers developed for the conifer *Cryptomeria japonica* to screen a panel of 14 different coniferous species. STSs developed for black spruce were successfully used for the screening of 12 additional conifer species and other Pinaceae (Perry & Bousquet 1998b).

Genomic DNA of six members of the Solanaceae were amplified with the 13 STS primer pairs developed in this study. The species were potato, tomato, tobacco, green pepper, petunia and *Solanum nigrum* (See Table 2.1). The same conditions used to amplify the potato cultivars were employed for all the species. The PCR products were analysed on 1.5% agarose gels (Figure 3.11 a - d) and reactions were repeated to confirm results.

Only two primer pairs, 2g4 (Figure 3.11 a) and 7e9 (Figure 3.11 c), showed no variation. All the different species yielded an amplification product of the same length, suggesting that these primers were probably developed from a very conserved region. According to the database similarity searches, 2g4 matched a putative chlorophyll binding protein and 7e9 a ubiquitin extension protein. These are so-called "house-keeping" genes and are expected to be found in all these species.

Two primer pairs, 3h7 (Figure 3.11 b) and 6b11 (Figure 3.11 c), amplified only potato DNA. The sequence similarity results (Table 3.1) (a translation initiation factor matched the sequence of 3h7 and a photosystem reaction centre protein matching 6b11) suggest that these should be common proteins. Since conditions determined for the potato were used for the amplification of all the Solanaceae species, the other species might yield a product using other PCR conditions.

Of the remaining nine polymorphic STSs, six showed variation in length between species: Figure 3.11a (6h9), Figure 3.11c (6b10, 6d6), Figure 3.11d (7a5, 7b4 and 3h5). Furthermore, markers 6h9 (Figure 3.11a), 3d11 (Figure 3.11b) and 3h5 (Figure 3.11d) produced more than one fragment in some species. This could be the result of an extra binding site for these primers in genomes other than potato. However, the additional fragments could again be due to the fact that PCR conditions optimised for potato were used to amplify all species.

Figure 3.11 (a) Amplification products using primer pairs **2g4** (left), **3d6** (middle) and **6h9** (right) on a 1.5% agarose gel. M = 1 kb ladder (given in bp); Po = potato; Sn = *Solanum nigrum*; P = *Petunia*; GP = green pepper; Tb = tobacco; To = tomato.

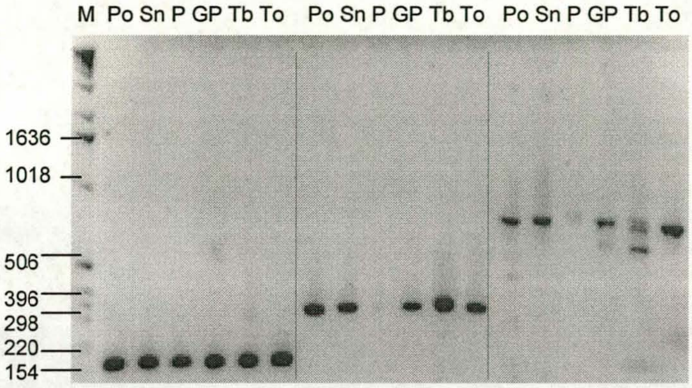

M Po Sn P GP Tb To Po Sn P GP Tb To Po Sn P GP Tb To
1636
1018
506
396
298
220
154

Figure 3.11 (b) Amplification products using primer pairs **3d11** (left), **3h7** (middle) and **3c10** (right) on a 1.5% agarose gel. M = 1 kb ladder (given in bp); Po = potato; Sn = *Solanum nigrum*; P = *Petunia*; GP = green pepper; Tb = tobacco; To = tomato.

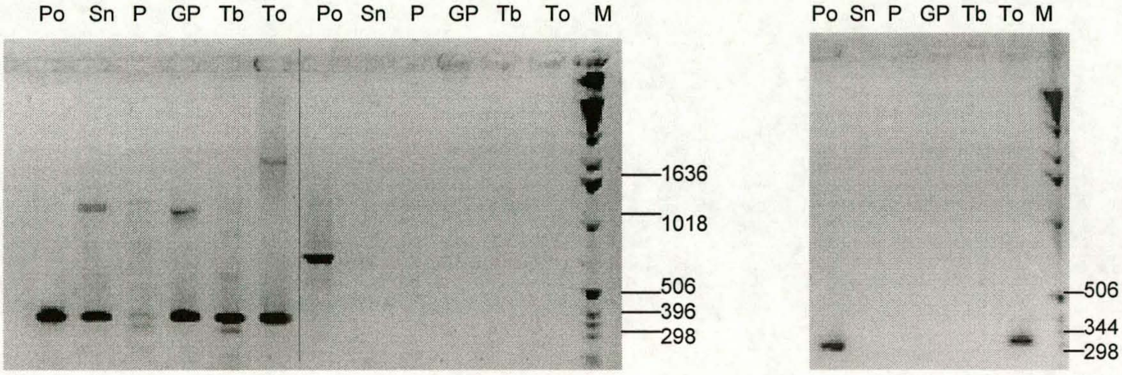

Po Sn P GP Tb To Po Sn P GP Tb To M
1636
1018
506
396
298
Po Sn P GP Tb To M
506
344
298

Figure 3.11 (c) Amplification products using primer pairs **6b10** (left), **6b11** (middle), **6d6** (right) and **7e9** (far right) on a 1.5% agarose gel. M = 1 kb ladder (given in bp); Po = potato; Sn = *Solanum nigrum*; P = *Petunia*; GP = green pepper; Tb = tobacco; To = tomato.

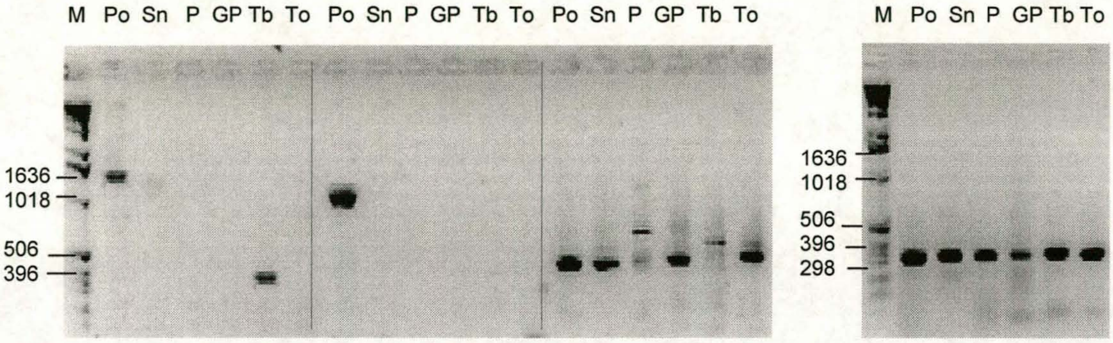
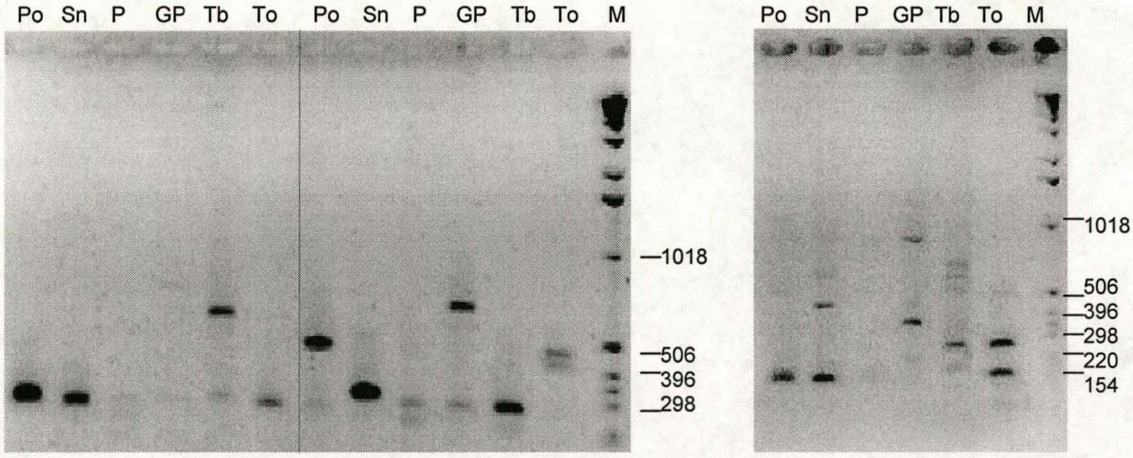

M Po Sn P GP Tb To Po Sn P GP Tb To Po Sn P GP Tb To
M Po Sn P GP Tb To
1636
1018
506
396
1636
1018
506
396
298

Figure 3.11 (d) Amplification products using primer pairs **7a5** (left), **7b4** (middle) and **3h5** (right) on a 1.5% agarose gel. M = 1 kb ladder (given in bp); Po = potato; Sn = *Solanum nigrum*; P = *Petunia*; GP = green pepper; Tb = tobacco; To = tomato.


Po Sn P GP Tb To Po Sn P GP Tb To M
Po Sn P GP Tb To M
1018
506
396
298
1018
506
396
298
220
154

The scoring of primer performance (Table 3.6) may provide a useful starting point when initiating studies involving one of the other Solanaceae species. These results indicate that STSs developed from potato cDNA can be used in these species. Primer pairs may even amplify plant species other than Solanaceae, especially those primers that produced PCR products in all of the species used in this study. Tsumura *et al.* (1997) found limited amplification success outside of the Taxodiaceae for primers developed for *Cryptomeria japonica*. Furthermore, polymorphisms can be detected by means of agarose gel electrophoresis without the necessity of manipulating PCR products.

The results in Table 3.6 were somewhat unexpected. Since the genomes of tomato and potato are similar (Tanksley *et al.* 1992), one might have expected that more markers should amplify tomato DNA than any of the other species. However, it was found that tobacco DNA was amplified with most markers (ten), followed by tomato and *S. nigrum* (nine), green pepper (eight) and *Petunia* (five).

In summary 13 STS primer pairs were developed from potato cDNA clones. Twelve of these could detect polymorphisms between potato cultivars by means of SSCP analysis and this provides a simple and cheap alternative to identify variation. Eleven markers could also amplify other solanaceous species, indicating that STSs developed for potato could also be used in studies analysing the genomes of other members of the Solanaceae family.

Table 3.6 Amplification trials for 13 potato-derived STS primer pairs in six solanaceous species. Each species-primer combination was scored for success (+) or failure (-) of amplification.

	Potato (*S. tuberosum)*	*S. nigrum*	*Petunia*	Green Pepper (*Capsicum annuum)*	Tobacco *(Nicotiana tabacum)*	Tomato *(Lycopersicon esculentum)*
2g4	+	+	+	+	+	+
3c10	+	-	-	-	-	-
3d6	+	+	-	+	+	+
3d11	+	+	+	+	+	+
3h5	+	+	-	+	+	+
3h7	+	-	-	-	-	-
6b10	+	-	-	-	+	-
6b11	+	-	-	-	-	-
6d6	+	+	+	+	+	+
6h9	+	+	-	+	+	+
7a5	+	+	-	-	+	+
7b4	+	+	+	+	+	+
7e9	+	+	+	+	+	+

CHAPTER 4

CONCLUSION

The ultimate goal in plant genome analysis is to isolate and tag all unique genes, estimated between 20 000 and 30 000 for crop plants (Kasha 1999). Only a small fraction of these have been identified in potato. Molecular markers which are rapid and inexpensive to analyse are needed in the attempt to reach this goal. The development of markers from cDNA clones has the advantage of directly pinpointing a gene and by analysing cDNA clones and not genomic sequences, the strategy is essentially neutral to genome size. Single nucleotide polymorphisms are the most common variation in a genome and methods to detect them are becoming increasingly important. Mapping with characterised markers would provide excellent landmarks to expedite potato genome analysis.

The primary aim of this study was to construct a cDNA library of the potato. This was accomplished by producing some 3000 cDNA clones. Although not all mRNAs are represented in such a small library, the library proved to be adequate for the development of STS markers. By employing techniques such as electroporation, or by using different competent cells or a phage vector, a more representative library could possibly be constructed.

Selected inserts were characterised by means of PCR and sequencing. DNA sequence data provide the highest level of identification of gene structure. Putative functions were assigned to ESTs by comparing the cDNA sequences to established database sequences. Combining the technology of automated sequencing with randomly selected cDNA clones allows one to construct a detailed picture of the transcriptional activity of a cell or tissue. In this study it was found that most of the ESTs showed similarities to genes that are abundantly expressed in leaves, such as photosynthesis-related proteins. Thus, to study gene expression in other types of tissue, for example potato tubers, a library constructed from tuber tissue should be analysed.

STS primers were designed based on the EST sequences and were subsequently used to screen a panel of potato cultivars. PCR products were analysed on non-denaturing acrylamide gels without the need for further manipulation or the use of radioisotopes, thus providing a cheap, safe and simple method to identify variation. Between two and six unique profiles per marker were detected by means of SSCP analysis.

Much time and money could be saved if primers generated for one species could be used in others. It was found that the markers developed for potato could be used in other Solanaceae as well. Variation was detected between different members of the Solanaceae family simply by means of PCR and standard agarose gel electrophoresis.

In this study, useful STS markers were developed from a potato cDNA library and can also be used in other Solanaceae species. This thesis is the first report on the application of cDNA-based STSs combined with SSCP analysis in potato. Markers generated can be used to map and tag traits of interest and may even be useful in diversity studies.

Future prospects

The STS markers developed in this study are currently being mapped and will be incorporated into the linkage map of potato. Although the present library was adequate for the development of markers, a more effective cloning system could expand the library and increase its value even more.

In future, the ESTs might be used in expression profiling using cDNA microarrays (Duggan *et al.* 1999). Using this technique, DNA targets, in the form of 3' ESTs, are arrayed onto glass slides and probed with labelled cDNAs. These cDNA microarrays are capable of profiling gene expression patterns of tens of thousands of genes in a single assay. cDNA chips can also be used for SNP screening (Brown 1999). Oligonucleotides for different alleles of each SNP can be arrayed on a chip which is then hybridised to labelled target DNA. Hence, the cDNA library constructed in this study might be appropriate for various applications in future studies.

REFERENCES

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. & Venter, J.C. 1991. Complementary DNA sequencing: Expressed sequence tags and the human genome project. *Science*, 252: 1651 - 1656.

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R.F., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. & Venter, J.C. 1992. Sequence identification of 2375 human brain genes. *Nature*, 355: 632 - 634.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403 - 410.

Altschul, S.F., Boguski, M.S., Gish, W. & Wootton, J.C. 1994. Issues in searching molecular sequence databases. *Nature Genet.*, 6: 119 - 129.

Altschul, S.F., Madden, T.L., Schåffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI–BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25: 3389 - 3402.

Altschul, S.F. 1999. BLAST Tutorial. http://www.ncbi.nlm.nih.gov/BLAST/tutorial /Altschul-1.html

Arumuganathan, K. & Earle, E.D. 1991. Nuclear DNA content of some important plant species. *P. Mol. Bio. Rep.*, 9: 208 - 218.

Aviv, H. & Leder, P. 1972. Purification of biologically active globin messenger RNA by chromatography on oligo thymidylic acid-cellulose. *Proc. Natl. Acad. Sci. U.S.A.,* 69: 1408 - 1412.

Beckman, J.S. & Soller, M. 1990. Toward a unified approach to the genetic mapping of eukaryotes based on sequence-tagged microsatellite sites. *BioTechnology*, 8: 930 - 932.

Belyavsky, A., Vinogradova, T. and Rajewsky, K. 1989. PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res.*, 17: 2919 - 2932.

Boguski, M.S., Lowe, T.M.J. & Tolstoshev, C.M. 1993. dbEST - database for "expressed sequence tags". *Nature Genet.*, 4: 332 - 333.

Bonierbale, M.W., Plaisted, R.L. & Tanksley, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosome evolution in potato and tomato. *Genetics*, 120: 1095 - 1103.

Bradshaw, H.D., Jr., Villar, M., Watson, B.D., Otto, K.G. & Stewart, S. 1994. Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. *Theor. Appl. Genet.*, 89: 167 - 178.

Bradshaw, J.E. 1994. Quantitative genetics theory for tetrasomic inheritance. In: Bradshaw, J.E. & Mackay, G.R. (eds.). *Potato genetics*. Oxon: CAB International.

Bradshaw, J.E. & Mackay, G.R. 1994. Breeding strategies for clonally propagated potatoes. In: Bradshaw, J.E. & Mackay, G.R. (eds.). *Potato genetics*. Oxon: CAB International.

Bradshaw, J.E., Stewart, H.E., Wastie, R.L., Dale, M.F.B. & Phillips, M. 1995. Use of seedlings progeny tests for genetical studies as part of a potato (*Solanum tuberosum* subsp. *tuberosum*) breeding programme. *Theor. Appl. Genet.*, 90: 899 - 905.

Bradshaw, J.E., Hackett, C.A., Meyer, R.C. Milbourne, D., McNicol, J.W., Phillips, M.S. & Waugh, R. 1998. Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to marker aided selection. *Theor. Appl. Genet.*, 97: 202 - 210.

Brown, T.A. 1999. *Genomes*, Oxford: βIOS Scientific Publishers.

Buetow, K.H., Edmonson, M.N. & Cassidy, A.B. 1999. Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genet.*, 21: 323 - 325.

Burglin, T.R. & Barnes, T.M. 1992. Introns in sequence tags. *Nature*, 357: 367 - 368.

Caetano-Anollés, G., Bassam, B.J. & Gresshoff, P.M. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology*, 9: 553 - 557.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonucleases. *Biochemistry,* 18: 5294-5299.

Collins, F.S., Guyer, M.S. & Chakravarti, A. 1997. Variations on the theme: cataloging human DNA sequence variation. *Science*, 278: 1580 - 1581.

Cooke, R., Raynal, M., Laudié, M., Grellet, F., Delseny, M., Morris, P-C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G., Li, Y-F., Mache, R., Krivitsky, M., Gy, I.J-J., Kreis, M., Lecharny, A., Parmentier, Y., Marbach, J., Fleck, J., Clément, B., Philipps, G., Hervé, C., Bardet, C., Tremousaygue, D., Lescure, B., Lacomme, C., Roby, D., Jourjon, M-F., Chabrier, P., Charpenteau, J-L., Desprez, T., Amselem, J., Chiapello, H. & Höfte, H. 1996. Further progress towards a catalogue of all *Arabidopsis* genes: Analysis of a set of 5000 non-redundant ESTs. *Plant J.*, 9: 101 - 124.

Cotton, R.G.H., Rodriguez, N.R. & Campbell, R.D. 1988. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. U.S.A.*, 85: 4397 - 4401.

Cox, R.A. 1968. The use of guanidinium chloride in the isolation of nucleic acids. In: Grossman, L. & Moldave, K., (eds.). *Methods Enzymol., 12.* Orlando: Academic Press.

Darnell, J.E., Philipson, L., Wall, R. & Adesnik, M. 1971. Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. *Science*, 174: 507 - 510.

Davis, T.M. and Yu, H. 1997. A linkage map of the diploid strawberry. *Fragaria vesca. J. Hered.*, 88: 215 - 221.

Doermann, P., Voelker, T.A. & Ohlrogge, J.B. 1995. Cloning and expression in *E. coli* of a novel thioesterase from *Arabidopsis thaliana* specific for long chain acyl carrier proteins. *Arch. Biochem. Biophys.*, 316: 612 - 618.

Domec, C., Garbay, B., Fournier, M., Bonnet, J., 1990. cDNA library construction from small amounts of unfractionated RNA: Association of cDNA synthesis with polymerase chain reaction amplification. *Anal. Biochem.*, 188: 422 - 426.

Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J.M. 1999. Expression profiling using cDNA microarrays. *Nature Gen. (Supplement)*, 21: 10 - 14.

Dudareva, N., Cseke, L., Blanc, V.M. & Pichersky, E. 1996. Evolution of floral scent in *Clarkia*: Novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell*, 8: 1137 - 1148.

Edwards, K., Johnstone, C. & Thompson, C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.*, 19: 1349.

Ehrenberg, L., Fedorcsak, I. & Solymosy, F. 1976. Diethyl pyrocarbonate in nucleic acid research. *Prog. Nucleic Acid Res. Mol. Biol.*, 16: 189 - 262.

Fargnoli, J., Holbrook, N.J. & Fornace, A.J. 1990. Low-ratio hybridization subtraction. *Anal. Biochem.*, 187: 364 - 373.

Froussard, P. 1992. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Res.*, 20: 2900.

Galau, G.A., Klein, W.H., Britten, R.J. & Davison, E.H. 1977. Significance of rare mRNA sequences in liver. *Arch. Biochem. Biophys.*, 179: 584 - 599.

Germano, J. & Klein, A.S. 1999. Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana* and *P. rubens*. *Theor. Appl. Genet.*, 99: 37 - 49.

Ghareyazie, B., Huang, N., Second, G., Bennet, J. & Khush, G.S. 1995. Classification of rice germplasm. I. Analysis using ALP and PCR-based RFLP. *Theor. Appl. Genet.*, 91: 218 - 227.

Gilpin, B.J., McCallum, J.A. & Frew, T.J. 1997. A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known function and expressed sequence tags (ESTs). *Theor. Appl. Genet.*, 95: 1289 - 1299.

Gordon, J.A. 1972. Denaturation of globular proteins. Interaction of guanidinium salts with three proteins. *Biochemistry,* 11: 1862-1870.

Grellet, F., Cooke, R., Raynal, M., Laudié, M. & Delseny, M. 1993. *Arabidopsis thaliana* systematic cDNA sequencing reveals a gene with homology with *Umbellularia californica* C12:0 ACP thioesterase. *Plant Physiol. Biochem*., 31: 599 - 602.

Gubler, U. & Hoffman, B. 1983. A simple and very efficient method for generating cDNA libraries. *Gene*, 25: 263 - 269.

Gurr, S.J., McPherson, M.J., Scollan, C., Atkinson, H.J. & Bowles, D.J. 1991. Gene expression in nematode-infected plant roots. *Mol. Gen. Genet.*, 226: 361 - 366.

Han, J.H., Stratowa, C. & Rutter, W.J. 1987. Isolation of full-length putative rat lysophospholipase cDNA using improved methods for mRNA and cDNA cloning. *Biochemistry*, 26: 1617 - 1625.

Hara, E., Kato, T., Nakada, S., Sekiya, S. & Oda, K. 1991. Subtractive cDNA cloning using oligo(dT)$_{30}$-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. *Nucleic Acids Res.*, 19: 7097 - 7104.

Hattori, M., Shibata, A., Yoshioka, K. & Sasaki, Y. 1993. Orphan peak analysis: a novel method for detection of point mutations using an automated fluorescence DNA sequencer. *Genomics*, 15: 415 - 417.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S. & Sasaki, T. 1998. A high–density rice genetic linkage map with 2275 markers using a single F$_2$ population. *Genetics*, 148: 479 - 494.

Hauge, B.M. & Goodmann, H.M. 1992. Physical mapping by random clone fingerprint analysis. In: Beckmann, J.S. & Osborn, T.C. (eds.). *Plant genomes: methods for genetic and physical mapping*. London: Kluwer Academic Publishers.

Hawkes, J.G. 1990. *The potato evolution, biodiversity and genetic resources.* Washington DC: Smithsonian Institution Press.

Hayashi, K. 1996. PCR SSCP. Single-strand conformation polymorphism analysis of PCR products. In: Landegren, U. *Laboratory protocols for mutation detection*. Oxford: Oxford University Press.

Hermsen, J.G.T. 1994. Introgression of genes from wild species, including molecular and cellular approaches. In: Bradshaw, J.E. & Mackay, G.R. (eds.). *Potato genetics*. Wallingford: CAB International.

Hernandez, P., Hemmat, M., Weeden, N.F., Dorado, G. & Martin, A. 1999. Development and characterization of *Hordeum chilense* chromosome specific STS markers suitable for wheat introgression and marker-assisted selection. *Theor. Appl. Genet.*, 98: 721 - 725.

Higuchi, R., Von Beroldingen, C.H., Sensabaugh, G.F. & Erlich, H.A. 1988. DNA typing from single hairs. *Nature*, 332: 543 - 546.

Hisada, S., Akihama, T., Endo, T., Moriguchi, T. & Omura, M. 1997. Expressed sequence tags of *Citrus* fruit during rapid cell development phase. *J. Amer. Soc. Hort. Sci.*, 122: 808 - 812.

Höfte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M.F., Charpenteau, J-L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D.Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., Marcillac, G., Gigot, C., Fleck, J., Philipps, G., Axelos, M., Bardet, C., Tremousaygue, D. & Lescure, B. 1993. An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.*, 4: 1051 - 1061.

Holland, P.M., Abramson, R.D., Watson, R. & Gelfand, D.H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' → 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, 88: 7276 - 7280.

Hornes, E. & Korsnes, L. 1990. Magnetic DNA hybridisation properties of oligonucleotide probes attached to superparamagnetic beads and their use in the isolation of poly(A) mRNA from eukaryotic cells. *Genet. Anal. Technol. Appl.*, 7: 145 - 150.

Hoog, C. 1991. Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucleic Acid Res.*, 19: 6123 - 6127.

href1. http://apps.fao.org/cgi-bin/nph-db.pl?subset=agriculture - FAOSTAT agricultural data. - June 1999.

href2. http://www.potatocongress.org/newsletter/Aug99_1html# Biotechnology - Agricultural biotech products on the market. - August 1999.

href3. http://www.bio.org/food&ag/approvedag98.html - Agricultural biotech products on the market. - August 1999.

href4. http://synteny.nott.ac.uk/perl/ace/tree/solgenes - SolGenes, potato MPI1 - April 1999.

href5. http://www.spg.wau.nl/pv/aflp/catalog.htm - A catalogue on chromosome-specific AFLP markers for alignment of potato maps. - April 1999.

href6. http://www.ncbi.nlm.nih.gov/SNP/ - Database of single nucleotide polymorphisms. - September 1999.

href7. http://www.ncbi.nlm.nih.gov/dbSTS/ - dbSTS: database of "Sequence Tagged Sites" - August 1999.

href8. http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary. html - dbEST: database of "Expressed Sequence Tags" - September 1999.

href9. http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast - Basic BLAST search - 1999.

href10. http://www.ncbi.nlm.nih.gov/ gorf/gorf.html - ORF Finder - June 1999.

href11. http://www.perkin-elmer.com - ABI PRISM® 377 sequencer specifications - September 1999.

href12. http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html - The statistics of sequence similarity scores. - April 1999.

href13. http://www.firstmarket/cutter/cut2.html - Webcutter 2.0 - September 1999.

Hussey, P.J. & Hunsperger, J. 1996. cDNA library construction. In: Foster, G.D. & Twell, D. (eds.). *Plant Gene Isolation*. London: Wiley and Sons Ltd.

Jakobsen, K.S., Breivold, E. & Hornes, E. 1990. Purification of mRNA directly from crude plant tissues in 15 minutes using magnetic oligo dT microspheres. *Nucleic Acids Res.*, 18: 3669.

Jepson, I., Bray, J., Jenkins, G., Schuch, W. & Edwards, K. 1991. A rapid procedure for the construction of PCR cDNA libraries from small amounts of plant tissue. *Plant Mol. Biol. Rep.*, 9: 131 - 138.

Joshi, C.P. 1987. An inspection of the domain between putative TATA box and translation start site in 79 plants. *Nucleic Acids Res.*, 15: 6643 - 6653.

Kasha, K.J. 1999. Biotechnology and world food supply. *Genome*, 42: 642 - 645.

Keith, C.S., Hoang, D.O., Barret, B.M., Feigelman, B., Nelson, M.C., Thai, H. & Baysdorfer, C. 1993. Partial sequence analysis of 130 randomly selected maize cDNA clones. *Plant Physiol.*, 101: 329 - 332.

Kieffer, B.L. 1991. Optimised cDNA size selection and cloning procedure for the construction of representative plasmid cDNA libraries. *Gene*, 109: 115 - 119.

Kijas, J.M.H., Fowler, J.C.S. & Thomas, M.R. 1995. An evaluation of sequence–tagged microsatellite site markers for the genetic analysis within *Citrus* and related species. *Genome*, 38: 349 - 355.

Kimmel, A.R. & Berger, S.L. 1987. Preparation of cDNA and the generation of cDNA libraries: overview. In: Berger, S.L. & Kimmel, A.R. (eds.). *Methods Enzymol., 152*. New York: Academic Press.

Ko, M.S.H. 1990. An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs. *Nucleic Acids Res.*, 18: 5705 - 5711.

Kondo, Y., Mori, M., Kuramoto, Y., Yamada, J., Beckmann, J.S., Simon-Chazottes, D., Montagutelli, X., Guenet, J. & Serikawa, T. 1993. DNA segments mapped by reciprocal use of microsatellite primers between mouse and rat. *Mamm. Genome*, 4: 571 - 576.

Kornberg, A. & Baker, T.A. 1992. In: Freeman, W.H. (eds.). *DNA Replication, Second Edition*. New York: Freeman.

Krug, M.S. & Berger, S.L. First-strand cDNA synthesis primed with oligo(dT). In: Berger, S.L. & Kimmel, A.R. (eds.). *Methods Enzymol., Guide to molecular cloning techniques*. New York: Academic Press.

Kurata, N., Nagamura, Y., Yamamoto, K., Harushima, Y., Sue, N., Wu, J., Antonio, B.A., Shormura, A., Shimuzu, T., Lin, S-Y., Inoue, T., Fukuda, A., Shimano, T., Kuboki, Y., Toyama, T., Miyamoto, Y., Kirihara, T., Hayasaka, K., Miyao, A., Monna, L., Zhong, H.S., Tamura, Y., Wang Z-X., Momma, T., Umehara, Y., Yano, M., Sasaki, T. & Minobe, Y. 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet.*, 8: 365 - 372.

Kwak, J.M., Kim, S.A., Hong, S.W. & Nam, H.G. 1997. Evaluation of 515 expressed sequence tags obtained from guard cells of *Brassica campestris*. *Planta*, 202: 9 - 17.

Lagercrantz, U., Ellengren, H. & Andersson, L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and animals. *Nucleic Acids Res.*, 21: 1111 - 1115.

Land, H., Grez, M., Hauser, H., Lindenmaier, W. & Schütz, G. 1981. 5' -Terminal sequences of eucaryotic mRNA can be cloned with high efficiency. *Nucleic Acids Res.*, 9: 2251 - 2266.

Lemmetty, A., Susi, P., Latvala, S., Lehto, K. & Martin, R.R. 1999. Detection of the putative causal agent of blackcurrant reversion disease. *Acta Hort.*, 471: 93 - 98.

Lim, C.O., Kim, H.Y., Kim, M.G., Lee, S.I., Chung, W.S., Park, S.H., Hwang, I. & Cho, M.J. 1996. Expressed sequence tags of Chinese cabbage flower bud cDNA. *Plant Physiol.*, 111: 577 - 488.

Liu, Q., Feng, J., Buzin, C., Wen, C., Nozari, G., Mengos, A., Nguyen, V., Liu, J., Crawford, L., Fujimura, F.K. & Sommer, S.S. 1999. Detection of virtually all mutations-SSCP - A rapid method for mutation scanning with virtually 100 percent sensitivity. *BioTechniques*, 26: 932.

Logemann, Z., Schell, J. & Willmitzer, L. 1987. Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.*, 163: 16 - 20.

Logemann, Z., Mayer, J.E., Schell, J. & Willmitzer, L. 1988. Differential expression of genes in potato tubers after wounding. *Proc. Natl. Acad. Sci. U.S.A.*, 85: 1136 - 1140.

Lowden, S. & Drinkwater, R. 1999. Detection of an SSCP in the 3' untranslated region of the bovine ETV4 gene. *Animal Genet.*, 30: 73.

Marks, R.J. & Brodie, B.B. 1998. *Potato cyst nematodes; biology, distribution and control.* Oxon: CAB Internatioanal.

Mara, M.A., Hillier, L. & Waterson, R.H. 1998. Expressed sequence tags - ESTablishing bridges between genomes. *Trends Genet.*, 14: 4 - 7.

McCombie, W.R., Adams, M.D., Kelley, J.M., FitzGerald, M.G., Utterback, T.R., Khan, M., Dubnick, M., Kerlavage, A.R., Venter, J.C. & Fields, C. 1992. *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nature Genet.*, 1: 124 - 130.

Meissner, P.S., Sisk, W.P. & Berman, M.L. 1987. Bacteriophage λ cloning system for the construction of directional cDNA libraries. *Proc. Natl. Acad. Sci. U.S.A.*, 84: 4171 - 4175.

Milbourne, D., Meyer, R.C., Collins, A.J., Ramsay, L.D., Gebhardt, C. & Waugh, R. 1998. Isolation, characterisation and mapping of simple sequence repeat loci in potato. *Mol. Gen. Genet.*, 259: 233 - 245.

Myers, R.M., Maniatis, T. & Lerman, L.S. 1987. Detection and ligation of single base changes by denaturing gradient gel electrophoresis. In: Wu, R. & Grossman, L. (eds.). *Methods Enzymol, 155.* San Diego: Academic Press.

Moore, G., Gale, M.D., Kurata, N. & Flavell, R.B. 1993. Molecular analysis of small grain cereal genomes: current status and prospects. *Bio/Technology*, 11: 584 - 589.

Moreno-Palanques, R.F. & Fuldner, R.A. 1994. Construction of cDNA libraries. In: Adams, M.D., Fields, C. & Venter, J.C. (eds.) *Automated DNA sequencing and analysis.* London: Academic Press.

Nakata, Y., Tang, X. & Yokoyama, K.K. 1997. Preparation of competent cells for high-efficiency plasmid transformation of *Escherichia coli*. In: Cowell, G & Austen, A. (eds.). *Methods in Molecular Biology, 69: cDNA library protocols*. Totowa: Humana Press Inc.

Nelson, C.E. 1995. The cause of calamity: the discovery of the potato blight in Ireland, 1845 - 1847, and the role of the National Botanic Gardens, Glasnevin, Dublin. In: Dowley, L.J., Bannon, E., Cooke, L.R., Keane, T. & O'Sullivan, E. (eds.). *Phytophthora 150*. Dublin: Boole Press Ltd.

Nelson, T., Harpster, M.H., Mayfield, S.P. & Taylor, W.C. 1984. Light regulated gene expression during maize leaf development. *J. Cell. Biol.*, 98: 558 - 564.

Newman, T., Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. & Somerville, C. 1994. Genes galore: A summary of methods for accessing results from large–scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.*, 106: 1241 - 1255.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, C.J. & Markham, A.F. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.*, 17: 2503 - 2516.

Ohlsson, R. 1989. Growth factors, protooncogenes and human placental development. *Cell. Differ. Development*, 28: 1 - 16.

Okayama, H. & Berg, P. 1982. High–efficiency cloning of full-length cDNA. *Mol.Cell Biol.*, 2: 161 - 170.

Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yakota, T. & Arai, K. 1987. High-efficiency cloning of full-length cDNA; construction and screening of cDNA expression libraries for mammalian cells. In: Wu, R. & Grossman, L. (eds.). *Methods in Enzymology, Recombinant DNA*. San Diego: Academic Press.

Olson, M., Hood, L., Cantor, C. & Botstein, D. 1989. A common language for physical mapping of the human genome. *Science*, 245: 1434 - 1435.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.*, 86: 2766 - 2770.

Parimoo, S., Patanjali, S.R., Shukla, H., Cahplin, D.D. & Weissman, S.M. 1991. cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.*, 88: 9623 - 9627.

Park, Y.S., Kwak, J.M., Kwon, O.Y., Kim, Y.S., Lee, D.S., Cho, M.J., Lee, H.H. & Nam, H.G. 1993. Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single–run partial sequencing. *Plant Physiol.*, 103: 359 - 370.

Patanjali, S.R., Parimoo, S. & Weissman, S.M. 1991. Construction of a uniform–abundant (normalised) cDNA library. *Proc. Natl. Acad. Sci. U.S.A.*, 88: 1943 - 1947.

Perry, D.J. & Bousquet, J. 1998a. Sequence–tagged-site (STS) markers of arbitrary genes: development, characterization and analysis of linkage in black spruce. *Genetics*, 149: 1089 - 1098.

Perry, D.J. & Bousquet, J. 1998b. Sequence–tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. *Theor. Appl. Genet.*, 97: 735 - 743.

Perry, D.J. & Furnier, G.R. 1996. *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proc. Natl. Acad. Sci. U.S.A.*, 93: 13020 - 13023.

Plomion, C., Hurme, P., Frigerio, J.M., Ridolfi, M., Pot, D., Pionneau, C., Avila, C., Gallardo, F., David, H., Neutelings, G., Campbell, M., Canovas, F.M., Savolainen, O., Bodenes, C. & Kremer, A. 1999. Developing SSCP markers in two *Pinus* species. *Mol. Breeding*, 5: 21 - 31.

Provan, J., Powell, W. & Waugh, R. 1996. Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). *Theor. Appl. Genet.*, 92: 1078 - 1084.

Rehbein, H., Mackie, I.M., Pryde, S., Gonzalessotelo, C., Medina, I., Perezmartin, R., Quinteiro, J. & Reymendez, M. 1999. Fish species identification in canned tuna by

PCR–SSCP. Validation by a collaborative study and investigation of intraspecies variability of the DNA–patterns. *Food Chem.*, 64: 263 - 268.

Reynolds, J.E., Head, S.R., McIntosh, T.C., Vrolijk, L.P. & Boyce-Jacino, M.T. 1997. Genetic bit analysis: a solid–phase method for genotyping single nucleotide polymorphisms. In: Caetano-Anollés, G. & Gresshoff, P.M. (eds.). *DNA markers, protocols, apllications, and overviews.* New York: Wiley-VCH.

Riesner, D., Steger, G., Zimmat, R. et al. 1989. Temperature gradient gel electrophoresis of nucleic acids - analysis of conformational transitions, sequence variations, and protein–nucleic acid interactions. *Electrophoresis*, 10: 377 - 389.

Rodriguez, I.R. & Chader, G.J. 1992. A novel method for the isolation of tissue–specific genes. *Nucleic Acid Res.*, 20: 3528.

Rouppe van der Voort, J.N.A.M., Van Eck, H.J., Draaistra, J., Van Zandvoort, P., Jacobsen, E., & Bakker, J. 1998. An online catalogue of AFLP markers covering the potato genome. *Mol. Breeding*, 4: 73 - 77.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim, N. 1985. Enzymatic amplification of β–globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* 230: 1350-1354.

Sambrook, J., Fritsch, E.F. & Maniates, T. 1989. *Molecular cloning. A laboratory manual.* New York: Cold Spring Harbor Laboratory Press.

Sargent, T.D. 1987. Isolation of differentially expressed genes. In: Berger, S.L. & Kimmel, A.R. (eds.). *Methods Enzymol: Guide to molecular cloning techniques.* New York: Academic Press.

Sasaki, T., Song, J., Koga-Ban, Y., Matsui, E., Fang, F., Higo, H., Nagasaki, H., Hori, M., Miya, M., Murayama-Kayano, E., Takiguchi, T., Takasuga, A., Niki, T., Ishimura, K., Ikeda, H., Yamamoto, Y., Mukai, Y., Ohta, I., Miyadera, N., Havukkala, I. & Minobe, Y. 1994. Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant Physiol.*, 6: 615 - 624.

Schuler, G.D. 1998. Electronic PCR: bridging the gap between genome mapping and genome sequencing. *Trends Biotech.*, 16: 456 - 459.

Schmidt, A., Grimm, R., Schmidt, J., Scheel, D., Strack, D. & Rosahl, S. 1999. Cloning and expression of a potato cDNA encoding hydroxycinnamoyl–CoA:tyramine N–(hydroxycinnamoyl)transferase. *J. Biol. Chem.*, 274: 4273 - 4280.

Sela, M., Anfinsen, C.B. & Harrington, W.F. 1956. The correlation of ribonuclease activity with specific aspects of tertiary structure. *Biochem. Biophys. Acta*, 26: 502-512.

Sheffield, V.C., Beck, J.S., Kwitek, A.E., Sandstrom, D.W. & Stone, E.M. 1993. The sensitivity of single–strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics*, 16: 325 - 332.

Sheiness, D. & Darnell, J.E. 1973. Polyadenylic acid segment in mRNA becomes shorter with age. *Natl. New Biol.*, 241: 265 - 268.

Shen, B., Carneiro, N., Torres-Jerez, I., Stevenson, B., McCreery, T., Helentjaris, T., Baysdorfer, C., Almira, E., Ferl, R.J., Habben, J.E. & Larkins, B. 1994. Partial sequencing and mapping of clones from two maize cDNA libraries. *Plant Mol. Biol.*, 26: 1085 - 1101.

Smith, J.S.C. & Smith, O.S. 1990. Fingerprinting in crop varieties. *Adv. Agron.*, 47: 85 - 140.

Soares M.B. 1994. Construction of directionally cloned cDNA libraries in phagemid vectors. In: Adams, M.D., Fields, C. & Venter, J.C. (eds.). *Automated DNA sequencing and analysis.* London: Academic Press.

Srivastava, A.K., McMillan, S., Jermak, C., Shomaker, M., Copelandyates, S.A., Mumm, S., Schlessinger, D. & Nagaraja, R. 1999. Integrated STS Yac physical, genetic, and transcript map of human Xq21.3 to Q23/Q24. *Genomics*, 58: 188 - 201.

States, D.J. & Botstein, D. 1991. Molecular sequence accuracy and the analysis of protein coding regions. *Proc. Natl. Acad. Sci. U.S.A.*, 88: 5518 - 5522.

Strachan, T. & Read, A.P. 1996. *Human molecular genetics.* Oxford: BIOS Scientific Publishers.

Tanksley, S.D., Ganal, M.W., Prince, J.P., De Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Röder, S., Wing, R.A., Wu, W. & Young, N.D. 1992. High–density molecular linkage maps of the tomato and potato genomes. *Genetics*, 132: 1141 - 1160.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.*, 17: 6463 - 6471.

Taylor, J.L., Fritzemeier, K-H., Häuser, I., Kombrink, E., Rohwer, F., Schröder, M., Strittmatter, G. & Hahlbrock, K. 1990. Structural analysis and activation by fungal infection of a gene encoding a pathogenesis–related protein in potato. *Mol. Plant-Microbe Interactions*, 3: 72 - 77

Taylor, M.A., George, L.A., Ross, H.A. & Davies, H.V. 1998. cDNA cloning and characterization of an α–glucosidase gene from potato (*Solanum tuberosum* L.). *Plant J.*, 13: 419 - 425.

Tragoonrung, S., Kanazin, V., Hayes, P.M. & Blake, T.K. 1992. Sequence–tagged–site-facilitated PCR for barley genome mapping. *Theor. Appl. Genet.*, 84: 1002 - 1008.

Tsumura, Y., Suyama, Y., Yoshimura, K., Shirato, N. & Mukai, Y. 1997. Sequence–tagged–sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers. *Theor. Appl. Genet.*, 94: 764 - 772.

Uchimiya, H., Kidou, S., Shimazaki, T., Aotsuka, S., Takamatsu, S., Nishi, R., Hashimoto, H., Matsubayashi, Y., Kidou, N., Umeda, M. & Kato, A. 1992. Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). *Plant J.*, 2: 1005 - 1009.

Van Gelder, W.M.J., Vinke, J.H. & Scheffer, J.J.C. 1998. Steroidal glycoalkaloids in tubers and leaves of Solanum species used in potato breeding. *Euphytica*, 39: 141 - 145.

Van Ness, J., Maxwell, I.H. & Hahn, W.E. 1979. Complex population of nonpolyadenylated messenger RNA in mouse brain. *Cell*, 18: 1341 - 1349.

Voelker, T.A., Worrell, A.C., Anderqson, L., Bleibaum, J., Fan, C., Hawkins, D.J., Radke, S.E. & Davies, H.M. 1992. Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. *Science*, 257: 72 - 74.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van der Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. & Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.*, 23: 4407 - 4414.

Wang, Z. & Brown, D.D. 1991. A gene expression screen. *Proc. Natl. Acad. Sci. U.S.A.*, 88: 11505 - 11509.

Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierrymieg, J. & Sulston, J. 1992. A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genet.*, 1: 114 - 123.

Weissenbach, J. 1993. A second generation linkage map of the human genome based on highly informative microsatellite loci. *Gene*, 135: 275 - 278.

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

Werner, D., Chemla, Y. & Herzberg, M. 1984. Isolation of poly(A)+ RNA by affinity chromatography. *Anal. Biochem.* 141: 329-336.

Wilcox, A.S., Khan, A.S., Hopkins, J.A. & Sikela, J.M. 1991. Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nucleic Acids Res.*, 19: 1837 - 1843.

Williams, J.G., Kubelik, A.R., Livak, J., Rafalski, J.A. & Tingey, S.V. 1990. DNA polymophisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, 18: 6531 - 6535.

Wu, R., Wu, T. & Ray, A. 1987. Adaptors, linkers, and methylation. In: Berger, S.L. & Kimmel, A.R. (eds.). *Methods Enzymol.: Guide to molecular cloning techniques*. New York: Academic Press.

Yip, S.P., Hopkinson, D.A. & Whitehouse, D.B. 1999. Improvement of SSCP analysis by use of denaturants. *BioTechniques*, 27: 20 - 24.