ISOLATION AND CHARACTERIZATION OF AN RNA POLYMERASE III ENCODED GENE OF *Pinus radiata* AND ITS USE IN PINE TRANSFORMATION

M.ANANDASAYANAN

A THESIS ON STUDIES CONDUCTED AT THE DIVISION OF MOLECULAR GENETICS, UNIVERSITY OF WAIKATO IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN MOLECULAR BIOLOGY

DIVISION OF MOLECULAR GENETICS

UNIVERSITY OF WAIKATO

HAMILTON

NEW ZEALAND

2006

Abstract

Several promotors such as the cauliflower mosaic virus 35S promoter (CaMV 35S) and its enhanced version, pEMU, the maize ubiquitin promoter, the alcohol dehydrogenase promoter (Adh) and rice actin promoter (Act1) are currently used in *Pinus radiata* (pine) transformation. These heterologous promoters were adopted for pine transformation for want of an endogenous promoter tailored specifically for the needs of pine. These promoters may not perform to the same extent in pine as in their homologous systems due to differences in quality and/or quantity of regulatory factors. Secondly, because of their heterologous origin, these promoters are open to silencing mechanisms that operate in plants against invasive DNA [Matzke & Birchler, 2005]. This could result in inactivation of these promoters at any time during the 30-year growth period of transformed pine which this poses a real threat to a forestry industry based on transgenic pine.

A pine promoter on the other hand, being endogenous, is less prone to silencing. In addition, confidence in its longevity (continued expression) can be easily established even before using it in transformation. The aim of this study was to isolate and validate pine promoters that can be used in pine transformation. As only a few pine sequences were available in the public domain for gene discovery in pine (at the beginning of this study), heterologous sequence information was used to screen the pine genome or its transcriptome for orthologs with desirable expression features.

The investigation proceeded along two lines. In the first approach, a putatively desirable gene was isolated and the expression profile of its promoter was then validated. This led to the characterization of *5Spr20*, a pine 5S rDNA paralog. *5Spr20* differs from all published 5S rDNA sequences and is therefore a novel pine gene. Analyses of its sequence using bioinformatics revealed that it is capable of initiating biologically active transcripts and *5Spr20* is therefore a functional gene. A recombinant *5Spr20* promoter consisting of the coding region and the immediately upstream region downregulated *gus* reporter activity by 90% by antisense activity in transient

expression studies in pine embryogenic cells. In stable expression studies, a *5Spr20* promoter-driven shDNA construct targeting *gus* completely silenced reporter activity in the model plant *Nicotiana benthamiana*. The *5Spr20* promoter appears to hold great promise for use in pine functional genomics and in gene downregulation applications.

In the second line of investigation, the expression profiles of pine orthologs of known heterologous genes were validated prior to gene isolation. Two pine genes that were identified as promising candidates are pine $tDNA^{Met\text{-}I}$ and an actin paralog pine, ActX. Both genes were strongly expressed in all vegetative tissues of pine. Several PCR-based methods were used to clone the upstream regions (containing putative promoter elements) but all attempts ended in failure, which is attributed to the presence of pseudogenes and regions homologous to walking/sequencing primers among paralogs.

The pine transcriptome was also screened unsuccessfully for ortholgs of desirable heterologous candidate genes like the ribosomal protein genes *MsRL5* of *Medicago sativa* and *AtL18* of *Arabidopsis thaliana* and genes for the second largest subunit of RNA polymerase II, gene T13794 and actin-2 of *A. thaliana*. Sequence heterogeneity, cell-specific expression and low transcript abundance are possible reasons for not being able to detect pine orthologs of these candidate genes in expression screens.

ACKNOWLEDGEMENTS

I am indebted to Professor.Richard Wilkins who guided me admirably through the difficult phases of my research. The experience I gained in his laboratory, whose activities range from P53 mutation analyses to microarray assay of bovine lactiferous expression and extremophile protein profiling has widened my horizons. I have gained more knowledge here than I have during my 12-year Molecular Biology career.

I am also thankful to Drs. Dave Musgrave and Ray Cursons of the Molecular Genetics academic staff who have been wonderful in discussions and brainstorming. I appreciate the help given by Dr.Barry O'Brien during the initial phases of my *in situ* hybridization work. I owe my thanks to Raewyn Towers, Laboratory Manager, for her continued support.

Finally, I wish to thank all members of Professor Wilkin's research group, especially Gregory Jacobson, Ivy Zhang, Greg Bailey, Tao Xie, Marcel Dinger, Linda Peters and Robert Brinzky, whose companionship and camaraderie were invaluable.

Thanks are also owed to Fletcher Challenge Forests Ltd (Molecular Biology Laboratory) and Carter Holt Harvey Forests Ltd (Tissue Culture Facility).

List of abbreviations

′ - Prime or Minutes
“ - Seconds
Act - Actin
AsRNA - Antisense RNA
BHLH - Basic Helix-turn-helix
CaMV 35S – Cauliflower Mosaic Virus 35S promoter
CDK - Cyclin dependant kinase
DCL - Dicer-like
DCR - Dicer
DEPC - Diethyl pyrocarbonate
ETS - External transcribed spacers
GTF - General transcription factor
HMG - High mobility group protein
hpRNA - Hairpin RNA
HTH - Helix-turn-helix
ICR - Internal control region
IGS - Intergenic spacers
ihpRNA - intron spliced hairpin RNA
miRNA - Micro RNA
NTS - Non-transcribed spacer
PIC - Preinitiation complex
Pol – RNA polymerase
PTGS - Post transcriptional gene silencing
rDNA - Ribosomal DNA
RdRp - RNA dependant RNA polymerase
RISC - RNA induced silencing complex
RLC - RISC loading complex
ShRNA - short-hairpin RNA
siRISC - Small interfering RNA-RISC
SiRNA - Small interfering RNA
SnRNA - Small nuclear RNA
SRB - Suppressor of Pol II truncation
TAF - transcription associated factors
TBP - TATA binding protein
TF - Transcription factor
TRNA - Transfer RNA
Ubi - Ubiquitin
UAS - Upstream activator sequences
UCE - Upstream control elements
UTR - Untranslated region
Vir - Virulence

Contents

List of figures

List of tables

CHAPTER I

INTRODUCTION

The flow of genetic information in organisms is regulated by processes operating during transcription, splicing, RNA transport, translation, post-translation and protein compartmentalization [Knowlton, 1995]. The primary stage at which gene expression is controlled in most eukaryotes is at transcription initiation [Holstege & Young, 1999]. This is reflected in the percentage of the genome (>20%) dedicated to encoding transcription factors in plants and other eukaryotes [Johnson, 1995; Bevan *et al*, 1998]. One of the major elements pivotal to successful transcription initiation is the gene promoter [Roa-Rodriguez, 2003]. It often consists of a minimal promoter responsible for basal transcription and *cis*-acting elements, which account for activated transcription.

Over the past twenty years, several promoters have been used to transform eukaryotic and prokaryotic cells [Potrykus & Spangenburg, 1996], an early example of which is the mobilization of recombinant plasmid features in bacteria using the *lac* operon promoter. Some of these promoters have been tailored to meet specific requirements of different plant groups, such as the maize *ubi* (-quitin) and rice actin (*Act1*) promoters for the transformation of monocotyledonous plants (patent US 5-510474, US 5-641876), opine and cauliflower mosaic virus (CaMV 35s) promoters (patent US 5-955646, US 5-352605) for dicotyledonous plants and the *rolC* promoter (patent application US2004/0168214) of *Agrobacterium rhizogenes* for phloem-specific expression in dicotyledonous plants [Noury *et al*, 2000; Heifetz *et al*, 2000].

Specific promoters that cater for tissue-specific or global expression in *Pinus radiata* (hereafter referred to as pine) transformation are not available in pine research. Any promoters that may have been developed for this purpose [Wood *et al*, 2000; Kirst *et al*, 2003; Egertsdotter *et al*, 2004] are clothed in secrecy as a result of their patentability. Pine transformation thus continues to be reliant on the use of the CaMV 35S promoter. Being a heterologous

promoter and of viral origin, CaMV 35S has problems of its own such as weak expression in certain tissues [Walter *et al*, 1997] and a tendency to be silenced in some crops such as cotton towards the tail-end of the vegetative phase [Llewellyn, 2003]. Therefore, there is a need to develop specific promoters for global or tissue-specific gene expression in pine. Such promoters are preferably obtained from the pine genome itself. The aim of this study is to isolate one or more pine promoters and evaluate their use in pine transformation by a forward or reverse genetics approach.

An appreciation of the factors determining the functionality of a promoter is essential in order to make a judicious choice of promoters and their subsequent manipulation. A promoter does not act in isolation to effect transcription but rather in concert with a host of other factors. The topology of the chromatin, transcription factors (TF) and cell cycle stage and epigenetic changes all play an equally important role in influencing transcription initiation [Mathieu *et al*, 2002]. This chapter reviews the various factors, whose interplay with one another and the promoter determine the efficiency of the promoter and its usefulness in transformation. The dissertation examines the various nuclear-imported proteins involved in transcription, the role of promoter elements, the cooperative interaction between the two which results in basal and activated transcription, the practical aspects of evaluating the biological efficacy of promoters and the use of promoters in functional genomics.

1. PROTEINS INVOLVED IN TRANSCRIPTION INITIATION

1.1. RNA polymerases

Eukaryotes differ from most prokaryotes where a single (core) polymerase transcribes all classes of genes. In eukaryotes, three DNA-dependant RNA polymerases transcribe subsets of genes, which differ in their gene structure and functions. A fourth RNA polymerase (Pol IV) is associated with a small interfering RNA (siRNA) pathway [Herr *et al*, 2005] which silences endogenous DNA.

The three DNA-dependent RNA polymerases differ in the class of genes they transcribe, the regulatory sequences they require for transcription initiation, their location in the nucleus and their susceptibility to inhibitors. On the other hand, they show affinities as in the shared subunits that make up these multi-subunit proteins, the TATA binding protein (TBP) which is required by all three as a component of the transcription complex [Huang & Maraia, 2001] and promoter elements which serve equally well for more than one polymerase [Margottini *et al*, 1991; Fung *et al*, 1995].

RNA polymerase I (Pol I) transcribes ribosomal DNA (rDNA) and is associated with the nucleolar organizing regions [Burton *et al*, 2005]. RNA Polymerase II (Pol II) resides in the nucleoplasm and is responsible for the synthesis of heterogeneous nuclear RNA - the precursor of messenger RNA (mRNA), intronic miRNA [Ying & Lin, 2004], small nucleolar RNA (snoRNA) and the U1-U6 small nuclear RNA (snRNA) of the spliceosome [Bartel, 2004]. RNA Polymerase III (Pol III) is localized in the nucleoplasm and transcribes transfer RNA (tRNA) and 5S rRNA, which have triphosphate 5′ ends that are not capped as in mRNAs and small nuclear RNA such as exonic miRNA [Ying & Lin, 2004], 7SK, 7SL and U4 snRNA [Reddy, 1988] and *Alu* sequences and adenovirus VA gene [Huang & Marai, 2001].

1.2. Transcription factors

Transcription factors are proteins that participate directly in the initiation of transcription, but which are not themselves part of RNA polymerase. They influence transcription by binding to specific DNA sequence motifs [Quandt *et al*, 1995], or target sequences in other transcription factors often by allosteric interaction [Wolffe, 1991]. Four classes of transcriptional factors are recognized; the general transcription factors (GTF) required for basal transcription [Reinberg *et al*, 1998], the activators and repressors, which bind to *cis*-acting elements, the coactivators and corepressors, which mediate transcriptional effects of activators and repressors and the architectural transcription factors involved in chromatin remodeling [Tsukiyama & Wu, 1996].

1.2.1. DNA binding domains of transcription factors

Specific DNA binding in most cases appears to result from cooperative and additive interactions between a DNA sequence and specific protein structural motifs; while one motif makes contact with a sequence-independent structure of the DNA such as the sugar-phosphate backbone, an associated structural motif provides specific binding contacts with base pairs in the major groove of the DNA binding site [Wolberger *et al*, 1991; Pabo & Sauer, 1992; Rhees *et al,* 1998].

Several categories of DNA binding motifs have been identified in transcription factors. The helix-turn-helix (HTH) motif consisting of two α helices separated by a turn of several amino acids is found in the engrailed repressor protein (EN) of *Drosophila* [Scott *et al*, 1998], the maize homeodomain protein knotted I (KN1) [Vollbrecht *et al*, 1991], KN1-like proteins such as rice OSH1, barley HVKNOX3 and *Arabidopsis* KNAT1 [Kerstetter *et al*, 1994] and most plant MYB proteins. Zinc finger motifs contain loops formed by coordination of a zinc atom with cystein and/or histidine (C/H) residues; the zinc fingers of transcription factor IIIA (TFIIIA) and petunia EPF family proteins (C_2H_2 type fingers) contain tandem repeats of the motif C-$X_{2\text{ or }4}$-C-X_3-F-X_5-L-X_2-H-X_3-H [Lee *et al*, 1989; Lewin, 1997] while the intracellular, steroid receptor proteins have a motif with the zinc binding consensus sequence C-X_2-C-$X_{1\text{-}3}$-C-X_2-C [Lewin, 1997]. Other common DNA binding motifs are the leucine zippers of opaque-2 (O2) of maize, HBP-1 of wheat and PosF21 of *Arabidopsis* [O'Neil *et al*, 1990; Foster *et al*, 1994], which possess a dimerization interface and the helix-loop-helix (HLH) motif of several plant bHLH protein regulators of anthocyanin biosynthesis [Koes *et al*, 1994].

Some DNA binding domains are unique to plants; the APETALA2 (AP2) and the related ethylene-responsive element binding protein (EREBP) multigene families [Riechmann & Meyerowitz, 1998] contain a conserved ~ 60-70 amino acid, AP2 DNA binding domain [Okamuro *et al*, 1997]. So are the family of viviparous1-related factors (VP1) which includes maize VP1 [Hill *et al*, 1996],

Arabidopsis ARF1 [Ulmasov *et al*, 1997] and monopteros proteins [Hardtke & Berleth, 1998].

1.2.2. Transcription regulatory domains of transcription factors

Transcription factors may also possess transcription regulatory domains responsible for positive or negative regulation of transcription; they exercise their activity when they are positioned in the vicinity of the target area. Such motifs may be an integral part of the transcription factor, or of an interacting protein which together with the bound protein forms a functional transcription factor [Holstege & Clevers, 2006]. The DNA binding and transcription regulatory domains are not interdependent as shown by swapping heterologous domains to produce chimaeric factors [Ptashne, 1989] in the yeast two-hybrid system [Fields & Song, 1989].

Acidic active domains exhibiting a relative abundance of negatively charged amino acids are present in maize opaque 2 (O2), viviparous 1 (VP1) and tomato heat shock (HSF) factors [Hope *et al*, 1988; Tamaoki *et al*, 1995]. Glutamine rich domains confer transcriptional activation to some factors; insertion of homopolymeric tracts of glutamine into GAL4/VP16 transcription factor resulted in 14-fold enhanced transcription [Gerber *et al,* 1994: Schwechheimer *et al*, 1998]. Similarly, the proline-rich region of plant G-box binding factor 1 (GBF1) was shown to be a potential activation domain [Schindler *et al*, 1992].

2. PROMOTER ELEMENTS

Most gene promoters are modular in nature. The core promoter is the minimal nucleotide sequence needed to assemble the general transcription factors and polymerase, specify the start site and initiate basal transcription [Washburn *et al*, 1997]. The upstream control elements, which occur at relatively fixed distances from the core promoter influence the efficiency and specificity of transcription. Apart from these generalities, the promoters of the three classes of genes transcribed by the different RNA polymerases differ in their architecture.

2.1. Promoter elements of Class I genes

The class I genes transcribed by Pol I are arranged as long, head-to-tail repeats separated by 2-30 kb long intergenic spacers (IGS). A typical repeat unit consists of the 17/18S, 5.8S and 25/28S coding sequences in that order, separated from one another by internal transcribed spacers (ITS) and flanked at the repeat ends by external transcribed spacers (ETS). The sequence of coding regions are highly conserved among eukaryotes while the ITS show little sequence homology [Jacob, 1995; Burton *et al*, 2005].

Deletion analysis and linker scanning in human and murine rDNA show that the sequence -45 to +18 and -39 to +9, respectively, forms the start site proximal core promoter responsible for accurate transcription. The regions -156 to -107 and -142 to -112, respectively, form the upstream control element (UCE) which acts cooperatively with the core promoter to increase transcription [Haltiner *et al*, 1986; Jones *et al*, 1988; Schnapps *et al*, 1999].

The core promoter in *Xenopus* is located between -147 and +4 positions. This sequence is duplicated two to six times within the IGS [DeWinter & Moss, 1987]. Separating the duplicated spacer promoters are regions of intermingled and related 60 or 81 bp repeat elements [Busby & Reeder, 1983], which occur in blocks of 6-12 units and share a 40 bp sequence with the core promoter repeats; the 40 bp sequence enhances efficiency of transcription initiation at the start site.

In *D. melanogaster* the core promoter spans the region -43 to +20. The IGS immediately upstream of the start site contains five to twelve 240 bp repeats, which overlap with spacer promoters. The repeats contain a 70 bp element, which includes a perfect copy of the -24 to +10 domain of the core promoter that stimulates transcription [Grimaldi *et al*, 1990; Jacob, 1995].

2.2. Promoter elements of Class II genes

Pol II transcribes primarily genes coding for proteins. A generalized gene would consist of a nontranscribed 5′ flanking region followed by an

untranslated region (5′ UTR), an open reading frame (ORF) often interrupted by introns and a 3′ UTR.

Three elements which constitute the core promoter are the TATA element, initiator element (Inr) and downstream promoter element (DPE) (Reinberg *et al*, 1998]. Many mammalian Pol II genes contain the characteristic TATA box located ~30 bp from the start site [Arkhipova, 1995] with the consensus sequence TATAAAA; in the lower eukaryote yeast, this occurs 40 to 120 bases upstream of the initiation site [Struhl, 1995]. The TATA motif in plants is of the consensus sequence CTATAAA(A/T)A [Washburn *et al*, 1997]. About 50 bases upstream at position -75 another promoter element, the CAAT box with the consensus sequence GGCCAATCT is found in several genes; This is replaced in plants by an AGGA box [Roa-Rodriguez, 2003]. A third conserved element observed in some genes is the GC box with the sequence GGGCGG on the nontemplate strand, which is recognized by factor SP1. The CAAT and GC box sequences may occur in either orientation. The 8 bp octamer element with the consensus sequence ATTTGCAT is another ubiquitous upstream element. Although it is generally stated that promoters recognized by Pol II contain a TATA box and at least one other important element upstream there are many exceptions and indeed, about one half of the known Class II promoters in *D. melanogaster* are TATA-less [Arkhipova, 1995]. The proximal promoter region of some genes such as *C-eta* and *Cdc2LI* genes is essential for gene expression even though they lack the canonical TATA, CAAT motifs and GC rich regions [Quan & Fisher, 1999; Kahle *et al*, 2005].

Several TAF subunits recognize the core promoter element Inr found in numerous genes [Lago *et al*, 2004] where it overlaps the transcription initiation site [Kaufman & Smale, 1994]. It consists of a pyrimidine-rich sequence with the approximate consensus sequence YY(A+1)N(T/A)YY [Javahery *et al*, 1994]. In *D. melanogaster*, Inr occurs in the region -5 to +5 [Arkhipova, 1995]; in plants it is located between -9 and +4 [Washburn *et al*, 1997]. In TATA box-deficient promoters, Inr is functionally analogous to TATA and capable of directing basal transcription from the precise start site [Martinez *et al*, 1994]. In

many TATA-containing genes, the presence of Inr enhances promoter strength [Kaufmann & Smale, 1994].

Several TATA-less genes contain a strand-specific, seven nucleotide downstream promoter element (DPE). In *D. melanogaster* it occurs 20 to 30 bp downstream of the start site [Burke & Kadonaga, 1997]. A ±3 nucleotide change in distance between DPE and Inr resulted in a seven- to eight-fold reduction in transcription and reduced binding of TFIID. Mutational analysis of the DPE motif (A/G)G(A/T)CGTG showed the importance of this conserved sequence to basal transcription of these genes [Burke & Kadonaga, 1996].

A later addition to the list of promoter elements is an element, which is recognized directly by TFIIB [Lagrange *et al*, 1998]. Positioned immediatly upstream of the TATA element, the TFIIB recognition element (BRE) has a putative consensus sequence (G/C)(G/C)(G/A)CGCC. It could play a part in determining directionality of the preinitiation complex [Reinberger *et al*, 1998].

2.3. Promoter elements of class III genes

The classical Type I and II genes transcribed by Pol III contain intragenic promoter elements. Type III genes contain promoters exclusively upstream of the coding sequence and still others are characterized by both intra- and extragenic elements [Huang & Maraia, 2001]. 5S rDNA (Type I gene) is of particular interest to this study and therefore it will be reviewed in greater detail.

2.3.1. 5S rDNA

Most 5S rRNA genes are arranged in tandemly repeated arrays consisting of a conserved 120 bp gene alternating with a nontranscribed intergenic spacer (IGS).

Brown and co-workers [Bogenhagen *et al*, 1980; Sakonju *et al*, 1980] demonstrated that the promoter of the 120 bp *Xenopus laevis* oocyte 5S rDNA resides within approximately 34 bp of its coding region. Plasmid constructs

containing progressive deletions of the cloned gene extending from its 5′ or 3′ flanking region into the gene were tested for transcriptional activity. Deletions in the region +1 to +50 in the 5′ half and +90 to +120 in the 3′ half of the gene did not affect transcription and a transcript of the correct size (120 nt) extending into the vector sequence was produced. Transcription was abolished when the deletions extended further inwards from either side.

In *X.laevis* 5S rDNA, this internal control region (ICR) resides in a region spanning positions +50 to +90 [Yang & Heyes, 2003]. Linker scanning (LS) mutagenesis and point mutation analysis demonstrate that the ICR is multipartite, consisting of three functional domains - The A box covering positions +50 to +64 which is the conserved class III promoter domain, a second 5S gene-specific domain, the C box encompassing base pairs +80 to +90 and an intermediate element (IE or I box) between positions +67 and +72 [Pieler *et al,* 1985, 1987].

In addition, 5S rRNA genes contain elements in their flanking regions, which modulate transcription and conserved nucleotides within their coding region which are critical to transcription [Wolffe, 1994]. The 5' flanking sequences are an absolute requirement 5S rDNA transcription in many species such as *Arabidopsis thaliana* [Cloix *et al*, 2003], *Bombyx mori* [Morton & Sprague, 1984], *Neurospora crassa* [Selker *et al*, 1986], *D. melanogaster* [Sharp & Garcia, 1988], *homo sapiens* [Hallenburg & Fredriksen, 2001], nematodes [Nelson *et al*, 1998] and loach [Felgenhauer et al, 1990].

Wormington *et al* [1981] detected a 2-fold reduction in transcription upon substituting nucleotides -26 to -11 upstream of *X. borealis* somatic 5S rDNA. This upstream region includes a conserved pentamer motif AAAGT (-18 to -14) overlapping a moderately conserved element Y-RRRR, both of which are found in 5S genes of other *Xenopus* species. Similar template titration studies [Oei & Pieler, 1990] with *X. laevis* genes, where nucleotides -34 to +5 were replaced with plasmid elements resulted in a 4-fold reduction in transcription efficiency.

Conserved upstream sequence elements are also found in human, hamster and mouse 5S rDNA [Nielsen *et al*, 1993]. The D-box, GGCTCTTGGGGC occurs at position -32 to -21 in all three species and its deletion reduces transcription efficiency *in vitro* to 10%. Three human genes possess the Sp1 binding sequence GGGCGG at positions -43 to -38 and another three contain an Sp1-like sequence GGGCCG in this region. All six genes contain another Sp1 binding site at position -245 and an activating transcription factor (ATF) [Bredow *et al*, 1990] recognition site at -202. In addition, Arnold *et al* [1987] found that transversion of the guanine residues at positions +87 and +89 resulted in a 8-17% reduction in transcription. Similar effects were observed in *X. laevis* [Pieler *et al*, 1985] and *D. melanogaster* [Sharp & Garcia, 1988].

D. melanogaster mutants lacking a TATA motif at -39 to-26 show less than 0.05% transcriptional activity. Nineteen genes in a cluster of twenty-four 5S rRNA genes containing a transition at position +86 were also transcriptionally inactive *in vitro.* [Sharp *et al*, 1984]. The *Drosophila* ICR deviates from the *Xenopus* paradigm in comprising four essential sequence elements located at +3 to +18, +37 to +44, +48 to +61 and +78 to +98 [Sharp & Garcia, 1988].

An analysis of 5S rDNA transcription in *A. thaliana* reveals that in addition to the tripartite ICR, major transcription control regions consisting of a TATA-like motif at positions -28 to -23, a GC sequence at -12 to -11, an AT rich region at -4 to -2 and a C residue at -1 are essential for accurate and efficient transcription [Cloix *et al*, 2003]; the TATA motif is important for re-inition rather than for single-round basic transcription. A TATA motif centred at -26 is also present in 5S rDNA of some other plants [Rafalski *et al*, 1982; Ellis *et al*, 1988; Hemleben & Werts, 1988]; in *Matthiola* & *Vigna* [Ellis *et al*, 1988; Hemleben & Werts, 1988] a conserved CCATATAT region occurs at position -31 to -24. However, most plant 5S rDNA do not contain a TATA box; rather, their common feature is the A/T content [Gerlach & Dyer, 1980; Frasch *et al*, 1989]. A generalized format for plant 5S rDNA extragenic promoter elements proposed by Venkateswarlu *et al* [1991] consists of a conserved C base at position -1, a G/C rich region centred at -13 and an A/T rich element centred at -26. The universality of the cytosine nucleotide one base before the start site

suggests that it may play a role in transcription initiation [Challice & Segall, 1989]; mutation of C to T decreases transcription by 70% in *A. thaliana* [Cloix *et al*, 2003]. *Nicotiana tabacum* is an exception where C is substituted by T in most sequences [Fulnecek *et al*, 2002].

Neurospora crassa 5S rDNA exhibits a TATA element (C/T)ATA(G/A) between -29 and -25 [Morzycka *et al*, 1985], which determines the transcription initiation site [Tyler, 1987]. Similarly, an absolute requirement for a critical TATAT element (-28 to -24) has been reported in *B. mori* [Morton & Sprague, 1984].

2.3.2. tDNA and other Type II genes

The tDNA bipartite promoter consists of well separated, 10 bp A- and B-boxes which are the cognate sequences for binding of TFIIIB and TFIIIC [Sprague, 1992; Hasegawa *et al,* 2003]. The A-box is invariantly intragenic and has the consensus sequence TGGCNNAGTGG; the consensus sequence of B-box, which is mostly intragenic is GGTCGANNC. The sequence between A- and B-boxes is not critical but the distance is important; transcription is prevented if the distance exceeds 100 bp, or the boxes are fused. Upstream sequences such as a CAA triplet at the transcription start site and a region of low duplex stability ~30 bp upstream of the coding sequence also appear to affect start site selection as well transcription efficiency [Yukawa *et al*, 2000; Hasegawa *et al*, 2003]. In the presence of upstream elements, deletion of the entire B Box and part of the A box, still allowed faithful initiation of the $tRNA^{Ala}$ gene in *B. mori* [Larson *et al*, 1985] and $tRNA^{Ser}$ in *X. laevis* [Park *et al*, 1995].

2.3.3. Type-III genes

Pol III also transcribes other small nuclear RNAs (snRNA) which do not have an ICR; instead their promoters lie in the 5′ flanking region and some have TATA motifs. Prototypes of this group include the human 7SK gene and the metazoan U6 snRNA [Mattaj *et al*, 1988; Lobo & Hernandez, 1989; Bredow *et al*, 1990]. Promoters of these genes contain a conserved proximal sequence element (PSE) at around position -60 and a TATA sequence near position -30,

which curiously, is the key determinant for pol III specificity [Hernandez *et al*, 1990].

3. ENHANCER ELEMENTS

Enhancers are *cis*-elements which stimulate (or repress) transcription. They differ from most other promoter elements in being orientation- and position-independent of the start site [Muller *et al,* 1988]; they can function from upstream, or downstream of the CDS, or downstream of the cap site, or from within the transcription unit [Gidekel *et al*, 1996]. A typical model is the intron-located enhancer of immunoglobulin kappa gene (V_K) [Queen & Baltimore, 1983], which lies 3 kb away from the gene and functions in either orientation to stimulate transcription from the V_K promoter. The enhancer effect is independent of intervening promoters and not altered when the distance between the promoter and enhancer is varied from 1.7 kb to l7.7 kb [Atchison & Perry, 1986].

Enhancer elements conferring spatial, temporal, or induced expression have been identified by loss/gain-of-function analysis [Donald & Cashmore, 1990; Atchinson & Perry, 1986; Murphy *et* al, 1989; Ellestrom *et al*, 1996; Gidekel *et al*, 1996; Ouwerkerk *et al*, 1999]. When constructs containing these elements are expressed in plants, they often duplicate their characteristic pattern of expression [Benfey *et al*, 1990b]. The looping model proposed [Picard & Schaffner, 1984] to explain stimulation of transcription at the promoter by a remote enhancer has taken precedence over a scanning model proposed by Moreau *et al* [1981].

4. PROMOTER EXPRESSION

The value of a promoter lies in its ability to initiate transcription, which culminates in the production of biologically active transcripts. This in turn depends on the presence of the appropriate promoter elements, transcription factors and a favourable chromatin configuration at the site of transgene insertion. The presence or absence of these factors determines the spatial, temporal, or null expression of a promoter. Therefore, a review of the process

transcription will be useful in interpreting the expression characteristics of promoters used in this study. Transcription is conveniently studied in two stages, namely,

(1) Basal transcription

(2) Activated transcription

4.1. Basal transcription

Transcription is preceded by the formation of a stable, binary preinitiation complex (PIC) at the core promoter, whose primary function is to provide RNA polymerase specificity to the gene and position the enzyme at the appropriate site.

4.1.1. Transcription initiation in Class I genes

Two transcription factors, the upstream binding factor (UBF-1) and the promoter selectivity factor for Pol I (SL1 in human) recruit Pol I to these promoters [Leblanc *et al*, 1993; Pikaard, 1994; Putnam *et al*, 1994]. Vertebrate SL1 comprises TBP and three TBP associated factors [Comai *et al*, 1994]. In humans, UBF-1 by itself binds to and protects positions -75 to -115; it also contacts the sequences around -21 in the core promoter. In the presence of UBF-1, SL1, which by itself does not exhibit sequence-specific DNA binding activity, extends the region protected in DNase-1 footprints from -165 to +1. This results in the formation of a binary complex that binds UCE and the core promoter and facilitates binding by Pol I and the Pol I associated factors TIF-IA and TIF-IC. SL1 probably confers polymerase specificity. Several other factors such as the core promoter binding factor (CPBF) may be also required for transcription initiation (Jacob, 1995).

4.1.2. Transcription initiation in Class II genes

Pol II requires at least six general transcription factors (GTF) for accurate initiation of transcription in humans [Reinberg *et al*, 1998]. Of these, TFIID consisting of TBP and 10-18 TAFs [Lago *et al*, 2004] is the core promoter recognition factor possessing a site-specific, DNA binding capacity. The initial step in PIC formation is the binding of TFIID to the TATA box by

means of TBP [Nikolov *et al*, 1992]. TFIIA associates with TBP to stabilize the TFIID-DNA complex. Subsequently, TFIID recruits TFIIB by interaction involving TBP and the amphipathic α helix domain of TFIIB. This DAB complex serves as a scaffold for Pol II escorted by TFIIF to enter the complex; Pol II is contacted by specific residues of TFIIB, while contacts with TFIIF is made by the zinc finger of TFIIB [Ha *et al,* 1995] This, is followed by the sequential association of TFIIE and TFIIH to complete PIC formation [Maldonado & Reinberg, 1995].

In more simplified systems, accurate transcription is achieved with just TBP and TFIIB alone [Parvin *et al*, 1994]; some *Drosophila* genes are transcribed by a subset of the GTFs [Drapkin *et al*, 1995]. However, in the case of TATA-less, Inr-containing promoters basal transcription has an absolute requirement for GTFs to recruit TFIID (or TBP) to the core promoter and additional factors such as TFII-I, which bind directly to Inr elements [Martinez *et al*, 1994].

4.1.3. Transcription initiation in Class III genes

4.1.3.1. 5S rDNA

More than 26 transcription factors involved in class III gene transcription have been purified [Huang & Maraia, 2001]. Classical Pol III promoters rely on two GTFs and one initiation factor for transcription. It is suggested that TFIIIB [Kassavetis *et al*, 1991] is the true initiating factor in yeast, though not in *H.sapiens* [Weser *et al*, 2003], while TFIIIA [Yang & Hayes, 2003] and TFIIIC [Matsutani, 2004] are assembly factors. TFIIIA is regarded as the gene-specific factor since it binds to 5S rDNA even in the absence of other transcription factors [Brown *et* al, 1996; Wyszko & Barciszewska, 1997].

While very little is known about Pol III transcription in plants [Cloix *et al*, 2003], *S.cerevisiae* has proved to be a good model applicable to most eukaryotic systems [Huang & Maraia, 2001]. The PIC in yeast commences with the sequestering of TFIIIA by the ICR where it binds to 3 sites within the A-, I- and C-boxes. This is followed by binding of TFIIIC to box A and box-C; footprinting shows that TFIIIC provides protection on both sides of

the region protected by TFIIIA [Geiduscheck & Kassavetis, 1995]. TFIIIC, then interacts through its N-terminal 165 amino acids with the amino-proximal half of Brf and recruits it to the pre-initiation complex, which is followed by the recruitment TFIIIB″ [Kassavetis *et al*, 1991]; this activates the DNA binding activity of TFIIIB, which by itself has low intrinsic affinity for DNA and positions it at the precise place to interact with Pol III. TFIIIB binds non-specifically to the 5′ region immediately upstream of the start point causing the DNA to bend. Along with Pol III which now joins the complex, TFIIIB and Pol III protect a region extending from position -40 to +15. The TFIIIB-DNA complex once established is difficult to disrupt and allows several rounds of reinitiation [Kassevetis *et al*, 1990].

The role of TBP in the transcription of TATA-less 5S rDNA has been resolved after several conflicting results. [White *et al*, 992; Willis, 1993]. The binding of TBP and thereby the TFIIIB assembly is thought to be brought about by protein-protein interaction involving Brf [Crighton *et al*, 2003], which binds directly with the N- and C- proximal lobes of TBP [Kumar *et* al, 1998]. Brf also binds to B″ and represents the Pol III specificity factor which recruits Pol III.

4.1.3.2. tDNA

In classical tRNA genes the core promoter elements are recognized by TFIIIC, and this is followed by the sequential binding of TFIIIB and Pol III. In yeast, TFIIIB acts as a true initiating factor because it can remain stably bound in a position-specific but sequence independent manner and facilitate Pol III recruitment [Wang & Roeder, 1995; Hasegawa *et al*, 2003]. Using templates containing azido-substituted nucleotides at definite positions to crosslink polypeptides, Geiduscheck and co-workers [Bartholomew *et al*, 1990] showed that at least four polypeptides of TFIIIC are bound to a region of 100 bp extending from -25 to most of the coding region (+75). The Brf and B″ subunits of TFIIB crosslink to -40 to +1 in the 5′ flanking region; Brf is closest to the start site and interacts with the 135 kD subunit of TFIIIC, which probably results in the stable binding of TFIIIB.

4.2. Activated transcription

Activated promoter expression is regulated at least at four levels - primarily by the transcription machinery itself [Holstege *et al*, 1998], by gene-specific activators and repressors [Ptashne & Gann, 1997], by the chromatin architecture [Tsukiyama & Wu, 1997] and by promoter elements.

4.2.1. Regulation by the transcription machinery

Studies in *S. cerevisiae* provide a paradigm for activated transcription. In yeast nuclei, a considerable part of the class II transcription initiation complex exists as preassembled holoenzymes [Greenblatt, 1997; Myers & Young, 1998]. Activators bind cooperatively to upstream sequences and recruit the initiation apparatus to the core promoter by interacting with subcomplexes of the holoenzyme [Kim *et al,* 1994]. Activator bypass experiments [Ptashne & Gann, 1997] and other studies [Yie *et al*, 1999] show that activators need to recruit only a few targets in the complexes to bring about the nucleation of a functional transcription complex.

The composition of holoenzymes is variable; in yeast, it is made up of at least 85 components [Holstege & Young, 1999] which include the core Pol II and nine SRB/mediator complexes, the SRB 10 (suppressor of Pol II truncation)/CDK (cyclin-dependent kinase) complex, chromatin remodelling Swi/Snf complex, a variable number of TAFs [Shen & Green, 1997] and subsets of GTFs and associated TAFs [Thompson & Young, 1995]. The existence of multiple forms of holoenzyme [Barberis & Gaudreau, 1998], activator-specific TAFs [Chen *et al*, 1994] cell-type specific TAFs [Verrijzer & Tjian, 1996], multiple TBPs [Hansen *et al,* 1997] and functional TFIIDs lacking TBP [Wieczorek *et al*, 1998] and the fact that activators need to interact with only a few components of the complex to recruit the preinitiation complex [Gonzalez-Couto *et al*, 1997] suggest that the composition of transcription initiation complex may be a major area of regulation. Myers *et al* [1999] found that a mutant holoenzyme lacking mediator 2 in its SRB/mediator complex caused a loss of activation by the recombinant transcription factor Gal4-VP16,

but not by Gcn4, implying that different activators require different holoenzymes.

4.2.2. Regulation by activators and repressors

Development of complex eukaryotes requires differential transcription of over 25,-35,000 protein coding genes [Schmidt, 2002; Ying & Lin, 2004] and expression of 3000 proteins in any given cell [Merrick, 2004] in precise spatial and temporal patterns. This is accomplished by employing combinatorial and synergistic control using a small number of ubiquitous, signal- and tissue-specific activators [Alvarez *et al*, 2003] (and also small RNAs such as miRNA and siRNA). The rate of PIC assembly and its activity is largely mediated through sequence-specific activators that recognize *cis*-acting, upstream activator sequences (UAS) [Yie *et al*, 1999]. A distinct set of transcription factors bind to the UAS to give rise to a stereo-specific interface ("pocket"), the enhanceosome, which is complementary to target surfaces on coactivators and the transcription machinery. Specific interactions within the enhanceosome, as well as with components of the transcription machinery results in cooperative DNA binding and transcriptional synergy (Figure-1.1).

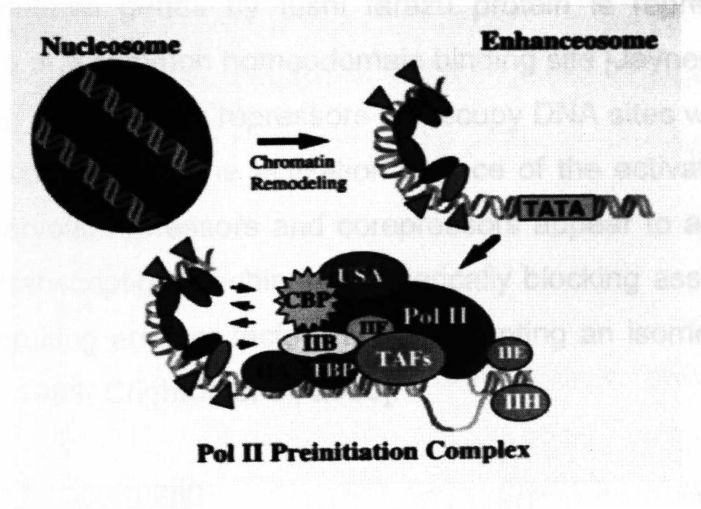


Figure-1.1. The enhanceosome (adapted from Carey, 1998). Combinatorial action of chromatin remodelling factors, activators and coactivators bound to the enhanceosome and interactions with the transcription machinery result in activated transcription. TAF-Transcription associated factor, TBP-TATA binding protein, CBP-CREB binding coactivator protein.

The efficiency of these plant transcription factors depends on the degree of interaction with other transcriptional activators [Yie *et al*, 1999], the number and type of binding sites in the UAS [Schwechheimer, 1998], rate of nuclear transport, oligomerization and post-translational modifications [Meshi & Iwabuchi, 1995].

The virus-induced expression of human interferonβ (IFNβ) promoter [Yie *et al*, 1999], anthocyanin expression in plants [Lloyd *et al*, 1992; Koes *et al*, 1994], abscisic acid response during seed development [Foster *et al*, 1994; Carson *et al*, 1995; Busk & Pages,1998], light, stress and hormone induced responses [Tabata *et al*, 1989; Meshi & Iwabuchi, 1995] and floral development regulated by MADS-box genes [Mesenguy & Dubois, 2003] are classic examples of combinatorial and synergistic interaction within and between activators and the preinitiation complex.

Repressors play a role opposite to that of activators. Johnson [1995] identified three types of repressors - those which repress by competitive DNA binding, by quenching and by direct binding to GTF. In competitive binding as exemplified by the fly engrailed homeodomain protein (En), activation of embryo developmental genes by fushi tarazu protein is repressed by En through occlusion of a common homeodomain binding site [Jaynes & O'Farrell, 1991]. Quenching occurs when repressors co-occupy DNA sites with activators and in doing so compromise the activation surface of the activator [Johnson, 1995]. Most eukaryotic repressors and corepressors appear to act directly on the unactivated transcription machinery by sterically blocking assembly, either directly or by recruiting another factor, or by preventing an isomerization step [Myers & Young, 1998; Crighton *et al*, 2003].

4.2.3. Regulation by chromatin

Transcriptionally competent regions tend to have a diffused, nuclease-sensitive chromatin structure [King, 2002], while chromatin of inactive regions are highly condensed [Felsenfeld, 1992]. Nucleosomal repression is the commonest global mechanism for negative regulation of transcription of many apparently unrelated genes [Cunliffe, 2003]. It restricts access of transcriptional regulatory

proteins to promoters, and hence the use of matrix attachment region (MAR) sequences in some transformation protocols [Petersen *et al*, 2002; Girard *et al*, 2004].

Several transcription factors are involved in changing chromatin structure to a repressive state such as the polycomb-group proteins, which repress homeotic genes in *Drosophila* [Zink & Paro, 1995; Wang *et al*, 2004] and the yeast Rap1 protein implicated in mating-type silencing [Hecht *et al*, 1995] and telomeric silencing [Thompson *et al*, 1994]. In many genes, remodelling of the ~200 bp histone octamer-bound nucleosome by protein complexes such as HMG proteins [Spiker, 1985], DEK protein, which binds preferentially to superhelical and cruciform DNA [Waldmann *et* al, 2004] histone acetylases and other enzymes [Logie & Peterson, 1997] is necessary [Giese *et al*, 1995] to permit binding of transcriptional activators in the region. In others, architectural proteins can be bypassed if strength of interactions can absorb the energy cost of DNA distortion.

4.2.4. Regulation by promoter elements

The modular elements that constitute the promoter such as the upstream elements, enhancer elements and in some cases 5′ UTR and introns [Last *et al*, 1991] have a dramatic influence on activated transcription. This knowledge is used to tailor the expression of a promoter through deletion, duplication or mutation of one or more of its component elements.

An exquisite example is the viral promoter CaMV 35S [Guiley *et al*, 1982; Roa Rodriguez, 2003]. The 35S promoter, consisting of 350 bp upstream of the +8 position [Benfey *et al*, 1990b] has been used succesfully to drive constitutive chimeric gene expression in many plant species, including those which are not natural hosts of the virus. Dissection of the native promoter shows a modular organization of several *cis*-elements which when acting alone produce low or tissue-specific expression but together act combinatorially and synergistically to produce strong and constitutive expression. The promoter contains a TATA box at -31 to -25, a tandem repeat of the pentanucleotide TGACG separated by 7 bp between -82 and -64 responsible for root expression [Benfey *et al*,

1989], a functional homolog of the octopine synthase (OCS) enhancer element at -79 to -64 overlapping the TGACG elements [Bouchez *et al*, 1989], a CAAT-like box at -85 whose deletion reduces transient activity to barely detectable levels and two proximal CAAT-like boxes at -64 and -57; three elements homologous to the SV40 core enhancer GTGGG(A/T)(A/T)(A/T)G occur at -143 and -104 orientated in the forward direction, and one at -125 in the reverse direction whose deletion produces 80% loss of transient activity [Ow *et al*, 1987]. Benfey *et al* [1990a,b] divided the B domain (-343 to -90) of the promoter into five subdomains (-343/301, -301/-206, -206/-155, -155/-106 and -106/-90) and used them in chimaeric constructs with the minimal CaMV promoter (-46 to +8) [Odell *et al*, 1985] or the A domain (-90 to +8) in transgenic studies; each subdomain displayed a different pattern of expression. A comparison of expression patterns conferred by each subdomain alone, in combination with the downstream domain, or in combination with other subdomains showed synergistic interactions among *cis*-elements within the enhancer.

The highly expressed and constitutive 300 bp early promoter of the 5243 bp SV40 genome is another multi-component structure [Zenke *et al*, 1986]. It contains an A/T-rich TATA-like element, two perfect 21 bp repeats containing multiple CCGCCC elements, a tandem duplication of a 72 bp enhancer sequence between -251 and -107 each repeat containing a 8 bp Pu/Py element GCATGCAT (Sph-motif) at -206/-199 and -133/-126 and a 8 bp core element GTGGAAAG at -247/-240 and -175/-168, respectively, followed by a distal 8 bp U/Y element ATGTGTGT that overlaps another core element [Weiher *et al*, 1983; Herr & Clarke, 1986]. Each functional element displays characteristic cell type-specific activity [Kanno *et al*, 1989; Macchi *et al*, 1989] but cooperate with one another synergistically to activate global transcription.

In the human interferonβ (IFNβ) promoter, the enhancer contains four positive regulatory domains to which six distinct proteins bind cooperatively [Yie *et al*, 1999]. Disruption to the precise context of the enhancer domains, or the absence/repositioning of an activator (except the architectural protein HMG 1(Y)) *in vitro*, abolishes synergy.

The synthetic, recombinant promoter pEmu [Last *et al*, 1991], which gives a high level of expression in monocots was constructed using multiple copies of *cis*-elements. The synthetic promoter comprises a cassette of six 42 bp anaerobic responsive elements (ARE) of maize and four 40 bp octopine synthase enhancers (OCS) of *A. tumefaciens* fused to a truncated maize alcohol dehydrogenase (*Adh1*) promoter containing the region -100 to +106 which includes a TATA box, transcription start site, intron and ATG translation initiation codon (patent AU-643521B2).

5. PLANT TRANSFORMATION

Once a suitable promoter is identified it should be tested for its biological efficacy *in vivo*. In preliminary promoter expression studies, organisms are usually transformed with chimaeric constructs containing the coding sequence of an assayable gene (reporter) under the control of the promoter and transformants are assayed for the gene product.

The ectopic expression of the gene is generally studied in two phases - as transient expression and as stable expression. Transient expression occurs when a construct which has penetrated the nucleus/organelle of a viable cell expresses itself while in an extrachromosomal state; this usually happens within 48-72 hours of plant transformation [Russell *et al*, 1993]. Transient expression is regarded as a measure of a promoter's strength in the assayed tissue [Martinez-Trujillo *et al*, 2003] since it reflects the interaction of the promoter and transcription factors, free of chromatin constraints. Following transient expression, a proportion of the constructs become stably integrated in the host genome. Expression during this phase, referred to as stable expression is more complex to analyze since it is influenced by additional factors such as chromatin structure at the site of integration, the intergrity, copy number and orientation of the construct entering the host genome and epigenetic changes.

Ideally, preliminary expression studies are best done in the host for which the promoter is intended. However, due to considerations such as the long

life cycle of a target crop and delayed phase changes (from vegetative to floral meristems), plants like *Nicotiana spp.*, *Arabidopsis and Petunia* are often used in preliminary transformation studies. These model plants which have a short life cycle have proved to be resilient enough to mimic expression patterns in other plant hosts. In this study, because of its short duration stable expression studies were conducted in *N. benthamiana*.

Two methods were used to transform plants in this study.

5.1. *Agrobacterium* based transformation

Several methods are available for transforming plants [Potrykus & Spangenburg, 1996]. Transformation mediated by the oncogenic, crown-gall phytopathogen *Agrobacterium tumefaciens* which elicits neoplastic growths on many plants is widely used in genetic transformation of most dicotyledonous and many tree species. Virulent strains of the bacterium possess a ~200 kb tumour inducing plasmid (Ti-plasmid) from which they transfer a single stranded copy of a ~ 23 kb segment, the transferred DNA (T-DNA), nicked from the bottom strand of Ti-plasmid to the host nucleus. T-DNA contains classical class II oncogenes *roi* (at the rooty locus) and *shi* (at the shooty locus) implicated in controlling hormone levels in tumour cells, and *nos* (nopaline synthase) or *ocs* (octopine synthase) opine genes. The transfer of T-DNA to the host is mediated mainly by (Vir) proteins of the virulence (*vir*) operons lying in a 30-40 kb region to the left of T-DNA. The ssT-DNA coated with VirE2 proteins and covalently bound at its 5′ end to VirD2 protein is exported (mediated by VirB and VirE1 proteins) through three membranes to reach the host nucleus, where it integrates into the top strand of the host DNA possibly by non-homologous end joining [Puchta, 2002]. Full integration of T-DNA into the host genome, which initially commences probably with synapses through micro-homologies allows these genes to be transcribed by plant RNA Pol II and this signals the onset of oncogenesis [Zambryski, 1993].

The T-DNA is flanked by 25 bp imperfect, direct repeats which are the only *cis*-element(s) required for the processing and transfer of T-DNA; any DNA

placed between these repeats will be imported into plant nuclei and function as T-DNA [Stachel *et al*, 1986; Zupan & Zambryski, 1997]. This knowledge is used to construct binary vectors for transformation where the gene(s) to be transformed is introduced between the left (LB) and right border (RB) repeats (or RB only) and transfer functions for mobilizing the recombinant T-DNA are provided in *trans* [An, 1985]. When an appropriate (disarmed) strain of *A.tumefaciens* containing the binary construct comes in contact with 'wounded', competent host cells, *vir* functions are induced by signal transduction resulting in transport of the recombinant T-DNA to host nucleus [Zupan & Zambryski, 1997].

Conifers and many monocotyledonous plants (the *Graminae* in particular), fall outside the natural host range of *Agrobacterium* and have proved to be recalcitrant to transformation by *Agrobacterium*, although in recent times successful transformation of rice [He *et al*, 2003], maize [Ishida *et al*, 1996], wheat and sugarcane [de la Riva *et al*, 1998] has been achieved. Birch [1997] believes that the problem lies in the competency and regenerability of host cells and not in the delivery of T-DNA [Grimsley *et al*, 1987], or integration [Narasimhulu *et al*, 1996], or expression [Shen *et al*, 1993]. Faced with this difficulty, much effort has been directed towards developing alternate methods for direct DNA delivery such as electroporation of protoplasts or intact tissue [Fromm *et al*, 1986; Dekeyser *et al*, 1989], polyethylene glycol (PEG) treatment of protoplasts [Shillito *et al*, 1985] and high-velocity bombardment of DNA coated microprojectiles into cells or intact tissue [Sanford *et al*, 1993]. Some dicotyledonous plants, which are not amenable to *Agrobacterium*-mediated methods because of host-specificity restrictions [Hinchee *et al*, 1988] or tissue culture limitations [Umbeck *et al*, 1987] may also be transformed using such methods.

5.2. Biolistic transformation

Particle bombardment technology is still the primary means of transforming pine [Charity *et al*, 2005; Gould *et al*, 2002] and several variants of the method exist [McCabe & Christou, 1993; Oard, 1993; Vain *et al*, 1993]. In

the biolistic (biological ballistic) method of DNA delivery introduced by Sanford and co-workers [Klein *et al*, 1987] and still used in pine, DNA is adsorbed onto microprojectiles (microcarriers) of gold or other inert metal in the presence of the polyamine spermidine and $CaCl_2$. Microcarriers are propelled at high velocity into target plant tissue by a shockwave created by compressed helium gas. Target cells or explants are often anchored in a medium such as agar, or on a support such as filter paper to absorb the shock of bombardment [Klein *et al*, 1987]. Penetration of the tissue is a random event and some of the projectiles enter the nucleus to release the bound DNA by desorption. During this extra-chromosomal phase the exogenous DNA comes under the control of the host's transcription machinery and expresses itself transiently. Some of this DNA may later stably integrate into the chromosomal DNA to give rise to stably transformed cells.

Efficiency of transformation is generally low because only a fraction of the bombarded cells may be penetrated. Iida *et al* [1990] observed that transient expression in bombarded suspension culture cells was of the order of 1-3 x 10^{-3}; transiently expressing cells appear to correspond to those cells in which the particle had entered the nucleus [Yamashita *et al*, 1991]. Stable transformation is at least one to two orders lower ie. < 5 x 10^{-4} [Russell *et al*, 1993]. This promoter-independent reduction may be mainly due to cell death caused by mechanical damage inflicted by microprojectiles. Plant cells usually form a callose plug at the point of injury as early as 10 minutes after injury [Nims *et al*, 1967]. Hunold *et al* [1994] observed that 98-99% of bombarded maize cells failed to close the lesion and succumbed to slow death within the first 48 hours.

5.3. Reporters

A convenient way to monitor promoter activity in transformed plants is to incorporate reporter genes into constructs containing the putative promoter. Reporter genes such as β-*glucuronidase* (*gus*) of *Escherichia coli* [Jefferson

et al, 1987], *luciferase* (*luc*) [Suzuki *et al*, 2005] and green fluorescent protein (*gfp*) of *Aequorea victoria* [Chalfie *et al*, 1994] produce easily assayable products. They differ in their sensitivity of detection; >100 molecules of GUS and >10,000 molecules of GFP need to be present in a cell for signal detection [Kohler, 1998]. The level of transient expression of the reporter is widely used to compare promoter activity [Martinez-Trujillo *et al*, 2003].

6. TRANSFORMATION AND GENE DOWN REGULATION

A potential application of transformation is in the selective down-regulation of endogenous gene expression to enhance crop performance. For example, a pine promoter selected in this study can be used to attenuate components of the lignin biosynthesis pathway to produce designer-wood of different industrial characteristics. Similarly, loss-of-function of genes associated with male cone formation can ablate pollen production and prevent spread of transgenic pollen to wild-type populations. Another significant use of gene down-regulation is in functional genomics where functions of novel genes are studied by targeted gene-knockouts.

Down regulation of specific characters can be accomplished in different ways. Zamecnik and Stephenson [1978] used an oligonucleotide (DNA) complementary to 13 nucleotides of the reiterated 3′ and 5′ terminal sequences of Rous sarcoma virus RNA to inhibit viral replication in infected fibroblast cultures. Another approach is the use of triple helix (triplex) formation [Volkmann *et al*, 1995]. The antisense RNA technology is widely used in plants. This, however, has been superseded by dsRNA-mediated post-transcriptional gene silencing (PTGS) or RNA interference (RNAi) technology, which provides more consistent results. Both methods were used in this study.

6.1. Antisense RNA strategy

The antisense RNA (as-RNA) strategy is similar to the endogenous

mechanism observed in several biological systems. An as-RNA forms complementary Watson-Crick pairing with bases of a target gene's RNA [Nellen & Lichtenstein, 1993]. The RNA duplex [Kim & Wold, 1985] is thought to be responsible, at least in part, for blocking target RNA maturation, transport, or translation [Van der Krol *et al*, 1988] possibly by steric hindrance or by recruiting dsRNA-specific RNases. In *E.coli*, replication of ColEI type plasmids is negatively controlled at the origin by an untranslated species of RNA (RNA I) which hybridizes with the primer precursor RNA II with which it shares a 106 base complementarity [Tomizawa, 1986]. Translation of P22 antirepressor protein in *Salmonella* phage P22 is inhibited by *sar* RNA (small as-RNA) which is complementary to the Shine-Delgarno sequence of antirepressor mRNA. mRNA of the constitutively expressed *eb4-psv* (pre-spore vesicle) gene of the slime mould *Dictyostelium discoideum* accumulates in aggregating cells, but is unstable in disaggregating cells where a developmentally regulated as-RNA derived from the same locus is detected [Hilderbrandt, 1992]. Induced murine erythroleukemia cells accumulate an induced as-RNA involved in the post-transcriptional down regulation of P53 [Khochbin & Lawrence, 1989]. Maize has an RNA complementary to α-tubulin mRNA which is expressed in tissues where α-tubulin mRNA is present at low levels [Dolfini *et al*, 1993].

Regulation by antisense RNA is achieved by direct introduction of antisense RNA into cells [Izant & Weintraub, 1984; Melton, 1985] or more commonly by generating antisense RNA *in vivo.* The latter involves transforming an organism with gene constructs containing the whole, or a part of the transcribed region of the target gene in reverse orientation. When integrated into the genome, the construct can be expected to provide sustained generation of antisense RNA. This method was used successfully to down-regulate polygalacturonase by 99% to delay fruit softening and increase shelf life of tomato [Smith *et al*, 1990], granule-bound starch synthase to produce amylose-free potato, cinnamyl alcohol dehydrogenase (CAD) to abolish lignin synthesis [in Mol *et al*, 1990] and chalcone synthase to reduce flower pigmentation [Van der Krol, 1988].

In several bacterial genes, the most effective antisense RNA appears to be complementary to the 5′ region, which generally includes the translation initiation site. In eukaryotes, however, a clear pattern has not emerged. Some authors have used sequences from the 5′ untranslated/ORF region [Izant & Weintraub, 1985; Melton, 1985; Chang & Stolzfus, 1987; Dwivedi, 1994; Klann *et al*, 1996; Euch *et al*, 1998] and others from the 3′ UTR/ORF region [Chang & Stolfus, 1987; Van der Krol *et al,* 1990a; Bourke & Folk, 1992; Temple *et al*, 1998], or from full length cDNA [Samac & Shah, 1994; Halpin *et al*, 1994; Landchutze *et al*, 1995; Shintani *et al*, 1997; Keller *et al*, 1999]. Some studies have found the 5′ region to be ineffective or less effective than the 3′ region [Van der Krol, 1990a; Bourke & Folk, 1992] and in others the 3′ region was ineffective [Melton, 1985].

The failure of antisense constructs selected from a different region of the gene to effect down-regulation, discrepancies in lengths of constructs used and intransigence of some genes to antisensing suggest that the mechanistic basis for antisensing is not merely base complementarity. Studies in prokaryotes suggest a possible role for high-order RNA structures and accessory proteins in enhancing or preventing RNA hybridization. A single-stranded loop in the secondary structure of RNA I may be an important determinant in hybrid formation with mRNA of RepAI in suppressing FII type plasmid replication in *E.coli* [Womble *et al*, 1985]. The low response of *ompA* (outer membrane protein) to antisensing is thought to be due to the propensity of *ompA* mRNA to form tertiary structures which render it resistant to hybridization with antisense RNA [Coleman *et al*, 1984]. The base paired region of bacterial *ompF* mRNA and antisense *mic* RNA (mRNA interfering complementary RNA) is stabilized by flanking stem-and-loop structures [Mizuno *et al*, 1984]. ColEI encoded protein Rom is required to stabilize RNA II:antisense RNA I duplexes in a stable complex [Tomizawa, 1990]. Similarly, in snRNP (small nuclear ribonucleoprotein) mediated splicing, proteins mediate hybridization of U1 RNA and pre-mRNA at the splice site [Lewin, 1997].

In addition, antisense constructs may be also subject to same selection pressure *in vivo* as sense constructs [Mol *et al*, 1990]. Although specific examples are lacking, it is reasonable to speculate that antisense transgenes are prone to effects of methylation, copy number and chromosomal position as are sense transgenes. Examples from sense gene transformation would help to illustrate this point. When the promoter inserts into a chromosomal region of different isochore context it may lead to methylation of CG or CAG islands in the promoter; such methylation is invariably associated with gene silencing. Maize *A1* gene was transformed into white flowering petunia to produce pelargonidin derivatives of different colours [Meyer *et al*, 1987]. Following a period of high light intensity and elevated temperatures, transgenic petunia producing fully pigmented flowers produced new flowers that were weakly coloured. Instability of *A1* transgene in single-copy transformants was found to be associated with methylation of the introduced DNA; hypermethylation was restricted to the transgene while the flanking plant DNA remained hypomethylated [Meyer & Heidmann, 1994]. Biolistics, and to a lesser extent *Agrobacterium* mediated transformation produce mutiple inserts in the host genome which can result in its silencing by methylation. Gelvin *et al* [1983] observed that T-DNA was inactivated in tumour lines of tobacco containing multiple copies of T-DNA. Inactivation was associated with hypermethylation of T-DNA and could be reversed by treatment with 5-azacytosine [Hepburn *et al*, 1983]. Similar repeat-induced gene silencing (RIGS) was demonstrated by increasing copy number of a transgene at a single locus in *Arabidopsis* [Assaad *et al*, 1993].

Until the underlying mechanism of antisense regulation is understood it may not be possible to define a universal as-RNA structure. It has to be done on a case-by-case basis, based on the gene region to be targeted, length of RNA hybrid and the potential of antisense transcript to form secondary structures.

6.2. RNA interference (RNAi) strategy – dsRNA induced gene silencing

The mechanistic basis of gene silencing mechanisms such as post-

transcriptional gene silencing (PTGS) [Hamilton & Baulcombe, 1999] used as a defensive mechanism against plant viruses, or cosuppresion which results in silencing of transgene and/or endogenous gene in plants [Van der Krol, 1990b] and quelling in fungi [Cogni *et al*, 1996] has long remained in the realm of speculation. For example, a deep violet flowering line of petunia transformed by petunia chalcone synthase gene (*chs)* gave several transformants bearing variegated flowers. When petals of these flowers were analysed, the steady-state level of *chs* mRNA was highly reduced in the white sectors of petals compared to the coloured sectors even though both had comparable mRNA in run-on experiments [Van der Krol *et al*, 1990a]. Kooter *et al* [1999] observed the extreme stoichiometric differences between antisense-*chs* (as-*chs*) and sense-*chs* transcripts in petunia transformed with as-*chs* and suggested an RNA titration model based on dsRNA to explain gene silencing (cosuppression) by substoichiometric levels of as-RNA. Later work by Mellors and co-workers [Bartel, 2004] shed light on the role of small interfering RNAs (siRNA), a class of small RNAs which display length and functional heterogeneity [Hamilton *et al*, 2002], in these highly conserved and related gene silencing phenomena.

Microinjection of dsRNA into *C. elegans* gut cytoplasm [Fire *et al*, 1998] was observed to have a profound impact on gene silencing; so was the feeding of worms with *E.coli* engineered to produce dsRNA, or soaking worms in dsRNA solution [Timmons & Fire, 1998; Tabara *et al*, 1999]. This effect termed RNA interference (RNAi) provides a paradigm for gene silencing/mRNA turnover mechanisms (Figure-1.2).

The process is initiated by long [Fire *et al*, 1998] or short dsRNA [Martinez *et al*, 2002] of endogenous (from retroelements, transposons, highly repeated sequences, pseudogenes and intergenic regions) or exogenous origin (viruses, transgenes). The dsRNA is processed into 21 to 25nt symmetric or asymmetric siRNA duplexes [Tang, 2005] possessing 5′ phosphate and 3′ hydroxyl termini with a 2nt overhang at the 3′ end by an RNase III-like ATP-dependent dsRNA-specific enzyme(s), called dicer (DCR) in animals and insects [Bernstein *et al*, 2001; Lee *et al,* 2004] or

dicer-like enzyme (DCL) in plants [Tang *et al,* 2003]; in *Arabidopsis* there

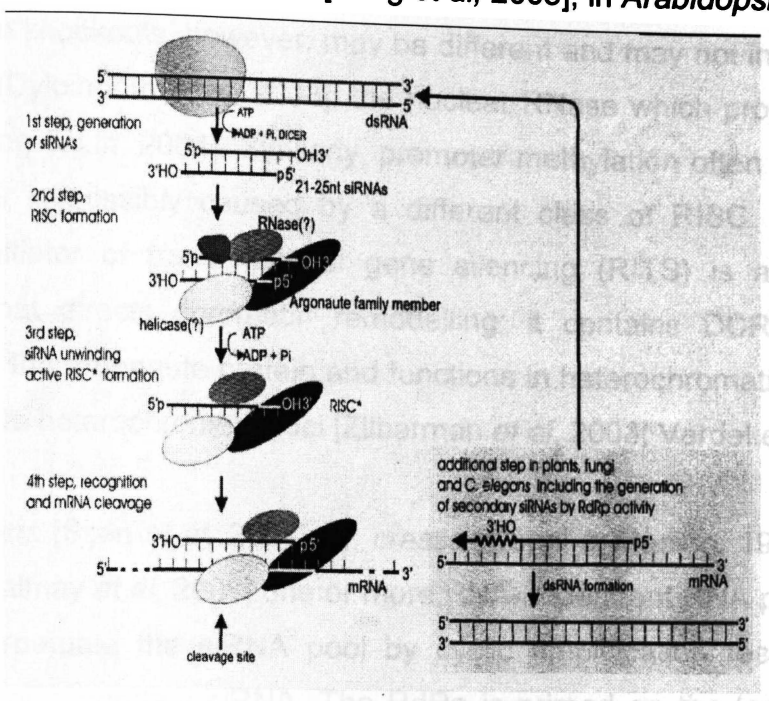


Figure-1.2. A schematic of RNAi (Adapted from Kulinska *et al*, 2003). The dsRNA is processed by dsRNA-specific dicer to siRNA. The siRNA becomes incorporated into the RNA-induced silencing complex (siRISC) where the helicase unwinds the duplex siRNA. The siRISC* may contain either the sense or antisense ss-siRNA. An siRISC* containing the antisense ss-siRNA targets the mRNA for cleavage. RdRp – RNA dependant RNA polymerase.

are at least four nuclear or cytoplasmic DCLs, each probably associated with a different size class of siRNAs [Papp *et al,* 2003], or with maturation of siRNA or miRNA, which follow distinct pathways [Ying & Lin, 2004]. The siRNA becomes incorporated into a multisubunit, ribonucleoprotein complex known as the RNA-induced silencing complex (siRISC) or RISC loading complex (RLC) [Tang, 2005]; unwinding of the siRNA duplex mediated by a RISC RNA helicase results in the formation of an active siRISC (siRISC*) complex [Kulinska *et al*, 2003]. A cleaving form of siRISC* containing ss-siRNA of the 21-22nt size class [Bartel, 2004], targets specific mRNAs sharing sequence complementarity with the guide strand of siRNA [Hutvagner & Zamore, 2002]. Its endonuclease slicer, an argonaute protein, cleaves the transcript between residues 10 and 11 from the 5′ end of siRNA complementarity [Elbashir *et al*, 2001]. Mutations in the central and 5′ region of shRNA lead to pronounced reduction in siRNA function [Pusch *et al*, 2003].

The processing of short hairpin constructs (shRNA) used in this study and in mammalian knockouts, however, may be different and may not involve dicer or drosha [Dykxhoorn *et al*, 2003], the nuclear RNase which processes pri-miRNA [Ying & Lin, 2004]. Similarly, promoter methylation often associated with PTGS is possibly caused by a different class of RISC. The RNA-induced initiator of transcriptional gene silencing (RITS) is a RISC-like complex that directs chromatin remodelling; it contains DCR-generated siRNA and the argonaute protein and functions in heterochromatin silencing by binding to heterochromatin loci [Zilberman *et al*, 2003; Verdel *et al*, 2004]

In *C. elegans* [Sijen *et al,* 2001], *N. crassa* [Cogni & Macino, 1999] and *A. thaliana* [Dalmay *et al*, 2001] one or more RNA-dependent RNA polymerase (RdRp) perpetuate the siRNA pool by cyclic amplification resulting in a population of secondary siRNA. The RdRp is primed on the target mRNA template by existing siRNAs to produce secondary siRNAs whose antisense strands are complementary to an upstream part of the initially targeted mRNA segment [Xie *et al*, 2004] or outside the region of homology [Klahre *et al*, 2002].

Plant siRNAs that are induced locally can spread cell-to-cell or systemically; they migrate through the plasmodesmata and the vascular system to reach distal tissues, their movement often facilitated by siRNA binding proteins [Yoo *et al*, 2004]. Similarly, RNAi spreads systemically in *C. elegans*. RNAi is also inherited by the offsprings in *C.elegans*; the inheritance is independent of the presence of the target locus in the progeny [Grishok *et al*, 2000].

RNA induced gene silencing can be effected in several ways [Dykxhoorn *et al*, 2003]. *In vitro* transcribed dsRNA [Donze *et al*, 2005] or synthetic siRNA, or siRNA generated *in vitro* by fragmenting long dsRNA with recombinant DCR [Buchholz *et al*, 2005] can be directly introduced into cells; or cells can be transformed with vectors containing tandem promoters that express the sense and antisense strands of the dsRNA from separate transcription units.

Very efficient gene silencing in plants is achieved by expressing self-complementary hairpin RNAs (hpRNA) within two complementary regions separated by a spacer [Pandolfini *et al*, 2003]. The hairpin is generally an inverted repeat of a 300-600bp fragment of the target gene sequence. The frequency of silencing is dramatically increased when an intron is used as the spacer and this led to the development of intron-spliced hairpin RNA (ihpRNA) constructs [Helliwell & Waterhouse, 2003]. A 90% reduction in GUS expression in *gus* transformed rice was obtained with a *Ubi-gus* ihpRNA construct [Wesley *et* al, 2001]. Potato virus Y (PVY) infection was completely suppressed in tobacco transformed with a CaMV-ihpRNA construct targeting the niaprotease gene of PVY genome [Smith *et al* , 2000]. Expression of the ethylene signalling (*EIN2*) and flower repression (*FLC1*) genes in *Arabidopsis* was completely suppressed by CaMV 35S-ihpRNA [Wesley *et al*, 2001]. Pandolfini *et al* [2003] used *rolC* promoter driven ihpRNA to target plum pox virus polyprotein where 80% of inoculated plants were virus free and 20% had only mild symptoms. The alcohol-inducible promoter-ihpRNA of Lo *et al* [2005] gave 100% silencing of *gus* transcripts.

In mammals and zebrafish, however, when dsRNA longer than 30nt is used for RNAi it results in activation of the dsRNA-dependent protein kinase (PKR) leading to general blockage of protein synthesis. Short hairpin RNA (shRNA) cassettes were developed to circumvent this interferon response. They often make use of RNA Pol III promoters such as U6 promoter, H1 promoter and $tRNA^{Val}$ promoter [Kim, 2003]. The shRNA expression cassette consists of a Pol III promoter followed by a 19-29nt gene-specific targeting sequence separated by a short spacer (loop) sequence of 9-12nt from the reverse complement sequence and terminating in a run of thymidine residues. The advantage of Pol III transcription system is that it allows the production of very short RNAs (shRNA) which terminate at 3-4 uridine residues.

7. OBJECTIVE OF THE STUDY

Almost all promoters that are used in the molecular transformation of pine since it commenced in New Zealand a decade ago are of heterologous origin, because endogenous pine promoters are still not available in the public domain. Promoters such as CaMV 35S promoter and its enhancer added version, maize ubiquitin promoter, pEMU and *chs* were selected based on their strong performance in their respective biological systems and also in transient expression assays in undifferentiated pine tissue [Walter *et al*, 1997].

From the discussion above, it can be anticipated that heterologous promoters may not perform to the same extent in pine as in their homologous systems due to differences in quality and/or quantity of regulatory factors. In the case of CaMV 35S promoter, which is routinely used in pine transformation there is an additional concern regarding its pathogen origin [Walter *et* al, 1997].

Furthermore, heterologous constructs may be open to silencing mechanisms that operate in plants against invasive DNA [Henikoff & Comai, 1998; Matzke & Birchler, 2005]. For instance in transgenic cotton bred to produce Bt (*Bacillus thuringiensis*) toxin under the control of CaMV 35s promoter against bollworms, it was observed that CaMV 35s promoter is silenced as the crop matures as a result of which Bt toxin expression is switched off [Llewellyn, CSIRO 2003]. Pine is grown for 25-30 years before it is milled and it will be catastrophic if a promoter used in its transformation ceases to be active midway through its growth. Many of the heterologous promoters that have been used in New Zealand are yet to be evaluated for their longevity since even the earliest pine transformants are barely 9 years old.

In view of these concerns, a pine promoter selected on the basis of its strength and expression pattern is preferred to a heterologous promoter for pine transformation. Since it is endogeneous to pine, its longevity can be quickly assessed by assaying non-transgenic plants of different ages for its gene product ie. there is no time lag as in the case of heterologous promoters which

have to be first transformed into pine and the transformants then grown to the required stage.

The aim of this study is to isolate highly expressed, endogenous promoter(s) from pine, which can be used to transform pine for global or stage/tissue-specific expression. Even though a great deal of work has been done in pine to identify genes and study gene expression, there is hardly any information available in the public domain that can be used to analyse the pine genome or its transcriptome. The approach taken in this study was to use heterologous sequence information to screen the pine genome or its transcriptome for orthologs with desirable expression features.

This can be done in one of several ways. The transcriptome can be screened with cDNA probes of heterologous candidate genes to confirm the expression pattern of their pine orthologs and then gene orthologs with the desired expression features can be isolated from the pine genome. Another approach is that taken in this study with the Pol III transcribed 5S rDNA gene family, where its multi-copy genes have nearly identical coding regions which display only 1-2 base differences and whose transcripts are indistinguishable. In this instance, a suitable member of the gene family was first isolated from the genome and then its expression characteristics were evaluated in transient and stable expression studies by forward and reverse genetics. Other candidate genes considered in this study include some single-copy genes which are strongly expressed in other plants and highly expressed members of housekeeping gene families such as actin and tRNA.

CHAPTER II

MATERIALS AND METHODS

1. PLANT NUCLEIC ACID EXTRACTION

1.1. RNA extraction

Total RNA was prepared by the method of Chang *et al* [1993]. Plastic-ware was treated with $CHCl_3$ and autoclaved. DEPC was added at 0.1% (v/v) to water and all solutions. Frozen tissue was homogenized to a fine powder. A one gram aliquot was incubated with 13ml of RNA extraction buffer at 65°C. 2-β mercaptoethanol (ME) in the extraction buffer prevents oxidation of phenolic compounds and Polyvinyl-pyrrolidone (PVP) adsorbs polyphenols. Cetylytrimethylammonium bromide (CTAB) solubilizes cell membranes and complexes with nucleic acids, separating them from polysaccharides and phenolic compounds; NaCl at 2 M concentration helps to eliminate polysaccharides and Proteinase K degrades cellular proteins. The homogenate was extracted with chloroform:isoamyl alcohol (24:1) and the RNA was precipitated with 1/4 volume 10 M LiCl overnight at 4°C and pelleted by spinning the tubes at 12,000 rpm for 30 minutes (') at 4°C. The pellet was resuspended in 500μl saline saturated SDS-Tris EDTA (SSTE) and extracted with $CHCl_3$:isoamyl alcohol mixture. RNA was precipitated with 1/10 volume 3 M sodium acetate (NaOAc), pH 8 and 2 volumes ethanol.

1.2. Genomic DNA Extraction

DNA was extracted by the method of Doyle & Doyle [1990]. Frozen tissue was homogenized and 1g was incubated with 13ml of DNA extraction buffer at 65°C. The homogenate was extracted with chloroform:isoamyl alcohol (24:1) mixture and the supernatant containing the nucleic acid-CTAB complex was re-extracted. The DNA was precipitated with an equal volume of cold (-20°C)

isopropanol (IPA), washed with 95% EtOH and air-dried. Residual polysaccharide contaminants were removed by a salt wash in 10ml 1 M NaCl.

1.3. Nuclear DNA isolation

(a) Floraclean™ (QBiogene) kit was used as per manufacturer's instructions. One gram of tissue was ground in liquid nitrogen, transferred to a Dounce homogenizer containing 10ml cold nuclear buffer and homogenized till the plunger passed smoothly. The homogenate was filtered through four layers of muslin and the filtrate centrifuged at 3500 rpm for 20' at 4°C. The supernatant was discarded and nuclei resuspended gently in 5 ml nuclear buffer and recentrifuged twice as above. The nuclei were lysed by resuspending the pellet in 1.5 ml nuclear resuspension buffer and adding 100µl nuclear lysis solution. The mixture was incubated at 55°C for 30'. A 25µl aliquot of protease mix was added and incubation was continued at 55°C for 4 hours. 500µl 'salt-out' mixture was mixed with the lysed nuclei and incubated overnight at 4°C. The mixture was spun at 13000 rpm for five minutes and the pooled supernatant was transferred to a Corex tube, diluted with 2 ml water, and mixed with 8 ml EtOH and incubated at -70°C overnight. Tubes were centrifuged at 13,000 rpm for 20' and the pellet was air dried and dissolved in 100µl TE.

(b) In a modification of the method of Murray and Thompson [1980], 1g of frozen tissue in liquid nitrogen was ground to fine powder and added to 40ml extraction buffer in a dounce homogenizer on ice. The mixture was stirred with a glass rod until all tissue was wet and then homogenized with 4-10 passes of the plunger. 400µl of NP-40 was added, mixed with the homogenate and the mixture filtered through four layers of muslin. The filtrate was centrifuged for five minutes at 4°C and the pellet resuspended in 4ml ice-cold wash buffer. 40µl NP-40 was added, mixed well and centrifuged at 4000 rpm for three minutes. The nuclear pellet was resuspended in 1ml wash buffer and allowed to come to room temperature. The suspension was transferred to a microfuge

tube, mixed with 200µl 5% sarkosyl and incubated at room temperature for 30'. 200µl 5 M NaCl was added and mixed by inversion; this was followed by adding 160µl 1% CTAB/0.7 M NaCl. The mixture was mixed by inversion and incubated at 60°C for 10'. A 1.56ml aliquot of $CHCl_3$:isoamyl alcohol mixture was added, mixed by inversion and centrifuged at 12,000 rpm for 15' at room temperature. Two-thirds volume of propanol (-20^{0}C) was added to the supernatant and incubated at -20°C for 10'. Tubes were centrifuged at 5000 rpm for 10' to pellet the DNA. Pellets were washed with 70% EtOH, air-dried and resuspended in TE buffer.

2. PLASMID DNA ISOLATION

2.1. Alkaline lysis method

Plasmid DNA was extracted according to Birnboim & Doly [1979]. Fresh Luria-Bertaini (LB) medium [Ausabel *et al,* 1998] was inoculated with an overnight culture of bacteria and shaken overnight at 225 rpm. Thirty milliliters of the culture was spun at 10,000rpm for 5minutes (') at 4°C. The pellet was resuspended in 2ml ice-cold resuspension buffer. A 4ml aliquot of freshly prepared lysis solution containing 0.2N NaOH and 1% SDS was added and incubated on ice for 5'; three milliliters of neutralizing solution was added, incubated on ice for 5' and centrifuged at 12,000rpm for 25'. RNase A was added to to the supernatant at a final concentration of 20µg ml^{-1} and incubated at 37°C for 20'. The preparation was extracted twice with Tris-EDTA (TE) saturated phenol-chloroform (1:1) and the DNA was precipitated in 2.5 volumes of ethanol.

2.2. Wizard™ [Promega] DNA purification system

Plasmid was isolated essentially by the alkaline lysis method and then purified using Promega™ DNA purification resin. A 100ml aliquot of overnight culture was pelleted by centrifuging at 8000 rpm for 10' at room temperature. The cell pellet was resuspended in 15ml resuspension buffer containing RNase A at 100µg ml^{-1}. Fifteen ml lysis solution was added, followed by 15ml neutralizing

solution, mixed immediately and the tube spun at 14,000 rpm for 15' at room temperature. The supernatant was filtered through Whatman® #1 filter paper; 0.5 volume IPA was added to the filtrate and centrifuged at 14,000 rpm for 15'. The DNA pellet was dissolved in two ml Tris-EDTA buffer and mixed with 10ml resin-guanidium hydrochloride slurry and loaded onto a maxicolumn. A vacuum was applied to the column and the column washed with 25ml wash solution. The resin was rinsed with 5ml 80% EtOH and the column centrifuged in a 50ml falcon tube at 2,500 rpm for 5'. The resin was dried under vacuum for 5'. A 1.5ml aliquot of water preheated to 70°C was added and after 1 minute the tube was centrifuged at 2,500 rpm for 5' to elute the DNA.

2.3. Alkaline-lysis/PEG procedure

A modified alkaline-lysis method was adopted to produce supercoiled plasmid DNA for cycle sequencing. The preparation contains. Bacteria was grown in Terrific broth resulting in a 8-fold increase in bacterial yield [Tartof & Hobbs, 1987]. The bacterial pellet from a 30ml culture was resuspended in 4ml lysis buffer and 6ml NaOH:SDS was added and mixed till the solution cleared. A 6ml aliquot of KOAc was added and incubated on ice for 5' before centrifugation. Following RNase A treatment, the supernatant was extracted twice with $CHCl_3$. The DNA was precipitated using IPA and immediately centrifuged for 10' at room temperature. The DNA pellet was washed with 70% EtOH and dried. The pellet was dissolved in 32µl nuclease-free water and 8µl 4 M NaCl and (followed by) 40µl autoclaved 13% PEG-8000 were added and mixed well. The mixture was incubated at 4°C for 20' and centrifuged at 12,000 rpm for 15' at 4°C to pellet the DNA.

2.4. Boiling method

The rapid plasmid preparation method [Ausabel *et al*, 1998] was used to screen large numbers of clones. A 1.5ml aliquot of an overnight bacterial culture was spun down and the pellet resuspended in 200µl STET. The suspension was boiled for 1' and centrifuged for 10' to pellet denatured cellular and chromosal debris which was removed with a tooth pick. 1/10 volume 3M

NaOAc and 2.5 volumes EtOH were added to the supernatant and spun for 10'. Plasmid DNA was resuspended in 29μl TE and 1μl RNase A (1mg ml^{-1}) was added.

3. GEL ELECTROPHORESIS

3.1. Agarose gel electrophoresis

DNA size was estimated from the distance migrated, which is inversely proportional to Log_{10} of its molecular weight [McDonell, 1977; Southern, 1979]. Ethidium bromide (EtBr) was incorporated into the gel and buffer at 0.5μg ml^{-1}. Enzyme digested plasmid was electrophoresed either in the absence of EtBr, or at EtBr concentrations in excess of 0.5μg ml^{-1}; under these conditions, the undigested closed circular supercoiled conformation (form I) migrates faster than form III and the nicked, relaxed circle (form II) migrates slower than forms I and III. Tris-Acetate-EDTA (TAE) was used as electrophoresis buffer when extracting bands for cloning. Separide™ gel matrix (Life Technologies) at 2% concentration in TAE was used to resolve dsDNA fragments differing by ~10bp.

3.2. Polyacrylamide gel electrophoresis (PAGE)

Gel casting solution of total monomer (T) concentration 6-12% was prepared from a sequencing gel mix containing 40% T (19:1 crosslinking). 1% (v/v) 10% ammonium persulphate and 0.02% (v/v) TEMED were added to promote self polymerization. Urea was added at 8.3M final concentration when running denaturing gels. Samples were electrophoresed in TBE buffer at 100-200V in a Biorad apparatus and visualised by EtBr- or silver staining, or autoradiography.

4. NUCLEIC ACID CLEANING

4.1. Ion exchange chromatography

Qiagen® tips contain anion-exchange silicagel resin covalently coated with a hydrophilic substance that prevents nonspecific binding. The elution point for dsDNA is 1.4-1.6 M NaCl at pH 7. DNA preparations (free of SDS or other anionic detergents which would complex and interfere with anion exchange groups of the resin) were made up to a final volume of 1ml, added to Quigen-tip

20 and pre-washed with QBT buffer. The solution was allowed to flow by gravity and columns washed with 1ml aliqouts of QC buffer. The tip was placed over a clean eppendorf tube and DNA eluted in 800µl QF buffer by gravity flow.

4.2. Spin column chromatography

MicroSpin™ columns (Pharmacia) pre-packed with Sephacryl resin in TE buffer were used for DNA cleaning. The exclusion limits of gel filtration resins, which are meaningful in continuous flow situations are not applicable here. The column was vortexed briefly to resuspend the resin. The cap was loosened one-fourth turn and the bottom closure snapped off. The Column was placed in a 1.5ml microfuge tube and spun at 3000 rpm (735 x *g*) for 1' to remove the storage buffer. The Column was placed in a fresh eppendorf tube and 25-100µl of sample applied slowly to the top center of the resin. The tube was spun at 3000 rpm for two minutes and the filtrate containing DNA collected in the eppendorf tube.

4.3. Diethyl aminomethanol (DEAE)-cellulose chromatography to clean tRNA

The observation that only 7% of rRNA adsorbed by a DEAE-cellulose column is eluted with NaCl gradients of 0.4 to 0.56 N while 100% of tRNA is eluted [Monier *et al*, 1960] and that rRNA is insoluble in 1 M NaCl is used in this method [Hecker *et al*, 1979; Anandaraj & Cherayil, 1974]. Three grams DEAE-ion exchange cellulose resin was equilibrated in three changes of 0.01 M Tris-HCl, pH 7 equilibrating buffer containing 0.1 M NaCl; 2µl DEPC was added to the mix. The slurry was poured into the barrel of a 10ml syringe closed at the bottom with a small circle of Whatman filter, till a two millilitre resin column was formed. The column was washed with ten column volumes of equilibrating buffer. An 0.1 M NaCl solution containing total RNAs was poured through the column and washed with ten column volumes of equilibrating buffer; rRNA and tRNA species bind to the column under these conditions. The tRNA was eluted by washing the column with 0.01 M Tris-HCl, pH 7 containing 1 M NaCl.

5. GEL EXTRACTION

5.1. Low melting point gel

Gels were cast from low melting point agarose and the DNA band of interest was excised with minimum exposure to shortwave UV [Sarkar & Sommers, 1991]. Five gel volumes of 20 mM tris-HCl, pH 8, 1 mM EDTA was added and the mixture was heated at 65°C for 10'. The mixture was extracted with phenol [Weislander, 1979] followed by $CHCl_3$:isoamyl alcohol mixture. The DNA was precipitated with two volumes of EtOH and dissolved in nuclease-free water.

5.2. Gel extraction with glass fibre

Nucleic acids bind specifically to the surface of glass fibres in the presence of chaotropic salts like guanidium thiocyanate [Vogelstein and Gillespie, 1979]. The Highpure™ (Boehrinnger Mannheim) PCR purification kit was used to extract DNA free of primer dimers <100 bp. 500µl binding buffer was added to 100µl melted agarose gel and the mixture was transferred to a High Pure™ filter tube and spun at 13,000 rpm for 30". The glass fleece washed with 500µl wash buffer. Washing was repeated with 200µl wash buffer and the filter tube was transferred to a fresh tube. 100µl water, pH 8 was added to the filter tube and the DNA eluted by spinning the tube at 13,000 rpm for 30".

5.3. Gel extraction with silica

The Prep-a-Gene (Bio-Rad) and Hi-Pure (Boehrinnger Mannheim) gel extraction kits make use of specially prepared silica matrix to selectively bind DNA. A 5µl aliquot of matrix was used per µg DNA extracted. Binding buffer was added and the mixture was heated at 50°C to dissolve the gel. The matrix was added, mixed and incubated at room temperature for 10'. The matrix was pelleted by centrifuging at 13,000 rpm for 30" and the pellet resuspended in 50 matrix volumes of binding buffer. The pellet was recovered by centrifugation and the rinse step repeated. The matrix pellet was washed thrice in 50 volumes

of wash buffer and resuspended in one pellet volume of elution buffer for 5 minutes at 50°C. The eluted DNA was recovered after centrifugation.

5.4. Extraction by boiling

DNA bands for PCR amplification were rapidly extracted from polyacrylamide gels by boiling [Upender *et al*, 1995]. A gel band < 4μl was excised along with the filter paper and soaked in 100μl water in an eppendorf tube. The mixture was overlaid with 2 drops of mineral oil and heated at 100°C for 15'. 2.5μl of 10 mM PCR-compatible dye cresol red, pH 8.5 was added to visualize the eluate.

6. CLONING

6.1. Ligation

Ligation was done for 16 hours in 10μl 1x ligation buffer [Sambrook *et al*, 1996] containing 1 Weiss unit of T4 ligase. Blunt-end ligation was done at room temperature while sticky-end ligation was at 4°C. A vector:insert molar ratio of 1:3 - 3:1 was used and the amounts of nucleic acid required calculated as,

$$\frac{\text{ng insert}}{\text{ng vector}} = \frac{\text{size of insert}}{\text{size of vector}} \times \text{insert:vector molar ratio}$$

6.2. T-vector cloning

The terminal deoxynucleotidyl transferase activity of *Taq* polymerase adds an A residue to most amplicons [Clarck, 1988] due to the lack of 3′→5′ exonuclease activity. Cloning such ligands to a T-vector possessing a 3′-terminal T overhang [Marchuk *et al*, 1991] can increase efficiency of cloning by 80% [Hadjeb & Berkowitz, 1996]. A 5μg aliquot of pSK was digested with 10units of *Eco*RV in 50μl volume for 2 hours at 37°C. The blunt-ended plasmid was precipitated and T-tailing was performed in 100μl PCR buffer with 2 mM dTTP and 5 units *Taq* polymerase. The mixture was incubated at 70°C for 24 hours and the DNA precipitated. To eliminate non-T tailed vectors from the plasmid pool, three Weiss units ligase was added and the DNA allowed to self-ligate. The mixture was electrophoresed to separate the fast moving linearized

plasmid from circularized concatemers of blunt ended molecules. The T-vector was then extracted from the excised gel slice.

6.3. Amplicons generated by primers containing restriction endonuclease sites

Restriction sites were added to the 5′ end of primers to enable cloning of PCR products. However, residual *Taq* contamination of amplicons can lead to complete or partial filling in of the 3′ recessed termini of digested amplicons and render them non-complementary to the vector ends [Kaufman & Evans, 1990]. The polymerase was therefore inactivated by protease digestion before restriction enzyme digestion. Proteinase K (20μg μl^{-1}) was added at 1.25μl per 25μl PCR product(s) and incubated at 65°C for 30'. The mixture was phenol extracted and the amplicons were purified using the Highpure™ system.

6.4. Dephosphorylation (using Gibco BRL kit)

Calf intestinal alkaline phosphatase (CIAP) was used to hydrolyze 5′-triphosphate groups of DNA, RNA and phosphorylated oligomers. CIAP was added to nucleic acid preparations at 1U 100$pmol^{-1}$ DNA 5′-protruding ends, or 1U $pmol^{-1}$ 5′-blunt or recessed ends (1μg 1kb DNA = 3.03 pmol of ends) and 1U 100$pmol^{-1}$ RNA ends. The reaction mix was made up in 1x phosphorylation buffer with nucleic acid, CIAP and water added to make up final volume. A CIAP (Gibco BRL) formulation containing 27.8U μl^{-1} was diluted in dilution buffer and the required amount added at a volume equal to, or greater than 1/10th volume of reaction mix. DNA with cohesive ends was incubated at 37°C for 30' while other nucleic acids were incubated at 50°C for 1 hour. CIAP was inactivated by heating the reaction mixture at 75°C for 10' [Ausabel *et al*, 1998].

7. TRANSFORMATION

7.1. XL 1-Blue electro competent cells

E. coli strain XL 1-blue ([F′, *proAB lacI*^{q}*ZΔM15*] *Tn10(Tet*^{r}*)*) maintained on LB Tetracyclin (10μg ml^{-1}) plates was used to prepare electro-competent cells

[Wilson and Gough, 1988]. Two hundred milliliters of fresh LB.Tet medium was inoculated with 2ml of an overnight culture of XL 1-blue and shaken at 250 rpm at 37°C to an OD_{600} of 0.4. The culture was incubated on ice for 30' and the cells pelleted by centrifuging at 5700 rpm for 15' at 4°C. The pellet was resuspended gently in 200ml ice-cold 10% glycerol and centrifuged at 5700 rpm for 15'. The pellet was washed again in 100ml 10% glycerol and centrifuged as above. The friable pellet was washed in 4ml glycerol, centrifuged and resuspended in 400µl (0.2% original volume) glycerol and stored as 40µl aliquots at -70°C [Dower *et al*, 1988].

7.2. DH5α chemically competent cells

E. coli strain DH5α ([F′, *lacI*^{q}ZΔM15] *thi*-1) was used for producing competetent cells for heat-shock transformation. A single colony of strain DH5α maintained on minimal plates supplemented with thiamine-HCl (for selection of F′ episome) was grown overnight in 25ml LB. A 2ml aliquot of overnight culture was used to inoculate 200ml of fresh LB and shaken at 250 rpm at 37°C to an OD_{600} of 0.4. The culture was incubated on ice for 2 hours and cells pelleted by centrifuging at 3000 rpm for 20' at 4°C. The pellet was resuspended in 10ml ice-cold trituration buffer and the volume was made up to 200ml and incubated on ice for 45'. The tubes were spun at 2500 rpm for 10' at 4°C and the cell pellet resuspended in 20ml ice-cold trituration buffer. Centrifugation was repeated and the pellet was resuspended in 1ml trituration buffer. 80% glycerol was added dropwise to a final concentration of 15% (v/v), mixed well and the cells stored as 200µl aliquots at -70°C [Ausabel *et al*, 1998].

7.3. Electroporation

Competent cells were thawed and placed on ice. Sterile cuvettes (0.1cm electrode gap) and the chamber slide were chilled [Dower *et al*, 1988]. The ionic concentration of the DNA preparations was reduced by EtOH precipitation [Bottger, 1988], or dilution [Wilson and Gough, 1988]. One to two µl DNA was mixed well with 40µl cells and placed briefly on ice [Chassy and Flickinger,

1987].The Gene Pulsar apparatus (Bio-Rad) was set to 25μF capacitance and 1.8kV voltage and resistance of pulse controller adjusted to 200Ω. The electroporation mixture was transferred to the cuvette and pulsed at a field strength of 25kV cm^{-1} for 4-5 milliseconds (msec). The cells were recovered in 1ml SOC medium immediately. Recovered cells were shaken at 37°C for 60' and plated for blue/white selection.

7.4. Heat shock transformation

A 200μl aliquot of DH5α competent cells (Invitrogen) was thawed on ice and 2μl of the ligation reaction was added. Cells were incubated on ice for 20' and then heat shocked at 42°C for 45". Tubes were transferred to ice for 2', 800μl SOC added and cells incubated at 37°C for one hour, with shaking.

7.5. Blue White selection [Ausabel *et al*, 1998]

Transformed cultures were plated on LB media containing 50μg ml^{-1} ampicillin, 40μg ml^{-1} X-gal and 100μg ml^{-1} IPTG and incubated at 37°C for 16 hours. The plates were stored at 4°C for several hours to accelerate blue colour development. Active blue colonies were blue in the center and dense blue at the periphery. White colonies occasionally showed a blue color in the center.

8. BLOTTING

8.1. Southern Blotting

Genomic digests were electrophoresed in a 0.7% agarose gel at 1.5V cm^{-1} for 16 hours. The gel was treated with 0.25 N HCl till tracking dyes changed colour to golden yellow. The gel was rinsed in distilled water and shaken in denaturing buffer (0.4 M NaOH/1.5 M NaCl) for 30' till dyes regained their colour.

In the alkali blotting method, capillary blot transfer was performed in transfer buffer containing 0.4 M NaOH/1.5 M NaCl, overnight. When transfer was done in 20xSSC, the gel, after denaturation, was rinsed in distilled water and shaken in neutralizing buffer for 30'. After transfer, the membrane was washed briefly in

2xSSC, air dried and the DNA fixed to the membrane either by baking at 80°C for 2 hours, or by UV cross-linking at 260-280 nm (short) wavelength for 30".

Capillary blotting was performed by placing the gel face down on a Whatman (MM1) filter wick whose ends dipped from a raised platform into the transfer buffer. A sheet of positively charged membrane (Hybond™ N^+, Amersham) cut to the size of the gel was wetted with distilled water and placed over the gel. Three sheets of Whatman 3MM paper of same size were wetted and placed over the membrane, followed by a stack of absorbent paper weighted down with a box of filter tips. Slightly positively charged membranes (Boehringer Manheim) were used with DIG-labelled probes.

8.2. Colony blotting

A circle of Hybond™ N^+ membrane (0.45μM pore size) was cut to the size of the colony plate and placed on the agar surface containing colonies. Three acentric punctures were made in the membrane to orientate the colonies. The membrane was stripped off the plate after 1' and placed, colony side up for 7' on a pad of filter paper soaked in denaturing buffer. It was transferred to another pad soaked in neutralizing solution for 3'. The neutralizing step was repeated, the blot washed briefly in 2x SSC, air-dried and UV cross-linked.

8.3. Northern blotting

RNA can be electrophoretically size fractionated under denaturing conditions [Lehrach *et* al, 1977] and blotted on to membranes. RNA samples were denatured at 65°C for 10' in 1x MOPS buffer containing 50% formamide and 6% formaldehyde (HCHO) and electrophoresed on 1.5% MOPS agarose gel containing 0.66 M HCHO [Fourney *et al*, 1988]. Electrophoresis was done at 0.75V cm^{-1}. The low concentration of HCHO used (0.66 M compared to the standard 2.2 M) produces less autofluoresence and allows RNA to be directly visualized using short UV radiation (254nm). Integrity of RNA was confirmed using the 18*S* (2.37kb) and 25*S* (6.33kb) rRNA bands as internal controls. Capillary blot transfer was done as previously described, in 20XSSC buffer.

8.4. siRNA blotting

Small RNA was resolved in a 15% polyacrylamide-7M urea gel in TBE buffer, electro-transferred to nylon membranes and UV cross-linked.

9. HYBRIDIZATION

9.1. Southern Hybridization

9.1.1. ^{32}P labelled probes

For hybridization with ^{32}P labelled probes, blots was prehybridized for >1 hour in 5-10 ml Church and Gilbert (CG) buffer in Hybaid tubes. The denatured probe was added directly to the buffer and the blots hybridized for >18 hours. Probe concentration was kept to <10ng ml^{-1} hybridization buffer to prevent competitive probe reannealing. Blots were washed thrice in 100ml 3x SSC, 0.1% SDS for 10' at 37°C, with shaking to remove nonspecifically bound probe. The blot was sealed between plastic covers and autoradiographed at -70°C with an intensifying screen, which produces >10-fold signal enhancement. T_m was calculated using the formula, 81.5°C+16.6 *log* molar concentration of monovalent cations+41(mole fraction of G+C)-0.62 (%formamide)-600/length of probe. An annealing temperature 10-20°C below T_m of native DNA produces maximum rate of hybridization [Wahl *et al.,* 1987]. Blots intended for reprobing were stripped with 1% SDS; a boiling solution of SDS was poured over the blot in a tray and the solution allowed to cool to room temperature.

9.1.2. DIG labelled probes

For hybridization with digoxigenin (DIG) labelled probes, blots were prehybridized in DIG EasyHyb buffer (Roche) which is used like formamide-based buffers. T_m was calculated as 49.82+0.41(%G+C)-600/l and hybridization temperature was set 20-25°C lower than T_m. Prehybridization was done at 37-42°C for 30'. dsDNA probes were denatured at 95° for 10', chilled directly on ice and added to 5ml preheated Easy Hyb buffer. The prehybridization buffer was discarded and blots hybridized with probe overnight. Blots were washed

twice in 2xSSC, 0.1% SDS for 5' at room temperature, and twice in 1xSSC, 0.1% SDS for 15'. Membranes were equilibrated in wash buffer for 1' and incubated in blocking solution for 60'. They were transferred to blocking solution containing alkaline phosphate conjugated anti-DIG Fab' fragments and incubated for 30'. The membranes were washed in wash buffer and equilibrated in detection buffer for 2'. The chemiluminiscent substrate CSPD was diluted 1:100 in detection buffer and blots incubated in it for 5'. The blots were sealed between plastic covers and autoradiographed at room temperature. The probe in Easy Hyb buffer was saved at -20°C for reuse and denatured at 68°C for 10', prior to use.

9.2. Northern hybridization

Northern hybridization was done in a manner similar to that described for Southern hybridization. The main differences were that solutions were made in DEPC treated water and T_m was calculated as 79.8°C + 18.5 log $[Na^+]$ + 0.58(%G+C) + 11.8(%G+C)^2 - 0.35(%formamide) - (820/l).

9.3. Colony hybridization

Oligonucleotides, or RNA end labelled with ^{33}P were used as probes. For oligos up to 18nt long, the T_m was calculated as 4(G+C)+2(A+T)°C. For oligos up to 70nt long T_m= 81.5+41(%GC)-675/l. Hybridization with oligos (15-25nt) was done at 5°C below T_m. T_m was reduced by 5°C for each 1% base mismatch. Blots were washed briefly, 3 times for 5' each. They were air dried and applied directly to the film and autoradiographed at room temperature. Intensifying screens were not used.

9.4. *In situ* hybridization

In situ hybridization techniques are used to localize gene expression at the cellular level [John *et al*, 1969; Pardue and Gall, 1969]. Pine tissue was cut into 5-7mm pieces and fixed in formaldehyde-acetic acid-alcohol (FAA) overnight at room temperature. The tissue was dehydrated through an ascending alcohol

series and left overnight in 95% EtOH containing 0.1% Eosin. The tissue was incubated in 100% EtOH and the dehydrated tissue was permeated with xylene by taking it through a graded series of xylene:EtOH mixtures. The 100% xylene incubation was for 1 hour and was repeated thrice. The tissue was cleared in xylene and embedded in paraplast inside cardboard boats and stored overnight at -20°C. Ten μm sections were cut in continuous ribbons and floated on a 42°C water bath. They were mounted on poly-L-lysine coated slides and baked overnight at 45°C. The paraplast was removed by washing slides twice in xylene for 10' followed by rinsing 15 times each, in a descending EtOH series. The hydrated sections were incubated with proteinase K (1μg ml^{-1}) in 100 mM Tris-HCl, pH 7.5, 50 mM EDTA at 37°C for 20'. The slides were treated with a 0.25% solution of acetic anhydride in 100 mM triethanolamine for 5' at RT.The tissue sections were dehydrated and dried under vacuum for 1 hour.

The tissue sections were overlaid with hybridization solution containing DIG-labelled probe and hybridized overnight at 42°C in a moist chamber. Slides were given two low stringency washes in 2xSSC at room temperature for 30' and at 65°C for 1 hour followed by a high stringency wash in 0.1xSSC at 65°C for 1 hour. The slides were equilibrated for 5' in antibody buffer and then incubated with AP conjugated DIG-antibody for 2 hours in a moist chamber. The slides were washed in antibody buffer and equilibrated in substrate buffer for 5'. Forty five μl NBT and 35μl BCIP were dissolved in 10ml substrate buffer and incubated with tissue in the dark till colour developed. Slides were washed in 95% EtOH for 1 hour, rinsed in water, sections dehydrated and mounted.

10. LABELLING

10.1. Random primer labelling

The *redi*prime labelling system (Amersham) was used to produce radio-labelled probes [Feinberg and Vogelstein, 1983]. Template DNA, purified, or in restriction enzyme buffers was diluted to 2.5 to 25ng in 45μl sterile water,

denatured at 95°C for 5' and added to the labelling mix; the proprietory mix consists of a buffered solution of dATP, dGTP and dTTP, exonuclease-free Klenow enzyme and random nonamers (9-mer). A 5μl aliquot of ^{32}P dCTP was mixed with the labelling mix and incubated at 37°C for 30' to 2 hours. The reaction was stopped with 5μl 200 mM EDTA and the probe denatured.

10.2. 5′ end labelling

Polynucleotide kinase (PNK) was used for end labelling. In the case of tRNAs, their 5′-triphosphate end was dephosphorylated and the tRNA then fragmented by alkaline hydrolysis to provide more 5′-hydroxylated ends for labelling [James, *pers. com*, 1992]. 5μg tRNA was dephosphorylated at 50°C for 1 hour and heated at 65°C for 10' to inactivate PNK. The mixture was made up to 120μl final volume in DEPC-water, an equal volume of 0.1 M NaHCO3, pH 9 added and incubated at 95°C. Empiricaly, a 3 minute hydrolysis is sufficient for most RNAs [Wilkins, 1999]. However, since this is not known in the case of tRNAs, aliquots of 30μl were removed at 3' intervals and chilled on ice. Aliquots were pooled and 1/10 volume 3 M NaOAc and 2.5 volumes EtOH added and incubated at -20°C for 10'. The RNA was pelleted by spinning at 13,000rpm for 30'. The dried RNA was labelled in a 50μl mix containing 43μl water, 5μl PNK buffer, 1μl PNK and 1μl γ^{33}P ATP at 37°C for 30'. Another 1μl PNK was added and incubation continued for 30'. The enzyme was inactivated by heating at 65°C for 10'.

Oligonucleotides are hydroxylated at their 5′-end and therefore, need no dephosphorylation. One femto mole (fmol) of oligonucleotide was labelled in a final volume of 20μl. The labelling mixture contained 15μl water, 2μl PNK buffer, 1μl (100 μM) oligonucleotide, 1μl PNK and 1μl γ^{33}P ATP. The mixture was incubated at 37°C for 60' and then heated at 65°C for 10' to inactivate the enzyme. The probe was denatured with 8μl 4 M NaOH prior to hybridization.

10.3. PCR labelling

10.3.1. DIG-labelling

DIG probes were generated using expand™ (Roche) high fidelity enzyme and a 10x dNTP labelling mix consisting of 2 mM each of dATP, dGTP, dCTP and 1.3 mM dTTP and 0.7 mM alkali-labile DIG-11-dUTP. 1ng of template was amplified in 25μl 1x PCR buffer (with $MgCl_2$) containing 200μm dNTP, 100 nM primers and 1.3U enzyme. PCR consisted of 30 cycles of 95°C x 30″, 55°C x 30″ and 68°C x 30″ following an initial denaturation at 95°C for 1′. The reaction was run alongside an unlabelled control reaction on agarose gel and labelling confirmed by gel shift of the slow moving labelled probe.

10.3.2. ^{32}P labelling

^{32}P labelled probes were produced according to Wilkins [1991]. ^{32}P dATP (3000Ci $mMol^{-1}$, 10Ci μl^{-1}) containing 0.33 μM dATP μl^{-1} was used to label probes. A cold dNTP mixture containing 2 mM dGTP, dCTP and dTTP and 0.05 mM dATP was prepared. 1μl of this mixture was added to 10μl of PCR mix containing 1μl 10x standard PCR buffer, 1.6μl forward and reverse primers (1 μM), 1μl template DNA (20ng), 0.25μl *Taq* (0.25U), 5μl ^{32}P dATP (1.7 μM) and 0.15μl water. The mixture is thus 6.7 μM with respect to dATP. The PCR mixture was subjected to 20 cycles of 95°C x 30″, 55°C x 30″ and 72°C x 30″

10.4. *In vitro* transcription

Sense and anti-sense probes labelled with the steroid hapten DIG were synthesized for *in situ* hybridization from cloned templates. Plasmid was prepared without the use of RNase A to avoid degradation of transcribed RNA by residual RNase activity. Plasmid was linearised at a site adjacent to the cloned insert corresponding to the 5′ or 3′ end of the RNA using a restriction enzyme that leaves a 5′-overhang, or blunt ends and gel extracted. One μg plasmid was used in a 20μl transcription reaction containing 1μl RNase inhibitor

(20U), 2μl NTP mix (10 mM), 2μl 10x buffer and 2μl RNA polymerase (40U) and incubated for 2 hours at 37°C. Template DNA was removed by incubating with 1μl RNase-free DNase I (10 units) at 37°C for 15 minutes and DNase was inactivated by heating at 95°C for 10 minutes. The NTP mix was made of 10 mM each ATP, CTP and GTP, 6.8 mM UTP and 3.5 mM DIG-UTP. Labelled probes were quantified by dot blotting. Two-and-a-half μl of labelled control RNA (250ng DIG-labelled neo-RNA) and 1μl of the labelled reaction was serially diluted 10-fold, dot blotted and visualized by chromogenic reaction.

11. POLYMERASE CHAIN REACTION

The polymerase chain reaction [Saiki *et al*, 1985] using *Taq* DNA polymerase alone [Saiki *et al*, 1987;Mullis, 1990], or in combination with thermostable, proof-reading DNA polymerases [Barnes, 1994], or reverse transcriptase [Veres *et al*, 1987] was used to amplify sequences from DNA and RNA [Keohavong *et al*, 1988].

11.1. Standard PCR

The primers were 18-24nt long, with closely matched T_m (within 2-3°C) and a GC content of 40-60%. The computer program Oligo 4.0 based on nearest-neighbour thermodynamic parameter [Rychlik & Rhoads, 1989] was used to design primers [Kwok *et al*, 1990]. Primers were used at 40-160 nM concentration. Depending on the number of PCR cycles, $1.6x10^3$-$4x10^4$ copies of template DNA were used per 25μl reaction which equates to 20-500ng pine genomic DNA, or 0.04-1pg plasmid DNA (5kb). To minimize mutations induced by *Taq* polymerase which has an error rate of $2x10^{-4}$ bp^{-1} [Dunning *et al*, 1988], the higher template copy number was used in conjunction with lower PCR cycles when preparing amplicons for sequencing. The dNTPs were used at 200 μM concentration; when amplifying long templates as in single primer PCR, or amplifying under nonspecific annealing conditions 300 μM was used.

A typical 25µl PCR mixture contained 2.5µl 10x PCR buffer, 2µl $MgCl_2$ (50 mM), 2µl dNTP (2.5 mM), 2µl primer (1 µM) and 0.5µl *Taq* polymerase (1U μl^{-1}). The number of cycles was limited to 30 [Bell & deMarini, 1991]. When primers failed to perform, matrix analyses was done with Mg^{2+}, annealing temperature and template concentration as variables.

11.2. Touchdown PCR

Touchdown PCR [Don *et* al, 1991] was done to amplify the target sequence preferentially over other sequences that shared partial homology with the primers. The annealing segment of TD-PCR started at a temperature a few degrees above the predicted T_m and sequential cycles were run at incrementally lower temperatures, when annealing temperature falls below T_m. of nonspecific amplification. This ensures that the first template-primer hybridization takes place between reactants with greatest complementarity and by the time the annealing temperature falls to T_m of nonspecific hybridization, the target amplicon would have started its exponential amplification and outcompetes nonspecific products.

11.3. Hot start PCR

This minimizes nonspecific priming, or primer dimer formation as a result of mixing reaction components at permissive temperatures (4-25°C) [Chou *et al,* 1992] and during incubation of PCR mix at temperatures below the T_m before the temperature rises above T_m in the first denaturing step [D'Aquilla *et al*, 1991]. A PCR master mix was made with all components except *Taq* polymerase and template dNA. The mix was dispensed into 0.2µl PCR tubes, template DNA and a granule of Ampliwax® bead were added and the tubes briefly heated at 75°C to melt the wax and allowed to cool. 0.5 µl *Taq* polymerase was deposited on the solidified wax and the reaction started.

11.4. Band-stab PCR

Band-stab PCR [Bjourson & Cooper, 1992] helps to amplify a specific single product from a mixture of PCR products of different molecular weights formed

under reduced-stringency primer annealing conditions. The PCR products were electrophoresed on 1-1.5% TAE agarose gel containing EtBr. Excess water on the gel surface was removed with filter paper and the appropriate band was stabbed with a small orifice pipette tip, swirled in fresh PCR mix and reamplified. A blank lane was stabbed as negative control.

11.5. Colony PCR

Colony PCR was used to screen transformed colonies for putative recombinants. Colonies were stabbed with the small end of sterile toothpicks and swirled in pre-made PCR mix; excess colony material was avoided as it could cause PCR inhibition [Costa & Weiner, 1995]. A non-recombinant pSK colony was used as negative control.

11.6. BigDye terminator cycle sequencing

The dye terminator cycle sequencing reaction differs from earlier versions [Murray, 1989] in that it uses differentially fluorescence labelled 2′,3′-dideoxyribonucleotides (ddNTP), and consequently reaction is performed in a single tube. Sequencing reactions were performed in a 10μl volume containing 4μl proprietory reaction mix containing AmpliTaq® polymerase, 1.6 pmol T7sequencing primer, template DNA (200-500ng dsDNA or 30-90ng PCR DNA) and nuclease-free water. Reactions were cycled 25 time in a heated lid MJ thermocycler, each cycle consisting of 96°C x 10″; 50°C x 5″; 60°C x 4′. 1μl 3 M NaOAc and 25μl 95% EtOH were added and the DNA was precipitated by centrifuging at 13,000rpm for 30' and washed with 70% EtOH.

11.7. Reverse transcriptase PCR (RT-PCR)

11.7.1. One-tube RT-PCR

Using Titan™ one step RT-PCR system, the first-strand cDNA synthesized by AMV reverse transcriptase was converted to dsDNA by Expand™ (Roche) high fidelity enzyme in a single optimized RT-PCR buffer. A 1μg aliquot of RNA was

added to 1.8μl DNase (10U μl^{-1}), 2.1μl RNasin and 0.4μl $MgCl_2$ (25 mM) and incubated at 37°C for one hour. The DNase was inactivated by heating the mixture at 95°C for 10'. A 25μl RT-PCR mix contained 0.2 mM dNTPs, 160 nM forward and reverse primers, 5 mM dithiothreitol (DTT), 2-5 Units RNase inhibitor, 0.5μl enzyme mix and 5μl (5x) buffer. 20-200ng DNase treated-total RNA was added and RT was performed with a single PCR cycle at 45-60°C for 30′. The resulting cDNA after initial denaturation at 94°C for 2′ was cycled 20 times through 94°C x 30″, 55-60°C x 15″, 68°C x 15″ for 20-25 PCR cycles. An RNA control tube was included after the RT segment of RT-PCR to detect amplification of any contaminating genomic DNA in RNA preparations.

11.7.2. Standard RT-PCR

Total RNA was reverse transcribed into sscDNA and dilutions of ss cDNA were amplified by standard PCR using *Taq* polymerase. Primer was added to DNase treated-RNA, made to a final volume of 15μl with DEPC-treated water and denatured at 70°C for 10 minutes. The mixture was quenched on ice and 5μl 5x buffer and 1.25μl RNasin® (20U μl^{-1}) (Invitrogen) were added and the mixture heated at 42°C for 5 minutes. 2.5μl of 40 mM sodium pyrophosphate (pre-heated to 42°C) and 15U (μg^{-1} RNA) AMV transcriptase were added to the (preheated) 25μl mixture. The mixture was incubated at 42°C for one hour. Aliquots of the reaction were then amplified directly in standard PCR.

12. METHODS TO ISOLATE GENE FLANKING SEQUENCES

12.1. Single primer PCR

Under low stringency-annealing conditions mispriming can occur in PCR, when a primer binds to an imperfect complementary sequence [Kwok *et al*, 1990] to generate multiple bands. When a single gene-specific primer is used under such conditions it could prime ss products from the true complementary site as well as from partially complementary loci. When two such sites lie in reverse orientation facing each other, within a distance that could be spanned by the

polymerase they would form dsDNA products. Some of these can be expected to result from priming between true and false priming sites.

12.2. Asymmetric PCR and ssDNA capture using paramagnetic beads

Genomic DNA was amplified by asymmetric PCR using an excess of gene-specific reverse primer. This results in the production of ssDNA amplicons extending into the 5′ flanking region. The ssDNA was affinity purified using biotin-labelled gene-specific primer bound to streptavidin (SA) coated paramagnetic particles. PCR products were heated and allowed to anneal to the trapping oligonucleotide by lowering the temperature. Magnetic particle bound amplicons were separated, washed and tailed with dGTP using terminal transferase (TdT). Strands were dissociated from the trapping oligo by heating and snap cooling. Gene specific and oligo-C primers were used to amplify the homopolymered ssDNA [Koenig, 1997].

Magnetic particles were prepared by washing 100 µl polystyrene paramagnetic particles thrice in 100 µl Ten100 buffer (10mM Tris, pH 7.5, 1 mM EDTA, 0.1 M NaCl) and separating them with a magnetic seperator (MPS). SA particles were incubated with 150 pmol biotin labelled oligonucleotide in 200 µl Ten100 buffer at RT for 10′ with gentle shaking. Using MPS, particles were separated and washed twice in Ten1000 (containing 1 M NaCl) buffer and in 0.5xSSC.

12.3. Thermal asymmetric interlaced PCR (TAIL-PCR) of $tDNA^{Met\text{-}i}$

TAIL-PCR [Liu *et al*, 1995] consists of three sequential rounds of PCR. 20-40 ng Pine DNA was amplified in primary PCR with *tDNA*$^{Met\text{-}i}$ reverse primer (PR-1) and arbitrary primer 5′ ngtcgaswganawgaa 3′ (Arb-1); the primary reaction was diluted a 1000-fold and amplified in secondary reaction with the sequencing primer 5′ cgatcctgggacctgtgga 3′ (PR-2) and Arb-1; 1000-fold dilution of the secondary reaction was reamplified using reverse forward primer 5′ ttccgctgcgccactctgat 3′ (PR-3) and Arb-1. The primary PCR consisted of an

initial denaturation at 95°C for 2′ followed by 5 high stringency cycles of (94°C x 30″; 57°C x 1′; 72°C x 2′), 1 low stringency cycle of (94°C x 30″; 30°C x 2′; 72°C x 2′) and 15 cycles of 2x (94°C x 30″; 62°C x 1′; 72°C x 2′) + 1x (94°C x 30″; 44°C x 1′; 72°C x 2′). The secondary PCR comprised 12 cycles of 2x (95°C x 30″; 66°C x 1′; 72°C x 2′) + 1x (95°C x 30″; 44°C x 1′; 72°C x 2′). Tertiary PCR consisted of 20 cycles of 95°C x 30″; 44°C x 1′; 72°C x 2′.

12.4 Enriched small-insert genomic library using wheat germ tRNA (WGT)

WGT (contains tRNA and traces of 5S rRNA) was used as the trapping ligand to isolate 5S rDNA and tDNA sequences. WGT (Sigma) was cross-linked to Hybond N^+ squares and unbound tRNA was removed by washing blots for two days in hybridization buffer containing 50% formamide at 37°C and boiling the blots in 1% SDS for 3′. Pine DNA was fragmented with *Hae*III and ligated to 5′ phosphorylated adaptors [Karagyozov *et al*, 1993]. The adaptor-ligated DNA fragments were amplified by 30 cycles of PCR using a 21-mer adaptor primer. Aliqouts of the reaction, as well as negative controls consisting of *Hae*III genomic digests mixed with adaptors in the absence of ligase were electrophoresed to confirm specificity of amplification. Amplified DNA fragments were denatured and hybridized with ligand-bound filters in formamide hybridization buffer at 37°C for 24 hours. Filters were washed at moderate stringency and bound amplicons dissociated by boiling in 1% SDS for 3′. The eluted ssDNA was precipitated with linear polyacrylamide and 1-4 µl aliquots used in a second round of enrichment with fresh ligand-bound nylon squares.

13. PLANT TRANSFORMATION

13.1. Biolistic transformation

Pine embryogenic tissue was dispensed in a circle on Lepoivre nutrient medium agar in petri dishes. DNA coated gold particles (Aldrich 1.5-3µ)

prepared as per Sanford *et al* [1993] were bombarded using a DuPont He1000 biolistic delivery system. The bombardment parameters used were, rupture disk pressure 15550 psi, gap distance 16 mm, macrocarrier travel distance 11 mm and microcarrier travel distance 6 cm.

13.2. Agrobacterium mediated transformation

Agrobacterium tumefaciens strain LBA4404 harbouring the binary vector for transformation was grown at 28°C in YEB broth [Ausabel *et al*, 1998] to OD A_{600} = 0.4. Leaves of *Nicotiana benthamiana* were cut to 1 cm squares and inoculated with the bacterial culture for 2' with gentle shaking. Excess liquid was blotted off and the leaf segments placed on Murashige and Skoog (MS) agar containing 100 ng ml^{-1} napthylacetic acid (NAA) and 1 µg ml^{-1} Benzylaminopurine (BAP). The plates were kept in the dark for two days.

13.3. Tissue culture of transformants

Agrobacterium inoculated explants were transferred to selection medium consisting of MS agar containing NAA and BAP as above and 100 µg ml^{-1} Kanamycin and 500 µg ml^{-1} cefotaxime and grown at 25°C in 16 hours light (3,000 lux). Shoots appearing from the callus were transferred to shooting medium which consists of MS agar containing 100 µg ml^{-1} Kanamycin, 500 µg ml^{-1} cefotaxime and 300 ng ml^{-1} BAP. Shoots that had reached a sizeable height were transferred to the rooting medium made up of MS agar containing Kanamycin and 100 ng ml^{-1} NAA. Rooted plants were transferred to soil and grown at 25°C (16 h 12,000-lux light/8 hr dark).

13.4. Growing T1 *gus* plants for transformation with shRNA vector constructs

The primary transformants (T0 generation) resulting from transformation of *N.benthamiana* with *gus* under the control of 35S promoter were selfed and harvested seeds were selected for kanamycin-resistant progeny on MS agar

containing Kanamycin at 25°C in 16 hours light (3,000 lux). Seedlings (T1 generation) were assayed histochemically for GUS expression and two T1 plants were transferred to soil and grown at 25°C (16 h 12,000-lux light/8 hr dark). Leaves of these plants were used for transformation with shRNA vector constructs.

14. GLUCURONIDASE ASSAY

14.1. Fluorometric GUS assay

Aliquots of 100 mg bombarded tissue were extracted in GUS extraction buffer [Jefferson, 1987] with Polyclar, centrifuged at 12000 rpm for 10′ and the clear supernatant spun through Sephadex G-25 columns [Gartland *et al,* 1995] precalibrated with haemoglobin marker whose M_R is close to that of GUS. This gave around 70 μl protein extract. The protein content of extracts was determined by Bradford assay [Bradford, 1987].

GUS was assayed fluorometrically [Gartland *et al,* 1995]. Aliquots of tissue extracts containing 25 μg protein were incubated with the substrate 1mM 4-methyl-umbelliferyl-β-D-glucouronide (MUG) at 37°C; duplicate aliquots of 100μl reaction mixtures were taken at 0′, 15′ and 30′, added to 1.9 ml 0.2 M Na_2CO_3 and their fluorescence read on a fluorometer precalibrated to read 4 pmol methyl umbelliferone (MU) per relative fluorescent unit. β-glucouronidase (GUS) cleaves the protein to liberate MU and GUS concentration is expressed as pmoles MU $minute^{-1}$ mg^{-1} protein.

14.2. Histochemical GUS assay

The GUS assay solution was made according to the method of Jefferson *et al* [1987]. Tissue samples were cut in 1-2 cm sections and vacuum infiltrated for 5′ in assay solution and incubated at 37°C for 1-2 days, destained in alcohol and viewd under a binocular microscope.

CHAPTER III

5S ribosomal DNA

1. INTRODUCTION

The 5S rDNA is a Pol III transcribed housekeeping gene whose 40 kilodalton transcript is an essential component of the ribosomal nucleoprotein, where it provides a structural scaffold through interactions with ribosomal proteins for the correct assembly of the large (60 S) subunit [Szymanski *et al*, 2003].

5S rDNA is one of the most abundant gene families in eukaryotic genomes ranging from around 150 copies per haploid genome in *S. cerevisiae* to 300,000 in the amphibian *Notophthalmus viridescens* [Long & Dawid, 1980]. The number of 5S rDNA units in the genomes of a range of plants has been assessed by means of reconstruction experiments and varies from 2000 to over 75,000 [Appels *et al*, 1980; Vakhitov *et al*, 1986; Gorman *et al*, 1991].

Saturation RFLP mapping, *in situ* hybridization and PFGE (pulse field gel-electrophoresis) have revealed the chromosomal location and long-range structure of eukaryotic 5S rDNA; their organization in plants and higher eukaryotes is different to that in bacteria [Nomura, 1976], yeast [Rubin & Sulston, 1973], *Dictyostelium* [Maizels, 1976] and organelle DNA [Bedbrook *et al*, 1977] where 5S rRNA genes are physically linked to other rDNA. In most plants, the 5S rDNA genes are clustered in tandem arrays, which are localized in one or a few chromosomes. In *Lycopersicum esculentum* nearly 1000 copies are located on the short arm of chromosome 1 [Lapitan *et al*, 1991], while in wheat, rye, barley and peas they are clustered in one to three major sites, all of which are separated from the 18S-5.8S-26S rRNA gene clusters [Appels *et al*, 1980; Ellis *et al*, 1988; Gerlach & Dyer, 1980]. The exceptions are flax with its 100,000 copies of 5S rDNA [Schneeberger *et al*, 1989] and pine [Gorman *et al*, 1991] where the genes are dispersed over many chromosomes. Only a few of the 5S rDNA loci produce active transcripts;

other loci are either transcriptionally silent or produce rapidly degraded transcripts [Cloix *et al*, 2001].

Each tandem array is made up of many copies of the 5S rDNA gene (repeat units) arranged in a head-to-tail fashion. A 5S rRNA gene consists of a 120 bp transcribed (coding) region and a nontranscribed spacer (NTS) region of variable length [Fulnecek *et al*, 2002]. The spacer region contains the external promoters and transcription stop signals, while the coding region contains the internal promoters and is transcribed in its entirety to the mature 5S rRNA; by analogy to a type Pol II transcribed gene, the spacer is equivalent to the upstream and downstream regions of the latter, while the coding region corresponds to the mature mRNA region between the transcription start site and the translation stop site. In most plants the nucleotide sequence and length of the transcribed region is highly conserved, whereas the NTS has evolved rapidly. The NTS displays length and sequence polymorphism, within and between plant species. Consequently, 5S rDNA paralogs belonging to different size classes are found in plants, such as the 250 and 500 bp repeats in *Arabidopsis* [Cloix *et* al, 2000], 650 and 870 bp repeats in *Laryx decidua* [Trontin *et* al, 1999] and 500 and 850 bp repeats in *P. radiata* [Moran *et al*, 1992]. The length polymorphisms among paralogs arise primarily from sequence differences in the spacer region due to deletions, duplications, point mutations and unequal crossing-over events [Sastri *et al,* 1992].

Heterogeneity of 5S rDNA transcripts has been reported in most organisms and developmental regulation of expression has been documented in *Xenopus* and *Arabidopsis* [Cloix *et al*, 2001]. Although it is stated that 5S transcripts of plants are highly conserved [Szymanski *et al*, 1995], sequence heterogeneity does exist as found in transcripts isolated from rice embryos [Hariharan *et al*, 1987] and from polysomes of flowers, stems, leaves and siliques of *Arabidopsis* [Cloix *et al*, 2002].

Plant 5S rDNA and their biogenesis have been studied primarily in angiosperms and less in gymnosperms. Investigations into the structure, organization and nuclear localization of gymnosperm 5S rDNA were mostly undertaken to produce diagnostic markers for molecular systematics

[Trontin *et al,* 1999; Wang *et al*, 2001; Besendorfer *et al*, 2004], physical mapping [Devey *et* al, 2002; Achere, 2004] and karyotyping [Lubaretz *et al*, 1996; Liu *et al*, 2003[A]].

Even less studied are the 5S rDNA genes of *Pinus spp*, and *P.radiata* in particular. Appels and co-workers [Moran *et al*, 1992] isolated five paralogs of pine 5S rDNA, which were used to classify pine species into old- and new-world species based on the presence of a 330 bp insertion in the 5S rDNA spacer region. Gorman *et al* [1991] used a 524 bp 5S rDNA gene to study the structure and chromosomal localization of 5S rDNA genes in Monterey pine (*P. radiata*). Smith & Devey [1994] extracted microsatellite sequences from the spacer regions of 5S rDNA clones for use in pine mapping. Liu *et al* [2003] sequenced several 5S rDNA clones in a phylogenetic study of five Asian pine species.

This chapter describes the isolation and characterisation of a novel *P. radiata* 5S rDNA gene and the use of its promoter in down-regulating gene expression in transformed plants.

2. MATERIALS AND METHODS

2.1. PCR – Reagents are same as in Materials and Methods (chapter II, section10.1). Temperature conditions were optimised for each reaction. The PCR segment of RT-PCR consisted of 25 cycles of 95°C x 40′; 57°C x 20′; 68°C x 30″, preceded by a 94°C x 2′ step.

2.2. Southern hybridization – Was performed as mentioned under Materials and Methods (chapter II, section 8.1). Hybridization was performed with a ^{32}P labelled, 344 bp *5Spr20*-specific *Taq*I fragment from the spacer region at 68°C and washed at 65°C in 3xSSC followed by1xSSC and 0.1xSSC.

2.3. *5Spr20* copy number determination - Slot blot hybridization was performed according to the method of Cullis *et al* [1984]. Two-fold serial dilutions of pine genomic DNA ranging from 9 pg - 580 ng and of standards ranging from 8-512 and 10^4 - 10^6 copies of *5Spr20*–specific *Taq*I probe were slot-blotted on +vely charged membranes and hybridized as in Southern hybridization.

2.4. Vector construction for transient expression studies - Transformation vectors were constructed by sequential cloning of the pol III terminator, *gus* coding sequence and *5Spr20* promoter in pSK. The *gus* sequence preceded by a monocot ribosomal binding sequence C(A/C)(A/G)(A/C)CAUGGCG [Joshi *et al,* 1997] was obtained from pGUS (Figure-3.1) (gifted by Geenz) while the terminator and promoter sequences were amplified from *5Spr20.*

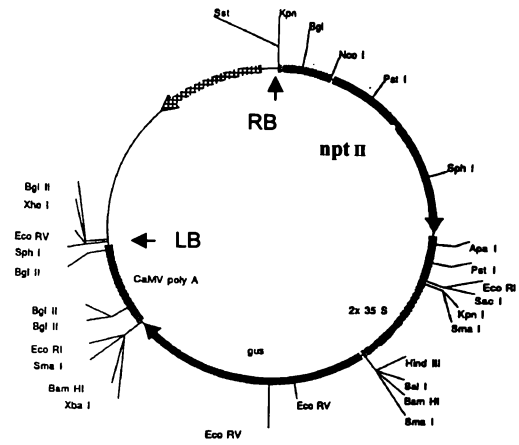


Figure-3.1. Binary vector pGUS. Figure shows the various restriction sites in the expression cassette which consists of a 2x35S promoter, gus ORF and CaMV terminator. RB = Right border of tDNA, LB = Left border of tDNA.

Primers 5′ ttgcaccctccccctctttt 3′ (containing a restriction site for *Xba*I) and 5′ tgacgcgatcgcacccgca 3′ (containing a *Sac*I site) were used to amplify a 91 bp fragment containing the putative transcription termination signal (section 3.4.1.1) which was ligated with pSK to give pT. The1.8 kb *Hin*dIII/*Eco*RI *gus* ORF in pGUS (Figure-3.1) was directionally cloned into pT to give pST. To prepare the antisense construct, a ~600 bp *Bam*HI/*Eco*RV fragment from the 5′ end of *gus* in pGUS was cloned into pT to give pAT. A 183 bp *Kpn*I/*Xho*I *5Spr20* promoter comprising the coding region and 63 bases of the immediate upstream spacer region was amplified from *5Spr20* using forward (F) 5′ taa*ggtacc*gccatggatgggcccctg 3′ and reverse primer (R) 5′ tcc*ctcgag*ggagggtgcaacactagga 3′ and cloned into the *Kpn*I/*Xho*I site of pST and pAT.

The ligated products failed to produce transformants and ligation was repeated over a range of ligand ratios without any success. *Kpn*I ends are

sometimes recalcitrant to ligation; therefore, ligations were repeated with ligands end-filled after *Kpn*I digestion and then digested with *Xho*I to produce fragments bearing a blunt and a cohesive end. The heat-shock method of transformation was also tested. However, all these attempts ended in failure.

The expression vectors were finally constructed by repeating the whole cloning exercise in the cloning vector pKS and the *5Spr20* promoter fragment was successfully cloned upstream of *gus* followed by the terminator to produce sense vector p5S and antisense construct p5A.

2.5. Short hairpin RNA vector (shDNA) for stable expression studies

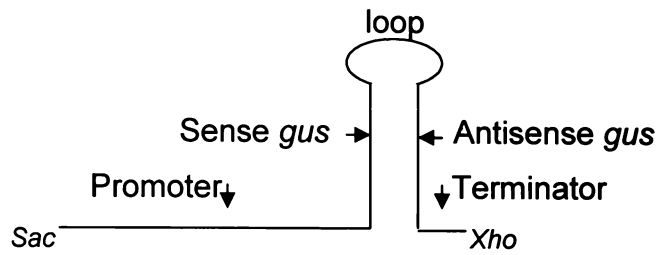


Figure-3.2. The shRNA cassette. The hairpin formed by complementary sense and antisense GUS sequences linked by a loop is flanked on the 5′ and 3′ ends by the 5Spr20 promoter and terminator, respectively.

The 21nt sequence 5′GCCGATGCAGATATTCGTAAT3′ from position 166 of the *gus* ORF and its complementary sequence for the hairpin were selected using the algorithm of Tuschl *et al* [Dykxhoorn *et al*, 2003]; the sequence 5′TTCAAGAGA3′ formed the loop region and five thymidine residues formed the terminator. The *5Spr20* promoter amplified with the same primers as in section 2.4 but containing *Sac*I and *Xba*I sites was cloned in pGEM-T to give pPro20. The cassette consisting of 5Spr20 promoter, 21nt sense *gus*, 9nt loop, 21 nt antisense *gus* and terminator (Figure-3.2) was amplified by overlap PCR [Gou & Liu, 2003] using the *5Spr20* promoter forward primer-F and two reverse hairpin primers P-1 and P-2. In the first round of PCR, pPro20 was amplified with the forward promoter primer and the reverse hairpin primer P-1 5′TCTCTTGAA-ATTACGAATATCTGCATCGGC-TCTAGAAGGGTGCAACA3′ containing the reverse sequence of the nine base loop, followed by 21 bases

of antisense *gus* and a 17 bp overlap with the 3′ end of *5Spr20* in pPro20. The diluted primary PCR was reamplified with forward promoter primer and reverse primer P-2 5′CTCGAG-AAAAA-GCCGATGCAGATATTCGTAAT-TCTCTTGAA-ATTAC containing an XhoI site, reverse terminator, followed by 21nt sense *gus* and a 14-base overlap with the loop and *gus* sequence. The cassette consisting of the promoter, hairpin and terminator was excised with *SacI-XhoI* and cloned in binary vector pGreen 0029-62 SK. The construct and helper plasmid pSoup were cotransformed into *Agrobacterium* strain LBA4404 and transformants were confirmed by back-transformation and sequencing of back transformants.

2.6. RNA isolation from *N. benthamiana*

Total RNA was extracted from leaves using TRIzol reagent (Invitrogen) according to manufacturer's instructions. The small RNA fraction which includes siRNA was isolated by the method of Llave *et al* [2000]. The total RNA solution was made to 0.5M NaCl and 10% PEG (8000) and incubated at 4°C for 30'. The mRNA and rRNA were precipitated by spinning at 12,000 rpm for 15'. Three volumes of EtOH was added to the supernatant and kept at –20°C for two hours. The mixture was spun at 12,000 rpm for 15' and the precipitate dissolved in DEPC-treated water.

3. RESULTS

3.1. Pine DNA

Pine genomic DNA was extracted from various tissues (chapter II, 1.2) for use in PCR and Southern analysis. All preparations contained high molecular weight DNA as evinced by a single migrating band (Figure-3.3) of >60 kb size.

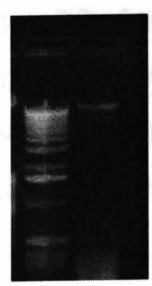


Figure-3.3 Pine genomic DNA. Lane 1. 1 Kb plus ladder, Lane 2. pine DNA. 4µl DNA was electrophoresed in 0.8% TAE agarose gel and stained with EtBr.

The quality of DNA obtained from different tissues varied according to the level of contaminants in the tissues. The DNA was tested for yield, degree of contamination and susceptibility to restriction endonucleases (Table 3.1). The fluorescence of residual material left in gel wells after electrophoresis was taken as a measure of contamination. DNA concentration was measured fluorometrically using Hoechst dye 33258 (bisbenzimide).

Table-3.1 DNA yield (μg/g fresh tissue) and quality

Tissue	Mature xylem	Juvenile plant	Mature needle
Yield	40	18	24
Polysaccharides	Low	Medium	High
Digestibility[1]	Good	Moderate	Low

[1] DNA was digested with 3units of several 6-base cutter enzymes μg^{-1} DNA. Polysaccharide contamination was estimated from residual fluorescence in gel wells.

DNA obtained from juvenile plants and xylem was routinely used in assays

3.2. Isolation of 5S rDNA

Several genomic clones exhibiting partial homology to five *P. radiata* [Moran *et al*, 1992], five Asian pine [Liu *et* al, 2003] and three gymnosperm [Brown & Carlson, 1996; Trontin *et al*, 1999; Besendofer *et al*, 2004] 5S rDNA genes were isolated during routine screening of an enriched *Hae*III subgenomic library (chapter II,12.4). Homology search of one genomic clone using BLASTn (basic local alignment search tool) revealed two high scoring segment pairs (HSPs +1 to +82 and -1 to -47), which bore very significant homology (score = 196-103, E value = 6e-48 to 7e-20) to 5S rDNA genes of gymnosperms. Seventy-six bases at the 3′ end of the clone shared 96-99% sequence identity to residues +1 to +76 of the coding sequence of 5S rDNA of several gymnosperms and angiosperms; in addition, the 47 bases immediately preceding the putative coding region showed absolute homology to sequences in identical locations in the nontranscribed region of five pine 5S rDNA genes (Figure-3.3).

ctctctctcctttccctttgattgctcaccaattcttccgctgggaggcaccacca
agtcgtggaagagggcgagctcctgtgccgaagcgtctcggatcgaaggccatggat
gggcccc^{-47}tggctcggcggtctcccttgaagaggggagggggtgagacctttgc^{+1}

gggtgcgatcataccagcgttaatccaccggatcccattagaactccgcagttaag
cgcgcttgggctagagtagt^{+76}

Figure-3.3 Partial sequence of the 5S rDNA clone. The region homologous to published 5S rDNA sequences is in blue. Bases +1 to +76 are homologous to the coding/ transcribed region of several angiosperm and gymnosperm 5S rDNA.

Therefore, this genomic clone possibly represents a putative pine 5S rDNA paralog and contains the partial sequences of its coding and spacer regions. 5S rRNA genes of most eukaryotes consist of the coding region flanked by nontranscribed spacer regions. Therefore, it is presumed that the 168 bases preceding the putative coding region of this clone belong to its spacer region.

The existence of 5S rDNA paralogs has been demonstrated in several plant species such as *Triticeae* [Appels *et al*, 1989], rice [Hariharan *et al*, 1987], pea [Ellis *et al*, 1988] and flax [Goldsborough *et al,* 1982]. In pine too, five paralogs have been reported [Moran *et al*, 1992]. Since, the 5′-most 121 bases of the spacer sequence of this genomic clone show low (<30%) homology to published pine sequences, it could be surmised that this genomic clone represents a novel pine 5S rDNA gene.

3.2.1. Amplification strategy for *5Spr20* gene isolation

The complete sequence of a novel pine 5S rDNA gene, *5Spr20*, comprising the complete coding and spacer regions was amplified from pine genomic DNA using PCR by exploiting the long-range structure of plant 5S rDNA genes [Gottlob-MacHugh *et al*, 1990]. If it is assumed that 5S rDNA genes are present as direct tandem repeats in the pine genome, as they are in angiosperms, it should be possible to amplify the complete gene of the above partial genomic clone using primers (based on the partial sequence) to read out into the unknown region in a manner reminiscent of inverse PCR [Silver & Kerrikatte, 1989].

3.2.2. Primer design and PCR amplification

The spacer region spanned by positions –48 to –168 of the genomic clone shows very low homology (0-30%) to sequences of five published pine 5S

rDNA paralogs. Therefore, this region is distinct from sequences of known paralogs and may be used to design gene-specific primers to amplify novel paralogs. Forward primer 5′ gctgggaggcaccaccaa 3′ and reverse primer 5′ ggaagaattggtgagcaatcaa 3′ (Figure-3.4) were used to amplify genomic DNA.

$^{-168}$ctctctctcctttccctt^{-150}ttgattgctcaccaattcttcc^{-128}gctgggagg
caccaccaagtcgtggaagagggcgagctcctgtgccgaagcgtctcggatcgaagg
ccatggatgggcccctggctcggcggtctcccttgaagaggggagggggtgaga^{-6}c

Figure-3.4. Partial sequence of the genomic clone showing the location of primers. The sequences in red are forward primer (-128 to –112) and reverse primer (-129 to -150) from the spacer region (See Section 4.1 for more details).

An amplicon ~ 800 bp long was amplified from genomic DNA at an annealing temperature of 58°C (Figure-3.5).

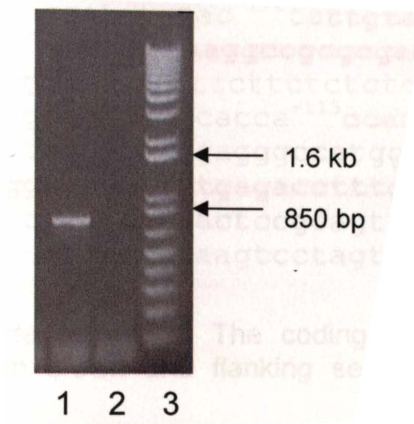


Figure-3.5. *5Spr20* amplification. Lane1. *5Spr20,* Lane 2. –ve control, Lane 3. 1Kb plus ladder. 20ng pine DNA was amplified using 25 cycles of 95°Cx40", 58°Cx15", 72°Cx1" and PCR was electrophoresed in 1% TAE gel.

The purified fragment was cloned and sequenced. Several clones displaying micro-heterogeneity arising from point mutations in the coding region and in the immediately upstream region were isolated. From among these, the genomic clone *5Spr20* containing all the features essential for transcription, such as a cytosine at the –1 position [Amerasinghe & Carlson, 1998] and the consensus sequences –SYYAA-C, S—RR-YAGU and –GRUGGG-GA at positions +50, +67 and +82, respectively [Wolters & Erdmann, 1988; Venkateswarlu *et al*,1991] was selected for further study.

3.3. *5Spr20* gene characterization

A comparison of the nucleotide sequences of *5Spr20* rDNA with its orthologs and paralogs in the plant kingdom reveals a high degree of conservation of

several features which may have implications for transcription, despite extensive sequence diversification among these genes.

3.3.1. *5Spr20* is a novel gene

5Spr20 gene is 820 bp long and comprises a 700 bp spacer and a 120 bp transcribed region. (Figures-3.6, 3.7).

Figure-3.6. The *5Spr20* nucleotide sequence: The coding sequence is in blue lettering, polythymidine regions in brown and flanking sequences conserved in published pine paralogs in red

5Spr20 exhibits length and therefore sequence polymorphism compared to published pine paralogs (see below) and is therefore a novel gene.

The subgenus *Pinus* is made up of new- and old-world species; 5S RNA genes of the new world species, of which *P. radiata* is a member, occur in two size classes of ~ 850 and 525 bp [Moran *et al*, 1992]. *5Spr20* is 820 bp long and therefore belongs to the long unit class. It is longer than pine 5S rDNA paralogs *5Spr9* (502 bp), *5Spr5* (523 bp), *5Spr7* (537 bp) and *5Spr6* (538 bp) and shorter than *5Spr10* (851 bp) [Moran *et al*, 1992].

3.3.2. *5Spr20* coding region

5Spr20 contains a putative full-length 120 bp coding region, which by analogy with angiosperm data would be expected to show high homology to coding

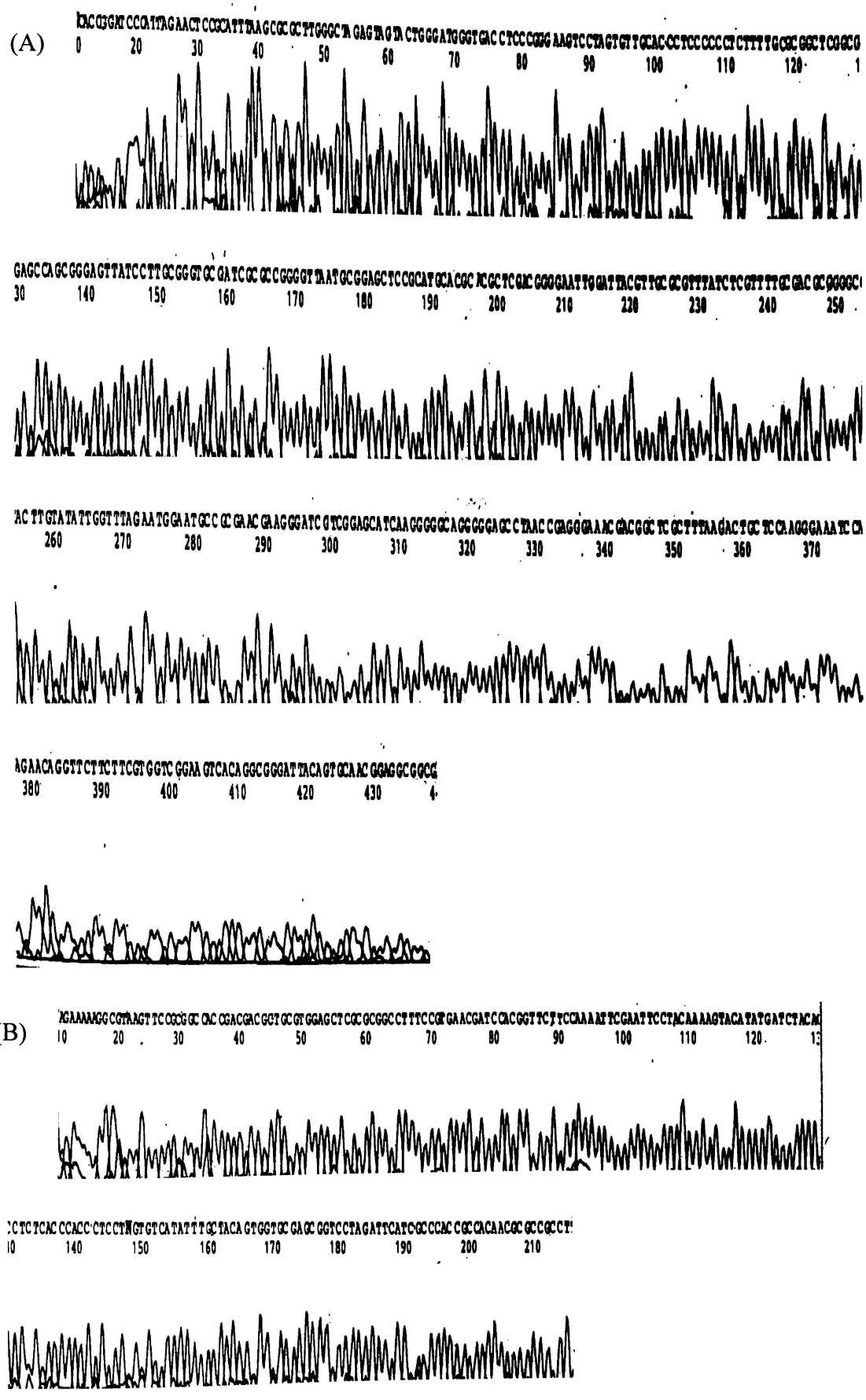


Figure-3.7. (A) Forward sequence of *5Spr20* from +25 to +120 and continuing into spacer region from –700. (B) Reverse sequence of positions –168 to –374 of the spacer region.

regions of several gymnosperm and angiosperm 5S rDNAs. The table below compares the coding regions of *5Spr20* with those of five published pine sequences of paralogs, and orthologs from *Laryx decidua* and *Gossypium .turneri* (Figure-3.8).

Base position	+1																			37
5Spr20	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	c	g	g	a	t	c	c	c	a	t	t	a	g	a
5Spr5[a]	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	c	g	g	a	t	c	c	c	a	t	**c**	a	g	a
5Spr6[a]	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	c	g	g	a	t	c	c	c	a	t	**c**	a	g	a
5Spr7[a]	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	c	g	g	a	t	c	c	c	a	t	**c**	a	g	a
5Spr9[a]	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	c	g	g	a	t	c	c	c	a	t	**c**	a	g	a
5Spr10[a]	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	c	g	**a**	a	t	c	c	c	a	t	**c**	a	g	a
Larix[b]	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	c	g	g	a	t	c	c	c	a	t	**c**	a	g	a
Gossypium[c]	g	g	g	t	g	c	g	a	t	c	a	t	a	a	c	**ac**	t	a	a	t	g	c	a	c	c	g	g	a	t	**c**	a	g	a

	50														64	67		72												80
5Spr20	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**a**	**a**	**g**	**c**	**g**	**c**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**t**	**a**	**g**	**a**	**g**
5Spr5[a]	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**a**	**a**	**g**	**c**	**g**	**c**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**t**	**a**	**g**	**a**	**g**
5Spr6[a]	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**a**	**a**	**g**	**c**	**g**	**c**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**t**	**a**	**g**	**a**	**g**
5Spr7[a]	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**a**	**a**	**g**	**c**	**g**	**c**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**t**	**a**	**g**	**a**	**g**
5Spr9[a]	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**a**	**a**	**g**	**c**	**g**	**c**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**t**	**a**	**g**	**a**	**g**
5Spr10[a]	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**t**	**a**	**g**	**c**	**g**	**c**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**t**	**a**	**g**	**a**	**g**
Larix[b]	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**a**	**a**	**g**	**c**	**a**	**c**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**t**	**a**	**g**	**t**	**g**
Gossypium[c]	a	c	t	c	c	g	c	**a**	**g**	**t**	**t**	**a**	**a**	**g**	**c**	**g**	**t**	**g**	**c**	**t**	**t**	**g**	**g**	**g**	**c**	**g**	**a**	**g**	**a**	**g**

	90																													
5Spr20	**g**	**g**	**g**	**t**	**g**	**a**	c	c	t	c	c	c	g	g	g	a	a	g	t	c	c	t	a	g	t	g	t	t	g	c
5Spr5[a]	**g**	**g**	**g**	**t**	**g**	**a**	c	c	t	c	c	c	g	g	g	a	a	g	t	c	c	t	a	g	t	g	t	t	g	c
5Spr6[a]	**g**	**g**	**g**	**t**	**g**	**a**	c	c	t	c	c	c	g	g	g	a	a	g	t	c	c	t	a	g	t	g	t	t	g	c
5Spr7[a]	**g**	**g**	**g**	**t**	**g**	**a**	c	c	t	c	c	c	g	g	g	a	a	g	t	c	c	t	a	g	t	g	t	t	g	c
5Spr9[a]	**g**	**g**	**g**	**t**	**g**	**a**	c	c	t	c	c	c	g	g	g	a	a	g	t	c	c	t	a	g	t	g	t	t	g	c
5Spr10[a]	**g**	**g**	**g**	**t**	**g**	**a**	c	c	t	c	c	c	g	g	g	a	a	g	t	c	c	t	a	g	t	g	t	t	g	c
Larix[b]	**g**	**g**	**g**	**t**	**g**	**a**	c	c	t	c	c	c	g	g	g	a	a	g	t	c	c	t	a	g	t	g	t	t	g	c
Gossypium[c]	**g**	**g**	**g**	**t**	**g**	**a**	c	c	**t**	g	g	g	a	a	g	t	c	c	**c**	g	t	g	t	t	g	c	a	c	c	**ct**

Figure-3.8. Multiple alignment of 5S coding sequences. Sequences in blue lettering correspond to A box = bases 50-64, I Box = 67-72 & C box = 80-90; a = Pine isolates - Moran *et al* (1992); b = *Larix decidua* - Trontin *et al* (1999); c = *Gossypium turneri* - Cronn *et al* (1996). Base differences are in red.

The putative internal promoters, namely the A-, I- and C-boxes are very highly conserved in all pine paralogs (Figure-3.8). The overall sequence identity is also high between *5Spr20* and its paralogs. They differ only by 1-2 residues at

positions +31, +39 and +54, respectively, the variation being due to point mutations by a-g and c-t base transition; *5Spr20* differs from *5Spr10* at position +31 by a guanine residue that forms part of a *Bam*HI restriction site, which explains some of the later Southern blot results (section 3.4.1). A BLASTn homology search revealed that the homology between coding sequences of *5Spr20* and orthologs in several higher plants is ~92-99%.

A triple G block (positions +1 to +3 in Figure-3.8) marks the beginning of the coding sequence; a pyrimidine-purine dinucleotide made up of the first G of the triple block (+1) and the preceding C at -1 are required for Pol III transcription [Fruscoloni *et al*, 1995]. The penultimate base of the coding sequence is an invariant thymidine residue. Six-base inverted repeats GGGTGC and GCACCC occur at the 5′ and 3′ ends of the gene, respectively, which are critical for the secondary structure of the transcript. In common with other paralogs, *5Spr20* has a high GC content of 76%. Normally one would anticipate problems in PCR amplification of such GC-rich templates; however, this posed no problem and the sequence was easily amplified from genomic DNA without the need for PCR enhancers or special polymerase enzyme mixtures.

3.3.3. *5Spr20* spacer region

The spacer regions of most 5S rDNA genes contain several regulatory elements which determine transcription initiation and termination.

3.3.3.1. Upstream elements

The flanking region of *5Spr20* is GC rich and has a GC content of 62%. A cytosine nucleotide is found at position -1 (Figure-3.9) as in 5S rDNA of many plant species [Amerasinghe & Carlson, 1998].

	-30 upstream region	-1 coding region
5Spr20	ttgaagaggggagggggtgagacctttgcgggtgcgatc	
5Spr9[1]	ttgaagaggggagggggcgagatctttgcgggtgcgatc	
Fir[2]	ttgaagaggggaggaggagtgatccttgcgggtgcgatc	
Flax[3]	ttcaaataagagtgatcgtgagaagtcgacgggtgcgatc	
Wheat[4]	ggataagggacgaagaccgggtaacatgtcggatgcgatc	
Petunia[5]	gataacagaagttagagcgcaaagaatgacagatgcgatc	

Pea[6]	aatatatacaattaaagcgcatttaatatcaggtgcgatc
Tobacco[7]	aataagaaaatttagagtgcaaggaatgtcagatgcgatc

Figure-3.9 Upstream sequence (-1 to -30) of plant 5S rRNA genes. Blue lettered sequence is the putative GC-rich region. The conserved cytosine residue is shown in red. 1 = Moran *et al*, 1992; 2 = Amerasinghe & Carlson, 1998, 3 = Goldsborough *et al,* 1982; 4 = Gerlach & Dyer, 1980; 5 = Frasch *et al*, 1989; 6 = Ellis *et al*, 1988; 7 = Venkateswarlu *et al*, 1991.

Similarly, a region consisting entirely of G residues centred at -15 occurs in *5Spr20* and other pine paralogs, resembling the G/C-rich region at position –13 in several angiosperm 5S rDNAs [Venkateswarlu *et al*, 1991]. The cytosine residue and the G/C region are both regarded as important transcriptional features of 5S rDNA genes and were shown to be essential for accurate and efficient transcription in *Arabidopsis* [Cloix *et al*, 2003].

3.3.3.2. Downstream elements

A stop signal consisting of four T residues is sufficient for accurate termination and production of 5S rRNA of the correct length in *A.thaliana* [Cloix *et al*, 2003]. A putative transcription stop signal [Gottlob McHugh *et al*, 1990] consisting of four T residues flanked by 6-7 C/G residues occurs eight bases downstream of the coding region at position -693 (Figure-3.6). Similar runs of thymidine are found at positions -153, -177, -257, -274 and -569. The context in which these elements occur, which could vary among genes is thought to determine Pol III termination [Bogenhagen & Brown, 1981; Gunnery & Mathew, 1999]. Whilst little is known of a gymnosperm consensus stop sequence, the 4T runs at positions -153 and -569 flanked by c/g rich or g+c dinucleotides, respectively, which are presumed to decelerate transcription also qualify as stop signals.

3.3.3.3. Repeat elements

Most plant 5S rDNAs contain several small stretches of sequences that are tandemly repeated elsewhere in the spacer but are not conserved between plant species (Figure-3.10). These repeat elements range in size from 8 bp in *Mathiola* [Hemleben & Werts, 1988] to 50 bp in pea [Ellis *et al,* 1988].

Species	Repeat	# Duplications
Tobacco[1]	aaatacat	2
Wheat[2]	gttgagaggg	3
Petunia[3]	ataacgtc	3

Figure-3.10. Repeat sequence elements in plant 5S rDNA. 1 = Venkatesvaralu *et al*, 1991; 2 = Gerlach & Dyer, 1980; 3 = Frasch *et al*, 1989.

5Spr20 also contains a 10 bp element GGCTCGGCGG at positions -685 and –46 (Figure-3.6).

A salient feature of *5Spr20* is the duplication of a 13 bp region straddling the transcription start site (ts) in the spacer region. The sequence TTGCGGGTGCGATC (-4 to +10) surrounding the start site is duplicated 657 bases upstream of the true start site. The 13 bp element is flanked by imperfect copies of sequences which exhibit ~60% identity at both sites (Figure-3.11).

	-668 -663	-662 -649	-648 -635
Upstream ts	**ggg**ag**ttatcc**	**ttgcgggtgcgatc**	gcgt**cagggttaat**
Core start site	**ggg**t**gagacct**	**ttgcgggtgcgatc**	atac**cagcgttaat**
	-15 -5	-4 +10	+11 +24

Figure-3.11. Comparison of the core start site region and the upstream duplication.

The long 5S rDNA paralog *5Spr10* also contains a similar duplication at more or less the same position in its spacer region (Figure-3.12).

Duplication in *5Spr20*	**ggg**ag**ttatcc**	**ttgcgggtgcgatc**	gcgt**cagggttaat**
Duplication in *5Spr10*	**ggg**ag**tgatcc**	**tcgcgggtgcgatc**	gcg**ccggcgttagt**

Figure-3.12. Duplication of the start site region in *5Spr20* and *5Spr10.*

The repeat is not found in the short 5S rDNA paralogs *5Spr-5, -6 -7* and -9.

3.3.3.4. Sequence homology with pine paralogs

The intercistronic region of *5Spr20* shows limited homology to those of its paralogs, which indicates that spacer regions are not subject to strong evolutionary constraints as are the coding sequence. Pairwise alignment was performed using gapped BLAST, where word size was decreased to 10,

filtering was turned off to accommodate low complexity thymidine repeats and E was set at default value. The homology between *5Spr20* and most pine 5S rDNA isolates appeared to be only 21-26% with the exception of *5Spr10*, with which *5Spr20* shared 55% homology.

3.3.3.5. Sequence/evolutionary relationship

As a result of the sequence divergence in their spacer region, pine 5S rDNA isolates fall into two phylogenetic clades with *5Spr20* showing closest relationship to *5Spr10*. CLUSTAL W [Thompson *et al*, 1994] was used to perform (progressive) multiple alignment of the spacer sequences of pine 5S rDNA paralogs; an outgroup was not included since this is an intra-species comparison. Since the sequences are <1000 bases the slow dynamic programming option was used. For the initial pairwise alignment k-tuple (word size) was set at 1 and window size at 5, for maximum sensitivity. An unrooted tree was created from similarity scores (percent identity) of the alignments; the tree has branch lengths proportional to estimated divergence from each branch node (Figure-3.13).

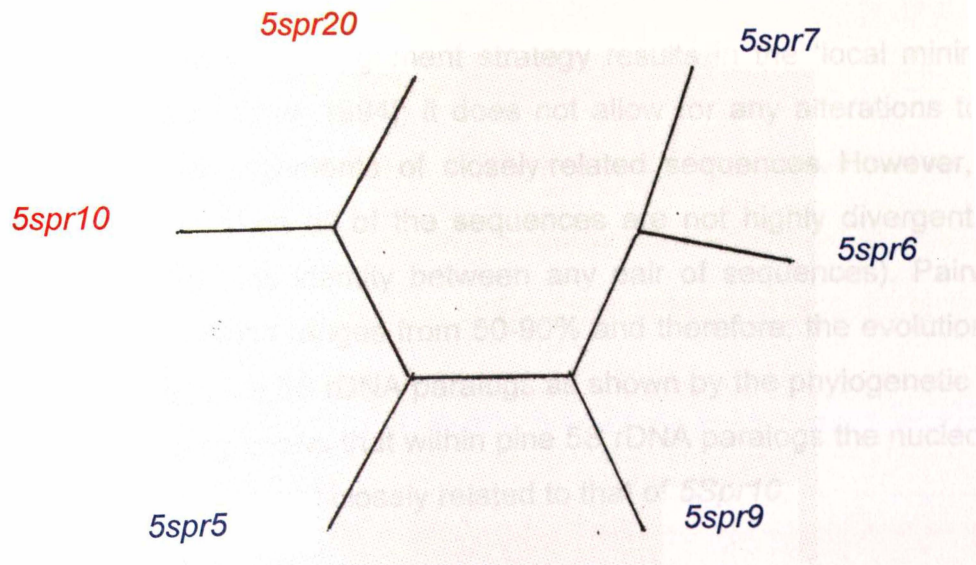


Figure-3.13 Unrooted tree showing branch lengths drawn to scale.

A rooted guide tree was computed from the neighbour-joining tree using the mid-point method and was subject to bootstrapping to give a measure of the reliability of the groupings within the tree. The consensus phylogenetic tree

derived from 100 iterations was visually displayed using PHYLIP (phylogeny inference package) package [Felsenstein, 1985] (Figure-3.14).

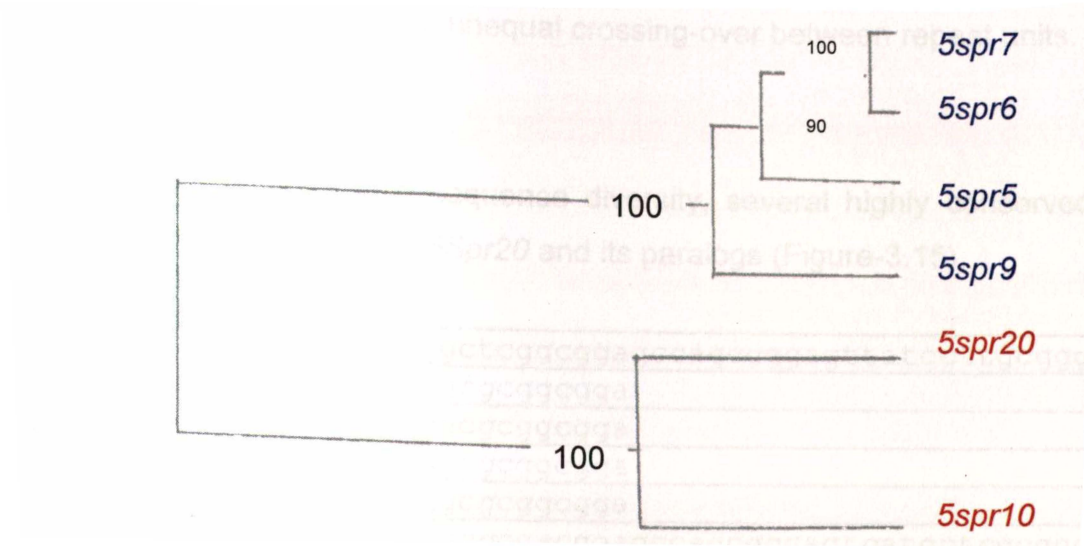


Fig-3.14 Consensus phylogenic tree of pine 5S rDNA paralogs based on spacer sequences. Numbers at the nodes indicate bootstrap values of 100 replicates. An outgroup is not included as this is a comparison of ligands within a species

The phylogenetic tree shows a clear division of the short and long paralogs into two clusters. Bootstrap values of >90 at the branch points are statistically significant and confirm that these nodes are valid.

The 'greedy' nature of the alignment strategy results in the 'local minimum problem' [Thompson *et al,* 1994]; it does not allow for any alterations to be made to the earlier alignments of closely related sequences. However, the method works well when all of the sequences are not highly divergent (ie. When there is >25-30% identity between any pair of sequences). Pairwise identity within the clades ranges from 50-90% and therefore, the evolutionary relationship among pine 5S rDNA paralogs as shown by the phylogenetic tree is valid. The grouping shows that within pine 5S rDNA paralogs the nucleotide sequence of *5Spr20* is most closely related to that of *5Spr10*.

3.3.3.6. Sequence homology with 5S rDNA orthologs

There is even less sequence homology between 5S gene spacer sequences of pine and most gymnosperms and angiosperms; 5S rDNA of the gymnosperm *Larix decidua* (larch) which comes closest to *5Spr20* shows a mere 8% base identity. Thus, the sequence divergence among the repeat

genes is more between species than within species. Such concerted evolution [Liao, 1999] is believed to result from DNA recombination, repair and replication mechanisms such as unequal crossing-over between repeat units.

3.3.3.7. Conserved regions

In spite of the spacer region sequence diversity, several highly conserved regions are recognized among *5Spr20* and its paralogs (Figure-3.15).

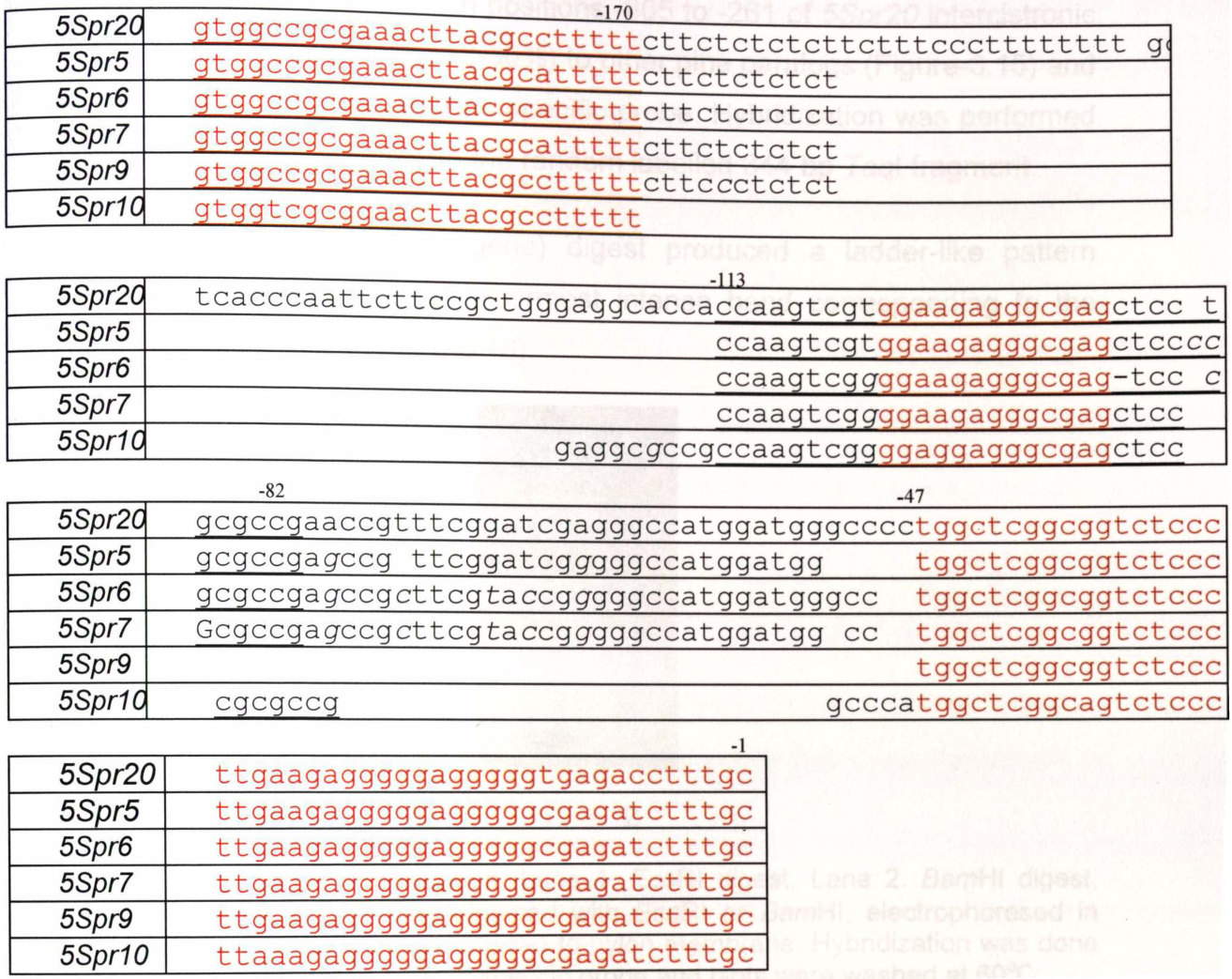


Figure-3.15 Comparison of spacer regions of pine 5S isoforms. Conserved regions are in red lettering. -1 indicates position immediately upstream of coding sequence.

The region immediately upstream of the coding sequence (-1 to -47) is almost identical in all isolates. Two regions which occur further upstream at positions -233 to -170 (64 bp) and -271 to -250 (22 bp) are also highly conserved (95%) in all pine isolates; the smaller 22 bp region has low duplex stability and is rich in adenine and thymidine bases.

3.4. Gene copy analysis

3.4.1. Southern Analysis

The multicopy nature of *5Spr20* was confirmed by Southern analysis. Southern blot analysis was performed with pine genomic DNA digested with enzymes, some of which cut within the repeat and others outside. Digestion was performed overnight using 20 U enzyme per μg DNA.

The 344 bp *Taq*I fragment from positions -605 to -261 of *5Spr20* intercistronic region exhibits low homology (3-30%) to other pine paralogs (Figure-3.15) and was therefore, used as the gene-specific probe. Hybridization was performed under stringent conditions with the random labelled 344 bp *Taq*I fragment.

The *Eco*RI (cuts within the gene) digest produced a ladder-like pattern extending upwards from the lowermost intense band corresponding to the monomer size (Lane 1, figure-3.16).

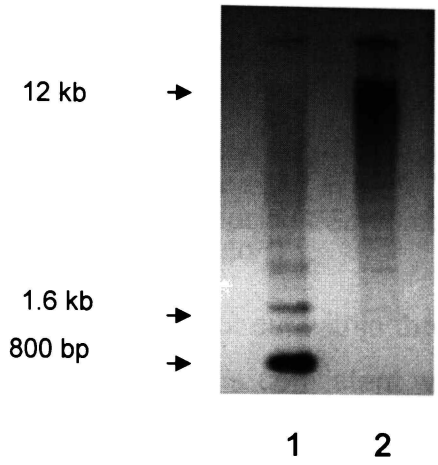


Figure-3.16. Southern hybridisation. Lane 1. *Eco*RI digest, Lane 2. *Bam*HI digest. 30μg pine genomic DNA was digested with *Eco*RI or *Bam*HI, electrophoresed in 0.8% TAE agarose gels and transferred to nylon membrane. Hybridization was done at 65°C using ^{32}P-labelled *5Spr20*-specific probe and blots were washed at 60°C.

*Bam*HI which has a unique cleavage site within the *5Spr20* coding sequence produced a similiar ladder of multimers (Figure-3.16, lane 2) with sizes that were in multiples of the monomer size; the monomer band is not visible in films because it constitutes <5% of the ladder (see below).

The *Eco*RI lane contained at least seven resolvable ladder bands and *Bam*HI lane contained 10 bands. This indicates that the minimum size of the longest tandem array of *5Spr20* genes is ~8 kb (820 bp x 10). Densitometer tracings of the *Eco*RI and *Bam*HI lanes show that nearly 74% and ≤ 5%, respectively, of *5Spr20* DNA is found in the monomer form.

In comparison to this, *Bam*HI digested pine DNA probed with the highly conserved coding region resulted in very intense ladder of bands (Figure-

3.17). This suggests that methylation is not restricted to *5Spr20* units alone but occurs in most pine 5S rDNA units.

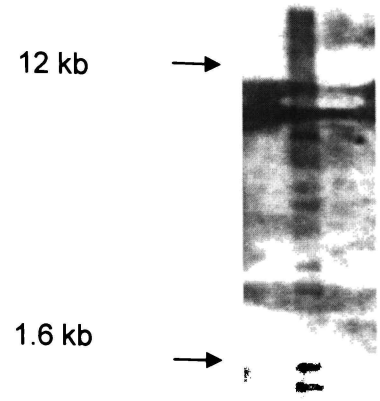


Figure-3.17. Southern hybridization with ^{32}P-labelled probe. 30μg *Bam*HI digested pine genomic DNA was electrophoresed in 0.8% TAE agarose gels and transferred to nylon membrane. Hybridization using *5Spr20* coding region probe was conducted at 65°C and blots were washed at 60°C.

*Hin*dIII and *Eco*RV (which do not cut within the gene) digests produced a high molecular weight smear which is consistent with enzymes cleaving outside a cluster of tandemly repeated genes (Figure-3.18).

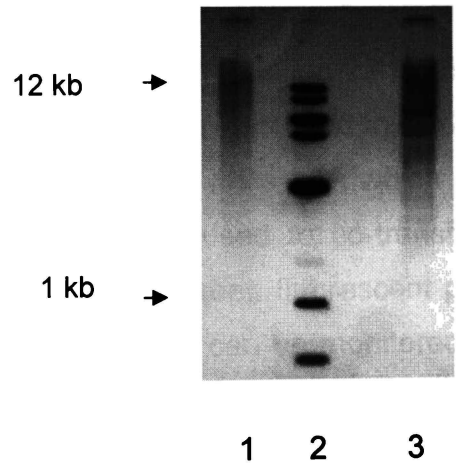


Figure-3.18. Southern hybridization with ^{32}P-labelled probe. Lane 1. *Eco*RV, Lane 2. 1 kb ladder, Lane 3. *Hin*dIII. 30μg digested pine genomic DNA was electrophoresed in 0.8% TAE agarose gels and transferred to nylon membrane. Hybridization using *5Spr20*-specific probe was conducted at 65°C and blots were washed at 60°C.

3.4.2. Slot-blot analysis

Gene reconstruction was done using slot blots (section 2.3) which accommodate a large number of DNA dilutions. A range of pine DNA and gene-specific probe dilutions were hybridised under stringent conditions and

their signal intensities were compared. Phosphor imaging of hybridized spots showed that ~18 pg genomic DNA was required to produce a detectable signal (Figure-3.19), whose intensity was nearly equal to that of 36 copies of the standard.

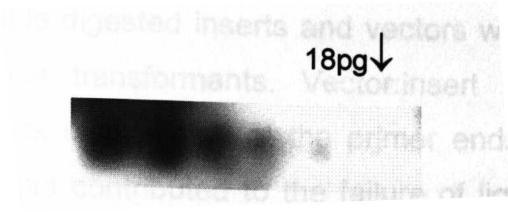


Figure-3.19 Slot-blot hybridization of pine genomic DNA. Section of the blot showing (from left to right) dots corresponding to ~0.6 ng to 18 pg genomic DNA hybridized with ^{32}P-labelled *5Spr20* gene-specific probe. Arrow shows least detectable signal from spot containing 18pg genomic DNA equal to 36 copies of the DNA standard.

The haploid pine genome (1C) weighs ~11 pg; therefore the number of *5Spr20* rDNA copies present per haploid genome is 11 x 36/18 = 22.

3.5. Transient expression studies

The strength of *5Spr20* promoter was evaluated by transient expression of GUS in pine embryogenic tissue transformed with sense expression vector p5S and antisense vector p5A (section 2.4). P5S, where *gus* is placed under the control of *5Spr20* promoter was used to study the ability of the promoter to drive GUS protein synthesis, while p5A which contains *gus* in the antisense orientation and pGUS were used to co-transform pine tissue in antisense studies. A reporter such as green fluorescent protein (GFP or mGFP) would be ideal since transformation can be monitored non-invasively [Chalfie *et al*, 1994] and it has been used successfully in pine [Tang *et al*, 2005]. However, because the *gus* reporter gene [Jefferson *et al*, 1987] was readily available in the laboratory it was used to evaluate *5Spr20* promoter efficiency.

3.5.1. Vector construction

3.5.1.1. Transcription termination signal

When the ligated product of the *5Spr20* terminator fragment (section 2.4) and pSK was transformed into XL 1-blue competent cells no transformants were obtained. A possible explanation is that ligation did not take place as a result

of residual *Taq* polymerase activity [Kaufman & Evans, 1990] and presence of residual protecting groups [Birren *et al.*, 1997].

Therefore, terminator amplicons were digested with Proteinase K to inactivate *Taq* polymerase. Double digested inserts and vectors were ligated but again they failed to produce transformants. Vector:insert ratio, wrong primer sequences and the lack of digestion at the primer ends are other possible reasons which may have contributed to the failure of ligation/transformation. Ligations were repeated at vector:insert ratios ranging from 1:7 to 3:1 and the sequence of both primers were scanned to confirm the correct orientation of the restriction sites. No band shift was observed when digested and undigested amplicons were electrophoresed in 10% polyacrylamide gels. Several endonucleases are affected when their restriction sites occur close to duplex ends [Kaufman & Evans, 1990; Moreira & Noren, 1995]. In order to ensure that terminator amplicons had ligatable ends, the amplicons were first cloned into pGEM-T vector and the excised with *SacI-XbaI*. The ~90 bp terminator fragment was gel purified and cloned into pSK to produce pT.

The insert was verified by PCR using terminator-specific primers (Figure-3.20) and by partial sequencing of the plasmid (Figure-3.21).

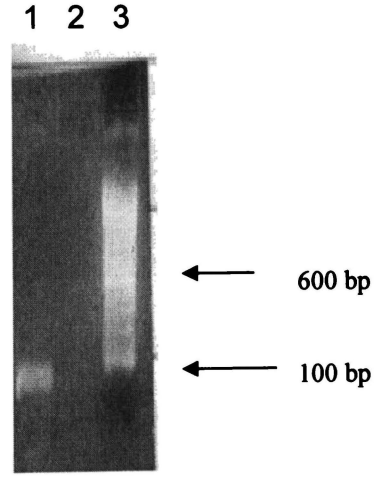


Figure-3.20. PCR of a putative clone with terminator-specific primers.PCR amplifications were electrophoresed in 2% TAE agarose gel. Lane 1. +ve clone producing a band ~90 bp, Lane 2. –ve water control, Lane 3.100 bp ladder.

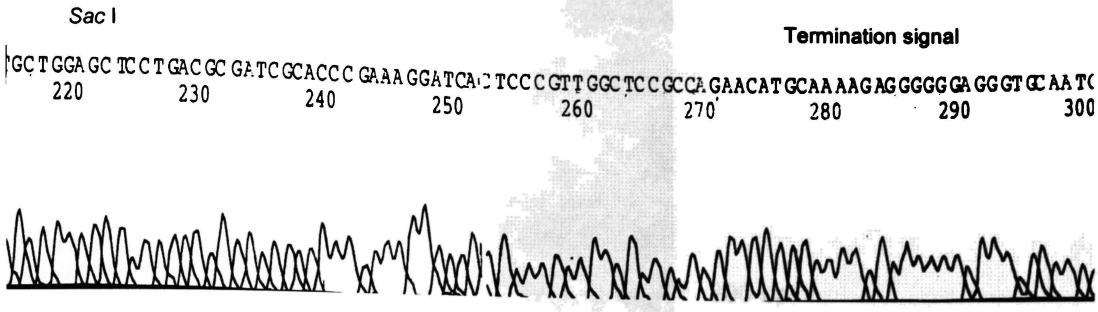


Figure-3.21. Sequence of the antisense strand of pT. Shows the *Sac*I site at position 220 and the poly-T terminator signal of *5Spr20* at position 280.

3.5.1.2. GUS coding sequence

3.5.1.2.1. Sense construct - pST

The advantage of blue/white selection is not available during the second cycle of cloning and therefore, transformants were screened by PCR and restriction analyses. *gus*-specific PCR primers 5′ atgttacgtcctgtagaaacc 3′ and 5′ ctttcttgtaacgcgctttccg 3′ (positions +1 to +21 and +104 to +125, respectively of the *uidA* ORF) targeting the 5′ end, and primers 5′ tgcgcgttggcggtaacaaga 3′ and 5′ tcattgtttgcctccctgct 3′ (positions +1682 to +1702 and +1793 to +1812, respectively) the 3′ end of *gus* insert were used to test for the ends of *gus* CDS (Figure-3.22).

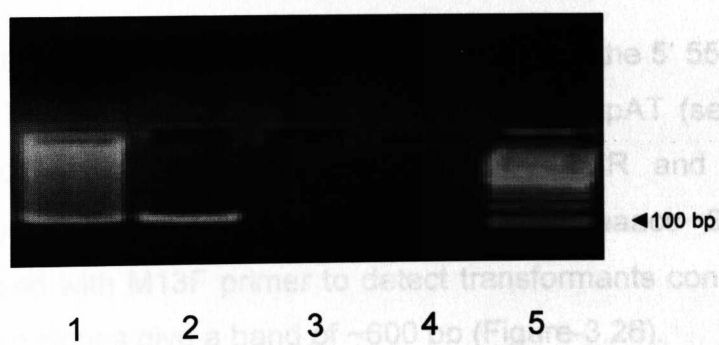


Figure-3.22. PCR of pST with gus-specific primers: Reactions were electrophoresed in 2% TAE agarose gels. Lanes 1 & 3 – PCR with 5' end gus primers; Lanes 2 & 4 – PCR with 3' end gus primers. 3 & 4 are negative (water) controls.

Restriction analysis was performed with *Eco*RV which cuts *gus* at two unique sites downstream of the ATG codon to give a fragment of approximately 250 bp (Figure-3.23).

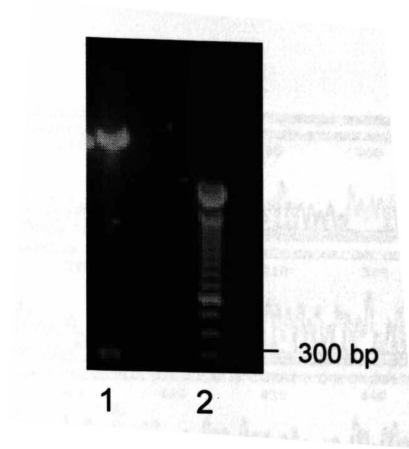


Figure-3.23 *Eco*RV restriction digestion of a putative clone. Lane 1. Positive clone, Lane 2. 100 bp ladder. Digest was electrophoresed in 1.5% TAE agarose gel.

One positive clone - pST was selected and partially sequenced to verify the junction sequences (Figure-3.24). The integrity of *gus* CDS was confirmed by sequencing from both ends (Figure-3.25).

(1) *Hin*dIII *Sal*I *Bam*HI *Sma*I
ggacagcccaagcttccaccatggcgtgcagggtcgacggatcccggg
(2) *Eco*RI *Sma*I *Bam*HI *Xba*I
ctgcaggaattccccgggggatcctctagagtcgaagctcgg

Figure-3.24 Partial sequence of junction fragments at the *gus* cloning sites in pST. (1) Restriction site *Hin*dIII at 5′ end and *Sal*I, *Bam*HI & *Sma*I sites flanking *gus* CDS. (2) Restriction sites *Eco*RI, *Sma*I, *Bam*HI and *Xba*I at the 3′ end of *gus*.

3.5.1.2.2. Antisense construct - pAT

The ~600 bp *Bam*HI-*Eco*RV fragment of pGUS containing the 5′ 556 bases of *gus* was cloned in an antisense orientation in pT to give pAT (section 2.4). Screening of positive transformants was done by PCR and restriction analysis. *gus* gene-specific primer 5′ atgttacgtcctgtagaaacc 3′ (section 3.5.1.2.1) was used with M13F primer to detect transformants containing the *gus* insert; positive clones give a band of ~600 bp (Figure-3.26).

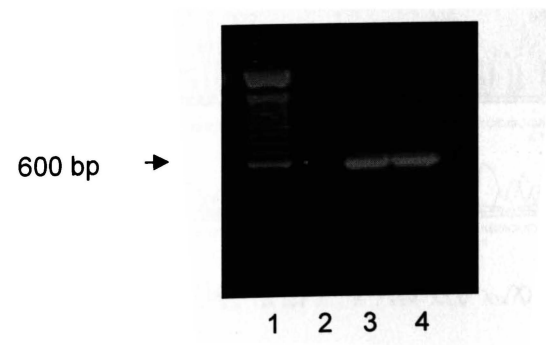
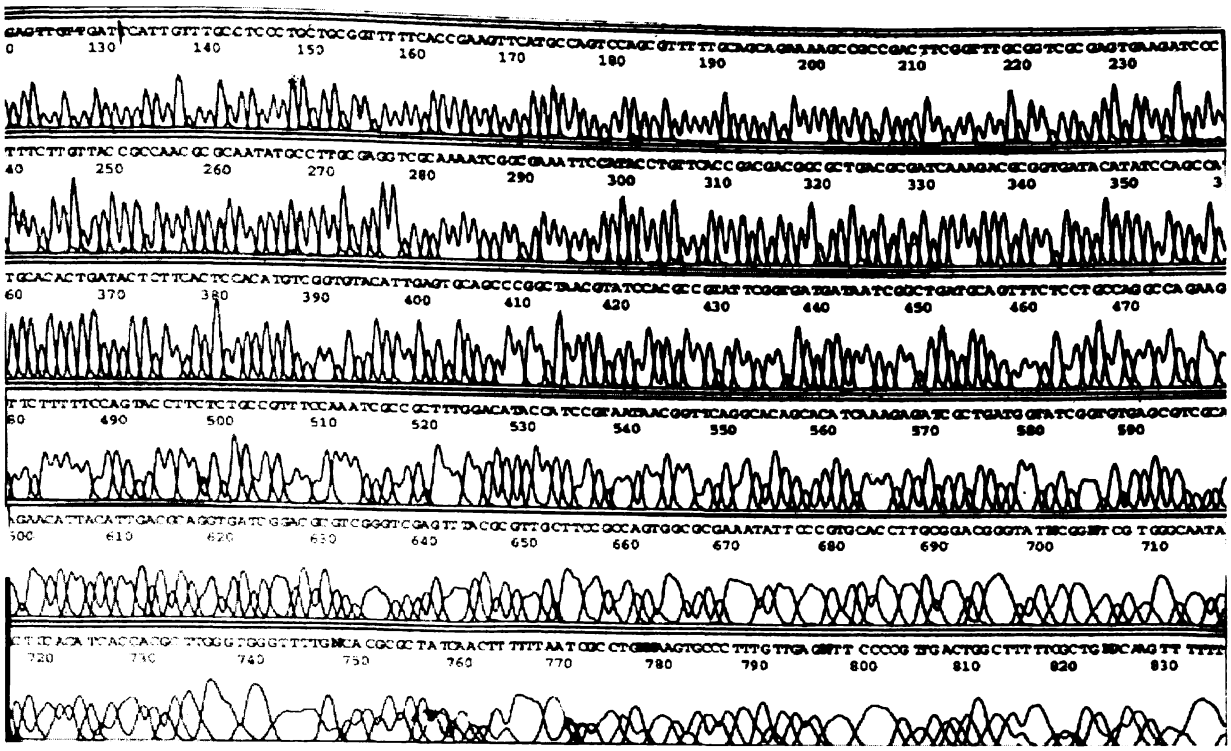


Figure-3.26. PCR of pAT clones. Lane 1. 100 bp ladder, Lane 2. –ve water control, Lanes 3 & 4 positive clones. Reactions were electrophoresed in 1.5% agarose gel.

(1) ↓ End of *gus* sequence



(2) ↓ Start codon

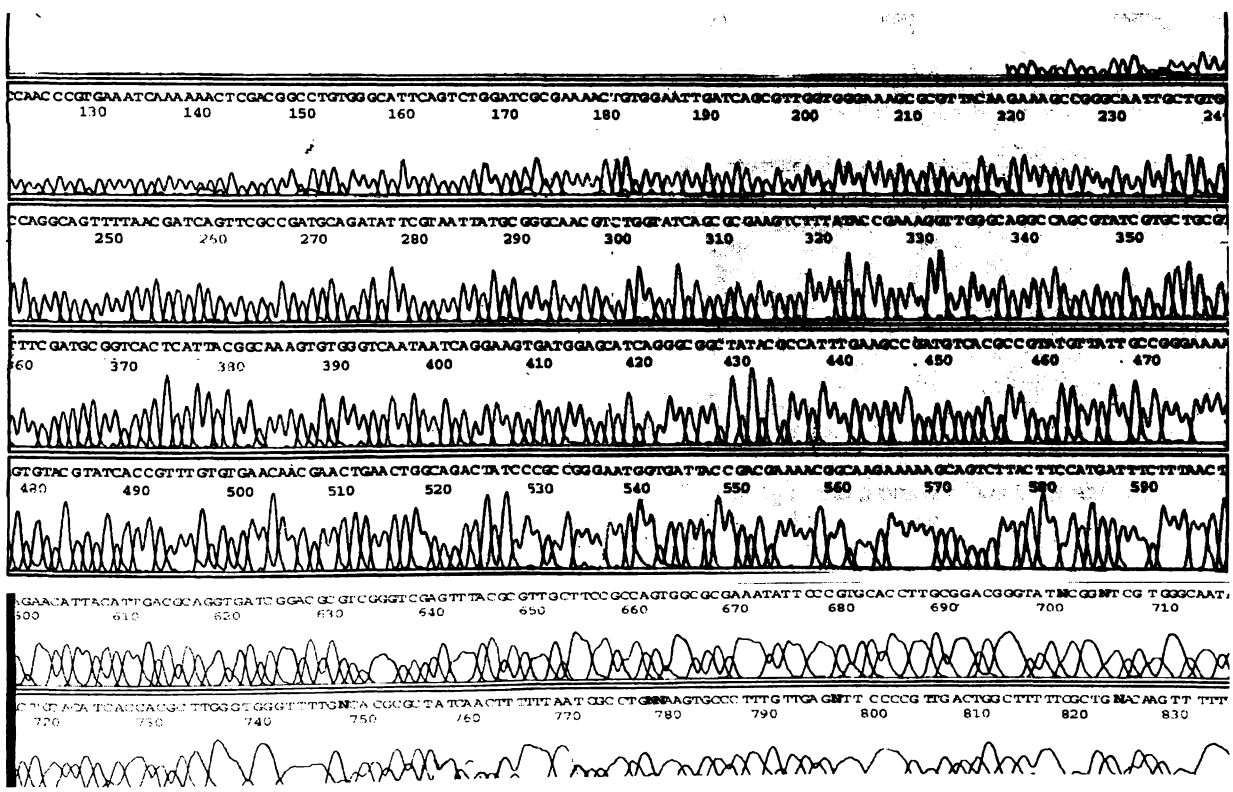


Figure-3.25 Sequence of *gus* in pST clone. (1) Reverse sequence of *gus* showing the last codon (2) Forward sequence showing start codon

Positive clones digested with *Spe*I and *Sal*I gave a ~600 bp fragment. One clone, pAT was selected for cloning the promoter (Figure-3.27).

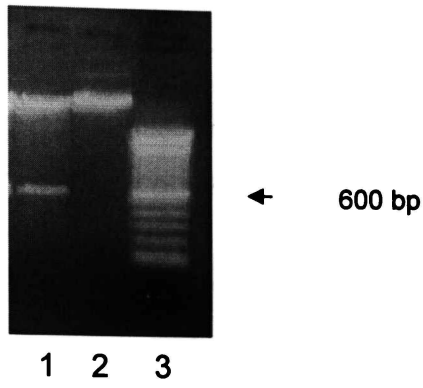


Figure-3.27. *Spe*I/*Sal*I digest of pAT. Lane 1. Digested plasmid, Lane 2. Undigested plasmid, Lane 3. 100bp Ladder Electrophoresed in 1.5% TAE agarose gel.

3.5.1.2.3. Cloning the promoter in pST and pAT

The third cycle of cloning involves placing the 5spr20 promoter upstream of the *gus* sequence in pST and pAT. Several problems were encountered at this stage (section 2.4) and these were resolved by constructing the expression vectors in cloning vector pKS. The *5Spr20* promoter fragment was successfully cloned upstream of *gus* followed by the terminator to produce sense vector p5S (Figure-3.28) and antisense construct p5A.

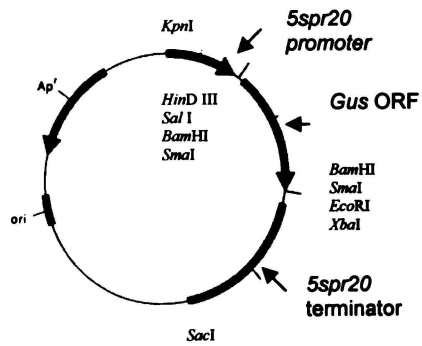


Figure-3.28 Sense vector p5S. Shows the expression cassette consisting of the promoter, *gus* ORF and the terminator flanked by *Kpn*I and *Sac*I restriction sites of pKS. Figure is not drawn to scale.

3.5.2. Biolistic transformation

The strength of *5Spr20* promoter was evaluated by transient GUS expression in biolistically transformed pine embryogenic tissue. Construct p5A was used to appraise the promoter's potential in gene downregulation and the expression vector p5S was used to study the promoter's ability to direct protein expression

in pine. pGUS, which contains *gus* under the control of duplicate CaMV 35S promoter was used as the positive control and non-bombarded tissue as negative control.

In 'sense' studies, pine embryogenic tissue was biolistically transformed with p5S and pGUS, separately, at equimolar concentrations and promoter strength was assessed on transient expression of GUS. In antisense studies, pine tissue was cotransformed with p5A and pGUS and the antisense effect was gauged from the reduction in *gus* expression in cotransformed tissue compared to the positive control. The four treatments were (1) pGUS (+ve control) (2) p5A +pGUS (3) p5S (4) non-bombarded (–ve control).

3.5.2.1. GUS expression analysis

Protein was extracted from the various treatments and protein content of extracts determined by the Bradford assay [1976] ranged from 0.5 to 1 $\mu g\ \mu l^{-1}$ (appendix I). β-glucouronidase (GUS) activity in the extracts was assessed by MUG assay (chapter II, 14.1), which is based on a linear rate of substrate hydrolysis as a function of time. The linearity of reaction over time was confirmed in preliminary time course assays using dilutions of a commercial GUS preparation containing 10^{-2} and 10^{-4} units of enzyme (Appendix I).

Enzyme assays were performed with extracts containing 25 μg protein and fluorescence was recorded at 0′, 15′ and 30′ (chapter II, 14.1). Readings taken at the last two time points were within the linear scale and were averaged to give the GUS concentration from which the background noise recorded at 0′ was subtracted. However, the assay readings were low and therefore the experiment was repeated using 50 μg protein in 400μl final volume and fluorescence was read only at 0′ and 30′ (Appendix I).

Table-3.2 shows the mean GUS concentration computed for each treatment. A low level of background activity was detected in non-bombarded tissue which may be due to intrinsic fluorescence [Gartland *et al*, 1995]. The readings were

corrected for background noise and expressed as a percentage of the GUS concentration in pGUS (positive control) bombarded tissue (in column 3 of Table 3.2).

Table-3.2 GUS activity in tissue extracts

Treatment	Mean	as % of pGUS[a]
pGUS	1075.4 (±87.2)	100
p5A/pGUS	125.2 (±24.5)	10.1
p5S	107.6 (±20.4)	8.5
Control	17.9 (±4.05)	

Table-3.2. Shows treatments and their respective mean GUS activity expressed in pmoles MU min^{-1} mg^{-1} protein. a = as % of pGUS activity was calculated after subtracting mean background activity (in control) from mean activity in treatments.

The transient GUS expression values show that the *5Spr20* promoter-antisense *gus* construct was very effective in suppressing GUS expression; GUS expression of cotransformed tissue was only ~10% of that in tissues transformed with pGUS only. However, the *5Spr20* promoter is inefficient in driving sense expression of GUS; GUS expression in tissues transformed with p5S was only 9% of that in CaMV 35S promoter (in pGUS) transformed tissue (Figure-3.29).

Bartlett's tests of raw data (Appendix I) showed that variances were not homoscedastic ($P<0.05$). Therefore, values were log-transformed (appendix I) to make variance independent of the mean and reduce heteroscedasticity [Sokal & Rohlf, 1995]. This data was used in one-way analysis of variance (ANOVA) to estimate variances between and within treatments (Table-3.3).

Table-3.3. ANOVA of log-transformed GUS activity

source of variation	sum of squares	df	mean square	F
Between treatments	13.9234	3	4.6411	74.99
Within treatments	1.7329	28	0.0619	
Total	15.6563	31		

Differences between treatments are highly significant ($p = 0.01$).

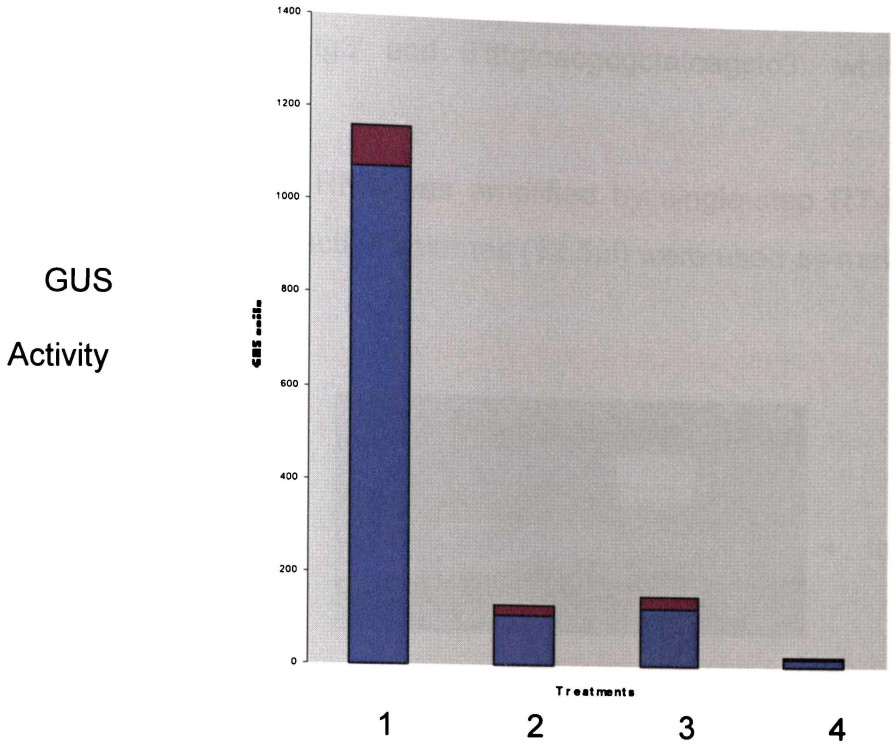


Figure-3.29. GUS activity of treatments. 1. pGUS (+ve control), 2. P5S, 3. P5A+pGUS 4. Non-bombarded control. GUS units expressed as pmoles MU min^{-1} mg^{-1} protein. Standard error of means shown as red blocks.

The Turkey test was applied to group means which differed significantly from one another (Table-3.4).

Table-3.4. Turkey comparison of treatment means

Treatment	Mean[a]	Homogeneous group
pGUS	3.0207	I
p5S	2.0157	I
p5A/pGUS	1.9721	I
Control	1.1605	I

a = means of log-transformed data. Treatment means that do not significantly differ are shown within the same homogenous group

3.5.2.2 *gus* transcript analysis

Total RNA from the different treatments was analysed to assess possible relationship in GUS expression at the translational (Figure-3.29) and transcriptional levels. The volume of transformed tissue was not sufficient to provide enough RNA for Northern hybridization and therefore, RT-PCR was used since the detection threshold of RT-PCR is at least 1000-fold lower than that of Northern hybridization which requires around 10^5-10^6 RNA molecules

[Hamoui *et al*, 1994]. Sense *gus* transcripts were analysed using primers 5'gcagatgaacatggcatggtg3' and 5'tttgtcacgcgctatcagctc3', which amplify a 3′ fragment of *gus*.

0.5 μg Total DNAsed RNA was amplified by single-step RT-PCR (chapter II, section 10.7.1); half reaction volumes (12.5μl) were used as material was limiting (Figure-3.30).

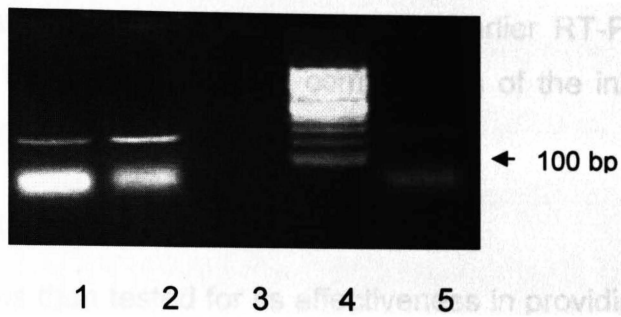


Figure-3.30. RT-PCR of transformed tissue. Lane 1. p5S, 2. pGUS, 3. Control pGUS without RT, 4. 100 bp ladder, 5. p5A+pGUS. 0.5 μg total RNA was amplified by single-step RT-PCR using 25 cycles of PCR and electrophoresed in 1.5% TAE gel.

Absolute comparisons cannot be made since PCR was not quantitative; a semi-quantitative analyses would have been possible had the reactions been co-amplified with primers to internal controls such as actin. However, based on intensity of fluorescence of PCR bands it appears that *gus* expression was weak in cotransformed treatment and stronger in pGUS and p5S treatments. The level of *gus* transcripts in pGUS and p5S treatments was more or less similar in most replicates (Appendix I).

Quantitative or semi-quantitative comparisons are valid only when amplification products are compared during the exponential phase of amplification. Therefore, RT-PCR was repeated with a subset of the replicates to see whether the previous RT-PCR was done in the exponential phase of amplification; since material was limiting, only one sample each of pGUS and p5S treatments which gave strongest bands in RT-PCR were chosen and PCR repeated with serial two-fold dilutions of their RT product. One μg total RNA from each sample was reverse transcribed in 25 μl reaction volume. Aliquots of 12.5, 6 and 3μl were made to 12.5 μl final volume in RT-PCR master mix and PCR performed as before. Figure-3.31 shows results obtained with pGUS treatment.

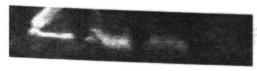


Figure-3.31 PCR of serial dilutions of pGUS RT product. 1. Undiluted, 2. 2x diluted, 3. 4x diluted, 4. –ve control. Smudges on the gel are due to undissolved EtBr. See text above for details.

There was a diminishing of products with dilution in both cases indicating that RT-PCR comparisons were made in the exponetial phase of PCR. Figure-3.31 refers to the trend observed with only the pGUS treatment, which gave a higher GUS transcript amplification than other treatments in earlier RT-PCR (Figure-3.30). Therefore, it can be assumed that all comparisons of the initial RT-PCR were done during the exponential phase of amplification.

3.6. Stable expression studies

The *5Spr20* promoter was then tested for its effectiveness in providing sustained gene downregulation in stably transformed plants. This was done by integrating a *gus* silencing cassette under the control of *5Spr20* promoter into the genome of a plant expressing GUS and monitoring GUS activity in stable (re-) transformants.

In view of time constraints on this study, stable expression studies were conducted in the model plant *Nicotiana benthamiana* instead of the preferred host pine since regeneration of transformants from pine (embryogenic) cells or tissues (such as cotyledons) is slow. For the same reason, an shRNA silencing cassette was used instead of the conventional ihpRNA cassette (chapter I, 6.2) and in preference to an antisense cassette as the latter gives inconsistent and low levels of silencing [Smith *et al*, 2001]. *Agrobacterium*-mediated transformation was used in preference to the biolistic method as the latter has a higher propensity to introduce multiple and fragmented transgene copies in transformants.

3.6.1. shDNA vector construction

The silencing shRNA cassette was amplified from *5Spr20* promoter contained in plasmid pPro20 by overlap PCR using two sequential amplifications [Gou & Liu, 2003] and cloned into the *SacI*/*XhoI* site of binary vector pGreen-0029-62 SK

which contains a kanamycin selection marker (section 2.5). Primary PCR was performed at an annealing temperature of 58°C using *5Spr20* forward primer and reverse primer P1 consisting of the reverse sequences of the loop and *gus* regions and a 17 bp overlap with the 3′ end of *5Spr20* promoter. An amplicon comprising the *5Spr20* promoter, 21 bp *gus* sense sequence and 9 bp loop sequence was generated and verified by sequencing (Figure-3.32).

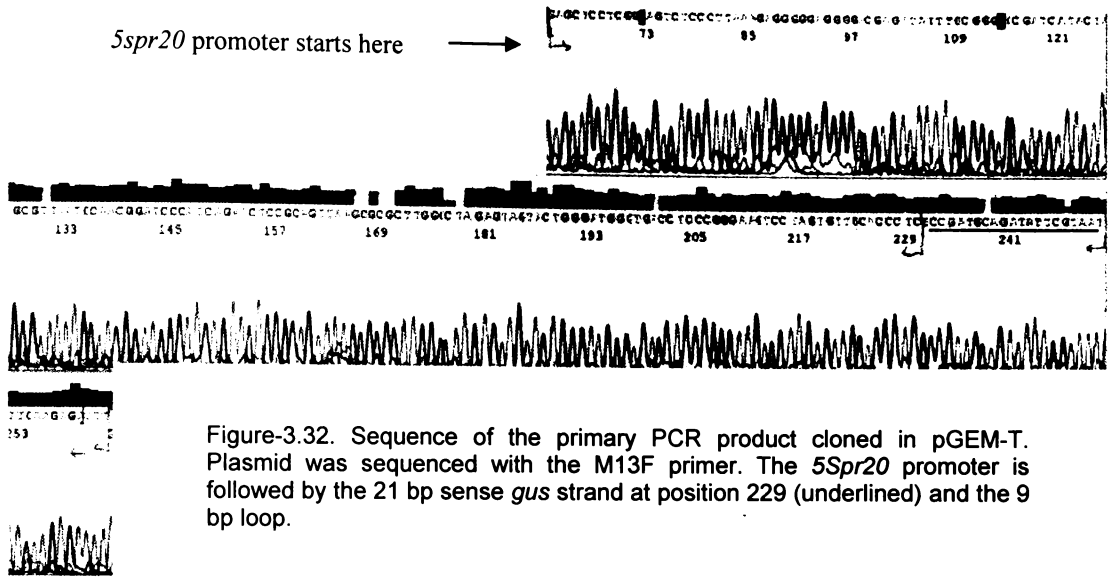


Figure-3.32. Sequence of the primary PCR product cloned in pGEM-T. Plasmid was sequenced with the M13F primer. The *5Spr20* promoter is followed by the 21 bp sense *gus* strand at position 229 (underlined) and the 9 bp loop.

The primary PCR product was used as the template in a second round of PCR using the same forward primer and a second reverse primer P2 containing the reverse sequence of the terminator, 21 bp *gus* sense sequence and a 14 bp overlap with the 3′ end of the primary PCR amplicon. This resulted in the final shRNA construct consisting of the *5Spr20* promoter, followed by sense *gus* strand, the 9 bp loop sequence, antisense *gus* strand and the poly-T terminator (Figure-3.33).

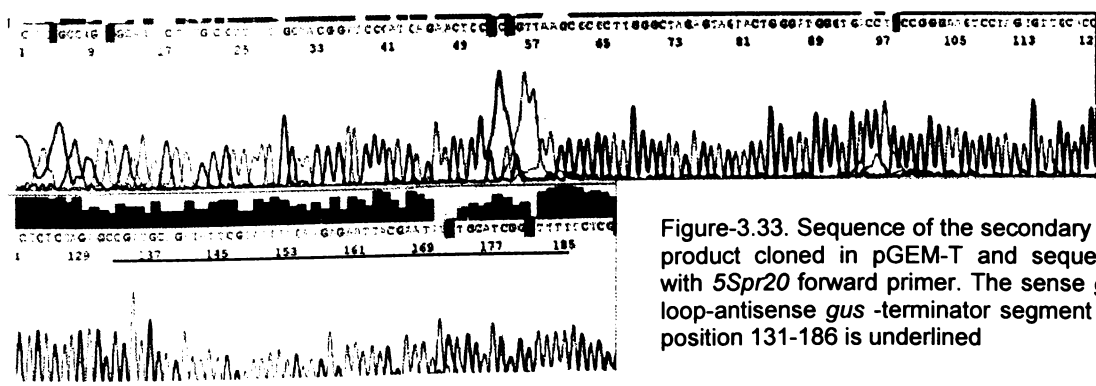


Figure-3.33. Sequence of the secondary PCR product cloned in pGEM-T and sequenced with *5Spr20* forward primer. The sense *gus* -loop-antisense *gus* -terminator segment from position 131-186 is underlined

The shRNA construct flanked by *Sac*I and *Xho*I ends was cloned in binary vector pGreen-0029-62 SK and transformed into *Agrobacterium*. Positive transformants growing on kanamycin selection medium were verified by PCR using forward and reverse primers of *5Spr20* promoter. The plasmid of a selected *Agrobacterium* clone, pSh was backtransformed into *E. coli* and plasmids of transformants were sequenced to confirm the sequence of shRNA construct in *Agrobacterium* pSh. This isolate was used in plant transformation.

3.6.2. GUS expressing stable transformants of *N.benthamiana*

GUS, which served well as a marker in transient expression studies was also used in stable expression studies; it ideally requires producing stable *gus* transformants, which are then stably re-transformed with a *5Spr20* promoter driven *gus* silencing cassette (cotransformation is another option).

Leaf discs of *N.benthamiana* were transformed with *Agrobacterium* harbouring the binary vector pGUS (section 2.4) which carries a 2x35S promoter driven-*gus* cassette and the kanamycin selection marker. Callus growing from inoculated leaf disks (Figure-3.34) produced three independent T0 plants (primary generation transformants) while all uninoculated control leaf disks died when grown on kanamycin selection medium (chapter II,13.3).

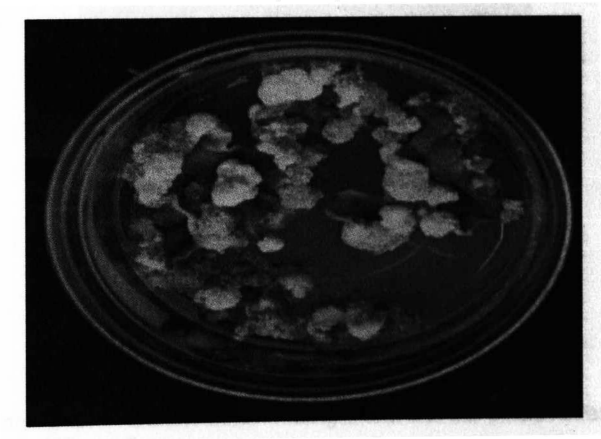


Figure-3.34. Callus development. Leaf disks were inoculated with pGUS and grown on MS-agar containing kanamycin selection marker.

Transformants were bioassayed for GUS expression at the three-leaf stage; leaf segments 1x0.5 cm were incubated in GUS assay buffer for 24 hours and examined for chromogenic reaction (chapter II, 14.2). Two GUS-expressing T0

plants were grown to maturity (in the glasshouse) and leaf samples were taken at monthly intervals for histochemical GUS assay. DNA was extracted from mature leaves and used in PCR with *gus*-specific primers 5′ gttacgtcctgtagaaacccca3′ and 5′ctccctgctgcggtttttcac3′. PCR and GUS assays both confirmed that the plant carried the *gus* gene(s) and expressed GUS (data not shown). GUS was expressed in all the leaves throughout the growth period.

Seeds of self-pollinated plants were grown on kanamycin selection medium and T1 plants which survived the selection pressure were again assayed for GUS expression at monthly intervals. Leaves obtained from a single GUS expressing T1 plant (which could be homozygous or heterozygous for the transgene) were used for retransformation with the shRNA construct.

3.6.3. Transformation of T1 GUS plants with *5Spr20*-shRNA construct

GUS expression was reconfirmed in leaves of T1 GUS-expressing *N.benthamiana* just prior to taking leaf discs for transformation with *Agrobacterium* pSh (section 3.6.1). 50 leaf discs were inoculated with pSH while five discs acted as the un-inoculated control. From callus cultured on kanamycin medium (chapter II, 13.3) 10 independent transformants were obtained from the pSH inoculated treatment, all of which grew from a single leaf disk (Figure-3.35).

Figure-3.35. shDNA transformants. Transformants arising from a GUS positive leaf disk transformed with pSH and grown on kanamycin-MS medium.

The uninoculated control (whose leaf disks are also kanamycin resistant by virtue of pGUS) produced two plants which arose from two leaf disks. Plants of the pSH treatment emerged 3-4 weeks later than the uninoculated treatment; apart from this, there were no apparent phenotypic differences between plants of the two treatments. The plants were separated and grown in individual pottles.

3.6.3.1. GUS assay

All 10 *5Spr20*-shRNA transformants tested negative for GUS when bioassayed at the three-leaf stage. The two plants of the 'untransformed' control treatment tested positive for GUS expression; any difference in the level of GUS expression between the two plants was not obvious when stained bioassay samples were examined visually.

Callus which grew healthily in the *5Spr20*-shRNA transformed treatment but which failed to produce shoots were subsequently assayed for GUS and found to be GUS positive. It is not clear why these active tissues did not produce new plants even if they were not transformed with *5Spr20*-shRNA. One explanation is that these foci were indeed transformed with *5Spr20*-shRNA, but failed to grow because of deleterious off-target effects [Snove *et al*, 2004].

The plants were bioassayed every three weeks; as the leaves were very small, a composite sample of leaf segments from all expanded leaves was taken and tested by the GUS histochemical assay. All plants transformed with the shRNA construct remained GUS negative, while the control plants were GUS positive (Figure-3.36).

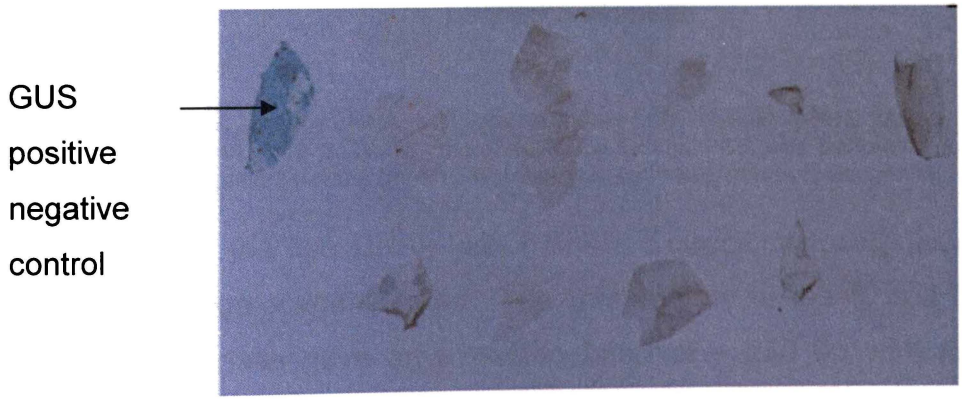


Figure-3.36. GUS histochemical assay. At top left is a leaf sample from an untransformed control plant which tested GUS positive (stains blue). Leaf samples from nine shRNA inoculated plants were GUS negative. Leaf segments were stained in GUS assay solution and destained in 70% EtOH.

Stem and petiole sections and whole leaves of all surviving plants, which were still small were assayed for GUS expression at the end of the study. GUS expression was not observed in plants transformed with the shRNA construct.

This suggests that *5Spr20*-shRNA is capable of sustained gene silencing of the target gene.

3.6.3.2. Molecular assay

pSH transformed plants were still small at the conclusion of this study and therefore molecular analysis could not be done on tissues of individual plants. A composite sample was therefore taken of leaves from eight surviving plants and DNA and RNA (section 2.6) were extracted. These were assayed by PCR for the presence of the genomic and transcript sequences of the target (*gus*) and the silencing construct (5Spr20-*shRNA*).

3.6.3.2.1. PCR

PCR amplification of genomic DNA extracted from plants of pSH transformed (pSH/pGUS) and control treatments (pGUS) with *gus*-specific primers (section 3.6.2) gave the expected ~ 120 bp amplicon (Figure-3.37), indicating that *gus* was present in plants of both treatments.

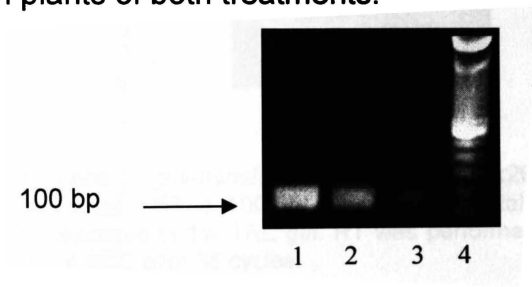


Figure-3.37. PCR of genomic DNA with *gus* primers. Lane 1. pSH/pGUS double transformant 2. uninoculated pGUS wild-type 3. Water control 4. 100bp ladder. PCR was performed for 35 cycles and reactions were electrophoresed in 1.5% TAE agarose gels and visualized by EtBr staining.

Amplification of the two DNAs with primers 5'gagctcctcggcagtctcccttaaga 3' and 5' aaaaaccgatgcagatattcgtaa 3' complementary to the 5' and 3' ends of the shRNA cassette gave an expected band of ~ 200 bp with the shDNA transformed DNA (Figure-3.38). This confirms the presence of *5Spr20*-shDNA in the genome of transformed plants.

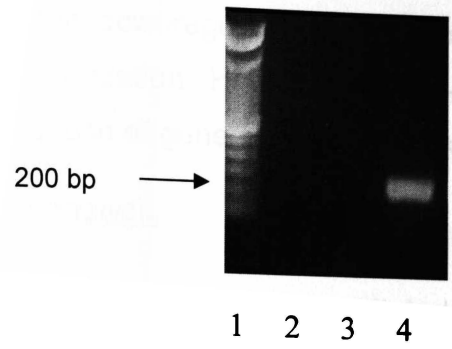


Figure-3.38. PCR of genomic DNA with shRNA primers. Lane 1. 100 bp ladder 2. Water control 3. *gus*-transformed control treatment, 4. *5Spr20*-shDNA transformed treatment. PCR was performed for 35 cycles and reactions were electrophoresed in 1.5% TAE agarose gels and visualized with EtBr staining.

3.6.3.2.2. RT-PCR

RT-PCR failed to produce visible amplification when one μg total RNA of shRNA transformed plants was amplified with primers 5′ gttacgtcctgtagaaacccca3′ and 5′ctccctgctgcggtttttcac3′ which target the 5′ and 3′ ends of *gus* (Figure-3.39).

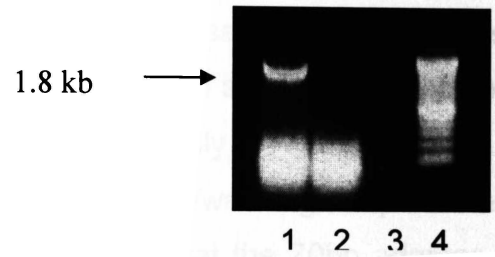


Figure-3.39. RT-PCR. Lane 1. *gus*-transformed control. 2. 5Spr20-shRNA transformed treatment, 3. Nontransformed control without RT, 4. 100 bp ladder. One μg total RNA was amplified by RT-PCR and reactions were electrophoresed in 1% TAE gel. RT was performed at 50°C and PCR was done at an annealing temperature of 55°C over 35 cycles.

But when RT-PCR was repeated with 10 μg total RNA from *5Spr20*-shRNA transformed plants a weak, but still visible amplicon of the expected size (~1.8 was observed (Figure-3.40).

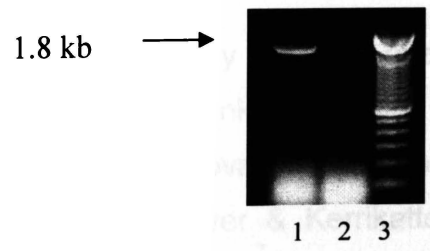


Figure-3.40. RT-PCR of total RNA with *gus* primers. Lane 1. 0.3 μg total RNA from control *gus* transformants, 2. 10 μg total RNA from shRNA transformed *gus* plants, 3. 100 bp ladder. RT was performed at 50°C and PCR was done at an annealing temperature of 55°C over 35 cycles.

This demonstrates that full-length transcripts were produced in both plants. But there is a massive reduction in the level of intact *gus* transcripts in the

silenced plants; this downregulation is probably due to siRNA-directed mRNA cleavage and degradation. However, siRNA-directed chromatin remodelling also can be the cause of gene silencing [Verdel *et al*, 2004].

3.6.3.2.3. siRNA analysis

The absolute proof for siRNA-mediated mRNA degradation is the accumulation of target-specific siRNA in silenced plants is [Zhou *et al*, 2004]. Therefore, snRNA analysis was done as per method of Llave *et al* [section 2.6]. Five μg small RNA was fractionated in 15% denaturing PAGE gels and electroblotted to a nylon membrane at 120V for one hour in TBE buffer. An adaptor consisting of the sense *gus*-loop-antisense *gus* sequence was random labelled with ^{32}P and used to probe the blot at 45°C. The blot was washed at 37°C in 1xSSC and autoradiographed. No signals were observed.

Further optimising of hybridisation conditions was not possible within the study period. The small quantity of snRNA used may be one reason for the failure to detect *gus*-siRNA; nominally 20-40 μg snRNA are used in siRNA assays. Secondly, the hybridisation/washing temperatures may not have been optimal. A third reason could be that the 70bp adaptor used as template for random labelling consists of two inverted repeats of the *gus* siRNA sequence. It is possible that the probe folded on itself during hybridisation thereby reducing probe concentration.

4. DISCUSSION

4.1. 5S rDNA organization and implications for PCR amplification of *5Spr20*

Cloning of most genes is fairly complex even when a partial sequence of the gene exists. Walking into the unknown region of a gene generally involves library construction or several innovative, but labour-intensive methods such as amplifying a circularised [Silver & Kerrikatte, 1989], or adaptor ligated gene segment [Seibert *et al*, 1995], etc. In contrast to this situation, the cloning of *5Spr20* was achieved by a simple PCR amplification strategy (section 3.2.2).

In most eukaryotic genomes with the exception of *Neurospora crassa*, 5S rDNA genes occur in several arrays as tandem, head-to-tail repeats; the genes in an

array are considered to be homogeneous [Moran *et al*, 1992]. This knowledge of the long-range structure of 5S rDNA genes was exploited to amplify *5S rDNA* from genomic DNA by PCR. The forward and reverse primers obtained from the partial sequence (section 3.2.4) were designed to face away from each other (Figure-3.41).

← Reverse primer Forward primer→

$^{-141}$caccaattcttccgctgg^{-123}g^{-122}ggcaccaccaagtcgtgg^{-105}

Figure-3.41. Figure shows reverse and forward primers directed away from each other to read into the unwalked gene regions.

These primers will not amplify the gene directly from genomic DNA if it is a single-copy or dispersed gene. But because 5S rDNA occurs as contiguous repeats, the primers will anneal to complementary sites at identical positions in each repeat during PCR (Figure-3.42).

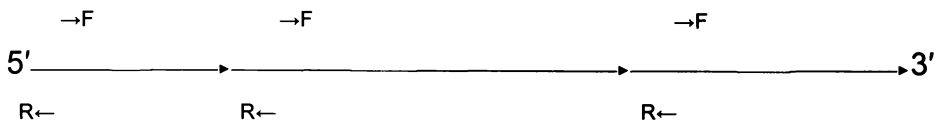


Figure 3.42. Shows three contiguous gene repeats. The small arrows are the forward (F) and reverse (R) primers, respectively, which anneal at identical locations in the three repeats during PCR. Red lines indicate the coding region.

The forward primer reads through the unknown 3′ end of the transcribed region of one repeat unit into the nontranscribed region of the adjacent repeat, while the reverse primer reads through the 5′ region of the nontranscribed region into the coding region of the preceding repeat unit. Thus, the entire sequence of a 5S rDNA unit can be amplified.

The walking primers were selected from the spacer region and not from the transcribed region because the latter is highly conserved among 5S rRNA genes [Appels *et al*, 1980; Ellis *et al*, 1980; Gerlach & Dyer, 1980; Schneeberger *et al*, 1989; Gorman *et al*, 1992]. If the primers were designed from the transcribed region, it would result in the coamplification of almost all 5S rRNA pine paralogs.

The forward and reverse primers were therefore obtained from the region between positions -141 and -105 of the *5Spr20* spacer region which shows only 2-35% sequence identity to published pine 5S rDNA paralogs. Despite selecting primers from an apparently unique, gene-specific region, a mixture of

transformants displaying sequence micro-heterogeneity was obtained. This suggests that even within a pine 5S rDNA cluster there may be genes which differ by minor point mutations; Cloix *et al* [2003] found a 1% divergence among 5S rDNA genes of the transcribed array on chromosome 5 locus of *A. thaliana*. Another explanation is that the primers could have annealed to partially homologous regions in yet undiscovered 5S rDNA arrays at other loci.

4.2. Gene characterization and putative promoter elements

The analysis of *5Spr20* gene structure revealed several regions of significance to promoter construction.

4.2.1 Conserved internal promoters

The overall homology between coding sequences of *5Spr20* and orthologs in several higher plants is 92-99%. Similar observations were made by Liu *et al* [2003] in a study of five Asian *Pinus* species. Their comparison of 120 independent *Pinus* 5S rDNA coding sequences showed that the genes had 97% sequence identity among *Pinus* species, while similarity between *Pinus* and other gymnosperm species was 94% and between gymnosperms and angiosperms ~88%. Because of the high degree of conservation in 5S rDNA coding sequences of many eukaryotes, there is also great consistency in the nature and requirements for the internal promoters in species as disparate as *Arabidopsis* [Cloix *et al*, 2003] and *Xenopus* [Majowski *et al*, 1987]. This could well be the case in pine and gymnosperms too where little is known of the functional elements required for transcription.

A comparison of the putative tripartite internal control region [Hori *et al,* 1984] of *5Spr20* and other 5S rDNA genes (Figure-3.8) shows that the gene-specific A- (+50 to +64), C- (+80 to +90) and I- (+67 to +72) promoter boxes [Pieler *et al,* 1985, 1987] are highly conserved in *5Spr20* [Erdman *et al*, 1985]. The sequence of the internal promoters and the distance between them were major considerations in selecting *5Spr20* from among several transformants (section 3.2.2) for further study. Deletion and mutation studies in several eukaryotes have demonstrated that basal transcription of 5S genes is a function of conserved split promoters and the distance between them. Transition mutations of A-box bases

51G, 52T, 53T, 54A, 56G, 58G and 61C in *Arabidopsis* reduce transcription efficiency by 40-65% and C-box bases 82T and 84T by 40-90% [Cloix *et al*, 2003]. Similarly, mutation analysis in *Xenopus* show that the guanosine residue at position +64 in the A-box and C and G residues at positions +67 and +70 of the I-box affect TFIIIA binding in a major way; TFIIIA is the gene-specific transcription factor that initiates assembly of the transcription preinitiation complex [Theunissen *et al*, 1998]. Thus *5Spr20* rDNA contains the appropriate internal promoters required for basal transcription.

4.2.2. Spacer region

The 5′ flanking spacer region plays a major role in activated transcription; for example, an *Arabidopsis* 5S gene construct consisting of the coding and downstream regions only showed low transcriptional activity [Cloix *et al*, 2003]. The spacer regions of pine paralogs show several conserved sequences resembling putative control elements in orthologs and repeat elements (section 3.3.3.3) analogous to those in other systems. The sequence differences observed among the paralogs (section 3.8.3.4) may also be implicated in gene expression.

4.2.2.1. Conserved motifs/regions

All pine paralogs including *5Spr20* possess a cytosine base preceding the transcription start site. The universality of cytosine one base before the start site (Figure-3.9) has led to the suggestion that it may play a role in transcription initiation, similar to that of the -1 nucleotide in yeast. Challice & Segal [1989] using deletion mutation demonstrated that mutation of the base preceding the transcription start site in yeast 5S rRNA genes reduced transcription by 15% compared to the wild type gene. In *Arabidopsis,* C at –1 is important for both efficient transcription and correct transcription initiation; a transition mutation, while maintaining the pyrimidine-purine dinucleotide (at-1 and +1) required for transcription [Fruscoloni *et al*, 1995], results in 30% decreased transcription and shorter transcripts initiating from position +30 [Cloix *et al*, 2003].

A guanine/cytosine-rich region at about -13, which is less conserved but clearly different from the composition of adjacent sequences is observed in most plants

[Venkateswarlu *et al*, 1991]. A GC dinucleotide occurs at positions –12 to –11 in *Arabidopsis*, which when mutated to AA reduces transcription efficiency by 50% [Cloix *et al*, 2003]. An analogous region consisting entirely of G residues and centred around -15 occurs in *5Spr20* and other pine paralogs (Figure-3.9).

A characteristic upstream motif associated with the majority of genes transcribed by Pol II is the TATA box of varying consensus sequence [Arkhipova, 1995; Struhl, 1995; Washburn *et al*, 1997], which is the cognate binding site for the TATA-binding protein, TBP. Similar TATA motifs are also found in several Pol III transcribed genes where they influence transcription. An upstream TATA box of consensus sequence (T/C)ATA(G/A) is found in all one hundred 5S rRNA genes of *Neurospora crassa*; the motif which is located at -29 to -24 is necessary for initiation of transcription from the correct start site, and its deletion results in reduced transcription [Tyler, 1987]. Similarly, 5S rDNA of the silkworm *Bombyx mori* contains a homologous TATA motif 28 to 24 nucleotides upstream of the transcription start site [Morton & Sprague, 1984]. The TATA motif of *Drosophila* 5S rDNA is positioned at -26 to -39 [Sharp & Garcia, 1988].

Among plant 5S rDNA, a TATATA motif occurs at –28 to –23 in *A.thaliana*; its deletion reduces transcription to 10% and mutation to a GAGAGA motif abolishes transcription altogether [Cloix *et* al, 2003]. The brassica species *Mathiola incana* possesses the sequence ATATAT at -29 to -24. A similiar conserved region is found at an identical location in *Vigna radiata.* TATA-like sequences are also observed at similiar locations in at least four other legumes - yellow lupin [Rafalski *et al*, 1982], pea [Ellis *et al,* 1988], mung bean [Hemleben & Werts, 1988] and soybean [Gottlob McHugh *et al,* 1990]. Neither a TATA box, nor TATA-like sequences are present in the upstream or downstream spacer regions of *5Spr20* or its paralogs. Neither, do they possess an adenine/thymine-rich element characteristic of most plant 5S rDNA such as the conserved ATAAG sequence centered around -27 in members of *Triticeae* [Scoles *et al*, 1988], or AAATAAA in flax [Goldsborough *et al,* 1982], or TTTAATA in *Lycopersicum esculentum* [Venkateswaralu *et al*, 1991]. However, a TTGAA sequence occurs at about the expected position (-30 to -26) in *5Spr20*; it differs from the immediately flanking sequences in its A/T composition and is fully conserved in

all pine paralogs and in the gymnosperm Douglas Fir (Figure-3.9). Possibly, it is the pine, or the gymnosperm equivalent of the angiosperm A/T rich element. Such regions of low duplex stability have implications for DNA unwinding during transcription [Lewin, 1994] and are probably related in function to the -10 hexamer box that is recognized in bacterial promoters [Venkateswaralu *et al*, 1991].

The region immediately upstream of the coding sequence (-1 to -47) is almost identical in all isolates (Figure-3.15). This is probably no coincidence since it corresponds to the region bound by TFIIIB. During the formation of a binary preinitiation complex which precedes transcription initiation, initiation factor TFIIIB binds to the 5′ region immediately upstream of the start point and protects a region of 45-50 bp on both strands of the DNA [Kassevetis *et al*, 1989]. This highly conserved (~95%) region which provides a uniform template for TFIIIB binding may be an important *cis*-element. A parallel exists in *A.thaliana* where substitution of all but the proximal 31bp of the 5′ flanking spacer does not affect transcriptional activity; but sequence alterations inside this region decrease transcription [Cloix, 2003].

Several upstream regulatory elements such as the pentamer motif AAAGT, and the GC motif CGGGGCGGGGC have been identified in Pol III genes [Sorenson & Frederiksen, 1991]. Although known *cis*-regulatory motifs are not observed in the highly conserved upstream regions of pine 5S rDNA spacers (corresponding to positions -1 to -47, -170 to -233 and -250 to -271 of *5Spr20* in Figure-3.15), their conserved status suggests that they have implications for possible binding sites to putative activators or repressors. The occurrence of similar highly conserved sequences within the spacer regions of Douglas fir and white spruce, taken together with the absence of upstream sequences conserved in angiosperms has led to speculations that the regulatory mechanism for 5S rDNA transcription in gymnosperms may be different [Amerasinghe & Carlson, 1998].

4.2.2.2. Repeat elements

Repeat sequence elements in the intergenic spacers of eukaryotic rDNA have been implicated in the enhancement of Pol I activity [Pikaard, 1993; Robinett *et*

al, 1997; Burton *et al*, 2005]. Similarly in the early promoter of Pol III transcribed gene SV40, the six GC-rich repeats present in the 21bp repeat region are essential elements of the early-early and late-early promoters [Baty *et al*, 1984]. Thus the presence of the repeat element GGCTCGGCGG in the spacer region of *5Spr20* (section 3.3.3.3) raises a similar possibility for *5Spr20* transcription efficiency. Alternatively, the presence of this 10 bp element may be due to tandem duplications coupled with unequal crossing over.

The significance of the duplication of the sequence TTGCGGTGCGATC surrounding the start site (Figure-3.11) is not clear, but there are strong parallels between this and spacer promoters of the major ribosomal RNA genes of several species. In *Drosophila,* five to 12 copies of a 70 bp element are observed in the intergenic spacer (IGS); the repeats contain a perfect copy of the -24 to +10 domain of the core promoter. Spacer initiated transcripts arising from these repeats are thought to stimulate transcription at the core promoter [Jacob, 1995]. In *Xenopus*, the region -147 to + 4 is repeated within the IGS as spacer promoters. These repeats, along with repeat enhancer elements present in the IGS are believed to regulate transcription at the core promoter [Busby & Reeder, 1983].

4.2.2.3. Sequence differences among paralogs

Sequence heterogeneity between *5Spr20* and the smaller 5S rDNA paralogs (*5Spr5, 6, 7 & 9*) are attributable to three factors. A ~360 bp insert at position -675 to -315 in *5Spr20* is the major reason for the size difference. Secondly, the region containing the dinucleotide simple sequence repeat $CT_{(9\text{-}16)}$ in the small 5S rDNA repeats [Smith & Devey, 1995], often suggested as the reason for polymerase slippage and the ensuing sequence heterogeneity, is absent in *5Spr20*. Point mutations account for the remaining sequence differences. Such major differences in spacer sequences of the short and long paralogs is reflected in the size classes being placed in different phylogenetic clades (Figure-3.14). Sequence differences between *5Spr20* and *5Spr10* which are of nearly equal length arise mainly from point mutations caused by insertions, deletions and transversion and transition events.

4.2.2.4. Gene copy analysis

4.2.2.4.1. Complex Souther hybridization patterns

Both *Bam*HI and *Eco*RI have a unique cleavage site within the *5Spr20* gene, *Bam*HI in the coding sequence and *Eco*RI in the spacer region. Therefore, each tandem gene copy can be expected to be cut at these sites to produce a single strong band in Southern blots from which the copy status of *5Spr20* can be estimated by gene titration. Instead, the two digests produced a ladder of multimers, where the monomer band was the least distinct in *Bam*HI and most dominant band in *Eco*RI lane (Figure-3.16).

Complex hybridization patterns unexpected of a simple repeating structure such as *5S* rDNA have been reported by several authors [Mascia *et al,* 1981; Goldsborough *et al*, 1982; Scoles *et al*, 1988; Lapitan *et al*, 1991]. These observations are explained on the basis of variable methylation, incomplete digestion and sequence heterogeneity arising from point mutations within the restriction sites. Using isoschizomers *Mbo*I and *Sau*3AI which share both recognition and cleavage specificity in the sequence GATC embedded in the restriction site of *Bam*HI but differ in their sensitivity to methylation at C, Gottlob-McHugh *et al* [1990] demonstrated extensive methylation of *Bam*HI sites (GGATCC) in soybean 5S rRNA genes; $C^{m4}C$ (5-methylcytosine-*dcm*) or $C^{hm5}C$ (5-hydroxymethylcytosine) methylation makes these sites insensitive to *Bam*HI; double digestion with *Sau*3AI and *Msp*I showed that differential methylation of overlapping *Msp*I and *Bam*HI recognition and not sequence heterogeneity was the cause of ladder-like hybridization patterns.

Methylation may have also affected the hybridization pattern of pine DNA digested with *Eco*RI, which is sensitive to ^{m5}C methylation. Plant DNA is highly methylated (>80%) at cytosine residues [Gruenbaum *et al*, 1981] and the level of methylation of 5S rDNA is higher than the average for the entire genome in angiosperms [Fulnecek *et al*, 2002]; methylation of 5S rDNA motifs mC_PC_p and mC_PG_p are estimated to be ~53-90%. The cytosine base in *Eco*RI site at position -266 of *5Spr20* (Figure-3.6) is flanked by a guanidine base (gaattcg); it is possible that some of the sites were methylated.

Densitometer tracings of the *Eco*RI and *Bam*HI lanes (Figure-3.16) showed that only 74% and ≤ 5%, respectively, of *5Spr20* DNA is found in the monomer form. This suggests that possibly *Bam*HI sites are extensively methylated and a significant number of *Eco*RI sites are either methylated or ablated by point mutations. The minor, but prominent band just above the monomer band in the *Eco*RI lane whose size is not a perfect multiple of the basic repeat size, may have arisen as a result of ablation of an existing *Eco*RI site combined with the creation of a new site by point mutation in some repeats; or because some repeats occur in tandem with junction sequences, or heterogeneous repeats [Moran *et al*, 1992]. The latter, however, is of rare occurrence [Flavell, 1986] since the force of concerted evolution acts to homogenize tandem repeats at a locus quickly. Secondly, if *5Spr20* arrays contained heterogeneous repeat classes differing in their length, it would have become apparent during PCR isolation of *5Spr20* (section 3.2.2) by the appearance of more than one amplicon. It also appears that methylation is widespread in most, if not all, pine 5S rDNA paralogs. This is seen when *Bam*HI digested DNA is probed with the coding sequence which is highly conserved in most paralogs; it produced a ladder of bands much more intense than bands obtained with a probe specific to *5Spr20* (Figure-3.17). The increased band intensity is likely due to detection of bands accumulating from 5S rDNA paralogs by the 5S rDNA-common probe.

4.2.2.4.2. *5Spr20* gene copy number

The minimum *5S rDNA* content of higher plants have been variously reported to be between 0.05 and 0.06% of the genome [Lapitan *et al*, 1991; Moran *et al*,1992], with a higher estimate of 0.7% for *A. thaliana* [Pruit and Meyerowitz, 1986]. Using a conservative estimate of 0.05%, the nuclear genome of pine (2C=22.2 pg) is expected to contain ~12 fg (12×10^{-15} g) of *5S rDNA*. If the average size of a pine *5S rDNA* repeat unit (the average of long and short units) is assumed to be 680 bp (section 3.3.1), then a single copy of diploid pine genome may contain a minimum of,

$(12 \times 10^{-15}\text{ g})\ (6 \times 10^{23})/680\text{ bp} \times 660\text{ da}$ or ~17,000 copies of *5S rDNA*.

Based on this calculation, 9-10^6 copies of the *5Spr20*-specific *Taq*I fragment were used as standards in gene reconstruction assays (section 2.3), which gave a value of 22 copies of *5Spr20* per haploid genome of pine (section 3.4.2). Using a similiar technique, Moran *et al* [1992] estimated that there are at least 300 copies of the long- and 7700 copies of the short-5S rDNA genes per haploid genome of *P.radiata*.

Thus, *5Spr20* does not appear to be highly duplicated in the pine genome. It constitutes only a mere 0.1% of the total pine 5S rDNA genes. But does it really matter if this candidate gene selected for transformation studies is under-represented in the pine 5S rDNA gene pool? Not necessarily as findings of Cloix *et al* [2001] and Mathieu *et al* [2003] seem to suggest; regardless of *5Spr20*'s contribution to the native 5S rRNA pool, its activity in transgenic pine will be determined mainly by the possession of the essential features of an active gene and its insertion into a favourable chromatin context. Cloix *et al* [2001] used a combination of YAC sequencing, RNA/DNA alignments, *in vitro* transcription and polysome RNA analyses to define the transcribed and non-transcribed 5S rDNA loci in *A.thaliana*. 5S gene arrays are found at 6 loci in chromosomes 3, 4 and 5 of which only genes on chromosome 4 and at the major locus of chromosome 5 produce mature transcripts (made up of 86% of a major transcript and 14% of minor transcripts), while the others are silent or produce rapidly degraded RNA. Alignment studies showed that many non-functional rDNA units do not have an RNA equivalent. The *in vitro* transcription studies showed that rDNA exhibiting mutations in important promoter elements failed to produce transcripts. Subsequent work by Mathieu *et al* [2003] suggests that within this mature rRNA pool, the minor RNA species are expressed more abundantly when the heterochromatin in which their genes lie is presented as loops that emanate from chromocenters and become repressed with heterochromatin establishment accompanied by DNA methylation and histone modification.

4.3. Promoter studies

4.3.1. Need for transformation studies

The usefulness of *5Spr20* promoter in pine transformation was studied by transient expression analysis, despite the fact that it is an endogenous gene.

Usually, the expression level of an endogenous gene is easily determined by assaying various plant tissues for the gene's transcripts using Northern analysis, RT-PCR or *in situ* hybridization; gene-specific sequences for use as probes or PCR primers in such assays can be obtained from the coding, or untranslated regions of Class II genes, even in case of highly conserved members of multigene families transcribed by Pol II.

In Class III genes such as 5S rDNA or tDNA, the main difference among members of a multi-copy family almost always resides in their nontranscribed regions. Therefore, very often, it is not possible to tag the transcripts of different paralogs by applying the same assay techniques used with Class II genes for two reasons. In 5S rRNA genes, the primary transcript and the mature RNA coterminate at the 5′ end of the coding region; the pre-RNA may extend 4 - 15 bases beyond the 3′ end of the mature RNA into the poly-T transcription stop signal [Weinmann & Roeder, 1974; Perry, 1976], but this region which is also highly conserved among paralogs is trimed back to the 3′ end of the coding region. Thus a gene-specific 5′ or 3′ transcribed region similiar to the UTR of Class II genes does not exist in 5S rRNA genes. Secondly, primary transcripts may undergo posttranscriptional editing to remove base differences within the coding region [Barciszewska *et al,* 1994; Szymanski *et al*, 1995]. Thus, the transcribed region of 5S rDNA genes does not bear sufficient sequence heterogeneity to discriminate transcripts of different paralogs.

Therefore, the expression profile of *5Spr20* promoter in pine cannot be evaluated by assaying its gene product. The alternative is to assay expression of a reporter placed under the control of *5Spr20* in transformed pine.

4.3.2. Promoter selection

Activated transcription in most Class III genes often requires one or more *cis*-acting elements, in addition to the minimal promoter responsible for the assembly of the preinitiation complex (PIC) [Huang & Maraia, 2001]. The minimal promoter for 5S rDNA resides within the coding sequence in *Xenopus* [Pieler *et al,* 1987]. In addition, a TATA, or TATA-like sequence essential for

accurate and activated transcription is located upstream of 5S rRNA genes in several species such as *A.thaliana, Bombyx mori* and *N. crassa* [Cloix *et al*, 2003]. Neither these, nor other well characterized *cis*-acting elements of Class I, II and III genes are found in the spacer region of *5Spr20*. Perhaps, activated transcription of 5S rRNA genes of pine and other gymnosperms depends on novel nucleotide motifs which are unique to these species. The presence of several conserved sequences and repeat elements within the spacer region of pine 5S rDNA paralogs and gymnosperm orthologs strongly argues for some, if not all, of these sequences to be activator elements.

One approach to define the promoter requirements for basal and activated transcription of *5Spr20* is through deletion analysis, where expression cassettes containing serially deleted upstream spacer regions linked to the *5Spr20* coding region – the 'minimal promoter' (which is analogous to the 46 bp minimal CaMV 35S promoter) are studied for their expression pattern. This helps to identify the *cis*-elements required for tissue-specific or global expression.

Midway through the studies, this research was transferred to a new location which lacked tissue culture facilities. Because of this and other constraints, a decision was taken to test only one promoter construct in transformation studies. Since a precedence does not exist for promoter dissection work in gymnosperms, several permutations and combinations are possible for the nature of *5Spr20* promoter to be tested in transformation studies. The entire native 5′ flanking region can be tested; or synthetic oligos representing the three highly conserved (95% identity) regions, namely the 64bp (-233 to –170) 22bp (-271 to –250) and 47bp (-47 to –1) elements in the upstream spacer can be fused into a synthetic promoter; or as was done in this study, a promoter comprising the entire coding region and 63 bases of the immediately upstream spacer region which is highly conserved (~90% identity) among pine paralogs can be used. There is good reason to believe that this region immediately flanking the 5′ end of the coding sequence may be an important *cis*-element; the proximal 47 bases show 98% sequence identity to those of its paralogs and may serve as

constant binding site for transcription factor TFIIIB, which is known to protect the region ~45-50 bp upstream of the transcription start site [Kassevetis *et al*, 1989].

4.3.3. Transient expression studies

4.3.3.1. Promoter cloning

Directional cloning of *5Spr20* promoter in pST and pAT (both pSK-based vectors) proved to be difficult and in spite of investigating several avenues, cloning of the promoter remained elusive; no transformants were obtained when ligated products were transformed into competent cells (section 3.5.1.2.3). On the other hand, (non-directional) cloning of the *5Spr20* PCR fragment in pGEM-T vector during initial gene isolation was accomplished with ease (section 3.2.2). This led to the speculation that a pause in replication fork movement was the likely reason for absence of viable transformants.

A pause in polymerization (replication) and the consequent termination of DNA polymerase activity [Samadashwily *et al*, 1997] may explain the absence of transformants in pSK-based vectors carrying the *5Spr20* promoter. DNA replication can terminate near specific sites charcterized by stable, or unusual secondary structures such as hairpins [Bedinger *et al,* 1989; Baran *et al*, 1991; Hacker & Alberts, 1994; Krasilnikov *et al*, 1997] where replication forks pause in plasmid ColE1 [Tomizawa, 1978] (whose origin of replication is included in pSK & pKS) and *E. coli* [Kuempel & Duerr, 1979]. In the model developed for polymerase release following termination of each Okazaki fragment, Hacker and Alberts [1994] demonstrated that a holoenzyme normally capable of associating *in vivo* with the leading strand for >5′, dissociates rapidly (half-life of 1″) when it is stopped by a hairpin helix.

During semidiscontinuous replication of bacterial DNA, while Pol III holoenzyme moves continuously along the template for the leading strand, the template for lagging strand is exposed as a single strand behind the primosome. Should adjacent mirror regions occur within this region it can be speculated that they have a chance to form stem structures depending on whether they can pair

before they are sequestered by single strand binding proteins (SSB); Samadashwily *et al* [1997] proposed a similiar paradigm for replication termination by triplex formation.

Analysis of *5Spr20* coding region using DNA Strider® shows that it contains mirror regions, 3-5 bases long and capable of forming stem structures with a free energy of ≥ - 6 kcal mol^{-1} in close proximity to one another in at least two locations. It is possible that when the polymerase hits these hairpin helices it senses encounter with duplex DNA and dissociates from the template.

Replication is unidirectional in pSK and proceeds from the *Xho*I cloning site of *5Spr20* to the *Kpn*I site *ie*. from 3′ to 5′ end of the cloned *5Spr20* promoter (Figure-3.43).

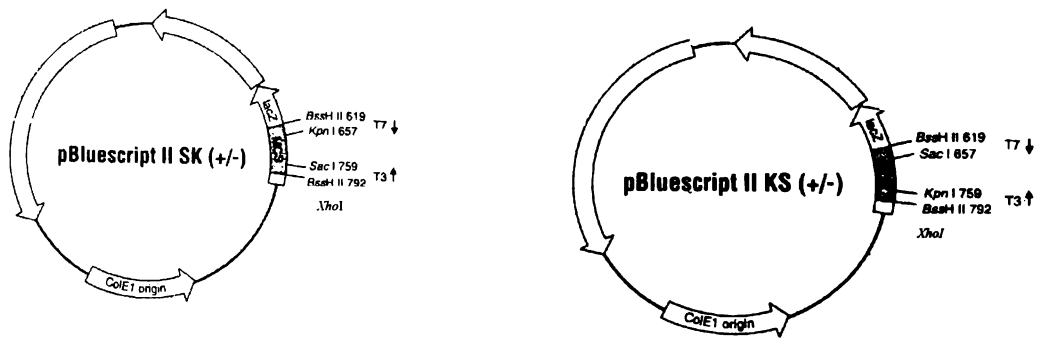


Figure-3.43. pSK. Arrow indicates direction of replication from *Xho*I to *Kpn*I site in pSK and and reversed in pKS in the direction *Xho*I to *Sac*I.

Reversing the direction of replication fork movement may aid unimpeded polymerase movement from the 5′ to 3′ end of *5Spr20* insert (from *Kpn*I to *Xho*I); Samadashwily *et al* [1997] found that stalling of polymerase within DNAs containing (CGG)n.(CCG)n and (CTG)n.(CAG)n repeats can be relieved by altering the repeat orientation relative to the replication origin. Thus, by reversing the orientation of the replication origin relative to *5Spr20* there is a chance that the sequential order in which mirror repeats occur in the antiparallel lagging strand template, or their duplex strength may offer less resistance to the progress of the fork. On this assumption, the expression vectors were recloned in pKS where the direction of replication is reversed with respect to the sequence of restriction sites in the MCS (Figure-3.43).

4.3.3.2. Antisense studies

For initial antisense assays using *gus* as the target gene, the antisense strand can be based on any region of the *gus* coding sequence This is because there are no clear-cut rules on the choice of the optimal gene region for maximum antisense effect. The 3′ transcript regions were found to be more effective by some authors [Van der Krol, 1990a; Bourke & Folk, 1992], the 5′ region by others [Melton, 1985] and the complete CDS by still others [Halpin *et al,* 1994; Keller *et al,* 1999]. In addition to sequence identity, factors such as secondary structure and access to the target region play a role in determining the outcome of antisense experiments [Dykxhoorn *et al*, 2003].

The *gus* antisense strand was obtained from the 5′ region of *gus* CDS primarily because it has convenient restriction sites which facilitate direct cloning into pT in an antisense orientation. The *gus* coding sequence contains an *Eco*RV site at position +556 and is flanked by a *Bam*HI site at the 5′ end of pGUS (Figure-3.1). The orientation of these sites is reversed in the cloning vector pKS. When cloned into the *Bam*HI-*Eco*RV site of pT (derived from pKS) the 5′ *Bam*HI - *Eco*RV *gus* fragment is effectively placed in an antisense orientation with respect to the terminator fragment at the downstream *Xba*I site and also the *5Spr20* promoter which will be cloned upstream (section 3.5.1.2.3).

In transient expression studies, the antisense 5′ *gus* fragment under the control of *5Spr20* promoter provided efficient suppression of GUS expression by pGUS. Expression was lowered by 90% in tissues cotransformed with p5A and pGUS. Interestingly, RT-PCR of cotransformed tissue shows that this was accompanied by a reduction in *gus* transcript levels (Figure-3.30). The negative control used in the experiment was nonbombarded tissue. It would have been preferred if the negative control too was subjected to bombardment (with irrelevant DNA). This would elicit any background fluorescence resulting from the bombardment effect and thus give a more accurate picture of the promoter's effectiveness. However, as mentioned earlier there were constraints on the number of bombardments

that could be conducted and therefore it was opted to have a nonbombarded negative control.

5Spr20 promoter appears to be a promising candidate for use in pine antisense work. These preliminary findings however, should be confirmed by comparing *5Spr20* promoter with other promoters in transient expression studies and its expression pattern verified in stably transformed plants.

4.3.3.3. Sense expression

RT-PCR experiments (Figure-3.29) show that *gus* transcript levels in p5S and pGUS treatments vary by a factor of at least 4-5. The difference between the treatments may reflect the relative transcriptional abilities of *5Spr20* and CaMV35S promoters; it is also likely that the lower transcription level of *5Spr20* was due to premature termination of Pol III transcription at polythymidine repeats. The *gus* CDS which was previously scanned for potential Pol III stop signals, shows runs of four T residues at positions 149, 1658 and 1741. The last two positions which are not perfect consensus signals are yet possible stop sites. Even imperfect stop signals are known to cause Pol III transcription termination in some instances, *albeit* inefficiently; Lewis and Manley [1986], while studying the translational efficiency of a chimeric adenovirus VA_1 promoter (Class III) fused to the herpesvirus thymidine kinase cmRNA found that several truncated transcripts were produced at the sequence CTT with 30% efficiency.

At the translational level (Table-3.2), the *5Spr20* promoter was <10% as efficient as the double CaMV 35S promoter. The anomaly in the transcriptional and translational efficiencies of p5S expression vector suggests that *5Spr20* transcripts are either not competent for polysome-mediated translation or are rapidly degraded by nonsense mediated decay.

4.3.4. Stable expression studies

Several options were considered for stable PTGS studies before deciding on *gus* as the target gene for downregulation. The easiest option is to insert *5Spr20* stably in *N.benthamiana* in a silencing cassette targeting an expressed native

gene. For this to be succesful, the nature of the resident gene and its expression pathway in *N.benthamiana* should be fully understood in order to avoid pleiotropic effects; secondly a convenient gene-specific assay should be available. A second option considered was the use of a virus silencing cassette targeting an RNA virus such as tobacco mosaic virus (TMV), where plants stably transformed with the *5Spr20*-driven silencing cassette are challenged with virions of TMV to assess resistance to TMV infection. However, *gus* was finally selected as the silencing target in order to keep the stable expression studies in line with transient expression studies.

There are two choices in the use of *gus* reporter for stable expression studies. One is to cotransform *N. benthamiana* with a *gus* expression cassette and a *5Spr20*-driven *gus* silencing cassette. This would require only a single round of transformation which suits the tight time frame in which the stable studies had to be completed. The resulting double-transformants can be screened using a combination of antibiotic selection, molecular analyses and GUS assay. Unfortunately the only binary vector available for shRNA delivery was pGreen-0029 which has the same kanamycin selection marker as the GUS expression vector, pGUS. Therefore, selection of cotransformants by differential antibiotic selection is not possible. Neither is it possible to screen plants at an early stage by PCR because of slow plant growth (section 3.6.3.2). But GUS assay is still possible using small leaf segments. Yet, when a putative transformant tests GUS negative it is not certain whether it is an authentic cotransformant where silencing was successful, or one in which the *gus* expression cassette had failed to integrate. Thus, a large number of plants including non-cotransformants have to be grown till such time molecular analysis can be done to confirm cotransformation. This would require more growth-cabinet space which was at a premium.

Therefore, stable expression studies were conducted in the classical way [Jones *et al*, 2001; Kadotani *et al*, 2004] using two rounds of transformation to suit the limited growth-cabinet space; in the first round of transformation GUS expressing transformants were raised which were transformed in the second round of transformation with the silencing cassette.

4.3.4.1. Selection of GUS expressing plants from T0 and T1 transformants

Transformed plants used in long-term studies are often selected on the basis of transgene copy number, gene orientation, inserted cassette integrity, homozygosity and transgene expression. A simplified approach based purely on the expression of the target gene, *gus* was adopted in this study to select GUS-expressing transgenics. A similar method was used by Kadotani *et al* [2004] to study GFP silencing in DCL (dicer-like) mutants of *Magnoportha oryzae*; they too used two rounds of transformation where transformants carrying the target gene *egfp* (enhanced GFP) were selected solely on the basis of fluorescence at 535 nm wavelength.

There are several reasons why this simplified approach was taken in this study. Multiple transgene inserts and invert-repeat inserts are thought to trigger repeat induced silencing and PTGS [Assaad *et al*, 1993]; it is for this reason that copy number and orientation of transgenes are analysed. Likewise, the functioning of the transgene cassette is determined to a large extent by its integrity and this is often verified with PCR primers targeting the 'junction' sequences [De Buck *et al*, 1999]. But despite such screening, there is no guarantee that selected plants will continue to express the transgene. For example, even single-copy transgenes without internal repeat structures are capable of activating gene silencing [Elmayan & Vaucheret, 1996; Jogernsen *et* al, 1996; Vaistij *et al*, 2002]. And as long as GUS is expressed by a plant, it matters little whether it originates from an intact GUS expression cassette or from an intact GUS ORF fragment integrated appropriately behind another promoter in the plant, since the ultimate target for silencing is the *gus* mRNA.

Therefore, GUS expressivity of a plant was the sole criterion used to select GUS plants for silencing, hence the use of GUS histochemical assay. However, if long-term studies are intended homozygation and molecular characterization of transformants are desirable.

4.3.4.2. Choice of construct for dsRNA-mediated silencing

The conventional construct used in plant dsRNA-mediated silencing is the intron-spliced hairpin RNA (ihpRNA) construct (chapter I, 6.2) [Smith *et al,* 2001;

Wesley *et al,* 2001; Stoutjesdijk *et* al, 2002; Byzova *et al*, 2003; Helliwell & Waterhouse, 2003; Padolfini *et al*, 2003; Lo *et al*, 2005]. The spliceosomal intron inserted as a spacer between complementary RNA strands is believed to contribute to the formation and stability of dsRNA resulting in PTGS with almost 100% efficiency [Smith *et al,* 2001]. In comparative studies the ihpRNA construct gave ~90% silencing while cosuppression and antisense constructs gave 0-30% silencing and intron-free hpRNA 50-70%, [Wesley et al, 2001].

However, construction of ihpRNA involves several cloning steps; a typical construct consists of the promoter, a sense strand (~400-800bp) of the target gene, the spliceosomal intron, the antisense strand and a fragment containing the terminator and polyadenylation signals. The ihpRNA construct can be generated in a single step with a PCR product by recombination cloning in the generic ihpRNA vector pHANNIBAL [Helliwell & Waterhouse, 2003]; however, it caters only for Class II promoters and cannot be used to expedite construction of a *5Spr20* vector.

shRNA constructs on the other hand are easy to construct; they can be cloned from amplicons generated by two cycles of overlap PCR [Gou & Liu, 2003], or by using synthetic oligonucleotides [Paul, 2005]. Because of the ease of cloning, shRNA constructs were used in stable expression studies. However, they have never been tested in plants. Secondly, they lack the advantage of hpRNA and antisense RNA which because of the longer target gene sequences they carry are able to present multiple putative siRNA sequences to the target mRNA.

4.3.4.3. Selection of the siRNA sequence used in shRNA construct

siRNA sequences for stable transformation were selected from the same 5′ region of GUS ORF which gave efficient gene downregulation in transient expression studies.

The algorithm of Tuschl *et al* [Dykxhoorn *et al*, 2003] was used to identify potential siRNAs from the 5′ GUS ORF region. A variety of siRNA formats differing in the size of the loop and length of the stem have proven successful

[Paul, 2005]. The selected *gus* sequence 5′ GCCGATGCAGATATTCGTAAT 3′ embodies several features of an ideal siRNA such as 50% GC content, an even representation of nucleotides, absence of repetitive sequences or stretches of a nucleotide especially G, or known sites for mRNA binding proteins [Dykxhoorn, 2003]. Stem lengths of 19-29 nt have been shown to silence genes effectively and loops of 4-23 nt have been described; the 9 loop UUCAAGAGA which was found to be most effective in a comparative study of 5, 7 and 9 nt long loops used with a 19nt stem [Brummelkamp et al, 2002] was chosen for stable expression studies.

The selection of an siRNA is still an empirical process. siRNAs that target different regions of the same gene vary markedly in their effectiveness [Du et al, 2004] and it is usual to test three or more siRNA constructs against the intended target. Furthermore, the *gus* siRNA sequence selected for this study does not conform to motifs such as $AAN_{19}TT$, $NAN_{19}NN$, $NARN_{17}YNN$ and $NANN_{17}YNN$ which are consistent with effective silencing. Therefore, it would be best to test a series of shRNA constructs covering the 5′ GUS ORF region. If all of them failed to downregulate *gus* it is a clear indication that the shRNA system is not suitable for plants. However, only a single shRNA construct was tested because of time and space considerations.

4.3.4.4. *5Spr20*-shRNA-mediated gene silencing

shRNA constructs are most likely to be genetically dominant as are hpRNAs [Helliwell & Waterhouse, 2003] and therefore phenotypes can be screened in T0 or T1 plants without the need to produce homozygous lines. The expression of GUS was completely suppressed in GUS expressing plants transformed with the *5Spr20* promoter driven shRNA cassette (section 3.6.3.1). GUS expression as measured by GUS histochemical assay, was not detected in leaves, petiole and stem of all eight transformed plants. The PTGS effect was absolute during the three months period when plants were monitored.

Even though *gus* expression was completely suppressed, it is still desirable if the construct had been compared with a control cassette driven by a standard

promoter such as CaMV 35S promoter. A control construct was not envisaged in the initial plan for two reasons. The effect of shRNA in plants was itself in doubt; secondly, Fulnecek *et al* [2002] showed that the base at -1 position critical for transcription of 5S rDNA is a guanosine in *Nicotiana* whereas it is a cytosine in *5Spr20*. Therefore, there was a low expectation of the functionality of *5Spr20* promoter in *N. benthamiana*. However, as soon as PTGS was detected at the first GUS assay of *5Spr20* transformants, work did indeed start on the construction of a CaMV 35S-silencing cassette; this was later abandoned when the futility of raising assayable plants before the study deadline was realized.

The level of PTGS induced by *5Spr20* promoter-shRNA is superior to or equivalent to that induced by ihpRNA cassettes under the control of standard promoters used in plant transformation such as the *Ubi, CaMV35S* or *RolC* [Smith *et al*, 2000; Wesley *et* al, 2001; Pandolfini *et al*, 2003; Lo *et al*, 2005]. However, the range of expression levels that normally characterises transformed populations was not observed among the *5Spr20*-shRNA transformed plants when GUS stained tissue was examined at x40 magnification. This can be attributed to the sensitivity of the assay method. In experiments on the silencing of *chalcone synthase* in *Arabidopsis* [Wesley *et al*, 2001], seeds of different transformed lines were indistinguishable to the naked eye; but perceptible differences were observed in pigmentation when seeds were examined under a microscope. Likewise, it is possible that subtle differences in GUS expression among *5Spr20*-shRNA transformants can be detected with a more sensitive assay such as MUG assay. For instance, Lo *et al* [2005] using CaMV 35S promoter in an inducible-silencing cassette, could not detect *gus* transcripts in Northern blots 24 hours after repeat induction with EtOH, but yet were able to detect residual levels of GUS using fluorometry.

The silencing effect of *5Spr20* promoter-shRNA is also similar to, or superior to that obtained with various class III promoters such as U6 and $tRNA^{val}$ used in human cells, which gave >90% silencing of lamin [Paul *et al* 2002] and mutant k-*ras* expression [Kawasaki & Taira, 2003], respectively. However, when the silencing efficiencies of U6, 7SL and 5S rDNA promoter constructs (all Class III promoters) were examined in HeLa cells, the 5S rDNA promoter

failed to silence human lamin expression [Paul, 2003]. This was in spite of 5S rDNA expressed siRNAs co-localizing with lamin transcripts in the cytoplasm. The anomaly was explained by accessibility of the target in particular locations. It is also possible that the RNAi-pathway in humans, which is still not well understood [Paul, 2005] is modulated by subcellular localization of components of the RNAi pathway. Such differences are possible between different biological systems; for instance, mammalian DCR (dicer) is localized in the cytoplasm while DCL1 in plants is nuclear [Bartel, 2004]. As a consequence, the processing of miRNA occurs in the nucleus in plants [Kurihara & Watanabe, 2004], while in animals the process commences in the nucleus and continues in the cytoplasm [Paul, 2005].

RT-PCR showed that *gus* transcript levels were almost non-existent or lowered several-fold in *5Spr20*-shRNA transformed plants (Figure 3.39); this proves that sequence-specific *gus*-downregulation by PTGS was effective. Yet significant levels of *gus* transcripts are present and it is not clear why corresponding GUS protein levels were not detected. Probably it may have another reason as suggested by Lo *et al* [2005]; in their study, when silencing was reversed *gus* transcripts reverted to normal levels in 72 hours, but GUS protein levels did not show comparable recovery and continued to decline to ~5% of pre-induction level; a translation suppression component associated with dsRNA-mediated PTGS was suggested by the authors.

The excellent results obtained with *5Spr20-gus* shRNA also leaves several questions unanswered which could be investigated in future studies. Was the absolute silencing of GUS in *5Spr20*-shRNA transformed plants caused by a single-copy of the transgene or by multiple copies as observed in the silencing of p53 in mammalian cells [Barton & Medzhitov, 2002]? The authors found that silencing of p53 by retroviral delivered Pol III promoter-driven shRNA cassette was highest in super-infected cells whose genomes had most copies of the integrated vector. Should Southern hybridisation reveal several copies of *5Spr20-gus* shRNA, what then is the effect of a single-copy of the silencing cassette?

Off-target effects [Snove *et al,* 2004], insertional mutation and somaclonal variation can also account for *gus* silencing and these possibilities should be eliminated. Prior to use, the *gus* sequence selected for the hairpin was validated by a BLAST search of EST, Unigene and other library databases at NCBI, which failed to reveal any homology between the *gus* sequence and unrelated sequences. Yet, it is still possible that there could have been unsuspected off-target effects. This possibility can be eliminated by an assay for *gus*-specific siRNA in silenced plants.

Accumulation of target-specific siRNA is the hallmark of dsRNA-mediated silencing. siRNA assay (section 3.6.3.2) was attempted only once and the failure to detect *gus*-specific siRNA in this assay can be attributed to insufficient snRNA, high hybridisation temperature and inappropriate probe. Clear evidence for accumulation of *gus*-specific siRNA in *5Spr20-gus* shRNA transformed plants is strong proof that silencing was due to sequence specific degradation by *gus*-siRNA.

CHAPTER IV

ALTERNATIVE PROMOTERS FOR GENE DOWN REGULATION AND EXPRESSION

1. INTRODUCTION

1.1. Candidate genes

Several other candidate genes were also investigated to identify useful pine promoters. These include the class III gene *transfer-DNA*$^{Met\text{-}I}$, three class II housekeeping genes and three genes that were proven to be good expressors in heterologous systems. The aim was to identify the pine orthologs of these candidate genes and clone their promoters.

1.1.1. Candidate gene transfer-DNA of initiator methionine (tDNA$^{Met\text{-}I}$)

Codon usage differs among species [Joshi *et al*, 1989]. However, an invariant feature of most eukaryotes is the use of AUG start codon for translation initiation. Transfer RNA$^{Met\text{-}i}$ (tRNA$^{Met\text{-}i}$) provides the first methionine of the nascent peptide chain and therefore, performs an essential function. The small gene size, a highly conserved sequence [Akama & Tanifuji, 1989] internalized promoters and constitutive expression [Sharp *et al*, 1985; Bourke & Folk, 1992] make *tDNA*$^{Met\text{-}i}$ an attractive candidate for promoter isolation. tDNA genes comprise a transcribed region, which is 72 to 113bp long and contains the internal (minimal) promoters (the A- and B- boxes) required for accurate and basal transcription and the flanking regions which contain the external promoters and signals for transcription termination.

Pol III gene promoters such as those of U6, H1 and tDNA^{Val} are very efficient in producing gene knockouts in humans [Kawasaki & Taira, 2003] and are widely used in human RNAi studies. As entire genomes of plant species become

sequenced the major task ahead is to define the functions of putative genes for which reverse genetics is widely used. A pine $tDNA^{Met-i}$ promoter can fulfil this niche role in pine functional genomics. Furthermore, a group from Poland reported their attempt at using tDNA promoters for protein coding sequences, but details are not available [Kaczmarek & Augustyniak, 1994].

1.1.2. Candidate genes from RNA polymerase II encoded housekeeping genes

An alternative approach is to select candidates from among housekeeping genes, which often exist as multigene families [Long & Dawid,1980]. Some of the best-known plant promoters come from multigene familes; the maize *Ubi* promoter was derived from the ubiquitin family while the rice *Act1* promoter comes from the actin superfamily. A problem often faced in selecting specific paralogs from large multigene families is that many isoforms have to be screened to eliminate low expressed genes and pseudogenes before a desirable gene can be identified. Three candidate genes were considered under this category.

(1) *Actin2* is a member of the *A.thaliana* actin superfamily. It exhibits strong and constitutive expression in vegetative tissues and flowers of *A.thaliana* [An *et al*, 1996]. Its expression pattern is different to that of other actin paralogs.

(2) *ActX* is an uncharacterized pine actin gene, whose full-length cDNA (a gift from Geenz Ltd) was isolated from a directionally cloned cDNA library. It may be possible to design gene-specific primers from its untranslated regions (UTR) for use in RT-PCR, PCR and hybridization.

(3) RNA Polymerase II is composed of 12 subunits [Sakurai & Ishihama, 2001], each of which is encoded by a single copy gene in yeast [Woychik & Young, 1998] and is required in every cell for the transcription of class II genes. *Rpb2* encodes the second largest subunit of Pol II in *A. thaliana* [Ulmasov & Guilfoyle, 1992].

1.1.3. Highly expressed heterologous candidate genes

Also considered for screening were three single-copy genes known to be strong expressors in heterologous hosts. Their pine orthologs, however, may not necessarily be strongly expressed in pine because of different requirements for activated transcription.

(A) EST *T13794* (or *T137*) is constitutively expressed in *A.thaliana* [An *et al* 1996]. *T137* lies in close proximity to *Actin3* in an antisense orientation and their TATA boxes are separated by only 270 bp. This is similiar to the divergent configuration of the human dihydrofolate reductase gene and a mismatch repair protein gene whose transcript start sites are 90 bp apart. Shimada *et al* [1989] showed that a promoter region as small as a 165 bp fragment between the genes is sufficient for accurate bidirectional transcription of the two human genes. The small space between *T137* and *Act3* promoters suggests that, as in the case of the human genes, the promoter sequences required for their expression may be very compact [An *et al*, 1996]. A putative pine *T137* ortholog will therefore, have the two desirable features of strong expression and a compact promoter, provided it retains the functionality and structure of its *Arabidopsis* counterpart.

(B) *AtL18* which encodes the cytosolic ribosomal protein (RP) L18 of *Arabidopsis* is expressed in leaves, stems, rosettes, siliques and roots and at all developmental stages [Baima *et al*, 1995].

(C) RP gene *MsRL5* of *Medicago sativa* is expressed strongly in the stems, leaves, roots, buds and flowers at petal burst and anthesis [Asemoto *et al*, 1994].

1.2. Screening for pine orthologs of candidate genes

Generally, the exons of gene orthologs display a fair degree of conservation while the introns and flanking regions are diverged. PCR primers and hybridisation probes were designed from the coding sequences of candidate genes and used to screen the pine transcriptome for orthologs. A positive signal

in the screen indicates not only the presence of putative ortholog(s) but also their spatial and temporal expression when transcriptomes of different tissues at various developmental stages are compared. Screening was performed under a range of stringent conditions on account of the sequence mismatches that may exist between orthologs of pine and heterologous species. Following this, the cloning of promising pine orthologs was attempted using several PCR-based methods.

1.3. RNA extraction

Nuclear RNA is often associated in ribonucleoprotein particle (RNP) complexes such as ribosomes and spliceosomes, or attached to nuclear pore complexes (NPC) [Galliurd & Strauss, 1990; Mattaj & Englmeier, 1998]; cytosolic RNA is found in polysomal arrays or bound to the endoplasmic reticulum. The RNA has to be separated from other 'contaminants', since the latter could interfere with various assays or trigger RNA degradation [Chang *et al*, 1993; Schultz *et al*, 1994; Chomczynski & Machey, 1995]. Polysaccharides and proteins inhibit reverse transcriptases [Schultz *et al*, 1994; Chomczynski & Machey, 1995] resulting in cDNAs of smaller modal size [Chang *et al*, 1993].

The presence of complex secondary products such as polysaccharides and phenolic compounds make routine isolation of plant RNA difficult [Hughes & Galau, 1988; Baker *et al*, 1990; Manning, 1990; Lopez-Gomez & Gomez-Lim, 1992]. Standard RNA extraction procedures using detergents, high molarity guanidium buffers, sodium perchlorate, chaotropic agents including boric acid, phenol extraction, or density gradient centrifugation often fail when working with plant tissues rich in secondary products [Levi *et al*, 1992]. Pine tissues contain a higher level of phenolics than deciduous trees [Schneiderbauer *et al*, 1991] and are rich in polysaccharides and RNases. Pine RNA extracted by phenol or guanidium hydrochloride methods [Baker *et al*, 1990; Schneiderbauer *et al*, 1991] result in a brownish precipitate; yields are less than 20 μg RNA g^{-1} tissue and the

A_{260}/A_{280} ratio <1.2. Commercial kits based on guanidium isothiocyanate/phenol give higher RNA yields but the quality is low [Chang *et al*, 1993].

The integrity of a total RNA preparation is deduced from the integrity of the major (25S and 18S) rRNA species when RNA is electrophoresed in native or denaturing gels. The hydroxyl group at the 2′ position of the sugar in RNA molecules participates in interactions with phosphates or bases that stabilize folded structures. As a result, mRNA can attain stable tertiary structure(s) and the resulting conformation polymorphisms will display different mobilties in gel electrophoresis. Therefore, denaturing gels are used to size fractionate total RNA for northern blots. Under denaturing conditions and in the presence of the zwitterionic MOPS, RNA migrates as linear monomers [Cantor & Smith, 1999].

2. MATERIALS AND METHODS

2.1. Plant material used in RNA extractions

To study temporal expression, RNA was isolated from three life stages of pine, namely juvenile plants cultured *in vitro*, semi-mature plants from the nursery and mature field plants. Spatial expression was monitored by assaying stems and leaves (needles). The material for expression analyses was collected from the elite pine clone 850.55. Using a single genotype for RNA studies minimises variability in nucleic acid quality, which could arise in RNA sourced from different genotypes which differ in the level of contaminants. RNA was also extracted from control plants such as *A. thaliana* and *Medicago sativa*.

2.2. Modified RNA extraction protocol

Total RNA isolated in this study (chapter II, 1.1) was low in snRNA. Therefore, the following modification was introduced to improve recovery of snRNA. Following serial extraction of the plant extract with $CHCl_3$, the clear supernatant was made up to 0.2 M NaOAc final concentration. 0.54 volume propanol (IPA)

was added with slow stirring, mixed by inversion and incubated at room temperature for 30 minutes. Tubes were centrifuged at 12,000rpm for 30 minutes and the RNA in the supernatant was recovered by adding IPA to 0.3- final volume [Von Ehrerenstein, 1967]. The supernatant was decanted completely and the pellet washed with 70% EtOH and dried at room temperature.

2.3.Probes used in hybridization

(A) A 72 bp probe of $tDNA^{Met-i}$ gene was amplified from pIBI [Akama & Tanifuji, 1989] using primers corresponding to the 5′ and 3′ end regions of *A.thaliana* $tDNA^{met-i}$. Riboprobes were also prepared from $tDNA^{Met-I}$ cloned in pKS.

(B) Digestion of plasmid pB293 containing the cDNA of *ActX* with *Xba*I gave a 1.2 kb fragment containing the entire ~1.1 kb CDS flanked by ~ 50 bases of UTR sequences. The full-length 3′ UTR of *ActX* was also amplified for use as gene-specific probe in Southern hybridization. In addition, a ~200 bp *Pst*I probe was also prepared from *ActX* CDS.

(C) Double digestion of pCDNA II [An *et al*, 1996] with *Xba*I and *Hind*III gave a ~1.6 kb cDNA probe of *Arabidopsis Act2*. Gene-specific probes for Southern hybridization were amplified from the ~300 bp 3′ UTR using forward primer 5′ ctaagctctcaagatcaaaggct 3′ containing the stop codon and reverse primer 5′ aacattgcaaagagtttcaaggt 3′ located 317 bases downstream of the stop codon.

(D) A genomic probe and a cDNA (~300 bp) probe of EST T137 were amplified from *A.thaliana* using primers used by An *et al* [1996].

(E) pL14 and p72/20 [Baima *et al*, 1995] were digested with *EcoRI* to release two non-overlapping cDNA probes of *AtL18* of *A. thaliana*.

(F) *BamHI* digest of pLuz5 [Asemoto *et al*, 1994] gave a 720 bp cDNA probe of ribosomal protein *MsRL5* of *Medicago*.

(G) The ~2 kb cDNA probe, corresponding to the 3′ 2/3rds of the coding region of the gene (*Rpb2*) for the second largest subunit of *Arabidopsis* Pol II was obtained by digesting pAtRP140 [Larkin & Guilafoyle, 1993] at its *Sma*I site.

2.4. Northern hybridization

Total RNA was quantitated fluorometrically and 30 μg total RNA from different pine tissues and control RNA were electrophoresed under denaturing conditions and blotted to positively charged nylon membrane (chapter II, 8.3). Hybridization was performed with ^{32}P- or DIG-labelled probes.

2.5 RT-PCR

2.5.1. RT-PCR of *tDNA*$^{Met-i}$

One μg DNased RNA was reverse transcribed using AMV transcriptase and *tDNA*$^{Met-I}$ antisense primer in a 25 μl reaction; five μl of cDNA (Kohchi *et al*, 1995) was amplified by 22 cycles of PCR using *tDNA*$^{Met-I}$ primers (chapter II, 11.7.2).

2.5.2. RT-PCR of *ActX*

RT-PCR was performed with 3′ UTR sense primer 5′ gtttttaagttcagtgcttaga 3′ and antisense primer 5′ tctagcaatacaaacacaaatt 3′. The PCR segment comprised 22 cycles of a reiterative program consisting of 95°C x 40″, 55°C x 15″, 72°C x 15″, preceded by an initial denaturation at 95°C for 2′.

2.6. Isolation of tRNA^{Met-i} coding region

Primers 5′ atcagagtggcgcagcgg 3′ and 5′ tatcagagccaggtttcgat 3′, corresponding to the 5′ and 3′ end regions, respectively, of *A.thaliana tDNA*$^{met-i}$ flanked by

EcoRI and *HindIII* restriction sites were used to amplify 100 ng pine DNA. PCR consisted of an initial denaturation step of 95°C x 2′ followed by 20 cycles of 95°C x 40″; 60°C x 15″; 72°C x 15″. The amplicon was cloned and sequenced.

2.7. Single primer PCR (SP-PCR)

2.7.1. SP-PCR of *tDNA*$^{Met\text{-}i}$

SP-PCR [Screaton *et* al, 1993] (chapter II, 12.1) optimization was conducted with internal primer 5′ tatcagagccaggtttcgat 3′ (position 72 to 53 of *tDNA*$^{Met\text{-}i}$) at annealing temperatures ranging from 45-55°C and template concentrations of 10 to 500 ng for 40-60 cycles, to determine the optimum annealing temperature for dsDNA formation. After optimization, PCR was performed at the annealing temperature of 50°C for 45 cycles. Each cycle consisted of 95°C x 40″; 50°C x 30″; 72°C x 2′. The PCR products were sequenced with the sequencing primer 5′ cgatcctgggacctgtggga 3′.

2.7.2. SP-PCR of pine *ActX*

The cDNA insert of pB293 contains a putative 183 bp 5′ UTR (Figure-4.1).

5′GGCACGAGCTTATTCTCTCTCACAACCCAAAGCCTTTGTATAACCACTCGAGCTGAGGTGA
GTGTTGGGAAAGTCTCTGCCGTTGTTGCTGCTGCTGCCCTTCTAAGCTGAGGAGGATCCGA
CGAGCAGGATCGATACCTGAATCTCTCTGCAATCGTCTTTACTCAAGTTATCTAGAGGAAG 3′

Figure-4.1. Sequence of *ActX* 5′ UTR. Underlined sequences refer to sequencing primer and internal primers, respectively. See text for explanation

Primers were designed from the distal 140 bases of the noncoding strand of 5′ UTR. Proximal primer 5′ agaagggcagcagcagcaacaa 3′ served as the internal primer (internal primer I) and distal primer 5′ agctcgagtggttatacaaa 3′ as sequencing primer. Internal primer II, 5′ tatcgatcctgctcgtcggat 3′ is located 11 bases from the proximal end of internal primer I (Figure-4.1).

2.8. Asymmetric PCR (A-PCR) of $tDNA^{Met-I}$ and biotinylated trapping oligomers

Primers 5' atcagagtggcgcagcgg 3' (primer-I) and 5' tatcagagccaggtttcgat 3' (primer II) were used to generate ssDNA amplicons of $tRNA^{Met-i}$ genes and their flanking regions by A-PCR. The target sequence was captured by an immobilized biotinylated oligonucleotide probe (primer-I) complementary to the 3′ end of the coding strand. The captured ssDNA was homopolymer-tailed and converted into dsDNA by amplifying the ssDNA with a gene-specific primer and a primer complementary to the homopolymer tail [Lee & Vacquier, 1992; Koenig, 1997].

Conditions for asymmetric PCR were optimised by testing Primers I and II at molar ratios ranging from 1:20 to 1:50 using 20-40 ng pine DNA. PCR consisted of 25 cycles, each made up of 95°Cx1′; 56°Cx20″; 72°x1′, preceded by an initial denaturation step of 95°Cx2′. The preparation of magnetic particle-linked trapping oligo is described in chapter II, 12.2.

The solution containing A-PCR products was denatured at 95°C for 3′, cooled to 80°C and trapping oligonucleotides were added. The solution was gradually cooled to 56°C and annealing continued for a further 10′. The ssDNA trapped on magnetic particles was washed and homopolymer tailed [Greiger *et al*, 1993]. The ssDNA was dissociated from trapping primer by boiling for 3′ and immedietly quenching on ice. The dG-tailed ssDNA was amplified by anchor-PCR [Loh *et al*, 1989] using an 18 nt oligo-dC primer complementary to the homopolymer tract and Primer II. A 15 μl aliquot of the supernatant was amplified using 30 cycles of 95°C x 40″; 56°C x 20″; 72°C x 2′.

2.9. Rapid amplification of genomic ends (RAGE) of $tDNA^{Met-i}$

The method of Mizobuchi & Frohman [1993] was used to isolate the upstream region of $tDNA^{Met-i}$. Five μg pine genomic DNA was digested with *EcoRI*, which cleaves outside the known $tDNA^{Met-i}$ sequence and ligated to 5 μg pSK linearized

with *EcoRI*. Two PCR amplifications were performed sequentially on the ligation mixture using pairs of nested primers. Since there may be two different orientations for each ligated insert, four reactions were performed using primers complementary to the plasmid and the known $tDNA^{Met\text{-}i}$ sequence.

In the first round of PCR, 100 ng of the ligated product was denatured at 95°C for 5′ in a PCR mix containing 25 pmoles of $tDNA^{Met\text{-}i}$ reverse primer and 25 pmole M13 forward or reverse primer and annealed at 55°C for 2′. *Taq* polymerase was added and initial extension was performed at 72°C for 40′. This was followed by 30 cycles of 94°C x 30″; 55°C x 30″; 72°C x 3′. The primary reaction was diluted a 1000-fold and 2 μl used in a second round of PCR with $tDNA^{Met\text{-}i}$ sequencing primer and T7 or T3 plasmid-primers.

3. RESULTS

3.1. RNA yield and purity

The quantity and quality of total RNA were determined by spectrophotometry and gel electrophoresis (Table-4.1). RNA was quantified using A_{260} values and its purity determined by A_{260}/A_{280} ratio, although the validity of this 'venerable and widely used' method [Warburg & Christian, 1942] has been questioned [Glasel, 1995].

Table-4.1 Total RNA yield (μg g^{-1} tissue) and purity

Stage	Juvenile		Semi-mature			Mature	
Tissue	1	2	1	2	3	1	2
Yield[1]	102	72	60	52	28	50	42
Purity	2.1	2.1	2.0	2.1	1.8	2.0	2.2

Table-4.1. A comparison of RNA isolated from different pine tissues. 1 = Stem, 2 = Needle, 3 = Root; [1] Yield based on 1 OD_{260} = 40 μg ml^{-1} RNA. RNA was prepared by a modification of the method of Chang *et al* [1993]. Purity is given by the A260/A/280 ratio.

When electrophoresed in native agarose gels, total RNA gave distinct 25*S* and 18*S* rRNA bands which confirmed that the RNA was intact and undegraded. There was no visible sign of contaminating DNA. (Figure-4.2).

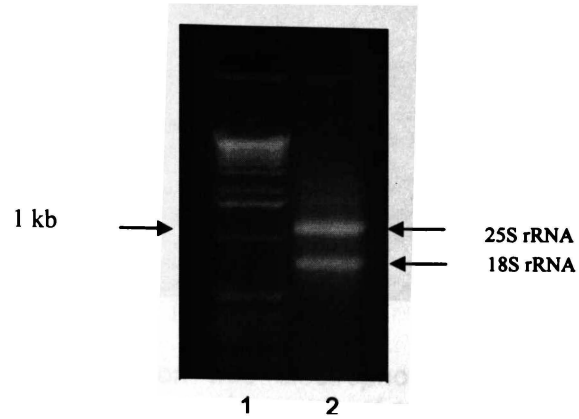


Figure 4.2. One μl total RNA was electrophoresed in 1.5% native TAE gel and visualized by EtBr staining. Lane 1. 1kb ladder, Lane 2. Total RNA.

3.2. Small nuclear RNAs

The small nuclear RNA (snRNA) fraction which contains tRNAs (and 5S rRNA and other snRNAs) is important for the assay of $tRNA^{Met\text{-}I}$. The concentration of snRNA in total RNA preparations was low as evinced by low intensity of fluoresence at the advancing front (~100 bp) of electrophoresed RNA preparations. This is probably due to the use of high molar LiCl in the RNA extraction protocol [Chang *et al*, 1993] which was used in this study. The snRNAs are solubilized at high LiCl concentrations [Avital & Elson, 1969]. Therefore, 2.5 M LiCl may have selectively precipitated the larger mRNAs and major rRNAs while the snRNAs were lost in the supernatant.

RNA was prepared by the modified RNA extraction protocol (Section 2.2.). The preparation obtained by this altered protocol was enriched in snRNA (Figure-4.3).

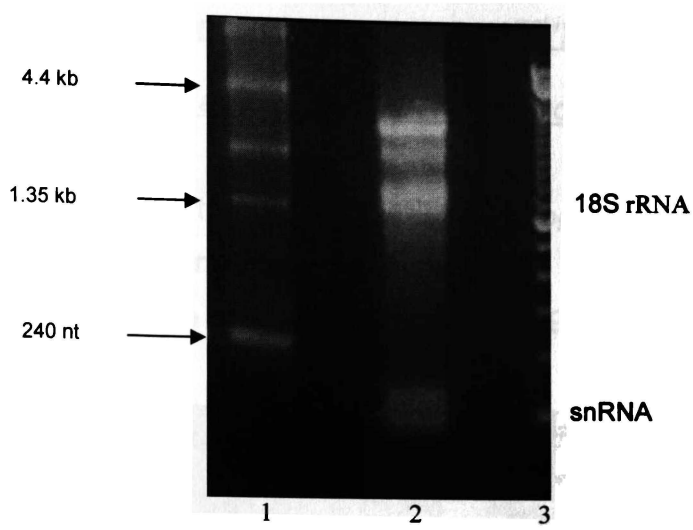


Figure-4.3. Total RNA preparation was electrophoresed in 1.5% agarose gel and visualized by EtBr staining. Shows good recovery of small RNA (snRNA). 1. RNA ladder, 2. 3µl RNA, 3. 100 bp DNA ladder

3.3 *Arabidopsis* and *Glycine max* RNA

RNA prepared from these species was used as positive control in RT-PCR and northern hybridization experiments of candidate housekeeping genes such as $tDNA^{Met\text{-}i}$, *Act 2* and Pol II. RNA was extracted using the modified method from newly emerging leaves of soybean seedlings and young *Arabidopsis* plants (Figure-4.4); juvenile tissue of *G.max* was used since mature or semi-mature tissues contain procyanidins (tannins) which lower the quantity and quality of the extracted RNA [Wang & Vodkin, 1994]. The A_{260}/A_{280} values of RNA prepared from the two species were 2.1 and 1.9, respectively.

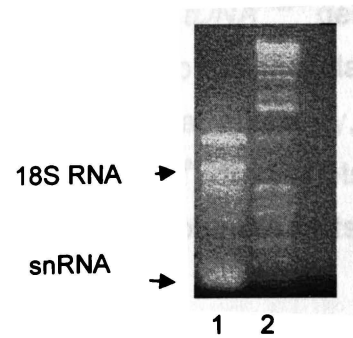


Figure-4.4. *Arabidopsis* total RNA. 1. *Arabidopsis* total RNA, 2. 1 Kb ladder. Total RNA was electrophoresed in 1.5% TAE gel and visualized with EtBr.

3.4. Transfer RNA-methionine initiator gene (*tDNA^{met-i}*) of pine

3.4.1. Expression of the putative pine *tDNA^{met-i}* ortholog

The expression profile of the pine *tDNA^{met-i}* ortholog was studied by northern hybridisation. Pine northern blots were prepared from RNA extracted from different pine tissues by the method of Chang *et al* [1993] (Figure-4.5).

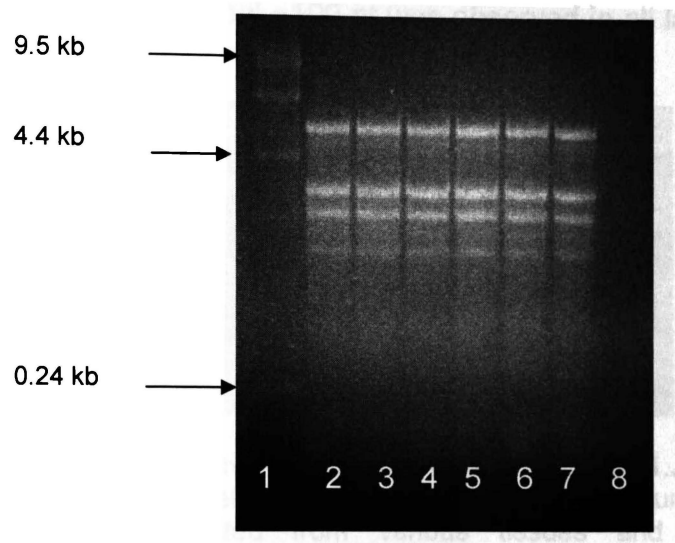


Figure-4.5. A Northern gel containing 30µg total RNA from different pine tissues (lanes 2-7) and 0.5 µg RNA from Arabidopsis (lane 8) run under denaturing conditions and visualized with EtBr staining.: 1. RNA ladder 2.Juvenile stem 3. Juvenile needle 4. Semi-mature stem 5. Semi-mature needle 6. Mature stem 7. Mature needle 8. Arabidopsis RNA.

Hybridization of the blots was performed over a range of stringency conditions using DIG- and later ^{32}P-labelled *tDNA^{met-i}* genomic probes and riboprobes. No signals were obtained, which led to the speculation that pine *tRNA^{met-i}* was a low-copy transcript with high functional efficiency, or had low sequence identity to plant orthologs. Therefore, *tRNA^{met-i}* transcripts in total RNA were first amplified by RT-PCR and then probed in Southern blots; yet no visible signals were seen (Data not shown).

A close examination of RNA gels showed that there was very little snRNA (≤150 nt) in the RNA preparations, which would explain why no *tRNA^{met-i}* was detected. The method of Chang *et al* [1993] possibly extracts the larger species of RNA preferentially, as a consequence of using high molarity LiCl to precipitate RNA. Therefore, fresh total RNA was prepared from pine tissues and control plants by the modified RNA extraction method (section 2.2) and northern blots prepared from these preparations were hybridized with *tDNA^{met-i}* probes at 68°C. A very intense hybridization signal ~100 nt was observed in all lanes of the blot (Figure-4.6).

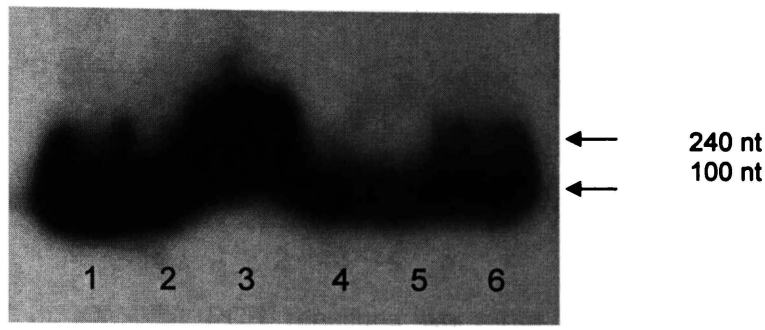


Figure-4.6. Northern blot hybridized with tDNA^{Met-i} probe. 1.Juvenile stem 2. Juvenile needle 3. Semi-mature stem 4. Semi-mature needle 5. Mature stem 6. Mature needle. Total RNA was extracted from various tissues and quantified. 30µg were electrophoresed per lane and transferred to nylon membrane. ^{32}P-labelled *tDNA^{met-i}* probe was hybridised at 68°C and washed at 65°C. Internal controls were not hybridised to show equal RNA loading but see EtBr stained ribosomal RNA bands in figure 4.5.

The steady-state level of tRNA^{met-i} appears to be high in all pine tissues as apparent from above autoradiograph which was exposed for only 30'. The hybridization signals detected in transcripts of all pine tissues confirm the constitutive expression of *tDNA^{met-i}*. The EtBr stained major ribosomal RNA bands in Figure-4.5 confirm equal loading of total RNA in each lane.

3.4.2 Isolation and characterization of pine tRNA^{Met-i} gene

The pine tDNA^{Met-i} gene is a suitable candidate to clone for promoter extraction because of its strong expression in all the tissues tested above. Primers designed from the fully conserved end sequences of *tDNA^{met-i}* of *A.thaliana* [Akama & Tanifuji, 1989] were used to amplify the pine ortholog from pine

genomic DNA. A single band of approximately 80 bp was amplified by PCR. When allowance is made for restriction sites introduced for cloning, the length of the amplicon is ~70 bp which is similar to the size of orthologs in other plant species (figure-4.7).

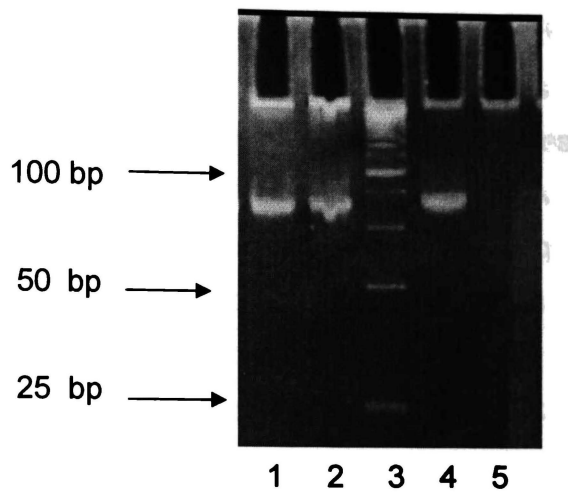


Figure-4.7. PAGE of *tDNA*$^{met\text{-}i}$: PCR reactions were run in 10% PAGE gels and visualized by EtBr staining. 1. *A. thaliana*, 2. *G. max*, 3. 25 bp ladder, 4. Pine, 5. Water.

The sequence of pine *tDNA*$^{met\text{-}i}$ shows absolute homology to orthologs from several plant species (Figure-4.8).

	1	36
Pine[a]	ATCAGAGT**GGCGCAGCGG**AAGCGTGGTGGGCCCATA	
Arabidopsis	ATCAGAGTGGCGCAGCGGAAGCGTGGTGGGCCCATA	
Soybean[b]	ATCAGAGTGGCGCAGCGGAAGCGTGGTGGGCCCATA	
Wheat[c]	ATCAGAGTGGCGCAGCGGAAGCGTGGTGGGCCCATA	
	37	72
Pine	ACCCACAGGTCCC**AGGATCGAAAC**CTGGCTCTGATA	
Arabidopsis	ACCCACAGGTCCCAGGATCGAAACCTGGCTCTGATA	
Soybean	ACCCACAGGTCCCAGGATCGAAACCTGGCTCTGATA	
Wheat	ACCCACAGGTCCCAGGATCGAAACCTGGCTCTGATA	

Figure-4.8. Multiple alignment of *tDNA*$^{Met\text{-}i}$ sequences [a]Underlined region shows the amplified segment: sequences in bold are the split promoters - the 5′ A Box and 3′ B-box. [b]Palmer & Folk [1987], [c]tRNA of wheat [Ghosh et al, 1982]

The amplified region of 33 bp between the primers shows complete identity to the corresponding region in *tDNA*$^{Met-i}$ orthologs. Since *tDNA*$^{Met-i}$ and its transcripts are fully conserved in all species where their sequence is known [Canaday *et al*, 1980; Sprinzl *et al*, 1998; Akama & Tanifuji, 1989], it is proposed that the above PCR sequence represents the complete genomic sequence of pine *tDNA*$^{Met-i}$. However, proof of its end sequences, which in this case were derived from conserved primers can only be obtained by sequencing genomic clones containing the gene, or by primer extension studies of tRNA^{Met-i}. The putative pine tDNA^{Met-i} has all essential features of an active tDNA gene - an A-box at position 9 which conforms to the consensus sequence RGYNNARYGG and corresponds to the 5′ part of the DHU (dihydrouridine) arm and loop of the mature tRNA and a B-box at position 50 of consensus sequence NNG(T/A)TCRANNC which corresponds to the TψC (pseudouridine) loop [Gauss & Sprinzl, 1981; Pavesi *et al*, 1994; Yukawa *et al*, 2000]

3.4.3. Cloning the flanking region of *tDNA*$^{Met-i}$

The genomic region flanking the 5′ end of *tDNA*$^{Met-i}$ may contain *cis*-acting elements as found in several plant tDNAs [Yukawa *et al*, 2000] that influence activated transcription and transcription re-initiation. Therefore several PCR-based methods, which enable walking into the unknown region of a gene were investigated in order to clone putative external promoter elements of *tDNA*$^{Met-i}$.

3.4.3.1. Single primer PCR of tDNA^{Met-i}

SP-PCR (section 2.7.1) uses a single gene-specific primer to amplify both gene-specific and –nonspecific amplicons and a second gene-specific primer to detect the authentic gene-specific product. Optimization studies showed that dsDNA amplification occured freely at annealing temperatures below 52°C, which in itself is in contrast to the findings of Screaton *et al* [1993] who reported it as a rare event. When reactions were performed under optimized conditions and the PCR

reaction containing a mix of heterogenous products were sequenced with a *tDNA*$^{Met-I}$-specific primer distal to the walking primer, no clean signals were obtained. In theory, the sequence primer should seek out the authentic gene amplicon from among several false primed amplicons.

These initial results suggested that there were several copies of pine *tDNA*$^{Met-I}$ gene which differed in their flanking sequences. Therefore the SP-PCR reaction was resolved in PAGE gels and band stabs [Weaver *et* al, 1994] of individual bands were reamplified and sequenced. They gave noisy sequences.

Subsequently, PCR reactions were cloned and about 20 clones were sequenced with the gene-specific primer and later with M13F primer. The sequences were characterized by high noise, or mixed signals of low amplitude (Figure-4.9), or non-specific extension products.

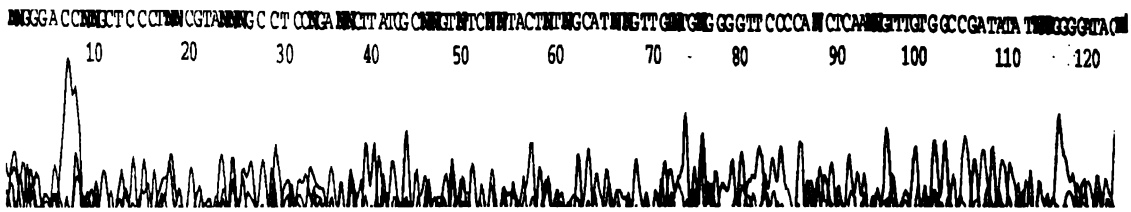


Figure-4.9. The sequence readout of a putative pine *tDNA*$^{Met-I}$ flanking sequence. The SP -PCR product was cloned in a T-vector and sequenced with the pine *tDNA*$^{Met-I}$-specific sequencing primer. The readout was noisy, possibly because of poor annealing of the sequencing primer.

3.4.3.2. Asymmetric PCR of tDNA^{Met-i} and biotinylated trapping oligomers

As SP-PCR was not successful in isolating *tDNA*$^{Met-i}$ flanking sequences, A-PCR using magnetic bead-based trapping oligonucleotides (section 2.8) was tested. During A-PCR optimization, all reactions gave the expected 72 bp band of *tDNA*$^{Met-i}$ produced by the reaction of stoichiometric amounts of forward and reverse primers and a smear of high- molecular weight ssDNA resulting from the

excess reverse primer. The highest yield of ssDNA was obtained using 40 ng pine DNA and a primer ratio of 50:1 (Figure-4.10).

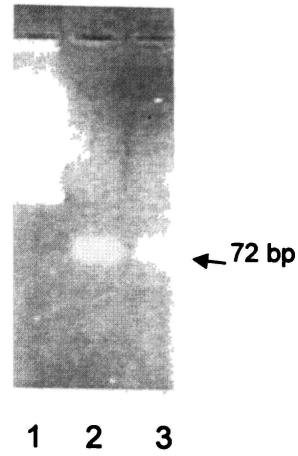


Figure-4.10. Asymmetric PCR performed at different primer ratios. The whole reaction was electrophoresed in 1%TAE and visualized by EtBr staining. 1. 1 kb extension ladder, 2. PCR primed with 1:20 primer ratio, 3. PCR primed with 1:50 primer ratio.

The PCR reaction affinity purified with antisense oligos annealed to magnetic beads and containing ss-DNA$^{Met\text{-}I}$ was homopolymer-tailed and subjected to PCR using oligo-dC primer directed at the homopolymer tail and the nested *tDNA*$^{Met\text{-}i}$ primer II. The PCR cycling was changed to a touchdown program [Don *et* al, 1991], commencing with an annealing temperature of 64°C and falling by 1°C decrements to 56°C to compensate for the higher T*m* of oligo-dC primer. The TD program produced a prominent ~100 bp band and a smear of high molecular weight bands. The 100 bp putative extension product was cloned and sequenced with the reverse primer (from the 3′ end of *tDNA*$^{Met\text{-}i}$). The proximal sequence of the readout was identical to the known 5′ sequence of *tDNA*$^{Met\text{-}i}$; however, it extended directly into a poly-C tract at the 5′ end of *tDNA*$^{Met\text{-}i}$ (Data not shown). Therefore, A-PCR was not successful in walking the *tDNA*$^{Met\text{-}i}$ 5′ flanking region.

3.4.3.3. Rapid amplification of genomic ends (RAGE) of tDNA$^{Met\text{-}i}$

The RAGE protocol (section 2.9) was then tested to clone the *tDNA*$^{Met\text{-}I}$ upstream region. A library consisting of digested pine DNA fragments cloned in pSK was

subjected to nested PCR using combinations of plasmid- and gene-specific primers. The primary RAGE reaction produced a smear and two prominent bands of ~ 400 and 500 bp. The secondary reaction using nested primers produced a band-shift, giving two smaller products ~200-300 bp (Figure-4.11). The two bands of secondary PCR were eluted and sequenced with the *tDNA*$^{Met\text{-}i}$-specific sequencing primer, which should seek out the authentic amplicons. However, a poor sequence readout was obtained.

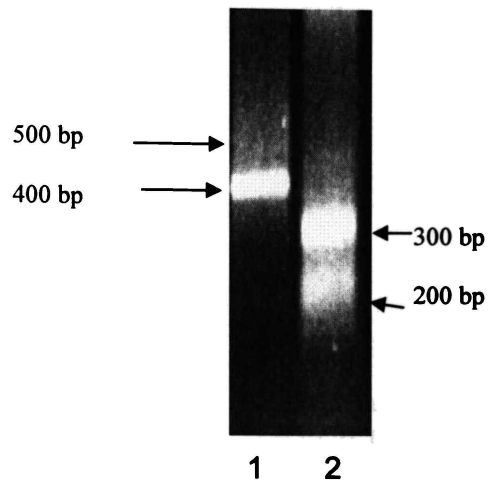


Figure-4.11. Primary and secondary PCR products of *tDNA*$^{Met\text{-}I}$ RAGE-PCR electrophoresed in 1.5% TAE agarose gel and visualized by EtBr staining. 1. Primary PCR 2. Secondary PCR. The 500 bp band is clearly seen in figure-4.12

A Southern hybridisation of RAGE primary PCR was performed to confirm whether the amplicons did in fact contain *tDNA*$^{Met\text{-}i}$-specific extensions. When Southern blots of primary RAGE products were probed with end-labelled *tDNA*$^{Met\text{-}i}$-specific sequencing primer under stringent conditions, it gave rise to a series of closely spaced, fine bands clustered around the ~300 to 500 bp region (Figure-4.12), indicating that there are indeed gene-specific products in the primary reaction.

RAGE was then repeated using different (eg. T7) nested primers, PCR cycles and template concentrations during secondary PCR. The amplicons were cloned

and sequenced but they did not give any 'valid' sequences that overlapped the known *tDNA*$^{Met-I}$ 5′ sequence.

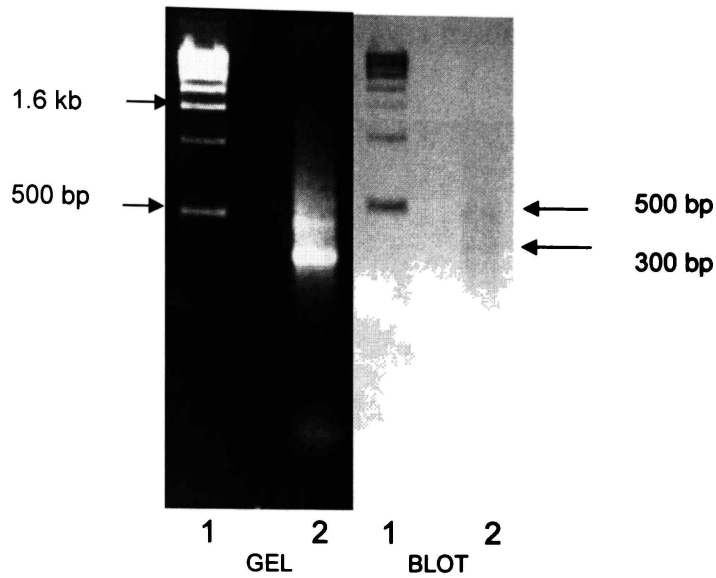


Figure-4.12 Hybridization of primary PCR products with gene-specific probe Figure shows the gel and corresponding blot; The PCR products were electrophoresed in TAE gels and alkali-transferred to positively charged membranes. The blot was hybridised with end labelled sequencing primer. 1. 1 kb ladder, 2. Primary PCR

3.4.3.4. Thermal asymmetric interlaced PCR (TAIL-PCR) of tDNA^{Met-i}

TAIL-PCR (chapter II, 12.3) which uses three nested *tDNA*$^{Met-I}$ primers and a constant arbitrary primer in three rounds of PCR to amplify gene-specific amplicons was then tested to isolate *tDNA*$^{Met-I}$ 5′ flanking regions. The primary reaction when electrophoresed, produced two bands of ~500 and 600 bp; however, no visible products were obtained in secondary or tertiary PCR.

3.4.3.5. Enriched small-insert genomic library of tDNA^{Met-i}

A final attempt at cloning the upstream region was made using the method of Karagyozov *et al* [1993], which was used to recover complementary genomic sequences from genomic DNA digests using short sequences such as simple sequence repeats (SSR) immobilized on solid supports (Chapter II, 12.4). In this

experiment, *HaeIII* digested and adaptor ligated genomic DNA was amplified by adaptor primers. The resulting library was enriched for *tDNA*$^{Met-I}$ fragments by adsorption with immobilized *tDNA*$^{Met-I}$–specific oligos. Two rounds of enrichment were performed. The primary amplified library contained fragments in the 0.2-0.8 kb range. The final enriched library was made up of fragments ~200-500 bp long (Figure-4.13).

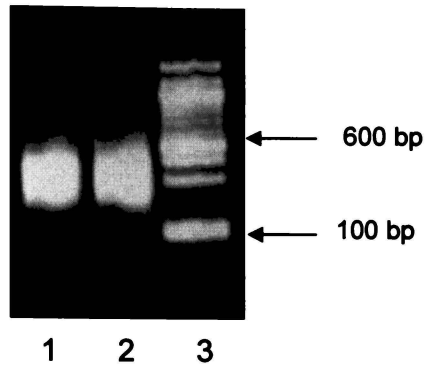


Figure-4.13. 1. The entire PCR reaction of secondary and primary enrichments were electrophoresed in 2% TAE gel and visualized by EtBr staining. 1. Secondary enrichment 2. Primary enrichment 3. 100bp ladder. See text for explanation.

The library was cloned and transformants grown on blue/white selection media were colony hybridized using end-labelled *tDNA*$^{Met-i}$ forward primer. The probe failed to distinguish the 'white' from the (control) 'blue' colonies even under stringent hybridization conditions. Therefore, transformants were selected by *Lacz* α-complementation alone and plasmids of selected clones were sequenced with M13 forward primer, but none of them contained known *tDNA*$^{Met-I}$ sequences. *In vitro* transcribed pine tRNA^{Met-i} was later used as a denatured trapping ligand to produce fresh libraries but again, none of the sequenced clones contained pine *tDNA*$^{Met-I}$ sequences.

3.4.4. Southern Hybridization of pine tDNA^{Met-i}

Southern hybridization was performed with ^{32}P-random labelled *tDNA*$^{met-i}$ probes to determine the copy status of pine tDNA^{Met-i}. This may, at least in part, provide reasons for failure of several attempts to walk the 5′ flanking region of the gene.

Hybridization and washing were performed over a range of stringency conditions but no signals were detected (Data not shown).

The efficiency of the probe was validated by hybridising the ^{32}P- labelled *tDNA^{met-i}* probe with a control blot carrying *Hind* III digested *A. thaliana* DNA; it gave at least three visible bands when hybridised at 55°C (Figure-4.14)

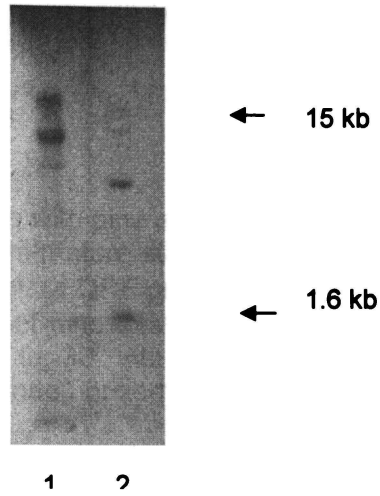


Figure-4.14. Southern blot of *Arabidopsis* DNA. 25 µg of DNA was digested with *Hind*III, electrophoresed in 0.8% TAE gel and transferred to positively charged nylon membrane. The blot was hybridised with a ^{32}P-labelled *tDNA^{Met-I}* probe at 55°C and washed at 45°C. Lane 1. EcoRI digested DNA, Lane 2. 1 kb extension ladder.

3.5. Actin gene *ActX* of Pine

The second pine gene considered for cloning was pine *ActX*, an actin gene. Sequencing of a cDNA clone of the gene (gifted by GEENZ Ltd) revealed a 200 bp 5′ UTR, a 1.1 kb ORF and a full length 3′ UTR ending in a poly-A tail. Homology search of the ORF sequence showed that it bore 85-90% identity to actin coding sequences of several angiosperm species.

3.5.1. Northern hybridisation of pine *ActX*

The expression of pine *ActX* was studied by Northern hybridisation. Northern blots were probed with a ~1.2 kb cDNA probe of *ActX*. Since it is an endogenous gene, hybridization (68°C) and washing (65°C) were performed at high

stringency. Very strong hybridization signals were observed in all lanes of the blot (Figure-4.15).

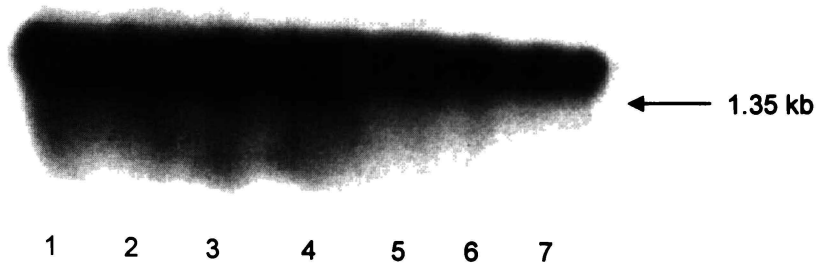


Figure-4.15. Northern blot hybridized with pine actin probe. 1.Juvenile needle 2. Juvenile stem 3. Semi-mature needle 4. Semi-mature stem 5. Mature needle 6. Mature stem 7. Arabidopsis total RNA. ^{32}P-labelled *ActX* ORF probe was hybridised at 68°C and washed at 65°C. Each lane contains 30 µg of pine total RNA; the *Arabidopsis* lane contains 15 µg total RNA. The film was exposed for 30'. Internal controls were not hybridised to show equal RNA loading but see EtBr stained ribosomal RNA bands in figure 4.5.

A BLASTn search showed that the 3' UTR of *ActX* is unique to the gene. Therefore it was used to probe fresh Northern blots carrying 30 µg total RNA of each pine tissue at 65°C to assay gene-specific expression of *ActX* (Figure-4.16). The steady state level of *ActX* mRNA was higher in the needles of mature, semi-mature and juvenile plants and moderate in stems of juvenile and mature plants. Overall, actin expression appears to vary by 10 to 40-fold among the tissues.

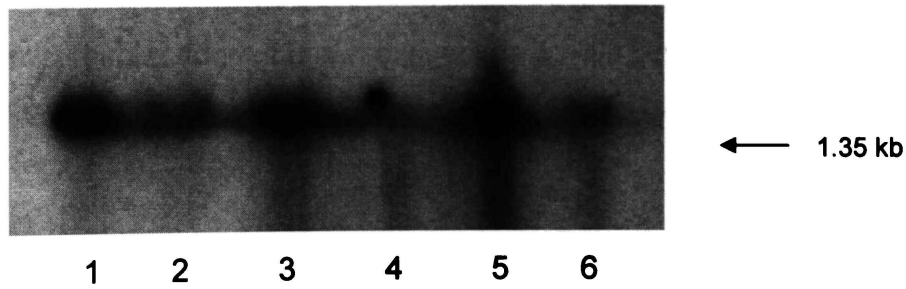


Figure-4.16. Northern hybridization with 3' UTR probe of pine actin. 1.Juvenile needle 2. Juvenile stem 3. Semi-mature needle 4. Semi-mature stem 5. Mature needle 6. Mature stem. ^{32}P-labelled *ActX* 3' UTR probe was hybridised at 65°C and washed at 65°C. Internal controls were not hybridised to show equal RNA loading but see EtBr stained ribosomal RNA bands in figure 4.5.

3.5.2. RT-PCR of pine *ActX*

An RT-PCR using *ActX*–specific primers and 1 µg total RNA from each tissue was also done to confirm *ActX*–gene specific expression (section 2.5.2). A single ~200 bp product of the expected size was amplified from 250 ng DNAsed-total RNA of all tissues (Figure-4.17) using gene specific primers obtained from *ActX* 3′UTR. It confirms that pine *ActX* is expressed in all pine tissues tested; the results parallel those of northern hybridization with a higher expression observed in the needles of all stages and in mature stem. These conclusions could have been stated with more confidence if quantitative-RT-PCR had been done.

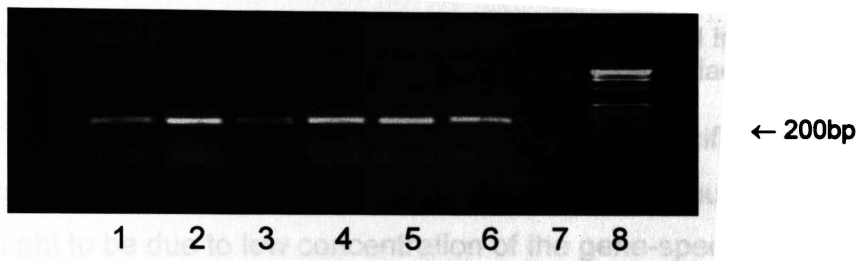


Figure-4.17. RT-PCR of total RNA using pine actin 3′UTR-specific primers. RT and PCR were performed at 50°C and 56°C, respectively. Reactions were electrophoresed in 1.5% TAE gel and visualised by EtBr staining. 1.Juvenile stem 2. Juvenile needle 3. Nursery stem 4. Nursery needle 5. Mature stem 6. Mature needle 7. Juvenile needle without RT (-ve control) 8. 1 kb ladder.

3.5.3. Cloning the pine *ActX* promoter using single primer PCR

ActX, because of its strong to moderate expression in all pine tissues, is an attractive candidate for cloning and promoter isolation. SP-PCR was used to walk into the promoter region of *ActX*.

During SP-PCR optimization, reactions primed with 4-40 pmole internal primer I (Section 2.7.2) gave three nonspecific bands, which also occurred in the negative control. In order to eliminate false positives, hotstart PCR was performed using ampliwax® beads, but nonspecific amplification persisted in the negative water control, which suggested internal priming of the primer [Kolmodin & Williams,

1997]. A new internal primer (internal primer II) was used, which produced four bands, ranging in size from ~0.6 to 1.1 kb (Figure-4.18).

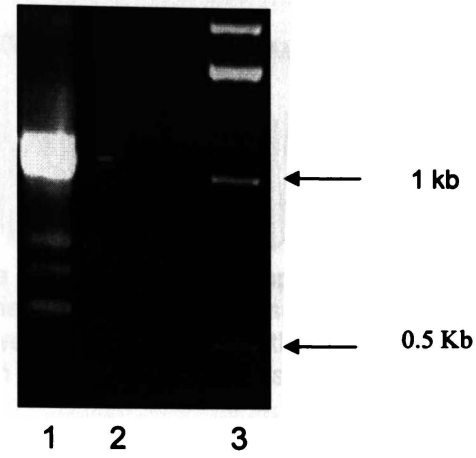


Figure-4.18. SP-PCR using Primer II. Reactions were electrophoresed in 1.5% TAE gel and visualised by EtBr staining. 1. PCR products 2. -ve control 3. 1 kb ladder

4μl of the PCR reaction was sequenced using the *ActX*-specific 'sequencing primer', external to the walking primer used in PCR; a poor readout was obtained which was thought to be due to low concentration of the gene-specific template in the heterogeneous PCR mix, or non-specific annealing of the sequencing primer, or due to the sequencing primer encountering extension products of several actin paralogs containing complementary regions. These various possibilities were validated by using a new sequencing primer which possessed a 'stronger' 3′ end, optimizing the concentration of the gene-specific template by sequencing 2, 4 and 8μl aliquots of SP-PCR products that had been concentrated, sequencing of DNA extracted from the individual bands and by sequencing band stabs of reaction products resolved in PAGE gels [Caetano-Anolles & Trigiano, 1998]. The sequence readout however was not improved by these changes (Data not shown).

The SP-PCR products were then cloned [Marchuk *et al*, 1991] and randomly selected clones were sequenced using the actin 'sequencing' primer and later M13 F primer. Most inserts gave very poor sequence readouts possibly because their inserts were products of two false priming sites. The few that gave good

sequence results with M13 F primer had no actin signature sequences, nor any overlap with the known 5′ sequence (Figure-4.19).

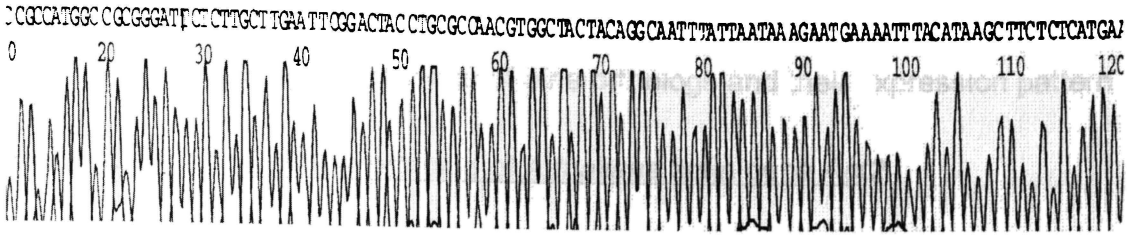


Figure-4.19 A clean sequence of a cloned SP-PCR amplicon. Assuming that the first base called is within 10-20 bp of the 3'end of the *ActX*-specific sequencing primer, the proximal bases of the sequence were compared with the known 34 bases 5′ to the sequencing primer. No overlap with known 5' sequences was observed; nor were any actin signature sequences seen.

3.5.4.Southern analysis of *ActX*

Blots probed under a range of stringency conditions with the 3′ UTR probe of *ActX* did not produce any hybridization signals. A *Pst*I fragment of *ActX* coding sequence (section 2.3) was then used to validate the blot; it produced several bands as expected of a conserved exonic sequence, but these were non-specific since they occurred at identical positions in the different digest lanes (Figure-4.20).

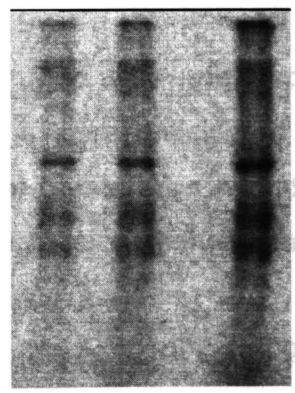


Figure-4.20. A Southern blot of pine DNA probed with *Pst*I fragment of *ActX* CDS. Lane 1. *Xba*I, Lane 2. *Hind*III, Lane 3. *EcoR*I. 30 μg of pine DNA digests were electrophoresed in 0.8% TAE gels, transferred to nylon membranes and probed with ^{32}P-labelled *ActX* *Pst*I fragment. Blots were hybridised at 60°C and washed at 45°C.

3.6. Pine orthologs of heterologous single-copy gene candidates

The pine transcriptome was also screened with exonic probes of candidate genes *Act2, T137, AtL 18* and *Rpb2* of *Arabidopsis* and *MsRL5* of alfalfa (section 1.1.2, 1.1.3) to verify the presence of pine orthologs and their expression pattern.

3.6. 1. Expression of pine ortholog of *A.thaliana* actin gene *Act2*

Northern hybridisation was performed with *Act2* cDNA probe at 68°C and the blots were washed stringently at 65°C. Very strong signals, similar to those obtained with pine actin probe were observed in all lanes (Figure-4.21).

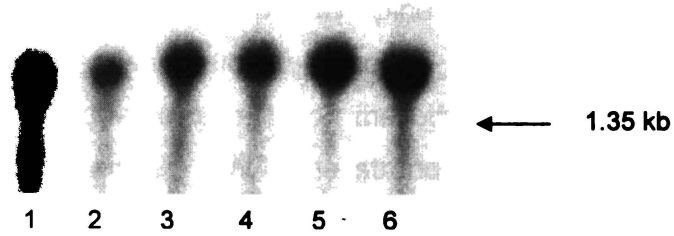


Figure-4.21. Northern blot of pine total RNA probed with *Act2* cDNA probe. 1.Juvenile stem 2. Juvenile needle 3. Semi-mature stem 4. Semi-mature needle 5. Mature stem 6. Mature needle. Blots were probed with ^{32}P-labelled *Act2* cDNA at 68°C and washed at 65°C.

Northern blots were then re-probed with a *Act2*-specific 3′ UTR probe. Possible mismatches between orthologs were anticipated and a range of hybridization temperature from 65-50°C were tested; no hybridization signals were detected even when low stringency washing conditions were adopted.

3.6.2. Expression of *T137*, RP and *Rpb2* pine orthologs

The pine northern blots failed to produce any signals when hybridized with the genomic and cDNA probes of *T137*, or the cDNA probes of MsRL5 and AtL18 ribosomal proteins and of *A.thaliana Rpb2*, at hybridization temperatures ranging from 65-50°C and washing temperatures of 60-37°C (data not shown).

4. DISCUSSION

4. 1. The $tRNA^{met-I}$ gene of pine

4. 1.1. Expression and copy number of *tDNA*$^{met-i}$

The high expression of $tRNA^{met-i}$ in all pine tissues (Figure-4.6) is most likely due to high copy number of the gene. In many protists, mycophytes, metazoans, arthropods, rodents and primates the total number of tRNA genes exceed by far the number of different tRNA sequences [Long & Dawid, 1980]. In angiosperms too, most known tRNA genes are found in multicopy gene families, although some tDNAs do exist as single-copy genes [Green *et al,* 1987; Akama & Tanifuji, 1989; Ramammonjisoa *et al*, 1998]. If the putative pine *tDNA*$^{met-i}$ proves to be a single-copy gene, its promoter is predicted to be an excellent candidate for constitutive expression in transgenic pine. If, on the other hand, pine *tDNA*$^{met-i}$ follows the norm of most eukaryotic *tDNA*, the strong signals in northerns is attributable to transcripts of several *tDNA*$^{met-i}$ paralogs, or of subsets of differentially expressed paralogs. The possibility also exists that most of the *tRNA*$^{met-i}$ expression is due to one or a few paralogs which are strongly and constitutively expressed; a similar situation is found in *A.thaliana* where *Act2* and *Act8* belonging to the 10-member actin superfamily encode nearly 80% of the plant's actin [Meagher *et al*, 2000].

The copy status of pine *tDNA*$^{met-i}$ could not be verified by Southern hybridisation using a probe which hybridises with *Arabidopsis* DNA (Figure-4.14), possibly because pine *tDNA*$^{met-i}$ is either a single-copy gene, or constitutes a small multicopy gene family which is dispersed through the pine genome [Long & Dawid, 1980]; considering the larger than usual size of pine nuclear DNA, it would appear that more DNA (>30 μg per lane) may be needed to be probed in Southern analysis in order to detect *tDNA*$^{met-i}$.

4. 1.2. Isolation of the 5′ flanking region of pine *tDNA*$^{Met\text{-}i}$

The gene (containing the internal promoters) and its 5′ flanking region may be needed for activated transcription. Almost all nuclear tDNA genes are characterized by internal promoters. Using hybrid tDNA constructs, Ciliberto *et al* [1982] demonstrated that these intragenic control regions (ICR) alone are sufficient to direct efficient transcription. Therefore, an expression cassete driven by the putative pine *tDNA*$^{Met\text{-}i}$ sequence alone containing the internalized promoters is sufficient to direct strong transcript expression in transgenic plants.

However, studies have shown that the 5′ flanking region of several tDNAs may modulate transcription in a positive [Schaack *et al*, 1985; Campbell *et al*, 1985; Arnold & Gross, 1987; Yukawa *et al*, 2000], or negative [DeFranco *et al*, 1981; Tapping *et al*, 1993] manner. Therefore, it is necessary to characterize the 5′ flanking region of the putative pine *tDNA*$^{Met\text{-}i}$ to confirm whether it is needed for activated transcription in transformed plants. Hence, many methods were tested for the isolation of the 5′ flanking region of pine *tDNA*$^{Met\text{-}i}$.

4. 1.2.1. Single primer PCR (SP-PCR) of *tDNA*$^{met\text{-}i}$

SP-PCR has three requirements - an internal primer (I) to initiate extension into the upstream region, a sequencing primer (S) to sequester the gene-specific product and some degree of overlap between the 'known' and 'unknown' regions to aid contig mapping (Figure-4.22).

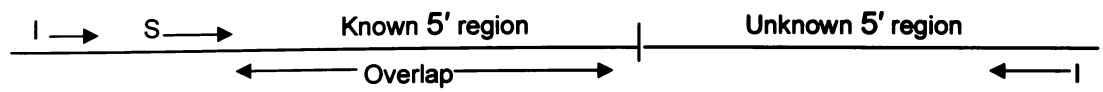


Figure-4.22. Shows the internal primer (I) annealing non-specifically to both strands of the DNA to give an amplicon. In this instance it has annealed to sites in the known and unknown regions to give an authentic walk into the unknown 5′ region. Nested sequencing primer (S) is used to sequence the gene-specific extension from among several false primed amplicons. The overlap region confirms authenticity of the gene-specific amplicon.

The application of this method to *tDNA*$^{Met\text{-}i}$ is constrained by the limited size of the known sequence, since all three requirements need to be fitted within the

short 72 bp *tDNA*$^{Met\text{-}i}$ sequence. The internal primer was therefore, selected from the very end (3′) of *tDNA*$^{Met\text{-}i}$ (position +72 to +53). The sequencing primer was pushed back as far as possible so that it overlaps the latter by 4 bases which, however, should not affect the primer's binding specificity [Kwok *et al*, 1990]. The sequencing primer was designed from position +56 to +37 (Figure 4.8) which still leaves a stretch of 36 5′ bases for contig mapping.

The two lines of defence used to authenticate *tDNA*$^{Met\text{-}i}$-specific amplicons were the use of a gene-specific sequencing primer, which seeks out the specific extension product, and a sequence overlap at the 5′ and 3′ ends of the 'known' and 'unknown' regions. This overlap is critical in the case of *tDNA*$^{Met\text{-}i}$ since its 5′ flanking region has no signature sequences except for a shared region of low duplex stability around positions -30 to -35 and a CAA triplet in the region of the presumed transcription site [Choisne *et al*, 1997]. Despite these precautions, none of the sequence-able SP-PCR products showed any sequence overlap.

4. 1.2.2. Asymmetric PCR of *tDNA*$^{met\text{-}i}$ and biotinylated trapping oligomers

The primers used in ssDNA amplification and trapping (Section 2.8) were designed to bias selection towards functional paralogs. Primer-I, 5' atcagagt**ggcgcagcgg** 3' contains the 10 bases constituting the A-box (shown in bold) and 5' sequences necessary for the stem structure of the tRNA cloverleaf, while Primer-II, 5' tatcagagccag**gtttcgat** 3' contains the amino acid acceptor terminus from the 3′ OH end and eight of the 11 bases forming the B-box (shown in bold).

The sequence of the putative extension product (Section 3.4.3.2) suggests that the 100 bp fragment is the amplification product of a tailed 72 bp ss*tDNA*$^{Met\text{-}i}$ template. The 95°C denaturing step used to denature ssDNA prior to trapping, also denatures the large fraction of 72 bp ds*tDNA*$^{Met\text{-}i}$ in the A-PCR products; during the following annealing step, the coding strands of this fraction compete

with ssDNA for binding to the trapping oligonucleotide. It is therefore, possible that a majority of the tailed products were either 72 nt, single-stranded *tDNA*$^{Met-i}$ which were tailed at their 3′ end, or double-stranded *tDNA*$^{Met-i}$ (forming an open triplex with the trapping oligonucleotide) which were tailed on both strands. During subsequent PCR, these tailed 72 bp fragments, are preferentially amplified because of their smaller size resulting in the prominent ~100 bp product.

This situation could have been avoided by fractionating the A-PCR products through hydroxylapatite columns to which, ss nucleic acids bind comparatively weakly at lower phosphate concentrations. The authentic *tDNA*$^{Met-i}$ extension product can be amplified from the homopolymer-tailed eluate, using nested-PCR. The weak link in the method is homopolymer tailing [Delort *et al,* 1989]. A suitable alternative would be single strand ligation-mediated (SLIC) anchor PCR [Edwards *et al*, 1998] of the ssDNA products, where their 3′ end is anchored covalently to an oligonucleotide of defined sequence [Tessier *et al*, 1986].

4. 1.2.3. Rapid amplification of genomic ends (RAGE) of *tDNA*$^{met-i}$

The RAGE protocol (Figure-4.23) is a modification of single-specific primer PCR (SSP-PCR) of Shyamala & Ames [1989] and RACE-PCR [Frohman *et al,* 1989].

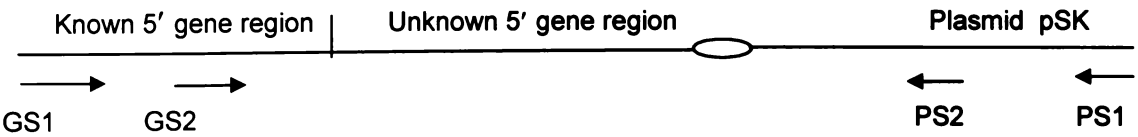


Figure-4.22 RAGE protocol. The ring denotes the EcoRI ligation site of the genomic fragment and pSK. Nested PCR is performed with GS1 and GS2 nested gene-specific primers and PS1 and PS2 nested plasmid-specific primers.

The band eluates of pine *tDNA*$^{met-i}$ RAGE secondary PCR sequenced poorly (Section 3.4.3.3), possibly, due to the eluates containing, instead of a single product, several pseudogenes and functional paralogs which had a common binding site for the *tDNA*$^{met-i}$–specific sequencing primer. The bands in the autoradiograph of primary PCR (Figure-4.12) may represent the different

paralogs, whose *Eco*RI site by which they are ligated to the constant *Eco*RI site of the vector is located at variable distances from the known 5′ end of the gene.

The marked band shift between visible products of primary and secondary nested PCRs (Figure-4.11) is too large to be accounted for by the distance differential between pairs of nested primers, which is at most around 40-50 bp. This suggests that the prominent ~200 bp band of secondary PCR is an artifact and the authentic *tDNA*$^{Met\text{-}i}$ amplicons are probably located in the 300 bp band and in the smear above it. The appearance of a prominent artifact in secondary PCR is difficult to explain since nested PCR is designed to eliminate such nonspecific amplification.

These results contrast with the findings of Mizobuchi & Frohman [1993], where a single intense band without artifacts was obtained which enabled direct sequencing of the PCR reaction; Southern blots of their primary PCR products too, gave only a single band. Cormack and Somssich [1997] were able to clone 500 and 900 bp of promoter regions from *Arabidopsis* and parsley, respectively, using three rounds of PCR in a modified RAGE protocol. Therefore, RAGE results of pine *tDNA*$^{Met\text{-}i}$ probably reflect the weakness of this method when applied to the analyses of putative multigene families.

4. 1.2.4. Thermal asymmetric interlaced PCR (TAIL-PCR) of *tDNA*$^{met\text{-}i}$

TAIL-PCR was used succesfully to amplify single-copy sequences in complex genomes ranging from *A.thaliana* (C=1.15 x 10^8) to *Triticum aestivum* (C=1.5x10^{10}) [Liu *et al*, 1995]. Among most PCR-based methods used for gene walking, TAIL-PCR is best suited for simultaneous amplification of more than one specific DNA fragment of interest [Liu *et al*, 1995]. Therefore, the failure of TAIL-PCR in this study cannot be attributed to the putative multigenic nature of *tDNA*$^{Met\text{-}I}$. The wrong choice of arbitrary primers was the probable reason for its failure to isolate *tDNA*$^{Met\text{-}i}$ sequences. The efficiency of arbitrary primers to

amplify specific products ranges from 50-80%, to nil in some cases [Liu *et al*, 1995].

4. 1.2.5. Enriched small-insert genomic library of *tDNA*$^{met-i}$

Membrane-bound oligonucleotides (30-40 nt long) have been used successfully to capture 250-900 bp long DNA fragments containing complementary sequences, from digested and denatured genomic DNA [Karagyozov *et al*, 1993; Edwards *et al*, 1996]. Therefore, there is a reasonable chance that immobilized 72 bp *tDNA*$^{Met-i}$ ligands can also be used to capture *tDNA*$^{Met-i}$ genomic fragments from an enriched small-insert genomic library (Wilkins *pers com.*).

Wheat germ-tRNA (WGT) (Sigma) contains most of the cytosolic tRNAs whose sequences are well conserved [Ghosh *et al,* 1982; Sprinzl *et al,* 1998]. Therefore, blots carrying bound WGT can be expected to capture several pine *tDNA* orthologs. The failure of WGT blots to retrieve tDNA fragments (Section 3.4.3.5) could be due to the closed structure assumed by immobolized ligands when they are adsorbed to the blot under conditions favouring their native configuration. Native tRNAS fold into a compact L-shaped tertiary structure offering at best only three, 3-7 nt single stranded regions for hybridization. Therefore, *in vitro* transcribed tRNA^{Met-i} was bound to membranes in the denatured form (in the presence of formamide) and used to create new enriched libraries. But none of the recovered sequences were related to *tDNA*$^{Met-i}$. Yet, some of these could have been *tDNA*$^{Met-i}$ flanking regions which were not detected during homology search because of flanking sequence diversity, or because of too short an overlap with the known *tDNA*$^{Met-i}$ sequence.

4. 2. Actin gene *ActX* of pine

4. 2.1. Northern hybridisation of pine *ActX*

The strong signal obtained in pine northern blots probed with *ActX* cDNA (Section 3.5.1) may be the result of closely related transcripts of a putative

multigene family hybridizing to the highly conserved ORF of *ActX*. Therefore, these results are not sufficient proof of constitutive expression of *ActX*, even if hybridization was performed under stringent conditions. The gene-specific expression of *ActX* can be demonstrated by differential washing, which was used to distinguish functional genes from closely related variants (eg. pseudogenes) [Sorenson & Fredrikson, 1991] through a gradual reduction of monovalent cations in the washing buffer. But, even then, there is no way of proving that the residual signals in a blot are due to a single class of hybrid molecules.

Therefore, a putative *ActX*–specific probe was used to probe northern blots. Northern hybridization using *ActX* 3′ UTR probe showed that *ActX* is expressed in all tissues tested in this study and that its spatial expression varied by about <40-fold. Several globally expressed genes show similar variation in their tissue-specific expression; the constitutively expressed *Act8* and *Act2* genes of *A.thaliana* show a 2- to16-fold variation in their expression levels compared to tissue-specific actin paralogs whose expression vary by an order of magnitude of >100 [Hightower & Meagher, 1985; An *et al*, 1996]. Therefore, *ActX* appears to be a good candidate gene, which merits further investigation

4. 2.2. RT-PCR of pine *ActX*

Although UTR sequences generally tend to be unique, the *ActX* 3′ UTR probe used to establish *ActX*-specific expression (section 3.5.1) shares high homology (~90% identity) with the 3′ UTR of partially and fully sequenced actin mRNAs of related conifers *P. taeda* and *Picea mariana*, respectively. If a similar degree of identity exist between 3′ UTRs of *ActX* and other pine actin paralogs, it would invalidate the findings of northern hybridization using *ActX*-specific 3′ UTR probes, since the expression profile can be attributed to more than one gene.

Allele-specific RT-PCR, similar to differential display used to detect RNA polymorphisms, is based on the ability of primers containing mutations at their 3′

end to specifically amplify alleles with sites complementary to their 3′ end [Newton *et al*, 1989]. The sequences of *ActX* 3′ UTR and the actin UTR sequences of *P.taeda* and *P.mariana* show polymorphisms due to transversion of two upstream and one downstream purine bases, which are separated by ~170 bp. Therefore, allele-specific sense primer 5′ gtttttaagttcagtgcttaga 3′ and antisense primer 5′ tctagcaatacaaacacaaatt 3′, displaying mismatches at their 3′ ends to corresponding sequences of *P.taeda* and *P.mariana* orthologs were used in RT-PCR (Section 3.5.2). Results of RT-PCR appear to support the findings of northern hybridisation conducted with the *ActX* 3′ UTR primer.

However, even the RT-PCR results are not conclusive for several reasons. Well-to-well temperature differences within the thermocycler block is a common source of variation in quantitative PCR; even though the MJ thermocyclers were validated using a thermocouple, it would be still desirable to monitor the uniformity of the reaction by coamplifying an internal standard like 18S rRNA.

Secondly, it is questionable whether the temperature profile of RT-PCR was adequate to ensure gene-specific amplification. The RT phase of the reaction was performed at the permissive temperature of 50°C and therefore, the gene-specificity of RT-PCR hinges on selective amplification by the sense primer. The annealing temperature which defines PCR specificity is predicted using various algorithms [Don *et al*, 1991; Wu *et al*, 1991; Ausubel *et al*, 1998] and determined by PCR optimization. In optimization reactions using pine DNA, the annealing temperature was raised to 55°C without loss in PCR yield. The exact annealing temperature, however, could be determined with greater precision by testing a battery of primers identical to the sense primer but containing all possible permutations of bases at their 3′ end, for their ability to amplify a product from *ActX* cDNA (in pB293) at different annealing temperatures. Real time PCR systems using bilabelled TaqMan™ probes [Holland *et al*, 1991], or molecular beacons, or the TagMan method [Hobart, 2000] are other options for confirming

unique expression of *ActX* since the proprietory algorithms of these systems are able to resolve single nucleotide polymorphisms (SNP) accurately.

4. 2.3. Isolation of the 5′ flanking region of pine *ActX* by single primer PCR

The gene walking primers for SP-PCR were designed from the 5′ UTR of *ActX*, because of its low homology (<30%) to eukaryotic actins and <75% identity to the ortholog in closely related *Picea rubens*. In addition, locating the primers close to the unknown region also helps to maximise new sequence information.

One advantage of SP-PCR is that the gene-specific extension product can be sequenced directly from the PCR reaction; Screaton *et al* [1993] sequenced 4μl of a 25μl reaction with good results. However, when SP-PCR products in this study were sequenced they gave a poor read-out. It was thought that the concentration of the specific product in the PCR reaction, which could not be quantitated (prior to sequencing) because of product heterogeneity, was too low for sequencing. The bad sequence read-out can also be due to nonspecific annealing of the sequencing primer. However, subsequent experiments failed to substantiate these suspicions (Section 3.5.3).

This led to the speculation that the sequencing primer was confronted by extension products of several actin paralogs which contained complementary regions. In spite of the remarkable divergence in the the flanking regions of actin genes, several conserved sequences still exist. For example, in *A.thaliana* the nontranscribed region (NTR) of *Act2* and *Act8* share >11 blocks of identical sequences, each 4-22 bases long and *Act1* and *Act3* share >14 regions of 4-25 bases which show absolute sequence identity; similarly, the 5′ UTR contains >6 regions of 3-6 contiguous, identical bases, some of which are separated from one another by only 1-2 bases [An *et al*, 1996], thus providing a suitable template for mismatch pairing. It is possible that sequencing primer II was selected from a region that was fully or partially conserved in the pine actin paralogs.

The failure to sequence DNA extracted from excised bands, or band stabs (Section 3.5.3) may be attributed to fragment allelism (amplicons showing size homogeneity and sequence heterogeneity) [Reiter, 2001]. Some of the cloned inserts, which gave good sequences but had no overlap with the known 5′ UTR sequence of *ActX* could still be authentic extension products of *ActX*. The absence of an overlap is explained by the presence of an intron in the genomic insert, immediately preceding the sequencing primer. Almost all known angiosperm actin genes possess a leader intron in their 5′ UTR. If it is assumed that pine actin genes also harbour split 5′ UTRs and if the sequencing primer is located immediately downstream of the intron, then the proximal bases read during sequencing belong to those of the intron sequence which are probably different to those of the 5′ end of *ActX* cDNA. This would also explain the absence of actin signature sequences in the inserts.

Another possibility is that the insert sequences represent the 5′ flanking regions of actin pseudogenes. This would explain the absence of conserved TATA sites which occur 10-40 bp upstream of multiple transcription start sites and actin signature sequences such as pyrimidine-rich regions in the insert sequence.

4. 2.4. Southern analysis of pine *ActX*

In Southern hybridizations performed to verify the single-copy status of *ActX*, the 3′ UTR probe of ActX cDNA failed to produce any signals (Section 3.2.1). The restriction enzymes used to digest pine DNA, except *XbaI*, do not cleave within the 3′ UTR of *ActX* cDNA and therefore, all lanes of the blot should exhibit at least one hybridizing signal. One possible explanation is that *ActX* possesses an intron in its 3′ UTR. However, this is not the case since both pine genomic DNA and pB293 (containing *ActX* cDNA) when amplified with 3′ UTR primers gave an amplicon of the same size.

A *Pst*I fragment of *ActX* CDS was then validated as an informative probe. Because of its high conservation, the exon probe can be expected to hybridize to several genomic fragments in the blot. However, it too failed to produce any 'authentic' signals, possibly because the *Pst*I ORF fragment was a spliced product of two adjacent exons. McDowell *et al* (1996) proposed a conserved structure for angiosperm actin genes, which comprises four exons containing codons 1-20, 21-151, 151-355 and 356-377. The *Pst*I fragment is positioned approximately 550-750 bp downstream of the ATG codon and therefore, corresponds to codon positions ~183-250 which are clear of any intron sequences. However, it is still possible that conifer actins do not conform to the angiosperm actin gene structure and that the *Pst*I fragment of *ActX* is split by an intron.

4.3. Actin gene *Act2* of *A.thaliana*

In most eukaryotes, actin is transcribed by multigene families [Baird & Meagher, 1987; Druoin & Dover, 1990; Thangavelu *et al*; 1993; Meagher *et al*, 2000] whose CDS differs by only ~0.4-12.3% RNS (replacement nucleotide substitution [Li *et al*, 1986]). The *Act2* ORF likewise, shows 80-92% homology to plant orthologs. The intense signals observed in northern blots (Section 3.6.1) is most likely due to hybridization with transcripts of several paralogs of a putative pine actin multigene family.

Therefore, the 3′ UTR of *Act2* was used as a probe to determine specific *Act2* homolog expression. The 3′ UTR probe, which does not contain an intron, unlike the 5′ UTR, failed to hybridize with pine northern blots. Earlier, attempts to produce a pine-specific 3′ UTR probe from pine DNA using *Act2* 3′ UTR primers also failed to produce an amplicon. The failure of both gene-specific PCR and northerns to detect homologous regions in pine DNA and RNA, respectively, indicates that *Act2* and its putative pine ortholog share little sequence homology in their 3′ UTR. Subsequent BLAST homology searches showed that the *Act2* 3′ UTR indeed showed <30% homology to published plant actin sequences,

including those of the 10-member *A.thaliana* multigene family. it shows <20% homology even to the 3′ UTR of *Act8*, which encodes an almost identical protein to that of *Act2* with only one conservative amino acid difference [An *et al*, 1996]

Since exons are more conserved than UTRs, the *Act2* ORF was scanned for gene-specific regions to design primers for RT-PCR of the pine ortholog; the sequence 5′ aagattaaggtcgttgcaccacct 3′, 150 bp upstream of the TAA codon shows sequence divergence from most other *Arabidopsis* actin genes at every silent nucleotide position; however, a second gene-specific primer sequence could not be found. The reverse primer used in 3′ UTR amplification could be used as the second primer, provided it is conserved in pine. However, when pine DNA was amplified with the two primers they failed to produce an amplicon because either, or both primer sequences were not present in pine DNA.

4.4. T137

A genomic probe of *T137*, comprising its 5′ UTR and first exon [An *et al*, 1996] did not give any signals when hybridized with pine northern blots (Section 3.1.1). The 5′ UTR of *T137* spans a 128 bp intron, which may have looped out during hybridization to produce a hybrid product of high free energy (low stability) that may account for the absence of hybridization signals.

However, when its cDNA probe was substituted, it too failed to elicit signals in northern hybridization. *T137* shows only limited homology (~68%) to known orthologs. It is possible that angiosperm and gymnosperm *T137* homologs share low homology. Alternatively, the *T137* region selected as probe shows little complementarity to the pine ortholog; for instance, when the same primers were used to amplify a genomic probe from pine DNA, they failed to produce a PCR product which suggests that there are dissimilarities between the putative orthologs in the primer regions and possibly in the wider probe region. On the other hand, pine may not possess a *T137* homolog at all. These reasons provide an explanation for the failure of northern assays to produce visible signals.

4. 5. Ribosomal protein (RP) genes

The large subunit of a eukaryotic ribosome such as that of yeast contains a single copy of each of the 45 different L-RPs (apart from L7/L12) [Ramakrishnan & White, 1998], which play a vital role in stabilizing the tertiary structure of rRNA in the ribosome. Furthermore, the MsRL5 protein has strong homology to the L5 family of RPs [Asemoto *et al*, 1994], and *AtL18* to orthologs in several species [Baima *et al*, 1995] Therefore, it is perplexing that no signals were obtained when the RP probes were hybridized to pine northern blots (Section 3.1.1).

One possibility is that the pine orthologs of *MsRL5* and *AtL18* are cell-specific and therefore, their transcript titre was too low to be detected in tissue RNA. Stage- or cell-specific expression of RPs has been demonstrated in diverse species [Larkin *et al*, 1989; Zhan & Scoll, 1990; Adams *et al*, 1992; Stafstrom & Sussex, 1992; Bonham-Smith *et al*, 1992; Williams & Sussex, 1995] leading to speculation of functionally distinct ribosome subclasses [Woolford, 1991].

Another explanation is that the pine RP orthologs are distantly related to *AtL18* and *MsRL5*. Fewer homologies are observed among related RPs of eukaryotes [Asemoto *et al*, 1994; Baima *et al*, 1994; Goddameir *et al,* 1996] unlike among those of prokaryotes and archae [Woolford, 1991]. When coding regions are compared, members of an RP gene family are almost always highly conserved, but differ from orthologs in other species [Zhang & Scoll, 1990; Bonham Smith *et al*, 1992; Braun *et al*, 1993; Cooke *et al*, 1997].

4. 6. Gene for the second largest subunit of RNA Polymerase II (*Rpb2*)

It is difficult to explain why this essential gene whose protein product is specific to RNA Pol II (unlike the smaller Rpbs) was not detected in northern hybridization. The most likely reason is sequence heterogeneity between the probe and its target. There is little consistency among *Rpb2* homologs in their gene structure [Sitzler *et al*, 1991; Huang & Maraia, 2001], TATA box requirements [Falkenburg

et al, 1987; Larkin & Guilfoyle, 1993; Sakurai & Ishihami, 2001] and coding sequence [Falkenburg *et al*, 1987; Sitzler *et al*, 1991].

A BLAST similarity of *Rpb2* orthologs showed that most of the homology resides in the 3′ half of their CDS between positions 2600-3200 and for this reason, the *AtRP140* cDNA probe (position ~1500-3500 nt of the mRNA) of Larkin & Guilfoyle [1993] was used in northern assays. Probably, northern assays of pine *Rpb2* may have yielded the desired results had RE fragments from the 600 bp homologous region, or even oligonucleotides targeted to conserved *Rpb2* functional domains such as nucleotide binding, zinc finger and ribonuclease motifs [Ulmasov & Guilfoyle, 1992; Larkin & Guilfoyle, 1993] had been used as probes.

A second possible explanation is that pine *Rpb2* produces a low abundance mRNA. Microarray analysis of >95,000 genes and ESTs of the human, mouse and rat transcriptomes shows that transcript copy numbers of genes range from <0.1 to >100 per cell [Lockhart & Winzeler, 2000]. Most eukaryotic *Rpb* genes are single-copy genes [Ulmasov *et al*, 1996] whose mRNA levels may vary by several orders of magnitude [Iyer & Struhl, 1996] and bear no relation to their protein levels. Using quantitative competitive PCR, Sakurai & Ishihami [2001] showed that the transcript level of *Rpb2* of *S.pombe* was 40-fold higher than those of the low abundance *Rpb 1*, *3*, *7* and *9* genes. However, *Rpb2* has a low protein-to-mRNA ratio. Conversely, it is possible that a pine *Rpb2* of high translation efficiency escaped detection in northern assays because of its low abundance.

CHAPTER V

DISCUSSION

1. AN ENDOGENOUS PROMOTER FOR PINE TRANSFORMATION

Standard breeding methods are based on recombination and segregation of entire genomes and lack the specificity to transfer a single desired gene or trait alone. Direct gene transfer methods have the potential to do so and therefore have special appeal for the improvement of pine, which is a predominantly outbred species.

In plant tranformation, generally, a gene cassette containing a promoter and the cDNA of the desired gene (optimized for codon usage) is delivered to competent cells of the recipient. Under ideal conditions, the cassette integrates into the nuclear genome and is processed along with the other host genes to provide the desired gene product. The modular nature of expression cassettes allows the promoter and other gene attributes contained within it to be tailored to suit the cellular environment of the host species.

The cassette element which primarily determines transgene expression is the promoter; it need not be necessarily native to the introduced gene, or to the host species but most importantly, should be recognized by the host's transcription machinery. Several heterologous promoters such as the maize ADH promoter [Ellis *et al*, 1987], monocot ABA (absicic acid) inducible promoter, cereal ubiquitin promoter [Noury *et al*, 2000], recombinant pEMU promoter [Last *et al,* 1991] and the promoter of Cauliflower mosaic virus (CaMV 35S) [Benfey *et al*, 1990a] have been tested, or used in pine transformation for the past nine years in New Zealand.

CaMV 35S which has proven to be a very efficient promoter is expressed in almost all tissues of many plant species. However, a major concern with CaMV 35S promoter is its viral origin. Several pathogenic ssRNA viruses and virus-like particles have been isolated from needle washes of gymnosperms

and water effluents of pine [Castello *et al,* 2000]; they have not proved to be major pathogens of the crop probably because they are kept in check by host resistance mechanisms. The CaMV 35S promoter too, because of its pathogen origin is subject to host resistance mechanisms at any stage of the transgenic plant's growth [Napoli *et* al, 1990]; posttranscriptional gene silencing represents an ancient antiviral response, similar to RNA interference (RNAi) in animals, to protect plants from viral infection [Baulcombe, 2004; Ratcliff *et al,* 2001]. Likewise, all heterologous promoters used in pine transformation are also open to silencing mechanisms that operate in plants against invasive DNA [Henikoff & Comai, 1998; Matzke & Birchler, 2005]. Should similar plant responses operate in pine, it could result in inactivation of the heterologous promoters used in transformation [Walter, 1997], which would pose a major risk to a forestry industry based on transgenic pine. There is also the unanswered question on the longevity of heterologous promoters currently used in transgenic pines in New Zealand. Thus, a strong pine promoter capable of sustained expression is desirable for pine transgenic research.

A pine promoter on the other hand, because it is endogenous to pine has a lower probability of being inactivated in transformed pine, provided the site of insertion, copy number and orientation with respect to the resident gene are optimal. Secondly, the longevity of the selected pine promoter can be evaluated even before it is used in pine transformation (by assaying the expression of the gene in wild-type pine plants of different ages), thereby increasing confidence in the promoter. The aim of this research was to isolate desirable pine promoters for use in pine transformation.

2. CANDIDATE GENES FOR PROMOTER ISOLATION

Fishing for genes in a complex genome like that of *P. radiata* is facilitated by the availability of pine DNA/RNA sequences. More than 4000 nuclear sequences of *P. radiata* (Monterey pine) and *P.taeda* (Loblolly pine), mainly EST and some genomic and RNA sequences are found at present in Genbank accessions [Izquierdo *et al,* 1997; McCallum *et al,* 1997; Bishop *et*

al, 1998; Loopstra *et al,* 1998; Walden *et al,* 1999; Arthur, 2001; Jones *et al,* 2001]. At the commencement of this study (1996), however, very few pine sequences were available in the public domain. Although, marker assisted selection (MAS), paternity testing, clonal indexing, genetic engineering and phylogenetic studies are widely used in *P.radiata*, only a few short nuclear sequences such as small rDNA genes and simple sequence repeats had been published by 1997 [Gorman *et al,* 1991; Smith & Devey, 1994]. Even now, many sequences obtained by high-throughput sequencing at Agri-Genesis Corporation (New Zealand) and North Carolina State University (United States of America) remain locked in private databanks [Wood *et al,* 2000; Sederoff *et al*, 2000; Kirst *et al*, 2003; Egertsdotter *et al*, 2004].

The availability of even a few pine ESTs would have enabled an early start to screening of genes; if these genes later proved to be poor candidates, one of several computational methods [Marcotte *et al*, 1999; Overbeek *et al,* 1999] such as phylogenetic profiling [Pellegrini *et al*, 1999] can be used to predict associated genes for further screening. In the absence of informative pine sequence data, heterologous sequence information was used to screen the pine genome or its transcriptome for orthologs with desirable expression features. This was done in one of two ways; in the first line of investigation, a suitable candidate gene was first isolated from pine, its promoter was characterized and the expression profile of the promoter was verified. In the second approach, the expression profiles of putative pine orthologs of selected heterologous gene candidates were first verified and then promoters of the selected pine orthologs were isolated.

Both approaches have an even chance of detecting desirable genes, although the second alternative is a more prudent method. Great care was exercised in selecting candidate genes in order to ensure a high probability of success. Some of the heterologous candidate genes were single-copy genes expressed constitutively in other plant species; their selection was based on the premise that their orthologs in pine may also perform in the same way. A list of such genes known to be strong expressors in other plants was prepared from a literature survey; the list was not extensive since not many single-copy

genes had been studied for spatial expression. Candidate genes were also selected from putative multi-gene and multi-copy housekeeping genes in the belief that one or more of their paralogs will be strongly or constitutively expressed.

The pine 5S rDNA paralog *5Spr20* is an example of a gene whose promoter was identified using the first approach, where the gene was isolated first and the expression of its promoter confirmed subsequently.

3. *5Spr20* OF PINE

Several partial sequences of pine 5S rDNA were obtained by panning pine genomic DNA with immobilized wheat germ tRNA (WGT) ligand (Chapter II, section 12.4). This sequence information was used to amplify pine 5S rDNA paralogs, from which *5Spr20* was selected for detailed study. This class III gene is compact, possesses internalized promoters and is probably expressed in most cells and is an excellent candidate promoter for selective downregulation of pine genes. The application of class III promoters for such purposes was first demonstrated by Jennings and Molloy [1987] using the adenovirus VA I gene; the gene promoter linked to SV40 antisense sequences resulted in transient inhibition of SV40 replicon function by more than 50%.

5S rRNA genes are the most duplicated of all eukaryotic genes; ~24,000 copies of the gene are found in *Xenopus* and 2000 in *H.sapiens* [Long & Dawid, 1980]. The high gene number is probably indicative of the heavy demands placed on it for ribosome biogenesis. Paradoxically, many of these genes prove to be pseudogenes. For instance, nearly 1300-1600 of the ~2000 5S rDNA of *H.sapiens* are known to be pseudogenes, or variants of the functional gene [Sorensen & Fredriksen, 1991]. In the oocyte-specific 5S rRNA genes of *Xenopus laevis*, every repeat unit is made up of a functional gene alternating with a pseudogene [Long & Dawid, 1980]. Of the 50 unique 5S rDNA sequences identified in five pine species only seven were found to be functional [Liu *et al*, 2003]. The number of 5S rRNA genes in a haploid pine genome is believed to be upwards of 8000 [Moran *et al*, 1992] and are

distributed among six or more homogenous arrays. It is likely that many of these genes are non-functional or pseudogenes. It is therefore essential that *5Spr20* meets the minimal chemical and structural requirements of a functional 5S rDNA gene [Cloix *et al*, 2003; Liu *et al*, 2003].

3.1. Functionality of *5Spr20* gene

The feature that distinguishes functional genes from pseudogenes is their biological activity, in other words their ability to produce functional transcripts capable of folding into conserved secondary and tertiary structures [Liu *et al*, 2003], that provide the domains to interact with ribosomal protein(s) [Fromont-Racine *et al*, 2003] and also 'negotiate' between peptidyltransferase and elongation factor 2 (eEF-2) binding domains in the 60S ribosome subunit [Bogdanov *et al*, 1995]. Functionality can be tested by *in vitro* experiments, or deduced from the nucleotide sequence.

3.1.2. Transcriptional ability of *5Spr20*

5S rDNA genes belong to type I group of Pol III genes [Willis, 1993] which are characterized by three intragenic promoter elements occupying positions +50 to 64, +67 to 72 and +80 to 90 (Pieler *et al*, 1987). A comparison with ortholog sequences reveals that the 3′ half of *5Spr20* which harbours A-, I- and C-promoter boxes conforms to the consensus sequence of 5S rDNA derived by Erdmann *et al* [1985]. All consensus residues of the promoter and the segments separating them are conserved in *5Spr20* (Figure-5.1).

	A-Box	I-Box
5Spr20	agttaagcgcgcttgggctagagtagt	
Triticum[a]	agttaagc--gcttgggc-agagtagt	
Arabidopsis[b]	agttaagcg--cttgggc-agagtagt	
Consensus[c]	-SYYAA-C--------GS--RR-YAGU	
	C- Box	
5Spr20	actgggatgggtgacctcccggga	
Triticum	act-ggatgggtgacctcc-ggga	
Arabidopsis	act-ggatgggtgacctcccggga	
Consensus	ASY--GRUGGG-GACY-Y--GSGA	

Figure-5.1, 5S rDNA promoter sequences. Sequence from +50 to +101 of *5Spr20* compared with the consensus sequence and two plant 5S rDNA sequences. The A-,

I- and C-promoter boxes are underlined. Consensus residues are in blue letters. a=Appels *et al*, 1980. b=Cloix *et al*, 2002. c=Erdmann *et al*, 1985.

Since these internal promoters, which alone are sufficient for transcription, are highly conserved and since there are no deletions within the coding region (Chapter III, section 3.3.2) it is inferred that *5Spr20* is an active gene capable of initiating full-length transcripts.

3.1.3. Ability of *5Spr20* transcript to form secondary structure

The model for the universal secondary structure of 5S rRNA proposed by Luerhrsen & Fox [1981] has largely remained unaltered over the past 20 years; the molecule folds over itself to produce five helices (I-V) alternating with five single stranded elements which constitute two internal loops (B & E), two external loops (C & D) and one junction loop (A) that serves to connect helices I, II and V (Figure-5.2).

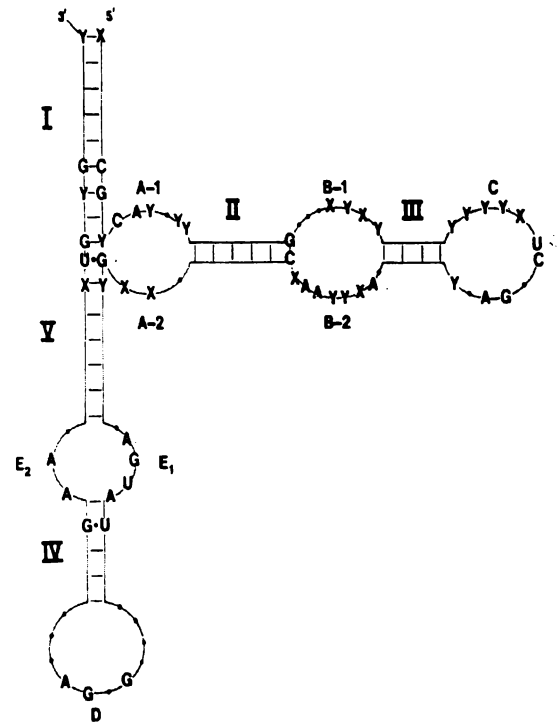


Figure-5.2. Secondary structure of 5S rRNA - The 5-helix model. Helical segments are labelled I, II, III, IV and V, loops as A, B, C, D and E, and non-contiguous single-strands forming a loop as -1 & -2 (adapted from Luerhrsen & Fox, 1981).

The higher order structure of 5S rRNA follows ineluctably from the primary

structure of the molecule. Using pairwise comparisons and comparative analysis of nine eubacterial and nine eukaryotic sequences, which included one plant sequence, MacKay *et al* [1982] proposed that several key elements are critical for eukaryotic 5S rRNA secondary structure. The derived RNA sequence of *5Spr20* contains almost all of these primary structural requirements. *5Spr20* possesses all but one of the absolutely conserved nucleotides, 13 of the 14 consensus Pu-Py residues and 11 of the 12 semiconserved nucleotides (Figure-5.3).

gggugcgaycauayyagcguuaaurcaccggaucycauyagaayuc
cgcauuuaagcgcgcuurrgcuagaguaguacugggaugrgugayc
ucccgggaaguycuagguygcacccuc

Figure-5.3. Secondary structure elements of 5Spr20 RNA. red letters denote conserved residues, blue letters consensus nucleotides and italics deviations.

Later, Barciszewska *et al* [1985] proposed a slightly different secondary structure model for plant 5S rRNA based on 12 plant 5S rRNA sequences. *5Spr20* contains 39 of the 40 absolutely conserved residues in the 5-helix secondary structure model of plant 5S rRNA (Figure-5.4).

	H (I)	S (A1)	H (II)	S (B1)	H (III)
5Spr20	gggugcgau	caua	ccagcgu	uaaucc	accgga
Consensus[1]	-g-ugcgau	caua	ccagcac	ua--gc	accgga

	S (C)	H (III)	S (B2)	H (II)
5Spr20	ucccauuagaac	uccgcauu	uaagc	gcgcuugg
Consensus	ucccaucagaac	uccg-agu	uaa-c	gugcuugg

	H (V)	S (E1)	H (IV)	S (D)
5Spr20	gcuagagu	agua	cugggaugg	guga
Consensus	gcgagag-	agua	cua-gaugg	guga

	H (IV)	S (E2)	H (V)	H (I)
5Spr20	ccucccgg	gaa	guccuagu	guugcacccuc
Consensus	cc-c-ugg	gaa	guccu-gu	guugca-----

Figure-5.4. Secondary structure elements of *5Spr20*. Red letters denote conserved residues, other letters consensus nucleotides, dashes variable positions and italics

deviations. H=helical region, S=single stranded region. consensus[1] = Derived by Barciszewska *et al* [1985].

The D-loop residues, which are conserved in *5Spr20* are thought to be of major functional importance. Cross-linking experiments suggest direct interaction between residues in the D-loop of 5S rRNA and two important regions of 25S rRNA, namely, the GTPase associated region involved in EF-G binding and the peptidyltransferease center [Dechampesme *et al*, 1999].

Since the derived RNA sequence of *5Spr20* contains all the structural features absolutely required for folding into the five-helix model it is inferred that the pine paralog is a functional gene.

5S rRNA structures tend to be identical within plant genera and show divergences between distantly related taxa. For example, 5S rRNA of *Graminae* are identical but differ from those of *Solanacea* at two positions [Barciszewska *et al*, 1985]. Similarly, *5Spr20* differs from the consensus plant sequence at seven positions, five of which are in the helical regions. Changes within helices are offset by compensating changes [Zhang *et al*, 2003] and by posttranscriptional editing [Szymanski *et al*, 1995] so that the universal folding arrangement is conserved.

3.2. Antisense potential of *5Spr20*

The antisense vector p5A containing the *5Spr20* promoter-*anti gus* cassette produced efficient suppression of GUS activity when cotransformed with expression vector pGUS (containing duplicate CaMV 35S promoter-*gus*construct); it downregulated GUS expression by as much as 90% (Chapter III, table-3.2), thereby demonstrating the potential of *5Spr20* promoter for pine antisense work.

The antisense effect is in realty a product of two interlinked factors, namely, the promoter which is the prerequisite for transcript formation and the antisense transcript which effects downregulation, each of which can be improved independent of the other to increase their gene downregulation efficiency.

3.2.1. Improving *5Spr20* promoter efficiency

The effectiveness of the *5Spr20* promoter component can be enhanced in several ways. Promoter efficiency may be increased by using multiple tandem copies of *5Spr20* promoter in expression cassettes, similar to the use of duplicated CaMV 35S promoter in pGUS used as the control in this study. An alternative approach would be to construct a composite synthetic promoter comprising pine 5S rDNA elements known to activate transcription. Similar constructs have been engineered for constitutive expression in monocots [Last *et al,* 1991] and phloem-specific expression in dicots [Heifetz *et al,* 2000]; the recombinant pEMU monocot promoter [Last *et al,* 1991] consists of a truncated alcohol dehydrogenase gene (*Adh1*) promoter to which six copies of the maize anaerobic responsive element (ARE) and four copies of *Agrobacterium tumifaciens* octopine synthase enhancer (OCS) have been added. The promoter gave a 10- to 50-fold higher expression than CaMV 35S promoter.

A consistent property of most strong/constitutive promoters, as exemplified by the CaMV 35S promoter (Chapter I, section 4.2.4) and the cassava vein mosaic virus [Verdaguer *et al,* 1998] is that their sequence consists of several *cis*-acting elements which act combinatorially and synergistically to produce activated transcription. Activated transcription in Type-I, Class III genes such as 5S rDNA is not as well understood as in Class II genes [Carey, 1998]. However, the ease with which polymerase specificity of some genes can be switched from Pol II to Pol III [Lobo & Hernandez, 1989], similarities in composition of the two polymerase enzymes [Margottini *et al*, 1991; Woychik *et al*, 1993; Hernandez, 1993; Roeder, 1996], polyadenylation of some naturally occuring Pol III transcripts [Singh *et al,* 1985; Kramerov *et al*, 1990] and initiation of transcription at some promoters from a common start site by both polymerases [Bentley *et al*, 1989; Piras *et al*, 1994] all seem to suggest that specificity of pathways of the two gene classes may not be rigorous; models of holoenzymes and enhanceosomes developed for the more intensively studied Pol II genes may also apply to 5S rRNA genes [Huang & Maraia, 2001].

Putative activator binding elements in the 5′ flanking region of *5Spr20* can be identified by functional studies of promoter deletion constructs, or by sequence analyses. An examination of *5Spr20* spacer sequence (Chapter III, Section 3.3.3.7) failed to reveal any motifs proposed for 5S rDNA genes [Sharp & Garcia, 1988; Bredow *et al*, 1990; Venkateswarlu *et al*, 1991], or canonical *cis*-acting elements of class I & II genes that have been determined by functional studies. Sequence analysis using the computer software FASTFAC [Quandt *et al*, 1995] against a database of transcription factor binding sites also failed to show any credible putative regulatory regions, although two CTTC motifs at positions – 420 and -417, similar to the GCRI binding site (RGCTTCCWC) of yeast [Huie *et al*, 1992], a CTCGT motif at position -573, resembling the upstream stimulatory factor (USF) binding site (NCACGTGN) of vertebrates and an AGGGGG motif at position – 488 homologous to the stress response element (STRE) of yeast (TMAGGGGN) [Schueller *et al,* 1992] were identified. Several TF-DNA binding sites listed in bioinformatics databases do not have a complete binding energy profile [Kotelnikova *et al*, 2004] and hence the above motifs need to be verified by footprinting and mutation analyses.

These partially homologous putative regulatory elements and other repeat elements (Chapter III, section 3.3.3.3) and several conserved sequences in the flanking region of Pine 5S rDNA paralogs (Chapter III, sections 4.2.2.1) offer exciting prospects for functional studies to define activator binding sites in *5Spr20*, which can be incorporated into a synthetic promoter construct. There was provision in this study to test only one promoter construct and in one type of pine tissue in transient assays (Chapter III, section 4.3.2). The recombinant *5Spr20* promoter used consists of the CDS containing the minimal promoters and the region immediately upstream of it, since the 47 nt sequence 5′ to the CDS was regarded as a putative *cis*-acting element because of its high conservation (Chapter III, section 4.3.2). However, this may not be the case and omission of the upstream region from p5A may even increase its antisense efficiency.

Future work should be directed at identifying the upstream regulatory elements of *5Spr20* which modulate transcription. Promoter deletion by *Bal31-*, or restriction digestion, or point mutant constructs produced by PCR can be evaluated in transient expression studies using both embryogenic (undifferentiated) and organogenic (differentiated) tissues. This will not only reveal the strength of the mutant promotors, but also their pattern of constitutive/spatial expression.

3.2.2. Improving antisense RNA strand performance

The second component of the antisense expression cassette which determines its success is the antisense strand. The principle underlying antisensing has not been established sufficiently to enable a common strategy and therefore the antisense component has to be found empirically, similar to other homology-dependent gene silencing strategies, on a case by case basis with regard to its sequence, length, dosage, propensity for secondary structure formation, accessibility to target site and cross-interference with nontargeted genes [Bosher & Labouesse, 2000]. Branch [1998] suggests that shorter antisense strands are preferred in order to reduce detrimental effects which may occur through unintended reaction with nontarget regions; but the antisense sequence can not be less than 17 nt since any sequence that is <17 nt long will have a high probability of occurring more than once in the pine genome; it is unlikely that a 17 nt sequence will find another perfectly matched template($1/4^{17}$ = $1.7x10^{-10}$) in the pine genome (1C=$1.1x10^{10}$bp), assuming random distribution of bases.

3.3 Ability of *5Spr20* promoter to generate translatable RNA

Eukaryotic mRNA is believed to be synthesized *in vivo* exclusively by Pol II. Such specialization with respect to RNA products and functional limitations may have arisen as a result of specific characteristics of polymerses such as intranuclear compartmentalization of Pol I within the nucleolus which denies mRNA processing proteins access to transcripts and the tendency of Pol III to terminate transcription in runs of thymidine residues, or limitations imposed by

transcript properties such as the m^7GpppN 5′ terminal cap, or mere perpetuation of some archaic properties.

Several authors have investigated the production of functional polypeptides encoded by polymerases other than Pol II. Fleischer and Grummt [1983] demonstrated competence of Pol I transcripts to support α-amanitin insensitive protein synthesis. Their findings were questioned by Lopata *et al* [1986] who found that transcripts initiated at the Pol I initiation site were not polysome attached, suggesting that protein expression was due to a cryptic Pol II promoter. Similarly, functional β-globin and thymidine kinase transcripts were generated by pol III [Carlson & Ross, 1984; Lewis & Manley, 1986] but the conclusions were held in dispute by Sisodia *et al* [1987]. More recently, Gunnery and Mathews [1995] used Pol III promoter of adenovirus type-2 VA RNA_I gene to generate polysome associated HIV type-1 *tat* transcripts which gave rise to TAT proteins, *in vivo*.

The expression vector p5s containing the *gus* ORF preceded by a consensus ribosome binding context and placed downstream of the *5Spr20* promoter produced very low levels of GUS expression compared to pGUS (Chapter III, table-3.2). The results, though were not totally unexpected.

But, high transcript levels accompanied by low protein expression indicates inefficient post-transcriptional processing of *5Spr20* sense transcripts. In the absence of corroborating evidence, we can only speculate on possible reasons.

3.3.1. Nucleolar localization of *5Spr20-gus* transcripts

Studies on ribosome biogenesis show that soon after synthesis, 5S rRNA transcripts in the nucleoplasm become transiently bound to La antigen, which acts in the termination of Pol III transcripts and then to a large (L) ribosomal protein (L5 in mammals and L1 in yeast) [Fromont-Racine *et al*, 2003]. This 5S RNP complex ('third ribosomal subunit'), directed by the nucleolar targeting signals (nts) located at the N- and C-termini of L5 migrates to the nucleolus [Rosorius *et al*, 2000], the site of major rRNA synthesis and

ribosomal subunit assembly where it is incorporated into the 60S subunit. This association is the signal for export of the ribosome subunit to the cytoplasm [Deshmuk *et al*, 1993; Marai, 2001].

Transcripts encoded by p5S are chimeras consisting of 5S coding region (5S rRNA) followed by *gus* RNA. It is speculated that the *gus* segment of the transcript (which includes the terminator-UUU extension of the primary transcript which provides the signal for La binding) complexes with La/ribosomal protein and is transported to the nucleolus. The chimeric 5S rRNA, however, is unable to assume the precise folded position in the ribosome because of steric hindrance of the *gus* sequence. Presumably, the aberrant folding prevents nuclear export signals (NES) such as the those which reside in the central region of L5 [Rosorius *et al,* 2000] from being recognized by putative exportins [Mattaj & Englmeier, 1998], or prevents the nuclear import signals of one or several ribosomal proteins from being covered up on assembly of the ribosomal subunit to prevent their recognition by the import machinery; this would result in the transcripts being confined to the nucleus and subsequently these may become the target of RNA surveillance mechanisms similar to nonsense mediated mRNA decay [Nazarenus *et al*, 2005]. Examples of nuclear localized transcripts are found in a *hox-1.3* homeobox and a p55 *Il-2* receptor gene, whose transcripts have an AUG-burdened 5′ UTR [*in* Kozak, 1989].

The majority of the sense-*gus* transcripts amplified from p5S transformed cells by RT-PCR (Chapter III, figure-3.30) are probably the nuclear confined *gus*, since RT-PCR does not distinguish between cytoplasmic and nuclear transcripts (unlike nuclear run off assays). The low, but significant GUS expression in p5S treatment (Chapter III, table-3.2) can be explained by 'leaky' export of transcripts to the cytoplasm. The Pol III encoded *tat* transcripts of Gunnery and Mathews [1995] which gave rise to TAT proteins are also chimeric structures (containing the adenovirus VA RNA forward of the *tat* sequence) but they differ from the *5Spr20-gus* chimera in that their export to the cytoplasm is not ribosome-dependent.

3.3.2. Inhibition of translation initiation

Alternatively, the chimeric transcript may have been exported to the cytoplasm, in a manner analogous to the nucleocytoplasmic transit of excess 5S rRNA transcripts in non-ribosomal 7S RNP particles during vitellogenesis in *Xenopus* oocytes [Dechampesme *et al*, 1999], but was poorly translated. According to the linear scanning model of translation initiation [Kozak, 2004], the 43S initiation complex which includes the small ribosomal subunit binds at, or near the 5′ cap end of mRNA and scans the sequence for the first AUG codon in an optimal context ('Kozak sequence') where it is joined by the 60S subunit to initiate translation. In 90% of eukaryotic mRNAs examined the AUG closest to the 5′m^7G cap is the initiating AUG [Merrick, 2004]. However, the model allows for ribosomes to reinitiate translation at a downstream AUG codon of a second ORF [Putterill & Gardner, 1989; Merrick, 2004] or by backscanning upto 50 nt [Peabody *et al,* 1986] subsequent to initial translation termination.

Several features in the chimeric *gus* transcript can inhibit cap-dependent translation, which accounts for 75-90% of eukaryotic translation initiation [Merrick, 2004]. Absence of a cap structure and the presence of strong secondary formation in its GC rich 5′ UTR (which in this case is the 5S rRNA - see Figure-5.2) mitigate vectorial scanning by ribosomes [Kozak, 1991]; the scanning mechanism is able to unwind secondary structures of low stability (-30 kcal mol^{-1}) but is blocked by more stable structures (-60 kcal mol^{-1}) [Gunnery *et al*, 1997]. In addition, ribosome binding at the ribosomal binding site just upstream of *gus* ORF could be negatively affected by two AUG codons present in the 5′ UTR (at positions +23 and +83 of 5S rRNA); Gunnery and Matthews [1995] have suggested that ribosomes may bind to uncapped mRNA in a random fashion and migrate to the nearest AUG where they initiate translation.

Alternative models for ribosome binding [Sonnenberg, 1993] (Figure-5.5), which may not necessarily be mutually exclusive for a given mRNA allow for ribosome bypass ('shunt') of a segment of the 5′ UTR [Futterer *et al*, 1993], or cap-independent internal ribosome entry (IRE) [Oh *et al*, 1992; Kozak 2003].

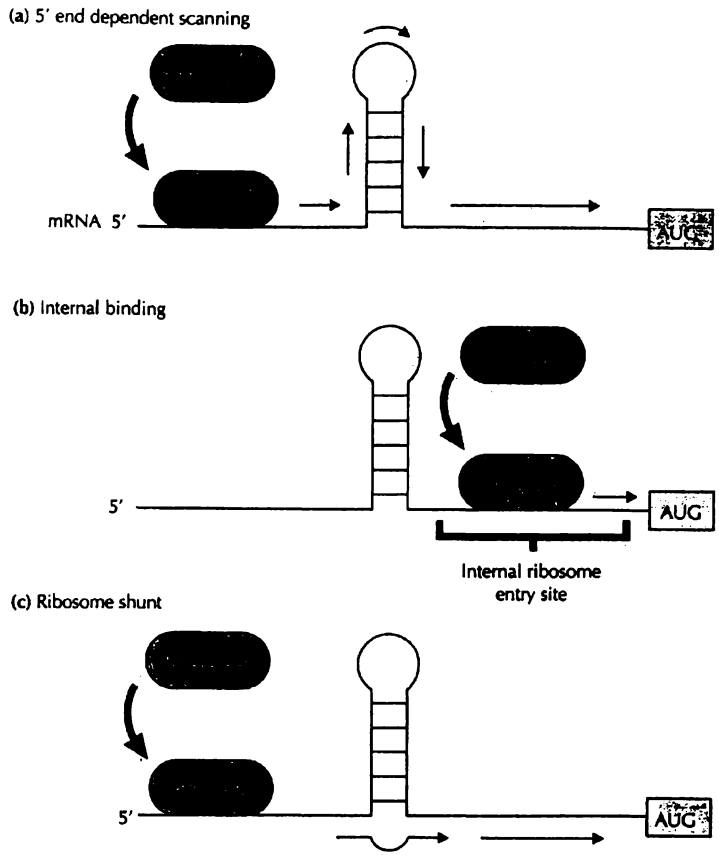


Figure-5.5. Alternative models for ribosomal binding. A. 5' end dependant scanning B. Internal binding C. Ribosome shunt (adapted from Sonnenburg [1993]

Evidently, these mechanisms which help ribosomes to skip the extensive secondary structure of the 5′ leader sequence, or bind directly to the ribosomal binding context of *gus* did not operate in the case of *5S-gus* transcripts.

Although only two possible reasons have been discussed for GUS attenuation in p5S treatment, several other mechanisms such as repressor binding at the 5′ UTR [Klausner, 1993], lack of CAP and 3′ sequence synergy required for closed-loop initiation [Gallie, 1993; Kozak, 2004], nonbinding of mRNA binding proteins [Svitkin, 1996], competition for limited initiation factors such as eLF-4E [Merrick, 2004] and controls at various post-translational stages may have contributed to low GUS expression.

3.4. *5Spr20*-driven gene downregulation in stable transformants

The results of stable expression studies appear to confirm the findings of transient expression studies and suggest that *5Spr20* promoter is capable of driving sustained and efficient gene downregulation when stably transformed in plants. These observations are, however, preliminary since the interpretation of stable expression results is complicated by factors such as transgene copy number, gene orientation, expression cassette integrity and transgene insertion site. Nevertheless, the results augur well for a role for *5Spr20* promoter in gene downregulation applications. What is required though, are more intensive trials where the silencing effect is studied in relation to the factors mentioned above so that an accurate conclusion can be drawn. Secondly, only eight transformants were studied in the present study. A bigger population should be analysed to improve confidence in the results. And most importantly, the accumulation of target-specific siRNA in silenced plants, the hallmark of PTGS/RNAi silencing [Zhou et al, 2004] should be demonstrated.

The stable expression experiment also shows that shRNA constructs, used exclusively in mammalian systems, perform equally well in plants. The absolute silencing obtained in *N. benthamiana* with the *5Spr20*-shRNA construct is similar to the total silencing of GFP obtained with a Pol III H1 promoter driven shRNA silencing cassette in transgenic mice and rats [Hasuwa *et al*, 2002]. Likewise, the silencing effect of *5Spr20*-shRNA construct is on par with that of CaMV 35S-driven ihpRNA constructs targeting PVY infection in tobacco [Smith *et al*, 2000], ethylene signalling (*EIN2*) and flower repression (*FLC1*) genes in *Arabidopsis* [Wesley *et al*, 2001] and *gus* in tobacco [Lo *et al*, 2005]. The only drawback is that shRNA constructs do not present several putative siRNA targets as do the longer hpRNA constructs. However, this is over-ridden by the fact that shRNA construction requires less time and is far more easier than standard ihpRNA constructs.

It would be useful to know whether the silencing caused by *5Spr20*-shRNA was due to siRNA directed-mRNA cleavage, translational suppression [Lo *et*

al, 2005] or methylation [Jones *et al,* 2001;Tang, 2005], or a combination of all.

The *5Spr20* promoter is eminently suited for two areas of pine research. Firstly, it can be used to downregulate specific genes/traits which results in improved crop performance or value-added produce; several areas of pine physiology research such as flower induction and lignin biosynthesis are prime targets for gene silencing. Pine uses about 10% of its carbon resources for cone (flower) production [Walden *et* al, 1997]; this energy can be diverted to vegetative growth by downregulating flower production. Flowering is determined by the expression of several floral meristem-identity genes and inactivation or co-suppression (by overexpression) of these genes can reverse these developmental switches [Busch *et al*, 1999]. Weigel & Nilsson [1995] used CaMV 35S promoter to ectopically express the primary flower-meristem-identity gene *lfy* (leafy) in diverse plants; strong and constitutive expression of the gene resulted in precocious flowering through overexpression. Such developmental switches are amenable to transcriptional silencing since naturally occuring *lfy* mutants result in loss-of-function phenotypes (Huala & Sussex, 1992). Similarly, pine timber quality which is determined by the ratio of syringyl:coniferyl monomers in lignin can be manipulated to produce designer-wood, by altering enzyme activity in the phenyl-propanoid pathway.

The biggest impact of *5Spr20* promoter on pine research will be in the area of functional genomics. The ultimate goal of current genome projects is to identify the biological function of every gene in the genome. Reverse genetics has emerged as the method of choice to elucidate the role of specific proteins in an organism. Loss-of-function studies have employed a plethora of techniques such as homologous recombination, insertion and transposon mutation, antisensing, cosuppression and chimeraplasty. Homologous recombination gives predictable targeted gene replacement in prokaryotes, moss and yeast, but is not suitable for plants where DNA integrates mainly via illegitimate recombination in an undirected sequence independent manner [Puchta, 2002]. Insertional mutations and transposable elements are used in

large-scale analyses of yeast genome [Wesley et al, 2001]. However, they require vast resources; in order to have a 90% chance of finding just one specific gene of 1 kb in *Arabidopsis* using T-DNA insertional mutagenesis would require the generation of about 350,000 independent transformants [Krysan *et al*, 1999]. Antisensing has the dubious distinction of being inconsistent [Helliwell & Waterhouse, 2003] and antisense and cosuppression constructs usually result in only a modest proportion of silenced individuals [Wesley et al, 2001].

In this respect, *5Spr20* promoter-driven PTGS/RNAi-mediated targeted gene knockouts hold great promise for the study of pine gene function by loss-of function. It provides consistent and very high-level of gene silencing, while still allowing a low but detectable heterogenous level of silencing within the transformed population which is useful in gene function studies [Wesley et al, 2001]. In *Caenorhabditis elegans* RNAi has led the way in functional genomics of virtually all of its 19,000 genes [Dykxhoorn, *et al*, 2003].

Thus a *5Spr20* promoter-shRNA construct will prove to be a useful tool in basic and applied pine research.

4. SCREENING FOR PINE ORTHOLOGS OF OTHER CANDIDATE GENES

4.1. Expression profile validation

A different rationale was adopted to investigate candidate genes *tDNA* $^{Met\text{-}I}$, *ActX, Act2*, the two ribosomal protein genes, *T137* and RNA Pol II subunit gene (Chapter IV, section 1.1). In this instance, the expression profiles of putative pine orthologs were validated before attempting to clone their genes and promoters. Northern hybridisation was used in preference to RT-PCR in expression screening on account of likely sequence differences between candidate genes and their pine orthologs.

None of the single-copy candidate genes had published orthologs in plant species and sequence homologies with orthologs in other eukaryotic groups ranged from 70% to less than 50%. For instance, the only ortholog sequences available for the candidate gene *AtL18* were *L18* sequences of rat, human

and *Xenopus*; a multiple alignment of these sequences provide at best, primers (21nt) with a 48-fold degeneracy for RT-PCR. Under such circumstances, Northern probes are better suited to detect micro-homologies than the shorter primers used in RT-PCR even though the latter assay is a 1000-fold more sensitive than Northern assays [Veres *et al*, 1987; Hamoui *et al*, 1994].

In the case of multi-copy candidate genes such as *tDNA* $^{Met\text{-}I}$ and actin, sufficient sequence information exists which enabled conserved or gene-specific probes or primers to be used in Northern and RT-PCR analyses.

Several reasons can be advanced to explain why the expression pattern of several pine orthologs could not be determined using heterologous probes/primers of single-copy candidate genes; sequence hetrogeneity of the probes (Chapter IV, sections 4.4, 4.5 & 4.6), primer selection from nonhomologous regions (Chapter IV, sections 4.4 & 4.6), cell-specific expression (Chapter IV, section 4.5) and low transcript copy number (Chapter IV, sections 4.5 & 4.6) are some of the possible reasons.

4.3. Gene isolation

Following expression profile validation, two candidate genes namely, *tDNA* $^{Met\text{-}I}$ and *ActX* were selected for promoter isolation (Chapter IV, sections 3.4.2., 3.5.3.), which can be approached in one of several ways. Library construction is the method of choice. However, PCR-based methods were preferred for several reasons.

4.3.1. Genomic/subgenomic libraries

The traditional method of gene isolation is by library construction where the genome is subcloned as discrete fragments in individual bacterial or bacteriophage clones and screened for genes, using specific or highly degenerate probes/primers. This method was not considered in view of the complexity of pine genome.

The ability of a genomic library to represent the entire genome depends primarily on the size of the unamplified library. Assuming that all genomic fragments can be cloned in the library vector, the minimum library titre (N) required to represent a cloned genome is given by,

$$N = \text{Log } (1-p)/\text{Log } (1-I/G)$$

where p = probability, I = genomic insert size and G = genome size [Clark & Carbon, 1976]. A haploid copy of pine genome is 11.1 pg [Miksche, 1985] or,

$$(11.1 \times 10^{-12} g)\ (6 \times 10^{23})/660 \text{ da} = \sim 10^{-10} \text{ bp}$$

Using a standard λ phage vector and genomic inserts of ~20 kb, a library titre of 2.4×10^6 pfu is required to capture a single-copy gene with 99% probability. Construction of a library of this size is not an easy task and screening of this library by PCR or hybridization would be time-consuming. If a library was available in the public domain it could have been used for gene isolation, but, no research laboratory has so far succeeded in creating a quality pine library [Gorman *et al*, 1992].

Genomic libraries are also beset with the problem of unclonable fragments so that a high titre library may still not contain a gene of interest. In addition, genes captured in slow growing library clones can be lost during library amplification.

A major consideration in not opting for a genomic library was because the only nucleic acid detection system available was the DIG system (see below), since the premises where the research was conducted for the first two years did not have a radioactive-user license. For these reasons the genomic library option was not considered.

A subgenomic library constructed from enriched genomic fragments may help to circumvent the need for library titres of the order of 10^6 pfu. But, this option was also not taken because of the sensitivity of the DNA labelling and detection system. The DIG detection system did not appear to be particularly sensitive for pine Southern hybridisation (and maybe for DNA hybridisation in

general); when a pine Southern blot which hybridized very strongly to a ^{32}P-labeled RFLP probe in a different laboratory was probed with the same probe labelled with DIG, in our hands, the signals obtained were extremely weak.

4.3.2. PCR-based methods

PCR-based methods have practically eliminated the need for library construction in gene expression studies [Frohman *et al*, 1988; Ohara *et al*, 1989]. Similarly, several variants of PCR have been used succesfully to capture promoter sequences and other regulatory elements from unknown regions ranging from a few 100 bases [Rudi *et al*, 1999] to several kilobases [Jones & Winistorfer, 1997] flanking the transcribed region of genes, eg. Inverse PCR [Ochman *et al*, 1988], ligation-mediated single-sided PCR [Fors *et al,* 1990], biotin-RAGE PCR [Bloomquist *et al*, 1992], targeted inverted repeat amplification [Jones & Winistorfer, 1993], supported PCR [Rudenko *et al*, 1993], adapter-ligated PCR [Seibert *et al*, 1995], RAGE [Cormack & Somssich, 1997], panhandle PCR [Jones & Winistorfer, 1997], multiplex restriction site PCR [Weber *et al*, 1998] or restriction cutting independent PCR [Rudi *et al*, 1999],

Because of its sensitivity, PCR is well suited for investigating complex genomes. Enrichment of amplicons using biotin-streptavidin bridges, nested primers and anchored primers often produce the target molecule in a form suitable for direct sequencing. Therefore, PCR-based methods were used for walking into the promoter regions of pine *tDNA^{met-i}* and *ActX*.

5. *tDNA^{met-i}*

In the case of pine actin gene *ActX* and the class III pine gene *tDNA^{Met-i}*, both of which probably belong to putative multigene families, the promoter region could not be recovered by any of the PCR methods tested. The experience with *tDNA^{Met-i}* exemplifies the complications in applying PCR-based gene walking methods to genes of multigene superfamilies whose known coding region is highly conserved.

tDNA^{met-i} was selected as a candidate gene for two reasons. Firstly, because it is constitutively expressed in all plant tissues [Sharp *et al*, 1985] where its transcripts are required in the initiation complex for binding of the smaller ribosomal subunit to mRNA; aminoacyl-tRNA^{met-i} is the only species that can enter the partial P site of 40S ribosomal subunit and recognize the AUG codon to initiate eukaryote translation. Secondly, antisense coding sequences attached to and transcribed by tDNA^{met-i} exert a strong suppressor effect as was demonstrated in antisense experiments of Bourque and Folk [1992] who showed that a soybean tDNA^{met-i} promoter elicited a 5-fold higher reduction of reporter gene expression than CaMV 35*S* promoter in carrot cells. Similarly, a chimaeric human *tDNA*$^{Met-i}$ fused to antisense templates inhibited replication of Moloney murine leukemia virus by 97% in mouse cells [Sullenger *et al*, 1990a] and HIV-1 virus by 99% in human cells [Sullenger *et al*, 1990b]. Further Choisne *et al* [1997] demonstrated that tDNAs producing suppressor tRNAs were very efficient in transactivating genes containing premature stop codons; a *gus* gene inactivated by a premature amber stop codon in transgenic tobacco was activated by introduction of a modified tDNA(Leu) expressing a strong amber suppressor tRNA.

5.1. Promoter isolation by PCR

Pine *tDNA*$^{met-i}$ was amplified using primers from the highly conserved 5′ and 3′ ends of the gene (Chapter IV, Section 3.4.2) and was found to be identical to homologs in other species [Palmer & Folk, 1987; Akama & Tanifuji, 1989]. The amplicon contains the internalized split-promoters, and these alone are sufficient for transcription [Ciliberto *et al*, 1982] and therefore could be tested directly in transformation studies.

However, the 5′ flanking region of many tDNAs is known to exert a modulatory effect on transcription and its re-initiation [DeFranco *et al*, 1981; Larson *et al*, 1983; Schaack *et al*, 1984; Campbell *et al*, 1985; Arnold & Gross, 1987; Yukawa *et al,* 2000]. Ulmasov & Folk [1995] using point mutations and deletions showed that replacing the naturally occuring 5′ upstream sequences negatively modulated *tDNA*^{Trp} expression. The succesful antisense

suppression of CAT activity using *tDNA*$^{Met-i}$ promoter by Bourque and Folk [1992] was probably due to the inclusion of 580 bp of 5′ flanking region with tRNA^{Met-i} gene in the expression cassette. Thus, the composition of a *tDNA*$^{met-i}$ 'promoter' would depend on whether its upstream region is required for activated transcription.

The PCR-based methods used to isolate the upstream region of pine *tDNA*$^{met-i}$ (Chapter IV, sections 3.4.3.1-3.4.3.4) rely on a gene-specific ('sequencing' or 'specific') primer located in the 3' half of gene to read into (as in SP-PCR) or amplify (as in RAGE and TAIL-PCR) the coding strand past the known 5′ end of the gene. The sense strands of these extension products extend from a point upstream of the gene, primed by a vector-/arbitrary-/internal primer to the 3′ half of the gene (or even beyond, if the internal primer in SP-PCR binds in the downstream region). The gene-specific primer was intentionally positioned 36 bases internal to the 5′ end of the gene to provide a sequence overlap between authentic extension products and the gene. None of the sequenced amplicons possessed this overlap region.

If pine *tDNA*$^{met-i}$ is a putative single-copy gene at least some, if not most, of the extension amplicons barring those produced by false priming sites (in SP-PCR) or by nonspecific amplification (TAIL-PCR) should give sequence readouts extending from the overlap region into the unknown 5′ region. The absence of authentic PCR 5′ extension amplicons suggests that the sequencing primer was overwhelmed by extension sequences bearing partial homology to pine *tDNA*$^{met-i}$. This is supported by observations made during SP-PCR optimization (Chapter IV, section 4.3.1) where amplification occured freely at annealing temperatures of <52°C which is in contrast to the findings of Screaton *et al* [1993]. Therefore, it is possible that there are many *tDNA*$^{met-I}$ or *tDNA*$^{met-i}$-like genes in the pine genome.

The tDNA genes are one of the most redundant of nuclear genes; their RNAs which act as the aminoacyl carriers of 20 standard amino acids and other rare ones (desmodsine, etc) are represented by >1300 tDNA genes in *H.sapiens*

and >1150 genes in *Xenopus* [Long & Dawid, 1980]. In most plants, tRNA genes exist as multigene families [Beier *et al*, 1991]. Sequences of several tRNA-hybridizing genomic regions reveal the occurance of pseudogenes (denoted by the symbol Ψ) intermingled with authentic genes.

A likely scenario for pine *tDNA*$^{met\text{-}i}$ is that it constitutes a multigene family where single- or oligo-copy functional genes exist along with a large number of pseudogenes in a situation analogous to the ~2000 human 5S rRNA genes (another class III gene) where three out of every four genes are pseudogenes [Sorenson & Fredriksen, 1991]. Pseudogenes may consist of either a partial sequence of a corresponding tRNA structure, or a complete *tDNA* sequence showing sequence heterogeneity [Sharp *et al*, 1981; Koski & Clarkson, 1982]. Green and Weil [1989] identified a cluster of four potentially functional *tDNA*^{Pro} genes and two pseudogenes (Ψ1 and Ψ2) in a 5.5 kb fragment of bean nuclear DNA; Ψ2 lacked the entire 5′ half (nucleotides 1-36) of functional genes and contained several nucleotide substitutions and deletions. Ψ tDNA^{Tyr} of *Nicotiana rustica* on the other hand, corresponds to the 5′ 37 nucleotides of the coding sequence and contains three point deletions and one substitution [Fuchs *et al*, 1992]. In other instances, differences between pseudogenes and active genes are subtle; Ψ tDNA^{Leu} of beans differs from wild-type gene by a single T insertion in the B-box internal promoter [Ramamonjisoa *et al*, 1998].

It is postulated that most of the pine *tDNA*$^{met\text{-}i}$ pseudogenes probably show greater complementarity to the 3′- rather than 5′ end of the functional gene and priming by the internal primer at several 3′ ends masks the identity of the functional gene. When primers (forward and reverse) complementary to 5' and 3' ends of pine tDNA$^{met\text{-}i}$ are used together in standard PCR, they amplify a uniform product from genomic DNA because functional copies (if more than one) are identical and pseudogenes lacking parts of 3' and/or 5' ends are not amplified. In the absence of the gene-specific forward primer (as in gene-walking PCR) preferential amplification of authentic gene(s) ceases and the reverse primer is free to initiate strands from 3' ends of functional and

pseudogenes alike. This would result in the generation of several heterogeneous amplicons. The sequencing primer used to read into the 5' region, which by necessity was selected from the 3' half of the noncoding strand would bind to most of these amplicons because of its complementarity to their 3′ region. As a result, the probability of detecting functional *tDNA*$^{met-i}$ gene sequences is greatly reduced. A large number of amplicons may have to be sequenced in order to find the authentic extension product. The large number of pseudogenes would also explain the failure of enriched small-insert libraries to capture authentic *tDNA*$^{Met-i}$ genes (Chapter IV, section 3.4.3.5).

In the case of single primer- and TAIL PCR, the effect is further compounded by partial complementarity between 5′ and 3′ ends of tDNA^{met-i} gene. The sequence of most tRNAs show several inverted repeats which account for the canonical secondary structure of tRNAs which consists of 4-5 stem structures showing typical Watson-Crick and G-U pairing (Figure-5.6).

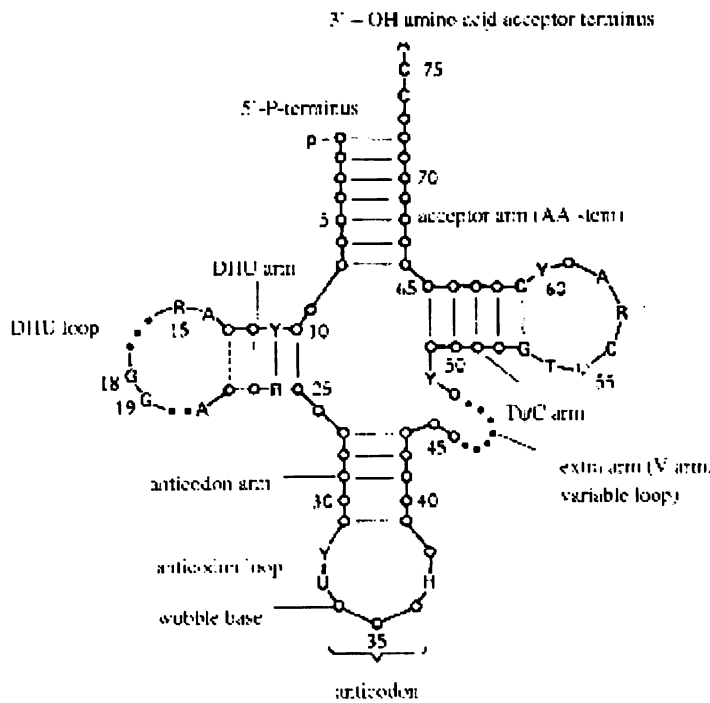


Figure-5.6. The secondary structure model of tRNA (adapted from Kahl, 1995). Shows complentary base pairing of 5′ and 3′ ends.

The stalk of the 'clover-leaf' is formed by base pairing of the 5′ and 3′ ends of the sequence. The sequence of pine tDNA^{met-i} (Chapter IV, section 3.4.2),

likewise, shows strict complementarity between bases constituting its ends (bases 1-7 and 65-71). When sub-optimal annealing temperatures are used to encourage mispriming as in single primer- and TAIL PCR, the imperfectly complementary 3′ site of the noncoding strand forms a second strong focus for reverse ('internal) primer binding and extension (probably by bubble formation) in the opposite direction into the 3′ flanking region.

In retrospect, the 'sequencing' or 'specific' primer could have been designed from the 5' end of pine *tDNA*$^{met\text{-}i}$. This would limit amplification (or 'reading') and hence, screening to authentic genes and pseudogenes which contain both the 3' and 5' ends of functional genes. The problem with this approach is that the sequence readout will be into a region with no signature or conserved sequences. With no contig formation with the known 5' end of *tDNA*$^{met\text{-}i}$, further evidence may be needed to show that it is part of the functional gene. Such proof may be obtained by using the reverse primer of *tDNA*$^{met\text{-}i}$ with a primer designed from the 5' end of the PCR amplified putative flanking region to amplify a genomic sequence showing continuation of the gene into its flanking sequence.

Another way would be to divide the PCR products resulting from these gene-walking methods, which are mini-libraries in their own right, into fractions and test aliquots of these for the presence of *tDNA*$^{met\text{-}i}$ by PCR. Fractions deemed to contain the known coding sequence can be transformed into an enriched library and screening continued till a clone containing the authentic product is detected.

The various PCR-based gene walking methods are not necessarily universal in their application and their success is dependent on the nature of the isolated gene. When dealing with putative multigene families containing many pseudogenes or gene look-alikes, methods employing stringent conditions may have a higher success rate in identifying the authentic gene. The RAGE protocol used in this study (Chapter IV, section 3.4.3.3) could have been tested at higher annealing temperatures by using longer primers; other restriction enzymes which gave a reasonable level of DNA digestion could

have been also tested. Another method worthy of consideration would have been that of Siebert *et al* [1995] which uses a very high annealing temperature for its 'suppression PCR' effect and has given good results with a variety of genomes.

Further attempts at $tDNA^{met\text{-}i}$ isolation would be justified if its copy-status can be established. Southern hybridization of pine genomic DNA with the $tDNA^{met\text{-}i}$ probe failed to reveal the copy status of the gene (Chapter IV, section 3.5.4); neither is the copy number known in other plant species. Pine $tDNA^{met\text{-}i}$ gene could be a single- copy gene such as $tDNA^{Leu}$ of *Phaseolus vulgaris* [Green *et al*, 1987], $tDNA^{Ala}$ of *A.thaliana* [Akama & Tanifuji, 1990] and $tDNA^{Thr}$ of beans [Ramamonjisoa *et al,* 1998]. Or, it could be an oligo- or multicopy gene like most other tDNA genes. The failure of Southerns to detect putative multiple gene copies is explained by the dispersed nature of plant tRNA genes; most plant tDNAs including $tDNA^{met\text{-}i}$ occur as individual genes on large DNA fragments of up to 14 kb [Bawnick *et al*, 1983; Waldron *et al*, 1985; Palmer & Folk, 1987; Reddy & Padyatta, 1988; Akama & Tanifuji, 1990; Beier *et al*, 1991].

If pine $tDNA^{met\text{-}i}$ is a putative multicopy gene, the strong constitutive expression observed in expression studies (Chapter IV, Section 3.4.1) may be due to spatial and temporal expression of different members of the family, or differential expression of two or more paralogs, or even strong constitutive expression of one or more paralogs similar to *Act2* and *Act8* of the 10-member actin superfamily of *A.thaliana*. The expression of individual $tDNA^{met\text{-}i}$ species can be resolved by using sequence-specific probes unique to individual $tRNA^{met\text{-}i}$. Introns have been used as gene-specific probe in *in situ* hybridization to detect intron-containing pre-tRNA [Bertrand *et al*, 1998] or excised introns [O'Connor & Peebles, 1991] but are not applicable to pine $tDNA^{met\text{-}i}$ since it is intron-free. One can also use 5′ leader and 3′ trailer sequence information of tRNA precursors to develop gene-specific probes; tRNAs are transcribed as a longer pre-tRNA from which 5′ and 3′ precursor sequences are removed by endonucleolytic cleavage during biogenesis of mature tRNA. A review of several clustered

and dispersed tRNA families [Green & Weil, 1989; Fuchs *et al*, 1992], however, reveals that these regions which number only a few bases in length and much smaller than those found in class II genes show little variation among the paralogs and therefore cannot be used to discriminate members of a family.

Probably the use of real-time PCR, using various concentrations of pine nuclear DNA and PCR amplified *tDNA*$^{met\text{-}i}$ as standards at different stringencies would help to clarify the copy status of authentic pine *tDNA*$^{met\text{-}i}$ and its pseudogenes.

6. THE IDEAL PINE PROMOTER

The study identified a promising endogenous promoter in *5Spr20*, which may be an excellent candidate for gene downregulation applications in pine. It gave superior silencing of *gus* reporter expression by antisensing during the transient expression stage in undifferentiated, primordial, embryogenic pine cells (equivalent to mammalian stem cells) and by dsRNA-mediated silencing during stable expression in *N. benthamiana*. Similar to other model plants such as *Arabidopsis* and tobacco, *N. benthamiana* often faithfully reproduces the transformation effects in target plant species. However, these preliminary findings still need to be verified in pine transformants for at least a two year cycle.

5Spr20 has the added advantage of belonging to a multi-copy gene family (~44 copies) in pine. Therefore, the addition of one or more *5Spr20* transgene copies to the pine genome has a very low probability of triggering transgene silencing resulting from repeat induced silencing (RITS).

The *5Spr20* promoter is, however, inefficient in directing protein synthesis unlike its counterpart class III, VA RNA_I gene promoter, which transcribed a translatable chimaeric protein coding sequence [Gunnery & Mathews, 1995]; yet, several exciting possibilities exist to engineer *5Spr20* into a recombinant promoter suitable for both sense- and antisense applications. Frenkel *et al*

[2004] note that the presence of numerous class III, tRNA-like sequences found in the region of Pol II promoters is an adaptation of Pol III promoters to binding by RNA polymerase II; therefore, subtle changes introduced into *5Spr20* promoter can make it capable of directing protein synthesis. For example, in the case of Pol III transcribed *Arabidopsis* U snRNA genes, it needs only a 10 bp insertion (or one helical turn) between two promoter elements to alter their polymerase specificity to Pol II [Waibel & Filipowicz, 1990].

Nevertheless, the ideal pine promoter is one which belongs to a Class II gene since it is inherently bi-functional; it is tailored for both gene expression and gene downregulation applications. The original proposal, in fact, was to use differential screening of pine cDNA libraries as the starting point for gene discovery. This was discounted later because of the unavailability of radioactive facilities for library screening. Using a differential (+ and -) library screening approach thousands of pine protein coding genes (assuming a minimum of 20,000 based on the human genome) can be reduced to an enriched library containing hundreds, if not tens, of constitutively/spatially expressed Class II pine transcript clones; sequential screening of a cDNA library from a selected tissue with labelled cDNA libraries prepared from different pine tissues and stages will result in a bank of potential candidate genes. The library is further reduced, if necessary, to a manageable size by eliminating clones representing housekeeping multigene families such as actin, tubulin, GAPDH and histones using conserved, or degenerate probes.

In this approach, the target gene pool is captured early and the nature of the constituent genes determined subsequently. In this study, however, a decision was taken to follow an approach, which was the exact opposite, where candidate genes were first determined and then the transcriptome/genome was screened for the genes.

Wisdom dictates that single-copy genes are easier targets to isolate because of their relatively unique sequence, than a member of a multigene family which shares varying degrees of homology with its paralogs. However, if

duplicated genes were to be totally excluded as candidates, a large number of protein coding genes will be lost to the selection process. A comparative analysis of the fully sequenced genomes of three model organisms shows that their proteomes are dominated by protein families [Rubin *et al*, 2000]; in *C.elegans* 8971 genes of a core proteome of 9453 are duplicated, 5536 of 8065 in *D.melanogaster* and 1858 of 4383 in *S.cerevisiae*.

The three candidate genes, Actin *ActX, tDNA*$^{Met\text{-}i}$ and *5Spr20* which were investigated in detail are all members of (putative) multigene families. It is a fair comment that in practice, fishing for a desirable gene within a multigene family can be a logistical nightmare. However, targeting a specific paralog is not entirely illogical if one considers the fact that many of the promoters in current use belong to multigene families; examples are the monocot ubiquitin promoter, rice actin promoter and the maize alcohol dehydrogenase promoter. The Agri-Genesis Corporation (NZ) using its extensive private database of pine ESTs identified yet another promoter of an abundantly expressed ubiquitin paralog - super ubiquitin, which belongs to a superfamily of chaperone proteins. Its 2.1 kb promoter embodying a 5′ UTR interrupted by a 839 bp intron shows a broader expression profile than CaMV 35S promoter [Wood *et al*, 2000].

BIBLIOGRAPHY

Acheré, V., P. Faivre-Rampant, S. Jeandroz, G. Besnard, T. Markussen, A. Aragones, M. Fladung, E. Ritte and J.-M. Favre (2004) A full saturated linkage map of *Picea abies* including AFLP, SSR, ESTP, 5S rDNA and morphological markers *Theoretical and Applied Genetics* 108:1602-13.

Adams, M., Sharp, G., Walker, A. and J.M.Varley (1992) Differential expression of translation-associated genes in benign and malignant human breast tumours. *British Journal of Cancer* 65:65-71.

Akama, K. and S.Tanifuji (1989) Nucleotide sequence of a methionine initiator tRNA gene of *Arabidopsis thaliana*. *Plant Molecular Biology* 13:599-600.

Akama, K. and S.Tanifuji (1990) Sequence analysis of three tRNA-Phe nuclear genes and a mutated gene, and one gene tRNA-Alu from Arabidopsis thaliana. *Plant Molecular Biology* 15:337-46.

Altschul, S., Stephen, F., Madden, L., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and J.Lipman (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. *Nucleic acids Research* 25:3389-402.

Alvarez, M., Rhodes, S. and J. Bidwell (2003) Context-dependent transcription: all politics is local. *Gene* 313, 43-57

Amarasinga, V. and J.E.Carlson (1998) Physical mapping and characterisation of 5S rRNA genes in Douglas Fir. *The Journal of Heredity* 89:495-500

An, G. (1985) Binary Ti vectors for plant transformation and promoter analysis. *Methods in Enzymology* 153:292-300.

An, Y, McDowell, J., Huang, S., McKinney, E. and R.Meagher (1996) Strong constitutive expression of the *Arabidopsis* ACT2/ACT8 actin subclasses in vegetative tissues. *The Plant Journal* 10:1:107-21.

Anandaraj, S. and J.Cherayil (1974) A simple method for large scale preparation of chick embryo tRNA. *Analytical Biochemistry* 58:190.

Appels, R., Gerlach, L.. and J.Peacock (1980) Molecular and chromosomal organization of DNAsequences encoding for ribosomal RNAs in cereals. *Chromosoma* 78:293-311.

Appels, R., Reddy, P., Moran, B. and C.Clark (1989) The molecular cytogenetic analysis of grasses and its application to studying relationship among species of *Triticeae*. *Genome* 31:122-33.

Appels, R. and B.Baum (1992) Evolution of the Nor and 5S DNA loci in the Triticea In Molecular systematics of plants Ed. Soltis *et al* p222, Chapman & Hall, NY.

Arkhipova, I.R. (1995) Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* 139:1359-69.

Arnold., G. and H.J.gross (1987) Unrelated leader sequences can efficiently promote human rRNA transcription. *Gene* 51:237-46.

Arthur, J. (2001) Genbank direct submission.

Asemoto, O., Breda, C., Sallaud, C., Kozak, I., Esnault, R. and A.Kondorosi (1994) Cloning and expression of a cDNA encoding a cytoplasmic L5 ribososmal protein from alfalfa (*Medicago sativa*). *Plant Molecular Biology* 26:1201-5.

Assaad, F., Tucker, K. and E.R.Signer (1993) Epigenetic repeat induced silencing (RIGS) in *Arabidopsis*. *Plant Molecular Biology* 22:1067-85.

Atchison, M. and R.P.Perry (1986) Tandem kappa immunoglobin promoters are equally active in the presence of the kappa enhancer: implications for models of enhancer function. *Cell* 46:253-62.

Ausabel, F.M., Brent, R., Kingston, R., Moore, D., Seidman, G. and K.Struhl (1998) Ed:Current Protocols in Molecular Biology. Wiley Interscience, USA.

Avital, S. and D.Elson (1969) A convenient method for preparing tRNA from *E. coli*. *Biochemica Biophysica Acta* 179:297-9.

Baima, S., Sessa, G., Ruberti, I. and G.Morelli (1995) A cDNA encoding *Arabidopsis thaliana* cytoplasmic ribososmal protein L18. *Gene* 153:171-4.

BairdW. and R.Meagher (1987) A complex gene superfamilyencodes actin in Petunia. *EMBO Journal* 6:3223-3231.

Baker, S., Rugh, C. and J.C.Kamalay (1990) RNA and DNA isolation from recalcitrant plant tisues. *Biotechniques* 9:268-72.

Baran N, Lapidot A. and H.Manor (1991) Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)n and d(GA)n tracts *Proceedings of the National Academy of Sciences* 88:507-11

Barberis, A. and L.Gaudreau (1998) Recruitment of the RNA polymerase II holoenzyme and its implication in gene regulation. *Biological Chemistry* 379:1397-405.

Barciszewska, M., T.Mashkova and J.Barciszewski (1985) The primary structureof maize and tobacco 5S rRNA. *Febs* 192:289-93.

Barciszewski, M.,Erdmann, V. and J.Barciszewski (1994) A new type of RNA editing: 5S ribosomal DNA transcripts are edited to mature 5S rRNA. *Biochemical and Molecular Biology International* 34:437-48.

Barnes, W.M. (1994) PCR amplification of up to 35kb DNA with high fidelity and high yield from lambda-bacteriophage templates. *Proceedings of the National Academy of Sciences* 91:2216-20.

Bartholomew, B., Kassavets, A., Braun, R. and P.Geiduscheck(1990) The subunit structure of *Saccharomyces cervisiae* transcription factor IIIC probed with a novel crosslinking reagent. *EMBO Journal* 9:2197-205.

Bartel, D. (2004) MicroRNAs:genomics, biogenesis, mechanism and functions. *Cell* 116:281-97.

Baty, D., Barrera-Saldana, H., Everett, R., Vigneron, M. and P.Chambon (1984) Mutational dissection of the 21 bp repeat region of the SV40 early promoter reveals that it contains overlapping elements of the early-early and late-early promoters. *Nucleic Acids Research* 12:915-32.

Bawnick, N., Beckman, S., Sarid, S. and V.Daniel (1983) Isolation and nucleotide sequence of a plant tRNA gene: petunia asparagine tRNA. *Nucleic Acids Research* 11:1117-9.

Bedbrook JR, Kolodner R. and L.Bogorad (1977) Zea mays chloroplast ribosomal RNA genes are part of a 22,000 base pair inverted repeat *Cell* 11:739-49

Bedinger P, Munn M. and B.Alberts (1989) Sequence-specific pausing during in vitro DNA replication on double-stranded DNA templates. *Journal of Bioogical Chemistry* 264:16880-6.

Beier, D., Stange, N., Gross, H. and H.Beier (1991) Nuclear tRNA-tyr genes are highly amplified at a single chromosomal site in the geneome of Arabidopsis. *Molecular General Genetics* 225:72-80.

Bell, A. and D.deMarini (1991) Excessive cycling converts PCR products to random length higher molecular weight fragments. *Nucleic Acid Research* 19:5079.

Benfey, N., Ren, L. and N.H.Chua (1989) The CaMV 35S enhancers contain at least two domains which can confer differential development and tissue specific expression patterns. *EMBO Journal* 8:2195-202.

Benfey, N., Ren, L. and N.H.Chua (1990a) Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. *EMBO Journal* 9:1685-96.

Benfey, N., Ren, L. and N.H.Chua (1990b) Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO Journal* 9:1677-84

Bentley, L., Brown, L. and M.Groudine (1989) Accurate TATA-dependant polymerase III transcription from promoters of c-myc gene in injected Xenopus oocytes. *Genes and development* 3:1179-89.

Bernstein E, Caudy AA, Hammond SM and G.Hannon (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference *Nature* 409:363-6

Bertioli, D. (1997) Rapid amplification of cDNA ends. In *PCR Cloning protocols* Ed. B.A.White, Human Press Inc. 490 pp.

Bertrand, E., Scott, F., Kendall, A. and R.engelke (1998) Nucleolar localization of early tRNA processing. *Genes and Development* 12:2463-8.

Besendorfer,V., Krajacic-Sokol,I., Jelenic,S., Puizina,J.,Mlinarec,J., Sviben,T. and D.Papes (2004) Two classes of 5S rDNA unit arrays of the silver fir Abies alba (unpublished)

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Van Staveren, M. and W.Steikman (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana. Nature* 391:485-8.

Birch, R.G. (1997) Plant transformation: Problems and strategies for practical application. *Annual Review of Plant Physiology and Plant Molecular Biology* 48:297-326.

Birnboim and Doly (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* 7:1513.

Birren, B., Green, D., Klapholz, s., Myers, M. and J.Roskams (1997) Ed:Genome analysis: Vol I Analyzing DNA. pp 675. Cold Spring Harbour Press, USA.

Bishop, S., Walter, C. and R.C.Gardner (1998) Genbank direct submission

Bjourson, A. and E.Cooper (1992) Band-stab PCR: a simple technique for the purification of individual PCR products. *Nucleic Acids research* 20:17:4675

Bloomquist, T., Johnson, C. and E.Mains (1992) Rapid isolation of flanking genomic DNA using biotin-RAGE a variation of single sided polymerase chain reaction. *DNA Cell Biology* 11:791-7.

Bogdanov, A, Dontsova, A., Dokudovskaya, S. and N.Lavrik (1995) Structure and function of 5S rRNA inb the ribosome. *Biochemistry and Cellular Biology* 73:869-76.

Bogenhagen, F., Sakonju, S. and D.D.Brown (1980) A control region in the center of the 5S rNA gene directs specific initiation of transcription: II.The 3' border of the region. *Cell* 19:27-35.

Bogenhagen, F. and D.D.Brown (1981) Nucleotide sequence in Xenopus 5S DNA required for transcription termination. *Cell* 24:261-70.

Bonham Smith, P., Oancia, T. and M.Moloney (1992) Cytoplasmic ribosomal protein S15a from *Brassica napus*. molecular cloning and developmental expression in mitotically active tissue. *Plant Molecular Biology* 18:909-19.

Bosher, M. and M.Labouesse (2000) RNA interference:genetic wand and genetic watchdog. *Nature:Cell Biology* 2:E31-6. Macmilan Magazines, UK.

Bottger, C. (1988) High efficiency generation of plasmid cDNA libraries using electro transformation. *Biotechniques* 6:878.

Bouchez, D., Tokuhisa, J., Llewellyn, D., Dennis, E. and J.G.Ellis (1989) The OCS element is a component of the promoters of T-DNA and plant viral genes. *EMBO Journal* 8:4197-204.

Bourque, E. and R.Folk (1992) Suppression of gene expression in plant cells utilizing antisense sequences transcribed by Pol III. *Plant Molecular Biology* 19:641-7.

Bradford, M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Analytical Biochemistry* 72:248.

Branch, D. (1998) A good antisense molecule is hard to find. *TIBS* 23:45-50.

Braun, P., Emerman, M. and U.K.Schmitz (1993) Primary structure and expression of a gene encoding the cytosolic ribosomal protein S4 from potato. *Biochimica Biophysica Acta* 1218:435-8.

Bredow, S., Surig, D., Muller, J., Kleinert, H. and J.Benecke (1990) Activating transcription factor (ATF) regulates human 7 S L RNA transcription by RNA Polymerase III *in vivo* and *invitro. Nucleic Acids Research* 18:6779-84.

Brown, M., Scroth, G., Gottesfeld, J. and D.Bazett Jones (1996) Protein and DNA requirements for transcription factor IIIA induced distortion of the 5S rRNA gene promoter. *Journal of Molecular Biology* 262:5:600-14.

Brown, G. and J.Carlson (1996) Picea glauca 5S ribosomal RNA gene, partial sequence. Direct submission to NCBI.

Buchholz, F., Dreschel, F., Ruer, M and R.Kittler (2005) Production of siRNA in vitro by enzymatic digestion of double stranded RNA. pp 87-99 *Gene silencing by RNA interference:Technology and application.*CRC Press, USA.

Burke, T. and J.T.Kadonaga (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA box deficient promoters. *Genes and Development* 10:6:711-24.

Burke, T. and J.T.Kadonaga (1997) The downstream core promoter element DPE is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes and Development* 11:22:3020-31.

Burton, R., Metz, E., Flower, J. and C.Wilett (2005) Unusual structure of ribosomal DNA in the copepod Tigriopus californicus: intergenic spacers lack internal subrepeats. *Gene* 344:105-13.

Busby, S. and R.H.Reeder (1983) Enhancer elements in the intergenic spacer regulate transciption *Cell* 34:989-96.

Busch, M., Bomblies, K. and D.Weigel (1999) Activation of a floral homeotic gene in *Arabidopsis. Science* 285:585-7.

Busk, P. and M.Pages (1998) Regulation of abscic acid induced transcription. *Plant Molecular Biology* 37:425-35.

Caetano-Anolles, G. and N.Trigiano (1998) Recovery of DNA amplification products from silver stained polyacrylamide gels. In *Methods in molecular Biology* Ed. White, Human Press, NJ 490 pp.

Campbell, K., Raymond, G. and J.Johnson (1985) In vivo modulation of yeast tRNA gene expression by 5 flanking sequences. *EMBO Journal* 4:2649-56.

Canaday, J., Guillemaut, P. and H.Well (1980) The nucleotide sequences of the initiator transfer RNAs from bean cytoplasm and chloroplasts. *Nucleic acids Research* 8:999-1008.

Cantor, R. and L.Smith (1999) Genomics. John Wiley and sons Inc. pp 596.

Carey, M. (1998) The enhanceosome and transcriptional synergy. *Cell* 92:5-8.

Carlson, P. and J.Ross (1984) Amanitin insensitive transcription of mouse b-globin 5 flanking and structural sequences correlated with mRNA expression. *Proceedings of the National Academy of sciences.* 81:7782-6.

Carson, C., Hattori, I., Vasil, I., Peterson, P. and D.R.McCarty (1997) The quiescent/colorless alleles of *viviparous1* show that the conserved B3 domain of VP1 is not essential for ABA regulated gene expression in the seed. *Plant Journal* 12:1231-40.

Castello, D., Lough, J. and S.Emerson (2000) Development of an RNA virus vector for use in conifers. *Proceedings of the Queenstown Molecular Biology Conference.*

Chalfie, M., Yuan, T., Euskirchen, G., Ward, W. and D.C.Prasher (1994) Green fluorescent protein as a marker for gene expression. *Science* 263:802-5

Challice, M. and J.Segall (1989) Transcription of the 5S RNA gene of *Saccharomyces cerivisae* requires a promoter element at +1 and a 14-base pair internal control region. *Journal of Biological Chemistry* 264:20060-7.

Chang, L. and C.M.Stoltzfus (1987) Inhibition of Rous sarcoma virus by antisense RNA. *Journal of Virology* 61:921-4.

Chang, S., Puryear, J. and J.cairney (1993) A simple and efficient method for isolating RNA from Pine trees. *Plant Molecular Biology reporter* 11:2:113-6.

Charity, J., Holland, L. and C. Walter (2005) Consistent and stable expression of the *npt*II, *uid*A and *bar* genes in transgenic *Pinus radiata* after *Agrobacterium tumefaciens*-mediated transformation using nurse cultures. *Plant Cell Reports* 23:606-16

Chassy, M. and L.Flickering (1987) Transformation of *Lactobacillus caseii* by electroporation. *FEMS Microbiological Letter* 44:173-5.

Chen, J., Attardi, L., Verrizer, C., Yokomori, K. and R.Tjian (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* 79:93-106.

Choisne N, Martin-Canadell A, and I.Small (1997) Transactivation of a target gene using a suppressor tRNA in transgenic tobacco plants. *Plant Journal* 11:597-604

Choisne N, Carneiro VT, Pelletier G and I. Small (1998) Implication of 5'-flanking sequence elements in expression of a plant tRNA(Leu) gene *Plant Molecular Biology* 36:113-23

Chomczynski, P. and K.Mackey (1995) Modification of the TRI reagent procedure for isolation of RNA from polysaccharide and proteoglycan-rich sources. *Biotechniques* 19:6:942-5.

Chou, Q., Russell, M., Birch, E. and W.Bioch (1992) Prevention of pre-PCR mispriming and primer dimerization improves low-copy number amplification. *Nucleic Acids Research* 20:1717-23.

Ciliberto, G., Traboni, C. and R.Cortese (1982) Relationship between the two components of the split promoter of eukaryotic tRNA genes. *Proceedings of the National Academy of Sciences* 79:1921-5.

Clarke, L. and J.Carbon (1976) A colony bank containing synthetic Col EI hybrid plasmids representative of the entire *E. coli* genome. *Cell* 9:91-4.

Clarck, M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research* 16:9677-86.

Cloix, C., Tutois, S. and S.Tourmente (2000) Analysis of 5S rDNA arrays in *A. thaliana*:physical mapping and chromosome specific polymorphism. *Genome Research* 10:679-691.

Cloix, C., Tutois, S. *et al.* (2001) Analysis of the 5S RNA pool in Arabidopsis: RNAs are heterogeneous and only two of the genomic 5S loci produce mature RNA. *Genome research* 12:132-44.

Cloix, C., *et al* (2002) Analysis of the 5S RNA pool in A thaliana: RNAs are heterogenous and only two of the genomic 5S loci produce mature 5S RNA. *Genome Research* 12:132-44.

Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M. and S.Tourmente (2003) In vitro analysis of the sequences required for transcription of the Arabidopsis thaliana 5S rRNA genes. *Plant Journal* 35(2):251-61.

Cogni, C., Ireland, T., Schumacher, M., Selker, U. and G.Macino (1996) Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoiplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *EMBO Journal* 15:3153-63.

Cogni., C. and G.Macino (1999) Gene silencing in N.crassa requires a protein homologous to RDRP. *Nature* 399:166-9.

Coleman, J., Green, P. and M.Inouye (1984) The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Cell* 37:429-36.

Colliver, S., Morris, P. and M.P.Robbins (1997) Differential modification of flavanoid and isoflavanoid biosynthesis with an antisense chalcone synthase construct in transgenic *Lotus corniculatus*. *Plant Molecular Biology* 509-22.

Comai, L., Tanese, N. and R.Tjian (1992) The TATA binding protein and associated factors are integral components of RNA polymerase transcription factor SL1. *Cell* 68:965-76.

Comai, L., Zomerdijk, J., Beckmann, H., Zhou, S., Admon, A. and R.Tjian (1994) Reconstitutionof transcription factor SL I: exclusive binding of TBP by SL I or TFIID subunits. *Science* 266:1966-72.

Cooke, R., Raynal, M., Laudie, M. and M.Delseny (1997) Identification of members of gene families in *Arabidopsis thaliana*by contig construction from partial DNA sequences:106 genes encoding 50 cytoplasmic ribosomal proteins. *Plant Journal* 11:5:1127-40.

Cormack, R and T.Somssich (1997) Rapid amplification of genomic ends as a simple method to clone flanking genomic DNA. *Gene* 194:273-6.

Costa, L. and P.Weiner (1995) Cloning and analysis of PCR generated DNA fragments. In PCR primer: A laboratory manual. Ed: Dieffenbacher, W. and S.Dverrksler. Cold Spring Harbour Press, USA.

Crighton, D., Zhang, C., Morton, J. and D.Johnson (2003) p53 represses RNA pol III transcription by targeting TBP and inhibiting promoter occupancy by TFIIIB. *EMBO Journal* 22:2810-20.

Cronn, C., Zhao, H., Paterson, A. and J.Wendel (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5s ribosomal DNA in diploid and alloploid cottons. *Journal of Molecular Evolution* 42:6:685-705.

Cullis, C., Rivin, c> and V.Walbot (1984) A rapid procedure for the determination of repetitive sequences in eukaryotic genomes. *Plant Molecular Biology reporter* 2:24-31.

Cunliffe, V. (2003) Memory by modification: the influence of chromatin structure on gene expression during vertebrate development. *Gene* 305:141-50.

Dalmay T, Hamilton A, Rudd S, Angell S. and D.Baulcombe (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus *Cell* 101:543-53

Dechampesme, A., Koroleva, O., Gas, N. and S.Camier (1999) Assembly of 5S ribosomal RNA is required at a specific step of the pre-RNA processing pathway. *Journal of Cell Biology* 145:1369-80.

DeFranco, D., Sharp, S. and S.Soll (1981) Identification of regulatory sequences contained in the 5-flanking region of *Drosophila* lysine tRNA2 genes. *Journal of Biological Chemistry* 256:12424-9.

Dekeyser, R., Claes, B., Marichal, M., van Montague, M. and A,Caplan (1989) Evaluation of selectable markers for rice transformation. *Plant Physiology* 90:217-23.

De la Riva, G., Gonzalez, J., Vazquez, R and C.Pardo (1998) Agrobacterium tumefaciens: a natural tool for plant transformation. *Electronic Journal of Biotechnology* 1:3:1-18

Delort, J., Dumas, B., Darmon, C. and J.Mallet (1989) An efficient strategy for cloning 5 extremeties of rare transcripts permits isolation of multiple 5 untranslated regions of rat tryptophan hydroxylase mRNA. *Nucleic Acids Research* 17:6439-48.

Deshmuk, M., Tsay, F., Paulovich, A. and L.Woolford (1993) Yeast ribosomal protein Lirequired for the stability of Newly sysnthesized 5S rRNA and the assembly of 60S ribosomal subunits. *Molecular and Cellular Biology*13:2835-45.

Devey, M., Bell. J., Uren, T. and G.Moran (2002) A set of microsatellite markers for fingerprinting and breeding applications in Pinus radiata. *Genome* 45:984-9.

DeWinter, R. and T.Moss (1987) Polymerase I mediated transcription in *Xenopus* system *Journal of Molecular Biology* 196:813-27.

Dolfini, S., Consonni, G., Mereghetti, M. and C.Tonelli (1993) Antiparallel expression of the sense and antisense transcripts of tubulin genes. *Molecular and General Genetics* 241:161-9.

Don, H., Cox, T., Wainwright, J. and J.S.Mattick (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 19:4008.

Donald, R. and A.R.Cashmore (1990) Mutations of either G-box or I-box sequences profoundly affects expression from *Arabidopsis* rbcs-1A promoter. *EMBO Journal* 9:1717-26.

Donze, O., Dudek, P. and D.Picard (2005) siRNA production by in vitro transcription. pp 59-72 *Gene silencing by RNA interference:Technology and application.*CRC Press, USA.

Doyle,J. and H.Doyle (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13-15.

Dower, J., Miller, F. and W.Ragsdale (1988) High efficiency transformation of *E.coli* by high voltage electroporation. *Nucleic Acids Research* 16:6127.

Drapkin, R., Merino, A. and D.Reinberg (1995) Regulation of RNA Polymerase II transcription. *Current Opinion in Cell Biology* 5:469-76.

Dunning, M., Talmud, P. and E.Humphries (1988) Errors in the polymerase reaction. *Nucleic Acids Research* 16:10395.

Dwivedi, U., Campbell, W., Yu, J., Raju, S. and G.K.Podila (1994) Modification of lignin biosynthesis in transgenic *Nicotiana* through expression of an antisense O-methyltransferase gene from *Populus*. *Plant Molecular Biology* 26:61-71.

Dykxhoorn, D., Novina, C. and P.Sharp (2003) Killing the messenger: short RNAs that silence gene expression. *Molecular Cell Biology* 4:457-67.

Edwards, D., Delort, B. and J.Mallet (1991) Oligonucleotide ligation to single stranded cDNAs A new tool for clning 5 ends f mRNAs and fr constructing cDNA libraries by in vitro amplification. *Nucleic Acids Research* 19:5227-32.

Edwards, M., Valdenaire, O. and J.Mallet (1998) Anchoring a defined sequence to the 5 end of mRNAs. In *Methods in Molecular Biology.* Ed. White, Human Press, pp 490.

Egertsdotter. U, van Zyl L, MacKay J, Peter G, Kirst M, Clark C, Whetten R and R.Sederoff (2004) Gene expression during formation of earlywood and latewood in loblolly pine: expression profiles of 350 genes Plant Biology 6:654-63

Elbashir, S., Martinez, J., Lendeckel, W. and T.Tuschi (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO JOURNAL* 20:6877-88.

Ellestrom, M., Stalberg, K., Ezcurra, I. and L.Rask (1996) Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm specific transcription. *Plant Molecular Biology* 32:1019-27.

Ellis, G., Llewellyn, J. and J.Peacock (1987) Maize Adhl promoter sequences control anaerobic regulation. *EMBO J* 6:11-6.

Ellis, T., Lee, D., Thomas, C., Simpson, P., Cleary, W. and K.W.Burcham (1988) 5S rRNA genes in *Pisum:*sequence, long range and chromosomal organisation. *Molecular and General Genetics* 214:333-342.

Elmayan, T. and H.Vaucheret (1996) Expression of single copies of a strongly expressed 35S transgene can be silenced post transcriptionally. *Plant journal* 8:787-97.

Erdmann, A., Wolters, J. and R.De Wachter (1985) Collection of 5S, 5.8S and 4.5S ribosomal RNA sequences *Nucleic Acids research* 13:r105-53.

Euch, C., Allemand, C., Pastuglia, M., Capelli, P. and L.Jouanin (1998) Expression of antisense chalcone synthase RNA in transgenic hybrid walnut microcuttings. Effect on flavanoid contentand rooting ability. *Plant molecular Biology* 38:467-79.

Falkenburg D, Dworniczak B, Faust, D. and E.Bautz (1987) RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit to the beta subunit of Escherichia coli RNA polymerase *Journal of Molecular Biology* 195:929-37

Feinberg, A.P. and B.Vogelstein (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* 132:6-13.

Felgenhauer P, Sedman J, Shostak N, Timofeeva M, Lind A, A.Bayev (1990) The 5'-flanking sequence of the loach oocyte 5S rRNA gene contains a signal for effective transcription *Gene* 90:243-8.

Felsenfeld, G. (1992) Chromatin as an essential part of the transcriptional mechanism. *Nature* 355:219-24.

Felsenstein, L. (1988) Phylogenies from molecular sequences: inference and reliability.*Annual Revew of Genetics*. 22:521-65.

Fields S and O.Song (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340:245-246.

Fire, A., Xu, S., Montgomery, K., Kostos, A., Driver, E. and C.Mello (1998) Ptent and specific genetic interference by double stranded RNA in *Caenorhabditis elegens. Nature*:391 806-11.

Flavell, B. (1986) Repetitive DNA and chromosomal evolution in plants. *Journal of the Royal society* 312:227-42.

Fleischer, S. and I.Grummt (1983) Expression of an mRNA coding gene under the control of an RNA Polymerase I promoter. *Embo Journal* 2:2319-22.

Fors, L., Saavedra, A. and L.Hood (1990) Cloning of the shark promoter using a genome walking technique based on PCR. *Nucleic Acids Research* 11:2793-9

Foster, R., Izawa, T. and N.H.Chua (1994) Plant bZIP proteins gather at ACGT elements. *FASEB Journal* 8:2:192-200.

Fourney, M., Miyakoshi, J. and C.Paterson (1988) Northern blotting: efficient RNA staining and transfer. *Focus* 10:1:5-6.

Frasch, M., Wenzel, W. and D.Hess (1989) The nucleotide sequences of nuclear 5S rRNA genes and spacer regions of *Petunia hybrida. Nucleic Acids Research* 17:2857.

Frenkel, F., Chaley, M. and K.Skryabin (2004) Evolution of tRNA-like equences nd genome variability. *Gene* 335:57-71.

Frohman, M., Dush, M. and G.Martin (1988) Rapid production of full length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences* 85:8998-9002.

Frohman, M.A. (1990) RACE:rapid amplification of cDNA ends In PCR protocols Ed. Innis *et al*, Academic Press, CA. pp 28-38.

Fromm, M., Taylor, L. and V,Walbot (1986) Stable transformation of maize after gene transfer by electroporation. *Nature* 319:791-3.

Fromont-Racine, M., Senger, B. and F.Fasiolo (2003) Ribosome assembly in eukaryotes *Gene* 313:17-42.

Fruscoloni, P., Zamboni, M., Panetta, G. and G.Tocchini (1995) Mutational analysis of the transcription start site of yeast tRNA-Leu3 gene. *Nucleic Acids Research* 23:2914-8.

Fuchs, T., Beier, D. and H.Beier (1992) The tRNA Tyr multigene family of Nicotinia rustica: genome organization, sequence analysis and expression in vitro. *Plant Molecular Biology* 20:869-878.

Fulnecek, J., Matyesek, R. and A.Kovarik (2002) Distribution of 5 methyl residues in 5S RNA genes in A.thaliana and S.cereale *Molecular Genetics and Genomics* 268:510-17

Fung, P., Gaertig, J, Grovosky, M. and M.Hallberger (1995) Requirement of a small cytoplasmic RNA for the establishment of thermotolerance. *Science* 268:1036-9.

Futterer, J., Lazlo, Z. and T.Hohn (1993) Nonlinear ribosome migration on cauliflower mosaic virus 35s RNA. *Cell* 73:789-802.

Gaillard, C. and F.Strauss (1990) Ethanol precipitation of DNA with linear polyacrylamide as carrier. *Nucleic Acids Research* 25:18:378

Gallie, R. (1993) The cap and poly-A tail function synergistically to regulate mRNA translational efficiency. *Genes and Development* 5:2108-16.

Gartland KM, Phillips JP, Vitha S. and K.Benes (1995) Fluorometric GUS analysis for transformed plant material. *Methods in Molecular Biology* 1995;44:195-9

Gauss, D. and M.Sprinzl (1981) Compilation of tRNA sequences. *Nucleic Acids research* 9:r1-23.

Geiduscheck, E. and G.Kassavetis (1995) Comparing transcriptional initiation by RNA Polymerases I and III. *Current Opinions in Cell Biology* 7:344-51.

Gelvin, S., Karchar, S. and V.J.DeRita (1983) Methylation of the T-DNA in *Agrobacterium tumefaciens* and in several crown-gall tumours. *Nucleic Acids Research* 11:159-74.

Gerber, P., Seipel, K., Hug, M., Rusconi, S. and W.Schaffner (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263:808-11.

Gerlach, L. and A.Dyer (1980) Sequence organization of the repeating units in the nucleus of wheat which contains 5S RNA genes. *Nucleic Acids Research* 8:4851-65.

Ghosh, P., Ghosh, K., Simsek, M. and L. Rajbandaray (1982) Nucleotide sequence of wheat germ cytoplasmic initiator methionine transfer ribonucleic acid. *Nucleic Acids Research* 10:3241-7.

Gidekel, M., Jimenez, B. and L.Herrera-Estrella (1996) The first intron of *Arabidopsis* gene coding for elongation factor 1 beta contains an enhancer like element. *Gene* 170:201-6.

Giese, K., Kingsley, C., Kirschner, J. and R.Grosshel (1995) Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1 induced DNA bending and multiple protein-protein interactions. *Genes and Development* 9:995-1008.

Girard-Reydet, C., Gregoire, D. and M.Mechali (2004) DNA replication initiates at domains overlapping with nuclear matrix attachment regions in the Xenopus and mouse c-myc promoter. *Gene* 332:129-38.

Giuliano, G., Pichersky, L., Malik, V. and A.R.Cashmore (1988) An evolutionarily conserved protein binding sequence of a plant light regulated gene. *Proceedings of the National Academy of Sciences* 85:7089-93.

Glasel, J. (1995) Validity of nucleic acid purities monitored by A260nm/A280nm absorbance ratios. *Biotechniques* 18:1:62-3

Goddameir, L., Rensing, A. and G.Feix (1996) Characterization of a maize ribosomal P2 protein cDNA and phylogenetic analysis of the P1/P2 family of ribosomal proteins. *Plant Molecular Biology* 30:655-8.

Goldsborough, P., Ellis, T. and G.Lomonosoff (1982) Sequence variation and methylation of the flax 5S RNA genes. *Nucleic Acids Research* 10:4501-4514.

Gonzalez-Couto, E., Klages, N. and M.Strubin (1997) Synergistic and promoter selective activation of transcription by recruitment of transcription factors TFIID and TFIIB. *Proceedings of The National Academy of Sciences* 94:8036-41.

Gorman, W., Teasdale, D. and A.Cullis (1991) Structure and organization of the 5S rRNA genes in *Pinus radiata. Plant Systematic evolution* 183:223-4.

Gorman, W., Cullis, A. and D.Teasdale (1992) Producing a P1 bacteriophage library from pine: Isolation and cloning of very high molecular weight DNA. *Biotechniques* 12:722-6.

Gottlob-McHugh, S., Levesque, M., McKenzie, K., Olson, M., Yarosh, O. and D.Johnson (1990) Organization of the 5S rRNA genes in soybean *Glycine max* and conservation of the 5S rDNA repeat structure in higher plants. *Genome* 33:486-94.

Gou, D. and L.Liu (2003] Gene silencing in mammalian cells by PCR based short hairpin RNA. *FEBS letters* 548:113-118.

Gould , J., Zhou, Y. and R.Newton (2002) Transformation and regeneration of loblolly pine:shoot apex inoculation with Agrobacterium. *Molecular Breeding* 10:131-41.

Green, G., Marechal, L. and P.Guillemaut (1987) A phaseolus vulgaris mitochondrial tRNA-Leu is identical to its cytoplasmic counterpart: sequencing and in vitro transcription of the gene corresponding to cytoplasmic tRNA-Leu. *Plant Molecular Biology* 10:13-19.

Green, A. and H.Weil (1989) Evolution of a plant nuclear tRNA-Pro gene clusters. *Plant Molecular Biology* 13:727-30.

Greenblatt, J. (1997) RNA polymerase II holoenzyme and transcriptional regulation. *Current Opinion in Cell Biology* 9:310-9.

Greiger, J., Bull, D., Eckels, M. and J.Gorski (1993) Amplification of complementary DNA from mRNA with unknown 5′ ends by one-way polymerase chain reaction. *Methods in Enzymology*218:321-35.

Grimaldi, C., Fiorentini, P. and P.DiNocera (1990) Spacer promoters are orientation dependent activators of pre-tRNA transcription in Drosophila melanogaster. *Molecular and Cellular Biology*. 10:4667-77.

Grimsley, N., Hohn, T., Davis, J. and B.Hohn (1987) Agrobacterium mediated delivery of infectious maize streak virus into maize plants. *Nature* 325:177-9.

Grishok, A., Tabara, H. and C.Mello (2000) Genetic requirements for inheritance of RNAi in *C. elegans*. *Science* 287:2494-7.

Gruenbaum, Y., Maney, T. and A.Razin (1981) Sequence specificity of methylation in higher plant DNA. *Nature* 292:860-2.

Guiley, H., Dudley, R., Jonard, G., Balaes,E. and K.E.Richards (1982) Transcription of cauliflower mosaic virus: Detection of promoter sequences and characterisation of transcripts. *Cell* 30:763-73.

Gunnery, S. and B.Mathews (1995) Functional mRNA can be generated by RNA Polymerase III. *Molecular and Cellular Biology* 15:3597-607.

Gunnery, S., Maivalis, U. and B.Mathews (1997) Translation of an uncapped mRNA involves scanning. *The Journal of Biological Chemistry* 272:21642-6.

Gunnery, S., Ma, Y. and B. Mathews (1999) Termination sequence requirements vary among genes transcribed by RNA Polymerase III. *Journal Of Molecular Biology* 286:3:745-57.

Ha, I., Roberts, S., Maldonado, E., Green, M. and D.Reinberg (1995) Multiple functional domains of human transcription factor IIB. Distinct interaction with two general transcription factors and RNA Polymerase II. *Genes and Development* 7:1021-10

Hacker KJ, and B.Alberts (1994) The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix. A model for polymerase release following the termination of each Okazaki fragment. *Journal of Biological Chemistry* 269:24221-8

Hadjeb, N. and A.Berkowitz (1996) Preparation of T-overhang vectors with high PCR product cloning efficiency. *Biotechniques* 20:1:20-2.

Hallenburg, C. and S.fredriksen (2001) Effect of mutations in the upstream promoter on the transcription of human 5S rRNA genes. *Biochemical Biophysical Acta.* 1520:169-73.

Halpin, C., Knight, M., Foxon, G., Campbell, M., Chabbert, B., Tollier, M. and W.Schuch (1994) Manipulation of lignin quality by down regulation of cinnamyl alcohol dehydrogenase. *The Plant Journal* 6:339-50.

Haltiner, M. (1986) Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. *Molecular Cell Biology* 6:227-35.

Hamilton, A. and D.Baulcomb (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286:950-2

Hamilton, A., Voinnet, O., Chappell, L, and D.Baulcombe (2002) Two classes of short interfering RNA in RNA silencing. *EMBO Journal* 21:4671-79.

Hamoui S, Benedetto JP, Garret M, Bonnet (1994) Quantitation of mRNA species by RT-PCR on total mRNA population *PCR Methods Application* 4:160-6

Hansen, S., Takada, S. and R.Tjian (1997) Transcription properties of a cell-type specific TATA-binding protein TRF. *Cell* 91:71-83.

Hardtke, C. and T.Berleth (1998) The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO Journal* 17:1405-11.

Hariharan, N., Reddy, S. and D.Padyatta (1987) 5S RNA genes in Rice embryos. *Plant molecular Biology* 9:443-51.

Hasegawa, K., Yukawa, Y. and M.Sugiura (2003) A tRNA-Leu like sequence located immediately upstream of an Arabidopsis clock-regulated gene is transcriptionally active *Gene* 307:133-9.

He, R., Wang, Y., Shi, Z. and G.He (2003) Construction of a genomic library of wild rice and Agrobacterium mediated transformation of large insert DNA linked to BPH resistance locus. *Gene* 321:113-21.

Hecker, L., Swartzbach, S. and W.Barmett (1979) Isolation of *Neurospora* mitochondrial tRNA. *Methods in Enzymology* 59:203.

Hecht, A., Laroche, T., Gasser, S. and M.Grunstein (1995) Histone H3 and H4 N termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell* 80:583-92.

Heifetz, B., Patton, A., Levin, Z. and J.Gielen (2000) Genbank direct submission.

Helliwell, C. and P.Waterhouse (2003) Constructs and methods for high throughput gene silencing in plants *Methods* 30:289-95.

Hemleben, V. and D.Werts (1988) Sequence and organization and putative regulatory elements in the 5S RNA genes of two higher plants (*Vigna radiata* and *Mathiola incana*) *Gene* 62:165-9.

Henikoff, S. and L.Comai (1998) Trans-sensing effects:the ups and downs of being together. *Cell* 93:329-32.

Hepburn, A., Clarke, L., Pearson, L. and J.White (1983) The role of cytosine methylation in the control of nopaline synthase gene expression in plant tumour. *Journal of Molecular and Applied Genetics* 2:315-29.

Hernandez, N., Lobo, S., Lister, J., Ifill, S., Ratnasabapathy, R., Sheldon, M. and L.Johal (1990) transcription of the human U2 and U6 RNA genes. *Molecular Biology Reporter* 14:167

Herr, W., and J.Klerk (1986) The SV40 enhancer is composed of multiple functional elements that can compensate for one another. *Cell* 45:461-70.

Herr, A., Jensen, M., Dalmay, T. and C.Baulcombe (2005) RNA polymerase IV directs silencing of endogenous DNA. *eScience*

Hilderbrandt, M and W.Nellen (1992) Differential antisense transcription from the EB4 gene locus. implications on the antisense mediated regulation of mRNA stability. *Cell* 69:197-204.

Hightower, C. and B.Meagher (1989) Divergence and differential expression of soybean actin genes. *EMBO Journal* 4:1-8.

Hill, A., Nantel, A., Rock, C. and R.S.Quatrano (1996) A conserved domain of the *viviparous* 1 gene product enhances the DNA binding activity of the bZIP protein EmBP-1 and other transcription factors. *Journal of Biological Chemistry* 271:3366-74.

Hinchee, M., Newell, C., Sato, S., Gasser, C., Fischhoff, D., Fraley, R. and R.B.Horsch (1988) Production of transgenic soybean plants using Agrobacterium-mediated DNA transfer. *Bio/Technology* 6:915-22.

Holland, M., Abramson, D., Watson, R. and H.Gelfand (1991) Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the national Academy of Sciences* 88:7276-80.

Holstege, C.P., Jennings, E., Wyrick, J., Lee, T., Hengartner, C., Golub, T., Lander, S. and R.A.Young (1998) Dissecting the regulatory circuitry of a eukaryotic geneome. *Cell* 95:717-28.

Holstege, C.P., and R.A.Young (1999) Transcriptional regulation: contending with complexity. *Proceedings of the National Academy Of Sciences* 96:2-4.

Holstege, P. and H.Clevers (2006) Transcription factor target practice. *Cell* 124:21-23.

Hope, A., Mahadevan, S. and K.Struhl (1988) Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* 333:635-40.

Hori, H., Lim, B. and S.Osawa (1984) Evolution of green plants as deduced from 5s rRNA sequences. *Proceedings of the National academy of Sciences* 82:820-3.

Huala, E. and M.sussex (1992) LEAFY interacts with floral homeotic genes to regulate Arabidopsis floral development. *Plant Cell* 4:901-8

Huang, Y. and R.Maraia (2001) Comparison of Pol III transcription machinery in S.pombe, S.cerevisiae and human. *Nucleic acids Research.* 29:2675-90.

Hughes, D. and G.Galau (1988) Preparation of RNA from cotton leaves and pollen. *Plant Molecular Biology Reporter* 6:253-7.

Huie A., Scott W., Drazinic M., Lopez C., Hornstra K., Yang P.and V. Baker (1992) Characterization of the DNA-binding activity of GCR1: In vivo evidence for two GCR1-binding sites in the upstream activating sequence of TPI of Saccharomyces cerevisiae *Molecular and Cellular Biology.* 12:2690-700.

Hunold, R., Bronner, R. and G.Hahne (1994) Early events in microprojectile bombardment:cell viability and particle location. *The Plant Journal* 5:593-604.

Hutvagner G and P. Zamore (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297:2056-60

Iida, A., Seki, M., Kamada, M., Yamada, Y. and M.Morikawa (1990) Gene delivery into cultured plant cells by DNA coated gold particles accelerated by gas pressure driven gene gun. *Theoritical and Applied Genetics* 80:813-6.

Ishida, Y., Saito, H., Komari, T. and T.Kumashiro (1996) High efficiency transformation of maize (*Zea mays*) mediated by *Agrobacterium tumefaciens. Nature and Biotechnology* 14:745-50.

Iwahura, H. and M.Itakura (1997) An end trimming method and its application to amplify adjascent cDNA and genomic DNA fragments by PCR. In *Methods in molecular biology Vol 67* Ed. A.White, Humana Press, NJ. pp 490.

Iyer, V. and K.Struhl (1996) Absolute mRNA levels and transcriptional initiation rates in *Saccharomyces cerevisiae Proceedings of the National Academy of Sciences* 93:5208-12.

Izant, J. and H.Weintraub (1984) Inhibition of thymidine kinase gene expression by antisense RNA: a molecular approach to genetic analysis. *Cell* 36:1007-15.

Izant, J. and H.Weintraub (1985) Constitutive and conditional suppression of exogenous and endogenous genes by antisense RNA. *Science* 229:345-52.

Izquierdo, Hamdorf, B, Sabridge, T., Teasdale, R. and A.Mouradov (1997) Genbank direct submission

Jacob, S.T. (1995) Regulation of ribosomal gene transcription. *Biochemical Journal* 306:617-26

Javahery, R., Khachi, A., Lo, K., Zenzie, B. and S.Smale (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Molecular and Cellular Biology* 14:116-27.

Jaynes, J. and H.OFarrell (1991) Active repression of Transcription by the engrailed homeodomain protein. *EMBO Journal* 10:1427-33.

Jefferson, R. (1987) Assaying chimeric genes in plants. The GUS gene fusion. *Plant Molecular Biology Reporter* 5:387-405.

Jefferson, A., Kavanagh, A. and W.Bevan (1987) GUS fusion: glucouronidase as a sensitive marker in higher plants. *EMBO Journal* 6:3901-7.

Jennings, A. and L.Molloy (1987) Inhibition of SV40 replicon function by engineered antisense RNA transcribed by RNA Polymerase III. *EMBO Journal* 6:3043-7.

John, A., Birnsteil, L. and K.W.Jones (1969) RNA-DNA hybrids at the cytological level. *Nature* 223:582-7.

Johnson, A.D. (1995) The price of repression. *Cell* 81:655-8.

Jones, M.H., Learned, M. and R.Tjian (1988) Analysis of clustered point mutations in the human ribosomal RNA gene promoter by transient expression *in vivo. Proceedings of the National Academy of Sciences* 85:669-73.

Jones, H. and C.Winistorfer (1997) Amplification of a human genomic DNA flanking a known site using panhandle PCR variant. *Biotechniques* 23:132-8.

Jones, F., Echt, S., Gardner, RC and S.Cato (2001) Genbank direct submission

Joshi, P., Zhou, H., Huang, X. and L.Chiang (1989) Context sequence of translation inition codon in plants. *Plant molecular Biology* 35:993-1001.

Kaczmarek, W. and J.augustyniak (1994) An attempt at application of tRNA genes as promoters of protein coding genes in transgenic plants. *Acta Biochemicha Poland* 41:125-6.

Kahl, G. (1995) Dictionary of gene technology. VCH, FRD. 550p

Kahle, A., Feng, Y. and M.Nelson (2005) Isolation and characterization of the human Cdc2LI gene promoter. *Gene*

Kanno, M., Frometal, C., Davidson, I. and P.Chambon (1989) The SV40 TC-II (xB) and the related H-2K enhansons exhibit different cell type specific and inducible protoenhancer activities, but the SV40 core sequence and the AP-2 binding site have no enhanson properties. *EMBO Journal* 8:4205-14.

Karagyozov, L., Kalcheva, D. and M.Chapman (1993) Construction of random small insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Research* 21:16:3911-2.

Kassavetis, G.A, Riggs, L., Negri, R., Nguyen, R. and P.Geiduscher (1989) Transcription factor IIIB generates extended DNA interaction in RNA polymerase III transcription complexes on tRNA genes. *Molecular and Cellular Biology* 9:2551-66.

Kassavetis, G.A, Braun, R., Nguyen, R. and P.Geiduscher (1990) *S.cerevisiae* TFIIIB is the transcription initiation factor proper of RNA polymerase II while TFIIIA and TFIIIC are assembly factors. *Cell* 60:235-45.

Kassavetis, A., Bartholomew, B., Blanco, A., Johnson, E. and P.Geiduscheck (1991) Two essential components of the *Saccharomyces cerevisiae* transcription factor IIIB: transcription and DNA binding properties. *Proceedings of the National academy of Sciences* 88:7308-12.

Kaufman, L. and A.Evans (1990) Restriction endonuclease cleavage at the termini of PCR products. *Biotechniques* 9:304-5.

Kaufmann, J. and S.T.Smale (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes and Development* 8:7:821-9.

Kawasaki, H. and K.Taira (2003) Short hairpin type of dsRNAs that are controlled by tRNA-Val promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. *Nucleic Acids Research* 31:700-7.

Keegan, L., Gill, G. and M.Ptashne (1986) Separation of DNA binding from the transcription activating function of eukaryotic regulatory protein. *Science* 231:699-704.

Keller R, Renz FS and J.Kossmann (1999) Antisense inhibition of the GDP-mannose pyrophosphorylase reduces the ascorbate content in transgenic plants leading to developmental changes during senescence. *Plant Journal* 19:131-41

Keohavong, P., Wong,C., Cha, S. and G.Thilly (1988) Enzymatic amplification and characterization of large DNA fragments from genomic DNA. *Gene* 71:211-6.

Kerstetter, R., Volbrecht, E., S.Hake (1994) sequence analysis and expression patterns divide the maize homeobox genes into two classes. *Plant Cell 6:1877-87.*

Khochbin, S. and J.Lawrence (1989) An antisense RNA involved in p53 mRNA maturation in murine erythroleukemia cells induced to differentiate. *EMBO Journal* 8:4107-14.

Kim, S.K. and B.J.Wold (1985) Stable reduction of thymidine kinase activity in cells expressing high levels of antisense RNA. *Cell* 42:129-38.

Kim, Y., Li, Y., Sayre, M. and D.Kornberg (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 77:599-608.

Kim, V. (2003) RNA interference in functional genomics and medicine. *Journal of Korean Medical Science* 18:309-18.

King, G. (2002) Through a genome darkly:comparative analysis of plant chromosomal DNA. *Plant Molecular Biology* 48:5-20.

Kirst, M., Sederoff, R., *et al* (2003) Apparent homology of expressed genes from wood forming tissues of loblolly pine with arabidopsis thaliana. Proceedings of the National Academy of Sciences 100:7383-8.

KlausnerD., Roualt, A. and B.Harford (1993) Regulating the fate of mRNA. The control of cellular iron metabolism. *Cell* 72:19-28.

Klann, E., Hall, B. and A.B.Bennett (1996) Antisense acid invertase (TIV1) gene alters soluble sugar composiiton and size in transgenic tomato fruit. *Plant Molecular Biology* 112:1321-30.

Klein, T.M., Wolf, E., Wu, R. and J.C.Sanford (1987) High velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327:70-3.

Knowlton, A.A. (1995) Current concepts in transcription, translation and regulation of gene expression. *Chest* 107:241-8.

Koenig, R. (1997) Magnetic bead technology aids sequence determinations for the 3 and 5 ends of viral RNA. *Biotechniques* 23:102-4.

Koes, R., Quattrocchio, F. and J.Mol (1994) The flavonoid biosynthetic pathway in plants:Function and evolution *Bioessays* 16:123-32.

Kohchi, T., Takemura, M., Nishi, A., Kami, C. and A.Yokota (1998) Functional analyses of genes isolated as possible meristem specific genes from a normalized cDNA library. *Proceedings of the 9th International Conference on Arabidopsis research, 1998.*

Kohler, R., Zipfel, W., Webb, W. and M.Hanson (1997) The green fluorescent protein as a marker to visualize plant mitochondria in vivo. *Plant Journal*11:613-21

Kolmodin, A. and F.Williams (1997) PCR In PCR cloning protocols. Ed. B.A.White Human Press, USA 490 pp.

Kooter, J. and J.Mol (1993) Trans-activation of gene expression in plants. *Current Opinions in Biotechnology* 4:166-71.

Kooter, J., Matzke, M. and P.Meyer (1999) Listening to the silent genes: Transgene silencing research identifies new mechanisms of gene regulation and pathogen control. *Trends in Plant. Science4:340-7*

Koski, A. and G.Clarkson (1982) Synthesis and maturation of *xenopus laevis* methionine tRNA gene transcripts in homologous cell-free systems. *Journal of Biological Chemistry* 257:4514-8

Kotelnikova, E., Makeev, V. and M.Gelfand (2005) Evolution of transcription factor DNA binding sites. *Gene (in press)*

Kozak, M. (1989) The scanning model for translation:an update. *Journal of cellular Biology* 108:229-41.

Kozak, M. (1991) An analysis of the vertebrate mRNA sequences: intimation of translation control. *Journal Of Cellular Biology* 115:887-903.

Kozak, M. (2003) Alternative ways to think about mRNA sequences and proteins that appear to promote internal initiation of translation. *Gene* 318:1-23.

Kozak, M. (2004) How strong is the case for regulation of the initiation step of translation by elements at the 3' end of eukaryotic mRNA. *Gene* 343:41-54.

Kramerov, A., Tillib, V., Shumayatski, P. and P.Geogive (1990) The most abundant nascent poly-A RNAs are transcribed by RNA polymerase III in murine tumour cells. *Nucleic acids Research* 18:4499-506.

Krasilnikov AS, Panyutin IG, Samadashwily GM, Cox R, Lazurkin YS. And S.Mirkin (1997) Mechanisms of triplex-caused polymerization arrest *Nucleic Acids Res*earch 25:1339-46

Kuempel PL and S.Duerr (1979) Chromosome replication in Escherichia coli is inhibited in the terminus region near the rac locus *Cold Spring Harbour Symposium on Quantitative Biology* 43 Pt 1:563-7

Kulinska, Z., Jarmolowski, A. and M.Figlerowicz (2003) RNA interference and its role in the regulation of eukaryotic gene expression. *Acta Biochemicha Polonica* 50:217-29.

Kumar, A.,Grove, A., Kassavetis and P.Geiduschek (1998) Transcription factor IIIB: the architecture of its DNA complex, and its roles in initiation of transcription by RNA Pol III. *Cold Spring harbour Symposium,* Cold Spring Harbour Laboratory Press pp121-9.

Kwok, S., Kellog, E., Levenson,c, and J.Sninsky (1990) Effects of primer template mismatches on the polymerase chain reaction. Human immunodeficiency virus I model studies. *Nucleic Acids Research* 18:999-1005.

Lago, C., Clerici, E. and M.Kater (2004) TBP-associated factors in Arabidopsis. *Gene* 342:231-41

Lagrange, T., Kapanidis, A., Tang, H., Reinberg, D. and H.Ebright (1998) New core promoter elemnt in RNA polymerase II dependant transcription: sequence specific DNA binding by transcription factor IIB. *Genes and Development* 12:34-44.

Landschutze, V., Willmitzer, L. and B.M.Rober (1995) Inhibition of flower formation by antisense repression of mitochondrial citrate synthase in transgenic potato plants leads to disintegration of the ovary tissue of flowers. *EMBO Journal* 14:660-6.

Lapitan, V., Ganal, W. and D.Tanksley (1991) Organization of the 5S ribosomal RNA genes in the genome of tomato. *Genome* 34:509-14.

Larkin, J., Hunsperger, J., Cully, D. and C.Silflow (1989) The organization and expression of maize ribosomal protein gene family. *Genes and Development* 3:500-9.

Larkin, R. and T.Guilfoyle (1993) The second largest subunit of RNA Polymerase II from *Arabidopsis thaliana. Nucleic Acids Research* 21:4:1038.

Larson, D., Bradford-wilcox, J., Young, I. and K.U.Sprague (1985) Drosophila genetics *Proceedings of the National Academy of sciences* 80:3416-20.

Last, D., Chamberlain, D., Chaudry, M., Larkins, P., Marsh, L. and E.S.Dennis (1991) pEmu: an improved promoter for gene expression in cereal cells. *Theoritical and Applied Genetics* 81:581-8.

Lawrence, G. and R.Appels (1986) Mapping the nucleolus organizer region, seed protein loci and isozyme loci on chromosome IR in rye. *Theoritical and Applied Genetics* 71:742-9.

Leblanc, B., Read, C. and T.Moss (1993) Recognition of the *xenopus* ribosomal core protein by the transcription factor xUBF involves multiple HMG box domains and leads to an xUBF interdomain interaction. *EMBO Journal* 12:513-25.

Lee, M.S, Gippert, P., Soman, V. and E.Wright (1989) Three-dimension solution structure of a single zinc-fingerDNA binding domain. *Science* 245:635-7.

Lee, Y. and V.Vacquier (1992) Reusable cDNA libraries coupled to magnetic beads. *Analytical Biochemistry* 206:207-10.

Lee, Y., Nakahara, J., Kim, K. and R.Carthew (2004) Distinct roles for Drosophila dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117:69-81.

Lehrach, H., Diamond, D., Wozney, M. and H.Boedtker (1977) Formaldehyde denatured gels for *Biochemistry* 16:4743.

Levi, A., Galau, G. and H.Y.Wetzstein (1992) A rapid procedure for the isolation of RNA from high phenolic containing tissues of pecan. *Hortscience* 27:1316-8.

Lewin, B. (1997) Genes VI. Oxford University Press, Oxford. 1366 p.

Lewis, E. and L.Manley (1986) Polyadenylation of a mRNA precursor occurs independent of transcription by RNA Polymerase II in vivo. *Proceedings of the National Academy of Sciences* 83:8555-9.

Li, H., Wu, L. and C.C.Luo (1986) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon usage. *Molecular Biology of Evolution* 2;150-74.

Liao, D (1999) Concerted evolution:molecular mechanism and biological implications. *American Journal of Human Genetics* 64:24-30.

Liu, Y., Mitsukawa, N. Ooosumi, T. and F.whittier (1995) Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *The Plant Journal* 8:457-463.

Liu,Z., Zhang,D., Wang,X., Ma,X. and X.Wang (2003) Intragenomic and interspecific 5S rDNA sequence variation in five Asian pines. American Journal of Botany 90:17-24

Liu Z., Zhang D, Hong DY, Wang XR (2003[A]) Chromosomal localization of 5S and 18S-5.8S-25S ribosomal DNA sites in five Asian pines using fluorescence in situ hybridisation. *Theoretical and Applied Genetics* 106:198-204.

Llave, C., Kasschau, K. and C.Carrington (2000) Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. *Proceedings of the Nationall Academy of Sciences* 97:13401-6

Llewellyn, D. (2003) CSIRO *Pers. Comm.*

Lloyd, A., Walbolt, V. and R.W.Davies (1992) *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1*. *Science* 258:1173-5.

Lo, C., Wang, N. and E.Lam (2005) Inducible double stranded RNA expression activates reversible transcript turnover and stable translational suppression of a target gene in transgenic tobacco. *FEBS letters* 579:1498-1502.

Lobo, M. and N.Hernandez (1989) A 7 bp mutation converts a human RNA Polymerase II snRNA promoter into an RNA Polymerase III promoter. *Cell* 58:55-67.

Lockhart, J. and A.Winzeler (2000) Genomics, gene expression and DNA arrays *Nature* 405:827-36.

Logie, C. and C.L.Peterson (1997) Catalytic activity of the Yeast SWI/SNF complex on reconstituted nucleosome arrays. *EMBO Journal* 16:6772-82.

Loh EY, Elliott JF, Cwirla S, Lanier L. and M.Davis (1989) Polymerase chain reaction with single-sided specificity: analysis of T cell receptor delta chain *Science* 243:217-20

Long, O. and B.Dawid (1980) Repeated genes in eukaryotes. *Annual Review of Biochemistry* 49:727-64.

Loopstra, A., Mouradov, A., Glassik, V. and D.Teasdale (1998) Genbank direct submission

Lopata, A., Cleveland, W. and B.Sollner-Webb (1986) RNA polymerase specificity of mRNA production and enhancer action. *Proceedings of the National Academy of Sciences.* 83:6677-81.

Lopez Gomez, R. and M.A.Gomez-lim (1992) A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango mesocarp. *Hortscience* 27:440-2.

Lubaretz , J. Fuchs , R. Ahne , A. and I. Schubert (1996) Karyotyping of three Pinaceae species via fluorescent in situ hybridization and computer-aided chromosome analysis. *Theoretical and Applied Genetics* 92:411-6.

Luehrsen, R. and E.Fox (1981) Secondary structure of eukaryotic cytoplasmic 5s ribosomal RNA. *Proceedings of the National Academy of Sciences* 78:2150-4.

Macchi, M, Bornet, J., Davidson, I., Fromental, C. and P.Chambon (1989) The SV40 enhanson binds ubiquituous and cell type specifically inducible nuclear proteins from lymphoid and nonlymphoid cell lines. *EMBO Journal* 8:4215-27.

Mackay, M., Spencer, F., Schnare, N. and W.Gray (1982) Comparative sequence analysis as an approach to evaluating structure, function and evolution of 5S and 5.8S ribosomal RNAs. *Canadian Journal of Biochemistry* 60:4:480-9.

Maizels, N. (1976) Dictyostelium 17S, 25S and 5S rDNAs lie within a 38,000 base pair repeated unit. *Cell* 9:431-8.

Majowski, K., Mentzel, H. and T.Pieler (1987) A split binding site for TFIIIC on Xenopus gene. *EMBO Journal* 6:3057-63.

Maldonado, E. and D.reinberg (1995) News on initiation and elongation of transcription by RNA Polymerase II. *Current Opinion in Cellular Biology* 7:352-6

Manning, K. (1990) Isolation of nucleic acids from plants by differential solvent precipitation. *Analytical Biochemistry* 195:45-50.

Marai, J. (2001) La Protein and the traficking of nascent RNA Pol III transcripts. *Journal of Cell Biology* 153:f13-17.

Marchuk, M., Drum, M., Saulino, A. and S.Collins (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research* 19:1154.

Marcotte, E., Pellegrini, M., Ng, H., Rice.W and D.Eisenberg (1999) Detecting protein function and protein-protein interactions from genome sequence. *Science* 285:751-3.

Margottini, F., Dujardin, G., Gerard, M., Egly, M., Huet, J. and A.Sentenac (1991) Participation of the TATA factor in transcription of the yeast U6 gene by RNA poymerase C. *Science* 251:424-6.

Martinez, E., Chiang, C., Ge, H. and R.G.Roeder (1994) TATA-binding protein associated facto in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO Journal* 13:13:3115-26.

Martinez, J., Urlaub, H., Luchrman, R. and T.Tuschi (2002) Single strand antisense siRNAs guide target RNA cleavage in RNAi. *Cell* 110:563-74.

Martinez-Trujillo, (2003) Functional analyses of the 5'untranslated region of the sucrose phosphate synthase rice gene *Plant Science* 165:9-20.

Mascia, P., Phillips, R. and L.Xiang (1981) Localization of the 5S rRNA genes and evidence for diversity in the 5S rDNA region of maize. *Gene* 15:7-20.

Mathieu, O., Jasencakova, Z., *et al* (2003) Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in arabidopsis. *The Plant Cell* 15:2929-2939.

Matsutani, S. (2004) Similarities in transcription factor IIIC subunits that bind to the posterior regions of internal promoters for RNA polymerase III. *BMC evolutionary Biology* 4:26-32

Mattaj, W., Dathan, A., Parry, D.Carbon, P. and A.Krol (1988) Changing the RNA Polymerase specificity of U snRNA promoters. *Cell* 55:435-42.

Mattaj, W. and L.Englemeier (1998) Nucleocytoplasmic transport. the soluble phase. *Annual Review of Biochemistry* 67:265-306.

Matzke, M. and J.Birchler (2005) RNAi-mediated pathways in the nucleus. *Nature Reviews Genetics* 6:24-35.

McCabe, D. and P.Christou (1993) Direct DNA transfer using electric discharge particle acceleration (ACCELL) technology. *Plant Cell, Tissue and Organ Culture* 33:227-36.

McCallum, A., Grant, E. and F.Moran (1997) Genbank direct submission

McDonell, W., Simon, N. and W.Studier (1977) Analysis of restriction fragments of T& DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *Journal of Molecular Biology* 110:119-46.

McDowell, J., Huang, S., McKiney, C., An, Y. and R.Meagher (1996) Structure and evolution of the actin gene family in *Arabidopsis thaliana*. *Genetics* 142:587-602.

McDowell, J., Huang, S., McKiney, C., An, Y. and R.Meagher (1996a) in *Arabidopsis thaliana* contains 10 actin genes encoding six ancient protein subclasses. *Genetics* 142:602-10.

Meagher, R., McKinney, C. and K.Muthugapatti (2000) The significance of diversity in the plant actin gene family. In Actin:*A dynamc framework for multiple plant cell functions.* Kluwer academic publishers 659 pp.

Melton, D. (1985) Injected antisense RNAs specifically block messenger RNA translation in vitro. *Proceedings of the National Academy of Sciences* 82:144-8

Merrick, C. (2004) Cap-dependant and cap-independent translation in eukaryotic systems. *Gene* 332:1-11.

Meshi, T. and M.Iwabuchi (1995) Plant transcription factors. *Plant Cell Physiology* 36:1405-20.

Messenguy, F. and E.Dubois (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* 316:1-21.

Meyer, P., Heidmann, I., Forkmann, G. and H.Saedler (1987) A new petunia flower color generated by transformation of a mutant by a maize gene. *Nature* 330:677-8.

Meyer, P. and I.Heidmann (1994) Epigeneitic variants of a transgenic petunia line showing hypermethylation in transgenic DNA:an indication for specific recognition of foreign DNA in transgenic plants. *Molecular and General Genetics* 243:390-9.

Miksche, P. (1985) Recent advances of Biotechnology of forest trees. *Forestry Chronicles* 61:449-53.

Mizobuchi, M. and A.Frohman (1993) Rapid amplification of genomic DNA ends. *Biotechniques* 1%:215-6.

Mizuno, T., Chou, M. and M.Inouye (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript. *Proceedings of the National Academy of Sciences* 81:1966-70.

Mol, J., Van der Krol, J., Tunen, R., Blokland, P., De Lange, P. and A.R.Stuitje (1990) Regulation of plant expression by antisense RNA. *FEBS* 2:427-30.

Monier, R., Stephenson, L. and C.Zamecknick (1960) The preparation and some properties of of a low molecular weight RNA from Baker's yeast. *Biochimica Biophysica Acta* 43:1.

Moran, F., Smith, D., Bell, C. and R.Appels (1992) The 5S rNA genes in Pinus radiata and the spacer regions as a probe for relationship between Pinus species. *Plant Systematic Evolution* 103:209-21.

Moreau, P., Hen, R., Everrett, R. and P.Chambon (1981) The SV 40 72 basepair repeat has a striking effect on gene expression both in SV 40 and other chimaeric recombinants. *Nucleic Acids Research* 9:6047-68.

Morreira, P. and S.Noren (1995) Minimum duplex requirements for restriction cleavage near terminin of linear DNA fragments. *Biotechniques* 19:130-8.

Morton, D. and K.U.Sprague (1984) In vitro transcription of a silkworm 5S RNA gene requires an upstream signal. *Proceedings of the National Academy of Sciences* 81:5519-22.

Morzycka-Wroblewska, E., Selker, E., Stevens, J. and R.Metzenberg (1985) Concerted evolution of dispersed *Neurospora crassa* 5S RNA genes: pattern of sequence conservation between allelic and nonallelic genes. *Molecular and Cellular Biology*5:46-51.

Muller, M., Gerster, T. and W.Schaffner (1988) Enhancer sequences and the regulation of transcription. *European Journal of Biochemistry* 176:485-95.

Mullis, K.B. (1990) The unusual origin of the polymerase chain reaction. *Scientific American.* 262:4:56-65.

Murphy, S., Moorfield, B. and T.Pieler (1989) Common mechanisms of promoter recognition by RNA polymerase II and III. *Trends in Genetics* 5:4:122-6.

Murray, G. and F.Thompson (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8:4321-5.

Murray, V. (1989) Improved double stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Research* 17:21-3.

Myers, C., Gustafsson, C., Hayashibara, K., Browns, P. and R.D.Kornberg (1999) Mediator protein mutations that selectively abolish activated transcription. *Proceedings of the National Academy of Sciences*96:67-72.

Myers,V. and R.A.Young (1998) RNA Polymerase II holoenzyes and subcomplexes. *Journal of Biological Chemistry* 273:27757-60.

Napoli, C., Lemieux, C. and R.Jorgenson (1990) Introduction of a chimaeric chalcone synthase gene in petunia results in reversible cosuppression of homologous genes. *Plant Cell* 2:279-89.

Narasimhulu, S., Deng, X., Sarria, R. and S.B.Gelvin (1996) Early transcription of *Agrobacterium* T-DNA genes in tobacco and maize. *Plant Cell* 8:873-86.

Nazarenus, T., Cedarberg, R., Forch, A. and A.Atkin (2005) Upf1P a highly conserved protein required for nonsense-mediated mRNA decay interacts with the nuclear proteins Nup100p and Nup116p. *Gene* 345:199-212.

Nellen, W. and C.Lichtenstein (1993) What makes an mRNA anti-sensitive. *Trends in Biochemical Science* 18:419-23.

Nelson, D., Linning, R. and B.Honda (1998) 5-flanking sequences required for efficient transcription *in vitro* of 5S RNA genes in the related nematodes *C.elegans and C.briggsae* 218:9-16.

Nielsen, N., Hallenbrg, C., Friedrikson, S., Sorensen, D. and B.Lomholt (1993) Transcription of human 5S rRNA genes is influenced by an upstream DNA sequence. *Nucleic Acids Research* 21:16:3631-6.

Nikolov, D., Hu, S., Lin, J., Gasch, A., Hoffman, A. and S.Burley (1992) Crystal structure of TFIID TATA box binding protein. *Nature* 360:40-6.

Nims, R.C., Halliwell, R. and D.W.Rosberg (1967) Wound healing in cultured tobacco cells following microinjection. *Protoplasma* 64:305-14.

Nomura, M. (1976) Organization of bacterial genes for ribosomal components: studies using novel approaches. *Cell* 9:633-44.

Noury M, Bassie L, Lepri O, Kurek I, Christou P. T.Capell (2000) A transgenic rice cell lineage expressing the oat arginine decarboxylase (adc) cDNA constitutively accumulates putrescine in callus and seeds but not in vegetative tissues. *Plant Molecular Biology* 43:537-44

Oard, J. (1993) Development of an airgun device for particle bombardment. *Plant Cell, Tissue and Organ culture* 33:247-50.

Ochman, H., Gerber, S. and L.Hartl (1988) Genetic application of an inverse polymerase chain reaction. *Genetics* 120:621-5.

O'Connor, P. and L.Peebles (1991) In vivo pre-tRNA processing *in Saccharomyces cerevisea. Molecular and Cellular Biology* 11:425-39.

Odell, J., Nagy, F. and N.Chua (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus. *Nature* 313:810-2

Oei, S. and T.Pieler (1990) A transcription stimulatory factor binds to the upstream region of *Xenopus* 5S RNA and tRNA genes. *Journal of Biological Chemistry* 265:11:7485-91.

Oh, K., Scott, P. and P.Sarnaw (1992) Homeotic gene Antepedia mRNA contains 5-noncoding sequences that confer translation initiation by internal ribosomal binding. *Genes and development* 6:1643-53.

Ohara, O., Dorit, L. and W.Gilbert (1989) One sided polymerase chain reaction: the amplification of cDNA. *Proceedings of the National Academy of Sciences* 86:5673-7.

Okamuro, J., Caster, B., Montagu, M. and K.D.Jofuku (1997) The AP2 domain of *APETALA2* defines a large new family of DNA binding proteins in *Arabidopsis. Proceedings of the National Academy of Sciences* 94:7076-81.

O'Neil, T., Hoess, H. and F.DeGrado (1990) Design of DNA-binding peptides based on the leucine zipper motif. *Science* 249:774-8.

Ouwerkerk, P., Hallard, D., Verpoorte, R. and J.Memelink (1999) Identification of UV-B light responsive regions in the promoter of tryptophan decarboxylase gene from *Catharanthus roseus. Plant Molecular Biology* 41:491-503.

Overbeek, R., Fonstein, M., Pusch, G. and N.Maltsev (1999) The use of gene clusters to infer functional coupling. *Proceedings of the Nataional Academy of Sciences* 96:2896-901.

Ow, D., Jacobs, J. and S.H.Hall (1987) Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporterof promoter activity. *Proceedings of the National Academy of Sciences* 84:4870-4.

Pabo, C.O. and R.T.Sauer (1992) Transcription factors: structural families and principles of DNA recognition. *Annual review of Biochemistry* 61:1053-95.

Padolfini, T., Molesini, B., Avesani, L. and A.Polverari (2003) Expression of self complementary hairpin RNA under the control of the RolC promoter confers systemic disease resistance to plumpox virus without preventing local infection. *BMC Biotechnology* 3:7-16.

Palmer, J. and W.Folk (1987) Isolation and sequence analysis of a nuclear tRNA-Met I gene from soybean. *Plant Molecular Biology* 8:47-51.

Papp, I., Mett, M., *et al* (2003) Evidence or nuclear processing of plant microRNA ans short interfering RNA precursors. *Plant Physiology* 132:1382-92

Pardue,M.L. and J.G.Gall (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proceedings of the National Academy of Sciences* 64:600-4.

Park, J. Choi, I., Kang, S., Lee, J., Hatfield, D. and B.Lee (1995) Upstream promoter elements are sufficient for selenocysteine tRNA[Ser]Sec gene transcription and to determine the transcription start point Gene 162:13-9.

Parvin, J., Shykind, B., Meyers, R., kim, J. and P.A.Sharp (1994) Multiple sets of basal factors initiate transcription by RNA Polymerase II. *Journal of Biological Chemistry* 269:28:18414-21.

Pavesi, A., Conterio, F., Bolchi, A. and S.Ottonello (1994) Identification of new tRNA genes in genomic DNA databases by a multistep weight matrix analysis of transcriptional control regions. *Nucleic Acids Research* 22:1247-56.

Paul CP, Good PD, Winer I, D.Engelke (2002) Effective expression of small interfering RNA in human cells.*National Biotechnology* 20:505-8.

Paul C (2005) Subcellular distribution of small interfering RNA: directed delivery through RNA polymerase III expression cassettes and localization by in situ hybridization. *Methods in Enzymology*. 392:125-45.

Peabody, S., Subramaniam, S. and P.Berg (1986) Efficiency of upstream reading frameson translation efficiency in simian virus 40 recombinants. *Molecular and Cellular Biology* 6:2704-11.

Pelligrini, M., Marcotte, E., Eisenberg, D. and O.Yeats (1999) Assigning protein functions by comparative genome analysis. Protein phylogenetic profiles. *Proceedings of the National Academy of Sciences* 96:4285-8.

Perry, P. (1976) transcription factors and gene activation *Annual Review of Biochemistry* 45:605-29.

Peterson, K., Leah, R. and V.Cameron (2002) Matrix attachment regions enhance transformation frequencies and reduce variance of transgene expression in barley. *Plant Molecular Biology* 49:45-58.

Picard, D. and Schaffner (1984) A lymphocyte-specific enhancer in the mouse immunoglobulin *k* gene. *Nature* 307:80-2.

Pikaard, C. (1993) Ribosomal gene promoter domains can function as artificial enhancers of RNA polymerase I transcription, supporting a promoter origin for natural enhancers in Xenopus. *Proceedings of the National Academy of Sciences* 91:464-8

Pieler, T., Oei, S., Hamm, J. and V.Erdmann (1985) Functional domains of the *Xenopus laevis* 5S gene promoter. *EMBO Journal* 4:3751-6.

Pieler, T., Hamm, J. and R.G.Roeder (1987) The 5S RNA gene internal control region is composed of three distinct sequence elements organized as two functional domains with variable spacing. *Cell* 48:91-100.

Pikaard, C.S. (1994) Ribosomal gene promoter domains can function as artificial enhancers of RNA polymerase I transcription, supporting a promoter origin from natural enhancers in *xenopus*. *Proceedings of the National Academy of Sciences* 91:464-8.

Piras, G., Kashanchi, F., Radonovich, F. and N.Brady (1994) Transcription of the human T-cell lymphotropic virus type I promoter by an a-amanitin resistant polymerase. *Journal of Virological Methods* 68:6170-9.

Potrykus, I. and G.Spangenburg (1996) Gene transfer to plants. Springer-Verlag, Berlin.

Promega (1996) Wizard plus DNA purification system. Revised technical bulletin.

Pruit, E. and M.Myerowitz (1986) Characterization of the genome of *A.thaliana. Journal of molecular Biology* 187:169-83.

Ptashne, M. (1989) How gene activators work. *Scientific American* 260:24-31.

Ptashne, M. and A.Gann (1997) Transcriptional activation by recruitment. *Nature* 386:569-77.

Puchta, H. (2002) Gene replacement by homologous recombination in plants. *Plant Molecular Biology* 48:173-82.

Pusch, O., Boden, D., Silbermann, R., Lee, F., Tucker, L. and B.Ramratnam (2003) Nucleotide sequence homology requirements of HIV-1-specific short hairpin RNA. Nucleic Acids Research 31:6444-9.

Putnam, C., Copenhaver, G., Denton, M. and C.S.Pikaard (1994) The RNA polymerase I transactivator upstream binding factor requires its dimerization domain and high-mobility-group (HMG) box 1 to bend, wrap and positively supercoil enhancer DNA. *Molecular and Cellular Biology* 14:6476-88.

Putterill, J. and C.Gardner (1989) Initiation of translation of the b-glucouronidase reporter gene at internal AUG codon in plant cells. *Plant Science* 62:199-205.

Quan, T. and J.Fisher (1999) Cloning and characterization of the human protein kinase C-eta promoter. *Journal of Biological Chemistry* 274:40:28566-74.

Quandt, K. Frech, K. Karas, H. Wingender, E. and T. Werner (1995) MatInd and MatInspector - New fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Research* 23, 4878-84.

Queen, C. and D.Baltimore (1983) Immunoglobulin gene transcription is activated by downstream sequence elements. *Cell* 33:741-8.

Rafalski, J., Weiwiorowski, M. and D.Soll (1982) Organisation and nucleotide sequence of nuclear 5S RNA genes in yellow lupin. *Nucleic Acids Research* 10:7635-42.

Ramakrishnan, V. and W.White (1998) Ribosomal protein structures. *TIBS* 23:208-12.

Ramamonjisoa, D., Kauffmann, S., Choisne, N. and A.Dietrich (1998) Structure and expression of several bean nuclear transfer RNA genes:relevance to the process of tRNA import into plant mitochondria *Plant molecular Biology* 36:613-25.

Reddy, R. (1988) Transcription of a U6 small nuclear RNA gene in vitro. Transcription of a mouse u6 small nuclear RNA gene in vitro by RNA polymerase III is dependant on transcription factors different from transcription factors IIIA, IIIB and IIIC. *Journal of Biological Chemistry* 263:31:15980-4.

Reddy, S. and D.Padyatty (1988) Effects of 5' flanking sequences and changes in the 5' internal control region on the transcription of rice tRNA-gly gene. *Plant Molecular Biology* 11:575-9.

Reichmann, J. and E.M.Meyerowitz (1998) AP2/EREBP family of plant transcription factors. *Biological Chemistry* 379:633-46.

Reinberg, D., Orphanides, G., Ebright, R., Akoulitchev, S., Carcamo, J., Cho, H., Cortes, P., Drapkin, R., Flores, O., Ha, I., Inostroza, J., kim, T., Kumar, P., Olave, I., Sheldon, M., Sheikhattar, R., Stone, N., Sun, X., Weis, L., Yeung, K. and L.Zawel (1998) The RNA Polymerase II transcription factors: Past, Present and future. *Cold Spring Harbour Symposium on Quantitative Biology* 83-103.

Reiter, R (2001) PCR based marker systems. In DNA based markers in plants 9-30. Ed: L.Phillips and K.Vasil Kluwer Academic Publishers, UK. 510 pp.

Rhees, S., Martin, G., Rosner, L. and D.R.David (1998) A novel DNA binding motifin Mar A:the first structure for an AraC family transcriptional activation. *Proceedings of the National Academy of Science* 95:18:10413-8.

Roa-Rodriguez, C. (2003) *Promoters used to regulate gene expression.* Cambia intellectual property resource, Canberra, Australia.

Robinett, C., Connor, A. and M.Dunaway [1997] The repeat organizer within the intergenic spacer of Xenopus rRNA genes. *Molecular and Cellular Biology* 17:2866-75.

Roeder, R (1996) Nuclear RNA polymerases: Role of general initiation factors and cofactors in eukaryotic transcription. *Methods in Enzymology* 273:165-71.

Rosorius, O., Fries, B., Stauber, H. and J.Haubert (2000) Human ribosomal protein L5b contains defined nuclear localization and export signals. *Journal of Biological Chemistry* 275:12061-8.

Rubin, M. and E.Sulston (1973) Physical linkage of the 5S cistrons to the 18S and 28S ribosomal RNA cistrons in *S.cerevisiae*. *Journal of Molecular Biology* 79:521-30.

Rubin, M., Yandell,D., Wortman, R. *et al* (2000) Comparative genomics of the eukaryotes. *Science* 287:2204-15.

Rudenko, N., Rommens, M. and J.Hille (1993) Supported PCR: an efficient procedure to amplify sequences flanking a known DNA segment. *Plant Molecular Biology* 21:723-8.

Rudi, K., Fosseim, T. and S.Jakobsen (1999) Restriction cutting independent method for cloning genomic DNA segments outside the boundaries of known sequences. *Biotechniques* 27:1170-7.

Russell, D., Wallace, K., Bathe, J., Martinell, B. and D.E.McCabe (1993) Stable transformation of *Phaseolus vulgaris* via electric discharge mediated particle acceleration. *Plant Cell Reports* 12:165-9.

Russell, J., Roy, M. and J.C.Sanford (1992) Physical trauma and tungsten toxicity reduce the efficiency of biolistic transformation. *Plant Physiology* 98:1050-6.

Rychlik, W. and E.Rhaodes (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Research* 17:8543-51.

Saiki, K., Scharf, S., Faloona, K., Mullis, B. Erlich, A. and N.Arnheim (1985) Enzymatic amplification of β-heamoglobin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *science* 230:1350-4.

Saiki, K., Gelfand, H., Stoffel, J., B.Mullis and A.Erlich (1988) Primer directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science* 259:487-91.

Sakonju, S., Bogenhagen, F. and D.D.Brown (1980) A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5' border of the region. *Cell* 19:13-25.

Sakurai, H. and A.Ishihama (2001) Transcription organization and mRNA levels of the genes for all 12 subunits of the fission yeast RNA Polymerase II. *Genes to Cells* 6:1:25-36.

Samac, D. and D.M.Shah (1994) Effect of Chitnase antisense RNAexpression on disease susceptibility of *Arabidopsis* plants. *Plant Molecular Biology* 587-96.

Sambrook, J., Fritsch, E. and T.Maniatis (1996) Molecular Cloning. Cold Spring Harbour Press.

Samadashwily, G., Raca, G. and M.Mirkin (1997) Trinucleotide repeats affect DNA replication *in vitro*. *Nature and Genetics* 17:298-304.

Sanford, J., Smith, F. and J.Russel (1993) *Methods in enzymology* 217:483-509.

Sarkar, G. and S.sommer (1991) Parameters affecting the susceptibility of PCR contamination to UV inactivation. *Biotechniques* 10:589-94.

Sastri, D., Appels, R. and B.Baum (1992) An overview of evolution in plant 5S DNA. *Plant systems and evolution* 183:169-81.

Schaack J and D. Soll (1985) Transcription of a Drosophila tRNAArg gene in yeast extract: 5'-flanking sequence dependence for transcription in a heterologous system *Nucleic Acids* 13:2803-14.

Schindler, U., Terzhagi, W., Beckman, H., Ecker, J. and A.R.Cashmore (1992) reversal of a standard paradigm *EMBO Journal* 11:1275-89.

Schmidt, R. (2002) Plant genome evolution:lessons from comparative genomics at the DNA level. *Plant Molecular Biology* 48:21-37

Schnapps, A., Clos, J., Hadell, W., Schreck, R., Cveck, A. and I.Grummt (1999) Isolation and functional characterization of TIF-IB, a factor that confers promoter specificity to mouse RNA polymerase I. *Nucleic Acids Research* 18:6:1385-93.

Schneiderbauer, A., Sandermann, H. and D.Ernst (1991) Isolation of functional RNA from plant tissues rich in phenolic compounds. *Analytical Biochemistry* 197:91-5.

Schneeberger, G., Creissen, P. and A.Cullis (1989) Chromosomal and molecular analysis of 5S RNA gene organization in flax. *Gene* 83:75-84.

Schueller C., Brewster L., Alexander R., Gustin C. and H. Ruis (1992)The HOG pathway controls osmotic regulation of transcription via the stress response element (STRE) of the Saccharomyces cerevisiae CTT1 gene

Schultz, J., Craig, R., Foster, D., Mumma, R. and J.I.Medford (1994) RNA Isolation from recalcitrant plant tissues *Plant molecular biology reporter* 12:4:310-6.

Schwechheimer, C., Smith, C. and M.W.Bevan (1998) The activities of acidic and glutamine-rich transcriptional activation domains in plant cells:design of modular transcription factors for high-level expression. *Plant Molecular Biology* 36:195-204.

Scoles, G., Gill, B. and R.Appels (1988) Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of *triticeae. Plant systematics and Evolution* 160:105-22.

Scott, P., Tamkun, W. and W.Hartzell (1989) The structure and function of the homeodomain. *Biochemica Biophysica Acta* 989:25-48.

Screaton, G., Bangham, C. and J.I.Bell (1993) direct sequencing of single primer PCR products: a rapid method to achieve short chromosomal walks. *Nucleic Acids research* 21:9:2263-4.

Seibert PD, Chenchik A, Kellogg DE, Lukyanov KA. and S.Lukyanov (1995) An improved PCR method for walking in uncloned genomic DNA *Nucleic Acids Research* 23:1087-8

Selker EU, Morzycka-Wroblewska E, Stevens JN, R.Metzenberg (1986) An upstream signal is required for in vitro transcription of Neurospora 5S RNA genes Molecular and General Genetics 205:189-92

Sharp, S., DeFranko, D., Schmidt, T., Wensink, C. and D.Soll (1981) The initiator tRNA genes of *Drosophila Melanogaster* evidence for a tRNA pseudogene. *Nucleic Acids research* 9:5867-74.

Sharp, S., Garcia, A., Cooley, L. and D.Soll (1984) Transcriptionally active and inactive gene repeats within the D.melanogaster 5S RNA gene cluster. *Nucleic Acids Research* 12:20:7617-32.

Sharp, J., Schaack, J., Burke, J. and D.Soll (1985) Structure and transcription of eukaryotic tRNA genes. *CRC Critical Review of Biochemistry.* 19:107-144.

Sharp, S. and D.Garcia (1988) Transcription of Drosophila melanogaster 5S RNA gene requires an upstream promoter and four intragenic sequence elements. *Molecular and Cellular Biology* 8:1266-74.

Shen, W., Escudero, J., Ramos, C., Hohn, B. and Z.Koukolikovanicola (1993) TDNA transfer to maize cells:histochemical investigation of glucouronidase activity in maize tissues. *Proceedings of the National Academy of Sciences*90:1488-92.

Shen, W.C. and M.R.Green (1997) Yeast TAF(II)145 functions as a core promoter selectivity factor, not a general coactivator. *Cell* 90:615-24.

Shillito, R., Saul, M., Muller, M. and I.Potrykus (1985) High efficiency direct gene transfer to plants. *Bio/Technology* 3:1099-103.

Shimada, T., Fuji, H. and H.Lin (1989) A 165 bp sequence between DHFR and the divergentlytranscribed upstream gene is sufficient for bidirectional transcriptional activity. *Journal of Biological Chemistry* 264:20171-4.

Shintani, D., Roesler, K., Shorrosh, B., Savage, L. and J.Ohlrogge (1997) Antisense expression and over expression of biotin carboxylase in tobacco leaves. *Plant Physiology* 114:881-6.

Shyamala V and G. Ames (1989) Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR *Gene* 84(1):1-8

Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, Timmons L, Plasterk RH, Fire A.(2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107(4):465-76. Silver, J. and V.Kerkatte (1989) Novel use of polymerase chain reaction to amplify cellular DNA adjascent to an integrated provirus. *Journal of Virology* 63:1924-8.

Singh, K., Carey, M., Sargosti, S. and M.Botchan (1985) Expression of enhanced levels of small RNA pol III transcripts encoded by the B2 repeats in Simian virus 40 transformed mouse cells. *Nature* 314:553-6.

Sisodia, S., Sollner-Webb, B. and W.Cleveland (1987) Specificity of RNA maturation pathwaysRNAS transcribed by RNA Polymerase III are not substrates for splicing or polyadenylation. *Molecular and Cellular Biology* 7:3602-12.

Sitzler S, Oldenburg I, Petersen G. and E. Bautz (1991) Analysis of the promoter region of the housekeeping gene DmRP140 by sequence comparison of Drosophila melanogaster and Drosophila virilis *Gene* 100:155-62.

Smith, C., Watson, C., Morris, P., Bird, C., Tucker, G., Schuch, W. and D.Grierson (1990) Inheritance and effect on ripening of antisense polygalactouronase genes in transgenic tomatoes. *Plant Molecular Biology*14:369-79.

Smith, N. and M.Devey (1994) Occurance and inheritance of microsatellites in *Pinus radiata. Genome* 37:977-83.

Sokal, H. and E.Rohlf (1995) Biometry. Freeman Press, USA. *pp 652.*

Soliner Webb, B. and J.Tower (1986) Transcription of cloned eukaryotic ribosomal RNA genes. *Annual Review of biochemistry* 55:801-30.

Sonnenberg, N. (1993) mRNA translation: influence of the 5 and 3 untranslated region. *Current opinion in genetics and development* 4:310-5.

Sorenson, P. and S.Fredriksen (1991) Characterization of human 5S RNA genes. *Nucleic Acids Research* 19:4147-51.

Southern, E.M. (1979) Gel elctrophoresis of restriction fragments *Methods in enzymology* 68:152-176.

Spiker, S. (1985) Plant chromatin structure. *Annual Review of Plant Physiology* 36:235-53.

Sprague, K. (1992) New Twists in class III transcription. *Current Opinion In Cell Biology* 4:475-9.

Sprinzl, M., Horn, C., Brown, M. and S.Steinberg (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research* 26:148-53.

Stachel, S., Timmermann, B. and P.Zambryski (1986) Generation of single stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens* plant cells. *Nature* 322:706-12.

Stafstrom, J. and I.M.Sussex (1992) Expression of a ribosomal protein gene in axillary buds of pea seedlings. *Plant Physiology* 100:1494-1502.

Stoutjesdijk PA, Singh SP, Liu Q, Hurlstone CJ, Waterhouse PA, Green AG (2002) hpRNA-Mediated Targeting of the Arabidopsis FAD2 Gene Gives Highly Efficient and Stable Silencing *Plant physiology* 129;1723-31.

Struhl, K. (1995) Yeast transcriptional regulatory mechanisms. *Annual Review of Genetics* 29:651-74.

Sullenger, A., Lee, C., Smith., A. and E.Gilboa (1990a) expression of chimeric tRNA driven antisense transcripts renders NIH 3T3 cells highly resistant to MMLV replication. *Molecular and Cellular Biology.* 10:6512-23.

Sullenger, A., Gallardo, F., Ungers, E. and E.Gilboa (1990b) Overexpressionof TAR sequences renders cells resistant to human HIV replication. *Cell* 63:601-8.

Suzuki, C., Nakajima, Y. and Y.Ohmiya (2005) A new additional reporter enzyme, dinoflagellate luciferase for monitoring gene expression. *Gene* 344:61-66.

Svitkin, Y., Ovchinnikov, P. and N.Sonenberg (1996) General RNA binding proteins render translation cap dependant. *EMBO Journal* 15:24:7147-55.

Szymanski, M., Barciszewska, M., Volker, A. and J.Barciszewski (1995) Are 5S ribosomal RNA primary transcripts edited. *Biochemistry and Molecular Biology International.* 37:3:413-22.

Szymanski, M., Barciszewski, V. and J.Barciszewski (2003) 5S rRNA: structure and interactions. *Biochemical journal* 371:641-51.

Tababra, H., Sarkissian, M., Fire, A. and C.Mello (1999) The rde-1 gene, RNA interference and transposon silencing in *C.elegans Cell:*99:123-32.

Tabata, T., Takase, H., Kawata, T., Nakayama, T.and M.Iwabuchi (1989) A protein that binds to a cis-acting element of wheat histone genes has a leucine zipper motif *Science* 245:965-7.

Tamaoki, M., Tsugawa, H., Yammamoto, N., Murakami, Y. and M.Matsuoka (1995) plant transcription factors *Plant Journal* 7:927-38.

Tang, L., Reinhart, J. and P.Zamore (2003) A biochemical framework for RNA silencing in plants. *Genes and Development* 17:49-63.

Tang, W., Newton, R. and V.Outhavong (2004) Exogenously added polyamines recover browning tissue into normal callus cultures and improve plant regeneration in pine. *Physiologia Plantarum* 122:386-95.

Tang, G. (2005) siRNA and miRNA:an insight into RISCs. *Trends in biochemical Sciences* 30:106-14

Tang, W., Weidner, D. and R.Newton (2005) Quantitative analysis of siRNA mediated GFP silencing in transgenic pine cells. *Plant Science* 168:741-6.

Tapping, R., Syroid, D., Bilan, P. and J.Capone (1993) The 5-flanking sequence negatively modulates the in vivo expression and in vitro transcription of a human tRNA gene. *Nucleic Acids Research* 21:4476-82.

Tartof, K.D. and C.A.Hobbs (1987) Improved media for growing plasmid and cosmid clones. *Bethesda Res. Laboratory. Focus* 9:12

Temple, S., Bagga, S. and C.S.Gopalan (1998) Down regulation of specific members of the glutamine synthetase gene family in alfalfa by antisense RNA technology. *Plant Molecular Biology* 37:535-47.

Tessier, C., Brosseau, R. and T.Vernet (1986)Ligation of single stranded oligonucleotides by T4 RNA ligase. *Analytical Biochemistry* 158:171-8.

Thangavelu, M., Belostotsky, M., Bevan, R.and J.Rogers (1993) Partial characterisation of the *Nicotiana tabacum* actin gene family: evidence for pollen specific expression of one of the gene family members. *Molecular and General Genetics* 240:290-5.

Theunissen, O., Rudt, F. and T.Pieler (1998) Structural determinants in 5S RNA and TFIIIA for 7S RNP formation. *European Journal of Biochemistry.* 258:758-767.

Thompson, D., Higgins, G. and J.Gibson (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids research* 22:4673-80.

Thompson, J., Ling, X. and M.Grunstein (1994) Histone H3 amino terminus is required for telomeric and silent mating locus in yeast. *Nature* 369:245-7.

Thompson, M. and R.A.Young (1995) General requirement for rNA polymerase II holoenzyme *in vivo. Proceedings of the National Academy of Sciences* 92:4587-90

Timmons, L. and A.Fire (1998) Specific interference by ingested dsRNA. *Nature* 395:854.

Tomizawa, J.I (1986) Control of ColEI plasmid replication: binding of RNA I to RNA II and inhibition of primer formation. *Cell* 47:89-97.

Tomizawa, J.I. (1990) Control of ColEI plasmid replication: Interaction of Rom protein with an unstable complex formed by RNA I and RNA II. *Journal of Molecular Biology* 212:695-708.

Trontin, J., Grandemange, C. and J.Favre (1999) Two highly divergent 5s rDNA unit size classes occur in composite tandem arrays in European larch (*Larix decidua*) and Japanese larch. *Genome* 42:837-848.

Tsukiyama, T. and C.Wu (1996) Purification and properties of an ATP-dependant nucleosome remodelling factor. *Cell* 83:1011-20.

Tyler, B.M. (1987) Transcription of *Neurospora crassa* 5S genes require a TATA box and three internal elements. *Journal of Molecular Biology* 196:801-11.

Ulmasov, T. and J.Guilfoyle (1992) Sequence of the fifth largest subunit of RNA polymerase II from plants. *The Journal of Biological Chemistry* 267:23165-9.

Ulmasov, B. and W.Folk (1995) Analysis of the role of 5 and 3 flanking sequence elements upon in vivo expression of the plant tRNA-Trp genes. *Plant Cell* :1723-34.

Ulmasov, T., Larkin, M. and J.Guilfoyle (1996) Association between 36- and 13.6 kd alpha-like subunits of *Arabidopsis thaliana* RNA polymerase II. *The Journal of Biological Chemistry* 5085-94

Ulmasov, T., Hagen, G. and T.Guifoyle (1997) ARF1, a transcription factor that binds to auxin response elements. *Science* 276:1865-8.

Umbeck. P., Johnson, G., Barton, K. and W.F.Swain (1987) Genetically transformed cotton (*Gossypium hirsutum*) plants. *Bio/Technology* 5:263-6.

Vain, P., Keen, N., Murillo, J., Nemes, C. and J.Finer (1993) Development of the particle inflow gun. *Plant Cell, Tissue and Organ Culture* 33:237-46.

Vakhitov, A., Gimalov, F. and P.Shumayatakii (1986) Determination of the number of copies of 5S rRNA ans trNA genes in wheat and Aegilops. *Genetika* 22:676-83.

Van der Krol, A., Mol, J. and A.R.Stuitje (1988) Modulation of eukaryotic expression by complementary RNA or DNA sequences *Biotechniques* 6:958-976.

Van der Krol, A., Mur, L., Lange, P., Mol, J. and A.R.Stuitje (1990a) Inhibition of flower pigmentation by antisense chs genes:promoter and minimal sequence requirements for the antisense effect. *Plant molecular Biology* 14:457-66.

Van der Krol, A., Mur, L., Beld, M., Mol, J. and A.R.Stuitje (1990b) Flavonoid genes in petunia:addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* 2:291-9.

Venkateswarlu, K., Lee, W. and N.Nazar (1991) Conserved upstream sequence elements in plant 5S ribosomal RNA coding genes. *Gene* 105;249-53.

Verdaguer B, de Kochko A, Fux CI, Beachy RN, Fauquet C.(1998) Functional organization of the cassava vein mosaic virus (CsVMV) promoter. *Plant Molecular Biology* 31:1129-39. Verdel, A. *et al* (2004) RNAi-mediated targeting of heterochromatin by RITS complex *Science* 303:672-6.

Veres, Q., Gibbs, A., Scherer, E. and T.Caskey (1987) The molecular basis of the sparse fur mouse mutation. *Science* 237:415-7.

Verrijzer, C.P. and R.Tjian (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trends in Biochemical Science* 9:338-42.

Vogelstein, B. and D.Gillespie (1979) Preparative and analytical purification of DNA from agarose *Proceedings of the National Academy of Sciences* 76:615-9.

Volbrecht, E., Veit, B. and S.Hake (1991) The developmental gene Knotted-1 is a member of a maize homeobox gene family *Nature* 350:241-3.

Volkmann, S., Jendis, J., Fraendorf, A. and K.Molling (1995) Inhibition of HIV-I reverse transcription by triple helix forming oligonucleotides with viral DNA. *Nucleic Acids Research* 23:1204-1212

Von Ehrenstein, G. (1967) Isolation of sRNA from intact *E.coli* cells. *Methods in enzymology* 12A:588

Wahl, M., Berger, L. and R.Kimmel (1987) Molecular hybridization of immobilized nucleic acids: Theoritical concepts and practical considerations. *Methods in enzymology* 152:399-407.

Waibel, F. and W.Filipowicz (1990) RNA-Polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. *Nature* 346:199-202.

Walden, A., Walter, C. and R.C.Gardner (1999) Genbank direct submission

Walden, A., Walter, C. and R.Gardner (1997) Big and Woody but no holding hands-A goal for conifer reproduction. . IUFRO genetics of Radiata pine, *FRI Bulletin 203*:310-19.

Waldmann, T., Scholten, I., Kappes, F., Hu, H. and R.Knippers (2004) The DEK protein - an abundant and ubiquitous constituent of mammalian chromatin. *Gene* 343 (1): 1-9

Waldron, C., Willis, N. and F.Gesteland (1985) Plant tRNA genes: putative soybean genes for tRNA-asp and tRNA-Met. *Journal of Molecular and Applied Genetics* 3:7-10.

Walter, C., Carson, M., Charity, J., Gardner, R., Wagner, A. and A.Walden (1997) Transformation and gene expression in Pinus radiata. IUFRO genetics of Radiata pine, *FRI Bulletin 203*:319-339.

Wang Z and R.Roeder (1995) Structure and function of a human transcription factor TFIIIB subunit that is evolutionarily conserved and contains both TFIIB- and high-mobility-group protein 2-related domains *Proceedings of the National Academy of Sciences* 92:7026-30

Wang, C. and L.O.Vodkin (1994) Extraction of RNA from tissues containing high levels of procyanidins that bind RNA *Plant Molecular Biology Reporter* 12:2:132-45.

Wang X., Szmidt, A. and O. Savolainen (2001) Genetic composition and diploid hybrid speciation of a high mountain pine, Pinus densata, native to the Tibetan plateau *Genetics* 159:337-46

Wang, S., Robertson, G., and J.Zhua (2004) A novel human homologue of Drosophila polycomblike gene is up-regulated in multiple cancers. GENE, 343 (1): 69-78

Warburg, O, and W.Christian (1942) Isolation and crystallization of Enolase. *Biochimica Z* 310:384-421.

Washburn, B., Davis, E. and S.Ackerman (1997) Coactivators and TAFs of transcription activation in wheat. *Plant Molecular Biology* 35:1037-43.

Weaver, R., Anolles, G. and L.M.Calahan (1994) Isolation and cloning of DNA amplification products from silver-stained polyacrylamide gels. *Biotechniques* 16:226-7.

Weber, K., Bolander, E. and G.Sarkar (1998) Rapid acquisition of unknown DNA sequences adjascent to a known segment by multiplex restriction site PCR. *Biotechniques* 25:415-9.

Weigel, D. and O.Neilsson (1995) Overexpression of homeotic genes *Nature* 377:495-7

Weiher, H., Konig, M. and P.Gruss (1983) Multiple point mutations affecting the simian virus 40 enhancer. *Science* 219:626-31.

Weinman, R. and G.Roeder (1974) Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes.*Proceedings of the National academy of Sciences* 71:1790-4.

Weser, S., Reimann, J. and W.Meissner (2003) Assembly and isolation of intermediate steps of transcription complexes formed on human 5S rRNA gene. *Nucleic Acids Research* 31:2408-16.

Wesley, .V, et al (2001) Construct design for efficient, effective and high throughput gene silencing in plants. *Plant Journal* 27(6):581-90

White, R., Jackson, S. and P.W.Rigby (1992) A role for the TATA-box-binding protein component of the transcription factor IID complex as a general RNA Polymerase III transcription factor. *Proceedings of the National academy of Sciences* 89:5:1949-53.

Wieczorek, E., Brand, M., Jacq, X. and L.Tora (1998) Functions of TAF_{II} containing complex without TBP in transcription by RNA polymerase II. *Nature* 393:187-91.

Wilkins, R. (1991) Radioactive labelling of DNA with PCR. *Gene Scene* March, 1991.

Williams, M. and I.Sussex (1995) Developmental regulation of ribosomal protein L16 genes in Arabidopsis thaliana *Plant Journal* 8:65-76

Willis, M. (1993) RNA polymerase III-Genes, factors and transcription specificity. *European Journal of Biochemistry* 212:1-11.

Wilson, A. and M.Gough (1988) High voltage *E.coli* transformation with DNA following ligation. *Nucleic Acids Research* 16:11820.

Wolberger, C. Vershon, K., Liu, B., Johnson, D. and C.O.Pabo (1991) Crystal structure of a Mat alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interaction. *Cell* 67:3:517-28.

Wolffe, P. (1991) RNA polymerase III transcription. *Current Opinion in Cell Biology* 3:461-6.

Wolters, J. and V.Erdmann (1988) Compilation of 5S rRNA and 5S rRNA gene sequences. *Nucleic Acids Research*14:r1-70.

Womble, D., Dong, X., Luckow, V., Wu, R. and R.H.Rowad (1985) Analysis of the individual regulatory components of the IncFII plasmid replication control system. *Journal of Bacteriology* 161:534-43.

Wood, M., DeAth, M., Maloney, M. and S.Rice (2000) Gymnosperm super ubiquitin - a strong constitutive promoter for gene expression in model plant species. *Proceedings of the ComBio conference, Wellington, December 2000.*

Woolford, L. (1991) The structure and biogenesis of yeast ribosomes. *Advances in Genetics* 29:63-83.

Wormington, W., Bogenhagen, D. and D.D.Brown (1981) A quantitative assay for *Xenopus* 5S RNA gene transcription *in vitro. Cell* 24:809-17.

Woychick, N. and R.Young (1992) Genes encoding transcription factor IIIA and the RNA Polymerase common subunit RPBC are divergently transcribed in *Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences* 89:3999-4003.

Wu, Y., Ugozzoli, L., Quian, J. and B.Wallace (1991) The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by PCR. *DNA Cell Biology* 10, 233-8.

Wyszko, E. and M.Barciszewska (1997) Purification and characterization of transcription factor IIIA from higher plants. *European Journal of Biochemistry* 249:107-12.

Xie, Z., Johansen, L., Lellis, A. and C.Carrington (2004) Genetic and functional diversification of small RNA pathways in plants. *PLOS Biology* 2:104-15.

Yang, Z. and J.Heyes (2003) Xenopus transcription factor IIIA and the 5S nucleosome:development of a useful in vitro system. *Biochemical Cell Biology* 81:177-184.

Yamashita, T., Iida, A. and H.Morikawa (1991) Evidence that more than 90% of glucouronidase expressing cells after particle bombardment directly receive the foreign gene in their nucleus. *Plant Physiology* 97:829-31.

Yie, J., Senger, K. and D.Thanos (1999) Mechanism by which the IFN enhanceosome activates transcription. *Proceedings of the National Academy of Sciences* 96:13108-13.

Ying, S. and S.Lin (2004) Intron derived microRNA-fine tuning of gene functions. *Gene* 342:25-8.

Yoo, B., Kragler, F., Lee, Y., Lough, T. and W.Lucas (2004) A systemic small RNA signalling system in plants. *The Plant Cell* 16:1979-2000.

Yukawa, Y., Sugita, M. and M.Suguira (2000) The TATA motif, the CAA motif and the poly T transcription termination motif are all important for transcription reinitiation on plant tRNA genes. *Plant Journal* 22:439-47.

Zambryski, P.C. (1993) Chronicles from the *Agrobacterium* plant cell DNA transfer story. *Annual Review of Plant Physiology and Plant Molecular Biology* 43:465-90.

Zamecnik, P.C. and M.L.Stephenson (1978) Inhibition of Rous Sarcoma virus replication and cell transformation by gene specific oligodeoxynucleotide. *Proceedings of the National Academy of Sciences* 75:280-4.

Zamora, D. Tuschi, T. Sharp, A. and P.Bartel (2000) RNAi:double stranded RNA directs the ATP dependent cleavage of mRNA at 21-23 nucleotide intervals. *Cell* 101:25-33.

Zenke, M., Grundstrom, T., Matthes, H., Schatz, C., Wildeman, A. and P.Chambon (1986) Multiple sequence motifs are involved in SV40 enhancer function. *EMBO Journal* 5:387-97.

Zhang, Y. and L.Scholl (1990) Two evolutionarily divergent genes encode a cytoplasmic protein of *Arabidopsis thaliana*. *Gene* 93:177-82.

Zilberman, D, Cao, X and S.Jacobsen (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation.*Science*. 299:716-9

Zink, D. and R.Paro (1995)Drosophila polycomb-group regulated chromatin inhibits the accessibility of transactivator to its target DNA. *EMBO Journal* 14:5660-71.

Zupan, J. and P.Zambryski (1997) The Agrobacterium DNA transfer complex. *Critical Reviews in Plant Science* 16:279-95.

Treatment means and standard error. Individual treatment readings obtained by Final (=pm/minute/25 ug) x (1000/25ug) x (1/30 minutes).

Treatment	Replicate	pmole/m/mg	**mean**	**std error**
pGUS	1	1404	1075.4	87.2
	2	923		
	3	848		
	4	1283		
	5	1463		
	6	980		
	7	878		
	8	825		
p5S	1	167	95.9	20.4
	2	74		
	3	48		
	4	212		
	5	56		
	6	79		
	7	158		
	8	68		
p5A/pGUS	1	72	120.0	24.5
	2	62		
	3	198		
	4	182		
	5	176		
	6	42		
	7	216		
	8	52		
Control	1	32	17.9	3.8
	2	10		
	3	29		
	4	36		
	5	10		
	6	8		
	7	12		
	8	6		
	one way ANOVA			
Source	SS	*df*	MS	F
Between	5956824	3	1985608	100.5
Within	553216	28	19757	
Total	6510040	31		

MUG time course assay

Stock solution containing 1U ul-1 b-glucouronidase was diluted 10-4 in GUS extraction buffer. 50 ul were incubated with 1 mM MUG and absorbance read at 0, 10, 20 and 30 minutes to confirm linearity of reaction.

Time (minutes)	Fluorescence (relative units)
0	0
10	260
20	510
30	790

Transient expression RT-PCR

Visual rating of *gus*_bands in RT-PCR (chapter V, Section 3.5.2) using a 1-6 scale.

Replicate	1	2	3	4	5	6	7	8
pGUS	+++++	++++	+	++++++	++++	+++++	++++	+++++
p5S	+++	++++	++++	++++	++++	++++	+++	+++++
pGUS/p5A	-	-	+	-	+	+	-	-

APPENDIX II

Pine RNA

Extraction buffer - 2% CTAB, 2% PVP (K30), 100 mM Tris-HCl, pH 8, 25 mM EDTA (di-sodium salt), 2 M NaCl, 0.5g l^{-1} spermidine
Autoclave and add β-mercaptoethanol and proteinase-K (20μg μl^{-1}) to 0.2 and 0.25%, respectively, final concentration just before use.

SSTE - 1 M NaCl, 0.5% SDS,10 mM Tris-HCl, pH 8,1 mM EDTA,Autoclave.

Prepare 10 M LiCl and 3 M NaOAc in DEPC-treated water and autoclave before use.

Pine DNA

Extraction buffer - 100 mM Tris-HCl, pH 8, 2% (w/v) CTAB, 1.4M NaCl, 20 mM EDTAAutoclave and add 0.2% ME and 0.25% Proteinase K (20μg μl^{-1}) just before use.

Nuclear extraction buffer - 50 mM Tris-HCl, pH 8, 5 mM EDTA, 350 mM Sorbitol, 10% PEG 8000. Autoclave and add 0.1% BSA & 0.1% ME just before use.

Nuclear wash buffer - 50 mM Tris-HCl, pH 8, 25 mM EDTA, 350 mM Sorbitol Autoclave and add 0.1% ME before use.

Plasmid preparation

TE buffer - 10 mM Tris-HCl, pH 8, 1 mM EDTA. Autoclave

TE saturated phenol Chloroform - Mix equal volumes of TE, pH 8 and phenol and allow phases to separate. Repeat the process till pH of upper phase is 8. Add an equal volume of Chloroform:isoamyl alcohol (24:1) to the lower phenol layer and store at 4^0C.

Resuspension buffer - 25 mM tris-HCl, pH 8, 10 mM EDTA, Autoclave.
Add 50 mM filter sterilized glucose.

Neutralizing solution

To 60ml 5 M KOAc add 11.5ml glacial acetic acid and 28.5ml water. This solution is 3 M with respect to potassium and 5 M with respect to acetate. Autoclave and store at 4^0C.

DNase-free RNase A

RNase A is dissolved in 10 mM Tris-HCl, pH 7.5, 15 mM NaCl to a final concentration of 10mg ml^{-1}, heated at 100^0C for 15 minutes and allowed to cool to room temperature. Store at -20^0C.

Wizard DNA purification system

Neutralizing solution -1.32 M KOAc, pH 4.8, Column wash solution -, 80 mM KOAc, 8.3 mM Tris-HCl, pH 7.5, 40 μM EDTA, 55% EtOH

Boiling method

STET buffer - 10ml 0.5 M EDTA, 5ml 1 M Tris-HCl, pH 8, 0.5ml Triton 100, 64.5ml H_2O Autoclave and add 20ml filter sterilized 40% sucrose.

Agarose gel electrophoresis

1X TAE - 40 mM tris-HCl, pH 7.8, 20 mM Acetic acid, 1 mM EDTA
1X TBE - 89 mM Tris-HCl, pH 8.3, 89 mM Boric Acid, 2 mM EDTA
5X Gel loading buffer - 50 mM Tris-HCl, pH 8, 5 mM EDTA, 25% glycerol,0.2% BPB, 0.2% Xylene cyanol

Migration of tracker dyes

Agarose 0.5-1.5%	Xylene cyanol: 4-5kb	BPB:400-500bp
Agarose 2-3%	Xylene cyanol: 750bp	BPB: 100bp

PAGE

Migration of tracker dyes

Acrylamide 5%	Xylene cyanol:260bp	BPB:65bp
Acrylamide 8%	Xylene cyanol:160bp	BPB:45bp
Acrylamide 12%	Xylene cyanol:70bp	BPB:20bp
Acrylamide 15%	Xylene cyanol:60bp	BPB:15bp

DNA cleaning by ion exchange chromatography

QBT buffer - 750 mM NaCl, 50 mM MOPS, pH 7, 0.15% Triton, 15% EtOH
QC buffer - 1 M NaCl, 50 mM MOPS, pH 7, 15% EtOH
QF buffer – 1.25 M NaCl, 50 mM Tris-HCl, pH 8.5, 15% Et OH

Gel extraction with glass fibre

Binding buffer - 3 M GuSCN, 10 mM Tris-HCl, pH 6.6, 5% EtOH
Wash buffer - 20 mM NaCl, 2 mM tris-HCl, pH 7.5, 80% EtOH

Gel extraction with silica

Binding buffer -6 M $NaClO_4$, 50 mM Tris-HCl, pH 8, 10 mM EDTA

Wash buffer -20 mM Tris-HCl, pH 7.5, 2 mM EDTA, 50% EtOH
Elution buffer - 10 mM Tris-HCl, pH 8, 1 mM EDTA

Cloning

10x ligation buffer - 660 mM Tris-HCl, pH 7.5, 50 mM $MgCl_2$, 10 mM DTE, 10 mM ATP

Dephosphorylation

Dilution buffer - 25 mM Tris-HCl, pH 7.6, 1 mM $MgCl_2$,0.1 mM $ZnCl_2$, 50% glycerol
Dephosphorylation buffer (1x) - 50 mM Tris-HCl, pH 8.5, 0.1 mM EDTA

5α competent cells

Minimal medium (M-9) - 6g Na_2HPO_4, 3g KH_2PO_4 , 0.5g NaCl , 1g NH_4Cl , 15g Agarose Dissolve in 1L water and after autoclaving add filter sterilized,

1 M $MgSO_4$ 2ml, 1 M $CaCl_2$ 0.1ml, 10ml 20% glucose , 1 M thiamine-HCl 1ml

Trituration buffer - 100 mM $CaCl_2$, 70 mM $MgCl_2$, 40 mM NaOAc, pH 5.5

Blue white selection

X-gal
X-gal is dissolved at 20mg ml^{-1} in N,N`dimethylformamide in a polypropylene tube, wrapped in aluminium foil to prevent light damage and stored at -20.
IPTG
2g is dissolved in 8ml water and volume adjusted to 10ml (840 mM), filter sterilized using a 0.22 micron filter and stored in 1ml aliquots at -20^{0}C.

Plates

To premade LB antibiotic plates (20ml) 40μl X-gal (and 4μl IPTG if added to the plate) stock solution(s) are added and spread over the entire surface of the plate using a sterile glass spreader. Plates are incubated at 37^{0}C for 3-4 hours to evaporate the formamide (which has low volatility). 200μl of transformation culture is then spread on the plate surface, allowed to be absorbed into the medium and incubated at 37^{0}C.

Southern blotting

Neutralizing solution - 1.5 M NaCl, 0.5 M Tris-HCl, pH 7.2, 0.001 M EDTA,
20x SSC - 3 M NaCl, 0.3 M Sodium citrate, pH 7

Northern blotting

10x MOPS/EDTA buffer - 0.2 M MOPS, 50 mM NaOAc, 10 mM EDTA, Adjust pH to 7 and autoclave.

RNA sample preparation -12.5µl deionized formamide, 2.5µl 10x MOPS buffer, 4µl formaldehyde (37%)6µl RNA sample. Heat at 65°C for 10minutes and chill on ice. Add 2.5µl 50%(v/v) glycerol containing 0.1 mg ml^{-1} bromophenol blue and 1µl EtBr.

RNA gel - 1.5g Agarose, 10 ml 10x MOPS/EDTA buffer, 87 ml Water.Dissolve agarose and cool to 50°C, add 5.1ml HCHO (37%) and pour immediately.

Digestion was performed with 6ug (human) DNA with 1ul type II restriction enzyme (10u ul) in 20ul final volume. reactions were stopped by addition of 0.1 volume 0.5M EDTA, pH 8 or by heating at 65°C.

Southern hybridization

Church and Gilbert buffer_- 0.5 M Na_2HPO_4, pH 7.2, 7% SDS, 1 mM EDTA

Detection wash buffer - 0. M Maleic acid, pH 7.5, 0.15 M NaCl, 0.3% (v/v) Tween 20
Maleic Acid buffer -0.1 M Maleic Acid, pH 7.5, 0.15 M NaCl
Blocking solution - Resuspend the blocking reagent in Maleic acid buffer to form a 1% (w/v) solution and store at 4°C.
Detection buffer - 0.1 M Tris-HCl, pH 9.5, 0.1 M NaCl

***In situ* Hybridization**

Fixative - Ethanol 50%' Acetic Acid 5%, Formaldehyde (37%) 3.7%, Water 41.3%
Proteinase K buffer - 100 mM Tris Hcl, pH 7.5, 50 mM EDTA
10x hybridization salts - 3 M NaCl, 100 mM Tris-HCl, pH 7.5, 10 mM EDTA
Hybridization buffer - 50% Formamide, 300 mM NaCl, 10 mM Tris-HCl, 1x Denhardt's solution,10% Dextran sulfate, 70 mM DTT, 40µg ml^{-1} salmon sperm DNA, 400ng ml^{-1} probe
Antibody buffer - 100 mM Tris-HCl, pH 7.5, 150 mM NaCl
Substrate buffer - 100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$
50x Denhardt's solution - 1g Ficoll, 1g PVP, 100ml water. BSA is filter sterilized and added at 1g per 100ml.

Salmon sperm DNA - DNA is added at 10mg per ml of water and stirred for 4 hours at room temperature to dissolve the DNA. The solution is made up to 0.1 M NaCl and extracted once with phenol. The aqueous layer is extracted with $CHCl_3$:isoamyl alcohol. DNA is sheared by passing the solution rapidly through a 17-gauge hypodermic needle 10 times. DNA is precipitated with 2 volumes of EtOH its concentration is adjusted to 10 mg ml^{-1}. The solution is boiled for 10 minutes and stored at -20^0C.

5′ end labeling

10x PNK buffer - 0.5 M tris-HCl, pH 8.2, 0.1 M $MgCl_2$, 1 mM EDTA, 50 mM DTT, 1 mM spermidine

Affinity capture

Hybridizing buffer - 50% formamide, 3x SSC, 40 mM sodium phosphate, 0.5% SDS
linear acrylamide - A 5% acrylamide solution (No *bis*-acrylamide added) is prepared in 40 mM Tri-HCl, pH 7.8, 20 mM NaOH, 1 mM EDTA. 1/100th volume of 10%APS and 1/1000th volume of TEMED are added and acrylamide allowed to polymerize for 30 minutes. When solution becomes viscous add 2.5 volumes EtOH and centrifuge at 13000rpm for 10 minutes. Dissolve the precipitate in 20 volumes H_2O by shaking overnight. The 0.25% linear acrylamide can be stored at 4°C for several years.

Standard RT-PCR

5x first strand buffer - 250 mM Tris-HCl, pH 8.3, 250 mM KCl, 50 mM MgCl2, 2.5 mM spermidine, 50 mM DTT, 5 mM dNTP

Bradford Assay

Bradford Reagent - Stock solution - Coomassie blue G-250 50mg, EtOH 95ml, Orthophosphoric Acid 205 ml. Store in dark at RT.
Working solution – Stock 15 ml, EtOH 7.1 ml, Phosphoric acid 15.5 ml, add water to 250 ml. Filter Whatman#1, store RT for 3-4 weeks.

BSA standard 1mg/ml at wavelength 595

MUG Assay

Gus extraction buffer – 5 ml 1 M sodium phosphate, pH 7 (NaH_2PO_4+ Na_2HPO_4), 0.07 ml ME, 2 ml 0.5 M EDTA, pH8, 0.33 ml 30% sarcosyl, 1 ml 10% triton x-100, 91.6 ml water.

Stop Buffer – 0.2 M anhydrous Na_2CO_3

MUG stock – 2 mM solution - 25 mg MUG (Boer-Mann or Sigma) in 25 ml gus extraction buffer. Store 4C.

MU stock – 1uM solution – 19.8g MU in 100 ml distilled water; store 4C away from light.