FURTHER CHARACTERISATION OF THE DOTHISTROMIN GENE CLUSTER OF *DOTHISTROMA PINI*

A thesis presented in partial fulfillment of the requirements
for the degree of Master of Science in Biochemistry
at Massey University, Palmerston North, New Zealand.

Olivia Rachel Teddy

2004

ABSTRACT

The polyketide dothistromin is a toxin produced by the filamentous fungus *Dothistroma pini* that is thought to play a role in causing *Dothistroma* needle blight in *Pinus radiata.* Dothistromin is structurally similar to aflatoxin B1 (AF), a highly carcinogenic toxin with no known function that is produced by the fungus *Aspergillus parasiticus* and also to versicolorin, an intermediate of the well characterised biosynthetic pathways of AF and sterigmatocystin (ST). The structural similarities between AF/ST and dothistromin suggest that genes homologous to AF biosynthetic genes will be involved in dothistromin biosynthesis. AF/ST biosynthetic genes of *A. parasiticus* and *A. nidulans* are clustered and hence it is likely that the dothistromin biosynthetic genes are also clustered in a similar manner. Two λ clones, λKSA and λCGV1 containing portions of the putative dothistromin cluster have been isolated in previous studies. Another λ clone λCGV2 was also identified using an aflatoxin gene probe but it is unknown whether it is part of the dothistromin biosynthetic cluster.

The λKSA clone contains part of a putative polyketide synthase pks^{dot} (64% identical to *A. parasiticus* AF biosynthetic gene *pksA).* Two crucial domains required for functioning are contained within λKSA, the β-keto acyl synthase (KS) and acyl transferase (AT) domains. The putative pks^{dot} is thought be involved in the beginning of the dothistromin biosynthetic pathway, working in a complex with a fatty acid synthase (FAS) to produce the intermediate noranthrone. A gene replacement construct was made using Multisite Gateway ™Recombination, replacing the AT and KS domains with an *hph* cassette. Disruption of the pks^{dot} gene will confirm it's involvement in dothistromin biosynthesis and could also confirm the role of dothistromin in pathogenicity as if the putative polyketide synthase (pks^{dot}) is involved in the first step of the dothistromin pathway thus a knockout would form a mutant devoid of any intermediates. Confirming the involvement of pks^{dot} would also provide evidence that like λCGV1, λKSA contains a portion of the dothistromin biosynthetic gene cluster.

As the positioning of the three lambda clones λKSA, λCGV1 and λCGV2 relative to one another in the *D. pini* genome was unknown Southern blot analysis was implemented to identify any relationship between the three lambda clones. No evidence was found to suggest the close linkage of the three lambda clones however this does not discount any linkage at all. Southern blot analysis did provide evidence that *ver-2* (77%

identity to melanin biosynthetic gene *phn1* of *Cochliobolus heterostrophus)* of λCGV2 is within close proximity to a putative *aflR* gene (regulatory gene for activating gene transcription in AF/ST biosynthesis) suggesting a regulatory role of this putative *aflR* gene in melanin biosynthesis and not dothistromin biosynthesis.

Further nucleotide sequencing of the λKSA clone revealed three putative dothistromin genes. Mox^{dot} and ord^{dot} have high amino acid identity to genes involved in the AF/ST pathways (70% identity to *moxY* and 51% identity to *avfA* of *A. parasiticus* respectively), suggesting similar roles in dothistromin biosynthesis. $Epox^{dot}$ showed high amino acid identity to an epoxide hydrolase of *A. niger (hyll)* suggesting it has a unique role in dothistromin biosynthesis as no homologs are seen in the AF/ST clusters. Southern blotting was also used to confirm the arrangements of genes from the λKSA clone within the *D. pini* genome.

Further characterisation of genes involved in dothistromin biosynthesis will firstly enable understanding of the role of dothistromin in needle blight and secondly will enable further comparative studies between AF/ST and dothistromin.

ACKNOWLEDGEMENTS

Firstly my biggest thank you has to go to my supervisor Rosie Bradshaw. Your constant encouragement and support was greatly appreciated. You went beyond of the call of duty at several times throughout the past 2 ½ years to see I finished this thesis, and that will never be forgotten. I feel privileged to have had the opportunity to get to know you on a both a personal and professional level.

Secondly I would like to thank the members of Fungal Jungle. Especially Phil who has been with me from the start, for being there and not knowing anything to begin with either (and for coining the term 'thesis madness' which I have frequently used in the last month) and to Arne who has provided many moments of light relief and also for doing some RT-PCR for me. Thanks to Shuguang, for your patience with me whilst I was writing up and for dealing with the constant bombardment of questions. Thanks to other members of IMBS especially Caryoln Young for your help during the beginning and the end and occasional time in between. Thanks to Carol Flyger, for helping with DIG labelling of my Southern blotting probes. Thank you to Kathyrn Stowell for making it possible for me to write up this thesis whilst starting a new job. Thanks to Asela and Vikki for being great friends and for helping me do one of my crucial ligations and transformations. Thanks to my other friends throughout the institute for being my lunch buddies and providing distractions from work. Thanks to my friends outside of the institute for your friendship and support. Thanks to Jason for your constant text messages, Emma for your always amusing adventure reports and giving me something to aspire to, Michelle for motivating me to write at night since you were doing work too, Molly for your emails, Shelly for being there and helping me through some tough times and lastly John for making me smile.

Another big thank you has to go to my Family, without you guys this would not have been possible. Thanks for the financial assistance and also for the emotional support, thanks for always believing in me, for never doubting me. Especially thanks to you mum, thanks being my friend and helping in everyway possible, thanks for the little cards, the endless phone calls, and dropping everything when I needed you. And lastly thanks to my cats, Samsara and Paige, for your cuddles and for making me laugh at your quirky personalities.

'It is in hoping that we dream, In dreaming that we seek, In seeking that we find our life's desire'

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACP	Acyl-carrier protein
Amp	Ampicillin
AF or AFB1	Aflatoxin B1
AT	Acyl transferase
AVF	Averufin
Bp	base-pairs (DNA)
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
cDNA	Complementary DNA
CSPD	Disodium 3-(4-methoxyspiro (1, 2- dioxetane-3, 2' – (5' chloro) tricyclo [3, 3.1.3.1] decan}-4-yl) phenyl phosphate
CTAB	Hexadecyltrimethylammonium bromide
CU	Cerato ulmin
DB	*D. pini* broth
DIG	digoxygenin
DH	Dehydratase
DM	*D. pini* media
DNA	Deoxyribonuclic acid
dNTP	deoxy-nucleotide tri phosphate
dot	Putative dothistromin gene
D. pini	*Dothistroma pini*
DSM	*D. pini* sporulation media
E. coli	*Escherichia coli*
ELISA	Enzyme-linked immunosorbent assay
EH	Epoxide hydrolase
ER	Enoyl reductase
Etbr	Ethidium bromide
FAS	Fatty acid synthase
Hph	Hygromycin B phosphate transferase gene
HPLC	High pressure liquid chromatography

IPTG	Isopropyl-β-D-galactoside
Kb	kilobase (DNA)
KDa	Kilodalton
KS	β-keto acyl synthase
Kv	Kilovolts
LB	Luira-Bertani
NA	Norsolorinic acid
PCR	Polymerase chain reaction
PEG	poly-ethyl glycol
PKS	Polyketide synthase
RACE	Rapid amplification of cDNA ends
RE	Restriction endonuclease
REMI	Restriction enzyme-mediated integration
Rpm	Revolutions per minute
RT-PCR	Reverse transcriptase polymerase chain reaction
SDS	Sodium dodecyl sulphate
ST	Sterigmatocystin
TAGKO	Transposition-arrayed gene knockouts
Taq	Thermus aquaticus
TE	Thioesterase
Tm	Melting temperature
UV	Ultra violet
VHA	Versiconal hemiacetal acetate
Xgal	5-bromo-4-chloro-3-idoly-β-D-galactopyranoside
Ω	Ohms
λ	Lambda
μf	Microfarad

1.0 INTRODUCTION

1.1 *DOTHISTROMA* NEEDLE BLIGHT IN *PINUS RADIATA*

The filamentous fungus *Dothistroma pini* Hulbury is a forest pathogen of *Pinus radiata* and other *Pinus* species, producing a necrotic disease known as *Dothistroma* needle blight. *D. pini* Hulbury was first reported in New Zealand in 1962 and positively identified in 1964 (Gadgil 1967). The disease is characterised by a red band on infected needles and also by small irregular shaped fruiting bodies. This fungus is known to cause premature defoliation leading to reduced photosynthetic ability and possibly tree death (Gadgil 1984).

Dothistroma needle blight has been recorded in most of the North Island of New Zealand (excluding the northern tip) as well as Southland, Otago, Westland, Nelson and parts of Marlborough (Gadgil 1984). Outbreaks have also been reported in several other countries including, England, Africa, Asia, America and parts of Europe (Gadgil 1984).

The forest industry in New Zealand currently contributes to 12% of New Zealand's total export earnings, with an annual export profit of $3.5 billion dollars. Radiata pine (*Pinus radiata*) makes up 90% of the 1.8 million hectares of managed tree plantations. The introduction and establishment of *D. pini* has created a significant threat to the forest industry, leading to loss of volume gain and hence resulting in huge economic losses (Carson, Dick *et al.* 1991).

A toxin dothistromin, secreted by *D. pini* has been implicated in needle blight, as injection of pure dothistromin into pine needles produced the symptoms of needle blight (Shain 1981). These observations lead to the conclusion that dothistromin is a phytotoxin responsible for the needle blight. Dothistromin toxicity has been demonstrated towards mature pine embryos and leaf callus, and a 40-kDa dothistromin binding protein has been detected in embryos, although its exact function in the disease is still to be established (Jones, Harvey *et al.* 1995). To elucidate the role that dothistromin has in *Dothistroma* needle blight, dothistromin deficient mutants need to be constructed, and these mutants examined for pathogenicity and virulence.

1.2 INFECTION

In favourable conditions, *D. pini* spores land on needles of infected susceptible hosts and germinate, penetrating through the stomata. Inside the tissue the fungus is able to spread laterally but is limited to an area a few millimeters from the point of penetration. Disruption of the mesophyll tissue then occurs through the release of the dothistromin toxin. After hyphal growth in the needle tissue, fruiting bodies (stromata) are produced which then push their way through the needle surface (Gadgil 1984). Asexual spores are liberated from these fruiting bodies in a film of water on the needle surface. When droplets of the film drip off the needles and are broken up on contact with other branches or the ground, these spores are then released into the air (Gadgil 1984).

Climatic conditions affect the incidence of needle blight. Conditions which are conducive to disease are warm temperatures, high light intensity and long leaf wetness periods (Carson, Dick *et al.* 1991). The age of Radiata trees also has an effect on disease susceptibility. Younger trees are more susceptible as their stomata are open pores, allowing easier access for the fungus (Gadgil 1967). The effect on tree growth does not become apparent until approximately 2 years after the first appearance of the disease (Gadgil 1984). The appearance of the disease is apparent by a distinctive reddish tinge, which occurs at the base of the crown and from here spreads upwards. Quite often there are clear divisions between infected and uninfected areas (Gadgil 1984). The extensive needle death seen in *D. pini* infection is thought to be due to benzoic acid production in cells adjacent to those killed by dothistromin, rather than a direct effect of the toxin itself (Franich, Carson *et al.* 1986).

As infected needles are shed prematurely a layer of needles bearing large numbers of *D. pini* stromata builds up on the forest floor. As *P. radiata* over 20 years are seldom infected with dothistromin and the shortest possible rotation of *P. radiata* is not likely to be less than 22 years, it is unlikely that a stand at the end of its rotation will have *D. pini* on shed needles producing viable conidia. This means there is little danger that regenerating seedlings will become infected as a direct result of growing in an area in which the crop had previously been infected (Gadgil 1970).

1.3 CURRENT CONTROL METHODS

Two methods currently being employed to reduce the level of blight infection are chemical control and breeding of resistant strains; these are outlined below.

1.3.1 Chemical Control

Dothistroma needle blight can be controlled by aerial spraying with copper-based fungicides. In the past both cuprous oxide and copper oxychloride have been used in operational spraying. Research trials have shown that both fungicides are equally effective in controlling the disease, but as cuprous oxide is more expensive only copper oxychloride has been used since 1972. The solubilisation of copper oxychloride into Cu^{2+} or its complexes allows *D. pini* conidia to readily take up the spray, thus enabling copper fungicides to be an efficient method of control. The extent of crown infection is used to determine whether spraying should be carried out. In an area where annual surveys show 15-30 % of crown infection by *D. pini*, one spray application of 1.6 kg/ha is applied in early summer. In areas with > 30% crown infection two spray applications are used, one in early summer and another in late summer (Dick 1989). Aerial spraying has been found to reduce, but not eliminate the disease, which brings into question the economic benefit of carrying out such a practice (Carson, Dick *et al.* 1991).

1.3.2 Resistant Strains

In stands of infected trees, some trees were observed to be resistant to *D. pini*. Resistance of *P. radiata* to *D. pini* is an additively inherited trait and may be partly due to the presence of oxidised resin acids at the needle surface around the stomata (Franich and Gadgil 1983). Programmes have now been implemented to selectively breed for resistant *P. radiata* strains. These programmes normally select for both disease resistant traits as well as non-disease traits such as growth and form. In New Zealand, this has involved using trees already improved for other traits to breed *D. pini* resistance (Carson and Carson 1989).

Using the *Dothistroma* resistant breed of *P. radiata* should allow a reduction in the stand mean infection that would in turn allow an annual saving in the amount of aerial spraying

carried out (Carson, Dick *et al.* 1991). The greatest gains of a resistant breed of *P. radiata* will be seen in areas of greatest risk (i.e. areas with a high incidence of needle blight), as loss of volume will be recovered (Carson, Dick *et al.* 1991). NZ Forest Research Institute has suggested that the *Dothistroma* resistant breed only be used in areas at risk from *D. pini* and standard breeds be planted at other sites. This suggestion was made on the basis that the resistant breed has a lower yield at normal sites (Dick 1989).

In New Zealand, the *D. pini* population has low genetic diversity, suggesting that a single strain was introduced into New Zealand and has propagated asexually ever since. One possible consequence of this is that any dothistromin resistant breed of *P. radiata* whilst having increased resistance to the New Zealand strain of *D. pini* may in fact be non-resistant to other strains from around the world (Bradshaw, Ganley *et al.* 2000).

1.4 THE BIOCHEMISTRY AND MOLECULAR BIOLOGY OF AFLATOXIN AND DOTHISTROMIN BIOSYNTHESIS

1.4.1 Polyketide Biosynthesis

Polyketides are a large and diverse family of compounds produced by both prokaryotes and eukaryotes for various biological roles (Hopward and Sherman 1990). Polyketides are synthesised by successive additions of simple carboxylic acids to form a chain analogous to the chain elongation steps in fatty acid biosynthesis (Mayorga and Timberlake 1992; Cane 1994). The β-carbon of these molecules always carries a keto group which is not reduced (unlike fatty acids)(Yu and Leonard 1995). It is these non-reduced keto groups that give the group the name polyketide. Aflatoxin (AF) and sterigmatocystin (ST) are both classified as polyketides as their carbon chains, assembled from simple acetate units, commonly exhibit keto groups at alternate carbons (Yu and Leonard 1995). The diversity of the polyketide structure arises through variation in the length of the starter and extender units for the carbon skeleton (Brown, Yu *et al.* 1996). Assembly of the initial carbon skeleton of polyketide is catalysed by a large enzyme known as a polyketide synthase (PKS) (Bingle, Simpson *et al.* 1999).

1.4.2 Dothistromin Biochemistry

Dothistromin is a polyketide and belongs to a group of compounds with a common structural feature of a furobenzofuran moiety, many of which are potent carcinogens (Gallagher and Hodges 1972). These include AF, ST and versicolorins (Figure 1.1). Dothistromin was first isolated from the pine tree pathogen *D. pini* and was subsequently isolated from several *Cercospora spp.* and *Mycosphaerella laricina* (Stoessl, Abramowski *et al.* 1990).

The tetrahydro-2-hydroxy-bisfuran moiety of dothistromin is similar to that of aflatoxin B1 (AFB1), from *Aspergillus flavus* (Figure 1.1). In a ^{13}C nuclear magnetic resonance study of dothistromin biosynthesis the labelling pattern in this bistetrahydrofurano side chain was identical to those found in AF and ST. This structural feature of AFB1 is considered to be responsible for its hepatotoxicity and carcinogenic characteristics (Elliott, Mason *et al.* 1989). Due to these observations dothistromin was tested and shown to be mutagenic, there is also some evidence to suggest that it is a potential carcinogen (Elliott, Mason *et al.* 1989).

Aside from dothistromin, other anthraquinones have been isolated in *D. pini.* AF and ST precursors including averantin, averufin and versicolorin are among these, all of which have anthraquinones fused to dihydrogen rings (Danks and Hodges 1974). These similarities suggest that the fusion of rings in dothistromin intermediates could occur in the same manner as AF intermediates. Also, the structural similarity of dothistromin to AF and ST suggests once again that these compounds could have common intermediates (Shaw, Chick *et al.* 1978) (Figure 1.1).

1.4.3 Aflatoxin and Sterigmatocystin

Aflatoxins are secondary metabolites produced by four species of *Aspergillus; A. flavus, A. parasiticus, A. nomius* and *A. tamarii.* These are widespread fungi that infect stored grains and nuts causing contamination of food with carcinogenic aflatoxins. As AFB1 is one of the most potent natural carcinogens known and is also a known mycotoxin, AF contamination of agricultural commodities is not only a serious health hazard but also an economic issue

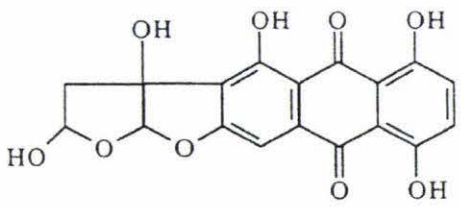
(Jelinek, Pohland *et al.* 1989; Bhatnagar and Cleveland *et al.* 2000). Because of the difficulty in effectively controlling AF contamination recent research efforts have been focused on developing an understanding of the molecular biology of the AF biosynthetic pathway.

The AFB1 pathway has been well characterised through a combination of biochemical and genetic studies (Figure 1.2). Genes in this pathway such as the gene responsible for the conversion of the intermediate versicolorin A into ST, and the genes responsible for the conversion of acetate to noranthrone, have been isolated, cloned and sequenced (Skory, Chang et al. 1992; Chang, Cary et al. 1995; Feng and Leonard 1995). Current evidence shows that the ST pathway in *A. nidulans* and the AF pathway share a common biosynthetic scheme (Keller and Hohn 1997).

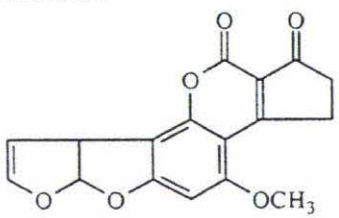
The genes responsible for the biosynthesis of AF and ST are clustered within the genome. Clusters of functionally related genes are a general feature of prokaryotic gene organisation, but are much less prevalent in eukaryotes. Fungal gene clusters are broadly defined as the close linkage of two or more genes that participate in a common metabolic or developmental pathway (Keller and Hohn 1997). Approximately 25 genes of the AF pathway are located within a 60-70 kb region (Yu, Chang et al. 2000) (Figure 1.3). Most of which encode enzymes with the necessary functions for aflatoxin biosynthesis. The functions of 19 genes have been assigned in AF biosynthesis whilst the functions of six genes remain to be assigned (Yu, Chang *et al.* 2004) (Figure 1.2). An ST cluster has also been found in *A. nidulans*, with 25 genes thought to be involved in ST synthesis identified (Figure 1.3). Eight of these genes have been shown to be required for ST biosynthesis and all 25 transcripts corresponding to the proposed genes are coordinately regulated (Brown, Yu *et al.* 1996). The order of homologous genes in this cluster differs from the AF cluster, but the gene products are very similar between the two (Brown, Yu *et al.* 1996). Genes involved in the bioconversion steps of the AF/ST biosynthetic pathway have been confirmed through either gene disruption or enzymatic studies (Yu, Chang *et al.* 2004). In both the AF and ST gene clusters there is a positive regulatory gene, *aflR,* for activating pathway gene transcription. The *aflR* gene encodes a sequence-specific zinc binuclear

Figure 1.1: Structures of Dothistromin, Aflatoxin, Sterigmatocystin and Versicolorin B

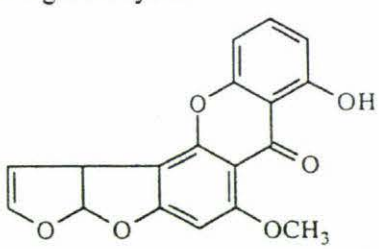
dothistromin



aflatoxin



sterigmatocystin



versicolorin B

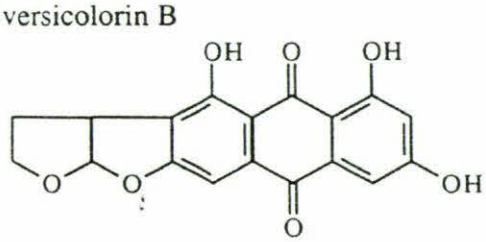


Figure 1.2: *A. parasiticus* AF Biosynthetic Pathway Gene Cluster with Comparison to the Biochemical Pathway

Figure 1.2 shows the action of genes on the aflatoxin biosynthetic pathway in *A. parasiticus* and *A. flavus*. The genes are listed with their function. The name in brackets corresponds to a new naming scheme (Yu, Chang *et al.* 2004). While the number corresponds to the pathway step they catalyse. Intermediates in bold are also found in *D. pini.* Genes are listed in the order they appear in the *A. parasiticus* AF gene cluster. The diagram is modified from *Seconi,* 2001 and Yu, Chang *et al.* 2004.

Gene	Product Function	Pathway step
norB (aflF)	Dehydrogenase	4
cypA (aflU)	P450 monooxygenase	Unassigned
aflT	Transmembrane protein	
pksA (aflC)	Polyketide synthase	2
nor-1 (aflD)	Reductase	4
fas-2 (aflA)	Fatty acid synthase-2	1
fas-1 (aflB)	Fatty acid synthase-1	1
aflR	Pathway regulator	
adhA (aflH)	Alcohol dehydrogenase	6 or 7
estA (aflJ)	Esterase	9
norA (aflE)	Aryl-alcohol dehydrogenase	4
ver-1 (aflM)	Dehydrogenase	12
verA (aflN)	Monooxygenase	12
avnA (aflG)	P450 Monooxygenase	5
verB (aflL)	Desaturase	11
avfA (aflI)	Oxidase	8
omtB (aflO)	Methyltransferase B	13
omtA (aflP)	Methyltransferase A	14
ordA (aflQ)	Oxidoreductase (p450 enzyme)	15
vbs (aflK)	Ver B synthase	10
cypX (aflV)	P450 Monooxygenase	Unassigned
moxY (aflW)	Monooxygenase	Unassigned
ordB (aflX)	Monooxygenase/oxidase	Unassigned
hypA (aflY)	hypothetical protein	Unassigned

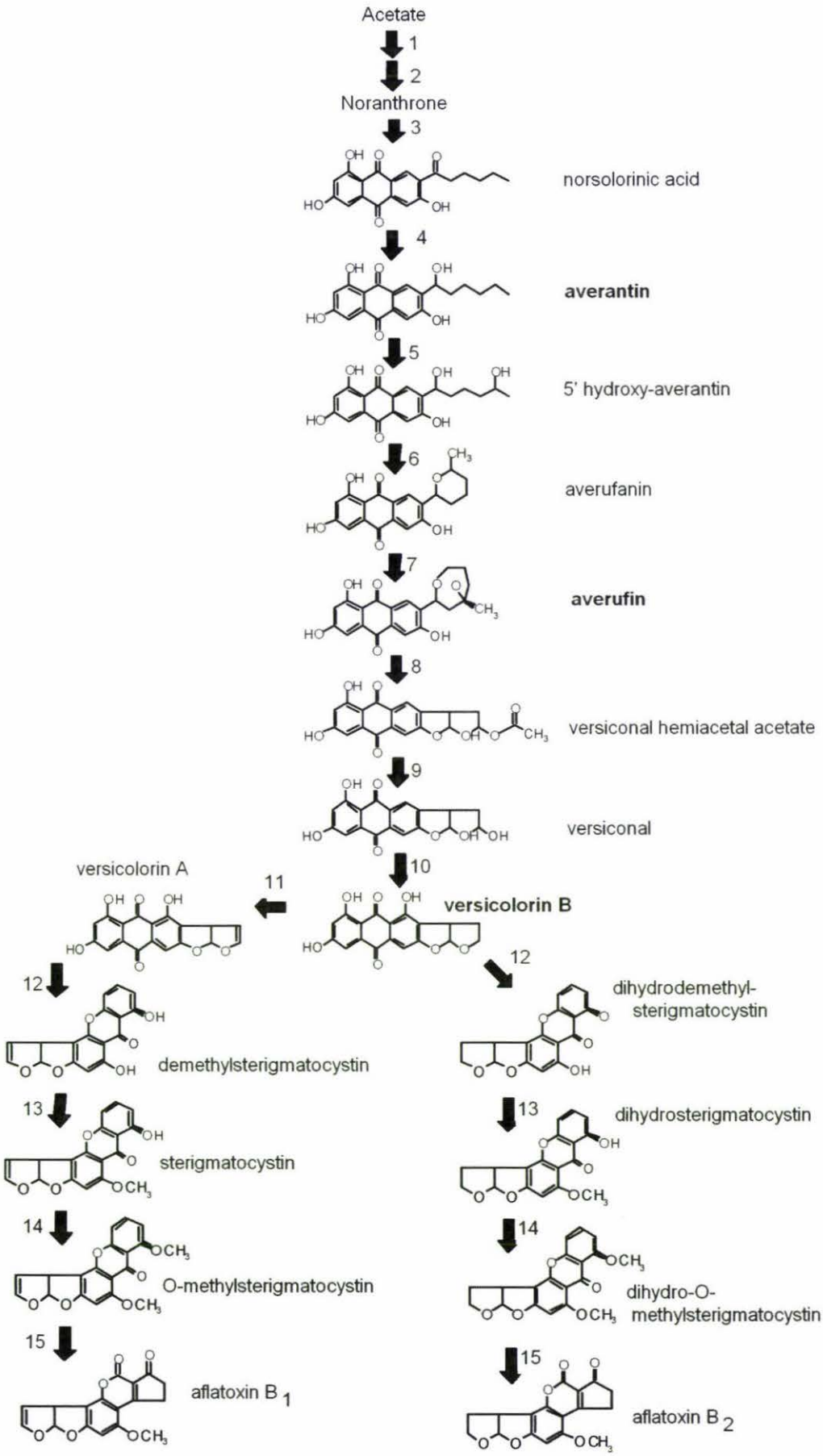
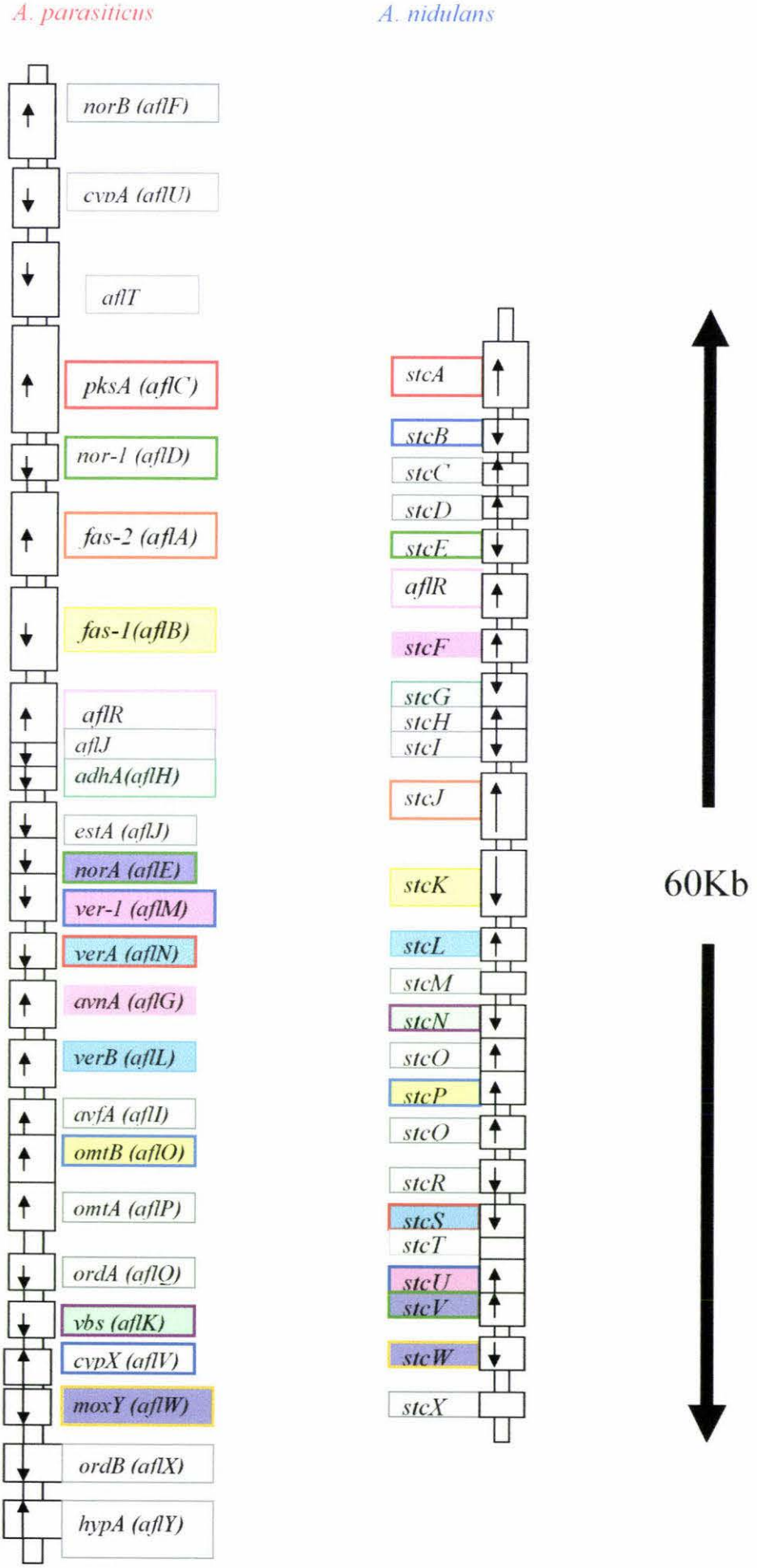

Acetate
1
2
Noranthrone
3
norsolorinic acid
4
averantin
5
5' hydroxy-averantin
6
averufanin
7
averufin
8
versiconal hemiacetal acetate
9
versiconal
10
versicolorin B
11
versicolorin A
12
demethylsterigmatocystin
13
sterigmatocystin
14
O-methylsterigmatocystin
15
aflatoxin B 1
12
dihydrodemethyl-sterigmatocystin
13
dihydrosterigmatocystin
14
dihydro-O-methylsterigmatocystin
15
aflatoxin B 2

Figure 1.3: Comparison Between *A. parasiticus* AF Gene Cluster and *A. nidulans* ST Gene Cluster.

The boxes in Figure 1.3 represent the position of genes within each cluster and arrows represent the direction of transcription. The coloured boxes contain the name of the gene. The names of homologs in each cluster are contained in the same coloured boxes. Grey boxes indicate homologs are unknown. Diagram modified from Seconi, 2001 and Yu, Chang *et al.* 2004.


A. parasiticus
A. nidulans
norB (aflF)
cypA (aflU)
aflT
pksA (aflC)
nor-1 (aflD)
fas-2 (aflA)
fas-1(aflB)
aflR
aflJ
adhA(aflH)
estA (aflJ)
norA (aflE)
ver-1 (aflM)
verA (aflN)
avnA (aflG)
verB (aflL)
avfA (aflI)
omtB (aflO)
omtA (aflP)
ordA (aflQ)
vbs (aflK)
cypX (aflV)
moxY (aflW)
ordB (aflX)
hypA (aflY)
stcA
stcB
stcC
stcD
stcE
aflR
stcF
stcG
stcH
stcI
stcJ
stcK
stcL
stcM
stcN
stcO
stcP
stcO
stcR
stcS
stcT
stcU
stcV
stcW
stcX
60Kb

DNA-binding protein and has been shown to be required for transcriptional activation of several of the structural genes (Yu, Chang *et al.* 2004).

The first step of the AF pathway involves the conversion of acetate to noranthrone (Figure 1.2). This involves firstly the conversion of acetate to hexanoic acid and is followed by the conversion of hexonyl CoA to noranthrone. A fatty acid synthase (FAS) has been identified as being responsible for the conversion of acetate to hexanoic acid. This FAS was identified in the aflatoxin gene cluster and shows significant similarity to the predicted amino acid sequence of the enoyl reductase and malonyl/ palmityl transferase domains of the FAS β-subunit encoded by the *fas1* of *Saccharomyces cerevisiae* (Yu, Chang *et al.* 1995). Through gene disruption and complementation studies, this gene, designated *fas-1*, has been shown to be required for norsolorinic acid (NA) biosynthesis and aflatoxin production in *A. parasiticus* (Payne and Brown 1998). A FAS has also been shown to have involvement in ST biosynthesis in *A. nidulans* (Brown, Adams *et al.* 1996).

A *pks* gene whose product is responsible for the conversion of hexanoic acid to noranthrone was isolated independently by two research groups in 1995 (Chang, Cary et al. 1995; Feng and Leonard 1995). The first reported *pks* for aflatoxin biosynthesis was identified in a cosmid clone that also contained the aflatoxin biosynthetic gene, *nor-1*. The *pks* gene was identified by its sequence similarity (80% amino acid similarity) to the β-ketosynthase and acyl transferase regions of the *wA* gene of *A. nidulans*, encoding the PKS for spore pigmentation (Mayorga and Timberlake 1992). Gene disruption of this putative pks, *pksA* in *A. parasiticus* showed the gene to be required for aflatoxin biosynthesis and not spore colour (Chang, Cary *et al.* 1995). In 1995 Feng & Leonard, also isolated a *pks* gene from *A. parasiticus* using PCR-derived PKS-specific probes to screen a cosmid library. Disruption of the *pksL1* gene resulted in transformants that failed to produce AF or any other AF intermediates (Feng and Leonard 1995). The amino acid sequences of *pksA* and *pksL1* have been aligned and these two genes are in fact the same, although they have been isolated from different *A. parasiticus* strains (Rosie Bradshaw personal communication). As hexonyl CoA is not detected as an intermediate in the AF pathway, it has been postulated that FAS and PKS act in a complex. The product of the conversion carried out by this complex is noranthrone. In fact, the first stable intermediate of the aflatoxin pathway is norsolorinic acid (NA) (Figure 1.2). The conversion of noranthrone to NA has

not yet been resolved but has been proposed to occur by a noranthrone oxidase, a monooxygenase, or to occur spontaneously (Payne and Brown 1998).

1.4.4 Fungal Polyketide Synthases

Polyketide synthases are generally one of two types: type I PKSs consist of very large multifunctional proteins, whereas type II PKSs consist of several monofunctional proteins (Yu and Leonard 1995). PKSs share conserved sequences and patterns with fatty acid synthases (FASs) (Feng and Leonard 1995).

Fungal PKS are large multifunctional proteins (type I PKS) encoded by a single gene and possessing up to eight types of functional domains: acyl transferase (AT), acyl carrier protein (ACP), ketosynthase (KS), ketoreductase, enoyl reductase, dehydratase, thioesterase and methyltransferase (Bingle, Simpson *et al.* 1999). A 'minimal PKS' contains an acyl transferase (AT) that selects the appropriate extender unit, a keto-acyl synthase (KS) that catalyses the condensation reaction, and an acyl-carrier protein (ACP) that holds onto the growing polyketide chain and takes the next extender unit from the AT domain ready for the next condensation reaction (Lal, Kumari *et al.* 2000).

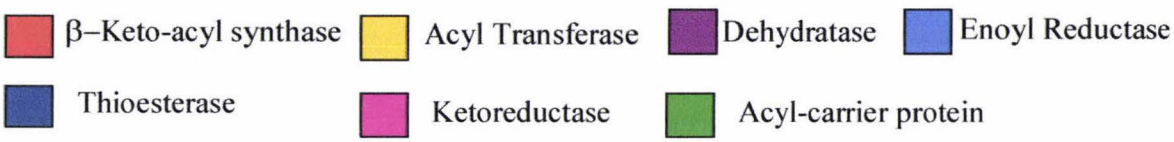
Several *pks* genes from filamentous fungi have been cloned. These include *A. nidulans* (Mayorga and Timberlake 1992; Yu and Leonard 1995), *A. parasiticus* (Feng and Leonard 1995), *Colletotrichum lagenarium* (Takano, Kubo *et al.* 1995) and *Cochliobolus heterostrophus* (Yang and Turgeon 1994). Figure 1.4 shows the arrangement of domains within each polypeptide. Three of these PKS proteins (*A. nidulans* WA and PKS^{St} and *C. lagenarium* PKSL1) contain duplicated ACP domains. The significance of having tandem ACP domains is not yet clear, especially as other PKS proteins in very similar pathways (such as the *A. parasiticus* PKS involved in the aflatoxin biosynthetic pathway (PKSL1)) only have one ACP domain (Yu and Leonard 1995). Not all *pks* genes have a TE domain in the same ORF. This domain is present in both *A. nidulans* PKS^{ST} and *A. parasiticus* PKSL1. The suggested function of this TE domain is to release the polyketide product from the polyketide synthase complex. The PKSL from *C. heterostrophus* contains all the domains for processing β-carbons except the TE domain. It is thought this domain may be encoded by a separate gene.

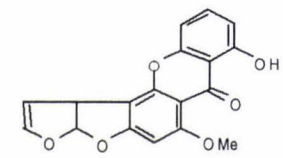
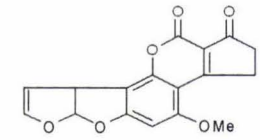
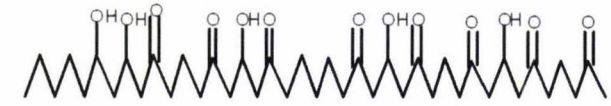
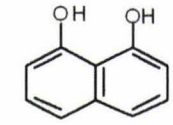
1.4.5 Dothistromin Molecular Biology

The structural and chemical similarities of dothistromin to the mycotoxins ST and AF produced by *Aspergillus* species and the sequences of genes involved in these pathways that have been cloned, have been used to isolate and characterise putative genes in the dothistromin biosynthetic pathway. Due to the similarities between AF and ST pathways with the dothistromin pathway, sequences of genes known to be on AF/ST pathways were used as probes for *D. pini* genomic library. Probing a *D. pini* genomic library with a fragment of the *Aspergillus parasiticus ver-1* gene identified two clones λCGV1 and λCGV2 (Gillman 1996). Sequencing of λCGV1 revealed four putative dothistromin pathway genes located within a 13.3kb genomic region: *dotA* (dothistromin ketoreductase), *dotB* (dothistromin oxidase), *dotC* (dothistromin toxin pump) and *dotD* (dothistromin thioesterase). The functions of the genes identified were deduced through sequence comparison to AF and ST cluster genes and other sequences encoding functional proteins (Table 1.1 and Figure 1.5). A Southern blot analysis of *D. pini* DNA restriction fragments confirmed the cluster and also indicated that a single copy of each gene was present. Partial characterisation of the λCGV2 clone revealed another potential *ver* gene, although this showed greater similarity to the melanin biosynthetic gene *phn* than the aflatoxin gene *ver-1* (Laarakkers 1999) (Table 1.1 and Figure 1.5). A Southern blot analysis of *D. pini* genomic DNA was also carried out using a probe (KS-2) encoding the highly conserved β keto-synthase domain from the polyketide synthase gene (*pksL1*) of *A. parasiticus*. This indicated the presence of a homologous sequence in *D. pini*. Subsequent isolation of the KS-2 hybridising lambda clone (λKSA) allowed a 2411bp fragment to be subcloned and sequenced (Morgan 1997). Sequence analysis revealed two functional protein domains, β keto-acyl synthase (KS) and acyl transferase (AT), both of which are present in polyketide synthases. The predicted amino acid sequence showed approximately 60% identity with *A. nidulans wA* and *A. parasiticus pksL1* (Morgan 1997). BLASTX searching also revealed similarities with a number of FASs, although the alignment was not as strong as those of PKSs. On the basis of these results, it is proposed that the 2.4 kb fragment encodes part of *D. pini* PKS (PKS^{Dot}) (Morgan 1997). Further characterisation of the λKSA clone enabled the completion of a double stranded sequence of the putative pks^{dot} gene region (Laarakkers, 1999 and Seconi, 2001) (Table 1.1 and Figure 1.5). Primers walking within

Figure 1.4: Comparison of PKSs in Different Species of Fungi

Figure 1.4 shows the arrangement of conserved domains within single PKS polyketide and the variety of final products resulting from the pathways the PKS acts on. Diagram from Seconi, 2001.



Organism	Gene	Protein structure	Product of Biosynthetic Pathway[1]
A. nidulans	*pks*ST (Yu & Leonard, 1995)	KS AT ACP ACP TE	 Sterigmatocystin
A. parasiticus	*pksL1* (Chang *et al*, 1995)	KS AT ACP TE	 Aflatoxin
C. heterostrophus	*pks1* (Yang *et al*, 1994)	KS AT DH ER KR ACP	 T-toxin
C. lagenarium	*pksL1* (Takano *et al*, 1995)	KS AT ACP ACP	 1, 8-dihydroxynapthalene
A. nidulans	*wA* (Mayorga and Timberlake1992)	KS AT ACP ACP	Unknown spore pigment

[1] This table shows the final product of the biosynthetic pathway these PKSs are on, not the intermediate immediately following the PKS catalysed reaction.

Table 1.1 Homologs of Putative Dothistromin Biosynthetic Genes

D. pini gene	Reference	Homologs	% aa identity (% similarity)	Organism	Function	Reference
dotA	Monahan (1998)	*ver-1*	80 (87)	*Aspergillus parasiticus*	Ketoreductase- AF biosynthesis	(Skory, Chang *et al.* 1992)
		stcU	79 (88)	*A. nidulans*	Ketoreductase- ST biosynthesis	(Keller, Kantz *et al.* 1994);
		BRM2	65	*Alternaria alternata*		(Brown, Yu *et al.* 1996)
		thr1	59 (73)	*Colletotrichum lagenarium*	Reductase- melanin biosynthesis	(Shimizu, Takana *et al.* 1997)
dotB	Monahan (1998)	*stcC*	24 (29)	*A. nidulans*	Oxidase- ST biosynthesis	(Brown, Yu *et al.* 1996)
		Chloro-peroxidase gene	24 (31)	*Caldariomyces fumago*	Chloroperoxidase- biosynthesis of chlorinated secondary metabolites	(Nuell, Fang *et al.* 1988)
dotC	Monahan (1998)	*aflT*	31	*A. parasiticus*	Putative AF pump	(Bhatnagar, Ehrlich *et al.* 2003)
		ToxA	31 (41)	*C. carbonum*	HC- toxin efflux pump	(Pitkin, 1996)
dotD	Monahan (1998)	TE[1] from:				
		stcA	37 (82)[2]	*A. nidulans*	PKS-ST biosynthesis	(Yu and Leonard, 1995)
		pksL1	35 (45)	*A. parasiticus*	PKS- AF biosynthesis	(Feng and Leonard, 1995)
		albl	43.6	*A. fumigatus*	PKS-conidial pigmentation biosynthesis	(Tsai, Chang *et al.* 1998)
		pks1	35 (42)	*C. lagenarium*	PKS-melanin biosynthesis	(Takano, Kubo *et al.* 1995)

Table 1.1 Homologs of Putative Dothistromin Biosynthetic Genes

D. pini gene	Reference	Homologs	% identity (% similarity)	Organism	Function	Reference
pks^{dot} (partial)	Seconi (2001) and Laarakkers (1999)	*pksA*	62 (74)	*A. parasiticus*	Polyketide synthase (PKS), AF biosynthesis	(Chang, Cary *et al.* 1995)
		stcA	63 (74)	*A. nidulans*	PKS, ST biosynthesis	(Yu and Leonard, 1995)
		Alb1(pksP)	42 (58)	*A. fumigatus*	PKS-conidial pigmentation biosynthesis	(Tsai, Chang *et al.* 1998)
		wA	42 (58)	*A. nidulans*	Green conidial pigment synthase	(Mayorga and Timberlake, 1992)
cyp^{dot}	Seconi (2001)	*cypX*	59 (74)	*A. parasiticus*	Cytochrome p450 monooxygenase,	(Yu, Chang *et al.* 2000)
		stcB	56 (68)	*A. nidulans*	AF biosynthesis Cytochrome p450 monooxygenase, ST biosynthesis	(Brown, Yu *et al.* 1996)

[1] Thioesterase domain

[2] Identities and similarities to TE domains of the gene only

subclones of the λKSA clone also revealed *cyp*^{dot} which has notable identity to known p450 cytochrome monooxygenases (Seconi, 2001) (Table 1.1 and Figure 1.5).

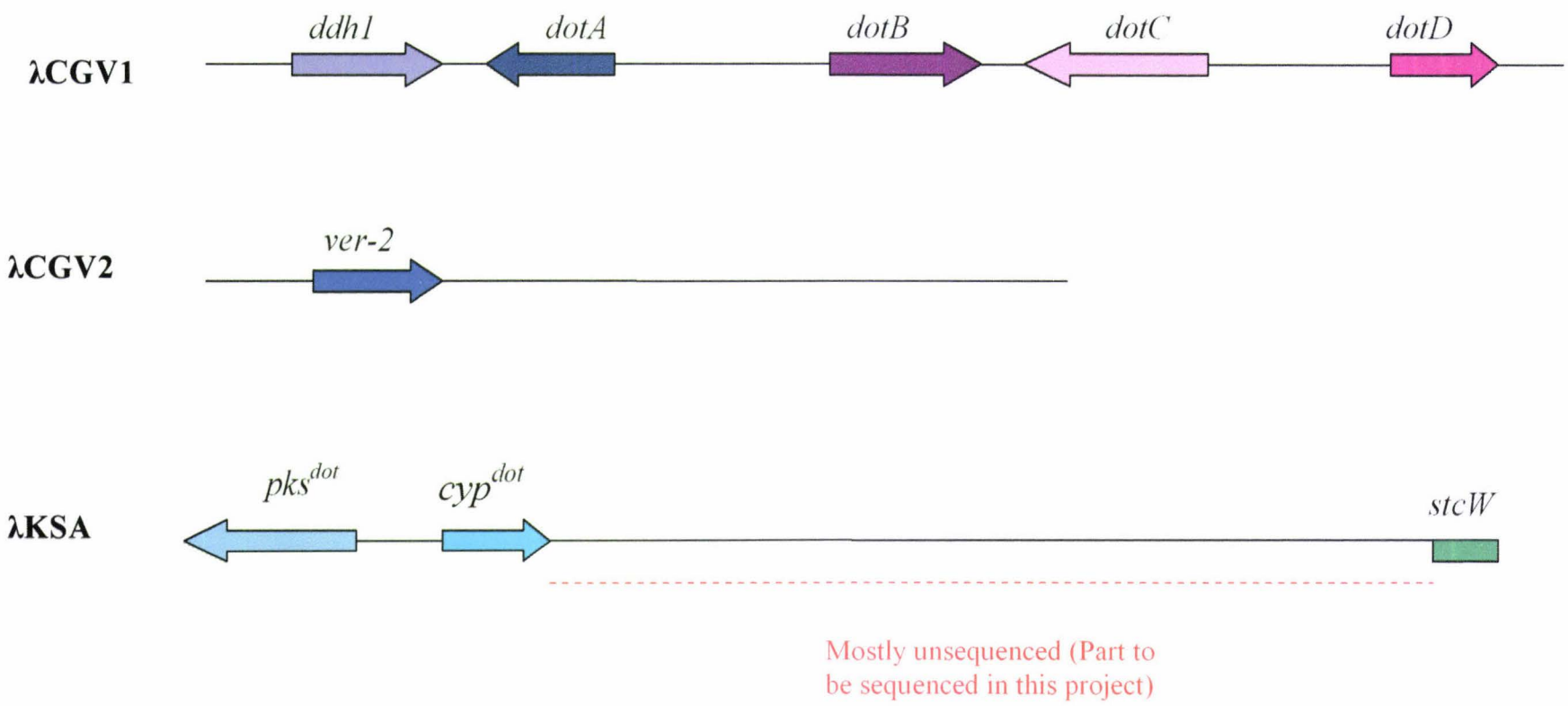
Although two genes have been found within the λKSA clone (*cyp*^{dot} and part of *pks*^{dot}) the λKSA clone remains partly uncharacterized (Figure 1.5). This study will investigate one of the subclones within the λKSA to identify other putative genes of the dothistromin biosynthetic gene cluster. Also to date, the positioning of the three lambda clones λKSA, λCGV1 and λCGV2 with respect to one another within the genome is unknown (Figure 1.5). This study will also attempt to identify the position of the three lambda clones with respect to one another. It is expected that at least λKSA and λCGV1 will be within close proximity to one another within the genome as both these clones contain genes which have high identity to genes within the AF/ST gene clusters. Due to the high similarity to AF/ST genes, the putative genes found within the λKSA and λCGV1 clones are postulated to be part of the cluster of genes required for dothistromin biosynthesis.

Sequencing of the λCGV1 clone revealed *dotD* which has high identity to the thioesterase (TE) domains of other fungal *pks* genes (Table 1.1). This suggests that like *C. heterostrophus*, the TE domain of PKS^{Dot} (the partial sequence of which has been identified on the λKSA clone) may be on a separate polypeptide to the rest of the PKS. The role of this domain is to release the polyketide product from the polyketide synthase complex (Yang and Turgeon 1994). Two different polyketide synthase gene sequences have been published for two different strains of *A. parasiticus* (*pksA* (Chang, Cary *et al.* 1995) and *pksL1* (Feng and Leonard 1995). While only one group has reported the presence of a thioesterase domain, both sequences are very similar and do in fact contain this domain (Bradshaw, Bhatnagar *et al.* 2002). Other fungal polyketide synthase genes such as the *wA* spore pigment polyketide synthase do not contain this thioesterase domain (Mayorga and Timberlake 1992). The similarity of *pks*^{dot} to *A. parasiticus pks* genes (*pksA* and *pksL1*), which have been shown to be involved in the first step of the aflatoxin pathway (conversion of acetate to noranthrone, which is thought to be two stage reaction) suggest that *pks*^{dot} may have a similar role in dothistromin biosynthesis. This study will investigate if *pks*^{dot} is essential for dothistromin biosynthesis in *D. pini.*

Figure 1.5 ***D. pini*** **clones showing putative dothistromin biosynthetic genes.**

A schematic diagram (not to scale) showing *D. pini* λ clones that contain open reading frames thought to be involved in dothistromin biosynthesis. Coloured arrows indicate putative dothistromin biosynthetic genes. Gene names are written below. The gene *ddh1* is not thought to be involved in the dothistromin biosynthetic pathway and is thought to designate one end of the dothistromin gene cluster.

The orientation of clones relative to each other within the genome is unknown at this time.


λCGV1
ddh1
dotA
dotB
dotC
dotD
λCGV2
ver-2
λKSA
pks^dot
cyp^dot
stcW
Mostly unsequenced (Part to
be sequenced in this project)

1.5 TARGETED GENE DISRUPTION

Targeted gene disruption in filamentous fungi has been demonstrated during the last decade by a number of groups (Upchurch, Walker et al. 1991; Feng and Leonard 1995; Shiotani and Tsuge 1995; Bowden, Smalley et al. 1996). Gene targeting involves replacing or disrupting a wild type gene in the organism with another gene (often one that allows positive selection of transformants). This disruption or replacement prevents the wild type gene from functioning. Homologous recombination of exogenous DNA permits the disruption and replacement of the wild type genes, with genomic segments that have been altered *in vivo*. This technique is a powerful tool for evaluating gene functions and constructing well defined mutations (Shiotani and Tsuge 1995). It permits the pathogenic effects of fungal species on plants to be observed as toxin deficient mutants are able to be generated. Gene disruption has been demonstrated in many filamentous fungi such as *Alternaria alternata* (Shiotani and Tsuge 1995) using a gene hygromycin B phosphotransferase from *E. coli* that confers resistance to the eukaryotic antibiotic hygromycin B1. This gene is placed under control of different promoters and is referred to as the *hph* cassette. The replacement is made possible by flanking regions around the *hph* cassette that is homologous to regions of the gene being targeted. Another commonly used marker in gene disruption of filamentous fungi provides resistance to phleomycin (*Phle*). Filamentous fungi have proven to be more recalcitrant to gene disruption in comparison to yeast. Homologous recombination and gene targeting using the *hph* cassette has been well developed in *A. alternata*. Shiotani and Tsuge (1995) found that both linear and circular molecules were successfully integrated into the genome. When using circular molecules the frequency of gene targeting was proportional to the length of the homologous sequence (Shiotani and Tsuge 1995).

To study the pathogenicity of fungal toxins, toxin deficient mutants have been generated using both UV mutagenesis and targeted gene disruption. Examples of this are seen in members of the fungal genus *Cercospora* that are pathogens of a variety of economically important crops including soybean and maize. One member of this fungal genus, *C. kikuchii,* infects soybeans causing the disease 'Purple Seed Stain'. Evidence suggests that cercosporin, a non-host specific phytotoxin, is responsible for the pathogenicity of the fungus. Cercosporin is light-activated and light has been shown to be important in the

development of disease (Upchurch, Walker *et al.* 1991). Treatment of soybean tissue with pure cercosporin elicits disease symptoms and cercosporin has been isolated from lesions of infected plants. Spontaneous and UV induced mutants in the cercosporin biosynthetic pathway which failed to produce cercosporin were used to infect soybean and were shown to be non-pathogenic. This suggests that the toxin plays a vital role in the pathogenicity of the fungus (Upchurch, Walker *et al.* 1991).

Examples also exist in which toxin deficient mutants retain wild type pathogenicity (Bowden, Smalley *et al.* 1996; Kawamura, Tsujimoto *et al.* 1999). *Ophitoma novo-ulmi* is a fungus that infects the xylem and is pathogenic to elm trees. It has been proposed that cerato-ulmin (CU) a hydrophobin produced by *O. novo-ulmi* is a pathogenicity factor in Dutch elm disease. CU^- mutants were made by targeting a gene involved in CU production (Bowden, Smalley *et al.* 1996). The CU^- mutants, whilst unable to produce CU, were still able to produce the symptoms of Dutch elm disease (i.e. still had the pathogenicity of wild type). This suggested that CU did not have a direct role in the pathogenic effects of the fungus (Bowden, Smalley *et al.* 1996).

Methods are also available for creating non-specific disruption mutants, where plasmid integration is able to occur at random sites within the genome. One of these methods restriction enzyme-mediated integration (REMI) insertional mutagenesis has been shown to increase the transformation frequency of plasmid integration. In *A. alternata* REMI has been optimised for integration of the plasmid pAN7-1 conferring resistance to hygromycin B. It was found that using the restriction enzyme *Hind* III was most effective for increasing the transformation frequency (Akamatsu, Itoh *et al,* 1997). Addition of the restriction enzyme presumably increases the transformation frequency by allowing integration of the plasmid where the genome has been nicked by the digestion enzyme. REMI can be used to create libraries of plasmid-tagged transformants which can then be screened for toxin production (Akamatsu, Itoh *et al,* 1997). Another method is transposition-arrayed gene knockouts (TAGKO) which involves using transposons carrying a bacterial and fungal drug resistance marker to mutagenise individual cosmids or entire libraries *in vitro*. *In vitro* transposition into cosmid libraries is used to create gene sequencing templates. Subsequent sequencing and analysis of these templates creates an annotated collection of unbiased insertional gene disruption vectors (Hamer, Adachi *et al,* 2001).

Gene disruption is able to give conclusive evidence as to whether a putative biosynthetic gene is involved in a proposed toxin pathway. An example of gene disruption being used for confirmation of biosynthetic action is the polyketide synthase (*pksL1)* gene of *A. parasiticus*. In this example, the *pksL1* gene was disrupted by gene targeting and the *pksL1*⁻ mutants did not produce NA (the first stable intermediate of the AF pathway) or AF, there was also no accumulation of any new intermediates. This indicated that it is highly likely that *pksL1* functions in polyketide biosynthesis during AF production, by converting acetate to noranthrone (Feng and Leonard 1995).

The transformation system using the pAN7-1 plasmid which contains an *hph* cassette originally developed in *A. nidulans* and *A. niger*, has been used successfully in *D. pini*. Most transformants contained a single copy of the pAN7-1 vector integrated at differing sites in each case (Bradshaw, Bidlake *et al.* 1997). Another plasmid which confers hygromycin resistance is pCWhyg. The *hph* cassette from this plasmid was successfully used in the disruption of *dotA* (a gene homologous to *ver-1* of the aflatoxin biosynthetic pathway) (Bradshaw, Bhatnagar *et al.* 2002). The disruption of this *dotA* gene leads to accumulation of an intermediate, versicolorin A. Pathogenicity testing of *dotA*⁻ mutants is currently in progress in Dr R Bradshaw's laboratory, Massey University.

1.6 Aims and Objectives

The primary focus of this project was to obtain a mutant blocked in the synthesis of dothistromin via targeted disruption of an early pathway gene pks^{dot}. This would allow direct assessment of the role of dothistromin in pathogenicity of *D. pini*. The putative polyketide synthase gene (pks^{dot}) was replaced using a disruption vector. This disruption vector was used to replace the two essential domains, KS (ketosynthase) and AT (acetyltransferase), with an *hph* cassette which confers hygromycin resistance. The phenotype of these pks^{dot-} mutants will be assessed with regards to growth rate, dothistromin deficiency and accumulation of pathway intermediates.

The secondary focus of this project was to further characterise genes involved in dothistromin biosynthesis. This was achieved by continuing sequencing of the λKSA clone (commenced by Seconi, 2001). Putative genes were identified through BLAST analysis (Altschul *et al.* 1990).

Thirdly, Southern blot analysis was used in an attempt to identify how the three lambda clones λKSA, λCGV1 and λCGV2 are positioned relative to one another within the *D. pini* genome. The positioning of putative genes (previously identified within the λKSA clone) within the genome was also confirmed.

2.0 METHODS AND MATERIALS

2.1 FUNGAL AND BACTERIAL STRAINS, LAMBDA CLONES AND PLASMIDS

Fungal and bacterial strains, Lambda (λ) clones and plasmids used in this study are listed in Table 2.1.

2.2 MEDIA

All media prepared with Milli-Q water and sterilized by autoclaving at 121° C for 15 minutes. Media were cooled to approximately 50°C before addition of antibiotics. See appendix six for list of media.

2.2.1 Luria Broth (LB)

(g/L): tryptone, 10; NaCl, 5, Yeast extract, 5; glucose, 1.

2.2.2 LB Solid Media

(g/L): tryptone, 10; NaCl, 5; yeast extract, 5; glucose, 1; agar, 15.

2.2.3 Selective LB Media

Supplements added at the following concentrations.
(μg/mL): ampicillin, 100; isopropylthio-β-D-galactoside (IPTG), 30; 5-bromo 4-chloro 3-indolyl-β-D-galactoside (X-gal), 60; kanamycin sulfate, 50; Chloramphenicol, 30.

2.2.4 *D. pini* Media (DM)

(g/L): Malt extract (Oxoid), 23; Nutrient Agar (Oxoid), 28

Table 2.1 **Lambda (λ) clones, plasmids, Fungal and Bacterial Strains**

Strain, Plasmid, λ Clone	Revevant Characteristics	Source or Reference
Fungal Strain		
Dothistroma pini		
NZE5 aka 3-3-1-1A	Single spore isolate, laboratory strain.	Kinleith, New Zealand (Hirst 1997)
NZE7	Single spore isolate, laboratory strain.	Te Mati Forest, New Zealand (2003)
Bacterial Strains		
Escherichia coli.		
XL1-Blue	*sup* E44 *hsd* R17 *rec* A1 *end* A1 *gyr* A46 *thi rel* A1 *lac* - F' [*pro* AB+ *lac* Iq Δ *(lac* Z) M15 *Tn* 10 (*tetr*)]	(Bullock, 1987)
Top10	F- *mcr* A Δ(mrr-*hsd* RMS-*mcr* BC) Φ80*lac* ZΔM15 Δ*lac* X74 *rec* A1 deoR *ara* D139 Δ(*ara-leu)* 7697 *gal* U *gal* K *rps* L (StrR) *end* A1 *nup* G	Invitrogen
DB3.1	F- *gyr* A462 *end* A1 Δ(*sr* 1-*rec* A) *mcr* B *Mrr hsd* S20(r_B^-, m_B^-) *sup* E44 *ara* 14 *gal* K2 *lac* Y1 *pro* A2 *rps* L20(Sm^r) *xyl* 5 Δ*leu mtl* 1	Invitrogen
KW251	F- *sup* E44 *sup* F58 *gal* K2 *gal* T22 *met* B1 *hsd* R2 *mcr* B1 *mcr* A- *arg* A81:*Tn* 10 *rec* D1014	Murray *et al. 1997*
Lambda Clones		
λCGV1	LambdaGEM-12 (Promega) containing genomic DNA (including *dotA*) from NZE1 (DP002).	(Gillman 1996)
λCGV2	LambdaGEM-12 (Promega) containing genomic DNA (including *phn* 1) from NZE1 (DP002).	(Gillman 1996)
λKSA	LambdaGEM-12 (Promega) containing genomic DNA (including pks^{dot}) from NZE1 (DP002).	(Morgan 1997)
Plasmids		
pCWhyg (pR174)	5 kb Hmb^R Amp^R	(Punt and van den Hondel 1992)
pAN7-1 (pR36)	6.5 kb Hmb^R Amp^R	(Punt and van den Hondel 1992)
pGEM-T	Amp^R, lacZ' (3.0kb)	Promega
pGEM-T easy	Amp^R, lacZ' (3.0kb)	Promega
pDONR P4-P1R	4.77 kb Cm^R, $kanamycin^R$, *ccd* B gene	Invitrogen
pDONR 221	4.76 kb Cm^R, $kanamycin^R$, *ccd* B gene	Invitrogen
pDONR P2R-P3	4.77 kb Cm^R, $kanamycin^R$, *ccd* B gene	Invitrogen
pDEST R4-R3	Amp^R, Cm^R, *ccd* B gene	Invitrogen
pMSGW	6 kb Amp^R	Invitrogen
pR145	PUC118 containing 2.0 kb *Bam* H1/*Eco* R1 *Ver 2* fragment from λCGV2	Laarakkers, 1999
pR162	PUC118 containing a 4.5 kb *Bam* HI fragment from λKSA	Laarakkers, 1999
pR162A	PUC118 containing a 4.5 kb *Bam* HI fragment from λKSA	This study

Strain, Plasmid, λ Clone	Revevant Characteristics	Source or Reference
pR163	pUC based plasmid containing 3.5 kb subclone from λKSA	Laarakkers, 1999
pR181	pUC based plasmid containing 0.9 kb subclone from λKSA	Laarakkers, 1999
pR209	pGEM-T based plasmid containing a 0.295 kb fragment from λKSA	Seconi, 2001
pR212	pGEM-T easy based plasmid containing a 0.324 kb *aflR* fragment from λCGV2	This study
pR213	pGEM-T easy based plasmid containing a 0.380 kb *ddhA* fragment from λCGV2	This study
pR215	pGEM-T easy based plasmid containing a 0.801 kb insert of *dotD* from λCGV1	This study
pR216	pGEM-T based plasmid containing a 1.315 kp insert of *pks 5 ' flank*	This study
pR217	pGEM-T based plasmid containing a 0.823 kb insert of *pks 3 ' flank*	This study
pR219	PUC118 containing a 7 kb *Bam* HI fragment from λKSA	This study
pR220	pDONR P4-P1R containing a 1.35 kb insert of *3 ' pks* flank and *att* B sites	This study
pR221	pDONR P2R-P3 containing a 0.85 kb insert of *5 ' pks* flank and *att* B sites	This study
pR225	pDONR 221 containing a 2.4 kb *hph* cassette insert with *att* B sites	This study
pR226	pDEST R4-R3 containing three inserts: 1.35kb *3' pks* flank; 0.85kb *5' pks* flank; 2.4 kb *hph* cassette	This study
pR227	pGEM-T based plasmid containing a 2.4 kb insert of an *hph* cassette and *att* B sites	This study

2.2.5 *D. pini* Broth (DB)

(g/L): nutrient broth (Oxoid), 50; Malt extract (Oxoid), 50.

2.2.6 *D. pini* Sporulation Media (DSM)

(g/L): Malt extract (Oxoid), 20; Yeast extract (Oxoid), 5; Agar, 15.

2.2.7 SOC Media

(g/L): Tryptone, 20; yeast extract, 5; NaCl, 0.5; 250 mM KCL. The pH was adjusted to 7.0. After autoclaving and cooling 1 mL of sterile 1 M glucose was added to each 50 mL aliquot and before use 250 µL of 2 M $MgCl_2$ was added to each 50 mL aliquot.

2.2.8 TB Top Agarose

(g/L): Tryptone, 10; NaCl, 5; Agarose, 8.

2.2.9 *D. pini* Top Media (DM Top)

(g/L): Malt extract, 50; Nutrient Agar, 11.2; Sucrose, 273.9 (0.8 M)

2.2.10 *D. pini* Media With Sucrose (DM Suc)

(g/L): Malt extract, 50; Nutrient Agar, 28; Sucrose, 273.9 (0.8 M)

2.2.11 Selective DSM

Media used to select for hygromycin resistant *D. pini* transformants contained (µg/ mL): Hygromycin B (Sigma), 70

2.3 GROWTH AND MAINTENANCE OF CULTURES

2.3.1 Growth and Maintenance of *E. coli* Cultures

E. coli cultures were grown at 37°C overnight on LB plates, or in LB broth, with the appropriate selection supplements (section 2.2.3). Plates were sealed with parafilm and stored at 4°C. All cultures were regularly subcultured by streaking onto fresh LB plates. For long-term storage glycerol stocks were prepared by growing *E. coli* to an OD of 0.2-0.6 (absorbance at 600 nm) then mixed with sterile glycerol (10% of total volume) and stored at -80°C, after being snap frozen in liquid nitrogen.

2.3.2 Growth and Maintenance of *D. pini* Cultures

Subcultured *D. pini* cultures were grown on DM or DSM (section 2.2.4 and 2.2.6) plates at 22°C in ambient light for 14-28 days. Three 2 x 2mm^2 pieces of *D. pini* mycelia (cut with a scalpel blade) were ground in 1 mL of sterile milli-Q water using a plastic grinder in a microcentrifuge tube. From this, 100 µL was spread onto DM plates which had been overlaid with a sterile cellophane disc. Plates were then sealed with parafilm and grown at 22° C for 7- 14 days. Alternatively one colony (approximately 1.6 cm in diameter) was ground in 4 mL of water. From this, 300 µL was used to inoculate 50 mL of DB broth (section 2.2.12) in 500 mL flasks. Flasks were incubated at 22° C shaking at 150 rpm for 6-7 days. Mycelium from flasks were harvested by vacuum pump.

2.3.3 Plating Bacteria for Bacteriophage

50 mL of LB broth (section 2.2.1) was inoculated with a single bacterial colony (KW251) supplemented with 0.2% maltose in a 250 mL flask and grown overnight to an OD_{600} of 2.0. Cells were then harvested by centrifugation at 4,000 g (6000 rpm, SS34) for 10 minutes and the pellet resuspended in sterile 0.01 M $MgSO_4$. Cells were stored at 4° C for up to 3 weeks.

2.3.4 Single Spore Purification

In order to obtain a genetically pure colony of *D. pini*, single spores were isolated and grown as described. A sterile loop of milli-Q water was wiped over the colony (growing on DSM; section 2.2.6) to be purified. This loop was then streaked across a new DSM plate (selective in the case of transformants), the plate was then turned approximately 90° and streaked again from the previous streak three times, flaming the loop between each of the streaks. This plate was grown up until a single colony was a sufficient size to repeat the streaking process. After streaking on two consecutive plates a single colony was isolated and considered to be pure.

2.3.5 Growth of *D. pini* From Transformed Protoplasts

D. pini protoplasts were gently spread onto 20 mL plates of DM Suc (section 2.2.10). 5 mL DM top agar (section 2.2.9) was added 24 hours after spreading over the protoplasts. The DM top agar contained hygromycin to give a final volume of 70 µg/ mL (per 25 mL). Protoplasts were grown for 14 days at 22°C. Colonies growing on hygromycin were subcultured and grown on DSM (section 2.2.11) for 14-21 days before finally subculturing onto DSM/ 70 µg/ mL hygromycin.

2.4 COMMON BUFFERS AND SOLUTIONS

All solutions were prepared with Milli-Q water and sterilized by autoclaving at 121°C for 15 minutes, unless otherwise stated. See appendix six for list of common buffers and solutions.

2.4.1 TE Buffer

10 mM Tris-HCL and 1 mM Na_2EDTA (TE 10:1) or 10 mM Tris-HCL and 0.1 mM Na_2EDTA (TE 10: 0.1) prepared from 1 M Tris-HCL (pH 8.0) and 250 mM Na_2EDTA (pH 8.0) stock solutions.

2.4.2 1 x TAE Buffer

40 mM Tris-HCL, 2 mM Na_2EDTA and 20 mM acetic acid.

2.4.3 1 x TBE Buffer

89 mM Tris-HCL, 2.5 mM Na_2EDTA and 89 mM boric acid (pH 8.3).

2.4.4 Ethidium Bromide

Agarose gels were stained in ethidium bromide prepared as follows: 1 µL of 10 mg/mL stock per 10 mL of Milli-Q water to give a final concentration of 1 µg/ mL.

2.4.5 Gel Loading Buffer (10x)

3 M urea, 50% (v/v) glycerol, 50 mM Tris acetate, 0.4% (w/v) bromophenol Blue and 0.4% (w/v) xylene cyanol.

2.4.6 STC Buffer

1.2 M sorbitol, 50 mM Tris-HCL (pH 8.0), 50 mM $CaCl_2$.

2.4.7 SM Buffer

10 mM NaCl, 8 mM $MgSO_4.7H_2O$, 50 mM Tris-HCL (pH 7.5), 0.4% gelatine.

2.4.8 PEG Solutions

20% PEG, 2 M NaCl in SM buffer (section 2.4.7) or 30% PEG, 30 mM, NaCl in SM buffer or 40% PEG, 50 mM NaCl in STC buffer (section 2.4.6)

2.4.9 ST Buffer

1.0 M sorbitol, 100 mM Tris-HCL (pH 8.0).

2.4.10 OM Buffer

1.6 M $MgSO_4.7H_2O$, with 10 mM Na_2HPO_4/100 mM $Na_2H_2PO_4$ buffer (pH 5.8).

2.4.11 Chemiluminescent Detection Reagents

Washing Buffer

0.1 M maleic acid, 0.15 M NaCl, 0.3% tween 20 (v/v)

Maleic acid Buffer

0.1 M maleic acid, 0.15 M NaCl. pH 7.5

Detection Buffer

0.1 M Tris-HCL, 0.1 M NaCl. pH 9.5

Blocking Solution

1% blocking reagent (Roche Molecular Biochemicals), in Maleic acid buffer

Antibody Solution

Anti-Digoxigenin-AP diluted 1: 10, 000 in Blocking Solution

2.4.12 Reagents for Southern Blotting

Denaturing Solution

0.5 M NaOH, 1.5 M NaCl.

Neutralising Solution

500 mM Tris (pH 7.4), 2 M NaCl.

20 x SSC

3 M NaCl, 0.2 M tri-sodium citrate

2 x SSC

0.5 M NaCl, 0.03 M sodium citrate.

Wash Solution One

1 x SSC, 0.1% (w/v) SDS.

Wash Solution Two (1 x SSC)

0.25 M NaCl, 0.015 M sodium citrate.

2.4.13 Hybridisation Solution

0.05 M HEPES, 3 x SSC, 0.018% (w/v) herring sperm DNA (Sigma), 1% (w/v) SDS, 0.02% (w/v) *Escherichia coli* tRNA, 2% (w/v) ficoll (Sigma 70), 2% (w/v) bovine serum albumin (Sigma), 2% (w/v) polyvinylpyrrolidone (Sigma PVP-10).

2.4.14 SEVAG

Chloroform: isoamylalcohol, 24: 1 (v/v)

2.4.15 TES Buffer

(g/L): 100 mM Tris, pH8.0, 10 mM EDTA, 2% SDS (w/v)

2.4.16 CTAB Buffer

2% CTAB (w/v), 1% PVP40 (w/v), 5 M NaCl, 0.5 M EDTA, 1 M Tris HCL

2.4.17 Lysis Buffer

40 mM tris-acetate, 20 mM sodium acetate, 1 mM EDTA, 1% SDS (w/v)

2.4.18 RNAse A (DNAse free)

10 mg/mL RNAse was dissolved in 0.01 M sodium acetate (pH 5.2) and placed in a boiling water bath for 15 minutes. This was cooled to room temperature and 0.1 volumes of tris-HCL (pH 7.4) was added and stored at – 20 °C.

2.5 DNA ISOLATION

2.5.1 A Simple Method for Extraction of Fungal Genomic DNA (Modified from Al-Samarrai and Schmid, 1999)

0.3 g of freeze dried mycelia were ground to a fine powder with liquid nitrogen in a sterile mortar and pestle. 5 mL of lysis buffer (section 2.4.17) was added to the ground mycelium in a 15 mL corex test tube. This was mixed vigorously by pipetting. After mixing 20 µL of RNAse (section 2.4.18) was added and the tube incubated at 37°C for 20 minutes. After the 20 minute incubation 1.65 mL of NaCl (5 M) was added and mixed. Centrifugation was then carried out at 20200 g (13000 rpm, SS34) for 20 minutes at 4°C. After centrifugation the supernatant was transferred to a fresh tube and 4 mL of chloroform and phenol (1:1) added. Centrifugation was then repeated as above. The aqueous phase (top layer) was removed to a clean corex tube and an equal volume of chloroform then added after which centrifugation was repeated. The aqueous phase was removed to a clean tube then DNA precipitated by adding 2 volumes of 95 % ethanol. After addition of ethanol centrifugation was carried out at 20200 g (13000 rpm, SS34) for 5 minutes. The supernatant was removed and 2 mL of lysis buffer (section 2.4.17) was added and the pellet resuspended. The resuspended DNA was then divided between 4 microcentrifuge tubes and 165 µL NaCl (5 M) added and mixed. After addition of NaCl an equal volume of chloroform was added and tubes inverted. Centrifugation was then carried out at 15110 g (13000 rpm, Biofuge 13), 4°C for 10 minutes. The aqueous layer was then transferred into a clean microcentrifuge tube and 2 volumes of 95% ethanol added and mixed thoroughly. Centrifugation was carried out for 2 minutes at 15110 g (13000 rpm, Biofuge 13) and the supernatant discarded. The remaining DNA pellet was washed twice in 70 % ethanol. After washing the ethanol was removed and the pellet air dried and resuspended in 50 µL of TE buffer (section 2.4.1).

2.5.2 CTAB DNA Extraction Method (Mclenachan Personal Communication)

0.5 g- 1.0 g of freeze dried mycelia were ground to a fine powder with liquid nitrogen in a sterile mortar and pestle. This was distributed to microcentrifuge tubes with approximately 10 mg per tube. 600 μL of CTAB buffer (section 2.4.16) was added and 2 μL of RNAse (20 mg/mL). The microcentrifuge tubes were then mixed thoroughly and placed at 37°C for 10 minutes, followed by incubation at 65°C for 45 minutes (inverting occasionally). After incubation at 65°C the tubes were cooled slightly then 600 μL of chloroform added. After addition of chloroform the microcentrifuge tubes were mixed gently then left to stand for two minutes, after standing they were centrifuged at 15110 g (13000 rpm, Biofuge 13) for 1 minute. After centrifugation the upper (aqueous) phase was transferred to a clean microcentrifuge tube, and then 600 μL of isopropanol was added and the tubes inverted once to mix. The microcentrifuge tubes were then incubated on ice for approximately 30 minutes to precipitate the DNA. Approximately 300 μL of liquid was then drawn off, then 600 μL of 70% ethanol added, followed by brief centrifugation (approximately 30 seconds at 15110 g (13000 rpm, Biofuge 13)). The pellet was then washed twice more in 70 % ethanol; after the last wash, ethanol was removed then microcentrifuge tubes were left to air dry before resuspending the DNA in TE buffer (section 2.4.1).

2.5.3 Isolation of High Molecular Weight DNA From Filamentous Fungi (Modified from Moller, E. M *et al,* 1992)

0.5 g- 1.0 g of freeze dried mycelia were ground to a fine powder with liquid nitrogen in a sterile mortar and pestle. 10 mL of TES (section 2.4.15) was added, along with 100 μg Proteinase K (Boehringer). The mixture was then incubated at 60°C for 1 hour with occasional shaking. After incubation the salt concentration was adjusted by adding 2.8 mL 5 M NaCl and 1.3 mL 10% CTAB (section 2.4.16). The mixture was then placed at 65° C for 10 minutes. 14 mL of SEVAG (section 2.4.14) was then added and the mixture incubated on ice for 30 min; after incubation centrifugation was carried out for 10 minutes at 20200 g (13000 rpm, SS34), 4° C. The supernatant was then transferred to a clean tube and 4.5 mL of 5 M sodium acetate added. After gentle mixing, the mixture was placed on

ice for 40 min before centrifugation for 10 minutes at 20200 g (13000 rpm, SS34), 4° C. The supernatant was transferred once again to a clean tube and 10.2 mL of isopropanol added to precipitate the DNA, after addition of isopropanol, the mixture was incubated on ice for 30 minutes. After incubation, centrifugation was carried out for 5 minutes at 20200 g (13000 rpm, SS34), 4° C. The supernatant was then removed and the remaining pellet washed twice with cold 70% ethanol (approximately 1 mL), centrifuging after each wash for 5 minutes at 20200 g (13000 rpm, SS34), 4°C. Excess ethanol was removed and the pellet air dried before resuspending the pellet in 700 µL T.E buffer (section 2.4.1).

2.5.4 λ DNA Preparation

5 mL of lysate from a single plaque (section 2.6.7) was centrifuged at 3020 g (5000 rpm, SS34) for 10 minutes to remove cell debris. To the supernatant, RNase (5 µL of 1 mg/mL; section 2.4.18) and DNase (0.5 µL of 10 mg/mL) were added and incubated for 37°C for 30 minutes. After incubation 5 mL of 30 % PEG (30 mM NaCl; section 2.4.8) solution was added and the mixture left to stand on ice for 1 hour. The mixture was then centrifuged at 7710 g (8000 rpm, SS34) for 15 minutes at 4°C. The phage pellet was then resuspended in 0.5 mL SM buffer (section 2.4.7). These phage were then transferred to a clean microcentrifuge tube with 5 µL of 10 % SDS and 10 µL of EDTA being added. This phage mixture was then incubated at 68°C for 15 minutes after which an equal volume of phenol was added and the mixture vortexed, then left standing for 5 minutes before vortexing again. This mixture was then centrifuged for 5 minutes at 15110 g (13000 rpm, Biofuge 13). The aqueous phase was taken and an equal volume of chloroform added, and the mixture centrifuged for 3 minutes at 15110 g (13000 rpm, Biofuge 13). The aqueous phase was again removed and an equal volume of isopropanol added. After incubation at -20°C for 20 minutes the mixture was centrifuged at 15110 g (13000 rpm, SS34) for 10 minutes. The supernatant was discarded and the pellet washed in 1 mL 70% ethanol then air dried at 37°C. After air drying the pellet was resuspended in 50 µL TE (10:0.1, section 2.4.1).

2.6 PURIFICATION OF DNA

2.6.1 Agarose Gel Purification of DNA Fragments

DNA containing the fragment to be purified was run on a 1-1.5% Sea Plaque (FMC) low melting point agarose gel in TAE buffer (section 2.4.2). After staining in ethidium bromide (section 2.4.4), the DNA was visualised under long wave UV light, and the appropriate band excised using a clean scalpel blade. The DNA was extracted from the agarose using a QIAquick Gel extraction kit (Qiagen) according to the manufacturer's instructions.

2.6.2 Purification of PCR Products From a PCR Reaction

DNA was purified from a PCR reaction either by gel purification (section 2.6.1), or using a QIAquick PCR purification Kit (Qiagen) according to the manufacturer's instructions.

2.6.3 Purification of Plasmid DNA from *E. coli*

Plasmid DNA was isolated from *E. coli* using a QIAGEN plasmid miniprep kit (≤20 μg plasmid DNA) or QIAGEN plasmid midikit (≤100 μg plasmid DNA) according to manufacturer's instructions. Freshly grown single colonies of transformed *E. coli* were used to inoculate LB broth (section 2.2.1) containing the appropriate selective antibiotic. The inoculated broth was then incubated (shaking) at 37°C overnight or according to instructions accompanying the kit before plasmid isolation was able to proceed.

2.6.4 Phenol/Chloroform Extraction

DNA samples were extracted with equal volumes of Tris-equilibrated phenol (pH 7.6) and chloroform, mixed thoroughly and centrifuged 15110 g (13000 rpm, Biofuge 13). The aqueous phase was re-extracted until a clear interface between the two phases was obtained (usually requires between 2-3 extractions). Samples were then extracted once with two volumes of chloroform. DNA was then isolated by ethanol precipitation (section 2.6.5).

2.6.5 Ethanol Precipitation

To concentrate DNA solutions, one tenth volume of 3 M sodium acetate (pH 5.2) and either 2.5 volumes 95% ethanol, or 0.6 volumes isopropanol, were added, mixed and left on ice for 15 minutes. The DNA was then pelleted by centrifugation at 15110 g (13000 rpm, Biofuge 13) for 10 minutes. The pellet was washed once with 70% ethanol, and then air dried. The DNA was resuspended in Milli-Q water or TE (10:0.1, section 2.4.1).

2.6.6 PEG Purification

PEG purification was carried out to improve purity of PCR products of 5'and 3' *pks* flanks and the *hph* cassette with their appropriate *att*B sites. 150 µL of TE, pH 8.0 (section 2.4.1) was added to 50 µL of PCR product followed by 100 µL of 30% PEG (section 2.4.8). This mixture was vortexed thoroughly and centrifuged immediately at 15,110 g (13000 rpm, Biofuge) for 15 minutes at room temperature. The supernatant was removed and the pellet dissolved in 50 µL of TE, pH 8.0 (section 2.4.1). The recovered PCR product was checked on an agarose gel to determine quality and quantified using fluorometric assay (section 2.7.1).

2.6.7 Purification of Lambda Stocks

Before lambda DNA could be purified it had to be repackaged. 25 µL of packagene extract (Promega packaging system) was added to 5 µL of each lambda clone DNA. To isolate a single pure plaque, 10-fold serial dilutions were made of lambda stocks in SM buffer (section 2.4.7). 100 µL of lambda suspensions were combined with 100 µL plating bacteria (section 2.3.3) and incubated for 20 minutes at 37° C to allow phage adsorption. 3 mL of molten 0.7% agar (47°C) was mixed gently with the lambda suspensions and this mixture was poured over a lower layer in a plate. Plates were then incubated overnight at 37°C. After incubation a single plaque was selected using a cut off pipette tip and placed in a microcentrifuge tube containing 1 mL SM buffer (section 2.4.7) plus 1 drop chloroform. This was left to stand at room temperature for 1 hour and stored indefinitely at 4°C.

Preparing a stock from a single plaque:

100 µL of diluted phage were added to 100 µL of plating bacteria (section 2.3.3) and left to absorb for 30 minutes at 37°C. This was then mixed with 3 mL of molten 0.7% agar and poured over plates containing bottom agar. The plates were left to incubate at 37°C for approximately 6-8 hours, or until confluent lysis was achieved. Once confluent lysis was achieved 5 mL of SM buffer (section 2.4.7) was added to each plate and the plates left overnight at 20°C. The SM buffer (section 2.4.7) was then removed from the plates and added to microcentrifuge tubes containing approximately 20 µL chloroform. These lysates were then stored at 4°C until required for DNA extraction.

2.7 DETERMINATION OF DNA CONCENTRATIONS

DNA was quantified by one of the following methods.

2.7.1 Determination of DNA Concentration By Fluorometric Assay

Fluorometric quantification was used for pure or impure DNA samples of concentrations thought to be higher than 50 ng/µL. DNA was quantified on a Hoefer DyNA Quant 200 Fluorometer according to the manufacturer's protocol. The scale of the fluorometer was set by firstly zeroing the fluorometer with 2 mL of a dye solution containing 1xTNE buffer (10 mM Tris-HCL, 1 mM Na_2EDTA and 100 mM NaCl, pH 7.4) and 0.1 µg/mL Hoescht dye and secondly by calibrating the fluorometer to 120 (Due to the high GC% of fungal Genomic DNA) using 2 µL of 100 µg/mL calf thymus DNA added to 2 mL of the dye solution previously described. Once the scale was set, 2 µL of sample DNA was added to 2 mL of the dye solution and resulting value was recorded as the concentration of DNA in ng/µL.

2.7.2 Determination of DNA Concentration By Gel Electrophoresis

Determination of DNA concentration by gel electrophoresis was used for samples thought to be lower than 100 ng/µL. A series of uncut lambda DNA concentration standards

(Invitrogen) were run on an agarose gel alongside the DNA sample. The DNA sample concentration was estimated by comparing the intensity of ethidium bromide fluorescence to that of the DNA concentration standards. This method was performed for most samples as it confirmed the presence of DNA and determined if more accurate quantification was required.

2.8 RESTRICTION ENDONUCLEASE DIGESTION OF DNA

Restriction digests were performed at 37°C, in the buffer specified by the manufacturer either overnight (with BSA added at 100 μg/ mL) or for 3 hours. Digestion was performed with 1-10 units of restriction enzyme per μg DNA. Digests were checked by running an aliquot on an agarose gel (section 2.9) after heating at 65°C for 5 minutes. If digestion was not complete, further restriction enzyme was added to the digestion mixture and incubation extended for 1 hour and the samples again checked for completion of digestion.

2.9 AGAROSE GEL ELECTROPHORESIS OF DNA

DNA fragments were size fractionated by electrophoresis through 0.8%- 1.5% agarose dissolved in 1 x TAE (section 2.4.2) or TBE (section 2.4.3) buffer at 60-100 volts. Dyes present in the loading buffer (section 2.4.5), which was added to DNA samples before loading, allowed estimation of DNA migration. After electrophoresis, agarose gels were stained in ethidium bromide for 15-30 minutes, washed in water, observed under short wave UV light and photographed.

2.10 DETERMINATION OF DNA MOLECULAR WEIGHTS

The size of DNA fragments was estimated by running the DNA sample on an agarose gel alongside known size ladders such as; *Hin*d III/ *Eco* RI double digest of lambda DNA, 1 kb ladder or 1 kb+ ladder (Invitrogen). The mobility of the standard markers from the wells were measured and used to determine the molecular weight of the unknown fragments by comparing the relative mobility between the lanes.

2.11 DNA SUB-CLONING TECHNIQUES

DNA fragments to be sub-cloned were purified (section 2.6.1) usually by gel purification, then ligated into the appropriate vectors according to section 2.11.1 (or manufacturers instructions) and transformed into a suitable *E. coli* host strain (section 2.2.12).
Transformants were selected on the appropriate antibiotic and where appropriate, blue/white colour selection was used for selection. If too many transformants were present to clearly distinguish between blue and white colonies, single colonies were selected and streaked in approximately 1 cm lines on ampicillin/Xgal/IPTG plates. Putative recombinants were then purified by streaking for single cells. The presence of insert was always confirmed by restriction enzyme digestion (section 2.8) or PCR (section 2.13) with isolated plasmid DNA.

2.11.1 Ligation Reactions

2.11.1.1 Standard Ligation

pGEM-T and pGEM-T Easy vectors:
Ligation reactions were set up in 10 µL with the required vector: insert ratio and final concentration of 1x ligation buffer (Invitrogen) and 3 U/µL of T4 ligase (Invitrogen). Ligation reactions were incubated for one hour at room temperature then incubated overnight at 4°C. Ligation reactions were checked for completion by removing a 1 µL aliquot and comparing this on an agarose gel with an aliquot taken before addition of ligase. Ligation mixtures and controls were used to transform *E. coli* (Section 2.12.1.3) and left to incubate at 37°C overnight. These *E. coli* were plated on LB/Amp/Xgal/IPTG media and transformants were identified. Selective transformant plates were stored at 4°C until needed.

pUC118:
Ligation reactions were set up in 20 µL with the required vector: insert ratios (either 1:3, 1:1.5 or 1:1) and final concentration of 1x ligation buffer, 1U of 40 U/µL T4 ligase and

approximately 100 ng/ µL of vector. All reagents were mixed well and briefly centrifuged, then incubated overnight at 4°C. Ligation reactions were checked for completion by removing a 1 µL aliquot and comparing this on an agarose gel with an aliquot taken before addition of ligase. Ligation mixtures and controls were used to transform *E. coli* (Section 2.12.1.3) and left to incubate at 37°C overnight. These *E. coli* were plated on LB/Amp/Xgal/IPTG media and transformants were identified. Selective transformant plates were stored at 4°C until required.

2.11.1.2 Gateway Recombination Reactions

BP recombination:

The MultiSite Gateway™ Three-Fragment Vector Construction Kit manual (Invitrogen) was used. Recombination reactions were set up according to the manufacturer's instructions in 16 µL volumes with the appropriate vector: insert ratio. In general approximately 300 ng of DONR vector was added and 40-100 fmol (approximately 50-160 ng) of PCR product along with 4 µL of 5x BP clonase reaction buffer and 4 µL of BP clonase enzyme mix. All mixtures were vortexed, briefly centrifuged then incubated at 25°C for between 9-18 hours. 2 µL of Proteinase K solution was added and mixtures incubated at 37° C for 10 minutes to stop the recombination reactions. Recombination reactions were stored at -20°C for a maximum of 1 week before transforming into TOP10 $CaCl_2$ competent cells (section 2.12.1.4).

LR recombination:

Reactions were set up following the MultiSite Gateway Three-Fragment Vector Construction Kit manual (Invitrogen). Recombination reactions were set up in 16 µL with the appropriate vector: insert ratio. In general 60 ng of DESTINATION vector was added and 20-25 fmol (approximately 90-120 ng) of each entry clone along with 4 µL of 5x LR Clonase reaction buffer and 4 µL of LR clonase enzyme mix. All mixtures were vortexed, centrifuged, then incubated at 25°C for 16 hours or overnight. 2 µL of Proteinase K (Invitrogen) solution was added and mixtures incubated at 37° C for 10 minutes to stop the

recombination reactions. Recombination reactions were stored at -20°C for a maximum of 1 week before transforming into TOP10 $CaCl_2$ competent cells (section 2.12.1.4).

2.11.2 Shot Gun Cloning

λ DNA was digested with *Bam*HI, *Eco*RI and *Xba*I (section 2.8) to produce 4 different sized insert fragments, and then a phenol/chloroform extraction (section 2.6.4) was carried out on the digested λ DNA. Ligation reactions were set up with different vector: insert ratios of pUC118 (section 2.11.1.1). Transformants were picked through blue/white selection then plasmid extraction (section 2.6.3). Digestion was used to confirm the size of inserts (section 2.8).

2.12 TRANSFORMATION PROTOCOLS

2.12.1 Transformation of *E. coli* by Electroporation

2.12.1.1 Preparation of Electroporation Competent *E. coli* Cells

One litre of LB broth (Section 2.2.1) was inoculated with 100 µL of an overnight culture of *E. coli.* XL-1 and grown at 37°C, with vigorous shaking to mid-log phase (OD_{600} 0.5-1.0). The cells were chilled on ice for 20 minutes then harvested by centrifugation for 10 minutes at 4080 g (5000 rpm, Sorvall GSA) (all centrifugation was performed at 4°C). The cells were washed (by resuspension, centrifugation at 4080 g (5000 rpm, Sorvall GSA) to pellet the cells and removal of supernatant) in 1.0 litre and 0.5 litre of ice cold sterile water, 20 mL of ice cold 10% (v/v) glycerol, then finally resuspended in 4 mL ice cold 10% (v/v) glycerol. Cells were stored at -80°C in 40 µL and 200 µL aliquots.

2.12.1.2 Preparation of $CaCl_2$ Competent Cells

50 mL LB broth (Section 2.2.1) was inoculated with 0.5 mL of a 5 mL overnight culture of the desired *E. coli* strain (i.e. TOP10), and grown at 37°C, with vigorous shaking to OD_{600}

0.45-0.60. The cells were chilled on ice for 10 minutes then harvested by centrifugation at 5860 g (6000 rpm, GSA) for 5 min (all centrifugation carried out at 4°C). The harvested cells were resuspended in 0.5 volume cold 0.1 M $CaCl_2$ and left to incubate on ice for 10 minutes. Cells were harvested again then resuspended in 1/20 volumes 0.1 M $CaCl_2$ and incubated on ice for approximately 1 hour before using for a transformation.

2.12.1.3 Electroporation of Competent *E. coli* Cells

2 µL of DNA (ligation mixture, controls or plasmid) was added to 40 µL of electroporation competent *E. coli* cells (section 2.12.1.1) in a 0.2 cm ice cold electroporation cuvette, gently mixed and tapped to the bottom. The cells were electroporated in a Biorad gene pulser set at 25 µF, 2.5 kV and 200 Ω. 200 µL of LB broth (section 2.2.1) was immediately added to the cells which were transferred into a microcentrifuge tube and incubated with shaking at 37°C for 30 minutes. Positive (circular plasmid DNA) and negative (cells only) controls were always included. Cells were spread onto LB agar containing appropriate supplements (section 2.2.3) and incubated overnight at 37°C.

2.12.1.4 Transformation of $CaCl_2$ Competent Cells

2 µL of either BP or LR recombination was added to 50 µL of $CaCl_2$ competent cells (TOP10) (section 2.12.1.2) in a microcentrifuge tube. This reaction was incubated on ice for 30 minutes, and then heat shocked at 42°C for 30 seconds without shaking. After heat shock cells were transferred to ice for approximately 2 minutes before addition of 250 µL of SOC medium (section 2.2.7). Cells were then incubated with shaking at 37°C for 1 hour. Positive (circular plasmid DNA) and negative (cells only) controls were also included. Cells were spread onto LB agar containing appropriate supplements (section 2.2.3) and incubated overnight at 37°C.

2.12.2 Transformation of *D. pini* protoplasts.

2.12.2.1 Preparation of Competent *D. pini* Protoplasts

100 µL of freshly grown *D. pini* mycelium ground in sterile milli-Q water was spread onto cellophane covered DM media (section 2.2.4). The cellophane plate was incubated at 22°C. After 5-6 days, a light covering of mycelium had grown on the cellophane. The cellophanes were removed carefully from DM media plates and were placed mycelia down in a sterile petri dish (approximately 3-4 cellophanes per Petri dish) containing 15 mL of sterile 20 mg/mL Glucanex (beta-glucanase, Chemcolour Industry, NZ) dissolved in OM buffer (section 2.4.10). The petri dish was sealed and gently shaken at 37°C overnight. Protoplast formation was confirmed by microscopic examination. 5 mL of the incubated protoplast solution was put in a 30 mL corex tube and overlayed with 2 mL ST buffer (section 2.4.9) and centrifuged at 4°C for 20 minutes at 4340 g (6000 rpm SS34). The interphase was then removed and was washed twice in 10 mL STC buffer (section 2.4.6) being pelleted in between washes by centrifuging for 15 minutes at 4340 g (6000 rpm SS34). (As the protoplasts did not successfully float, instead of overlaying with ST buffer, 20 mL of STC buffer was used to pellet the protoplasts. This pellet was then washed twice in 10 mL of STC buffer (section 2.4.6). The pellet resulting after washing was resuspended in 400 µL STC buffer (section 2.4.6) and stored at 4°C optimally for no longer than 4 hours before use however they are able to be stored overnight. The concentration of the protoplasts was determined by counting with a haemocytometer.

2.12.2.2 Transformation of *D. pini* protoplasts

For each transformation, between 8 x 10^7 and 1 x 10^8 protoplasts in 150 µL of STC buffer (section 2.4.6) were mixed with 5 µg of DNA and incubated at 22°C for 20 minutes. In three steps 250, 250 and 850 µL of 40% polyethyleneglycol (PEG) 6000 solution in STC buffer (section 2.4.8) were added carefully and thoroughly mixed with DNA/ protoplast mixture and the final suspension was incubated for a further 20 minutes at 22°C. The suspension was then diluted in 5 mL of STC buffer and the protoplasts were collected by

centrifugation at 4°C for 10 minutes at 1085g (3000 rpm, SS34) and resuspended in 500 µL of STC (section 2.4.6). 100 µL of this suspension was spread over a plate containing 20 mL DM Suc (section 2.2.10). Selective antibiotics (in most cases 70µg/mL hygromycin) were added 24 hrs later by overlaying the plates with 5 mL DM top agar (section 2.2.9) to which the appropriate concentration of antibiotic had been added.

2.13 AMPLIFICATION OF DNA BY THE POLYMERASE CHAIN REACTION (PCR)

Due to the nature of this research, a number of different PCR reactions were used. PCR reactions were set up on ice using a master mix which contained all common reagents for the PCR reactions. Other reagents were pipetted separately. PCR reactions were in a total volume of 50 µL in 0.2 mL PCR tubes. Negative controls of no DNA, only one primer and a positive control were included in each experiment. The reactions were stored at -20°C. All reactions used *Taq* polymerase and *Taq* buffer from Invitrogen.
Primers used for PCR are included in Table 2.2.

2.13.1 Reagents and Cycling Conditions for Basic PCR

The final concentrations of each component in 1 reaction were 1 x *Taq* buffer, 1.5 mM $MgCl_2$, 0. 2 mM dNTPs, 1 unit *Taq* DNA polymerase, 0.2 µM of each primer and 1-3 ng of plasmid template or 20 ng of genomic template. After mixing, the PCR reaction tubes were placed in a thermal cycler (Techne) and subjected to an initial denaturation step of 1 min at 94°C, then 30 cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec (adjusted depending on primers used and specificity wanted), extension at 72°C for 90 sec (depending on expected product size).
For specific reagents and cycling conditions refer to Table 2.3 – 2.7

2.14 DNA SEQUENCING

Automatic sequencing reactions were preformed using the ABI PRISM ™BigDye™ Terminator Cycle Sequence Ready Reaction Kit version 3.1 (Perkin Elmer) with 200 ng plasmid DNA or 2 ng/100 bp of PCR product, 3.2 pmol primer and 8 μL of terminator ready reaction mix in a total volume of 20 μL. Samples were run on an ABI3730 Genetic Analyzer (Applied Biosystems Inc.).

Primers used for sequencing are included in Table 2.2.

Table 2.2 PCR and Sequencing Primers

Primer	Size (nt)	Tm' (°C)	Sequence (5' to 3')	Source/ Reference
pUC/M13 Forward	22	70	GCC AGG GTT TTC CCA GTC ACG A	Perkin Elmer
Puc/M13 Reverse	24	70	GAG CGG ATA ACA ATT TCA CAC AGG	Perkin Elmer
M13 Reverse- gateway	17	50	CAG GAA ACA GCT ATG AC	This Study
SP6	22	58	TTT AGG TGA CAC TAT AGA ATA C	Promega
T7	23	66	TAA TAC GAC TCA CTA TAG GGC GA	Promega
ITS5	22	62	GGA AGT AAA AGT CGT AAC AAG G	White *et al, 1990*
ITS4	20	58	TCC TCC GCT TAT TGA TAT GC	White *et al*, 1990
pksCB2	18	54	TGA AGA AGT ATG TCG CCG	This Study
OTpks KO/2	25	76	CTG TCT AGA GTG TTG GTC TCC ATC C	This Study
OTpks KO/3	26	76	CAC TCT AGA CAG AAG ACA CTC ATT GC	This Study
OTpks KO/4	19	58	GGC TCC GTC TTC TCT TCA A	This Study
OT R126A Fwd1	20	64	GTC GCC GAC AAT GCC TCC AA	This Study
OT R162A Rev1	22	62	TTC TCG ACA ACA ATC GTC CAT A	This Study
OT R162A Fwd2	23	68	CAA GAT GAC TCT CGG AGT TTC AG	This Study
OT R162A Rev2	21	64	GTG TGG TAA CGA GCG TTG CAA	This Study
OT R162A Fwd3	22	68	CGC CTG CAA TGC CAA TCC AAT C	This Study
OT R162A Rev3	23	68	CAC CGT CTC AAT TTC CGA ATC CA	This Study
OT R162A Fwd4	21	66	TCG AAA GTG CGC CAG GAG AAG	This Study
OT R162A Rev4	21	64	GCC TTG GAT TGG ATG GGT ACA	This Study
OT R162A Fwd5	22	66	TAC AAT CGC GCC GTG TAT ACT C	This Study
OT R162A Rev5	24	70	GCA GAC ATG AGA GTA AGT CCA GTT	This Study
OT R162A Fwd6	22	68	AGA ACC CTG ACT GGC CGA GAT T	This Study
OT R162A Rev6	22	66	TTC GCC CCA TCG AAC AAG AAT C	This Study
OT R162A Fwd7	24	70	GTG CTT CAT CCA TAC AGA CTG TTC	This Study
OT R162A Rev7	24	72	CGT GGT ACC TGA CAT GAG AGC TTA	This Study
R162A Rev Out	22	68	ATC TTC GTC CGT ACG GGA GTC T	This Study
R162A Fwd Out	22	64	TGG GCA TAT TGA TGG CTT GAT C	This Study
R181Revch2	22	62	CAC ATT GTT GCA GAG TCA ATC A	This Study
R181 Ovlp	22	64	CAG ACA TCC GTT GAA GAA CAC A	This Study
R181 Fwd2	21	68	GGA GTC CGG CGA TGC ACA ATG	This Study
gateway pks CB2-attB4	47	136	GGG GAC AAC TTT GTA TAG AAA AGT TGT GAA GAA GTA TGT CGC CGG CA	This Study
gateway OT pks ko/2-attB1	46	136	GGG GAC TGC TTT TTT GTA CAA ACT TGG TGT TGG TCT CCA TCC AGT C	This Study
gateway OT pks ko/3-attB2	49	146	GGG GAC AGC TTT CTT GTA CAA AGT GGC AGA AGA CAC TCA TTG CGA AGA G	This Study
gateway OT pks ko/4-attB3	47	134	GGG GAC AAC TTT GTA TAA TAA AGT TGG GCT CCG TCT TCT CTT CAA TG	This Study
gateway hph-attB1	49	144	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CTA GAG TCG ACG GTA TCG A	This Study
gateway hph-attB2	51	150	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TCT AGA AAG AAG GAT TAC CTC	This Study
OT phnA 3' Out	21	64	ATG GGT CAA CGG CAA GGT CAT	This Study
OT phnA 5' Out	19	62	CCA TTG CGG CAC CGT ACC A	This Study
Inverse AFLR141	22	66	GGT TGT TAC CTG ACC CCT CAA T	This Study
Inverse AFLR465	22	62	CGG GGA ATA GCA GAG GCT TGA T	This Study

' Melting temperature calculated as Tm (°C) = 2 (A+T) + 4(G+C) from Itakura *et al. (1984)*

² Underlined bases indicate an enzyme recognition site

Primer	Size (nt)	Tm' (°C)	Sequence (5' to 3')	Source/ Reference
Epoxide 3' end	21	66	TCT CCC AAG GCG CGA AAT GAC	This Study
Mox-dot 5' end	22	66	AAT CGG TGC GGG AAT TTC AGG A	This Study
Mox-dot 3' end	23	68	ACT ACC AGA GTT TCC GGC TAT CA	This Study
3' hph Out	23	68	TCC TTG AAC TCT CAA GCC TAC AG	This Study
5' hph Out	22	68	GAA TCT CCG GTG GTG TGG AAG A	This Study
hph Fwd 1	22	68	TCT TCC ACA CAC CAC CGG AGA TTC	This Study
hph Fwd 2	23	68	ATT TCA TAT GCG CGA TTG CTG AT	This Study
hph Rev 1	23	68	CTG TAG GCT TGA GAG TTC AAG GA	This Study

' Melting temperature calculated as Tm (°C) = 2 (A+T) + 4(G+C) from Itakura *et al. (1984)*

² Underlined bases indicate an enzyme recognition site

Table 2.3 Reagents and Cycling Conditions for Amplification of *Pks*^{dot} Flanking Regions and Addition of *att* B Sites

Reagents/ Cycle conditions	Amplification of PKS flanking regions		Amplification/ addition of *att* B sites
	5' *pks* flank	3' *pks* flank	5' *pks* flank plus *att* B4 and *att* B1 sites
***Taq* DNA polymerase**	2 units	2 units	2 units *platinum taq*
***Taq* buffer**	1 X	1 X	1 X
$MgCl_2$	4 mM	4 mM	1.5 mM
dNTPs	0.2 mM	0.2 mM	0.2 mM
Template	25 ng genomic *D. pini*	25 ng genomic *D. pini*	3 ng of *5' pks* flank ligated in pGEM-T Easy vector
Primers	0.2 µM pks CB2 0.2 µM pks KO/2	0.2 µM pks KO/3 0.2 µM pks KO/4	0.2 µM gateway pks CB2-attB4 0.2 µM gateway OT pks KO/2-attB1
Cycle			
intial denaturation	3 min at 94°C	3 min at 94°C	2 min at 94°C
	30 cycles of:	30 cycles of:	30 cycles of:
denaturation	45 sec at 94°C	25 sec at 94°C	30 sec at 94°C
annealing	45 sec at 48°C	30 sec starting at 58°C and decreasing by -1°C increments to 48°C	30 sec at 58°C
extension	90 sec at 72°C	90 sec at 72°C	60 sec at 75°C

Final concentrations of all reagents are shown except for T*aq* DNA polymerase and template, which are given as amounts per 50 µL reaction

Table 2.4 Reagents and Cycling Conditions for Gateway Disruption Construct R226

Reagents/ Cycle conditions	**Checking integrity of disruption vector R226**			
	Checking 5' *pks* flank	**Checking 3' *pks* flank**	**Checking *hph* cassette**	**Checking orientation of inserts**
***Taq* DNA polymerase**	2 units	2 units	2 units	2 units
***Taq* buffer**	1 X	1 X	1 X	1 X
$MgCl_2$	0.3 mM	3.0 mM	3.0 mM	3.0 mM
dNTPs	0.2 mM	0.2 mM	0.2 mM	0.2 mM
Template	20 ng R226	20 ng R226	20 ng R226	20 ng R226
Primers	0.2 µM of each gateway pks CB2-attB4 OT pks KO/2-attB1	0.2 µM of each gateway OT pks KO/3-attB2 gateway OT pks KO/4-attB3	0.2 µM of each gateway hph-attB1 gateway hph-attB2	0.2 µM of each in pairs indicated 3' hph out and pks CB2 3' hph out and OT pks KO/4 5' hph out and pks CB2 5' hph out and OT pks KO/4
Cycle				
intial denaturation	2 min at 94°C	2 min at 94°C	2 min at 94°C	2 min at 94°C
	30 cycles of:	30 cycles of:	30 cycles of:	30 cycles of:
denaturation	30 sec at 94°C	30 sec at 94°C	30 sec at 94°C	30 sec at 94°C
annealing	30 sec at 58°C	30 sec at 58°C	30 sec at 58°C	30 sec at 58°C
extension	2.5 mins at 75°C	2.5 mins at 75°C	2.5 mins at 75°C	2.5 mins at 75°C

Final concentrations of all reagents are shown except for T*aq* DNA polymerase and template, which are given as amounts per 50 µL reaction

Table 2.5 Reagents and Cycling Conditions for Re-amplification of RT-PCR Products

Reagents/ Cycle conditions	RT-PCR products and Re-amplification of RT-PCR products		
	epox^{dot}	*mox*^{dot}	*ord*^{dot}
***Taq* DNA polymerase**	2 units	2 units	2 units
***Taq* buffer**	1 X	1 X	1 X
$MgCl_2$	3.0 mM	3.0 mM	3.0 mM
dNTPs	0.2 mM	0.2 mM	0.2 mM
Template	1 ng of RT-PCR product	1 ng of RT-PCR product	1 ng of RT-PCR product
Primers	0.2 µM of each epoxide 3' end OT R162A Rev4	0.2 µM of each Mox-dot 5' end Mox-dot 3' end	0.2 µM of each OT R162A Fwd1 OT R162A Rev6
Cycle			
intial denaturation	2 min at 94°C	2 min at 94°C	2 min at 94°C
	30 cycles of:	30 cycles of:	30 cycles of:
denaturation	45 sec at 94°C	45 sec at 94°C	45 sec at 94°C
annealing	45 sec at 62°C	45 sec at 62°C	45 sec at 62°C
extension	2.5 mins at 72°C	2.5 mins at 72°C	2.5 mins at 72°C

Final concentrations of all reagents are shown except for T*aq* DNA polymerase and template, which are given as amounts per 50 µL reaction

Table 2.5 Reagents and Cycling Conditions for Re-amplification of RT-PCR Products

Reagents/ Cycle conditions	**RT-PCR products and Re-amplification of RT-PCR products**		
	epox ^{dot}	***mox*** ^{dot}	***ord*** ^{dot}
***Taq* DNA polymerase**	2 units	2 units	2 units
***Taq* buffer**	1 X	1 X	1 X
$MgCl_2$	3.0 mM	3.0 mM	3.0 mM
dNTPs	0.2 mM	0.2 mM	0.2 mM
Template	1 ng of RT-PCR product	1 ng of RT-PCR product	1 ng of RT-PCR product
Primers	0.2 µM of each epoxide 3' end OT R162A Rev4	0.2 µM of each Mox-dot 5' end Mox-dot 3' end	0.2 µM of each OT R162A Fwd1 OT R162A Rev6
Cycle			
intial denaturation	2 min at 94°C	2 min at 94°C	2 min at 94°C
	30 cycles of:	30 cycles of:	30 cycles of:
denaturation	45 sec at 94°C	45 sec at 94°C	45 sec at 94°C
annealing	45 sec at 62°C	45 sec at 62°C	45 sec at 62°C
extension	2.5 mins at 72°C	2.5 mins at 72°C	2.5 mins at 72°C

Final concentrations of all reagents are shown except for T*aq* DNA polymerase and template, which are given as amounts per 50 µL reaction

Table 2.6 Reagents and Cycling Conditions for Southern Blotting Probes

Reagents/ Cycle conditions	PCR of Probes for Southern Blotting			
	5' *pks*	***mox*^{dot}**	***ord*^{dot}**	***epox*^{dot}**
***Taq* DNA polymerase**	2 units	2 units	2 units	2 units
***Taq* buffer**	1 X	1 X	1 X	1 X
$MgCl_2$	3.0 mM	3.0 mM	3.0 mM	3.0 mM
dNTPs	0.2 mM	0.2 mM	0.2 mM	0.2 mM
Template	1 ng of pR216	1 ng of RT-PCR product	1 ng of RT-PCR product	1 ng of RT-PCR product
Primers	0.2 µM of each T7 Sp6	0.2 µM of each Mox-dot 5' end Mox-dot 3' end	0.2 µM of each OT R162A Fwd1 OT R162A Rev6	0.2 µM of each epoxide 3' end OT R162A Rev4
Cycle				
intial denaturation	2 min at 94°C	2 min at 94°C	2 min at 94°C	2 min at 94°C
	30 cycles of:	30 cycles of:	30 cycles of:	30 cycles of:
denaturation	30 sec at 94°C	30 sec at 94°C	30 sec at 94°C	30 sec at 94°C
annealing	30 sec at 58°C	30 sec at 58°C	30 sec at 58°C	30 sec at 58°C
extension	2.5 mins at 75°C	2.5 mins at 75°C	2.5 mins at 75°C	2.5 mins at 75°C

Final concentrations of all reagents are shown except for T*aq* DNA polymerase and template, which are given as amounts per 50 µL reaction

Table 2.6 Reagents and Cycling Conditions for Southern Blotting Probes

Reagents/ Cycle conditions	PCR of Probes for Southern Blotting				
	cyp ^{dot}	*pR209*	*ddhA*	*ver-2*	*aflR*
***Taq* DNA polymerase**	2 units	2 units	2 units	2 units	2 units
***Taq* buffer**	1 X	1 X	1 X	1 X	1 X
$MgCl_2$	3.0 mM	3.0 mM	3.0 mM	3.0 mM	3.0 mM
dNTPs	0.2 mM	0.2 mM	0.2 mM	0.2 mM	0.2 mM
Template	1 ng of pR163	1 ng of pR209	1 ng of pR213	1 ng of pR145	1 ng of pR212
Primers	0.2 µM of each dmoIR dmoCB	0.2 µM of each T7 Sp6	0.2 µM of each T7 Sp6	0.2 µM of each M13 Fwd M13 Rev	0.2 µM of each T7 Sp6
Cycle					
intial denaturation	2 min at 94°C	2 min at 94°C	2 min at 94°C	2 min at 94°C	2 min at 94°C
	30 cycles of:	30 cycles of:	30 cycles of:	30 cycles of:	30 cycles of:
denaturation	30 sec at 94°C	30 sec at 94°C	30 sec at 94°C	30 sec at 94°C	30 sec at 94°C
annealing	30 sec at 58°C	30 sec at 58°C	30 sec at 58°C	30 sec at 58°C	30 sec at 58°C
extension	2.5 mins at 75°C	2.5 mins at 75°C	2.5 mins at 75°C	2.5 mins at 75°C	2.5 mins at 75°C

Final concentrations of all reagents are shown except for T*aq* DNA polymerase and template, which are given as amounts per 50 µL reaction

Table 2.7 Reagents and Cycling Conditions Checking Orientation of pR162A

Reagents/ Cycle conditions	Amplification between R181 and R162A
***Taq* DNA polymerase**	1 unit
***Taq* buffer**	1 X
$MgCl_2$	1.5 mM
dNTPs	0.1 mM
Template	3 ng genomic DNA
Primers	0.08 µM R181 Ovlp 0.08 µM R162A Rev out
Cycle	
intial denaturation	2 min at 94°C
	30 cycles of:
denaturation	30 sec at 94°C
annealing	30 sec at 60°C
extension	90 sec at 72°C

Final concentrations of all reagents are shown except for T*aq* DNA polymerase and template, which are given as amounts per 50 µL reaction

2.15 SOUTHERN BLOTTING AND HYBRIDISATION

2.15.1 Southern Blotting

The DNA blotting method used was based on that of Southern (1975). DNA (digested with appropriate enzymes) to be transferred onto a nylon membrane was separated by gel electrophoresis, through a 1-1.5% agarose gel, stained with ethidium bromide (section 2.6.1) and photographed with a ruler alongside the gel. The DNA in the gel was then denatured by gentle agitation for 45 minutes in denaturing solution (section 2.4.12). The gel was neutralized by washing twice in neutralising solution (section 2.4.12) for 25 minutes per wash. Before placing on the blotting apparatus, the gel was washed in 2 x SSC (section 2.4.12) for 5 minutes. The blotting apparatus was constructed as described by Ausubel *et al* (1994). After blotting overnight, the apparatus was disassembled and the Hybond –N membrane (Amersham) was treated with shortwave UV light for 3 minutes to crosslink the DNA to the membrane. The membrane was stored between two acetate sheets at 4° C.

2.15.2 DIG Labelling of DNA Probe

DNA required for probing the membrane was labeled with Digoxygenin (DIG)-11-dUTP with a DIG-random labeling kit (Roche) according to manufacturer's instructions.

2.15.3 Preparation of [α-^{32}P]dCTP- Labelled DNA Probe

Probe DNA (30 ng) was labelled with [α-^{32}P]dCTP (3 000 Ci/mmole, Amersham) using High Prime kit (Roche) according to manufacturer's instructions. However when fresh isotope was used only 3 μL of [α-^{32}P] was used and the template DNA was in a final volume of 13 μL. The reaction was also incubated for at least 30 minutes at 37°C. Unincorporated nucleotides were removed using ProbeQuant G-50 Micro Columns (Amersham) as described by the manufacturers. The labelled probe was boiled before addition to the pre-hybridisation blot.

2.15.4 Southern Blot Hybridisation of DIG labelled Probe

The membrane was placed in a hybridisation tube and washed with 50 mL of 2 x SSC for a few minutes. The 2 x SSC was gently poured out and 20 mL of DIG Easy Hyb solution (Roche) (20 mL/100 cm^2 of membrane) was added to the tube. The membrane was placed in a rotary oven at 40°C and prehybridised for one hour.

The labeled probe was denatured by heating for 5 minutes in a boiling water bath and plunging immediately into ice. After prehybridisation, the prehybridisation solution was replaced with 20 mL of fresh DIG Easy Hyb buffer that contained 2.5 ng of the denatured probe per mL of DIG Easy Hyb. The contents of the hybridisation tube were mixed briefly and incubated overnight at the hybridization temperature (T_{opt} = T_m – 20 to 25°C, where T_m= 49.82 + 0.41 (%GC)- (600/I) and I= length of hybrid in base pairs.

After overnight incubation, excess probe was washed off the blot as follows:

The membrane was removed from the hybridisation tube with forceps and placed in 200 mL of wash solution one (section 2.4.12) (prewarmed to 68°C) and incubated on a shaking platform at 68°C for 5 minutes. Wash solution one was replaced with 200 mL of fresh warmed wash solution one and incubated on a shaking platform at 68°C for 5 minutes. Wash solution one was replaced with 200 mL of warmed wash solution two (section 2.4.12) and incubated on a shaking platform at 68°C for 15 minutes. This step was repeated with 200 mL of wash solution two. The membrane was then wrapped in gladwrap and stored at 4°C until needed for chemiluminescent detection.

2.15.5 Southern Blot Hybridisation of [α-^{32}P]dCTP- Labelled DNA Probe and Detection

The membrane to be probed was placed in a hybridisation tube and pre-hybridised with 20 – 30 mL of hybridisation solution (section 2.4.13) for 2 hours prior to the addition of the labelled probe. The membrane was hybridised to boiled labelled probe in a hybridisation solution volume of 10 mL. Hybridisation was carried out at 65°C overnight. Following hybridisation, the filters were washed three times in 2 x SSC (section 2.4.12) (heated to

55°C) for 15 minutes at room temperature before exposing to X-ray film (Fuji). Membranes were exposed to X-ray film for 2-7 days depending on strength of isotope at -80°C. The film was then developed using an automated developer (100Plus Automatic X-ray Processor, All Pro Imaging).

2.15.5 Chemiluminescent Detection of DIG Labelled DNA Probe

The membrane was washed for 1-5 minutes in washing buffer (section 2.4.11). The membrane was then incubated in 1 x blocking solution (section 2.4.11) for 30 minutes and then incubated in Antibody solution (section 2.4.11) for 30 minutes. Membrane was washed twice for 15 minutes in washing buffer and then equilibrated for 1-5 minutes in washing buffer (section 2.4.11) and then equilibrated for 1-5 minutes in detection buffer (section 2.4.11).

The wet membrane was placed onto acetate paper with the DNA side facing up. 1 mL CSPD ready-to-use lumigen was dispensed over the surface of the membrane and a second acetate sheet was placed over top. Lumigen was spread evenly and gently over the surface and left for 1 minute. The top acetate sheet was then wiped firmly to remove the excess lumigen from the membrane. The membrane was left between the two acetate sheets with the thin layer lumigen for 5 minutes. The membrane was then removed from the acetate and blotted on clean 3MM filter paper and placed between two fresh sheets of acetate and incubated for 15 minutes at 37°C. The membrane was then exposed to X-ray (Fuji) overnight and developed using an automated developer (100Plus Automatic X-ray Processor, All Pro Imaging).

2.15.6 Removal of DIG Labelled DNA From Southern Blots (Stripping)

Membranes to be stripped were washed thoroughly in double distilled water, then washed for 15 minutes at 37°C in 0.2 M NaOH containing 0.1% (v/v) SDS (repeated twice). Membranes were then rinsed thoroughly in 2 x SSC (section 2.4.12). Stripping was confirmed by carrying out chemiluminescent detection (section 2.15.5).

2.15.7 Removal of [α-^{32}P]dCTP- Labelled DNA From Southern Blots (Stripping)

Membranes to be stripped were placed in a container containing 0.1% SDS at 100°C and left shaking for 30 minutes (repeated three times). Stripping was confirmed by exposing membranes to X-ray film (Fuji) overnight at -80°C and developing using an automated developer (100Plus Automatic X-ray Processor, All Pro Imaging).

3.0 SEQUENCING OF λKSA AND IDENTIFICATION OF PUTATIVE DOTHISTROMIN BIOSYNTHETIC GENES

3.1 INTRODUCTION

Southern blot analysis of *D. pini* genomic DNA was carried out using a probe (KS-2) encoding the highly conserved β keto-synthase domain from the aflatoxin polyketide synthase gene (*pksL1*) of *A. parasiticus*. This indicated the presence of a homologous sequence in *D. pini*. A hybridising clone λKSA, was subsequently isolated from a *D. pini* genomic library. Double digestion of the λKSA clone with the restriction enzymes *Bam*HI and *Eco*RI allowed ligation of these *Bam*HI/*Eco*RI subclones into separate vectors, to enable sequencing to be carried out. Prior to this study sequencing of the λKSA clone was approximately half complete (Morgan, 1997 and Seconi, 2001); two *Bam*HI fragments of 4.5 Kb and 3.0 Kb remained to be sequenced. The λKSA subclones sequenced thus far include pR209, pR156, pR163 all of which have been sequenced on both strands (Morgan, 1997; Laarakkers, 1999; Seconi, 2001) and pR181 which has been completely sequenced on one strand and partially sequenced on the other strand (Seconi, 2001). These subclones are all indicated on figure 3.1. Sequencing of these fragments was achieved by using universal forward and reverse primers and primer walking with custom primers. PCR was used to determine overlapping sequences and consequently determined the orientation of the subclones with respect to one another.

Analysis of λKSA had revealed a putative *pks* gene and this is contained within subclones pR209, pR156 and pR163. BLAST analysis of the partial PKS^{DOT} (1-1426 aa) sequence showed the first 1426 amino acids to have 62% identity (74% similarity) to PKSA (≡ PKSL1), a PKS involved in AF synthesis in *A. parasiticus* and 63% (74%) to STCA, a PKS involved in ST synthesis in *A. nidulans* (Seconi, 2001). A partial ORF homologous to the *A. nidulans* monooxygenase *stcW* gene (Laarakkers, 1999) has also been identified within the subclone R161 and provisionally called mox^{dot}. Further sequencing and analysis should identify other genes in this dothistromin biosynthetic gene cluster and verify the presence of mox^{dot}.

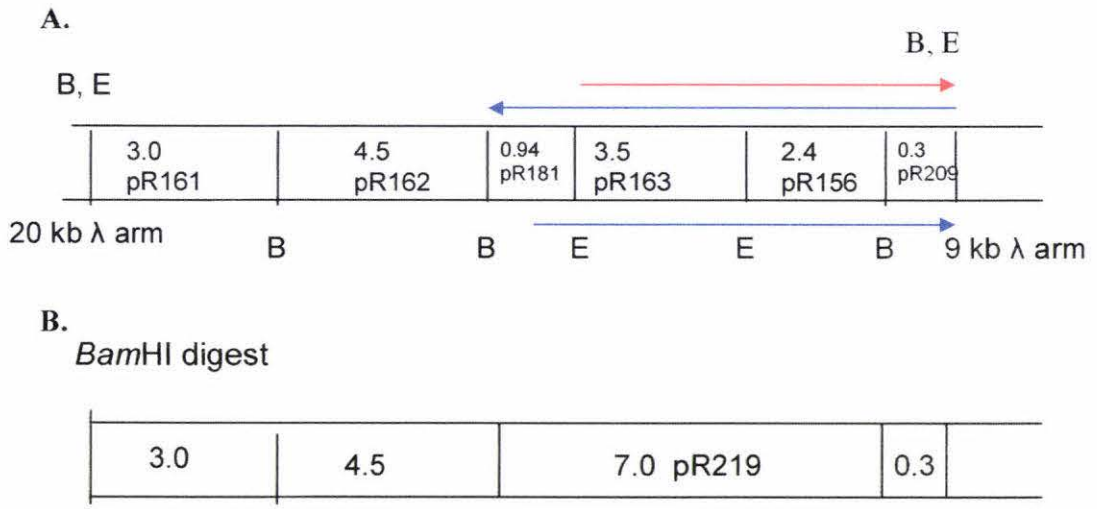


Figure 3.1 λKSA Clone

Schematic of the λKSA restriction map, Showing **A.** *Bam*HI (B) and *Eco*RI (E) restriction sites and **B.** fragment sizes produced by a *Bam*HI digest. Subclones (pR161 etc) are indicated and also the size of each of these subclones. Regions of the λKSA subclones sequenced prior to this project are indicated by the blue arrows. The partial *pks* gene is indicated by the red arrow. All sizes are given in kb.

3.2 SEQUENCING OF λKSA

To continue sequencing the λKSA clone, fresh subclones had to be obtained as old stocks of these had degraded. The λKSA clone DNA was repackaged and single plaques obtained (Section 2.6.7). λ DNA was digested with *Bam*HI with the aim of obtaining each of the 4.5 Kb, 3.0 Kb and 7.0 Kb fragments (Figure 3.1b). After phenol/chloroform extraction of the digested DNA a shot gun cloning approach was used (section 2.11.2). Digested λ DNA was ligated in pUC118 vectors (section 2.11.1.1) using three different vector to insert ratios to achieve the most efficient ligation reaction. After transforming *E. coli (XL-1)* (section 2.12.1.3) transformants were screened to identify those obtaining inserts. Digestions were performed to check insert sizes (section 2.8). After screening approximately 50

transformants, a subclone containing the 4.5 Kb *Bam*HI fragment (pR162A) was identified and also a 7.0 Kb fragment which after sequencing was shown to contain R181 (0.94 Kb), R163 (3.5 Kb) and R156 (2.4 Kb) subclones. The plasmid containing the 7.0 Kb *Bam*HI fragment is now referred to as pR219.

3.2.1 Sequencing of pR162A

pR162A was named to distinguish it from the old 4.5 Kb pR162 subclone that had been lost. Sequencing of this subclone was completed on both strands using universal forward and reverse pUC/M13 primers and primer walking with custom primers (Table 2.1 and Figure 3.2). PCR amplification was used to determine the overlapping sequences between pR161 & pR162A and pR181 & pR162A and consequently determined the orientation of the pR162A subclone. The overlap between pR181 and pR162A was confirmed by amplifying genomic DNA with primers R181 Ovlp and R162A Fwd out and then by sequencing the PCR product. PCR and sequencing of the overlap between pR161 and pR162A and sequencing of pR161 have been carried out in another project (Jin Ping, unpublished data).

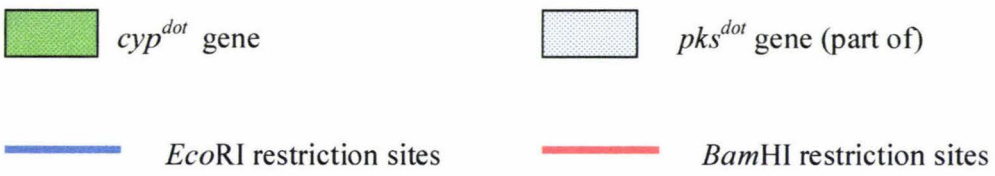
3.2.2 pR162A Contains Three Putative Dothistromin Biosynthetic Genes

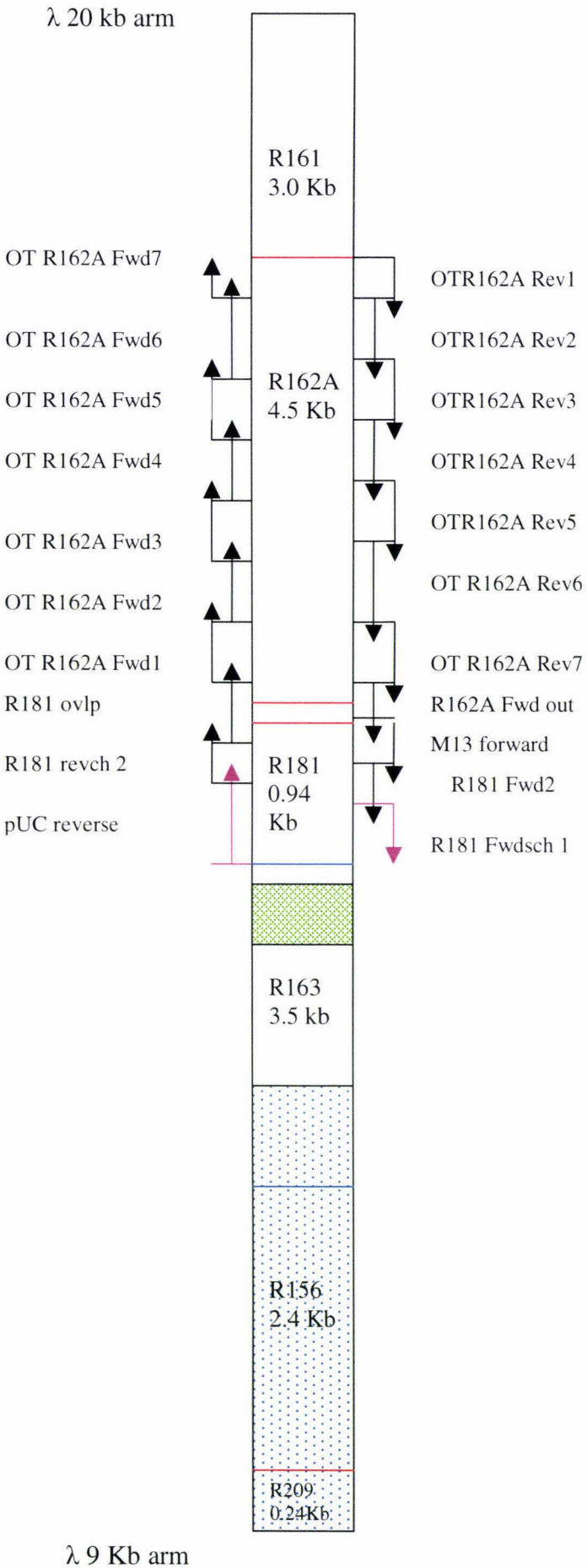
Sequence analysis identified two entire open reading frames and one partial open reading frame within the 4.5 Kb pR162A subclone (Figure 3.3), all of which display significant identity to previously described genes (Table 3.1). Two of these genes show similarity to genes located within the ST and AF biosynthetic gene clusters, indicating that these genes may be involved in dothistromin biosynthesis. Inspection of sequence databases was performed using the BLAST search programme (National Centre for Biotechnology Information). Introns predicted on the basis of nucleotide sequence were confirmed using RT-PCR (section 2.13 and Table 2.5). Sequence analysis of each gene is subsequently discussed in this chapter and summarized in Table 3.1. The possible functions of gene products discussed in this chapter are only speculative the roles of these genes in dothistromin biosynthesis need to be confirmed. The gene names (*ord*^{dot}, *mox*^{dot}, *epox*^{dot})

Figure 3.2 Double Stranded Sequencing Strategy for pR162A

Figure 3.2 shows the direction and approximate location of sequencing primers used for double stranded sequencing of pPR162A and overlapping regions. Arrows indicate direction of sequencing; the primers are indicated by labels adjacent to arrows. Pink arrows indicate sequencing carried out in previous studies and black arrows indicate sequencing carried out in this study. Specific location and sequence of primers can be found in appendix 3.

Legend

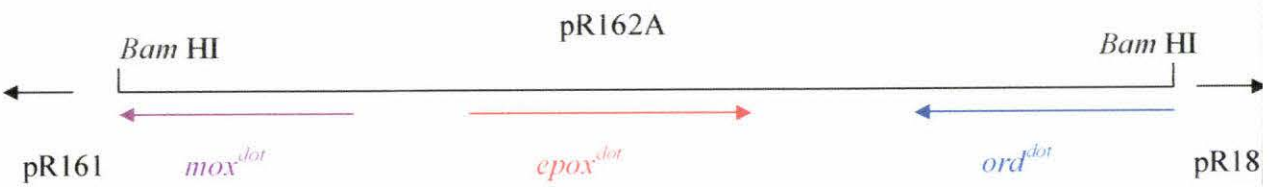



λ 20 kb arm
R161
3.0 Kb
OT R162A Fwd7
OT R162A Fwd6
OT R162A Fwd5
OT R162A Fwd4
OT R162A Fwd3
OT R162A Fwd2
OT R162A Fwd1
R181 ovlp
R181 revch 2
pUC reverse
R162A
4.5 Kb
OTR162A Rev1
OTR162A Rev2
OTR162A Rev3
OTR162A Rev4
OTR162A Rev5
OT R162A Rev6
OT R162A Rev7
R162A Fwd out
M13 forward
R181 Fwd2
R181 Fwdsch 1
R181
0.94
Kb
R163
3.5 kb
R156
2.4 Kb
R209
0.24Kb
λ 9 Kb arm

will be renamed in keeping with the published dot gene nomenclature (*dotA*, *dotB* etc) once their position in the gene cluster is known and verified.

Figure 3.3 Putative Genes Identified In pR162A

A schematic Map of the three genes identified in the pR162A subclone is shown. Arrows indicate position and direction of each gene.



3.3 SEQUENCE OF THE *MOX*^{DOT} GENE

The *mox*^{dot} (monooxygenase – dothistromin) gene displays significant sequence identity to many monooxygenase genes, especially those encoding proteins involved in ST/AF biosynthesis. The monooxygenase which displays the highest identity to the predicted *mox*^{dot} gene product is a monooxygenase present in *A. oryzae* which displays 71% amino acid identity (84% similarity). Functions and references for all genes showing significant similarities are shown in Table 3.1.

The *mox*^{dot} ORF is estimated to be approximately 1148 bp in length (966 bp without intron sequences (section 3.3.1), encoding a deduced polypeptide product of 322 amino acid residues, comparable with the gene products described above. The *mox*^{dot} translation initiation sequence (CCACTATGGCTC) was estimated at position 3179 and is consistent with the consensus sequence for filamentous fungi [CNNNCA(A/C)NATGGC] (Bruchez and Eberle, 1993). A putative termination codon has not been identified as this gene continues into another subclone R161, which has not been sequenced as part of this thesis. The nucleotide sequence and deduced amino acid sequence for *mox*^{dot} is shown in Figure

Table 3.1 Summary of *D. pini* Genes Present in pR162A

Gene products displaying the highest identity and similarity to the deduced amino acid sequences of the three putative genes identified are shown.

D. pini gene	Similar gene/protein	Putative function of gene products	% aa identity (% aa similarity)	Genbank accession number (reference)
mox^{dot}	*moxY- Aspergillus parasiticus* *moxY- A. oryzae* *stcW- A. nidulans*	Monooxygenase- AF biosynthesis Monooxygenase-Sugar utilisation Monooxygenase- ST biosynthesis	70 (84) 71 (84) 63 (75)	AAS66023 (Yu, Chang *et al.* 2000) BAC20334 QOO730 (Brown, Yu *et al.* 1996)
ord^{dot}	*avfA- A. flavus* *avfA- A. parasiticus* *avfA- A. oryzae* *stcO- A. nidulans*	Averufin oxidase- AF biosynthesis Averufin oxidase- AF biosynthesis P450 monooxygenase Oxido-reductase- ST biosynthesis	51 (66) 51 (67) 51 (66) 47 (61)	AAF26224 (Yu, Woloshuk *et al.* 2000) AAS66010 (Yu, Woloshuk *et al.* 2000) BAC45244 QOO710 (Yu, Chang *et al.* 2000)
epox^{dot}	*hyll- A. niger* Hypothetical protein AN3017.2 – *A. nidulans*	Epoxide hydrolase- catabolic role Epoxide hydrolase	45 (60) 44 (61)	CAB59812 (Arand, Hemmer *et al.* 1999) EAA63588

3.4. Like StcW of *A. nidulans* and MoxY of *A. parasiticus* no characteristic motifs which are found in some cytochrome p450 type monooxygenases were identified within the putative Mox^{dot} protein.

3.3.1 The Partial *mox^{dot}* Gene Contains Three Introns

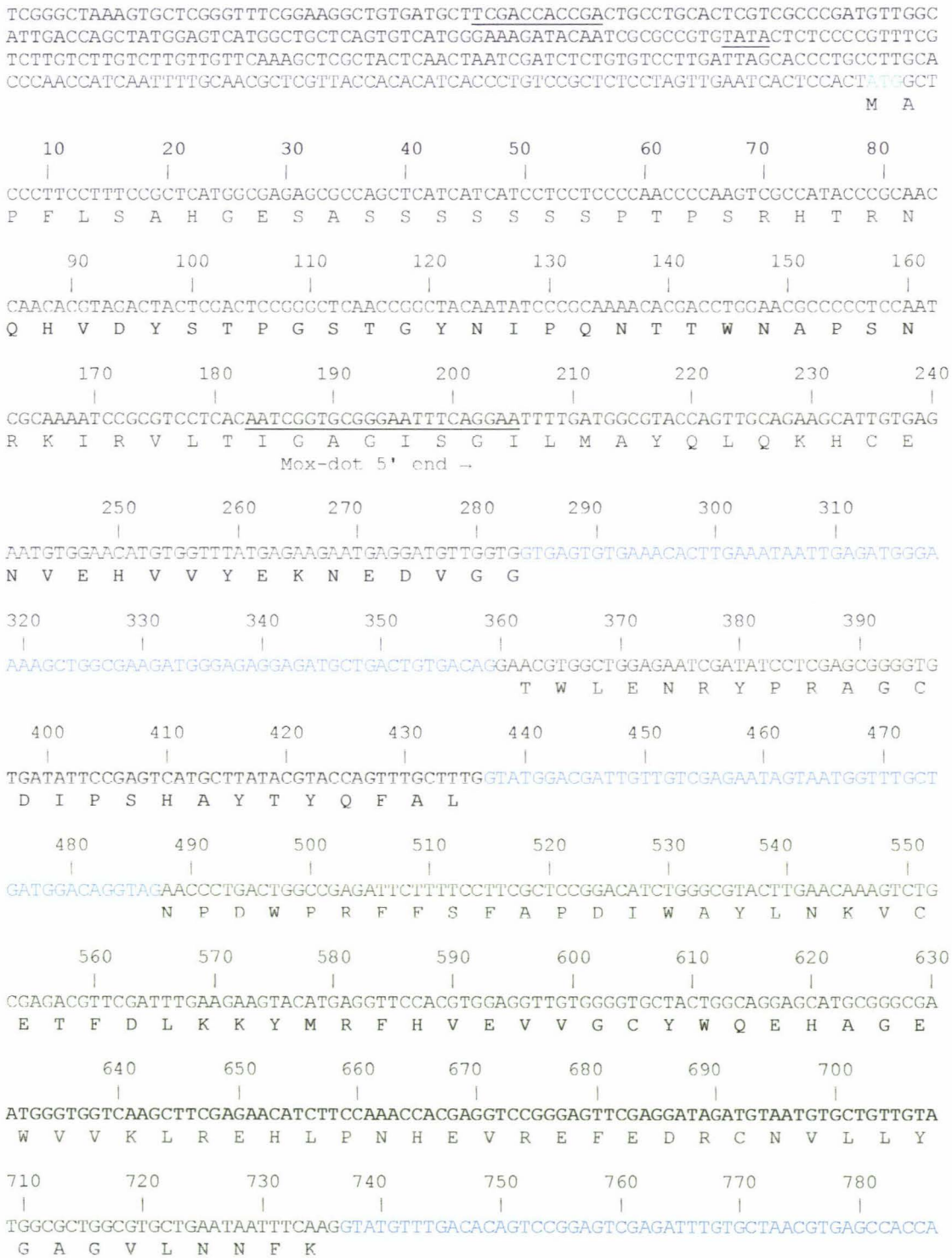
Putative intron regions were identified by the presence of 5' [GT(A/G/T)NGTY] and 3' (YAG) fungal consensus intron splice site sequences (Balance, 1986; Bruchez *et al.*, 1993), as well as continuance of amino acid similarity with other sequences. Three putative introns were identified, the first 75 nt in length and starting at position 3462, the second 51 nt in length and starting at pR162A nucleotide position 3614, and the third 51 nt in length and starting at position 3914 (Figure 3.4 and 3.5). Intron positions were confirmed by direct sequencing of a RT-PCR product amplified from *D. pini* RNA using primers mox^{dot} 5' end and mox^{dot} 3' end (courtesy of Arne Schwelm). See appendix 3 for RT-PCR sequences. The positions of all three introns are conserved between *mox^{dot}* and the ST gene *stcW* whilst the AF gene *moxY* does not contain any introns.

3.3.2 Amino Acid Sequence Comparison

The deduced amino acid sequence for *mox^{dot}* was aligned with those of *stcW* and *moxY* from *A. nidulans* and *A. parasiticus* respectively and also to a *moxY* from *A. oryzae* using CLUSTAL X (Figure 3.5). This showed high identity between the genes throughout the sequence, except for a lack of unity at the very beginning of the alignment; it appears the start position of *mox^{dot}* differs from that of *stcW (A. nidulans) and moxY (A. parasiticus* and *A. oryzae).* The lack of unity at the beginning of the alignment could be explained by the presence of a non-consensus intron; however this would have to be confirmed by analysis of cDNA or RACE. Due the extent of sequence identity observed between the proteins it is likely that like *stcW* and *moxY*, *mox^{dot}* is a monooxygenase.

Figure 3.4 Partial Nucleotide and Deduced Amino Acid Sequence of the Putative *Dothistroma pini mox*^{dot} Gene

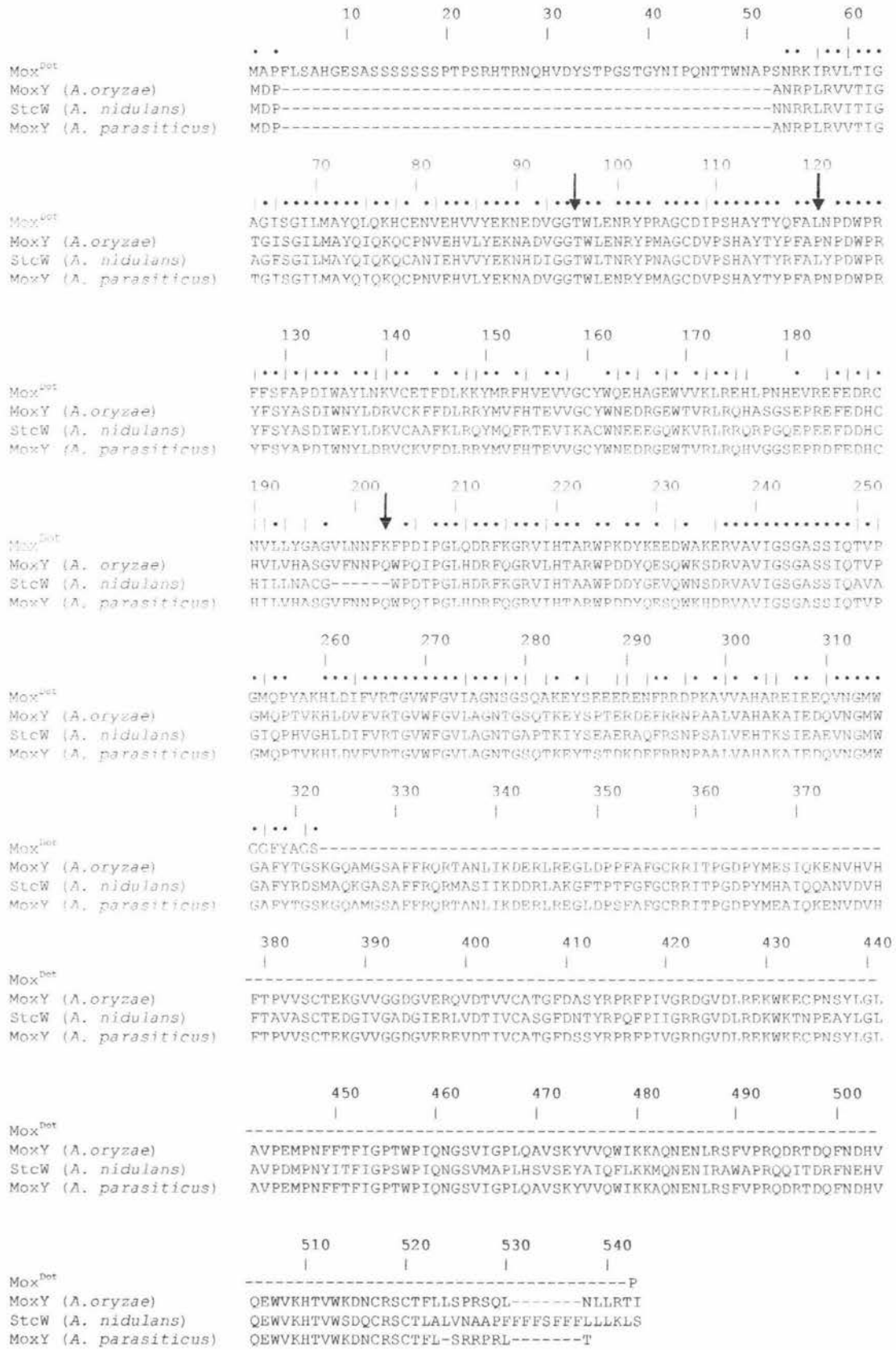
The coding sequence, promoter regions and the deduced amino acid sequence of *mox*^{dot} are shown. The nucleotide sequence starts at position 3179 relative to the start of pR162A sequence. The initiation codon is indicated in green and intron consensus sequences are shown in blue. The positions of primers used in RT-PCR are indicated. Possible TATA and AFLR binding sites are indicated by underlining.


TCGGGCTAAAGTGCTCGGGTTTCGGAAGGCTGTGATGCTTCGACCACCGACTGCCTGCACTCGTCGCCCGATGTTGGC
ATTGACCAGCTATGGAGTCATGGCTGCTCAGTGTCATGGGAAAGATACAATCGCGCCGTGTATACTCTCCCCGTTTCG
TCTTGTCTTGTCTTGTTGTTCAAAGCTCGCTACTCAACTAATCGATCTCTGTGTCCTTGATTAGCACCCTGCCTTGCA
CCCAACCATCAATTTTGCAACGCTCGTTACCACACATCACCCTGTCCGCTCTCCTAGTTGAATCACTCCACTATGGCT
M A
10 20 30 40 50 60 70 80
CCCTTCCTTTCCGCTCATGGCGAGAGCGCCAGCTCATCATCATCCTCCTCCCCAACCCCAAGTCGCCATACCCGCAAC
P F L S A H G E S A S S S S S S S S P T T P S R H T R N
90 100 110 120 130 140 150 160
CAACACGTAGACTACTCGACTCCGGGCTCAACCGGCTACAATATCCCGCAAAACACGACCTGGAACGCCCCCTCCAAT
Q H V D Y S T P G S T G Y N I P Q N T T W N A P S N
170 180 190 200 210 220 230 240
CGCAAAATCCGCGTCCTCACAATCGGTGCGGGAATTTCAGGAATTTTGATGGCGTACCAGTTGCAGAAGCATTGTGAG
R K I R V L T I G A G I S G I L M A Y Q L Q K H C E
Mox-dot 5' end →
250 260 270 280 290 300 310
AATGTGGAACATGTGGTTTATGAGAAGAATGAGGATGTTGGTGGTGAGTGTGAAACACTTGAAATAATTGAGATGGGA
N V E H V V Y E K N E D V G G
320 330 340 350 360 370 380 390
AAAGCTGGCGAAGATGGGAGAGGAGATGCTGACTGTGACAGGAACGTGGCTGGAGAATCGATATCCTCGAGCGGGGTG
T W L E N R Y P R A G C
400 410 420 430 440 450 460 470
TGATATTCCGAGTCATGCTTATACGTACCAGTTTGCTTTGGTATGGACGATTGTTGTCGAGAATAGTAATGGTTTGCT
D I P S H A Y T Y Q F A L
480 490 500 510 520 530 540 550
GATGGACAGGTAGAACCCTGACTGGCCGAGATTCTTTTCCTTCGCTCCGGACATCTGGGCGTACTTGAACAAAGTCTG
N P D W P R F F S F A P D I W A Y L N K V C
560 570 580 590 600 610 620 630
CGAGACGTTCGATTTGAAGAAGTACATGAGGTTCCACGTGGAGGTTGTGGGGTGCTACTGGCAGGAGCATGCGGGCGA
E T F D L K K Y M R F H V E V V G C Y W Q E H A G E
640 650 660 670 680 690 700
ATGGGTGGTCAAGCTTCGAGAACATCTTCCAAACCACGAGGTCCGGGAGTTCGAGGATAGATGTAATGTGCTGTTGTA
W V V K L R E H L P N H E V R E F E D R C N V L L Y
710 720 730 740 750 760 770 780
TGGCGCTGGCGTGCTGAATAATTTCAAGGTATGTTTGACACAGTCCGGAGTCGAGATTTGTGCTAACGTGAGCCACCA
G A G V L N N F K

```
 790        800        810        820        830        840        850        860
   |          |          |          |          |          |          |          |
TTCCCCGATATTCCAGGCCTGCAGGATCGCTTCAAGGGCCGGGTCATTCACACCGCCCGGTGGCCAAAGGACTACAA
 F  P  D  I  P  G  L  Q  D  R  F  K  G  R  V  I  H  T  A  R  W  P  K  D  Y  K

     870        880        890        900        910        920        930        940
      |          |          |          |          |          |          |          |
GGAGGAAGACTGGGCAAAGGAGCGAGTCGCCGTAATTGGATCTGGTGCTTCATCCATACAGACTGTTCCAGGCATGCA
 E  E  D  W  A  K  E  R  V  A  V  I  G  S  G  A  S  S  I  Q  T  V  P  G  M  Q

       950        960        970        980        990       1000       1010
        |          |          |          |          |          |          |
GCCATACGCAAAACACCTCGACATCTTCGTCCGTACGGGAGTCTGGTTCGGGGTGATAGCCGGAAACTCTGGTAGTCA
 P  Y  A  K  H  L  D  I  F  V  R  T  G  V  W  F  G  V  I  A  G  N  S  G  S  Q
                                                    Mox-dot 3' end ←

1020       1030       1040       1050       1060       1070       1080       1090
|          |          |          |          |          |          |          |
AGCCAAAGAGTACAGCGAAGAAGAGCGAGAGAACTTCCGTCGCGATCCCAAAGCCGTCGTTGCTCACGCCAGAGAGAT
 A  K  E  Y  S  E  E  E  R  E  N  F  R  R  D  P  K  A  V  V  A  H  A  R  E  I

 1100       1110       1120       1130       1140
  |          |          |          |          |
CGAAGAGCAGGTGAACGGCATGTGGGGTGGATTCTATGCTGGATCCCCGGG
 E  E  Q  V  N  G  M  W  G  G  F  Y  A  G  S  P
```

Figure 3.5 Alignments of the Deduced Amino Acid Sequence of Mox^{Dot} with Other Monooxygenases

The deduced amino acid sequence of the putative *D. pini* gene product (Mox^{Dot}) was aligned with that of the *A. parasiticus moxY* gene (MoxY), *A. nidulans stcW gene* (StcW) and *A. oryzae moxY* gene (Mox-Y). See table 3.1 for references. Sequence alignment was performed using CLUSTAL X. Amino acid identity is indicated by dots (all three sequences share same residue) and lines (two of three residues share same residue). The Mox^{Dot} amino acid sequence is truncated compared to those of *A. parasiticus, A. nidulans* and *A. oryzae* as the mox^{dot} gene continues into another subclone R161, which has not been sequenced as part of this thesis. Intron positions are indicated by black arrows. The positions of the introns are also conserved in StcW of *A. nidulans.*


MoxDot MAPFLSAHGESASSSSSSSPTPSRHTRNQHVDYSTPGSTGYNIPQNTTWNAPSNRKIRVLTIG
MoxY (A.oryzae) MDP----------ANRPLRVVTIG
StcW (A. nidulans) MDP----------NNRRLRVITIG
MoxY (A. parasiticus) MDP----------ANRPLRVVTIG
MoxDot AGISGILMAYQLQKHCENVEHVVYEKNEDVGGTWLENRYPRAGCDIPSHAYTYQFALNPDWPR
MoxY (A.oryzae) TGISGILMAYQIQKQCPNVEHVLYEKNADVGGTWLENRYPMAGCDVPSHAYTYPFAPNPDWPR
StcW (A. nidulans) AGFSGILMAYQIQKQCANIEHVVYEKNHDIGGTWLTNRYPNAGCDVPSHAYTYRFALYPDWPR
MoxY (A. parasiticus) TGTSGILMAYQTQKQCPNVEHVLYEKNADVGGTWLENRYPMAGCDVPSHAYTYPFAPNPDWPR
MoxDot FFSFAPDIWAYLNKVCETFDLKKYMRFHVEVVGCYWQEHAGEWVVKLREHLPNHEVREFEDRC
MoxY (A.oryzae) YFSYASDIWNYLDRVCKFFDLRRYMVFHTEVVGCYWNEDRGEWTVRLRQHASGSEPREFEDHC
StcW (A. nidulans) YFSYASDIWEYLDKVCAAFKLRQYMQFRTEVIKACWNEEEGQWKVRLRRQRPGQEPEEFDDHC
MoxY (A. parasiticus) YFSYAPDIWNYLDRVCKVFDLRRYMVFHTEVVGCYWNEDRGEWTVRLRQHVGGSEPRDFEDHC
MoxDot NVLLYGAGVLNNFKFPDIPGLQDRFKGRVIHTARWPKDYKEEDWAKERVAVIGSGASSIQTVP
MoxY (A. oryzae) HVLVHASGVFNNPQWPQIPGLHDRFQGRVLHTARWPDDYQESQWKSDRVAVIGSGASSIQTVP
StcW (A. nidulans) HTLLNACG------WPDTPGLHDRFKGRVIHTAAWPDDYGEVQWNSDRVAVIGSGASSIQAVA
MoxY (A. parasiticus) HTLVHASGVFNNPQWPQTPGLHDRFQGRVIHTARWPDDYQESQWKHDRVAVIGSGASSIQTVP
MoxDot GMQPYAKHLDIFVRTGVWFGVIAGNSGSQAKEYSEEERENFRRDPKAVVAHAREIEEQVNGMW
MoxY (A.oryzae) GMQPTVKHLDVFVRTGVWFGVLAGNTGSQTKEYSPTERDEFRRNPAALVAHAKAIEDQVNGMW
StcW (A. nidulans) GIQPHVGHLDIFVRTGVWFGVLAGNTGAPTKIYSEAERAQFRSNPSALVEHTKSIEAEVNGMW
MoxY (A. parasiticus) GMQPTVKHLDVFVRTGVWFGVLAGNTGSQTKEYTSTDKDEFRRNPAALVAHAKAIEDQVNGMW
MoxDot CGFYAGS----------
MoxY (A.oryzae) GAFYTGSKGQAMGSAFFRQRTANLIKDERLREGLDPFFAFGCRRITPGDPYMESIQKENVHVH
StcW (A. nidulans) GAFYRDSMAQKGASAFFRQRMASIIKDDRLAKGFTPTFGFGCRRITPGDPYMHAIQQANVDVH
MoxY (A. parasiticus) GAFYTGSKGQAMGSAFFRQRTANLIKDERLREGLDPSFAFGCRRITPGDPYMEAIQKENVDVH
MoxDot ----------
MoxY (A.oryzae) FTPVVSCTEKGVVGGDGVERQVDTVVCATGFDASYRPRFPIVGRDGVDLREKWKECPNSYLGL
StcW (A. nidulans) FTAVASCTEDGIVGADGIERLVDTIVCASGFDNTYRPQFPIIGRRGVDLRDKWKTNPEAYLGL
MoxY (A. parasiticus) FTPVVSCTEKGVVGGDGVEREVDTIVCATGFDSSYRPRFPIVGRDGVDLREKWKECPNSYLGL
MoxDot ----------
MoxY (A.oryzae) AVPEMPNFFTFIGPTWPIQNGSVIGPLQAVSKYVVQWIKKAQNENLRSFVPRQDRTDQFNDHV
StcW (A. nidulans) AVPDMPNYITFIGPSWPIQNGSVMAPLHSVSEYAIQFLKKMQNENIRAWAPRQQITDRFNEHV
MoxY (A. parasiticus) AVPEMPNFFTFIGPTWPIQNGSVIGPLQAVSKYVVQWIKKAQNENLRSFVPRQDRTDQFNDHV
MoxDot ----------P
MoxY (A.oryzae) QEWVKHTVWKDNCRSCTFLLSPRSQL-------NLLRTI
StcW (A. nidulans) QEWVKHTVWSDQCRSCTLALVNAAPFFFFSFFFLLLKLS
MoxY (A. parasiticus) QEWVKHTVWKDNCRSCTFL-SRRPRL-------T

3.4 *ORD*^{DOT} ENCODES A PUTATIVE OXIDO-REDUCTASE

The *ord*^{dot} (oxido-reductase – dothistromin) gene displays similarities to oxido-reductases involved in ST /AF synthesis. The deduced amino acid sequence of *ord*^{dot} displays the highest identity and similarity to that of the *A. parasiticus avfA* gene (51% amino acid identity, 67% similarity). Functions and references for all genes with high similarity to *ord*^{dot} are shown in Table 3.1.

The *ord*^{dot} ORF is estimated to be 865 bp in length encoding a deduced polypeptide product of 288 amino acid residues which is a similar length to those of the genes described in Table 3.1. The translation initiation codon (ATG) was determined at position 189 bp as it fitted in the context of the predicted amino acid sequences of other oxido-reductase genes. The termination codon (TAG) is at nucleotide position 1043. The nucleotide sequence and deduced amino acid sequence for *ord*^{dot} is shown in Figure 3.6.

The cDNA sequence of *ord*^{dot} derived from primers OT PR162A Fwd1 and OT PR162A Rev 6 was identical to that shown in Figure 3.6 suggesting a lack of introns. This is consistent with the *A. parasiticus avfA*, *A. nidulans stcO* and *avfA* genes from *A. flavus* and *A. oryzae*, all of which do not appear to contain introns. However further analysis through cDNA/RACE is required to confirm the lack of introns in the regions upstream and downstream of the RT-PCR primers indicated in Figure 3.6.

The deduced amino acid sequence of *mox*^{dot} was aligned with those of the *A. nidulans stcO*, *A. parasiticus avfA*, *A. flavus avfA* and *A. oryzae avfA* using CLUSTAL X (Figure 3.7). A high level of aa identity was shown between *ord*^{dot} and other oxido-reductases except towards the termination codon.

Figure 3.6 Nucleotide and Deduced Amino Acid Sequence of the Putative *Dothistroma pini ord*^{dot} Gene

The coding sequence, promoter and terminator regions and the deduced amino acid sequence of *ord*^{dot} is shown. The nucleotide sequence starts at position 189 relative to the start of pR162A sequence. Initiation and terminator codons are indicated in green.
The positions of primers used in RT-PCR are indicated.

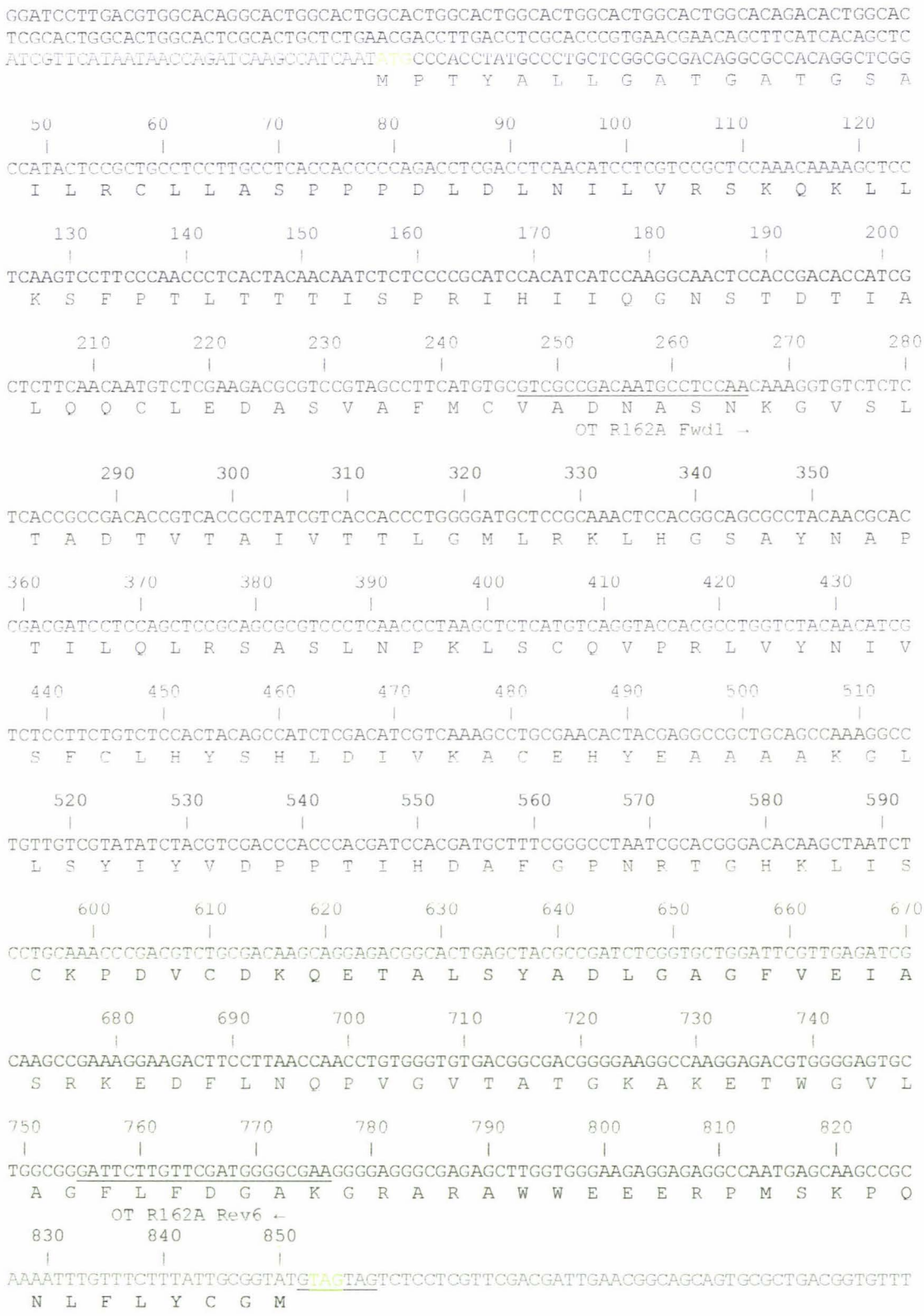
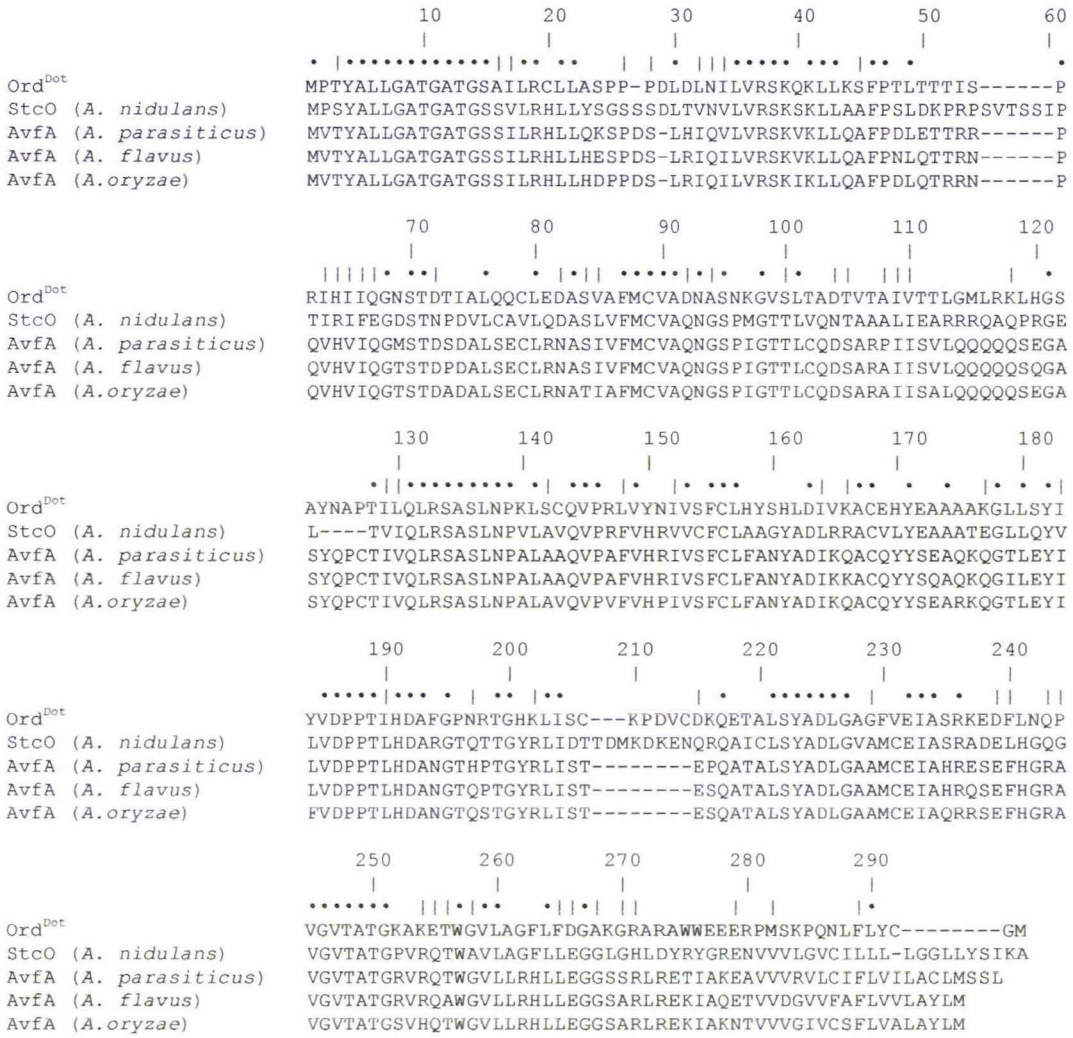

GGATCCTTGACGTGGCACAGGCACTGGCACTGGCACTGGCACTGGCACTGGCACTGGCACAGACACTGGCAC
TCGCACTGGCACTGGCACTCGCACTGCTCTGAACGACCTTGACCTCGCACCCGTGAACGAACAGCTTCATCACAGCTC
ATCGTTCATAATAACCAGATCAAGCCATCAATATGCCCACCTATGCCCTGCTCGGCGCGACAGGCGCCACAGGCTCGG
M P T Y A L L G A T G A T G S A
50 60 70 80 90 100 110 120
CCATACTCCGCTGCCTCCTTGCCTCACCACCCCCAGACCTCGACCTCAACATCCTCGTCCGCTCCAAACAAAAGCTCC
I L R C L L A S P P P D L D L N I L V R S K Q K L L
130 140 150 160 170 180 190 200
TCAAGTCCTTCCCAACCCTCACTACAACAATCTCTCCCCGCATCCACATCATCCAAGGCAACTCCACCGACACCATCG
K S F P T L T T T I S P R I H I I Q G N S T D T I A
210 220 230 240 250 260 270 280
CTCTTCAACAATGTCTCGAAGACGCGTCCGTAGCCTTCATGTGCGTCGCCGACAATGCCTCCAACAAAGGTGTCTCTC
L Q Q C L E D A S V A F M C V A D N A S N K G V S L
OT R162A Fwd1 →
290 300 310 320 330 340 350
TCACCGCCGACACCGTCACCGCTATCGTCACCACCCTGGGGATGCTCCGCAAACTCCACGGCAGCGCCTACAACGCAC
T A D T V T A I V T T L G M L R K L H G S A Y N A P
360 370 380 390 400 410 420 430
CGACGATCCTCCAGCTCCGCAGCGCGTCCCTCAACCCTAAGCTCTCATGTCAGGTACCACGCCTGGTCTACAACATCG
T I L Q L R S A S L N P K L S C Q V P R L V Y N I V
440 450 460 470 480 490 500 510
TCTCCTTCTGTCTCCACTACAGCCATCTCGACATCGTCAAAGCCTGCGAACACTACGAGGCCGCTGCAGCCAAAGGCC
S F C L H Y S H L D I V K A C E H Y E A A A A K G L
520 530 540 550 560 570 580 590
TGTTGTCGTATATCTACGTCGACCCACCCACGATCCACGATGCTTTCGGGCCTAATCGCACGGGACACAAGCTAATCT
L S Y I Y V D P P T I H D A F G P N R T G H K L I S
600 610 620 630 640 650 660 670
CCTGCAAACCCGACGTCTGCGACAAGCAGGAGACGGCACTGAGCTACGCCGATCTCGGTGCTGGATTCGTTGAGATCG
C K P D V C D K Q E T A L S Y A D L G A G F V E I A
680 690 700 710 720 730 740
CAAGCCGAAAGGAAGACTTCCTTAACCAACCTGTGGGTGTGACGGCGACGGGGAAGGCCAAGGAGACGTGGGGAGTGC
S R K E D F L N Q P V G V T A T G K A K E T W G V L
750 760 770 780 790 800 810 820
TGGCGGGATTCTTGTTCGATGGGGCGAAGGGGAGGGCGAGAGCTTGGTGGGAAGAGGAGAGGCCAATGAGCAAGCCGC
A G F L F D G A K G R A R A W W E E E R P M S K P Q
OT R162A Rev6 ←
830 840 850
AAAATTTGTTTCTTTATTGCGGTATGTAGTAGTCTCCTCGTTCGACGATTGAACGGCAGCAGTGCGCTGACGGTGTTT
N L F L Y C G M
ACGACAGTTATGGTGAGTTTGGCCGCGGTAGTCTTGGTGCAGTATACGGGAACGATGAACAGATAGGAGTGCGACAGA
AACCATATTGGCAGCGTTAAAAAGGCTTGTTCGAGTCTTGATGTAGTCGCATGTCAAAACTTCGAATTTT

Figure 3.7 Alignments of The Deduced Amino Acid Sequence of Ord^{Dot} With Other Oxido-reductases

The deduced amino acid sequence of the putative *D. pini ord^{dot}* gene (Ord^{dot}) was aligned with those of the *A. parasiticus avfA* gene (AvfA), *A. nidulans stcO gene* (StcO), *A. flavus avfA* gene (AvfA) and *A. oryzae avfA* gene (AvfA). See Table 3.1 for references. Sequence alignment was performed using CLUSTAL X. Amino acid identity is indicated by dots (all three sequences share same residue) and lines (two of three residues share same residue).


Ord^Dot MPTYALLGATGATGSAILRCLLASPP-PDLDLNILVRSKQKLLKSFPTLTTTIS------P
StcO (A. nidulans) MPSYALLGATGATGSSVLRHLLYSGSSSDLTVNVLVRSKSKLLAAFPSLDKPRPSVTSSIP
AvfA (A. parasiticus) MVTYALLGATGATGSSILRHLLQKSPDS-LHIQVLVRSKVKLLQAFPDLETTRR------P
AvfA (A. flavus) MVTYALLGATGATGSSILRHLLHESPDS-LRIQILVRSKVKLLQAFPNLQTTRN------P
AvfA (A.oryzae) MVTYALLGATGATGSSILRHLLHDPPDS-LRIQILVRSKIKLLQAFPDLQTRRN------P
Ord^Dot RIHIIQGNSTDTIALQQCLEDASVAFMCVADNASNKGVSLTADTVTAIVTTLGMLRKLHGS
StcO (A. nidulans) TIRIFEGDSTNPDVLCAVLQDASLVFMCVAQNGSPMGTTLVQNTAAALIEARRRQAQPRGE
AvfA (A. parasiticus) QVHVIQGMSTDSDALSECLRNASIVFMCVAQNGSPIGTTLCQDSARPIISVLQQQQQSEGA
AvfA (A. flavus) QVHVIQGTSTDPDALSECLRNASIVFMCVAQNGSPIGTTLCQDSARAIISVLQQQQQSQGA
AvfA (A.oryzae) QVHVIQGTSTDADALSECLRNATIAFMCVAQNGSPIGTTLCQDSARAIISALQQQQQSEGA
Ord^Dot AYNAPTILQLRSASLNPKLSCQVPRLVYNIVSFCLHYSHLDIVKACEHYEAAAAKGLLSYI
StcO (A. nidulans) L----TVIQLRSASLNPVLAVQVPRFVHRVVCFCLAAGYADLRRACVLYEAAATEGLLQYV
AvfA (A. parasiticus) SYQPCTIVQLRSASLNPALAAQVPAFVHRIVSFCLFANYADIKQACQYYSEAQKQGTLEYI
AvfA (A. flavus) SYQPCTIVQLRSASLNPALAAQVPAFVHRIVSFCLFANYADIKKACQYYSQAQKQGILEYI
AvfA (A.oryzae) SYQPCTIVQLRSASLNPALAVQVPVFVHPIVSFCLFANYADIKQACQYYSEARKQGTLEYI
Ord^Dot YVDPPTIHDAFGPNRTGHKLISC---KPDVCDKQETALSYADLGAGFVEIASRKEDFLNQP
StcO (A. nidulans) LVDPPTLHDARGTQTTGYRLIDTTDMKDKENQRQAICLSYADLGVAMCEIASRADELHGQG
AvfA (A. parasiticus) LVDPPTLHDANGTHPTGYRLIST--------EPQATALSYADLGAAMCEIAHRESEFHGRA
AvfA (A. flavus) LVDPPTLHDANGTQPTGYRLIST--------ESQATALSYADLGAAMCEIAHRQSEFHGRA
AvfA (A.oryzae) FVDPPTLHDANGTQSTGYRLIST--------ESQATALSYADLGAAMCEIAQRRSEFHGRA
Ord^Dot VGVTATGKAKETWGVLAGFLFDGAKGRARAWWEEERPMSKPQNLFLYC--------GM
StcO (A. nidulans) VGVTATGPVRQTWAVLAGFLLEGGLGHLDYRYGRENVVVLGVCILLL-LGGLLYSIKA
AvfA (A. parasiticus) VGVTATGRVRQTWGVLLRHLLEGGSSRLRETIAKEAVVVRVLCIFLVILACLMSSL
AvfA (A. flavus) VGVTATGRVRQAWGVLLRHLLEGGSARLREKIAQETVVDGVVFAFLVVLAYLM
AvfA (A.oryzae) VGVTATGSVHQTWGVLLRHLLEGGSARLREKIAKNTVVVGIVCSFLVALAYLM

3.5 *EPOX*^{DOT} IS A PUTATIVE EPOXIDE HYDROLASE GENE

The predicted polypeptide encoded by the *epox*^{dot} (epoxide- dothistromin) gene displays significant sequence identity to epoxide hydrolase enzymes. The deduced amino acid sequence of *epox*^{dot} displays; 44% amino acid identity (59% similarity) to that of the *Aspergillus niger EH* (epoxide hydrolase) gene and 44% amino acid identity (60% similarity) to the *A. nidulans* hypothetical protein AN3017.2. Searching the *A. nidulans* genome revealed that the hypothetical protein AN3017.2 is listed as an epoxide hydrolase. Functions and references for these genes are shown in Table 3.1.

The *epox*^{dot} ORF is 1314 bp in length (1263 bp without intron sequence, section 3.5.1), encoding a deduced polypeptide product of 438 amino acid residues. The translation initiation codon was estimated at position 1690 as it was consistent with the initiation codons of *A. niger* and *A. nidulans*. The termination codon (TAG) was estimated at nucleotide position 3004 making the coding sequence approximately the same size as the epoxide hydrolase of *A. nidulans*. The nucleotide sequence and deduced amino acid sequence for *epox*^{dot} is shown in Figure 3.8.

3.5.1 The Putative *epox*^{dot} Gene Contains One Intron

Putative intervening regions were identified by the presence of 5' [GT(A/G/T)NGTY] and 3' (YAG) fungal consensus intron splice site sequences (Ballance, 1986; Bruchez and Erberle, 1993), as well as continuance of amino acid similarity with other sequences. One putative intron was identified, 54 nt in length starting at nucleotide position 2795 relative to the beginning of pR162A. This intron position was confirmed by direct sequencing of an RT-PCR product which was amplified from *D. pini* RNA using primers OT PR162A Rev4 and epoxide 3' end (courtesy of Arne Schwelm). The position of this intron allows preservation of the Epox^{Dot} sequence with the epoxide hydrolases of *A. nidulans* and *A. niger*. The presence of only one intron does differ from *A. niger* epoxide hydrolase which has 9 exons and 8 introns which is an unusually high number for an *A. niger* gene (Arand, Hemmer *et al,* 1999). One of the introns (no. 8) of *A. niger* is consistent with the intron

found in *epox*^{dot}. It seems unusual that other introns were not found within *epox*^{dot} however cDNA has confirmed the absence of four of the introns found within *A. niger*.

3.5.2 Amino Acid Sequence Comparison

The deduced amino acid sequence for *epox*^{dot} was aligned to epoxide hydrolase (EH) peptides from *A. nidulans* and *A. niger* (Figure 3.9). This identified high identity between gene products through most of the sequence with the similarity reducing towards the end of the *epox*^{dot} gene.

Figure 3.8 Nucleotide and Deduced Amino Acid Sequence of the *Dothistroma pini* Putative Gene *epox*^{dot}

The coding sequence, promoter and terminator regions and the deduced amino acid sequence of *epox*^{dot} is shown. The nucleotide sequence starts at position 1508 relative to the start of pR162A sequence. Initiation and terminator codons are indicated in green and introns are indicated in blue. The positions of primers used in RT-PCR are indicated. Possible AFLR and TATA binding motifs are indicated through underlining.

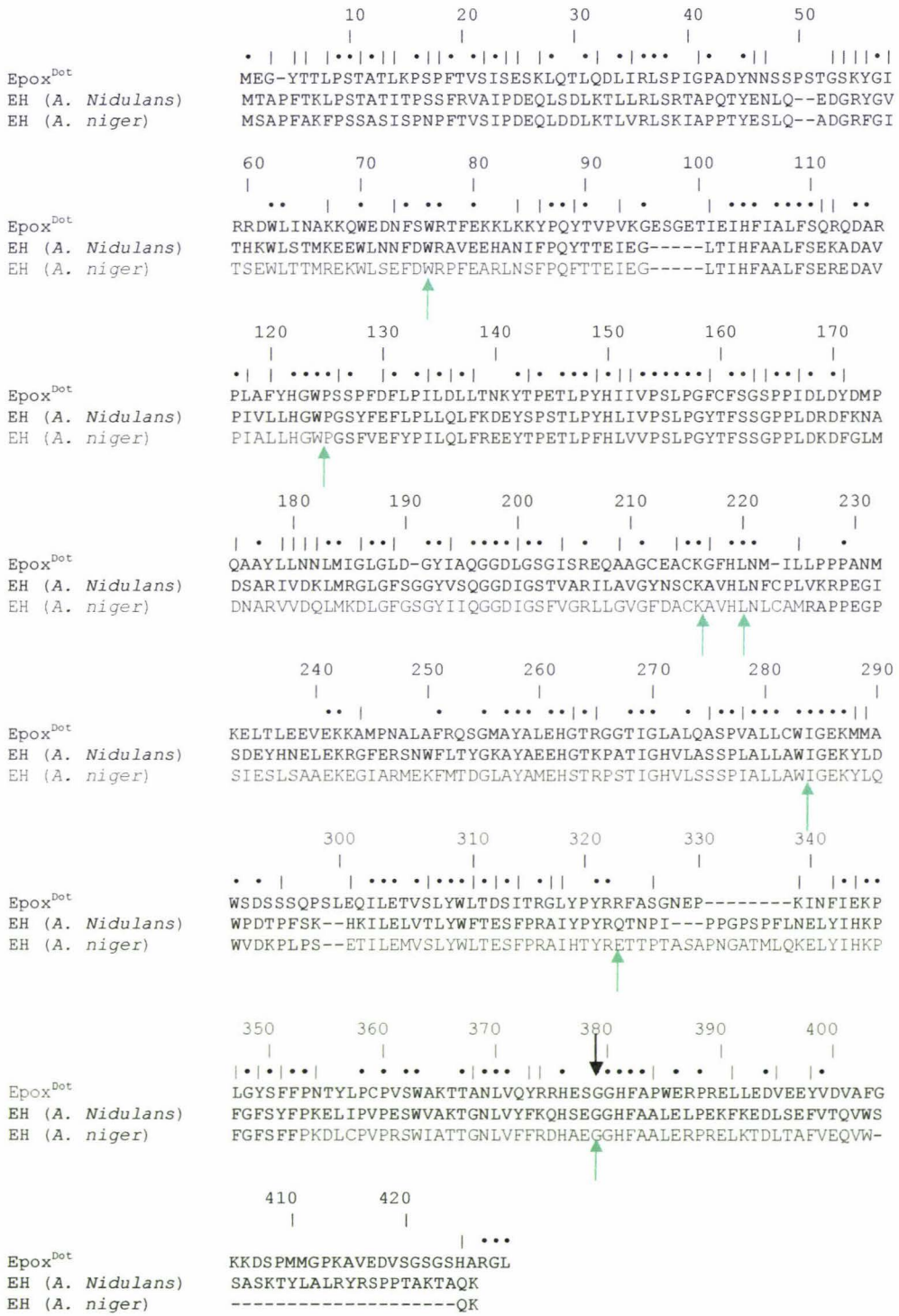
```
TGCCAACATCGGGCGACGAGTGCAGGCAGTCGGTGGTCGAAGCATCACAGCCTTCCGAAACCCGAGCACTTTAGCCCG
AGCCATCTAGTTCCGAGTTGCACTAGCCGAATGAGGAAGGATATGGATTCCTATGGGCAGACGTCAGCACATCACTGC
TGCCACGATTCAGTCGTAGACCTCGTCGCCGCTGTAAAGCTGCCCATACAACACCATATTGTCACAGCCATCGTCTTC
TTGGCACCAACACAGAATACTTATCCCTCAACTAAATCAACCTCATCACCACCAACACCACAATCGATATGGAAGGCT
                                                                  M  E  G  Y

        20        30        40        50        60        70        80
        |         |         |         |         |         |         |
ATACCACCCTCCCCTCCACCGCGACTCTCAAGCCCAGTCCCTTCACCGTCTCAATTTCCGAATCCAAACTCCAAACCC
  T  T  L  P  S  S  T  A  T  L  K  P  S  P  F  T  V  S  I  S  E  S  K  L  Q  T  L

90        100       110       120       130       140       150       160
|         |         |         |         |         |         |         |
TCCAAGACCTGATCCGCCTCTCCCCCATCGGACCCGCAGACTACAACAATTCCTCCCCTTCCACTGGCTCAAAGTACG
  Q  D  L  I  R  L  S  P  I  G  P  A  D  Y  N  N  S  S  P  S  T  G  S  K  Y  G

  170       180       190       200       210       220       230       240
   |         |         |         |         |         |         |         |
GTATCCGCCGTGACTGGCTCATCAATGCAAAGAAACAATGGGAAGACAACTTCTCCTGGCGCACTTTCGAGAAGAAGT
  I  R  R  D  W  L  I  N  A  K  K  Q  W  E  D  N  F  S  W  R  T  F  E  K  K  L

    250       260       270       280       290       300       310       320
     |         |         |         |         |         |         |         |
TGAAGAAGTACCCGCAGTATACCGTCCCTGTGAAGGGAGAGAGTGGGGAGACGATCGAAATTCACTTCATCGCGCTGT
  K  K  Y  P  Q  Y  T  V  P  V  K  G  E  S  G  E  T  I  E  I  H  F  I  A  L  F

      330       340       350       360       370       380       390       400
       |         |         |         |         |         |         |         |
TCAGTCAAAGGCAAGACGCAAGGCCGTTGGCGTTTTACCATGGATGGCCCTCTTCCCCGTTCGATTTTCTGCCGATTC
  S  Q  R  Q  D  A  R  P  L  A  F  Y  H  G  W  P  S  S  P  F  D  F  L  P  I  L

        410       420       430       440       450       460       470
        |         |         |         |         |         |         |
TGGATCTTTTGACAAACAAATACACGCCCGAGACACTGCCGTATCACATCATCGTCCCGAGTCTGCCGGGTTTCTGCT
  D  L  L  T  N  K  Y  T  P  E  T  L  P  Y  H  I  I  V  P  S  L  P  G  F  C  F

480       490       500       510       520       530       540       550
|         |         |         |         |         |         |         |
TCAGCGGCTCCCCACCCATAGATTTGGATTACGACATGCCCCAAGCAGCCTACCTCCTCAACAACCTCATGATCGGCC
  S  G  S  P  P  I  D  L  D  Y  D  M  P  Q  A  A  Y  L  L  N  N  L  M  I  G  L

  560       570       580       590       600       610       620       630
   |         |         |         |         |         |         |         |
TTGGATTGGATGGGTACATTGCCCAAGGCGGCGATCTGGGATCGGGGATCTCGAGGGAGCAGGCCGCGGGTTGTGAAG
  G  L  D  G  Y  I  A  Q  G  G  D  L  G  S  G  I  S  R  E  Q  A  A  G  C  E  A
   OT R162A Rev4 →

    640       650       660       670       680       690       700       710
     |         |         |         |         |         |         |         |
CTTGCAAGGGATTTCATCTTAATATGATTTTGCTGCCGCCGCCGGCGAACATGAAAGAGTTGACGTTGGAGGAGGTGG
  C  K  G  F  H  L  N  M  I  L  L  P  P  P  A  N  M  K  E  L  T  L  E  E  V  E

      720       730       740       750       760       770       780       790
       |         |         |         |         |         |         |         |
AGAAGAAGGCTATGCCCAACGCCCTGGCGTTCCGTCAAAGTGGCATGGCCTATGCCCTGGAGCACGGAACCCGCGGCG
  K  K  A  M  P  N  A  L  A  F  R  Q  S  G  M  A  Y  A  L  E  H  G  T  R  G  G

        800       810       820       830       840       850       860
        |         |         |         |         |         |         |
GCACGATTGGATTGGCATTGCAGGCGAGTCCCGTGGCGCTGTTATGTTGGATCGGGGAGAAGATGATGGCCTGGAGTG
  T  I  G  L  A  L  Q  A  S  P  V  A  L  L  C  W  I  G  E  K  M  M  A  W  S  D
```

```
870        880        890        900        910        920        930        940
 |          |          |          |          |          |          |          |
ATTCCTCTTCTCAGCCGTCTCTCGAACAAATCTTGGAGACCGTCTCCCTATACTGGCTCACCGATTCCATCACGCGGG
  S  S  S  Q  P  S  L  E  Q  I  L  E  T  V  S  L  Y  W  L  T  D  S  I  T  R  G

  950        960        970        980        990        1000       1010       1020
   |          |          |          |          |          |          |          |
GTCTTTACCCCTACCGTCGCTTCGCTTCTGGTAACGAACCCAAAATTAATTTCATCGAGAAGCCGCTGGGGTATTCGT
  L  Y  P  Y  R  R  F  A  S  G  N  E  P  K  I  N  F  I  E  K  P  L  G  Y  S  F

    1030       1040       1050       1060       1070       1080       1090       1100
     |          |          |          |          |          |          |          |
TTTTCCCGAATACGTACCTCCCTTGCCCTGTGAGCTGGGCGAAGACGACGGCGAATTTGGTCCAGTACCGCAGACATG
  F  P  N  T  Y  L  P  C  P  V  S  W  A  K  T  T  A  N  L  V  Q  Y  R  R  H  E

      1110       1120       1130       1140       1150       1160       1170       1180
       |          |          |          |          |          |          |          |
AGAGTAAGTCCAGTTGAATGTTGGTTTGAGTGTACTTGCCGACTGACGTGTGTAGGTGGAGGTCATTTCGCGCCTTGG
  S                                                          G  G  H  F  A  P  W
                                                                        Epoxide 3' end ←

        1190       1200       1210       1220       1230       1240       1250
         |          |          |          |          |          |          |
GAGAGGCCAAGGGAGTTGTTGGAAGATGTGGAGGAGTATGTTGATGTGGCGTTCGGGAAGAAGGATTCCCCGATGATG
E  R  P  R  E  L  L  E  D  V  E  E  Y  V  D  V  A  F  G  K  K  D  S  P  M  M

1260       1270       1280       1290       1300       1310
 |          |          |          |          |          |
GGTCCGAAAGCTGTGGAAGATGTCAGCGGGAGTGGAAGCCATGCAAGAGGGTTGTAGATTGCGGAGGCTCAGCTCTTG
G  P  K  A  V  E  D  V  S  G  S  G  S  H  A  R  G  L

ATTGATGGTACATACCATGGGGCGTTTCATCCCGGCTTTACGAAAAAGAGCTGAAACTCCGAGAGTCATCTTGAAAAT
TCGAAGTTTTGACATGCGACTACATCAAGACTCGAACAAGCCTTTTTAACGCTGCCAATATGGTTTCTGTCGCACTCC
TATCTGTTCATCGTTCCCGTATACTGCACCAAGACTACCGCGGCCAAACTCACCATAACTGTCGTAAACACCGTCAGC
GCACTGCTGCCGTTCAATCGTCGAACGAGGAGACTACTACATACCGCAATAAAGAAACAAATTTTGCGGCTTGCTCAT
TGGCCTCTCCTCTTCCCACCAAGCTCTCGCCCTCCCCTTCGCCCCATCGAACAAGAATCCCGCCAGCACTCCCCACGT
CTCCTTGGCCTTCCCC
```

Figure 3.9 Alignment of the Deduced Amino Acid Sequence of *epox*^{dot} to Other Epoxide Hydrolases

The deduced amino acid sequence of the putative *D. pini* gene (Epox^{dot}) was aligned with those of the *A. niger epoxide hydrolase* gene (EH) and *A. nidulans* hypothetical protein AN3017.2 (epoxide hydrolase-EH). Sequence alignment was performed using CLUSTAL X alignment. Amino acid identity is indicated by dots (all three sequences share same residue) and lines (two of three residues share same residue).

The position of the putative intron found in Epox^{dot} is indicated by a black arrow and the positions of the intron sequences of *A. niger* (as designated by Arand, Hemmer *et al.* 1999) are indicated by green arrows (The first intron of *A. niger* is not shown).


Epox^Dot
EH (A. Nidulans)
EH (A. niger)
MEG-YTTLPSTATLKPSPFTVSISESKLQTLQDLIRLSPIGPADYNNSSPSTGSKYGI
MTAPFTKLPSTATITPSSFRVAIPDEQLSDLKTLLRLSRTAPQTYENLQ--EDGRYGV
MSAPFAKFPSSASISPNPFTVSIPDEQLDDLKTLVRLSKIAPPTYESLQ--ADGRFGI
RRDWLINAKKQWEDNFSWRTFEKKLKKYPQYTVPVKGESGETIEIHFIALFSQRQDAR
THKWLSTMKEEWLNNFDWRAVEEHANIFPQYTTEIEG-----LTIHFAALFSEKADAV
TSEWLTTMREKWLSEFDWRPFEARLNSFPQFTTEIEG-----LTIHFAALFSEREDAV
PLAFYHGWPSSPFDFLPILDLLTNKYTPETLPYHIIVPSLPGFCFSGSPPIDLDYDMP
PIVLLHGWPGSYFEFLPLLQLFKDEYSPSTLPYHLIVPSLPGYTFSSGPPLDRDFKNA
PIALLHGWPGSFVEFYPILQLFREEYTPETLPFHLVVPSLPGYTFSSGPPLDKDFGLM
QAAYLLNNLMIGLGLD-GYIAQGGDLGSGISREQAAGCEACKGFHLNM-ILLPPPANM
DSARIVDKLMRGLGFSGGYVSQGGDIGSTVARILAVGYNSCKAVHLNFCPLVKRPEGI
DNARVVDQLMKDLGFGSGYIIQGGDIGSFVGRLLGVGFDACKAVHLNLCAMRAPPEGP
KELTLEEVEKKAMPNALAFRQSGMAYALEHGTRGGTIGLALQASPVALLCWIGEKMMA
SDEYHNELEKRGFERSNWFLTYGKAYAEEHGTKPATIGHVLASSPLALLAWIGEKYLD
SIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGHVLSSSPIALLAWIGEKYLQ
WSDSSSQPSLEQILETVSLYWLTDSITRGLYPYRRFASGNEP--------KINFIEKP
WPDTPFSK--HKILELVTLYWFTESFPRAIYPYRQTNPI---PPGPSPFLNELYIHKP
WVDKPLPS--ETILEMVSLYWLTESFPRAIHTYRETTPTASAPNGATMLQKELYIHKP
LGYSFFPNTYLPCPVSWAKTTANLVQYRRHESGGHFAPWERPRELLEDVEEYVDVAFG
FGFSYFPKELIPVPESWVAKTGNLVYFKQHSEGGHFAALELPEKFKEDLSEFVTQVWS
FGFSFFPKDLCPVPRSWIATTGNLVFFRDHAEGGHFAALERPRELKTDLTAFVEQVW-
KKDSPMMGPKAVEDVSGSGSHARGL
SASKTYLALRYRSPPTAKTAQK
----------QK

3.6 UNTRANSLATED REGIONS OF THE THREE PUTATIVE DOTHISTROMIN GENES

3.6.1 5' Promoter Analysis

The 5' untranslated regions of the three putative genes described in the previous sections were searched for TATA and CAAT motifs which are thought to be important transcription initiation sites in filamentous fungi (Hamer and Timberlake, 1987). A TATA motif was found in *mox*^{dot} 168 bp upstream of the of the proposed translational start site. The closest putative TATA motif to the proposed translational start site of *epox*^{dot} was 948 bp upstream and the closet to the proposed translational start site of *ord*^{dot} was 674 bp upstream (within the pR181 subclone) which are further away from the translational start site than usual. No CAAT motifs were identified.

AflR is involved in the regulation of ST/AF biosynthesis. It is a GAL4-like transcription factor, and is required for ST/AF production (Chang, Bhatnagar *et al,* 1993). Potential binding sites for this AflR regulatory protein ($TCGN_5CGA$) (Fermandes *et al,* 1998) are evident in both *epox*^{dot} and *mox*^{dot} at 295 bp and 267 bp upstream of the predicted translational start sites respectively. The putative biosynthetic regulatory motif found in other putative dothistromin genes ($TCGN_{11}CGA$) (Astin, 2000) was present 432 bp upstream of the *ord*^{dot} proposed translational start site in the subclone pR181. The sequence of this possible regulatory protein binding site is TCGGACATGGCCCACGA.

3.6.2 3' Untranslated Region

One polyadenylation signal sequence (AATAAA; Proudfoot and Brownlee, 1976) was observed in the 3' untranslated region of the *epox*^{dot} gene, which is located 308 bp from the termination codon and has been tentatively assigned. The *ord*^{dot} gene did not have any identifiable polyadenylation signal sequences and as the *mox*^{dot} gene continues into another subclone (pR161), none were identified in the sequence available from this study.

3.7 DISCUSSION

3.7.1 AF/ST Like Genes Present in pR162A

The three putative dothistromin genes (*ord*^{dot}, *mox*^{dot} and *epox*^{dot}) show significant similarities to genes involved in ST/AF biosynthesis.

3.7.1.1 mox^{dot}

The putative *mox*^{dot} gene displays high amino acid identity to the *moxY* and *stcW* genes that are required for ST/AF production (Table 3.1, Section 3.4). The *moxY* and *stcW* genes each encode monooxygenases (probably flavin-requiring) (Keller *et al,* 2000). The functions of *moxY* and *stcW* are yet to be assigned as no conclusive results have been obtained from experiments with *A. parasiticus* and *A. nidulans* although some suggestions have been made through gene knockouts. A knockout of *stcW,* (Keller, Watanabe *et al,* 2000) makes mutants unable to produce ST but able to accumulate averufin. The same results were obtained with a *stcB* (P450 monooxygenase) inactivation mutant. These results suggest a possible role of *stcW* and *stcB* in a dimer or in conjunction with each other in the steps following the averufin intermediate of the AF/ST pathway. However a clear role in biosynthesis is yet to be demonstrated (see Figure 1.2 and Figure 3.10) (Keller, Watanabe *et al*, 2000). *Mox*^{dot} also displays high amino acid similarity to other monooxygenases involved in sugar utilisation in *A. oryzae.* Due to the high amino acid identity displayed to monooxygenases present in *Aspergillus* species it is likely that *mox*^{dot} also functions as a monooxygenase in dothstromin biosynthesis.

Like *stcW, mox*^{dot} contains three introns, none of which are found in *A. parasiticus* (Yu, Chang *et al,* 2000). As the entire sequence of *mox*^{dot} has not been completed within this project it is unknown whether the *mox*^{dot} gene is of similar length and whether it encodes a similar sized protein to those of *A. parasiticus* and *A. nidulans.* The conversion of AVF to VHA is thought to be a two step enzymatic process in *A. nidulans*, possibly requiring both *stcW* and *stcB* (Keller, Watanabe *et al,* 2000), homologs of which have been found in *A.*

parasiticus (*moxY* and *cypX* respectively). As two enzymes may be required to work together in *A. nidulans* and *A. parasiticus* it is also possible that the putative *mox*^{dot} gene could require another gene to be functional in dothistromin biosynthesis. It is possible that, in *D. pini*, *cyp*^{dot} which shows high amino acid identity to both *cypX* (59% identity, 68% similarity) and *stcB* (56% identity, 68% similarity) (Seconi, 2001), works in conjunction with *mox*^{dot}. To identify the function of *mox*^{dot} gene inactivation will need to be carried out. This will also identify the step in the dothistromin pathway that this putative monooxygenase is required for.

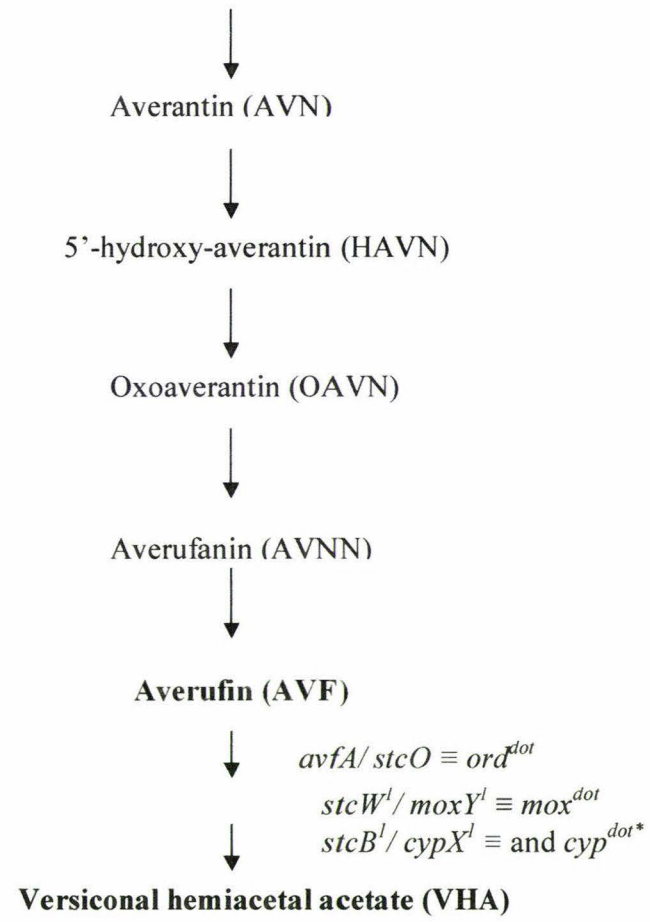
3.7.1.2 ord^{dot}

The putative *ord*^{dot} gene also displays high amino acid identity to genes required for ST/AF production (Table 3.1). *avfA* and *stcO* found in *A. parasiticus* and *A. nidulans* respectively show high amino acid identity to *ord*^{dot}. *avfA* and *stcO* act as oxidases in the AF/ST pathway. They are specifically involved in the conversion of averufin (AVF) to versiconal hemiacetal acetate (VHA) (Figure 3.10). Gene complementation experiments using AVF-accumulating mutant strains have demonstrated that *avfA* encodes an oxidase which is necessary for the conversion of AVF to VHA which is thought to be a two step process with 1-hydroxyversicolorone as a stable intermediate (Yabe, Chihaya *et al*, 2003).

Ord^{dot} encodes a 288 amino acid polypeptide product which is of similar length to the *avfA* and *stcO* gene products of *A. parasiticus* and *A. nidulans* (285 aa and 290 aa respectively). *ord*^{dot} also appears to contain no introns which is consistent with the *avfA* and *stcO* genes. Due to the similarity of *ord*^{dot} to *avfA* and *stcO* it seems likely that *ord*^{dot} could also act as an oxidase in dothistromin biosynthesis, however this needs to be confirmed through gene inactivation studies. Unlike *avfA*, *ord*^{dot} does appear to have a possible AflR binding site approximately 432 bp upstream of the predicted translational start site. It is possible that *avfA* is transcribed together with *omtB* (another gene in the aflatoxin gene cluster) and processed afterwards into two mRNA species (Yu, Woloshuk *et al.* 2000). It appears that like the sterigmatocystin gene cluster, the order of genes in the dothistromin gene cluster is

Figure 3.10 Proposed Conversion of Averantin to Versiconal Hemiacetal in AF-ST Pathway

Figure 3.10 shows the possible sites of action of the ord^{dot}, mox^{dot} and cyp^{dot} gene products and their putative ST/AF homologs. [1] indicates the functions of genes are unassigned.



*Sequenced by Seconi, 2001

Figure modified from Keller, Watanabe *et al.* 2000 and Yu, Chang *et al.* 2004.

likely to differ from those in the aflatoxin gene cluster. However the genes found within in the cluster may encode proteins with similar functions to those encoded by genes in the AF and ST gene clusters.

3.7.1.3 epox^{dot}

Most epoxide hydrolases (EHs) are members of the α/β hydrolase fold family. Amino acid similarity between the different members of this family is usually very low and mostly restricted to the α/β hydrolase fold domain. This family of enzymes hydrolyses their substrates in a two-step mechanism. The two step hydrolytic cleavage starts with the attack of an aspartate group, reacting as a nucleophile to yield a covalent glycol-monoester intermediate. In the second catalytic step, a water molecule finally hydrolyses the monoester intermediate to the *vic*-glycol product. See Figure 3.11 for a basic epoxide hydrolase reaction. The α/β hydrolase-fold enzymes hydrolyse their substrates by the action of a catalytic triad in the two-step mechanism (reviewed in Arand, Hemmer *et al,* 1999). The residues within the catalytic triad vary between different species. Most recent work has been focused on EH from microbial and mammalian sources. Mammalian EHs appear to be optimised for a rapid first step in the enzymatic hydrolysis which suits their function as detoxifying enzymes, since the genotoxic agent is rapidly inactivated by trapping it in the form of the ester intermediate (Arand, Hemmer *et al,* 1999). In contrast to these mammalian EHs the *A. niger* EH seems to be optimised for rapid product formation, through the presence of a different catalytic triad which suggests a catabolic role in the fungus (Arand, Hemmer *et al,* 1999).

Figure 3.11 Basic Epoxide Hydrolase Reaction

The putative *epox*^{dot} gene displays significant amino acid identity to the *A. niger* EH (See Table 3.1), indicating that perhaps *epox*^{dot} has a catabolic role in *Dothistroma pini.* Studies of epoxide hydrolases in fungi have identified epoxide activity associated with both the cell debris (microsomal mEH) and soluble (sEH) fractions demonstrating the occurrence of two distinct EH activities with different substrate specificity (Moussou, Archelas *et al,* 1999). The EH found within *A. niger* although having a sequence identity of approximately 25 % to mammalian mEHs, does not have a N- terminal anchor and so is actually found in the soluble fraction of *A. niger* lysates, this is consistent with the EH activity of *Alternaria alternata* and *Fusarium solani pisi* (Pinot, Cladas *et al,* 1997 and references within; Arand, Hemmer *et al.* 1999). Like mammalian sEH activity, the activity of *A. alternata* EH can be enhanced by clofibrate (drug known to induce EH in mammals). Studies in the fungus *A. alternata* indicate a relationship between AAL toxin production and expression of EH activity. Studies following the kinetics of growth, EH activity and toxin production in *A. alternata* firstly show that pH and medium composition markedly influence AAL production and EH activity and, secondly, that EH activity is coincident with AAL toxin production. On both glucose (day 6) and pectin (day 12) AAL production occurred in parallel with EH activity (peaking coincidently with the second peak of EH activity on pectin and the only peak of EH activity seen on glucose) (Moriseau, Ward *et al,* 1999). This observation was also supported by a non-toxin producer of *A. alternata* which had very low EH activity which remained constant over the entire sampling period (Moriseau, Ward *et al,* 1999). Another study in *Alternaria alternata* f. sp. *Lycopersici* using trans-diphenylpropan oxide (tDPPO) as a substrate showed EH activity was mainly located in the soluble fraction and that an increase of toxin production between days 3 and 9 in an AAL liquid culture was concomitant with a period of high EH activity, whilst remaining constant in a non-toxin producing species. This suggests once again, the possibility for a relationship between epoxide hydrolase activity and toxin biosynthesis. sEH may be important in the early stages of toxin biosynthesis as the activity of this enzyme increased before toxin production could be detected (Pinot, Cladas *et al,* 1997). The possible role of sEH activity in toxin biosynthesis suggest that the EH present in *D. pini* is also likely to be a sEH, like those present in *A. alternata* and *F. solani pisi.*

The products of mEH metabolism have been implicated as ultimate reactive carcinogens that are responsible for polycyclic aromatic hydrocarbon-initiated carcinogenesis, and also in the production of other xenobiotic compounds such as 1,3-butadiene, benzene and aflatoxin B_1. The implication of mEH in the production of reactive carcinogens and xenobiotic compounds also indicate that the putative *epox*^{dot} could have a role in dothistromin biosynthesis (Fretland and Omiecinski, 2000).

AFB_1 formed via the aflatoxin (AF) pathway generates a highly reactive epoxide (through a microsomal cytochrome P450 (CYP450) dependent epoxidation reaction) that is responsible for nucleic acid alkylation (AFB,-8,9-exo-epoxide (AFBO)). Chemical and enzymatic epoxidations of AFB_1 yield exo and endo AFB_1 epoxide stereoisomers, with the endo-epoxide being much more mutagenic. Detoxification of AFBO by a specific alpha class glutathione S-transferase (GST) is an important protective mechanism in mice and accounts for some resistance to the carcinogenic effects of aflatoxin (Eaten and Gallagher, 1994 and references within).

Involvement of an epoxidation reaction in aflatoxin carcinogenesis, suggests that there could be an epoxidation reaction forming a reactive epoxide intermediate in dothistromin synthesis and hence an epoxide hydrolase could be required to hydrolase the reactive intermediate. As an epoxide hydrolase gene has not been identified in the AF or ST gene clusters, [the epoxide hydrolase identified in *A. nidulans* (hypothetical protein AN30172) is found on a different chromosome to the ST cluster and hence not likely to be involved in ST biosynthesis] the putative *epox*^{dot} gene (if involved in dothistromin biosynthesis) may be involved in the dothistromin pathway after the formation of veriscolorin, as the AF/ST and putative dothistromin pathways appear to differ from this point on.

4.0 SOUTHERN BLOTTING ANALYSIS

4.1 INTRODUCTION

Two lambda clones, λCGV1 (Gillman, 1996) and λKSA (Morgan, 1997) were previously isolated from a library of *D. pini* genomic DNA. Each of these λ clones had been subcloned and most of these subcloned regions sequenced and analysed to identify putative genes of the dothistromin biosynthetic cluster. Identification of putative genes was enabled through BLAST searches and mainly revealed matches to genes present in the gene clusters of AF and ST synthesis. The genes identified in each of these lambda clones so far (in this study and previous work) are outlined in Figures 4.1 and 4.2. Another lambda clone, λCGV2 (Gillman, 1996) had also been identified but had only been partly characterised, however a 3 kb subcloned section of this λCGV2 clone which has been characterised also revealed matches to an *aflR* gene involved in AF and ST synthesis as well as to a melanin keto-reductase.

What is unknown at this stage is how these three gene clusters relate to one another. Comparing the putative genes identified so far to those of *A. nidulans* and *A. parasiticus/flavus* does suggest that these gene clusters could be linked on the genome, however what distance and in what orientation these clusters are linked is unknown. A comparison of the clustering of *D. pini* biosynthetic genes to those of *A. nidulans* and *A. parasiticus/ flavus* is shown in Figure 4.3. Long Range PCR was attempted to identify how the three clusters of *D. pini* genes are located with respect to each other. Primers were designed from each end of the λCGV1 and λKSA clones and from the characterised 3 kb subclone of λCGV2 and internal control primers were also designed within the λKSA clone to ensure the PCR reactions were working correctly. Unfortunately this method did not produce reliable results. Southern blotting was then implemented to attempt to identify the relative positions of the three lambda clones. Whilst carrying out Southern blotting to identify positioning of each of the lambda clones in relation to each other, it was also decided to design probes to confirm the genomic positioning of genes identified within the λKSA clone. This should also enable identification of areas within the λKSA in which sequencing is not accurate.

Figure 4.1 λCGV1 Clone

A schematic diagram (not to scale) showing the *D. pini* λCGV1 biosynthetic gene cluster. Arrows indicate putative dothistromin biosynthetic genes and the direction refers to the direction of synthesis. The putative *D. pini* gene cluster is compared to homologs in *A. parasiticus* and *A. nidulans*. Gaps indicate no homologs yet identified (Skory, Chang *et al.* 1992; Keller, Kantz *et al.* 1994; Brown, Yu *et al.* 1996; Bhatnagar, Yu *et al.* 2003; Yu and Lenonard, 1995; Feng and Leonard, 1995).

% identity of predicted amino acid sequences to AF/ST gene products

	ddhA	***dotA***	***dotB***	***dotC***	***dotD***
A. parasiticus aflatoxin		*verI* (80%)	-	*aflT* (31%)	*pksL1* (35%)
A. nidulans sterigmatocystin		*stcU* (79%)	*stcC* (24%)	-	*stcA* (38%)
putative activity		Keto - reductase	oxidase	toxin pump	Thio - esterase

Figure 4.2 λKSA Clone

A schematic diagram (not to scale) showing the *D. pini* λKSA biosynthetic gene cluster. Arrows indicate putative dothistromin biosynthetic genes and the direction refers to the direction of synthesis. The putative *D. pini* gene cluster is compared to AF/ST homologs in *A. parasiticus* and *A. nidulans*. *Epox*^{dot} has no known homologues (Chang, Cary *et al.*1995; Yu and Leonard, 1995; Yu, Chang *et al.* 2000; Brown, Yu *et al.* 1996; Yu, Woloshuk *et al.* 2000).

% identity of predicted amino acid sequences to AF/ST gene products

	***pks*^{dot}** (part of)	***cyp*^{dot}**	***ord*^{dot}**	***epox*^{dot}**	***mox*^{dot}**
A. parasiticus aflatoxin	*pksA* (62%)	*cypX* (59%)	*avfA* (51%)	-	*moxY* (70%)
A. nidulans sterigmatocystin	*stcA* (63%)	*stcB* (56%)	*stcO* (47%)	-	*stcW* (63%)
putative activity	polyketide synthase (KS, AT domains)	P450 mono-oxygenase	Oxido-reductase	Epoxide hydrolase	mono-oxygenase

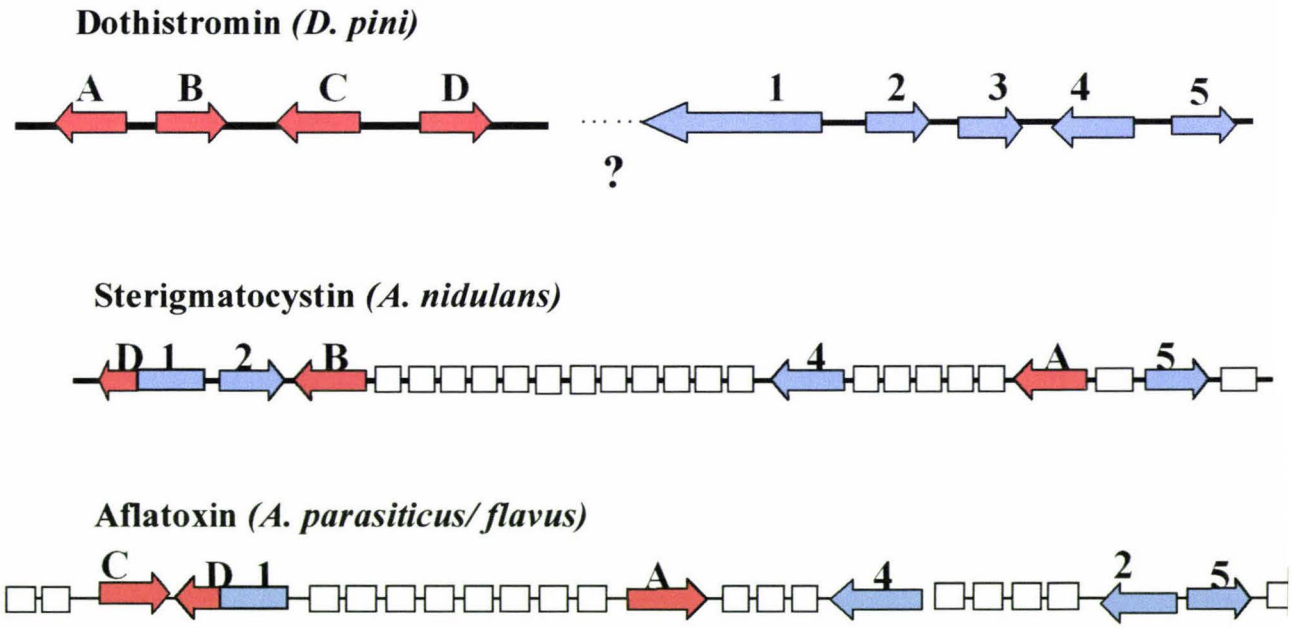
Figure 4.3 Comparisons of Gene Clusters

A schematic diagram (not to scale) comparing the *D. pini* cluster of genes to those found in *A. nidulans* and *A. parasiticus/flavus* (Yu, Chang *et al.* 2004).

Red arrows- genes present in λCGV1 (A-D; *dotA* - *dotD* respectively)

Blue arrows – genes present in λKSA (1- pks^{dot}; 2- cyp^{dot}; 3- ord^{dot}; 4 – $epox^{dot}$; 5 –mox^{dot})

□ indicates additional genes in AF/ST cluster



4.2 ISOLATION OF GENOMIC DNA

To enable Southern blotting to be carried out genomic DNA was isolated from *D. pini*. A number of methods were compared to optimise the amount and quality of genomic DNA isolated.

The first DNA extraction method used was modified from Al-Samarrai and Schmid, 1999 (A simple method for extraction of fungal genomic DNA; section 2.5.1). This method produced 19.3 µg of DNA per 100 mg of freeze dried mycelia used. Although this was sufficient genomic DNA the quality was very poor making digestion with restriction enzymes (section 2.8) almost impossible (incomplete digestions were produced) even after additional cleaning of the DNA using phenol/chloroform extraction (section 2.6.4). The next method to be trialed was the CTAB method (section 2.5.2). This method produced a highly variable yield of DNA (2 - 125 µg DNA/100 mg freeze dried mycelium) which although clean was not very practical for producing the quantities of genomic DNA required for Southern blotting. The final method to be trialed and successfully used was modified from Moller, Bahnweg *et al,* 1992 (Isolation of high molecular weight DNA from filamentous fungi; section 2.5.3). This method produced on average 186.2 µg DNA/100 mg of freeze dried mycelium (ranging from 135.8 µg to 236.6 µg of DNA). After an additional phenol/chloroform extraction (section 2.6.4) and ammonium acetate/ethanol precipitation the genomic DNA was able to be digested successfully with restriction enzymes (section 2.8).

4.3 PROBES FOR SOUTHERN BLOTTING

To identify the relative positions of the lambda clones two probes were used from each of the clones. From λCGV1 (Figure 4.4) fragments of the putative genes *dotD* and *ddhA* were used as probes and from λCGV2, the putative genes *aflR* and *ver-2* were used as probes. Several probes were used for the λKSA clone (Figure 4.5) to enable the positioning of this lambda clone in relation to λCGV1 and λCGV2 as well as probes to distinguish the order of putative genes within this λ clone. The probes include the following putative genes; 5' region of the pks^{dot} gene, mox^{dot}, ord^{dot}, $epox^{dot}$, cyp^{dot} and also R209 which is a 0.7 kb fragment at the end of the λKSA clone (containing part of the pks^{dot} gene). See table 2.6 for

PCR reactions used to isolate probes. The above gene-specific fragments were hybridised to Southern blots containing 5 µg *D. pini* genomic DNA digested separately with *Bam*HI, *Eco*RI, *Not*I, *Sal*I and *Xho*I (section 2.8). Four identical blots were produced and probed. DIG labeling (sections 2.15.2, 2.15.4, 2.15.5 and 2.15.6) was initially used to label the first four probes – mox^{dot}, cyp^{dot}, ord^{dot} and 5' *pks*. As these did not all work successfully, and due to the amount of background from non-specific binding, it was decided to use radioactive labeling (sections 2.15.3, 2.15.5 and 2.15.7) for the remaining probes and to repeat those which did not work.

The *D. pini* genomic Southern blots produced are displayed in Figure 4.6 (a-j), a summary of these results is also included in Table 4.1.

Figure 4.4 λCGV1 Restriction Map (not to scale)

The λCGV1 restriction map was computer generated and based on the complete nucleotide sequence available. Putative genes and direction of transcription are indicated by red arrows and probe positions are indicated by blue lines. All fragment sizes shown are in kb. λ restriction sites are shown in green.

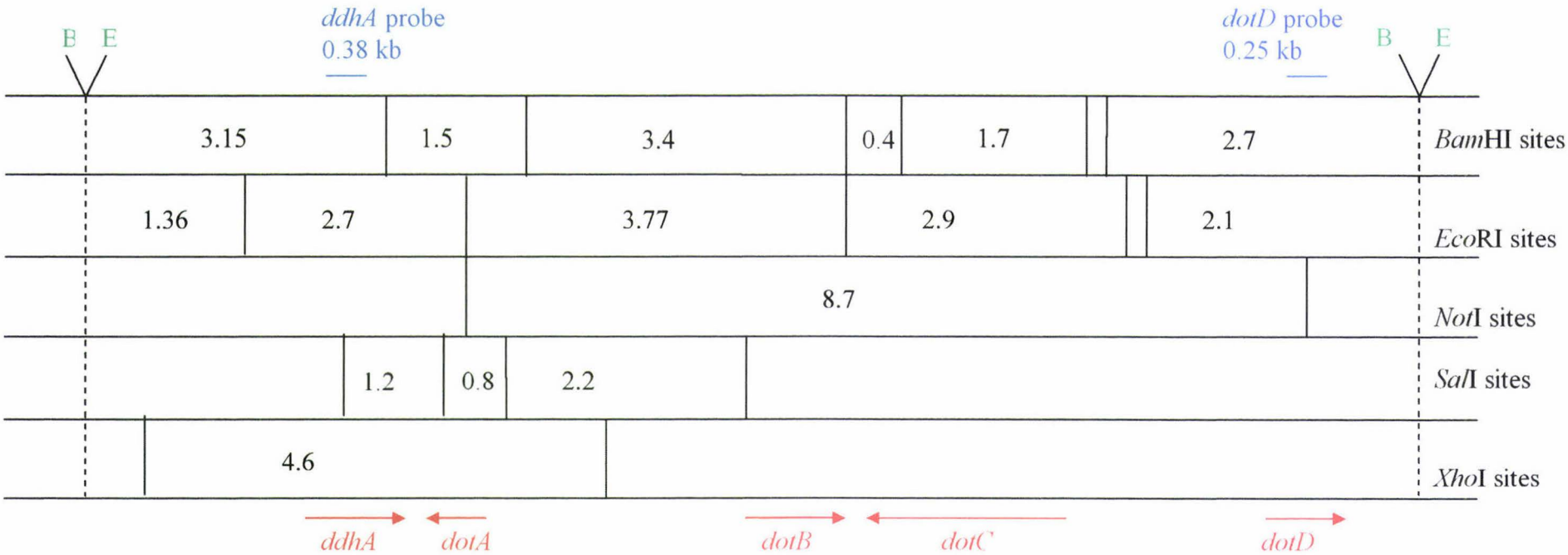

B E
ddhA probe
0.38 kb
dotD probe
0.25 kb
B E
3.15
1.5
3.4
0.4
1.7
2.7
BamHI sites
1.36
2.7
3.77
2.9
2.1
EcoRI sites
8.7
NotI sites
1.2
0.8
2.2
SalI sites
4.6
XhoI sites
ddhA
dotA
dotB
dotC
dotD

Figure 4.5 λKSA Restriction Map (not to scale)

The λKSA restriction map was computer generated and based on the complete nucleotide sequence available at the time of writing. Subclones (R161 etc) are indicated and probes used for Southern blotting are indicated by blue lines. All fragment sizes shown are in kb. λ restriction sites are shown in green. The asterisk (*) indicates the end of the original *Bam*HI/ *Eco*RI lambda arm. This has recently been extended from a subgenomic library clone (unpublished data). Putative genes and the direction of transcription are indicated by red arrows.

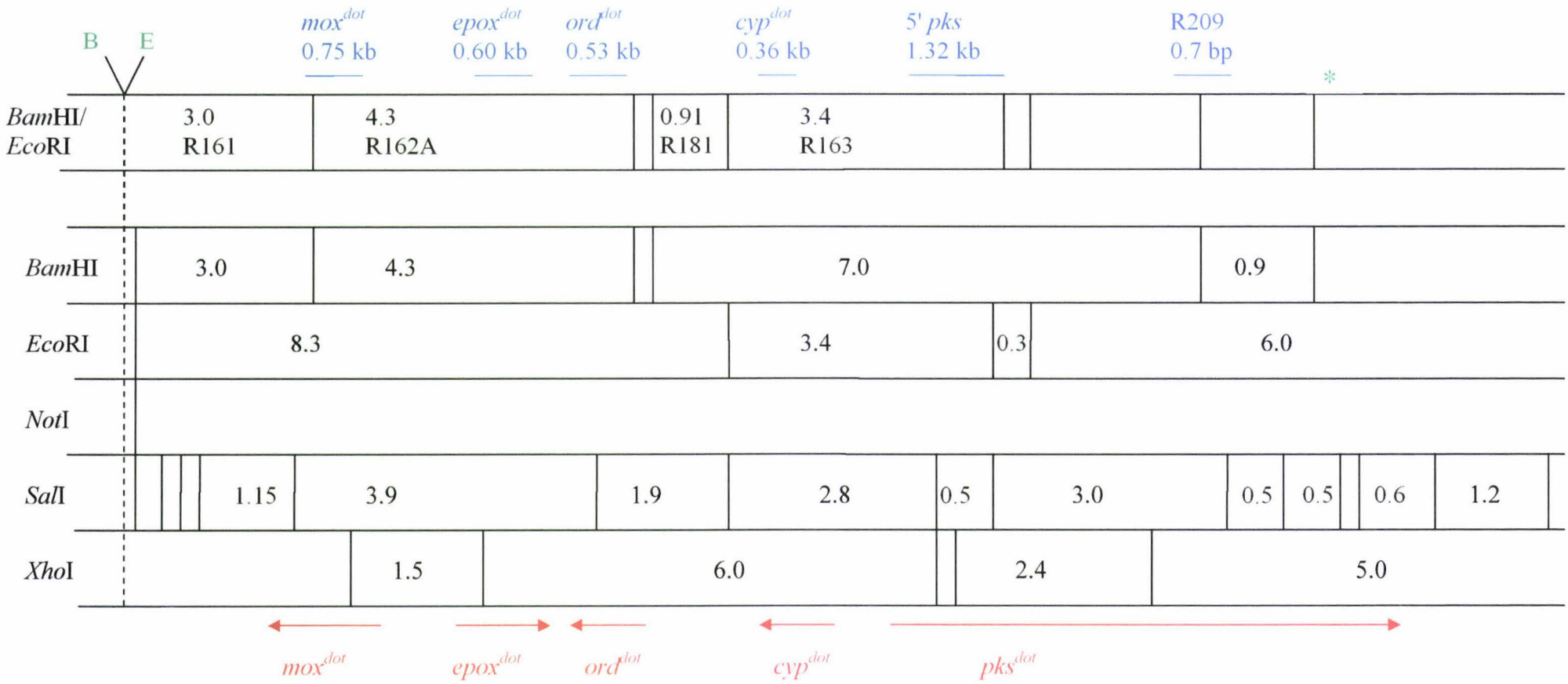
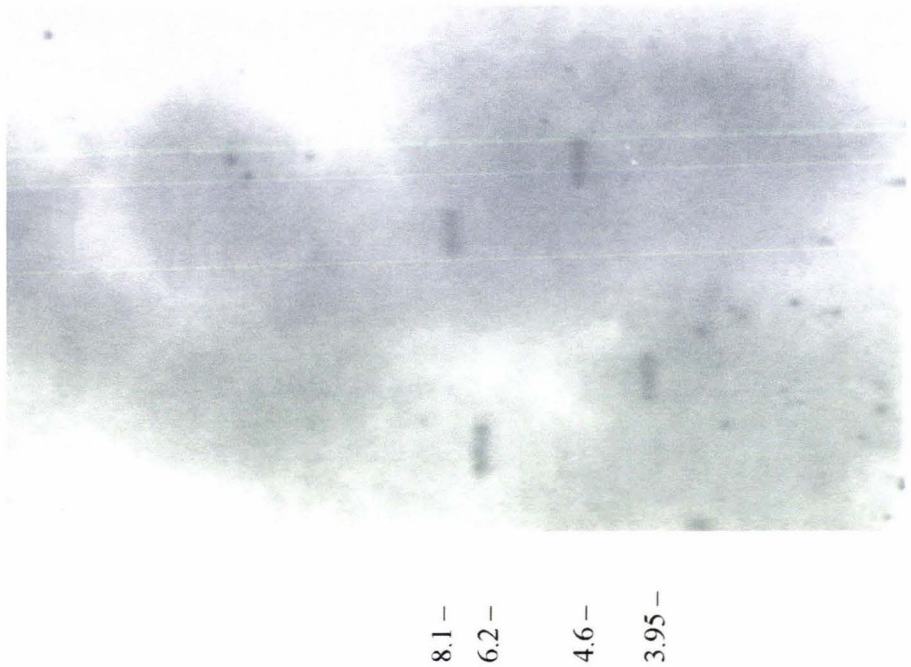

B
E
mox dot
0.75 kb
epox dot
0.60 kb
ord dot
0.53 kb
cyp dot
0.36 kb
5' pks
1.32 kb
R209
0.7 bp
*
BamHI/
EcoRI
3.0
R161
4.3
R162A
0.91
R181
3.4
R163
BamHI
3.0
4.3
7.0
0.9
EcoRI
8.3
3.4
0.3
6.0
NotI
SalI
1.15
3.9
1.9
2.8
0.5
3.0
0.5
0.5
0.6
1.2
XhoI
1.5
6.0
2.4
5.0
mox dot
epox dot
ord dot
cyp dot
pks dot

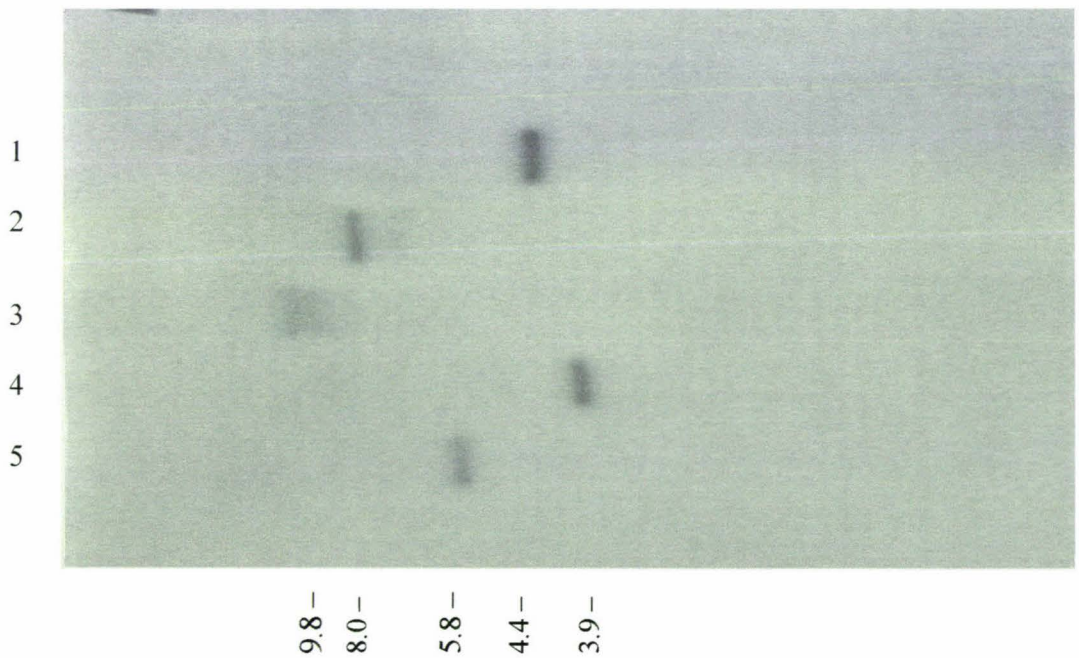
Figure 4.6 *Dothistroma pini* Genomic Southern Blots

Each blot contained 5 µg *D. pini* genomic DNA digested separately with *Bam*HI **(lane 1)**, *Eco*RI **(lane 2)**, *Not*I **(lane 3)**, *Sal*I **(lane 4)** and *Xho*I **(lane 5)** and hybridised with **a)** mox^{dot} 0.75 kb **b)** $epox^{dot}$ 0.60 kb **c)** ord^{dot} 0.53 kb **d)** cyp^{dot} 0.36 kb **e)** 5' *pks* 1.32 kb **f)** R209 0.7 kb **g)** *ddhA* 0.38 kb **h)** *dotD* 0.25 kb **i)** *ver2* 2.0 kb **j)** *aflR* 0.324 kb. Four identical blots were produced. Approximate sizes (kb) of bands are indicated, the sizes shown represent the closest band on the Southern blot.

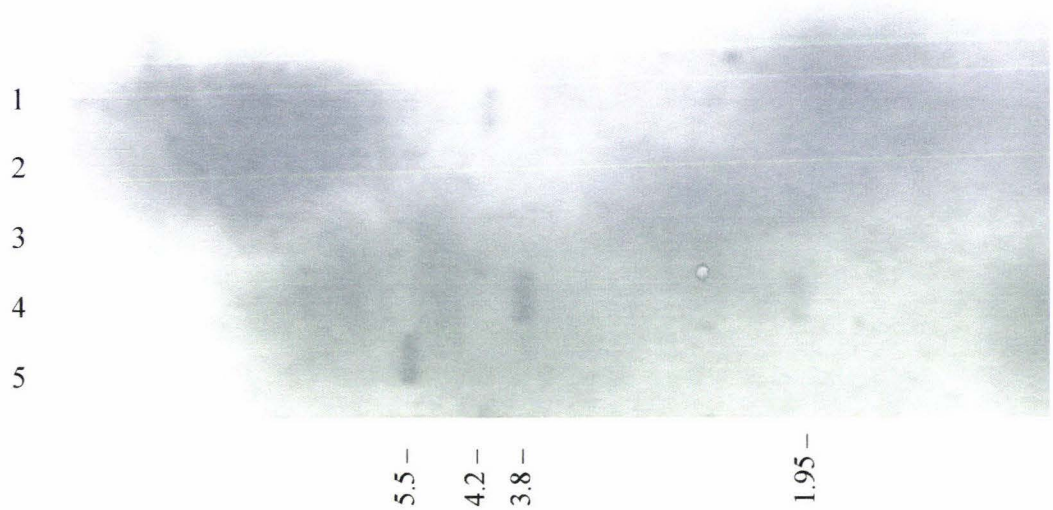
a) mox^{dot}



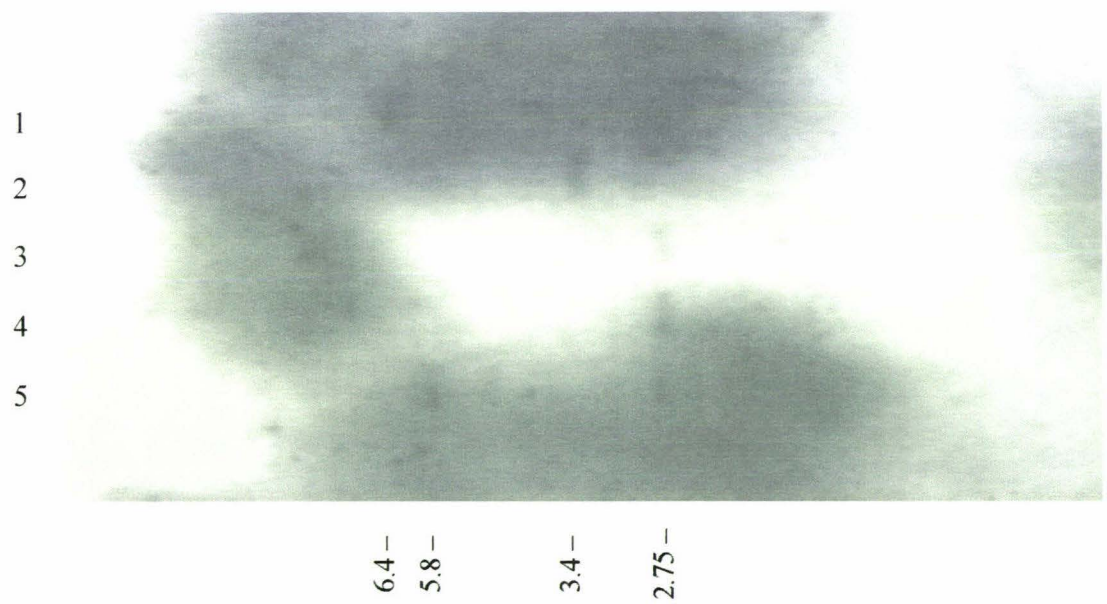
b) epox^{dot}



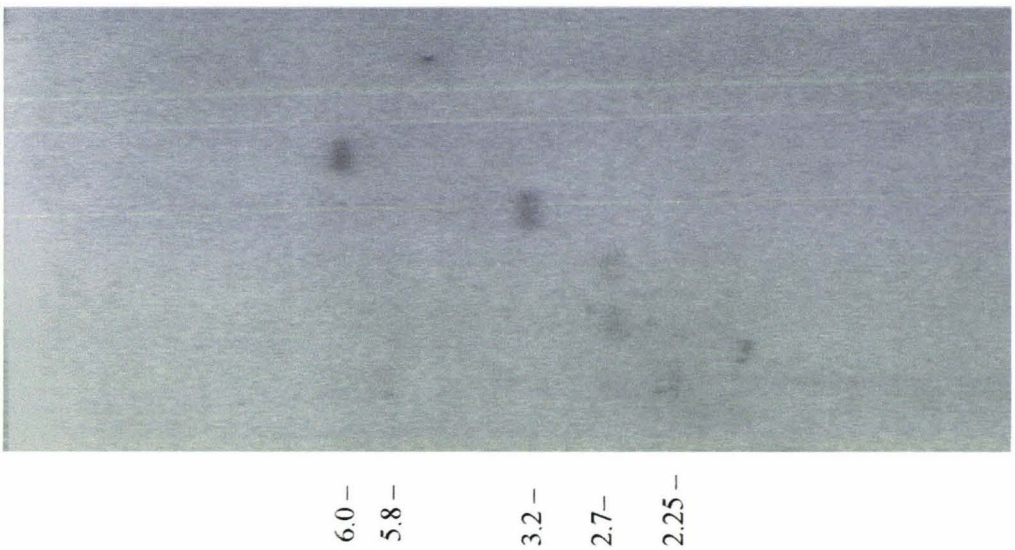
c) ord^{dot}



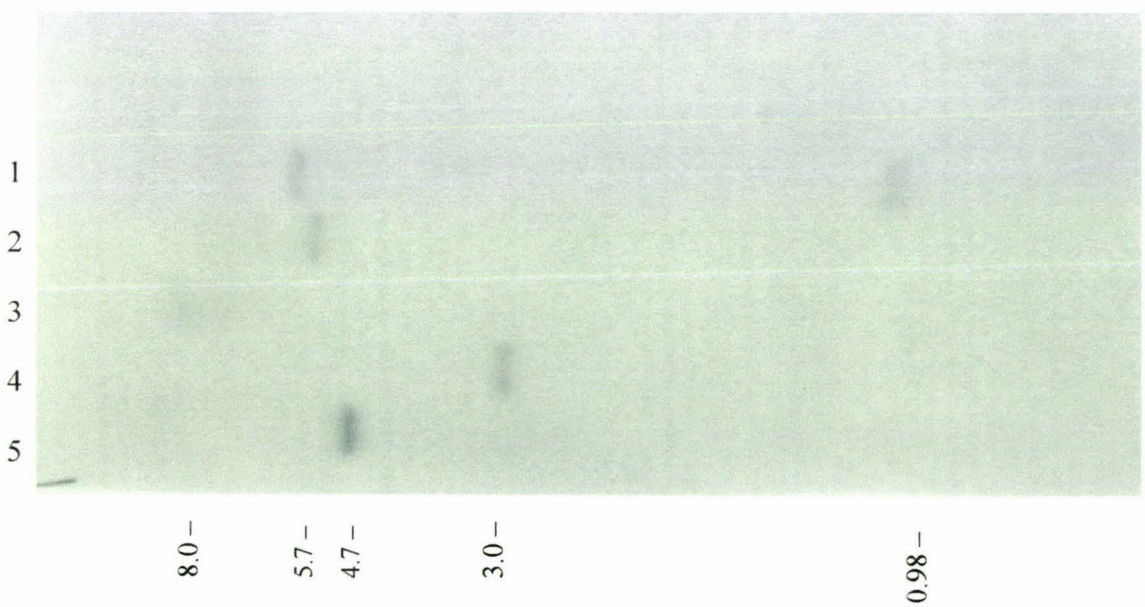
d) cyp^{dot}



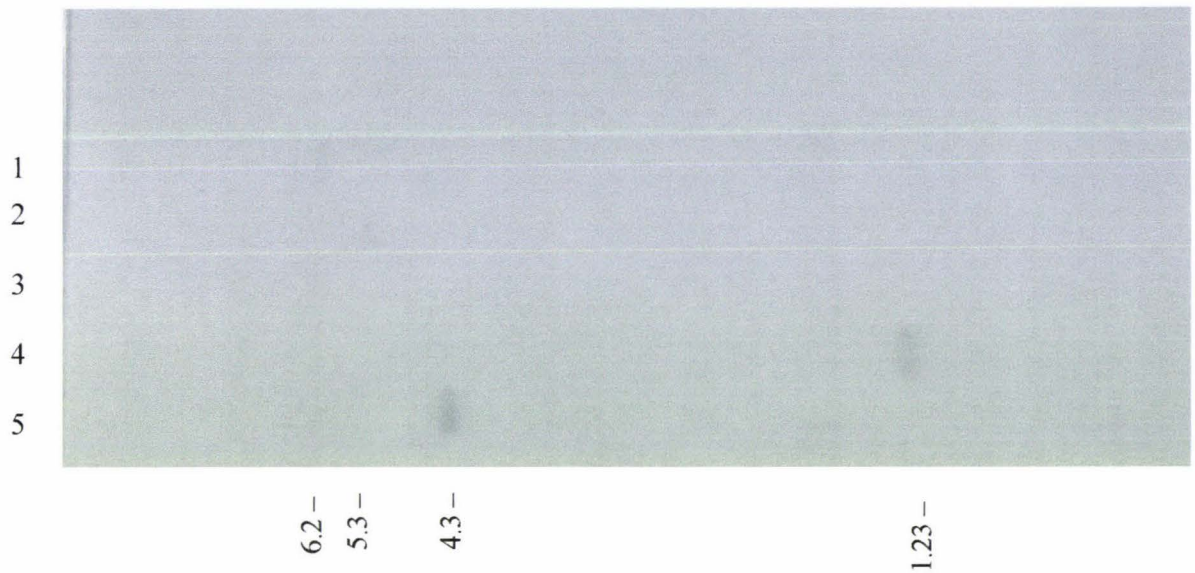
e) **5'** ***pks***



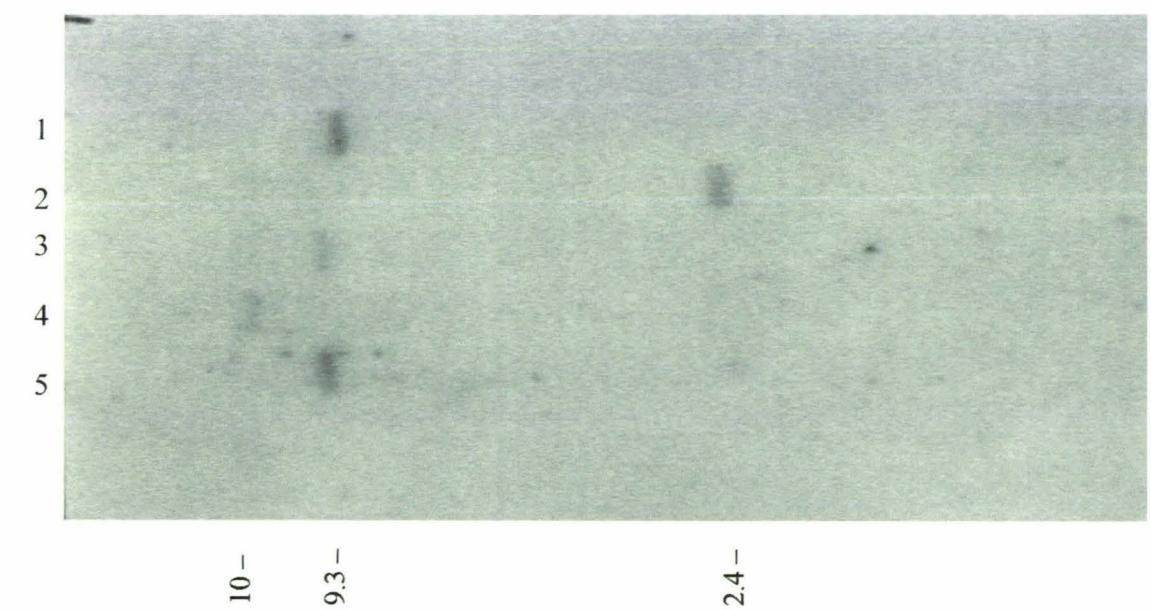
f) R209



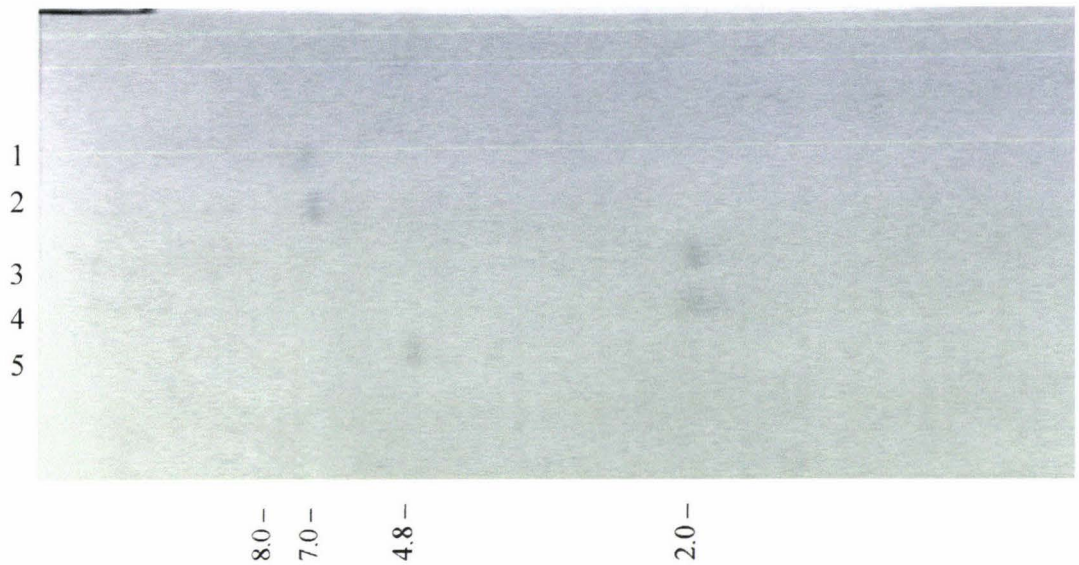
g) ddhA



h) dotD



i) ver-2



j) aflR

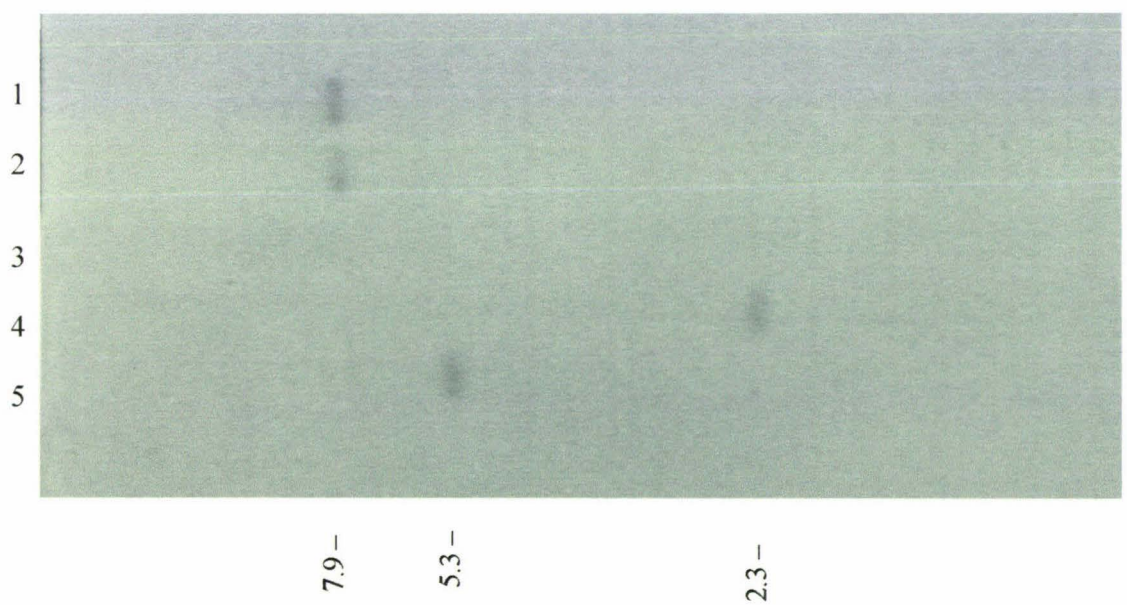


Table 4.1 *Dothistroma pini* Genomic DNA Southern Blotting Summary

Most of the fragment sizes produced through Southern blotting (figure 4.6; a-j) are within 10 % (allowing for error of size estimation) of the expected fragment sizes indicated on the restriction maps (figures 4.4 and 4.5) except those which lie at the end of a lambda arm. The sizes of these fragments are unknown and are indicated by an asterisk (*). The results produced from the Southern blots which do not lie within 10 % of expected fragment size are indicated by blue font and these are explained within the discussion. Those fragments which are seen faintly on the original autoradiograph but not scanned diagrams contained within this thesis are indicated by a question mark (?). A dash (-) indicates no hybridisation seen. All sizes are in Kb. The blot used is indicated by A, B, C or D and the order in which blots were probed are indicated in brackets (1), (2) or (3). DIG and radioactive (α-^{32}P) probes are indicated. KSA, CGV1 and CGV2 refer to the lambda clone in which the shown probe is located.

Note: Only a partial restriction map of λCGV2 is available and hence the expected fragment sizes produced from probing with *aflR* and *ver-2* are unknown and the map not included within this thesis.

Digests (Lane)	**KSA** *mox*^{dot} Blot A (1) DIG	**KSA** *epox*^{dot} Blot A (2) α-^{32}P	**KSA** *ord*^{dot} Blot C (1) DIG	**KSA** *cyp*^{dot} Blot B (1) DIG	**KSA** 5' *pks* Blot B (2) α-^{32}P	**KSA** R209 Blot C (2) α-^{32}P	**CGV1** *ddhA* Blot C (3) α-^{32}P	**CGV1** *dotD* Blot D (2) α-^{32}P	**CGV2** *ver-2* Blot B (3) α-^{32}P	**CGV2** *aflR* Blot A (3) α-^{32}P
*Bam*H1 (1)	4.4	4.4	4.2	6.0	6.0	5.7 0.98	6.2 *	8.8 *	8	7.9
*Eco*R1 (2)	8.0	8.0	-	3.2	3.2	5.3 6.2 ?	5.3	2.4 *	7	7.9
*Not*I (3)	9.8 ?	9.8	-	2.7	2.7	8 *	- *	9.3	2	-
*Sal*I (4)	3.9	3.9	3.8 1.95	2.65	2.65	3	1.23	10 + *	2.2	2.3
*Xho*I (5)	6.2 *	5.8	5.5	2.75 5.8	2.25? 5.8	4.7	4.3	9.3 *	4.8	5.3

4.4 DISCUSSION

4.4.1 Genomic Organisation of The Dothistromin Genes Within λKSA

Southern blot analysis of *D. pini* genomic DNA was performed to verify that the λKSA clone truly represents the genomic organisation of putative genes and is not the result of cloning artefacts produced in the construction of the *D. pini* genomic library. This analysis also indicated the copy number of each gene. *D. pini* mox^{dot}, ord^{dot}, $epox^{dot}$, cyp^{dot}, 5' *pks* and R209 gene-specific fragments (Figure 4.5) were used as probes and hybridised to Southern blots containing 5μg of *D. pini* genomic DNA digested separately with *Bam*HI, *Eco*RI, *Not*I, *Sal*I and *Xho*I.

Unfortunately, some questionable bands were produced (indicated through blue font in Table 4.1). However the presence of these bands are resolved in the following explanations.

The R209 probe hybridised to the expected 0.98 kb fragment in the *Bam*HI digest however the second fragment produced was smaller than expected (5.7 kb instead of the predicted 7.0 kb) (figure 4.5 and table 4.1). It appears that there may be another *Bam*HI restriction site; this is also supported by the fragments produced with hybridisation of cyp^{dot} and *5' pks* to the *Bam*HI digest. The 6.0 kb fragment produced with hybridisation of *5' pks and* cyp^{dot} (table 4.1) is also smaller than expected. When examining the available sequence of λKSA there are several degenerate *Bam*HI sites which differ from the restriction site by one base pair. The sequencing in the area of these possible additional *Bam*HI restriction sites is not completely accurate hence it is possible that one of these *Bam*HI sites is real. In accordance with this S. Zhang (personal communication) has identified regions within the λKSA clone sequence with sequence ambiguities (based on data from previous students).

Hybridisation with the R209 probe to the *Eco*RI digest also produced some unexpected results. The 6.2 kb fragment produced (table 4.1) is expected (this is not clearly seen in figure 4.6 e but is faintly visible on the original autoradiograph). As the more clearly seen fragment is 5.3 kb (table 4.1 and figure 4.6 f) it appears that once again there may be an extra *Eco*RI site present which is not observed due to sequencing errors. The presence of

both fragments could then be explained through a partial digest of *Eco*RI in blot C. A partial digest of *Eco*RI in blot C is also supported with the *ddhA* hybridisation producing one 5.3 kb fragment (figure 4.6g and table 4.1) instead of the expected 2.7 kb and approximate 1.36 kb fragments (figure 4.4). As the 1.36 kb fragment is at the end of a lambda arm it could be slightly longer hence when taken in addition with the 2.7 kb fragment could explain the presence of the 5.3 kb fragment produced in hybridisation with *ddhA*.

When using the cyp^{dot} and *5' pks* probes there is hybridisation to an unexplained ~ 2.7 kb fragment within the *Not*I digest and also in the *Xho*I digest (table 4.1). There is a possibility that there are degenerate *Not*I sites within this region which do not appear to be present through sequencing error (several sites which differ by only one base pair were identified, two of these producing a 3 kb fragment). However it seems most likely that there was some plasmid contamination in the *Not*I and *Xho*I digests of blot B as the ~ 2.7 kb fragments produced are plasmid sized. The 2.75 kb fragment produced with hybridisation of cyp^{dot} to the *Sal*I digest is not thought to be plasmid contamination as, firstly, this is the expected fragment size and, secondly, the same result was seen when repeating this experiment using radio-active hybridisation on a different blot (results not included). The cyp^{dot} radioactive Southern results were not included as there was also some cross-hybridisation to the λ size markers thus titrating the probe, and reducing hybridisation to the genomic *D. pini* region. This resulted in very faint bands being produced and hence scanning of the image was not very successful. The 2.25 kb fragment produced from cross-hybridisation of *5' pks* probe to *Xho*I digested *D. pini* genomic DNA is significantly smaller than the 2.7 kb fragment thought to be produced due to plasmid contamination and hence is proposed to be a real product. However there is one discrepancy that does not support the hypothesis of plasmid contamination. The *ver-2* probe also used on blot B did not reveal any plasmid size fragments in the *Not*I and *Xho*I digests. This could be explained by the *ver-2* probe being slightly cleaner than the cyp^{dot} and 5' *pks* probes in terms of having less (or no) plasmid contamination and therefore being less likely to hybridise to plasmid contamination within the blot itself.

Overall, Southern analysis clearly showed that the λKSA sequence and organisation, whilst containing a few putative sequencing errors, is representative of the *D. pini* genome. Results also indicate that all genes are present in single copies; this is especially evident in

*Xho*I digests for each of the hybridisations (figure 4.6 lane 5). As the enzyme *Not*I was not expected to cut within λKSA clone it was hoped that the same size fragment would be hybridised by each of the probes illustrating that the genes are clustered. Only two probes within the λKSA, R209 and *epox*^{dot} clearly cross-hybridised to the *Not*I digest. These fragments are predicted to be 8 kb and 9.8 kb respectively. However there is some inaccuracy in the methods used for calculating the size of these large fragments hence it is possible the fragments are much larger. This would explain why hybridisation to *Not*I fragments was not seen with all the probes as larger fragments are harder to transfer and therefore may not be seen.

4.4.2 Linkage of The Lambda Clones

As the lambda clones so far identified (λKSA, λCGV1 and λCGV2) do not overlap each other (Laarakkers, 1999) it was anticipated that Southern blot analysis would show how closely the lambda clones are located to one another in the *D. pini* genome. This was approached by designing probes for Southern blotting at each end of two lambda clones (λKSA and λCGV1) and within the 3 kb subcloned region of λCGV2. If clones are within close proximity to each other the same size restriction fragment would cross-hybridise to probes within each of the clones. The Southern blots did not reveal any large fragments which hybridised to all of the clones and did not reveal any fragments which suggested there was a junction between any two of the lambda clones. This does not necessarily indicate the three lambda clones are not linked within the same gene cluster. Due to the similarity of genes within the each of the lambda clones to genes within the AF/ST gene clusters it is likely that these three lambda clones are linked but are too far apart to be shown through Southern blotting with the restriction digests chosen for this study. No fragments over 10 kb were seen in this study and hence it is likely that the three lambda clones (λKSA, λCGV1 and λCGV2) are over 10 kb apart. Since the AF/ST clusters are approximately 70 kb in length, this is a distinct possibility.

4.4.3 The λCGV2 Lambda Clone

The two clones λCGV1 and λCGV2 were identified by screening a *D. pini* genomic library using the heterologous probe *ver*-1 from *A. parasiticus* which encodes a putative aflatoxin

ketoreductase but which also displays significant amino acid identity to ketoreductase genes involved in melanin biosynthesis (Gillman, 1996). Southern blot analysis confirmed that λCGV1 contained a sequence hybridising to the *A. parasiticus ver-1* gene fragment. The *ver-1* gene fragment found in λCGV1, later termed *dotA* after completion of sequencing, (Monahan, 1998) displays 80.2 % amino acid identity to the *A. parasiticus* AF biosynthetic gene *ver-1*. Disruption via homologous recombination of *dotA* confirmed the involvement of this putative gene in the dothistromin biosynthetic pathway (Seconi, 2001).

Partial characterisation of the second clone, λCGV2, revealed that the other potential *ver-1* gene (referred to as *ver-2)* showed greater similarity to the melanin biosynthetic gene *phn* (polyhydroxynapthalene) from *Cochliobolus heterostrophus* (with 77% amino acid identity and 86% similarity) than to the aflatoxin gene *ver-1*. Hence this gene appears to encode an enzyme required for melanin biosynthesis rather than being required for dothistromin biosynthesis. An *aflR* like sequence was also found within the 3.0 kb characterised subclone of λCGV2. Southern blotting carried out using the *ver-2* gene and possible *aflR* gene has confirmed that it is likely that both of these genes lie within the same region of the genome (figure 4.6 i & j and table 4.1). The same size fragments hybridised for most of the restriction digests with both the *aflR* and *ver-2* probes indicating linkage of these two genes. However no linkage to λKSA or λCGV1 was detected. The close proximity of *aflR* and *ver-2* within the *D. pini* genome suggests that either *aflR* is required for melanin biosynthesis and not dothistromin synthesis or that the dothistromin gene cluster also contains a gene required for melanin biosynthesis. The latter would suggest some gene duplication within the cluster.

4.4.4 Limitations of Southern Blotting

Southern blotting analysis can be inaccurate, firstly due to the clarity of blots. DIG labeling of probes especially interfered with the clarity of blots creating a lot of background from non-specific binding which made it particularly difficult to distinguish real fragments. Radioactive labeling of blots, although not creating background, requires a subjective opinion for determining how long to expose blots to X-ray film. This is governed by the 'hotness' of a probe after labeling with α-^{32}P. Not leaving blots exposed to X-ray film for long enough can result in a weak signal transfer and hence fragments may not be seen.

However as the radioactive signal deteriorates very slowly, the blot is able to be re-exposed to new X-ray film. Secondly Southern blotting can be inaccurate due to the methods used to deduce fragment sizes; hence a 10% difference was accepted between the expected fragment size and the fragment sizes determined from the Southern blots. Some errors in size prediction can definitely be seen within this study when two or more probes were expected to hybridise to the same size fragment. This can be seen with the three probes *mox*^{dot}, *epox*^{dot} and *ord*^{dot}. These three probes should have hybridised to the same fragments in four of the restriction digests (*Bam*HI, *Eco*RI, *Not*I and *Sal*I; table 4.1). The probes used in the same blot (*mox*^{dot} and *epox*^{dot}) gave the same size fragments but when deducing the size of fragments of *ord*^{dot} which was hybridised to a different blot, different sizes were seen. This was also seen with the three probes *cyp*^{dot}, 5' *pks* and R209 within the *Bam*HI digest.

Transfer of large fragments is also not very efficient, depurination could have been used to overcome this. This is not normally required for fungal genomic DNA, however since no, fragments larger than approximately 10 kb were transferred in this study, perhaps depurination could have resulted in larger fragment transfer and hence the positioning of the three lambda clones within the *D. pini* genome may have been revealed. Using a lower percentage agarose gel (0.7- 0.8 %) would have also resolved the larger fragments. However a 1% gel was used to enable clarity of smaller fragments required for confirming the genomic positioning of genes previously identified within λKSA. Another limitation of Southern blotting is that there may be more than one copy of a gene within the genome and hence it is possible that the wrong gene is cross-hybridised. In this case it is unlikely as no evidence for cross-hybridisation of the *ver-2* probe to *ver-2* of λCGV2 as well as *dotA (ver-1)* of λCGV1 (Table 3.1 and Figure 4.6i).

4.4.5 Summary

Southern blotting has shown that the organisation of genes within the λKSA clone is representative of the *D. pini* genome. It has also indicated that all genes within the λKSA clone are present in a single copy and indicated regions within the unpublished λKSA clone where further sequence analysis is required to remove errors. Sequencing of the entire cluster is currently in progress and will enable an accurate restriction map to be

constructed. Southern blotting has also confirmed suggestions from examination of the λCGV2 clone that *ver-2* and *aflR* are located together in the genome, providing the possibility that *aflR* could be involved in melanin biosynthesis or that *ver-2* could be a melanin biosynthesis gene located within the dothistromin gene cluster.

5.0 TARGETED REPLACEMENT OF A PUTATIVE DOTHISTROMIN BIOSYNTHETIC GENE *PKS^{DOT}*

5.1 INTRODUCTION

Dothistromin is a polyketide containing a common furobenzofuran moiety like other related polyketides, including AF and ST. The first step of synthesis (conversion of acetate to noranthrone) is catalysed by a fatty acid synthase (FAS) to form hexanoic acid and subsequently a polyketide synthase (PKS) extends the hexonyl starter unit to produce noranthrone. Probing of a *D. pini* genomic library with the highly conserved β-keto acyl synthase (KS) domain from the polyketide synthase gene (*pksL1*) of *A. parasiticus*, indicated the presence of a homologous gene sequence in *D. pini*. Cloning, isolation and sequencing of a library clone containing the putative *pks* gene (λKSA) confirmed the presence of part of a *pks* gene (*pks^{dot}*) with two functional domains, β-keto acyl synthase (KS) and acyl transferase (AT) (Morgan 1997).

Sequence analysis carried out by Seconi, 2001 (see section 4.3; Seconi, 2001) of the partial *pks^{dot}* sequence revealed the first 1426 amino acids have 62% identity (74% similarity) to PKSA (*pksL1* = *pksA)*, a PKS involved in AF synthesis in *A. parasiticus*, and 63% (74%) to STCA, a PKS involved in ST synthesis in *A. nidulans*. This suggests that the PKS of *D. pini* is involved in dothistromin synthesis. Sequencing of the λKSA clone in previous studies has provided enough sequence to enable a *pks^{dot-}* mutant to be constructed. It is hypothesised that a non-functional *pks^{dot}* gene will disrupt the dothistromin biosynthetic pathway at the two step conversion of acetate to noranthrone. This mutant construct will confirm if Pks^{Dot} is on the dothistromin biosynthetic pathway.

5.2 DISRUPTION OF *PKS*^{DOT} BY HOMOLOGOUS RECOMBINATION

5.2.1 Disruption Plan A

The disruption strategy was developed following the formation of a previous disruption mutant (*dotA)* by Seconi (2001). The original disruption strategy involved using PCR to amplify regions from genomic DNA that flank the KS and AT domains. It was thought that replacement of the KS and AT domains with the 2.3 Kb *hph* cassette would make the *D. pini* PKS non-functional. One primer in each amplification reaction was designed to contain an *Xba*I site to enable the two flanking regions to be joined (Figure 5.1). Each flanking region (gel extracted) was to be ligated into a pGEM- T Easy vector and then through digestion and ligation both flanking regions would be combined in one vector. Once in this vector an *Xba*I digest was to be performed to enable the *hph* cassette to be placed in the middle of the flanking regions (Figure 5.1). For further details see sections 2.6.1, 2.8, 2.11.1.1 and 2.12.1.3.

However this disruption plan had to be amended. After ligating the 5' and 3' *pks* flanking regions into pGEM-T easy, sequencing of the 3' *pks* plasmid revealed an extra *Xba*I site (Figure 5.2) which was not predicted from the original gene sequencing (Morgan, 1997) due to a sequencing error. The sequencing error was checked to ensure it was not just a PCR artefact, by sequencing an independent PCR product obtained using a genomic DNA template. Discovery of this extra *Xba*I site meant the disruption strategy would have to be redesigned (section 5.2.2).

Figure 5.1 Disruption Plan A

Amplification of the genomic regions flanking the KS and AT domains and ligation of the *hph* cassette in pGEM-T Easy vector containing 5' and 3' *pks* flanking regions.

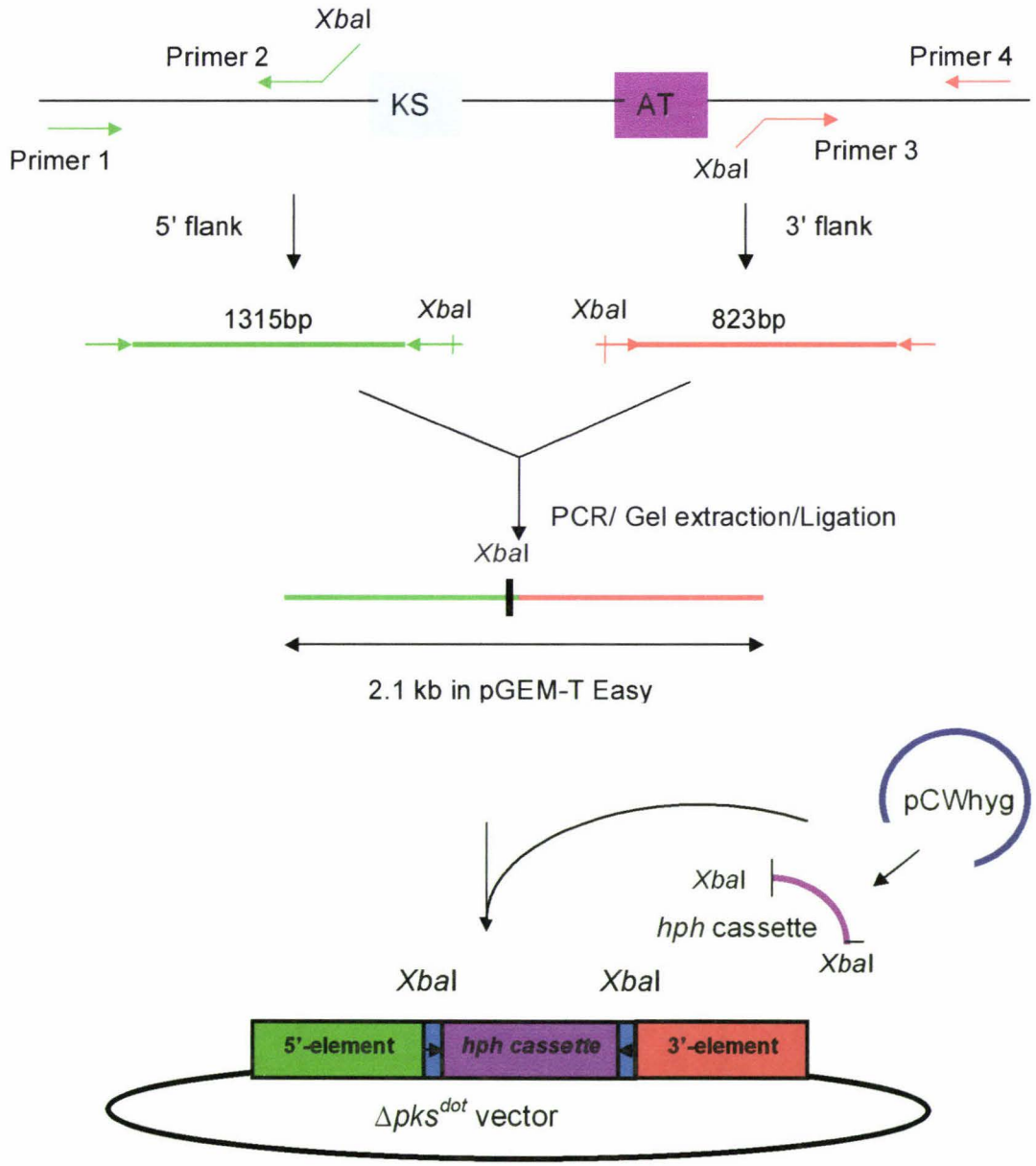
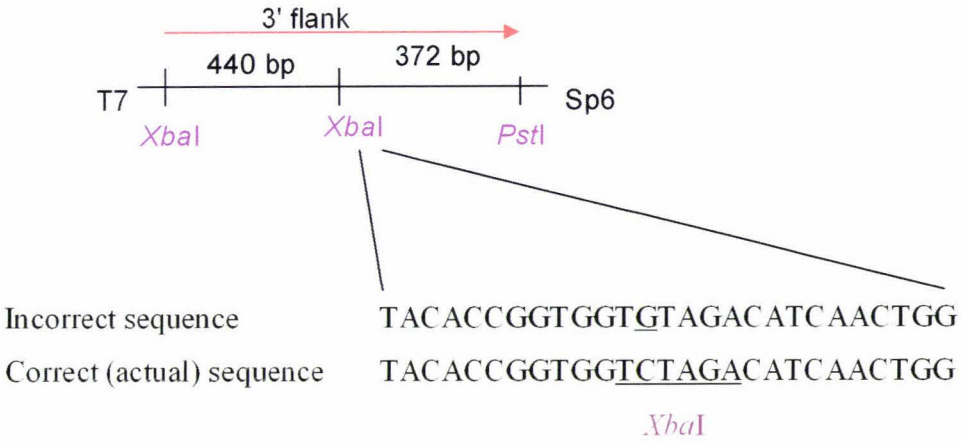


Figure 5.2 3' *pks* fragment with relevant restriction sites indicated



With the discovery of the extra *Xba*I site the 3' flank would be cut into two fragments, leaving the option of using part of the 3' flank or using a partial digest to obtain the whole 3' *pks* flank. As neither of these methods were ideal plan B was implemented.

5.2.2 Disruption Plan B (MultiSite Gateway Recombination)

Gateway Technology enables transfer of heterologous DNA sequences between vectors through specific recombination sequences (*att* sites). Two recombination reactions are the basis of the MultiSite Gateway system. Firstly BP recombination, which facilitates recombination of an *att*B substrate (i.e. PCR product with the correct *att*B sites) with an *att*P substrate (donor vector) to create an *att*L- containing entry clone. Secondly LR recombination which facilitates recombination of *att*L-containing entry clones with an *att*R-containing destination vector to enable the formation of a disruption vector.

Step 1 Addition of attB sites

Using the flanking templates previously made for disruption strategy A, PCR was used to add recombination *att*B sites to each of the flanking regions contained within pGEM-T easy vectors. The primers used for isolation of the flanking regions were extended with the addition of the correct *att*B sites. Gateway OT pks KO/2-attB1 and gateway CB2-attB4 (table 2.2) were used to add *att*B1 and *att*B4 sites to the 5' flank creating a 1355 bp fragment. Primers gateway OT pks KO/3-attB2 and OT pks KO/4-attB3 (table 2.2) were used to add *att*B2 and *att*B3 sites to the 3' flank creating an 870 bp fragment. Primers gateway hph-attB1 and gateway hph-attB2 (table 2.2) were designed to enable amplification of the *hph* cassette from the pCWhyg vector template and also allow addition of the *att*B1 and *att*B2 sites, to create a 2400 bp *hph* cassette. Due to difficulties encountered in recombining the PCR product of the *hph* cassette + *att*B sites into its appropriate pDONR vector, the *hph* cassette + *att*B sites was firstly ligated into a pGEM-T easy vector (section 2.11.1.1). These enabled the *hph* cassette + *att*B sites to be in a supercoiled form which enabled more efficient recombination into the appropriate pDONR vector.

Step 2 Recombination into pDONR vectors

After addition of the correct *att*B sites BP recombination was carried out with each of the ***att*B**-flanking region and the appropriate ***att*P**-containing donor vectors. (Table 5.1)

Table 5.1: PCR products and their appropriate pDONR vectors

PCR product	pDONR™ vector
*att*B4-5' flank-*att*B1	pDONR™ P4-P1R (*att*P4 and *att*P1R)
*att*B1- *hph* cassette - *att*B2 (in pGEM-T easy)	pDONR™ 221 (*att*P1 and *att*P2)
*att*B2- 3' flank - *att*B3	pDONR™ P2R-P3 (*att*P2R and *att*P3)

Each of the pDONR™ vectors contain *att*P regions which enables recombination with the *att*B sites of the PCR products. (BP clonase reaction: *att*B X *att*P → *att*L X *att*R)

BP Recombination reactions were set up as outlined in section 2.11.1.2 and transformed into $CaCl_2$ competent TOP10 *E. coli* (section 2.12.1.3). Figure 5.3 outlines the BP recombination used.

After carrying out the three BP recombination reactions, the following three entry clones were created:
*att*L4- 5' flank- *att*R1
*att*L1 – *hph* cassette – *att*L2
*att*R2- 3' flank- *att*L3

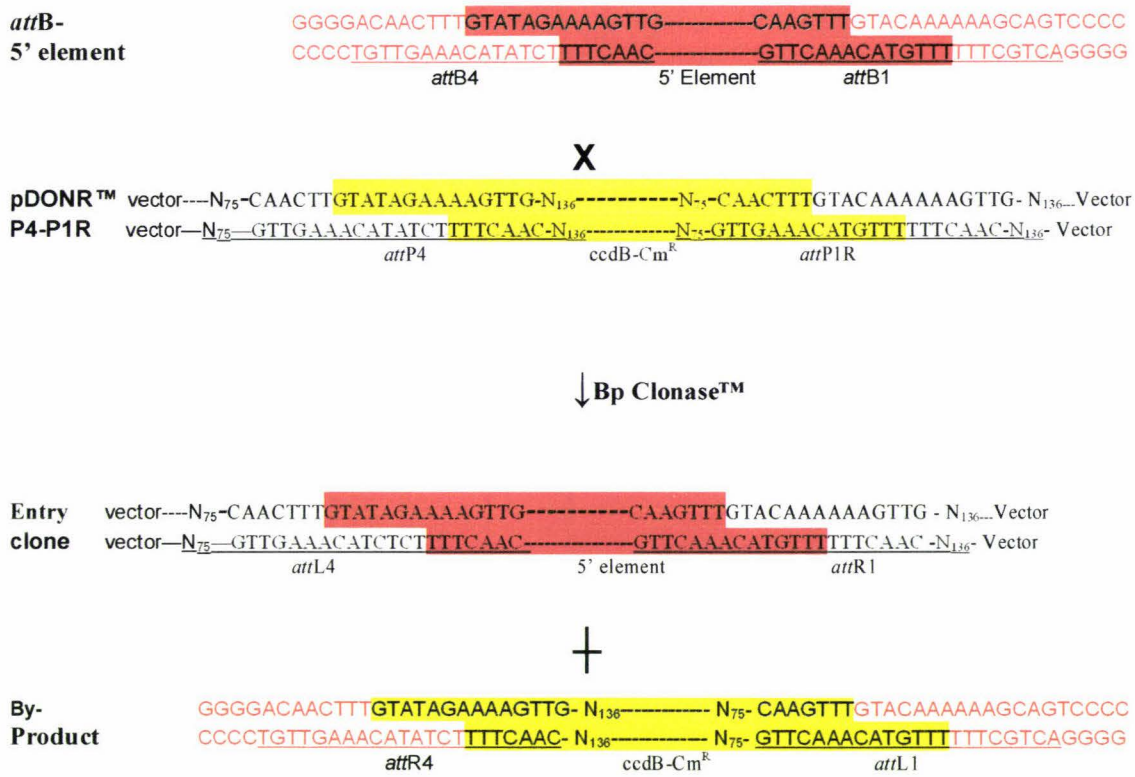
After carrying out *E. coli* transformations (section 2.12.1.4), the numbers of transformants for each of the three entry clones were reasonably low. There were approximately 5-10 colonies per transformation using 56 ng of the 3' flank PCR product and 90 ng of the 5' flank PCR product and 300 ng of the appropriate pDONR vector (this is compared to the 40 + colonies seen in the positive controls for each of the transformations which used 100 ng of control plasmid). There was a slightly increased number of transformants seen with the recombination reaction between *hph* cassette in pGEM-T easy and pDONR 221 which can be accounted for by the increase in the amount of DNA used (165 ng of *hph* pGEM-T easy, the amount of pDONR vector remained constant at 300 ng). Plasmid DNA was isolated (section 2.6.3) from approximately 5 transformants (for each of the three transformation events) and most of these appeared to contain the correct insert on PCR screening (Table 2.3). Transformants were PCR screened using the primers originally used to add the *att*B sites (Table 2.2). After this initial screening one of each of the entry clones were then sequenced using M13 forward and M13 reverse gateway primers before continuing with the three way LR recombination (section 2.14).

Step 3 Three way Recombination into pDEST vector

After generation of the three entry clones an LR recombination was performed to simultaneously transfer the three DNA fragments into the pDEST™ R4-R3 destination vector to create a disruption vector with the following structure:
*att*B4- 5' element- *att*B1- *hph* cassette- *att*B2- 3' element- *att*B3

Figure 5.3 An Example of a BP Recombination Reaction

Figure 5.3 shows an example BP recombination reaction, between a *att*B4 and *att*B1-flanked PCR product (i.e. 5' *pks* flank) and pDONR™ P4-P1R to create an entry clone and a by- product. ccdB- gene allows for negative selection of pDONR vectors. cm^R (chloramphenicol resistance gene) – used for counter selection. Loss of ccdB-cm^R allows for selection of insert within donor vector.



Modified from Invitrogen Multisite Gateway Three-Fragment Vector Construction manual.

A three way recombination reaction was set up as outlined in section 2.11.1.2 and transformed (sections 2.12.1.2 and 2.12.1.4). An overview of the LR recombination reaction is shown in Figure 5.4. After carrying out the three way recombination reaction (approximately 100 ng of each of the entry clones and 60 ng of the pDEST vector were used in the recombination reaction) and transformation, eight transformants were obtained. Plasmid DNA was isolated (section 2.6.3) from all eight transformants. PCR screening eliminated 4 of the transformants, as these did not contain all three inserts (section 2.13.1, table 2.4). Initial sequencing showed three of the four remaining transformants contained the correct sequence. Details and confirmation of procedures are given in the next section.

Step 4 Confirming Disruption vector

Firstly, after obtaining disruption vector DNA (section 2.6.3) PCR was carried out to ensure that the three inserts were all contained. This was done by using the primers designed to add the *att*B sites. For details see Table 2.4 and Figure 5.5. To ensure the three inserts were present in the correct order PCR was also carried out between the *hph* cassette and each of the 5' and 3' *pks* flanks, for details see table 2.4 and Figure 5.6. Primers were designed from each end of the *hph* cassette (5' *hph* out and 3' *hph* out) and used in conjunction with primers at each end of the 5' and 3' flanking regions (OT pks CB2 and OT pks KO/4 respectively). After confirming the disruption vector through PCR, sequencing was carried out using primers within the flanking regions and *hph* cassette to ensure its integrity (Sequence of disruption vector is included in appendix 4) (section 2.14). After confirmation from PCR and sequencing plasmid 4 (shown in Figures 5.5 and 5.6) was designated pR226.

Figure 5.4 LR Recombination Reaction

Schematic outlining the three was LR recombination reaction. Diagrams indicate which att sites interact to produce the recombination product.

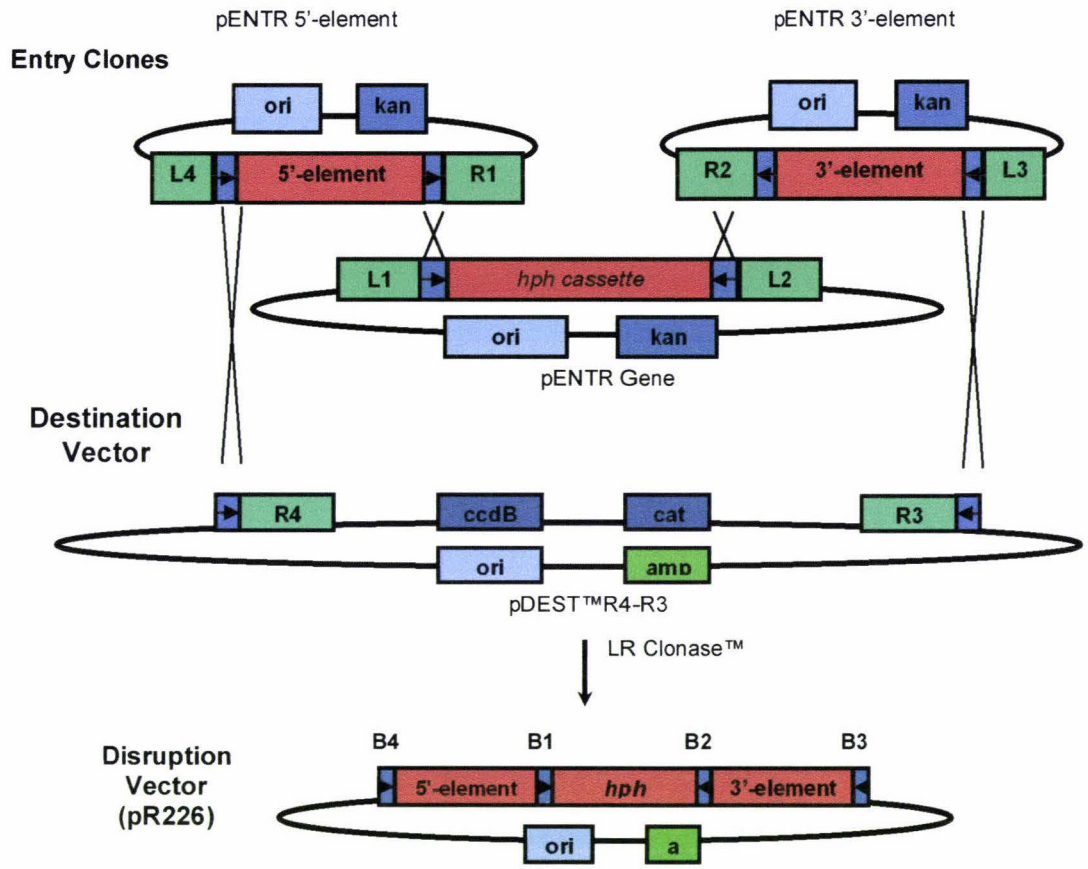


Diagram modified from Invitrogen Multisite Gateway Three-Fragment Vector Construction manual.

Figure 5.5 – Determining if the Three Inserts Are Contained Within the *pks* Disruption Vector

Agarose gel electrophoresis (section 2.9) was used to determine the presence of inserts within the five potential *pks* disruption vector constructs (plasmids 1-5). Primers gateway OT pks KO/2-attB1 and gateway pks CB2-attB4 (Table 2.2) were used determine if the 5' *pks* flank was present; Primers OT pks KO/3-attB2 and OT pks KO/4-attB3 (table 2.2) were used to determine if the 3' *pks* flank was present; and primers gateway hph-attB1 and gateway hph-attB2 (Table 2.2) were used to determine if the *hph* cassette was present. Plasmids 2 and 4 contain the expected size for all three inserts.

Lane	PCR Primers (Table 2.2)	DNA	Expected fragment size
2	M13 Rev gateway and M13 fwd	pDEST Vector	1.87 kb
3	gateway OT pks KO/2-attB1 and gateway pks CB2-attB4	Plasmid 1	1.3 kb
4	OT pks KO/3-attB2 and OT pks KO/4-attB3	Plasmid 1	0.85 kb
5	gateway hph-attB1 and gateway hph-attB2	Plasmid 1	2.45 kb
6	gateway OT pks KO/2-attB1 and gateway pks CB2-attB4	Plasmid 2	1.3 kb
7	OT pks KO/3-attB2 and OT pks KO/4-attB3	Plasmid 2	0.85 kb
8	gateway hph-attB1 and gateway hph-attB2	Plasmid 2	2.45 kb
9	gateway OT pks KO/2-attB1 and gateway pks CB2-attB4	Plasmid 3	1.3 kb
10	OT pks KO/3-attB2 and OT pks KO/4-attB3	Plasmid 3	0.85 kb
11	gateway hph-attB1 and gateway hph-attB2	Plasmid 3	2.45 kb
12	gateway OT pks KO/2-attB1 and gateway pks CB2-attB4	Plasmid 4	1.3 kb
13	OT pks KO/3-attB2 and OT pks KO/4-attB3	Plasmid 4	0.85 kb
14	gateway hph-attB1 and gateway hph-attB2	Plasmid 4	2.45 kb
15	gateway OT pks KO/2-attB1 and gateway pks CB2-attB4	Plasmid 5	1.3 kb
16	OT pks KO/3-attB2 and OT pks KO/4-attB3	Plasmid 5	0.85 kb
17	gateway hph-attB1 and gateway hph-attB2	Plasmid 5	2.45 kb

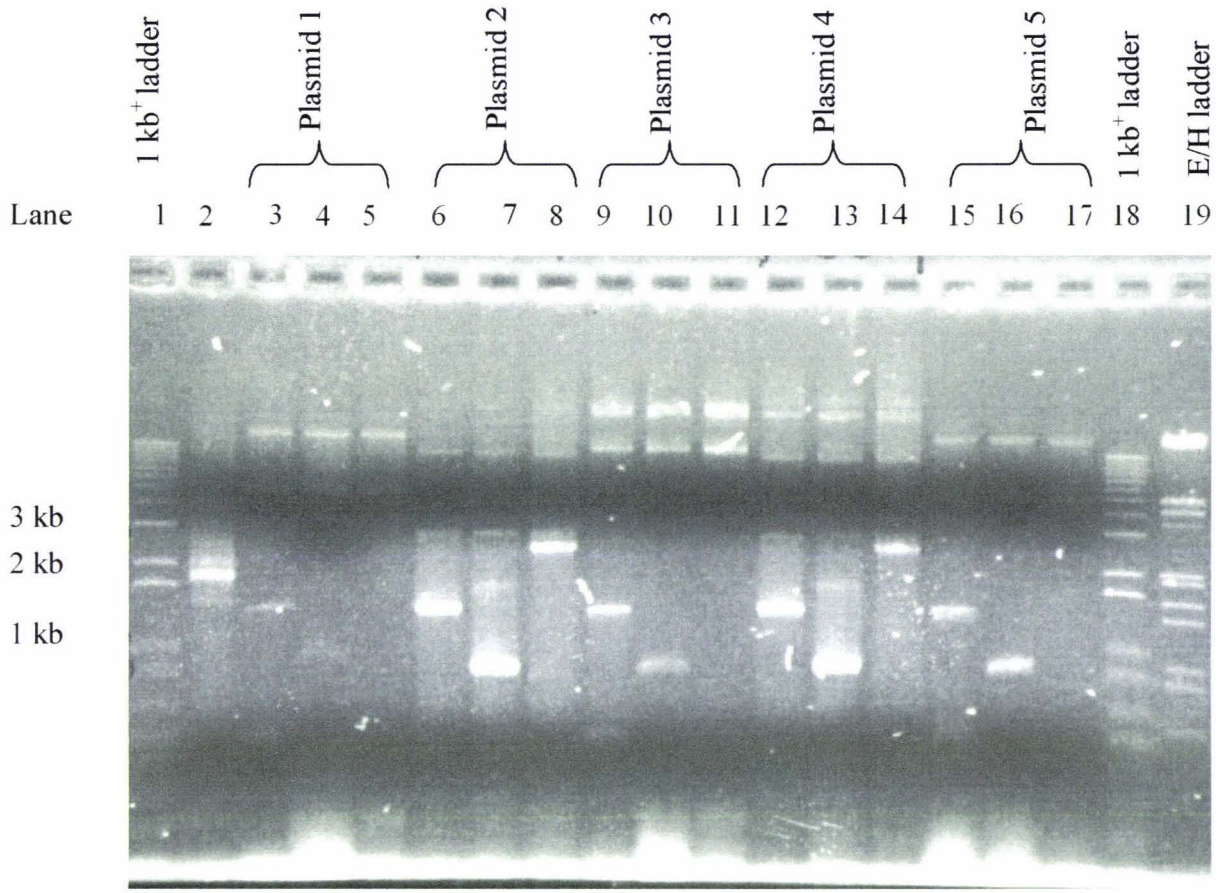

1 kb+ ladder
Plasmid 1
Plasmid 2
Plasmid 3
Plasmid 4
Plasmid 5
1 kb+ ladder
E/H ladder
Lane
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
3 kb
2 kb
1 kb

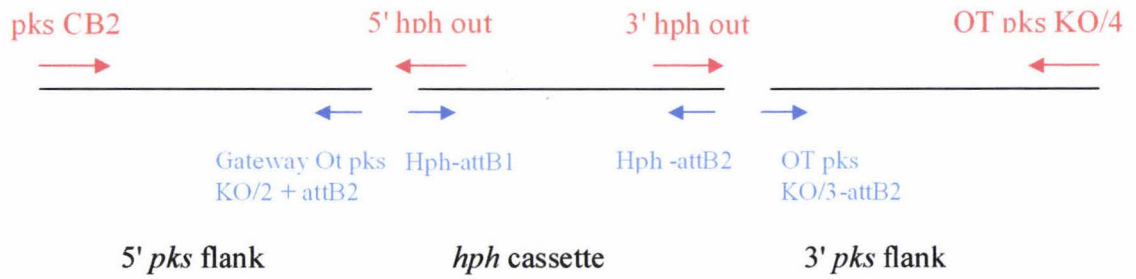
Figure 5.6 – Checking the Orientation of Inserts in the *pks* Disruption Vector

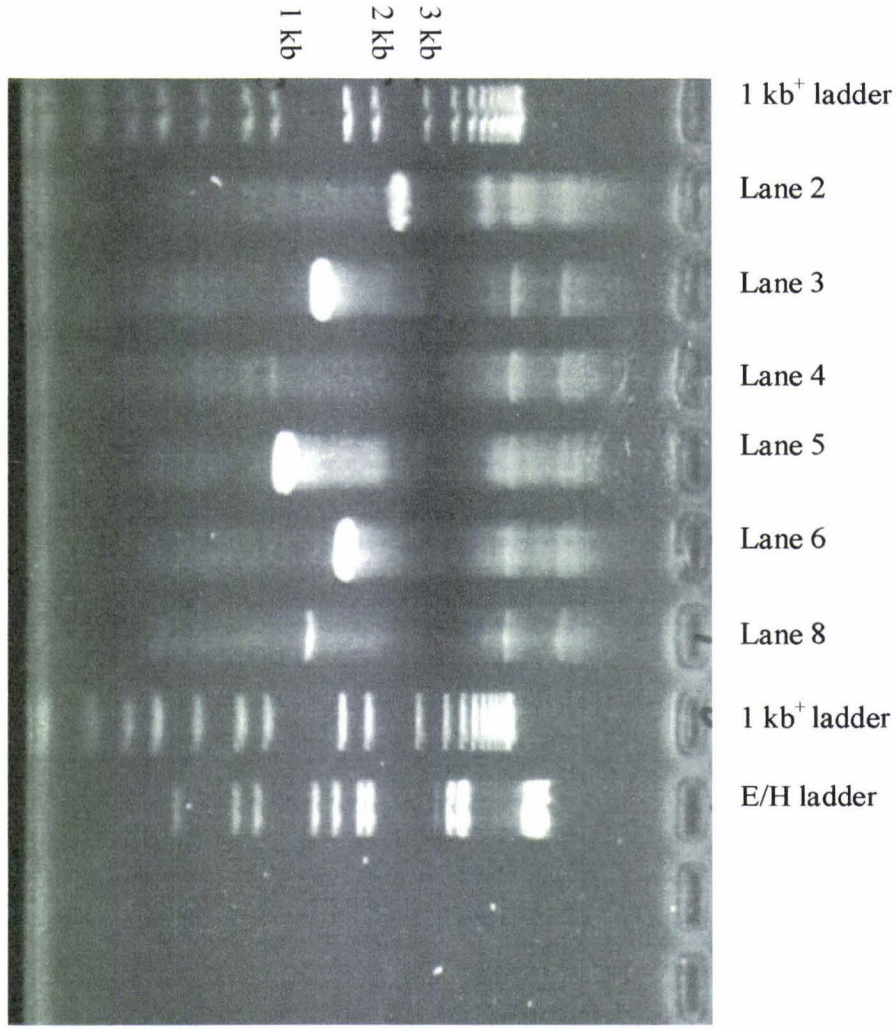
Agarose gel electrophoresis (section 2.9) was used to determine the orientation of inserts in the *pks* disruption vector construct. Plasmid 4 appears to contain the three inserts (5' flank, *hph* cassette, 3' flank) in the correct order. This was confirmed by sequencing (section 2.14).

Lane	PCR Primers (table 2.2)	DNA	Expected fragment size
2	gateway hph+attB1 and hph-attB2 (checking presence of *hph* cassette)	Plasmid 4	2.45 kb
3	gateway pks CB2+attB4 and gateway OT pks KO/2+attB2 (checking presence of 5' *pks* flank)	Plasmid 4	1.3 kb
4	3' hph out and pks CB2	Plasmid 4	No product
5	3' hph out and OT pks KO/4	Plasmid 4	1 kb
6	5' hph out and pks CB2	Plasmid 4	1.6 kb
7	5' hph out and OT pks KO/4	Plasmid 4	No product

Note: Product seen in lane 8 is probably due to non-specific binding of primers (the product seen is much fainter than other products seen). Sequencing has confirmed the integrity of plasmid four.

Schematic of Primer design:




1 kb
2 kb
3 kb
1 kb+ ladder
Lane 2
Lane 3
Lane 4
Lane 5
Lane 6
Lane 8
1 kb+ ladder
E/H ladder

5.3 GENERATION OF *PKS*^{DOT} MUTANTS

5.3.1 Optimisation of *D. pini* Transformation System

A transformation system for *D. pini* has been developed (Bidlake, 1996) and optimised (Seconi, 2001). However the system required minor modifications before working efficiently due to new enzyme batches for cell wall digestion.

A trial was carried out to determine the most efficient enzyme for the cell wall digestion of *D. pini* protoplasts, as a new batch of glucanex (beta-glucanase, Chemcolour Industry NZ) was available, along with the batch of glucanex previously used by Seconi (2001) and a new enzyme yatalase (Chitinase, Chitobiase, β-1, 3-glucanase, TAKARA BIO INC). A mixture of the glucanex batches (1:1) was most efficient for protoplast isolation (section 2.12.2.1). The duration of the incubation and the concentration of this enzyme solution were also optimised. An incubation time of 16-18 hours at 37° C in 20 mg/mL glucanex (dissolved in OM buffer) was found to produce the highest yield of protoplasts. A trial was also performed to determine the most efficient way of obtaining protoplasts. It was found that growing ground mycelia on cellophane sheets yielded higher numbers of protoplasts than using a liquid spore culture (approximately 3 weeks old). A preliminary comparison of these methods suggested that cellophanes produce 5.0 x 10^8 protoplasts/ mL whereas a liquid spore culture produces 2.1 x 10^7 protoplasts/ mL.

The method previously used by Seconi (2001) for isolation of *D. pini* protoplasts (section 2.12.2.2) also had to be modified, as protoplasts would not float on ST buffer (section 2.4.9) as expected but instead were pelleted down by centrifugation with STC buffer (section 2.4.6). See discussion for further details.

The regeneration frequency of wild type NZE5 protoplasts transformed with water and treated with PEG according to the transformation protocol (section 2.12.2) was extremely low at only 0.004%, this could be explained by protoplasts being pelleted down instead of being floated when overlaid with ST buffer (section 2.4.9). Extra material (i.e. mycelial

debris and spores) could have been present, making protoplast counting inaccurate and hence protoplast numbers may have been much smaller than estimated.

5.3.2 Transformation of NZE5 Protoplasts With *pks*^{dot} Disruption Vector

The transformation system for *D. pini* uses positive selection of transformants based on hygromycin resistance. Previous studies showed that 50 µg/mL hygromycin B (HmB) is sufficient to inhibit wild type growth of NZE5 (Seconi, 2001). Two plasmids contain the *E. coli* hygromycin B phosphotransferase *(hph)* gene; PAN7-1 and pCWhyg. PCWhyg used in this project was constructed from an *A. niger glaA* promoter, *E. coli* hygromycin B phosphotransferase *(hph)* gene and the *trpC* terminator from *A. nidulans.* pCWhyg has been shown to confer hygromycin resistance to *D. pini* (Seconi, 2001) and was used as a positive transformation control.

Due to the low number of protoplasts which regenerated, it was not surprising that no transformants were obtained either with the positive controls (pCWhyg or PAN7-1) or with the disruption vector construct (*pks*^{dot}). These procedures were repeated several times, without successful transformation of *D. pini*, even after protoplast numbers were increased through pelleting. Contamination also hindered the transformation procedure this was probably due to open centrifuge tubes being used during the centrifugation steps of the protocol (section 2.12.2).

5.4 DISCUSSION

5.4.1 Construction of *pks*^{dot} Disruption Vector

Targeted gene disruption in filamentous fungi is a powerful mechanism to identify the function of proteins encoded by specific genes (Shiotani and Tsuge, 1995). Homologous recombination of exogenous DNA permits the disruption and replacement of wild type genes with genomic segments that have been altered *in vivo*. Multisite gateway technology was implemented to allow the *in vivo* alteration of the polyketide synthase gene (*pks*^{dot})

through replacement of the two functional domains KS and AT with a hygromycin resistance gene.

Implementing Multisite gateway technology for construction of the disruption vector required setting up this procedure for future use by other laboratory members. This firstly involved creating stocks of all the vector components (i.e. pDONR vectors, the pDEST vector and pMS/GW control vector). Secondly, the procedures used had to be fine tuned to work as efficiently as possible. This involved establishing the required length of time for recombination reactions (section 2.11.1.2) (overnight incubation was most successful) and also the amounts of these recombination reactions required to be added to subsequent transformation reactions (section 2.12.1.4). The most efficient protocol for *E. coli* transformation was determined (chemical transformation rather than electroporation) and also methods for making Top10 competent cells (section 2.12.1.2) (these need to be made up fresh for each transformation as they do not store well). Thirdly, procedures for checking inserts within pDONR and pDEST vectors also had to be established as restriction digestion was not successful. Instead PCR was found effective for a preliminary screen before sequencing (it is important to make sure each construct is correct before proceeding to the next step).

Some problems were encountered whilst using the multisite gateway technology. Firstly some difficulty was experienced whilst carrying out one of the BP recombination reactions. Recombination of the *hph* cassette PCR product into pDONR 221 was initially unsuccessful. In order to overcome this problem, the *hph* cassette was firstly ligated by TA cloning into pGEM-T easy. The pGEM-T easy vector containing the *hph* cassette was subsequently used in the BP recombination reaction with the pDONR vector. The supercoiled nature of the *hph* cassette in pGEM-T easy seemed to aid the BP recombination. Now that the *hph* cassette has been ligated into pDONR 221 this is available for future laboratory members to use in LR recombination reactions, in order to construct disruption vectors of other *D. pini* genes. Secondly some difficulty was encountered whilst purifying PCR products (with their appropriate att sites) to use in BP recombination reactions. The recommended procedure of PEG purification (section 2.6.6)

was unsuccessful and resulted in loss of products due to the formation of a clear pellet. This procedure was successfully replaced by gel purification (section 2.6.1).

5.4.2 Generation, Transformation and Regeneration of *D. pini* Protoplasts

The inability to transform *D. pini* was an unexpected problem as the protocol followed was previously used to obtain a *dotA* knockout mutant in NZE5 (Seconi, 2001). The fact that protoplasts would not float when overlaid with ST buffer but remained in the bottom phase was especially unexpected. Futile attempts were made to enable the protoplast to float, by swelling protoplasts through altering the $MgSO_4$ concentration of the OM buffer (section 2.4.10) (reducing from 1.6 M to 1.4 M). As this problem was also experienced with the newly isolated wild type strain NZE7, it is unlikely that this could be due to a change in *D. pini* growth characteristics seen in the laboratory-cultured NZE5 wild type strain. The pathogenicity of these laboratory strains had not been tested to date. Due to time constraints the method for protoplast generation was unable to be improved any further. Recent work conducted within the laboratory suggests that the best way to obtain protoplasts is by growing macerated mycelium (rather than spores) in flasks of media rather than on cellophane covered plates. Protoplast numbers were increased by growing macerated mycelium in flasks and harvesting by centrifugation. Altering the amount of time for glucanex digestion and also the temperature at which digestion was carried out also seems to have a positive effect even with a reduced amount of glucanex (10 mg/mL).

Linearising the disruption construct may have also enabled a more efficient transformation to be carried out. Difficulty was encountered whilst trying to identify an enzyme which would cut the disruption vector once to enable linearisation. All the enzymes screened, cut the disruption vector in several places and thus were not of any use. The large size of the disruption vector (approximately 8800 bp; the insert being 4530 bp) means there may not be an enzyme which will enable linearisation. However as only a limited number of enzymes were screened in this study, further investigation is required.

Due to time constraints the transformation system could not be developed any further. As the pks^{dot} knockout mutant is hypothesised to be involved in the two step conversion of

acetate to noranthrone (the first step in the proposed dothistromin biosynthetic pathway) the construction of this mutant will be invaluable in determining the role of dothistromin in respect to pathogenicity and virulence. Due to the importance of obtaining a knockout mutant of *pks*^{dot} future lab work would involve firstly successfully establishing the transformation system. After obtaining transformants, PCR screening could be used to identify those where integration of the knockout construct in the genome has occurred. This can be confirmed through Southern blotting analysis. To determine the function of Pks^{Dot}, mutants can be examined for dothistromin production (through competitive enzyme-linked immunosorbant assay (ELISA)) or HPLC, accumulation of putative dothistromin intermediates and growth rate effects.

6.0 CONCLUSIONS AND FUTURE WORK

The main aims of this project were to elucidate *pks*^{dot} function and to further characterise genes involved in dothistromin biosynthesis through sequencing of a subclone (R162A) and through identifying the relative positions of the three lambda clones λKSA, λCGV1 and λCGV2.

None of the clones isolated from *D. pini* so far are large enough to contain the entire putative dothistromin biosynthetic cluster. The dothistromin biosynthetic cluster is expected to be 60-70 kb based on the sizes of AF and ST clusters identified in *A. parasiticus* and *A. nidulans*. In an attempt to piece together the λKSA, λCGV1 and λCGV2 clones in a single gene cluster, Southern blot analysis was implemented. Genomic *D. pini* digests were hybridised with two probes from each of the clones (one at either end of λKSA and λCGV1 and within the 3 kb characterised subclone of λCGV2) to see if any hybridised to the same fragment. No probes hybridised to the same restriction fragment indicating that the three lambda clones λKSA, λCGV1 and λCGV2 are not closely linked within the *D. pini* genome. However this does not indicate that the clones are not part of the dothistromin biosynthetic cluster as they could be as much as 40 kb apart and still be contained within the same cluster. Southern blotting confirmed that the putative *aflR* gene (encodes regulatory protein for activating gene transcription in AF/ST biosynthesis) and *ver-2* (thought to encode an enzyme required for melanin biosynthesis) are linked within the genome (suggested by the λCGV2 clone). It is therefore possible that the putative *aflR* is not responsible for regulating genes within the dothistromin biosynthetic cluster but instead involved in melanin biosynthesis. However in both *A. parasiticus* and *A. nidulans, aflR* does not regulate melanin biosynthesis, thus leading to the conclusion that perhaps the λCGV2 clone is not part of the dothistromin biosynthetic cluster, or that perhaps the characterised 3.0 kb end of λCGV2 defines one end of the dothistromin gene cluster. Alternatively it is possible that the dothistromin cluster includes a melanin biosynthetic gene, which could indicate that the cluster was formed from duplication of genes encoding proteins with other functions, which have been adapted through evolution for dothistromin biosynthesis.

Screening of a *D. pini* cosmid library is currently underway (Zhang, personal communication) using probes from different putative dothistromin biosynthesis genes. This will indicate which of the genes are linked and hopefully provide evidence linking genes which are currently on different lambda clones. Sequencing of the cosmid library clones will enable identification of new genes within the dothistromin gene cluster and in doing so is likely to divulge the whole gene cluster and order of genes within.

D. pini is expected to have a regulatory protein homologous to AflR found in the AF/ST gene clusters. Within *D. pini* two different regulatory sequences similar to those bound by AflR have been identified. The $TCGN_5CGA$ regulatory sequence has been identified upstream of two genes of the λCGV1 clone (*dotA* and *dotD)* and one gene within the λKSA clone (ord^{dot}). The regulatory sequence $TCGN_{11}CGA$ (which is identical to the AflR binding site in the AF/ST biosynthetic genes) has been identified upstream of all genes within the λCGV1 cluster (*dotA-D* and *ddhA)* and four genes within the λKSA clone ($epox^{dot}$, mox^{dot}, pks^{dot}, cyp^{dot}). The identification of both regulatory binding sites upstream of genes within both the λKSA and λCGV1 indicates that there may be two different regulatory proteins. These different regulatory binding sites may be utilised at particular stages of the dothistromin pathway or may also reflect the role of dothistromin in pathogenicity, enabling genes to be switched on in plants.

Disruption of *dotA* of the λCGV1 clone has confirmed the involvement of this gene in dothistromin biosynthesis (Seconi, 2001) and hence it is likely other genes within this lambda clone are also involved in dothistromin biosynthesis. Disruption of the pks^{dot} gene will confirm involvement in the dothistromin biosynthetic pathway and in doing so will also indicate whether other genes within the λKSA clone are involved as well.

This study resulted in the construction of a disruption vector, replacing two essential domains of the pks^{dot} gene (KS and AT domains) with a hygromycin resistance cassette (*hph).* Homologous recombination with the disruption plasmid (pR226) was intended to replace the wild type copy of the putative pks^{dot} gene resulting in a non-functional gene. Due to time constraints and difficultly in transforming wild type *D. pini* with the disruption plasmid a knockout was not achieved. However, recent work within the laboratory has improved the efficiency of the transformation system and attempts to obtain a pks^{dot} disruption mutant are currently in progress. This will enable the function

of *pks*^{dot} to be identified through examining mutants for dothistromin production and intermediate production. As *pks*^{dot} is postulated to be involved in the first step of the dothistromin biosynthetic pathway (working in a complex with a FAS in two part conversion) these dothistromin deficient mutants will allow the role of dothistromin in pathogenicity to be studied. Infecting *P. radiata* with the dothistromin deficient mutant will allow the pathogenicity of these mutants to be investigated. This will enable the role of dothistromin in needle blight to be deduced. Further knockouts of putative dothistromin genes will elucidate the biosynthetic pathway of dothistromin.

The order of genes within the dothistromin biosynthetic cluster is not thought to be correlated to the action of gene products, as homologs present in both AF and ST pathways are organised differently, yet catalyse equivalent steps on the biosynthetic pathways. The order of the homologs found within the λKSA and λCGV1 clones do not correspond to the order of genes in either the *A. parasiticus* AF cluster or the *A. nidulans* ST cluster. To further characterise genes involved in the dothistromin biosynthetic pathway primer walking with custom primers was implemented to obtain the full sequence of a subclone pR162A (part of the λKSA clone). PCR amplification was also used to determine the orientation of the pR1612A subclone relative to the adjacent subclone pR181. Sequence analysis of the approximate 4.5 kb subclone pR1612A identified two entire open reading frames (*ord*^{dot} and *epox*^{dot}) and one partial open reading frame (*mox*^{dot}), all of which display predicted amino acid identity to previously described genes. Mox^{Dot} displays high amino acid identity to the *moxY* and *stcW* gene products of ST/AF production which both encode monooxygenases, therefore Mox^{Dot} is also thought to be a monooxygenase. The functions of MoxY and StcW are yet to be confirmed (it is suggested that they are involved in the conversion of averufin (AVF) to versiconal hemiacetal acetate (VHA)) hence further investigation into *mox*^{dot} may shed light on the function of these homologs. Ord^{Dot} also displays high amino acid identity to products of genes of the ST/AF gene clusters (*avfA and stcO* respectively), both of which act as oxidases, once again suggesting a similar role of Ord^{Dot} in dothistromin biosynthesis. *AvfA* has been shown to encode an oxidase which is necessary for the conversion of AVF to VHA and hence Ord^{Dot} could possibly have the same role. To confirm the functions of *mox*^{dot} and *ord*^{dot} genes inactivation studies need to be carried out.

The *D. pini* $epox^{dot}$ gene found in this study does not display any homology to genes involved in AF/ST biosynthesis hence is probably a novel gene in an AF-type gene cluster. It displays high amino acid identity to the epoxide hydrolase of *A. niger* and hence is thought to be involved in a step which is unique to the dothistromin biosynthetic pathway. The dothistromin biosynthetic pathway diverges from the AF and ST pathways after the formation of versicolorin B, therefore $epox^{dot}$ is likely to be involved in the dothistromin biosynthetic pathway after this point. A gene disruption study is currently underway to reveal the function of $epox^{dot}$ within dothistromin biosynthesis.

As several genes within the λKSA clone have now been identified (mox^{dot}, $epox^{dot}$and ord^{dot} in this study and cyp^{dot} & pks^{dot} in previous studies), Southern blot analysis was used to confirm that the arrangement of genes within the λKSA clone is truly representative of the *D. pini* genome and not a result of cloning artefacts produced through construction of the *D. pini* genomic library. In doing so, areas within the λKSA clone which may contain sequencing errors were also identified. Further sequencing of the λKSA clone is currently in progress to remove areas of ambiguity. Southern blotting can also be used in the same manner to confirm the genomic organisation of other putative dothistromin genes.

Dothistromin research is able to contribute to the understanding of the biosynthesis and function of similar compounds ST and AF, as the formation of these secondary metabolites are thought to be the same until the formation of versicolorin B. The biological role (if any) of the metabolites AF and ST have not yet been determined, while dothistromin has been proposed as a pathogenicity factor in *Dothistroma* needle blight caused by infection of *P. radiata* with *D. pini*. The role of dothistromin as a pathogenicity factor should be authenticated through disruption of the pks^{dot} gene. Deducing the role of dothistromin in pathogenicity may enhance our understanding of why some *Aspergillus* species produce the highly toxic and carcinogenic metabolites AF and ST as no biological role has been attributed to these compounds.

REFERENCES

Akamatsu, H., Itoh, Y., Kodama, M., Otani, H., and Kohmoto, K. (1997) AAL-toxin-deficient mutants of *Alternaria alternata* tomato pathotype by restriction enzyme-mediated integration. *Phytopathology* **87**: 967-972.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic loacl alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Arand, M., Hemmer, H., Durk, H., Baratti, J., Archelas, A., Furstoss, R., and Oesch, F. (1999) cloning and molecular characterisation of a soluble epoxide hydrolase from *Aspergillus niger* that is related to mammalian microsomal epoxide hydrolase. *Biochemical Journal* **344**: 273-280.

Astin, J.W. (2000) Investigation of dothistromin regulation in *Dothistroma pini. MSc thesis, Massey University New Zealand.*

Ballance, D.J. (1986) Sequences important for gene expression in filamentous fungi. *Yeast* **2**: 229-236.

Bhatnagar, D., Cleverland, T.E., Brown, R.L., Cary, J.W., Yu, J., and Chang, P. (1997) Preharvest Aflatoxin Contamination: Elimination Through Biotechnology. In *Ecological agriculture and sustainable development.* Vol. 1. Dhaliwal, G.S., Arora, R., Randhawa, N.S. and Dhawan, A.R. (eds). Ludhiana, India: Indian Ecological Society, pp. 100-129.

Bidlake, A.T. (1996) Development of an efficient transformation system for *Dothistroma pini. MSc thesis, Massey University New Zealand.*

Bingle, L.E.H., Simpson, T.J., and Lazarus, C.M. (1999) Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genetics and Biology* **26**: 209-223.

Bowden, C.G., Smalley, E., Guries, R.P., Hubbes, M., Temple, B., and Horgen, P.A. (1996) Lack of association between cerato-ulmin production and virulence in *Ophiostoma novo-ulmi. Molecular Plant-Microbe Interactions* **9**: 556-564.

Bradshaw, R.E., Bidlake, A., Forester, N., and Scott, D.B. (1997) Transformation of the fungal forest pathogen *Dothistroma pini* to hygromycin resistance. *Mycological Research* **103**: 325-332.

Bradshaw, R.E., Ganley, R.J., Jones, W.T., and Dyer, P.S. (2000) High levels of dothistromin toxin produced by the forest pathogen *Dothistroma pini*. *Mycological Research* **104**: 325-332.

Bradshaw, R.E., Bhatnagar, D., Ganley, R.J., Gillman, C.J., Monahan, B.J., and Seconi, J.M. (2002) *Dothistroma pini*, a forest pathogen, contains homologs of aflatoxin biosynthetic pathway genes. *Applied and Environmental Microbiology* **68**: 2885-2892.

Brown, D.W., Adams, T.H., and Keller, N.P. (1996a) *Aspergillus* has distinct fatty acid synthases for primary and secondary metabolism. *Proceedings of the National Academy of Science of the United States of America*: 14873-14877.

Brown, D.W., Yu, J.H., Kelkar, H.S., Fernandes, M., Nesbitt, T.C., Keller, N.P., Adams, T.H., and Leonard, T.J. (1996b) Twenty-five coregulated transcripts define a sterigmatocystin gene cluster in *Aspergillus nidulans*. *Proceedings of the National Academy of Science of the United States of America* **93**: 1418-1422.

Bruchez, J.J.P., and Eberle, V.E.A. (1993) Regulatory sequences involved in the translation of *Neurospora crassa* mRNA: kozak sequences and stop codons. *Fungal Genetics Newsletter* **40**: 85-88.

Bullock, W.O., Fernandes, J.M., and Short, J.M. (1987) XL-1 blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with b-galactose selection. *Biotechniques* **5**: 376-378.

Cane, D.E. (1994) Polyketide biosynthesis: molecular recognition or genetic programming. *Science* **263**: 338-441.

Carson, S.D., and Carson, M.J. (1989) Breeding for resistance In forest trees, a quantitative genetic approach. *Annual Review of Phytopathology* **27**: 373-395.

Carson, S.D., Dick, A.M.P., and West, G.G. (1991) Benefits of the *Dothistroma*-resistant breed of radiata pine. *New Directions in Forestry: Costs and Benefits of Change. Australia and New Zealand Institutes of Forestry Conference., Christchurch, New Zealand.*

Chang, P., Bhatnagar, D., Cleverland, T.E., Bennett, J., W, Linz, J., Woloshuk, C.P., and Payne, G.A. (1993) Cloning of the *Aspergillus parasiticus apa-2* gene associated with the regulation of aflatoxin biosynthesis. *Applied and Environmental Microbiology* **59**: 3273-3279.

Chang, P., Cary, J.W., Yu, J., Bhatnagar, D., and Cleverland, T.E. (1995) The *Aspergillus parasiticus* polyketide synthase gene *pksA,* a homolog of *Aspergillus nidulans wA*, is required for aflatoxin B1 biosynthesis. *Molecular and General Genetics* **248**: 270-277.

Danks, A.V., and Hodges, R. (1974) Polyhydroxyanthraquinones from *Dothistroma pini. Australian Journal of Chemistry* **27**: 1603-1606.

Dick, A.M.P. (1989) Control of dothistromin needle blight in the *Pinus radiata* stands of Kinleith forest. *New Zealand Journal of Forestry Science* **19**: 171-179.

Eaton, D.L., and Gallagher, E.P. (1994) Mechanisms of aflatoxin carcinogenesis. *Annual review of Pharmacology and Toxicology* **34**: 135-172.

Elliott, G.S., Mason, R.W., Ferry, D.G., and Edwards, I.R. (1989) Dothistromin risk assessment for forestry workers. *New Zealand journal of Forestry Science* **19**: 163-170.

Feng, H.G., and Leonard, T.J. (1995) Characterisation of the polyketide synthase gene (*pksL1*) required for aflatoxin biosynthesis in *Aspergillus parasiticus. Journal of Bacteriology* **177**: 6246-6254.

Fernandes, M., Keller, N.P., and Adams, T.H. (1998) Sequence-specific binding by *Aspergillus nidulans* AflR, a C6 zinc cluster protein regulating mycotoxin biosynthesis. *Molecular Microbiology* **28**: 1355-1365.

Franich, R.A., and Gadgil, P.D. (1983) Fungistatic effects of *Pinus radiata* needle epicuticular fatty and resin acids on *Dothistroma pini. Physiological Plant Pathology* **23**: 183-195.

Franich, R.A., Carson, M.J., and Carson, S.D. (1986) Synthesis and accumulation of benzoic acid in *Pinus radiata* needles in response to tissue Injury by dothistromin, and correlation with resistance of *P. radiata* families to *Dothistroma pini. Physiological and Molecular Plant Pathology* **28**: 267-286.

Fretland, A.J., and Omiecinski, C.J. (2000) Epoxide hydrolases: biochemistry and molecular biology. *Chemico-Biological Interactions* **129**: 41-59.

Gadgil, P.D. (1967) Infection of *Pinus radiata* needles by *Dothistroma pini*. *New Zealand Journal of Botany* **5**: 498-503.

Gadgil, P.D. (1970) Survival of *Dothistroma pini* on fallen needles of *Pinus radiata*. *New Zealand Journal of Botany* **8**: 303-309.

Gadgil, P.D. (1984) *Dothistroma* needle blight. *Forest Pathology in New Zealand* **5**: 1-8.

Gallagher, R.T., and Hodges, R. (1972) The chemistry of dothistromin, a difuroanthraquinone from *Dothistroma Pini*. *Australian Journal of Chemistry* **25**: 2399-2407.

Gibson, I.A.S. (1972) *Dothistroma* blight of *Pinus radiata*. *Annual Review of Phytopathology* **10**: 51-72.

Gillman, C. (1996) Identification of dothistromin biosynthetic genes. *MSc thesis, Massey University New Zealand.*

Hamer, J.E., and Timberlake, W.E. (1987) Functional organisation of the *Aspergillus nidulans trpC* promoter. *Molecular and Cellular Biology* **7**: 2352-2359.

Hammer, L., Adachi, K., Montenegro-Chamorro, M.V., Tanzer, M.M., Mahanty, S.K., Lo, C., Tarpey, R.W., Skalchunes, A.R., Heiniger, R.W., Frank, S.A., Darveaux, B.A., Lampe, D.J., Slater, T.M., Ramamurthy, L., DeZwaan, T.M., Nelson, G.H., Shuster, J.R., Woessner, J., and Hammer, J.E. (2001) Gene discovery and gene function assignment in filamentous fungi. *Proceedings of the National Academy of Science of the United States of America* **98**: 5110-5115.

Hirst, P. (1997) Genetic diversity of *dothistromin pini* in new zealand. *MSc thesis, Massey University New Zealand.*

Hopward, D.A., and Sherman, D.H. (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annual Reviews in Genetics* **24**: 67-90.

Itakura, K., Rossi, J.J., and Wallace, R.B. (1984) Synthesis and use of synthetic oligonucleotides. *Annual review of Biochemistry* **53**: 323.

Jelinek, C.F., Pohland, A.E., and Wood, G.E. (1989) Worldwide occurrence of mycotoxins in foods and feeds- an update. *Journal of Association Of Analytical Chemistry* **72**: 223-230.

Jones, W.T., Harvey, D., Jones, S.D., Sutherland, P.W., Nicol, M.J., Sergejew, N., Debnam, P.M., Cranshaw, N., and Reynolds, P.H.S. (1995) Interaction between the phytotoxin dothistromin and *Pinus radiata* embryos. *Phytopathologia* **85**: 1099-1104.

Kawamura, C., Tsujimoto, T., and Tsuge, T. (1999) Targeted disruption of a melanin biosynthesis gene affects conidial development and UV tolerance in the Japanese pear pathotype of *Alternaria alternata*. *Molecular Plant-Microbe Interaction* **12**: 59-63.

Keller, N.P., and Kantz, N.J. (1994) *Aspergillus nidulans verA* is required for production of the mycotoxin sterigmatocystin. *Applied and Environmental Microbiology* **60**: 1444-1450.

Keller, N.P., and Hohn, M.T. (1997) Review- metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology* **21**: 17-29.

Keller, N.P., Watanabe, C.M.H., Kelkar, H.S., Adams, T.H., and Townsend, C.A. (2000) Requirement of monooxygenase-mediated steps for sterigmatocystin biosynthesis by *Aspergillus nidulans*. *Applied and Environmental Microbiology* **66**: 359-362.

Laarakkers, S. (1999) Characterisation of a putative dothistromin biosynthetic cluster. *MSc thesis, Massey University New Zealand.*

Lal, R., Kumari, R., Kaur, H., Khanna, R., Dhingra, N., and Tuteja, D. (2000) Regulation and manipulation of the gene clusters encoding type-1 PKSs. *Trends in Biotechnology* **18**: 264-274.

Mayorga, M.E., and Timberlake, W.E. (1992) The developmentally regulated *Aspergillus nidulans* wA gene encodes a polypeptide homologous to polyketide and fatty acid synthases. *Molecular and General Genetics* **235**: 205-212.

Moller, E.M., Bahnweg, G., Sandermann, H., and Geiger, H.H. (1992) A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies and infected plant tissues. *Nucleic Acids Research* **20**: 6115-6116.

Monahan, B. (1998) Identification of putative dothistromin genes. *MSc thesis, Massey University New Zealand.*

Morgan, B.S. (1997) Isolation of a polyketide synthase gene from *Dothistroma pini*. *MSc thesis, Massey University New Zealand.*

Morisseau, C., Ward, B.L., Gilchrist, D.G., and Hammock, B.D. (1999) Multiple epoxide hydrolases in *Alternaria alternata* f. sp. *lycopersici* and their relationship to medium composition and host-specific toxin production. *Applied and Environmental Microbiology* **65**: 2388-2395.

Moussou, P., Archelas, A., Furstoss, R., and Baratti, J.C. (2000) Clues for the existence of two different epoxide hydrolase activities in the fungus *Beauveria bassiana*. *Enzyme and Microbial Technology* **26**: 414-420.

Murray, N.E.W., Brammar, W.J., and Murray, K. (1977) Lambdoid phages that simplify the recovery of *in vitro* recombinants. *Molecular and General Genetics* **150**: 53-58.

Nuell, M.J., and Fang, G. (1988) Isolation and nucleotide sequence of the chloroperoxidases gene from *Caldariomyces fumago*. *Journal of Bacteriology* **170**: 1007-1011.

Payne, G.A., and Brown, M.P. (1998) Genetics and physiology of aflatoxin biosynthesis. *Annual Review of Phytopathology* **36**: 329-362.

Pinot, F., Cladas, E.D., Schmidt, C., Gilchrist, D.G., Jones, A.D., Winter, C.K., and Hammock, B.D. (1997) Characterization of epoxide hydrolase activity in *Alernaria alternata* f. sp. *lycopersici.* Possible involvement in toxin production. *Mycopathologia* **140**: 51-58.

Pitkin, J.W., Panaccione, D.G., and Walton, J.D. (1996) A putative cyclic peptide efflux pump encoded by the *toxA* gene of the plant pathogenic fungus *Cochliobolus carbonum*. *Microbiology* **142**: 1557-1565.

Punt, P.J.A., and Van Den Hondel, A.M.J.J. (1992) Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. *Gene* **56**: 117-124.

Seconi, J.M. (2001) Confirmation of the presence of a dothistromin biosynthetic gene cluster in the fungal forest pathogen *Dothistroma pini*. *MSc thesis, Massey University New Zealand.*

Shain, L. (1981) Induction of *Dothistroma* blight symptoms With dothistromin. *Physiological Plant Pathology* **19**: 49-55.

Shaw, G.J., Chick, M., and Hodges, R. (1978) A ^{13}C NMR study of the biosynthesis of the anthraquinone dothistromin by *Dothistroma pini*. *Phytochemistry* **17**: 1743-1745.

Shimizu, K., Takana, C., and Tsuda, M. (1997) Cloning of *Brn1* a reductase gene involved in melanin biosynthesis in *Cochliobolus heterostrophus*. *Journal of General and Applied Microbiology* **43**: 145-150.

Shiotani, H., and Tsuge, T. (1995) Efficient gene targeting in the filamentous fungus *Alternaria alternata*. *Molecular and General Genetics* **248**: 142-150.

Skory, C.D., Chang, P., Cary, J., and Linz, J.E. (1992) Isolation and characterisation of a gene from *Aspergillus parasiticus* associated with the conversion of versicolorin A to sterigmatocystin in aflatoxin biosynthesis. *Applied and Environmental Microbiology* **58**: 3527-3537.

Stoessl, A., Arbramowski, Z., Lester, H.H., Rock, G.L., and Towers, G.H.N. (1990) Further toxic properties of the fungal metabolite dothistromin. *Mycopathologia* **112**: 179-186.

Takano, Y., Kubo, Y., Shimizu, K., Mise, K., Okuno, T., and Furusawa, I. (1995) Structural analysis of *PKS1* a polyketide synthase gene involved in melanin biosynthesis in *Collectotrichum lagenarium*. *Molecular and General Genetics* **249**: 162-167.

Tsai, H., Chang, Y.C., Washburn, R.G., Wheeler, M.H., and Kwon-Chung, K.J. (1998) The developmentally regulated *alb1* gene of *Aspergillus fumigatus:* Its role in modulation of conidial morphology and virulence. *Journal of Bacteriology* **180**: 3031-3038.

Upchurch, R.G., Walker, D.C., Rollins, J.A., Ehrenshaft, M., and Daub, M.E. (1991) Mutants of *Cercospora kikuchii* altered in cercosporin synthesis and pathogenicity. *Applied and Environmental Microbiology* **57**: 2940-2945.

White, T.J., Bruns, T.D., Lee, S.B., and Taylor, J.W. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenies. Innis, G., Sninsky, White (ed). PCR protocols: A guide to methods and applications: Academic Press, New York, pp. 315-322.

Yabe, K., Chihaya, N., Hamamatsu, S., Sakuno, E., Hamasaki, T., Nakajima, H., and Bennett, J., W (2003) Enzymatic conversion of averufin to hydroxyversicolorone and elucidation of a novel metabolic grid involved in aflatoxin biosynthesis. *Applied and Environmental Microbiology* **69**: 66-73.

Yang, G., and Turgeon, B.G. (1994) Toxin-deficient mutants from a toxin-sensitive transformant of *Cochliobolus heterostrophus*. *Genetics* **137**: 751-757.

Yu, J., Chang, P., Cary, J.W., Wright, M., Bhatnagar, D., Cleverland, T.E., Payne, G.A., and Linz, J.E. (1995) Comparative mapping of aflatoxin pathway gene clusters in *Aspergillus parasiticus* and *Aspergillus flavus*. *Applied and Environmental Microbiology* **61**: 2365-2371.

Yu, J., and Leonard, T.J. (1995) Sterigmatocystin biosynthesis in *Aspergillus nidulans* requires a novel type 1 polyketide synthase. *Journal of Bacteriology* **177**: 4792-4800.

Yu, J., Chang, P., Bhatnagar, D., and Cleverland, T.E. (2000a) Genes encoding cytochrome P450 and monooxygenase enzymes define one end of the aflatoxin pathway gene cluster in *Aspergillus parasiticus*. *Applied and Microbiological Biotechnology* **53**: 583-590.

Yu, J., Woloshuk, C.P., Bhatnagar, D., and Cleverland, T.E. (2000b) Cloning and characterisation of *avfA* and *omtB* genes involved in aflatoxin biosynthesis in three *Aspergillus* species. *Gene* **248**: 157-167.

Yu, J., Chang, P., Ehrlich, K.C., Cary, J.W., Bhatnagar, D., Cleverland, T.E., Payne, G.A., Linz, J.E., Woloshuk, C.P., and Bennett, J., W (2004) Clustered pathway genes in aflatoxin biosynthesis. *Applied and Environmental Microbiology* **70**: 1253-1262.

APPENDIX 1: FLANKING REGIONS OF *PKS*^{DOT} USED FOR THE *PKS*^{DOT} DISRUPTION VECTOR

Flanking regions for the disruption vector were designed in order to replace the KS and AT domains with the *hph* cassette.

Primers pksCB2 and gateway pksCB2
OT pks KO/2 and gateway OT pks KO/2
OT pks KO/3 and gateway OT pks KO/3
OT pks KO/4 and gateway OT pks KO/4

KS binding domain is shown in blue

CCGATCTTGACGGCTTTCCTGGAACAGTCTCATTATGTCGTGTAAGTCGTACTGATCGATCGGA
CGCAGACCGTTGCTAAGAATTGCCAGCCGAGCACAGATGATCCAGACACTGCCTCCGGCGGAGC
ACAAGGCAGCTCGAACCTTCGACTTAGCAGATATGCTGAAGAAGTATGTCGCCGGCAAGCTGAA
CCCTGCCTTCCAGACGGCCCTCAGCTGCATCACACAACTCGGTGTGTTCATGCGAGAGTTCCAT
GACTTCACCAAGCCATATCCACGACACGATAGCAGCTACGTGTTGGGTATCTGCACCGGCTCGC
TTGCCGCTGCGGCCGTCAGCTCCAGCAACTCTTTATCCGAGCTTCTGCCTATTGCCGTGCAAAC
GGCTTTGATCGCCTTTCGCCTCGGTCTGTGCGTCACAGACATGCGAGATCGTCTCGAAAGCTCG
GAAGAAGACCGCACCCAGCCTTGGTCGGTAGTTCTGTTCGACACAGACGAGCAGACTGTTACCA
AGGCCATCAAAGACTTCTGCACATCCAACGTCCTCCCGAAGACCAAGCAGCCCTGGATTACCTC
CGCGTCATCGAAGACAATCTCCATCAGCGGAGCACCACGTGTGCTGAAGAAGTTGTCGCAAGAG
CCTGCGCTCAAGGACAAAAAGACCAGACAGATCCCGATTTATGTCCCAGCACACAATTCTGCCC
TCTTCACCCCCGAAGATGTCAAGTCTATCCTTGAGACTACCCCTGTCGACACTTGGAGCAACTA
CCCTACCAAGCTTCCATTCATCTCGAGCGTTTCTGGCAAGATGGCTTGGGCAGACAACTACCTT
GCAGTAATCCATCTCGCCCTCAATCAGTGCCTGCTCGAGAGCATTGGCTGGGGTAAGGTCGAGA
CTGAGCTCCCAAGGCTCCTCAAGTCTCGCGGCGCAGAAAATGTGCTCATCACGCCAATCACCAC
CTCTGCCGACCGTGCTCTGTCGGCTGCACTCAGCCCAACGATCTCCAATATCGAGGTCGAGAAG
CCCACAATCAACGAATCTTTCGCCCATAGACCTGGTTCGGGAAAGAGCAAGCTTGCTATCGTCT
CAATGTCTGGTCGCTTCCCAGAGGCACAAAGCACCGACGCTTTTTGGGATCTGCTCTACAAGGG
GCTTGACGTGGTCAAGGAAGTGCCCAAACGTCGTTGGGACGTCGAGACTCACGTCGACCCAACT
GGGCGTGCTCGCAACAAGGGTGCGACCAAATGGGGCTGCTGGCTCGATTTCGCCGGAGAATTCG
ACCCTCGCTTCTTCAGCATCTCACCCAAGGAAGCACCACAAATGGACCCCGCCCAGCGTATGGC
TTTGATGTCGACCTGGGAAGCGATGGAACGTGGTGGCATCGTCCCGGATACTACGCCCTCAACA
CAAAGAAATCGAATTGGTGTCTTCCATGGCGTTACCTCCAACGACTGGATGGAGACCAACACGG
CTCAAAACATTGATACCTACTTCATCACCGGCGGTAACCGCGGTTTCATCCCCGGCCGTATTAA
CTTCTGCTTTGAATTCTCTGGACCCAGCTTCACCAATGACACGGCCTGCTCCAGTTCGCTCGCA
GCGATCCATT

5' flanking region

GGAGTCTCGCCGAGATCACGGCGGGCAAGAATGTCGAAGTCGCATGCGTTAATGGGCCTGAAGAC
ACTGTCCTCTCCGGCACTGTCGAGGAAATTGGAGAGGTGCAGAAGACACTCATTGCGAAGAGCCAT
CAAGGCTACACTCTTGAAGTTGCCCTTCGCGTTTCATTCGGCGCAGGTACAACCTATCCTCCGAG
GACGTTCTGAAGAACTTGCGGCTGGAGCTACTTTTGAGAAGCCCAAGCTTGCGGTCATTTCCCCG
CTACTGGGCAGTGTGGTCGACGACGAAGGAGTCGTTGGACCCAACTACCTTGCACGCCACTGCCG
TGAGGCGGTCGGAATGGTCAAAGCCCTCGGAGTGGCGAAGGAGAAGGGTATAATCAACGAGAAGA
CCTTCGTCATTGAGATTGGTCCTAAGCCGCTTCTCTGCGGAATGATCAAGAACATACTCGGCCAG
AACATCGTAGCCTTGCCTACGTTGAAGGACAAGGGTCCAGACGTCTGGCAGAACCTCTCGAACAT
CTTCACGACGCTCTACACCGGTGGTCTAGACATCAACTGGACTGCCTTCCACGCCCCCTTCGAGC
CCGCGAAGAAGGTCCTGCAACTTCCTGATTATGGCTGGGATCTCAAGGATTACTTCATCCAGTAT
GAAGGCGATTGGGTTCTGCATCGGCACAAGATCCACTGCAACTGTGCAGATGCTGGAAAGGATGT
GCATAACACTTCGCACTACTGTCCTGGCAAACACACCTTCGCTGAGAATGTTGTCGTTCCTGGTG
GGGCTCAGAAGGCCGTTCAGGAAGCACCTGCGGCGAAGACAGAGACGAAGAAGATGTCGAAGCTG
GATCCTACCAAGGAGGCGTATCCGGGCATTCCGCTCACCACGACCGTGCACAAGGTCATTGAAGA
GAAGACGGAGCCTCTGGGAGCGCAGTTCACGGTCGAGACGGATATCTCCCGCAAGGATGTCAACA
GCATCGCTCAAGGTCACACTGTTGACAGCATTCCCCTCTGCACGCCATCATTCTATGCGGATATT
GCGCTTCAGGTGTTACGGATCGAGTTGGCAAATACGCCATGGACCGCATCCGTGCTGGACATCCC
GGTGCCGGTGCTATTGATGGAAGGGTCGA

3' flanking region

APPENDIX 2: PARTIAL SEQUENCE OF ΛKSA INCORPORATING pR162A, pR181, pR163, pR156 AND OVERLAPS

Nucleotides 1- 5000 bp sequenced in this study

```
         10        20        30        40        50        60        70
         |         |         |         |         |         |         |
CGAGCTCGGTACCCGGGGATCCAGCATAGAATCCACCCCACATGCCGTTCACCTGCTCTTCGATCTCTCT
GCTCGAGCCATGGGCCCCTAGGTCGTATCTTAGGTGGGGTGTACGGCAAGTGGACGAGAAGCTAGAGAGA

         80        90        100       110       120       130       140
         |         |         |         |         |         |         |
GGCGTGAGCAACGACGGCTTTGGGATCGCGACGGAAGTTCTCTCGCTCTTCTTCGCTGTACTCTTTGGCT
CCGCACTCGTTGCTGCCGAAACCCTAGCGCTGCCTTCAAGAGAGCGAGAAGAAGCGACATGAGAAACCGA

         150       160       170       180       190       200       210
  Mox-dot 3' end →|         |         R162A Rev Out →       |         |
TGACTACCAGAGTTTCCGGCTATCACCCCGAACCAGACTCCCGTACGGACGAAGATGTCGAGGTGTTTTG
ACTGATGGTCTCAAAGGCCGATAGTGGGGCTTGGTCTGAGGGCATGCCTGCTTCTACAGCTCCACAAAAC

         220       230       240       250       260       270       280
         |         |         |         |         |         |         |
CGTATGGCTGCATGCCTGGAACAGTCTGTATGGATGAAGCACCAGATCCAATTACGGCGACTCGCTCCTT
GCATACCGACGTACGGACCTTGTCAGACATACCTACTTCGTGGTCTAGGTTAATGCCGCTGAGCGAGGAA
                  OT R162A Fwd7 ←

         290       300       310       320       330       340       350
         |         |         |         |         |         |         |
TGCCCAGTCTTCCTCCTTGTAGTCCTTTGGCCACCGGGCGGTGTGAATGACCCGGCCCTTGAAGCGATCC
ACGGGTCAGAAGGAGGAACATCAGGAAACCGGTGGCCCGCCACACTTACTGGGCCGGGAACTTCGCTAGG

         360       370       380       390       400       410       420
         |         |         |         |         |         |         |
TGCAGGCCTGGAATATCGGGGAACTGGTGGCTCACGTTAGCACAAATCTCGACTCCGGACTGTGTCAAAC
ACGTCCGGACCTTATAGCCCCTTGACCACCGAGTGCAATCGTGTTTAGAGCTGAGGCCTGACACAGTTTG

         430       440       450       460       470       480       490
         |         |         |         |         |         |         |
ATACCTTGAAATTATTCAGCACGCCAGCGCCATACAACAGCACATTACATCTATCCTCGAACTCCCGGAC
TATGGAACTTTAATAAGTCGTGCGGTCGCGGTATGTTGTCGTGTAATGTAGATAGGAGCTTGAGGGCCTG

         500       510       520       530       540       550       560
         |         |         |         |         |         |         |
CTCGTGGTTTGGAAGATGTTCTCGAAGCTTGACCACCCATTCGCCCGCATGCTCCTGCCAGTAGCACCCC
GAGCACCAAACCTTCTACAAGAGCTTCGAACTGGTGGGTAAGCGGGCGTACGAGGACGGTCATCGTGGGG

         570       580       590       600       610       620       630
         |         |         |         |         |         |         |
ACAACCTCCACGTGGAACCTCATGTACTTCTTCAAATCGAACGTCTCGCAGACTTTGTTCAAGTACGCCC
TGTTGGAGGTGCACCTTGGAGTACATGAAGAAGTTTAGCTTGCAGAGCGTCTGAAACAAGTTCATGCGGG

         640       650       660       670       680       690       700
         |         |         |         |         |         |         |
AGATGTCCGGAGCGAAGGAAAAGAATCTCGGCCAGTCAGGGTTCTACCTGTCCATCAGCAAACCATTACT
TCTACAGGCCTCGCTTCCTTTTCTTAGAGCCGGTCAGTCCCAAGATGGACAGGTAGTCGTTTGGTAATGA
```

```
                                OT R162A Fwd6 ←
         710       720       730       740       750       760       770
    OT R162A Rev1 →  |         |         |         |         |         |
ATTCTCGACAACAATCGTCCATACCAAAGCAAACTGGTACGTATAAGCATGACTCGGAATATCACACCCC
TAAGAGCTGTTGTTAGCAGGTATGGTTTCGTTTGACCATGCATATTCGTACTGAGCCTTATAGTGTGGGG

         780       790       800       810       820       830       840
         |         |         |         |         |         |         |
GCTCGAGGATATCGATTCTCCAGCCACGTTCCTGTCACAGTCAGCATCTCCTCTCCCATCTTCGCCAGCT
CGAGCTCCTATAGCTAAGAGGTCGGTGCAAGGACAGTGTCAGTCGTAGAGGAGAGGGTAGAAGCGGTCGA

         850       860       870       880       890       900       910
         |         |         |         |         |         |         |
TTTCCCATCTCAATTATTTCAAGTGTTTCACACTCACCACCAACATCCTCATTCTTCTCATAAACCACAT
AAAGGGTAGAGTTAATAAAGTTCACAAAGTGTGAGTGGTGGTTGTAGGAGTAAGAAGAGTATTTGGTGTA

         920       930       940       950       960       970       980
         |         |         |         |         |         |         |
GTTCCACATTCTCACAATGCTTCTGCAACTGGTACGCCATCAAAATTCCTGAAATTCCCGCACCGATTGT
CAAGGTGTAAGAGTGTTACGAAGACGTTGACCATGCGGTAGTTTTAAGGACTTTAAGGGCGTGGCTAACA
                                                  Mox-dot 5' end ←

         990       1000      1010      1020      1030      1040      1050
         |         |         |         |         |         |         |
GAGGACGCGGATTTTGCGATTGGAGGGGGCGTTCCAGGTCGTGTTTTGCGGGATATTGTAGCCGGTTGAG
CTCCTGCGCCTAAAACGCTAACCTCCCCCGCAAGGTCCAGCACAAAACGCCCTATAACATCGGCCAACTC

         1060      1070      1080      1090      1100      1110      1120
         |         |         |         |         |         |         |
CCCGGAGTCGAGTAGTCTACGTGTTGGTTGCGGGTATGGCGACTTGGGGTTGGGGAGGAGGATGATGATG
GGGCCTCAGCTCATCAGATGCACAACCAACGCCCATACCGCTGAACCCCAACCCCTCCTCCTACTACTAC

         1130      1140      1150      1160      1170      1180      1190
         |         |         |         |         |         |         |
AGCTGGCGCTCTCGCCATGAGCGGAAAGGAAGGGAGCCATAGTGGAGTGATTCAACTAGGAGAGCGGACA
TCGACCGCGAGAGCGGTACTCGCCTTTCCTTCCCTCGGTATCACCTCACTAAGTTGATCCTCTCGCCTGT

         1200      1210      1220      1230      1240      1250      1260
         |  OT R162A Rev2 →  |         |         |         |         |
GGGTGATGTGTGGTAACGAGCGTTGCAAAATTGATGGTTGGGTGCAAGGCAGGGTGCTAATCAAGGACAC
CCCACTACACACCATTGCTCGCAACGTTTTAACTACCAACCCACGTTCCGTCCCACGATTAGTTCCTGTG

         1270      1280      1290      1300      1310      1320      1330
         |         |         |         |         |         |         |
AGAGATCGATTAGTTGAGTAGCGAGCTTTGAACAACAAGACAAGACAAGACGAAACGGGGAGAGTATACA
TCTCTAGCTAATCAACTCATCGCTCGAAACTTGTTGTTCTGTTCTGTTCTGCTTTGCCCCTCTCATATGT

         1340      1350      1360      1370      1380      1390      1400
         |         |         |         |         |         |         |
CGGCGCGATTGTATCTTTCCCATGACACTGAGCAGCCATGACTCCATAGCTGGTCAATGCCAACATCGGG
GCCGCGCTAACATAGAAAGGGTACTGTGACTCGTCGGTACTGAGGTATCGACCAGTTACGGTTGTAGCCC
OT R162A Fwd5 ←

         1410      1420      1430      1440      1450      1460      1470
         |         |         |         |         |         |         |
CGACGAGTGCAGGCAGTCGGTGGTCGAAGCATCACAGCCTTCCGAAACCCGAGCACTTTAGCCCGAGCCA
GCTGCTCACGTCCGTCAGCCACCAGCTTCGTAGTGTCGGAAGGCTTTGGGCTCGTGAAATCGGGCTCGGT
```

```
         1480       1490       1500       1510       1520       1530       1540
         |          |          |          |          |          |          |
TCTAGTTCCGAGTTGCACTAGCCGAATGAGGAAGGATATGGATTCCTATGGGCAGACGTCAGCACATCAC
AGATCAAGGCTCAACGTGATCGGCTTACTCCTTCCTATACCTAAGGATACCCGTCTGCAGTCGTGTAGTG

         1550       1560       1570       1580       1590       1600       1610
         |          |          |          |          |          |          |
TGCTGCCACGATTCAGTCGTAGACCTCGTCGCCGCTGTAAAGCTGCCCATACAACACCATATTGTCACAG
ACGACGGTGCTAAGTCAGCATCTGGAGCAGCGGCGACATTTCGACGGGTATGTTGTGGTATAACAGTGTC

         1620       1630       1640       1650       1660       1670       1680
         |          |          |          |          |          |          |
CCATCGTCTTCTTGGCACCAACACAGAATACTTATCCCTCAACTAAATCAACCTCATCACCACCAACACC
GGTAGCAGAAGAACCGTGGTTGTGTCTTATGAATAGGGAGTTGATTTAGTTGGAGTAGTGGTGGTTGTGG

         1690       1700       1710       1720       1730       1740       1750
         |          |          |          |          |          |          |
ACAATCGATATGGAAGGCTATACCACCCTCCCCTCCACCGCGACTCTCAAGCCCAGTCCCTTCACCGTCT
TGTTAGCTATACCTTCCGATATGGTGGGAGGGGAGGTGGCGCTGAGAGTTCGGGTCAGGGAAGTGGCAGA

         1760       1770       1780       1790       1800       1810       1820
 OT R162A Rev3 →     |          |          |          |          |          |
CAATTTCCGAATCCAAACTCCAAACCCTCCAAGACCTGATCCGCCTCTCCCCCATCGGACCCGCAGACTA
GTTAAAGGCTTAGGTTTGAGGTTTGGGAGGTTCTGGACTAGGCGGAGAGGGGGTAGCCTGGGCGTCTGAT

         1830       1840       1850       1860       1870       1880       1890
         |          |          |          |          |          |          |
CAACAATTCCTCCCCTTCCACTGGCTCAAAGTACGGTATCCGCCGTGACTGGCTCATCAATGCAAAGAAA
GTTGTTAAGGAGGGGAAGGTGACCGAGTTTCATGCCATAGGCGGCACTGACCGAGTAGTTACGTTTCTTT

         1900       1910       1920       1930       1940       1950       1960
         |          |          |          |          |          |          |
CAATGGGAAGACAACTTCTCCTGGCGCACTTTCGAGAAGAAGTTGAAGAAGTACCCGCAGTATACCGTCC
GTTACCCTTCTGTTGAAGAGGACCGCGTGAAAGCTCTTCTTCAACTTCTTCATGGGCGTCATATGGCAGG
                OT R162A Fwd4 ←

         1970       1980       1990       2000       2010       2020       2030
         |          |          |          |          |          |          |
CTGTGAAGGGAGAGAGTGGGGAGACGATCGAAATTCACTTCATCGCGCTGTTCAGTCAAAGGCAAGACGC
GACACTTCCCTCTCTCACCCCTCTGCTAGCTTTAAGTGAAGTAGCGCGACAAGTCAGTTTCCGTTCTGCG

         2040       2050       2060       2070       2080       2090       2100
         |          |          |          |          |          |          |
AAGGCCGTTGGCGTTTTACCATGGATGGCCCTCTTCCCCGTTCGATTTTCTGCCGATTCTGGATCTTTTG
TTCCGGCAACCGCAAAATGGTACCTACCGGGAGAAGGGGCAAGCTAAAAGACGGCTAAGACCTAGAAAAC

         2110       2120       2130       2140       2150       2160       2170
         |          |          |          |          |          |          |
ACAAACAAATACACGCCCGAGACACTGCCGTATCACATCATCGTCCCGAGTCTGCCGGGTTTCTGCTTCA
TGTTTGTTTATGTGCGGGCTCTGTGACGGCATAGTGTAGTAGCAGGGCTCAGACGGCCCAAAGACGAAGT

         2180       2190       2200       2210       2220       2230       2240
         |          |          |          |          |          |          |
GCGGCTCCCCACCCATAGATTTGGATTACGACATGCCCCAAGCAGCCTACCTCCTCAACAACCTCATGAT
CGCCGAGGGGTGGGTATCTAAACCTAATGCTGTACGGGGTTCGTCGGATGGAGGAGTTGTTGGAGTACTA
```

```
        2250      2260      2270      2280      2290      2300      2310
    OT R162A Rev4 →         |         |         |         |         |
CGGCCTTGGATTGGATGGGTACATTGCCCAAGGCGGCGATCTGGGATCGGGGATCTCGAGGGAGCAGGCC
GCCGGAACCTAACCTACCCATGTAACGGGTTCCGCCGCTAGACCCTAGCCCCTAGAGCTCCCTCGTCCGG

        2320      2330      2340      2350      2360      2370      2380
        |         |         |         |         |         |         |
GCGGGTTGTGAAGCTTGCAAGGGATTTCATCTTAATATGATTTTGCTGCCGCCGCCGGCGAACATGAAAG
CGCCCAACACTTCGAACGTTCCCTAAAGTAGAATTATACTAAAACGACGGCGGCGGCCGCTTGTACTTTC

        2390      2400      2410      2420      2430      2440      2450
        |         |         |         |         |         |         |
AGTTGACGTTGGAGGAGGTGGAGAAGAAGGCTATGCCCAACGCCCTGGCGTTCCGTCAAAGTGGCATGGC
TCAACTGCAACCTCCTCCACCTCTTCTTCCGATACGGGTTGCGGGACCGCAAGGCAGTTTCACCGTACCG

        2460      2470      2480      2490      2500      2510      2520
        |         |         |         |         |         |         |
CTATGCCCTGGAGCACGGAACCCGCGGCGGCACGATTGGATTGGCATTGCAGGCGAGTCCCGTGGCGCTG
GATACGGGACCTCGTGCCTTGGGCGCCGCCGTGCTAACCTAACCGTAACGTCCGCTCAGGGCACCGCGAC
OT R162A Fwd3 ←

        2530      2540      2550      2560      2570      2580      2590
        |         |         |         |         |         |         |
TTATGTTGGATCGGGGAGAAGATGATGGCCTGGAGTGATTCCTCTTCTCAGCCGTCTCTCGAACAAATCT
AATACAACCTAGCCCCTCTTCTACTACCGGACCTCACTAAGGAGAAGAGTCGGCAGAGAGCTTGTTTAGA

        2600      2610      2620      2630      2640      2650      2660
        |         |         |         |         |         |         |
TGGAGACCGTCTCCCTATACTGGCTCACCGATTCCATCACGCGGGGTCTTTACCCCTACCGTCGCTTCGC
ACCTCTGGCAGAGGGATATGACCGAGTGGCTAAGGTAGTGCGCCCCAGAAATGGGGATGGCAGCGAAGCG

        2670      2680      2690      2700      2710      2720      2730
        |         |         |         |         |         |         |
TTCTGGTAACGAACCCAAAATTAATTTCATCGAGAAGCCGCTGGGGTATTCGTTTTTCCCGAATACGTAC
AAGACCATTGCTTGGGTTTTAATTAAAGTAGCTCTTCGGCGACCCCATAAGCAAAAAGGGCTTATGCATG

        2740      2750      2760      2770      2780      2790      2800
        |         |         |         |         |    OT R162A Rev5 →
CTCCCTTGCCCTGTGAGCTGGGCGAAGACGACGGCGAATTTGGTCCAGTACCGCAGACATGAGAGTAAGT
GAGGGAACGGGACACTCGACCCGCTTCTGCTGCCGCTTAAACCAGGTCATGGCGTCTGTACTCTCATTCA

        2810      2820      2830      2840      2850      2860      2870
        |         |         |         |         |         |         |
CCAGTTGAATGTTGGTTTGAGTGTACTTGCCGACTGACGTGTGTAGGTGGAGGTCATTTCGCGCCTTGGG
GGTCAACTTACAACCAAACTCACATGAACGGCTGACTGCACACATCCACCTCCAGTAAAGCGCGGAACCC
                                                  Epoxide 3' end ←

        2880      2890      2900      2910      2920      2930      2940
        |         |         |         |         |         |         |
AGAGGCCAAGGGAGTTGTTGGAAGATGTGGAGGAGTATGTTGATGTGGCGTTCGGGAAGAAGGATTCCCC
TCTCCGGTTCCCTCAACAACCTTCTACACCTCCTCATACAACTACACCGCAAGCCCTTCTTCCTAAGGGG

        2950      2960      2970      2980      2990      3000      3010
        |         |         |         |         |         |         |
GATGATGGGTCCGAAAGCTGTGGAAGATGTCAGCGGGAGTGGAAGCCATGCAAGAGGGTTGTAGATTGCG
CTACTACCCAGGCTTTCGACACCTTCTACAGTCGCCCTCACCTTCGGTACGTTCTCCCAACATCTAACGC
```

```
         3020       3030       3040       3050       3060       3070       3080
         |          |          |          |          |          |          |
GAGGCTCAGCTCTTGATTGATGGTACATACCATGGGGCGTTTCATCCCGGCTTTACGAAAAAGAGCTGAA
CTCCGAGTCGAGAACTAACTACCATGTATGGTACCCCGCAAAGTAGGGCCGAAATGCTTTTTCTCGACTT

         3090       3100       3110       3120       3130       3140       3150
         |          |          |          |          |          |          |
ACTCCGAGAGTCATCTTGAAAATTCGAAGTTTTGACATGCGACTACATCAAGACTCGAACAAGCCTTTTT
TGAGGCTCTCAGTAGAACTTTTAAGCTTCAAAACTGTACGCTGATGTAGTTCTGAGCTTGTTCGGAAAAA
 OT R162A Fwd2 ←

         3160       3170       3180       3190       3200       3210       3220
         |          |          |          |          |          |          |
AACGCTGCCAATATGGTTTCTGTCGCACTCCTATCTGTTCATCGTTCCCGTATACTGCACCAAGACTACC
TTGCGACGGTTATACCAAAGACAGCGTGAGGATAGACAAGTAGCAAGGGCATATGACGTGGTTCTGATGG

         3230       3240       3250       3260       3270       3280       3290
         |          |          |          |          |          |          |
GCGGCCAAACTCACCATAACTGTCGTAAACACCGTCAGCGCACTGCTGCCGTTCAATCGTCGAACGAGGA
CGCCGGTTTGAGTGGTATTGACAGCATTTGTGGCAGTCGCGTGACGACGGCAAGTTAGCAGCTTGCTCCT

         3300       3310       3320       3330       3340       3350       3360
         |          |          |          |          |          |          |
GACTACTACATACCGCAATAAAGAAACAAATTTTGCGGCTTGCTCATTGGCCTCTCCTCTTCCCACCAAG
CTGATGATGTATGGCGTTATTTCTTTGTTTAAAACGCCGAACGAGTAACCGGAGAGGAGAAGGGTGGTTC

         3370       3380       3390       3400       3410       3420       3430
         |       OT R162A Rev6 →|          |          |          |          |
CTCTCGCCCTCCCCTTCGCCCCATCGAACAAGAATCCCGCCAGCACTCCCCACGTCTCCTTGGCCTTCCC
GAGAGCGGGAGGGGAAGCGGGGTAGCTTGTTCTTAGGGCGGTCGTGAGGGGTGCAGAGGAACCGGAAGGG

         3440       3450       3460       3470       3480       3490       3500
         |          |          |          |          |          |          |
CGTCGCCGTCACACCCACAGGTTGGTTAAGGAAGTCTTCCTTTCGGCTTGCGATCTCAACGAATCCAGCA
GCAGCGGCAGTGTGGGTGTCCAACCAATTCCTTCAGAAGGAAAGCCGAACGCTAGAGTTGCTTAGGTCGT

         3510       3520       3530       3540       3550       3560       3570
         |          |          |          |          |          |          |
CCGAGATCGGCGTAGCTCAGTGCCGTCTCCTGCTTGTCGCAGACGTCGGGTTTGCAGGAGATTAGCTTGT
GGCTCTAGCCGCATCGAGTCACGGCAGAGGACGAACAGCGTCTGCAGCCCAAACGTCCTCTAATCGAACA

         3580       3590       3600       3610       3620       3630       3640
         |          |          |          |          |          |          |
GTCCCGTGCGATTAGGCCCGAAAGCATCGTGGATCGTGGGTGGGTCGACGTAGATATACGACAACAGGCC
CAGGGCACGCTAATCCGGGCTTTCGTAGCACCTAGCACCCACCCAGCTGCATCTATATGCTGTTGTCCGG

         3650       3660       3670       3680       3690       3700       3710
         |          |          |          |          |          |          |
TTTGGCTGCAGCGGCCTCGTAGTGTTCGCAGGCTTTGACGATGTCGAGATGGCTGTAGTGGAGACAGAAG
AAACCGACGTCGCCGGAGCATCACAAGCGTCCGAAACTGCTACAGCTCTACCGACATCACCTCTGTCTTC

         3720       3730       3740       3750       3760       3770       3780
         |          |       OT R162A Rev7 →|          |          |          |
GAGACGATGTTGTAGACCAGGCGTGGTACCTGACATGAGAGCTTAGGGTTGAGGGACGCGCTGCGGAGCT
CTCTGCTACAACATCTGGTCCGCACCATGGACTGTACTCTCGAATCCCAACTCCCTGCGCGACGCCTCGA
```

```
         3790      3800      3810      3820      3830      3840      3850
         |         |         |         |         |         |         |
GGAGGATCGTCGGTGCGTTGTAGGCGCTGCCGTGGAGTTTGCGGAGCATCCCCAGGGTGGTGACGATAGC
CCTCCTAGCAGCCACGCAACATCCGCGACGGCACCTCAAACGCCTCGTAGGGGTCCCACCACTGCTATCG

         3860      3870      3880      3890      3900      3910      3920
         |         |         |         |         |         |         |
GGTGACGGTGTCGGCGGTGAGAGAGACACCTTTGTTGGAGGCATTGTCGGCGACGCACATGAAGGCTACG
CCACTGCCACAGCCGCCACTCTCTCTGTGGAAACAACCTCCGTAACAGCCGCTGCGTGTACTTCCGATGC
                                   OT R162A Fwd1 ←

         3930      3940      3950      3960      3970      3980      3990
         |         |         |         |         |         |         |
GACGCGTCTTCGAGACATTGTTGAAGAGCGATGGTGTCGGTGGAGTTGCCTTGGATGATGTGGATGCGGG
CTGCGCAGAAGCTCTGTAACAACTTCTCGCTACCACAGCCACCTCAACGGAACCTACTACACCTACGCCC

         4000      4010      4020      4030      4040      4050      4060
         |         |         |         |         |         |         |
GAGAGATTGTTGTAGTGAGGGTTGGGAAGGACTTGAGGAGCTTTTGTTTGGAGCGGACGAGGATGTTGAG
CTCTCTAACAACATCACTCCCAACCCTTCCTGAACTCCTCGAAAACAAACCTCGCCTGCTCCTACAACTC

         4070      4080      4090      4100      4110      4120      4130
         |         |         |         |         |         |         |
GTCGAGGTCTGGGGGTGGTGAGGCAAGGAGGCAGCGGAGTATGGCCGAGCCTGTGGCGCCTGTCGCGCCG
CAGCTCCAGACCCCCACCACTCCGTTCCTCCGTCGCCTCATACCGGCTCGGACACCGCGGACAGCGCGGC

         4140      4150      4160      4170      4180      4190      4200
         |         |         |         |         |         |         |
AGCAGGGCATAGGTGGGCATATTGATGGCTTGATCTGGTTATTATGAACGATGAGCTGTGATGAAGCTGT
TCGTCCCGTATCCACCCGTATAACTACCGAACTAGACCAATAATACTTGCTACTCGACACTACTTCGACA
          R162A Fwd Out ←

         4210      4220      4230      4240      4250      4260      4270
         |         |         |         |         |         |         |
TCGTTCACGGGTGCGAGGTCAAGGTCGTTCAGAGCAGTGCGAGTGCCAGTGCCAGTGCGAGTGCCAGTGT
AGCAAGTGCCCACGCTCCAGTTCCAGCAAGTCTCGTCACGCTCACGGTCACGGTCACGCTCACGGTCACA

         4280      4290      4300      4310      4320      4330      4340
         |         |         |         |         |         |  BamHI
CTGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCTGTGCCACGTCAAGGATCCTA
GACACGGTCACGGTCACGGTCACGGTCACGGTCACGGTCACGGTCACGGACACGGTGCAGTTCCTAGGAT

         4350      4360      4370      4380      4390      4400      4410
         |         |         |         |         |         |         |
CCACGTCAGTGTCTTCTTCGGCAAATATTCTTTGGTCGGAGGTGGGAGAGCAGTCTGCAATAGCCATTCG
GGTGCAGTCACAGAAGAAGCCGTTTATAAGAAACCAGCCTCCACCCTCTCGTCAGACGTTATCGGTAAGC

         4420      4430      4440      4450      4460      4470      4480
         |  BamHI  |         |         |         |         |         |
AAGCTGTAATTTTCGGGATCCCGAAGGACGTCAGTCGGTGCGAAGGCGATGCGGGAAGATGCAGACGGCT
TTCGACATTAAAAGCCCTAGGGCTTCCTGCAGTCAGCCACGCTTCCGCTACGCCCTTCTACGTCTGCCGA

         4490      4500      4510      4520      4530      4540      4550
         |         |         |         |         |         |         |
GGAAATGTCAAGTCCGAGAGCCTACCAGTGGATGCTACCGCGTGTCGGGCAGGGATCGCAGATGGGTAAT
CCTTTACAGTTCAGGCTCTCGGATGGTCACCTACGATGGCGCACAGCCCGTCCCTAGCGTCTACCCATTA
```

```
         4560       4570       4580       4590       4600       4610       4620
         |          |          |          |          |          |          |
ATGAATCGTGGGCCATGTCCGAGCCGTCACGACTATTCGTGTGTTCTTCAACGGATGTCTGCCATCATGA
TACTTAGCACCCGGTACAGGCTCGGCAGTGCTGATAAGCACACAAGAAGTTGCCTACAGACGGTAGTACT
                                       R181 Ovlp ←

         4630       4640       4650       4660       4670       4680       4690
         |          |          |          |          |          |          |
ATGGGCCTAAGTATGGCGCGCATTGCCAGAGGATAAGGGGGGTTCTGCTGTCCATAGGAAGAAGCGACGT
TACCCGGATTCATACCGCGCGTAACGGTCTCCTATTCCCCCCAAGACGACAGGTATCCTTCTTCGCTGCA

         4700       4710       4720       4730       4740       4750       4760
         |          |          |          |          |    R181 Fwd 2 →    |
TGCACACCAACAACACCGACAATCCTCGATGCATGTGCACATGTCCCCGGAGTCCGGCGATGCACAATGC
ACGTGTGGTTGTTGTGGCTGTTAGGAGCTACGTACACGTGTACAGGGGCCTCAGGCCGCTACGTGTTACG

         4770       4780       4790       4800       4810       4820       4830
         |          |          |          |          |          |          |
GACGGAATGCTGCCGCGAGGAGGGCTGTACTATGCATGCATGACTGAGTATACCAGCCAGATGGGGTACG
CTGCCTTACGACGGCGCTCCTCCCGACATGATACGTACGTACTGACTCATATGGTCGGTCTACCCCATGC

         4840       4850       4860       4870       4880       4890       4900
         |          |          |          |          |          |          |
ACTTGGCAGAGATGCGGTTCTGATGGTCGGGTGTGTGTTGTTGTGATGAAGAGTGAGGTGACCACGGGTG
TGAACCGTCTCTACGCCAAGACTACCAGCCCACACACAACAACACTACTTCTCACTCCACTGGTGCCCAC

         4910       4920       4930       4940       4950       4960       4970
         |          |          |          |          |          |          |
AAGAAGACTGCAGACGAAGTGGCAGTGGAAGTGGGTGAGAAAGATGATGTAAAACAGTCCATCAATATCG
TTCTTCTGACGTCTGCTTCACCGTCACCTTCACCCACTCTTTCTACTACATTTTGTCAGGTAGTTATAGC

         4980       4990       5000       5010       5020       5030       5040
         |          |          |          |          |          |          |
CGTGCAGATACCCCTAAAACTTGATTGACTCTGCAACAATGTGTGTTTCGTTTTCCCCGTCTACACACGA
GCACGTCTATGGGGATTTTGAACTAACTGAGACGTTGTTACACACAAAGCAAAAGGGGCAGATGTGTGCT
                        R181 revch 2 ←

         5050       5060       5070       5080       5090       5100       5110
         |          |          |          |          |          |          |
CACGGCCTCGACCTTTACTTCACCCAATGCACAACAACATCCATTCCTTCATGCTCATGCCTTCATCGTC
GTGCCGGAGCTGGAAATGAAGTGGGTTACGTGTTGTTGTAGGTAAGGAAGTACGAGTACGGAAGTAGCAG

         5120       5130       5140       5150       5160       5170       5180
         |          |          |          |          |          |          |
ACTTCGCACCGCCTGTCTCTCGGTACACCGGCAACGAAACTGTCCACCACCGCCATGCTCTCAGGCGTCG
TGAAGCGTGGCGGACAGAGAGCCATGTGGCCGTTGCTTTGACAGGTGGTGGCGGTACGAGAGTCCGCAGC

         5190       5200       5210       5220       5230       5240       5250
         |          |          |          |          |          |          |
CCGACTTGGCCAGCTTGACTCCGCGCAGCTCTCGGAAGAAGTGAGCAACGGCAAGACGCATTTCCAACCT
GGCTGAACCGGTCGAACTGAGGCGCGTCGAGAGCCTTCTTCACTCGTTGCCGTTCTGCGTAAAGGTTGGA

         5260       5270       5280       5290       5300       5310       5320
         |          |          |          |          |          |          |
GCCCAGATGAATGCCCAAACACTGCCTGGAACCTTGGCCGAACGGGTTGAACGCCATCTTCTGCTTCTCA
CGGGTCTACTTACGGGTTTGTGACGGACCTTGGAACCGGCTTGCCCAACTTGCGGTAGAAGACGAAGAGT
```

```
         5330      5340      5350      5360      5370      5380      5390
         |         |         |         |         |         |         |
CTGAATTCGAGCCTGTTCTCGGGCAGGAAGCGCGTGTGGTCGAATCTGTCGCAACATTAGCATCTCCCCC
GACTTAAGCTCGGACAAGAGCCCGTCCTTCGCGCACACCAGCTTAGACAGCGTTGTAATCGTAGAGGGGG

         5400      5410      5420      5430      5440      5450      5460
         |         |         |         |         |         |         |
GAGCATCCACCACACATACACAACTCACGTCTGTGCCTCCTTCCACGTGGCACCATCCGTGTGCAAAGTC
CTCGTAGGTGGTGTGTATGTGTTGAGTGCAGACACGGAGGAAGGTGCACCGTGGTAGGCACACGTTTCAG

         5470      5480      5490      5500      5510      5520      5530
         |         |         |         |         |         |         |
CAATTCTGCGTGTCGACGACAGTGCCTTCGGGGATGTAGTATCCACCAAGGTTGGCACCACCAGCTGGCG
GTTAAGACGCACAGCTGCTGTCACGGAAGCCCCTACATCATAGGTGGTTCCAACCGTGGTGGTCGACCGC

         5540      5550      5560      5570      5580      5590      5600
         |         |         |         |         |         |         |
GGCTGCGTGGAAGTCCTCCGGGAGCTGCACCGTACAGTCGCATGCTTTCGTCAATGATGGCGTTCATCAG
CCGACGCACCTTCAGGAGGCCCTCGACGTGGCATGTCAGCGTACGAAAGCAGTTACTACCGCAAGTAGTC

         5610      5620      5630      5640      5650      5660      5670
         |         |         |         |         |         |         |
CGGCAGTCCCTCACATACGGTATCCGTCACTTCTCCATCGATCGACGCGACTTCCTCCTCCAGCTGCTTC
GCCGTCAGGGAGTGTATGCCATAGGCAGTGAAGAGGTAGCTAGCTGCGCTGAAGGAGGAGGTCGACGAAG

         5680      5690      5700      5710      5720      5730      5740
         |         |         |         |         |         |         |
TGGAGCTCAGGTCGGCTGAGAACGAGGTAGATCAGGAACGTCAGCGAAATGGCAGTCGGGTCGGAGCCGG
ACCTCGAGTCCAGCCGACTCTTGCTCCATCTAGTCCTTGCAGTCGCTTTACCGTCAGCCCAGCCTCGGCC

         5750      5760      5770      5780      5790      5800      5810
         |         |         |         |         |         |         |
CGAGCAACAGAGCACCGGCGTCAGTGATGATGTCCGTATCAGTCAACGTCGCAGCATCGCCCTCGCCCTC
GCTCGTTGTCTCGTGGCCGCAGTCACTACTACAGGCATAGTCAGTTGCAGCGTCGTAGCGGGAGCGGGAG

         5820      5830      5840      5850      5860      5870      5880
         |         |         |         |         |         |         |
CTGCAACGCCTTTGCGAACAAGTTCTGGTTGAACTCGCCAGCCTCCTTGTCCTGACGAGCGCTCTTCACC
GACGTTGCGGAAACGCTTGTTCAAGACCAACTTGAGCGGTCGGAGGAACAGGACTGCTCGCGAGAAGTGG

         5890      5900      5910      5920      5930      5940      5950
         |         |         |         |         |         |         |
ACGCCAGCACCGGCCTTGAACATTTTCTCCTGCGAGTAGAAAATGTCGTTCATGCGCGTGTTGACCTTTC
TGCGGTCGTGGCCGGAACTTGTAAAAGAGGACGCTCATCTTTTACAGCAAGTACGCGCACAACTGGAAAG

         5960      5970      5980      5990      6000      6010      6020
         |         |         |         |         |         |         |
CGAGTACCCGACCGACGTAGTAAAGAGGGGGAATGAACATCTTGAGGAGGTGTGCGAGGTCGCCTTTACG
GCTCATGGGCTGGCTGCATCATTTCTCCCCCTTACTTGTAGAACTCCTCCACACGCTCCAGCGGAAATGC

         6030      6040      6050      6060      6070      6080      6090
         |         |         |         |         |         |         |
CTTCTCAAGCATGAGCACGAATGGATCTTTGATACCCTTCTCGACAGTTCCGTGTCCGCCATTGAACGTC
GAAGAGTTCGTACTCGTGCTTACCTAGAAACTATGGGAAGAGCTGTCAAGGCACAGGCGGTAACTTGCAG
```

```
        6100       6110       6120       6130       6140       6150       6160
        |          |          |          |          |          |          |
AGGCGGCAGACAATCTCGTTGGCCATCAGCGTCCACCAGCCCATGAGCTCGGTGTCGCCATTGGCAGCCT
TCCGCCGTCTGTTAGAGCAACCGGTAGTCGCAGGTGGTCGGGTACTCGAGCCACAGCGGTAACCGTCGGA

        6170       6180       6190       6200       6210       6220       6230
        |          |          |          |          |          |          |
CTTCCTTCATCTTCTGCACGGCCATGCTGATGGTCTTGTTGATCGTGCTTTCCCAATTCTTTCTCAGCTC
GAAGGAAGTAGAAGACGTGCCGGTACGACTACCAGAACAACTAGCACGAAAGGGTTAAGAAAGAGTCGAG

        6240       6250       6260       6270       6280       6290       6300
        |          |          |          |          |          |          |
GACAAGAGTAAAGCCCTTGGCATACAAGCTTGCGGCGCTGGCTGTGCTTCTTCTGGTCGCGAAAGTTGAA
CTGTTCTCATTTCGGGAACCGTATGTTCGAACGCCGCGACCGACACGAAGAAGACCAGCGCTTTCAACTT

        6310       6320       6330       6340       6350       6360       6370
        |          |          |          |          |          |          |
GATGTTGTCAACCGGGCCTGGAGAAAGGAGACGATAGAACGGAGCTTTTGTGAACACTGTACCCATGCGG
CTACAACAGTTGGCCCGGACCTCTTTCCTCTGCTATCTTGCCTCGAAAACACTTGTGACATGGGTACGCC

        6380       6390       6400       6410       6420       6430       6440
        |          |          |          |          |          |          |
TGCACCTCCCGTGCCGCCGCTGGATCTGCAACATCCACCTCCTTGGGACCGATCCTCACCATGGGGCCAT
ACGTGGAGGGCACGGCGGCGACCTAGACGTTGTAGGTGGAGGAACCCTGGCTAGGAGTGGTACCCCGGTA

        6450       6460       6470       6480       6490       6500       6510
        |          |          |          |          |          |          |
ACTTTTGGTGCAGTGAATCAACATAGTAGATGCGGTTGCCAGCGAAGACGGAGTACGTCAGGCGCAAGTC
TGAAAACCACGTCACTTAGTTGTATCATCTACGCCAACGGTCGCTTCTGCCTCATGCAGTCCGCGTTCAG

        6520       6530       6540       6550       6560       6570       6580
        |          |          |          |          |          |          |
TGTCAACTTGGCATACCATGGTCCTGGAATCTTGCTCAGAGGAGAAAAGTATGCTGTCGTGATGATCTGC
ACAGTTGAACCGTATGGTACCAGGACCTTAGAACGAGTCTCCTCTTTTCATACGACAGCACTACTAGACG

        6590       6600       6610       6620       6630       6640       6650
        |          |          |          |          |          |          |
AACATCTATTAGCAGCGTTTCCACCGCGAAAATTGACGTCTACTCACAGAAACAATGTTATACAGAACAA
TTGTAGATAATCGTCGCAAAGGTGGCGCTTTTAACTGCAGATGAGTGTCTTTGTTACAATATGTCTTGTT

        6660       6670       6680       6690       6700       6710       6720
        |          |          |          |          |          |          |
ACGCAGCAGCGACAAGTGCCAGCGAAAACGGCAATGGAGCACCAGCAGTGGCGTCCATAATCCACTTGTA
TGCGTCGTCGCTGTTCACGGTCGCTTTTGCCGTTACCTCGTGGTCGTCACCGCAGGTATTAGGTGAACAT

        6730       6740       6750       6760       6770       6780       6790
        |          |          |          |          |          |          |
GAGCTCTCCTGCCATGGTGTCGGAGTGGCAGTGAAGACCAGCAGACGTTGAAAGAGAATGACAAACGGCG
CTCGAGAGGACGGTACCACAGCCTCACCGTCACTTCTGGTCGTCTGCAACTTTCTCTTACTGTTTGCCGC

        6800       6810       6820       6830       6840       6850       6860
        |          |          |          |          |          |          |
TCAAGAACAGTAAGTAACTTGTCAATGCTATCATGCAATCCGTTGATACTACACAAGCCTCGCAGAACGA
AGTTCTTGTCATTCATTGAACAGTTACGATAGTACGTTAGGCAACTATGATGTGTTCGGAGCGTCTTGCT
```

```
         6870       6880       6890       6900       6910       6920       6930
         |          |          |          |          |          |          |
CTTGTCATCAACAAAGCGGTATTCCGCACGCTCTGAGTGATCCTTATGTCGGTTTGCGATACTTCGGACT
GAACAGTAGTTGTTTCGCCATAAGGCGTGCGAGACTCACTAGGAATACAGCCAAACGCTATGAAGCCTGA

         6940       6950       6960       6970       6980       6990       7000
         |          |          |          |          |          |          |
TCAGGCTCGGTCAAGTGAATCCAGCCCGAAACAGCCAGCGCAAGCCCACGGCAGTACCCGATTGGATCGA
AGTCCGAGCCAGTTCACTTAGGTCGGGCTTTGTCGGTCGCGTTCGGGTGCCGTCATGGGCTAACCTAGCT

         7010       7020       7030       7040       7050       7060       7070
         |          |          |          |          |          |          |
TTGATCGATACCGCGCAAACTGCTGCTTGCCGTCGGCCTGCACGTGTATGCCATGCGAATGCGCTAACTC
AACTAGCTATGGCGCGTTTGACGACGAACGGCAGCCGGACGTGCACATACGGTACGCTTACGCGATTGAG

         7080       7090       7100       7110       7120       7130       7140
         |          |          |          |          |          |          |
TGCCTCAGCTATCGTAATCGCCAAACCCTTCGGCCTGGCCAGCCGTCCTCACAGTGGCCCGACTGTTGCT
ACGGAGTCGATAGCATTAGCGGTTTGGGAAGCCGGACCGGTCGGCAGGAGTGTCACCGGGCTGACAACGA

         7150       7160       7170       7180       7190       7200       7210
         |          |          |          |          |          |          |
GGCGGACAGCGACCGTATCTCGCAGGCAAGACCCGAAGACAAATTGACTTTGATACGCAGGCCCAGTACA
CCGCCTGTCGCTGGCATAGAGCGTCCGTTCTGGGCTTCTGTTTAACTGAAACTATGCGTCCGGGTCATGT

         7220       7230       7240       7250       7260       7270       7280
         |          |          |          |          |          |          |
CTTTCCTCGCTGTCTTCCTCGACCTGTTGACTGCCATCGATCCCAGCACCGCTCACCATACAGGCTGTCC
GAAAGGAGCGACAGAAGGAGCTGGACAACTGACGGTAGCTAGGGTCGTGGCGAGTGGTATGTCCGACAGG

         7290       7300       7310       7320       7330       7340       7350
         |          |          |          |          |          |          |
CCTGTCTCTACACCTACACTCATTCTTTCCAGGTGTGCTTTGCGCCTCGTTGATGACACACTAATTTCGA
GGACAGAGATGTGGATGTGAGTAAGAAAGGTCCACACGAAACGCGGAGCAACTACTGTGTGATTAAAGCT

         7360       7370       7380       7390       7400       7410       7420
         |          |          |          |          |          |          |
TACTCAGGCGACTTTCTTCGCCATAACCTCCCCACACAGTCAATATGACACACTCCAACGCAACGAGGGT
ATGAGTCCGCTGAAAGAAGCGGTATTGGAGGGGTGTGTCAGTTATACTGTGTGAGGTTGCGTTGCTCCCA

         7430       7440       7450       7460       7470       7480       7490
         |          |          |          |          |          |          |
ACTGGTCTTTGGTGACCAGACCTATGACTTTGTGCCTAAGCTTCGAGAACTGTTCCACGTCAAGGATAAC
TGACCAGAAACCACTGGTCTGGATACTGAAACACGGATTCGAAGCTCTTGACAAGGTGCAGTTCCTATTG

         7500       7510       7520       7530       7540       7550       7560
         |          |          |          |          |          |          |
CCGATCTTGACGGCTTTCCTGGAACAGTCTCATTATGTCGTGTAAGTCGTACTGATCGATCGGACGCAGA
GGCTAGAACTGCCGAAAGGACCTTGTCAGAGTAATACAGCACATTCAGCATGACTAGCTAGCCTGCGTCT

         7570       7580       7590       7600       7610       7620       7630
         |          |          |          |          |          |          |
CCGTTGCTAAGAATTGCCAGCCGAGCACAGATGATCCAGACACTGCCTCCGGCGGAGCACAAGGCAGCTC
GGCAACGATTCTTAACGGTCGGCTCGTGTCTACTAGGTCTGTGACGGAGGCCGCCTCGTGTTCCGTCGAG
```

```
         7640      7650      7660      7670      7680      7690      7700
         |         |    pks CB2/gateway pks CB2 →          |         |
GAACCTTCGACTTAGCAGATATGCTGAAGAAGTATGTCGCCGGCAAGCTGAACCCTGCCTTCCAGACGGC
CTTGGAAGCTGAATCGTCTATACGACTTCTTCATACAGCGGCCGTTCGACTTGGGACGGAAGGTCTGCCG

         7710      7720      7730      7740      7750      7760      7770
         |         |         |         |         |         |         |
CCTCAGCTGCATCACACAACTCGGTGTGTTCATGCGAGAGTTCCATGACTTCACCAAGCCATATCCACGA
GGAGTCGACGTAGTGTGTTGAGCCACACAAGTACGCTCTCAAGGTACTGAAGTGGTTCGGTATAGGTGCT

         7780      7790      7800      7810      7820      7830      7840
         |         |         |         |         |         |         |
CACGATAGCAGCTACGTGTTGGGTATCTGCACCGGCTCGCTTGCCGCTGCGGCCGTCAGCTCCAGCAACT
GTGCTATCGTCGATGCACAACCCATAGACGTGGCCGAGCGAACGGCGACGCCGGCAGTCGAGGTCGTTGA

         7850      7860      7870      7880      7890      7900      7910
         |         |         |         |         |         |         |
CTTTATCCGAGCTTCTGCCTATTGCCGTGCAAACGGCTTTGATCGCCTTTCGCCTCGGTCTGTGCGTCAC
GAAATAGGCTCGAAGACGGATAACGGCACGTTTGCCGAAACTAGCGGAAAGCGGAGCCAGACACGCAGTG

         7920      7930      7940      7950      7960      7970      7980
         |         |         |         |         |         |         |
AGACATGCGAGATCGTCTCGAAAGCTCGGAAGAAGACCGCACCCAGCCTTGGTCGGTAGTTCTGTTCGAC
TCTGTACGCTCTAGCAGAGCTTTCGAGCCTTCTTCTGGCGTGGGTCGGAACCAGCCATCAAGACAAGCTG

         7990      8000      8010      8020      8030      8040      8050
         |         |         |         |         |         |         |
ACAGACGAGCAGACTGTTACCAAGGCCATCAAAGACTTCTGCACATCCAACGTCCTCCCGAAGACCAAGC
TGTCTGCTCGTCTGACAATGGTTCCGGTAGTTTCTGAAGACGTGTAGGTTGCAGGAGGGCTTCTGGTTCG

         8060      8070      8080      8090      8100      8110      8120
         |         |         |         |         |         |         |
AGCCCTGGATTACCTCCGCGTCATCGAAGACAATCTCCATCAGCGGAGCACCACGTGTGCTGAAGAAGTT
TCGGGACCTAATGGAGGCGCAGTAGCTTCTGTTAGAGGTAGTCGCCTCGTGGTGCACACGACTTCTTCAA

         8130      8140      8150      8160      8170      8180      8190
         |         |         |         |         |         |         |
GTCGCAAGAGCCTGCGCTCAAGGACAAAAAGACCAGACAGATCCCGATTTATGTCCCAGCACACAATTCT
CAGCGTTCTCGGACGCGAGTTCCTGTTTTTCTGGTCTGTCTAGGGCTAAATACAGGGTCGTGTGTTAAGA

         8200      8210      8220      8230      8240      8250      8260
         |         |         |         |         |         |         |
GCCCTCTTCACCCCCGAAGATGTCAAGTCTATCCTTGAGACTACCCCTGTCGACACTTGGAGCAACTACC
CGGGAGAAGTGGGGGCTTCTACAGTTCAGATAGGAACTCTGATGGGGACAGCTGTGAACCTCGTTGATGG

         8270      8280      8290      8300      8310      8320      8330
         |         |         |         |         |         |         |
CTACCAAGCTTCCATTCATCTCGAGCGTTTCTGGCAAGATGGCTTGGGCAGACAACTACCTTGCAGTAAT
GATGGTTCGAAGGTAAGTAGAGCTCGCAAAGACCGTTCTACCGAACCCGTCTGTTGATGGAACGTCATTA

         8340      8350      8360      8370      8380      8390      8400
         |         |         |         |         |         |         |
CCATCTCGCCCTCAATCAGTGCCTGCTCGAGAGCATTGGCTGGGGTAAGGTCGAGACTGAGCTCCCAAGG
GGTAGAGCGGGAGTTAGTCACGGACGAGCTCTCGTAACCGACCCCATTCCAGCTCTGACTCGAGGGTTCC
```

```
         8410       8420       8430       8440       8450       8460       8470
         |          |          |          |          |          |          |
CTCCTCAAGTCTCGCGGCGCAGAAAATGTGCTCATCACGCCAATCACCACCTCTGCCGACCGTGCTCTGT
GAGGAGTTCAGAGCGCCGCGTCTTTTACACGAGTAGTGCGGTTAGTGGTGGAGACGGCTGGCACGAGACA

         8480       8490       8500       8510       8520       8530       8540
         |          |          |          |          |          |          |
CGGCTGCACTCAGCCCAACGATCTCCAATATCGAGGTCGAGAAGCCCACAATCAACGAATCTTTCGCCCA
GCCGACGTGAGTCGGGTTGCTAGAGGTTATAGCTCCAGCTCTTCGGGTGTTAGTTGCTTAGAAAGCGGGT

         8550       8560       8570       8580       8590       8600       8610
         |          |          |          |          |          |          |
TAGACCTGGTTCGGGAAAGAGCAAGCTTGCTATCGTCTCAATGTCTGGTCGCTTCCCAGAGGCACAAAGC
ATCTGGACCAAGCCCTTTCTCGTTCGAACGATAGCAGAGTTACAGACCAGCGAAGGGTCTCCGTGTTTCG

         8620       8630       8640       8650       8660       8670       8680
         |          |          |          |          |          |          |
ACCGACGCTTTTTGGGATCTGCTCTACAAGGGGCTTGACGTGGTCAAGGAAGTGCCCAAACGTCGTTGGG
TGGCTGCGAAAAACCCTAGACGAGATGTTCCCCGAACTGCACCAGTTCCTTCACGGGTTTGCAGCAACCC

         8690       8700       8710       8720       8730       8740       8750
         |          |          |          |          |          |          |
ACGTCGAGACTCACGTCGACCCAACTGGGCGTGCTCGCAACAAGGGTGCGACCAAATGGGGCTGCTGGCT
TGCAGCTCTGAGTGCAGCTGGGTTGACCCGCACGAGCGTTGTTCCCACGCTGGTTTACCCCGACGACCGA

         8760       8770       8780       8790       8800       8810       8820
         |          |          |          |          |          |          |
CGATTTCGCCGGAGAATTCGACCCTCGCTTCTTCAGCATCTCACCCAAGGAAGCACCACAAATGGACCCC
GCTAAAGCGGCCTCTTAAGCTGGGAGCGAAGAAGTCGTAGAGTGGGTTCCTTCGTGGTGTTTACCTGGGG

         8830       8840       8850       8860       8870       8880       8890
         |          |          |          |          |          |          |
GCCCAGCGTATGGCTTTGATGTCGACCTGGGAAGCGATGGAACGTGGTGGCATCGTCCCGGATACTACGC
CGGGTCGCATACCGAAACTACAGCTGGACCCTTCGCTACCTTGCACCACCGTAGCAGGGCCTATGATGCG

         8900       8910       8920       8930       8940       8950       8960
         |          |          |          |          |          |          |
CCTCAACACAAAGAAATCGAATTGGTGTCTTCCATGGCGTTACCTCCAACGACTGGATGGAGACCAACAC
GGAGTTGTGTTTCTTTAGCTTAACCACAGAAGGTACCGCAATGGAGGTTGCTGACCTACCTCTGGTTGTG
                                          OT pks KO/2 /gateway OT pks KO/2 ←

         8970       8980       8990       9000       9010       9020       9030
         |          |          |          |          |          |          |
GGCTCAAAACATTGATACCTACTTCATCACCGGCGGTAACCGCGGTTTCATCCCCGGCCGTATTAACTTC
CCGAGTTTTGTAACTATGGATGAAGTAGTGGCCGCCATTGGCGCCAAAGTAGGGGCCGGCATAATTGAAG

         9040       9050       9060       9070       9080       9090       9100
         |          |       KS Domain       |          |          |          |
TGCTTTGAATTCTCTGGACCCAGCTTCACCAATGACACGGCCTGCTCCAGTTCGCTCGCAGCGATCCATT
ACGAAACTTAAGAGACCTGGGTCGAAGTGGTTACTGTGCCGGACGAGGTCAAGCGAGCGTCGCTAGGTAA

         9110       9120       9130       9140       9150       9160       9170
         |          |          |          |          |          |          |
TGGCTTGCAACTCGCTCTGGCGCGGCGATTGCGATACTGCTGTGGCGGGTGGCACGAACATGATCTTCAC
ACCGAACGTTGAGCGAGACCGCGCCGCTAACGCTATGACGACACCGCCCACCGTGCTTGTACTAGAAGTG
```

```
         9180      9190      9200      9210      9220      9230      9240
         |         |         |         |         |         |         |
ACCTGATGGTCACGCTGGTCTCGACAAAGGGTTCTTCCTGTCCCGTACTGGTAACTGTAAGCCTTTCGAT
TGGACTACCAGTGCGACCAGAGCTGTTTCCCAAGAAGGACAGGGCATGACCATTGACATTCGGAAAGCTA

         9250      9260      9270      9280      9290      9300      9310
         |         |         |         |         |         |         |
GACAAGGCTGACGGATACTGTCGTGCTGAGGGTGTTGGTACCGTTATGGTCAAGAGGCTCGAAGATGCTC
CTGTTCCGACTGCCTATGACAGCACGACTCCCACAACCATGGCAATACCAGTTCTCCGAGCTTCTACGAG

         9320      9330      9340      9350      9360      9370      9380
         |         |         |         |         |         |         |
TTGCGGACGGAGATCCAATCCTTGGCACGATCCTCGACGCGAAGACGAACCACTCCGCCATGAGCGACTC
AACGCCTGCCTCTAGGTTAGGAACCGTGCTAGGAGCTGCGCTTCTGCTTGGTGAGGCGGTACTCGCTGAG

         9390      9400      9410      9420      9430      9440      9450
         |         |         |         |         |         |         |
TATGACTCGCCCCTTCGTCCCAGCCCAGATCGACAACATGGAAGCTTGCCTCAGCACCGCTGGAGTGGAC
ATACTGAGCGGGGAAGCAGGGTCGGGTCTAGCTGTTGTACCTTCGAACGGAGTCGTGGCGACCTCACCTG

         9460      9470      9480      9490      9500      9510      9520
         |         |         |         |         |         |         |
CCTACCTCTCTCGACTACATTGAGATGCACGGTACTGGTACTCAAGTCGGCGACGCAGTCGAGATGGAGT
GGATGGAGAGAGCTGATGTAACTCTACGTGCCATGACCATGAGTTCAGCCGCTGCGTCAGCTCTACCTCA

         9530      9540      9550      9560      9570      9580      9590
         |         |         |         |         |         |         |
CTGTTCTCAGCGTCTTTGCGCCGAATGAGCAGTTCCGCGGCAAGGACCAGCCTCTGTATGTCGGCTCCGC
GACAAGAGTCGCAGAAACGCGGCTTACTCGTCAAGGCGCCGTTCCTGGTCGGAGACATACAGCCGAGGCG

         9600      9610      9620      9630      9640      9650      9660
         |         |         |         |         |         |         |
CAAGGCCAACATCGGACACGGTGAGGGTGTGTCTGGTGTCACCAGTTTGATCAAGGTCCTTCTCATGATG
GTTCCGGTTGTAGCCTGTGCCACTCCCACACAGACCACAGTGGTCAAACTAGTTCCAGGAAGAGTACTAC

         9670      9680      9690      9700      9710      9720      9730
         |         |         |         |         |         |         |
CAGACCAACCACTATCCGCCGCATTGCGGTATCAAGCCTGGAAGCAAGATCAACCACAATTACCCGGATC
GTCTGGTTGGTGATAGGCGGCGTAACGCCATAGTTCGGACCTTCGTTCTAGTTGGTGTTAATGGGCCTAG

         9740      9750      9760      9770      9780      9790      9800
         |         |         |         |         |         |         |
TTGCGGCAAGAAATGTGCACATCGCGTTTGAGCCGAAACCGTTCTTGAGACGGGAGGGCAAGTTAAGACG
AACGCCGTTCTTTACACGTGTAGCGCAAACTCGGCTTTGGCAAGAACTCTGCCCTCCCGTTCAATTCTGC

         9810      9820      9830      9840      9850      9860      9870
         |         |         |         |         |         |         |
GGTTTTGATCAATAACTTCAGTGCTGCAGGTGGCAATACTGCGCTTCTCATTGAGGATGCGCCTGACAGG
CCAAAACTAGTTATTGAAGTCACGACGTCCACCGTTATGACGCGAAGAGTAACTCCTACGCGGACTGTCC

         9880      9890      9900      9910      9920      9930      9940
         |         |         |         |         |         |         |
ATGCCGCTCTCAGGACAAGATCCTCGCACGACTCAGACTGTCACGATCTCGGGACATGTTGGCAAGTCTC
TACGGCGAGAGTCCTGTTCTAGGAGCGTGCTGAGTCTGACAGTGCTAGAGCCCTGTACAACCGTTCAGAG
```

```
         9950       9960       9970       9980       9990       10000      10010
         |          |          |          |          |          |          |
TCAGCAACAATGTCGCCAACTTGCTCGCACATCTGAAGAAGAATCCTACCATCGATCTCTCACAGCTCGC
AGTCGTTGTTACAGCGGTTGAACGAGCGTGTAGACTTCTTCTTAGGATGGTAGCTAGAGAGTGTCGAGCG

         10020      10030      10040      10050      10060      10070      10080
         |          |          |          |          |          |          |
CTACACGGTCAGTGCACGAAGATGGCATCACCTCCATCGTGTTGCTGTCGCGGGTACTACCGTCGCAGAT
GATGTGCCAGTCACGTGCTTCTACCGTAGTGGAGGTAGCACAACGACAGCGCCCATGATGGCAGCGTCTA

         10090      10100      10110      10120      10130      10140      10150
         |          |          |          |          |          |          |
ATTACCGCGAAGTTGGAGAAAGCCATTGAGAATAAGGAAGGTGTCAACAGACCTAAGGCGAAGCCTTCGG
TAATGGCGCTTCAACCTCTTTCGGTAACTCTTATTCCTTCCACAGTTGTCTGGATTCCGCTTCGGAAGCC

         10160      10170      10180      10190      10200      10210      10220
         |          |          |          |          |          |          |
TCTTCTTCGCCTTCACAGGTCAAGGATCTCAGTACCTCGGCATGGGCAAGCAACTCTACGACTCTTATCC
AGAAGAAGCGGAAGTGTCCAGTTCCTAGAGTCATGGAGCCGTACCCGTTCGTTGAGATGCTGAGAATAGG

         10230      10240      10250      10260      10270      10280      10290
         |          |          |          |          |          |          |
AATGTTCAGATCCGAGCTTCAAGGCTACGATCGCTTGGCACAATCGCAAGGCTTCCCAAGCTTTGCACAC
TTACAAGTCTAGGCTCGAAGTTCCGATGCTAGCGAACCGTGTTAGCGTTCCGAAGGGTTCGAAACGTGTG

         10300      10310      10320      10330      10340      10350      10360
         |          |          |          |          |          |          |
ATCTTCACCGAGACGAAGGGAGATGTTGAACAGAATCTTCCAGTGGTCGTGCAGCTTGCTATTACATGCT
TAGAAGTGGCTCTGCTTCCCTCTACAACTTGTCTTAGAAGGTCACCAGCACGTCGAACGATAATGTACGA

         10370      10380      10390      10400      10410      10420      10430
         |          |          |          |          AT Domain             |
TGCAAATGGCTCTCTTCAACCTCGTCACCTCCTTCGGAATCAAGGCCTCTGCCGTTGTCGGCCACTCGCT
ACGTTTACCGAGAGAAGTTGGAGCAGTGGAGGAAGCCTTAGTTCCGGAGACGGCAACAGCCGGTGAGCGA

         10440      10450      10460      10470      10480      10490      10500
         |          |          |          |          |          |          |
GGGCGAGTACGCTGCGCTGTATGCAGCTGGTGTGTTGAGTGCCAGCGACACGATCTACCTGGTCGGCAAA
CCCGCTCATGCGACGCGACATACGTCGACCACACAACTCACGGTCGCTGTGCTAGATGGACCAGCCGTTT

         10510      10520      10530      10540      10550      10560      10570
         |          |          |          |          |          |          |
CGTGCCGAGCTTCTCCAGGATCATTGCCAGAGGGGTACGCATGCGATGCTTGCGTGCAAGGCGAGTGAGT
GCACGGCTCGAAGAGGTCCTAGTAACGGTCTCCCCATGCGTACGCTACGAACGCACGTTCCGCTCACTCA

         10580      10590      10600      10610      10620      10630      10640
         |          |          |          |          |          |          |
GGAGTCTCGCCGAGATCACGGCGGGCAAGAATGTCGAAGTCGCATGCGTTAATGGGCCTGAAGACACTGT
CCTCAGAGCGGCTCTAGTGCCGCCCGTTCTTACAGCTTCAGCGTACGCAATTACCCGGACTTCTGTGACA

         10650      10660      10670      10680      10690      10700      10710
         |          |          OT pks KO/3 /gateway OT pks ko/3 →          |
CCTCTCCGGCACTGTCGAGGAAATTGGAGAGGTGCAGAAGACACTCATTGCGAAGAGCATCAAGGCTACA
GGAGAGGCCGTGACAGCTCCTTTAACCTCTCCACGTCTTCTGTGAGTAACGCTTCTCGTAGTTCCGATGT
```

```
         10720     10730     10740     10750     10760     10770     10780
         |         |         |         |         |         |         |
CTCTTGAAGTTGCCCTTCGCGTTTCATTCGGCGCAGGTACAACCTATCCTCCGAGGACGTTCTGAAGAAC
GAGAACTTCAACGGGAAGCGCAAAGTAAGCCGCGTCCATGTTGGATAGGAGGCTCCTGCAAGACTTCTTG

         10790     10800     10810     10820     10830     10840     10850
         |         |         |         |         |         |         |
TTGCGGCTGGAGCTACTTTTGAGAAGCCCAAGCTTGCGGTCATTTCCCCGCTACTGGGCAGTGTGGTCGA
AACGCCGACCTCGATGAAAACTCTTCGGGTTCGAACGCCAGTAAAGGGGCGATGACCCGTCACACCAGCT

         10860     10870     10880     10890     10900     10910     10920
         |         |         |         |         |         |         |
CGACGAAGGAGTCGTTGGACCCAACTACCTTGCACGCCACTGCCGTGAGGCGGTCGGAATGGTCAAAGCC
GCTGCTTCCTCAGCAACCTGGGTTGATGGAACGTGCGGTGACGGCACTCCGCCAGCCTTACCAGTTTCGG

         10930     10940     10950     10960     10970     10980     10990
         |         |         |         |         |         |         |
CTCGGAGTGGCGAAGGAGAAGGGTATAATCAACGAGAAGACCTTCGTCATTGAGATTGGTCCTAAGCCGC
GAGCCTCACCGCTTCCTCTTCCCATATTAGTTGCTCTTCTGGAAGCAGTAACTCTAACCAGGATTCGGCG

         11000     11010     11020     11030     11040     11050     11060
         |         |         |         |         |         |         |
TTCTCTGCGGAATGATCAAGAACATACTCGGCCAGAACATCGTAGCCTTGCCTACGTTGAAGGACAAGGG
AAGAGACGCCTTACTAGTTCTTGTATGAGCCGGTCTTGTAGCATCGGAACGGATGCAACTTCCTGTTCCC

         11070     11080     11090     11100     11110     11120     11130
         |         |         |         |         |  XbaI    |         |
TCCAGACGTCTGGCAGAACCTCTCGAACATCTTCACGACGCTCTACACCGGTGGTCTAGACATCAACTGG
AGGTCTGCAGACCGTCTTGGAGAGCTTGTAGAAGTGCTGCGAGATGTGGCCACCAGATCTGTAGTTGACC

         11140     11150     11160     11170     11180     11190     11200
         |         |         |         |         |         |         |
ACTGCCTTCCACGCCCCCTTCGAGCCCGCGAAGAAGGTCCTGCAACTTCCTGATTATGGCTGGGATCTCA
TGACGGAAGGTGCGGGGGGAAGCTCGGGCGCTTCTTCCAGGACGTTGAAGGACTAATACCGACCCTAGAGT

         11210     11220     11230     11240     11250     11260     11270
         |         |         |         |         |         |         |
AGGATTACTTCATCCAGTATGAAGGCGATTGGGTTCTGCATCGGCACAAGATCCACTGCAACTGTGCAGA
TCCTAATGAAGTAGGTCATACTTCCGCTAACCCAAGACGTAGCCGTGTTCTAGGTGACGTTGACACGTCT

         11280     11290     11300     11310     11320     11330     11340
         |         |         |         |         |         |         |
TGCTGGAAAGGATGTGCATAACACTTCGCACTACTGTCCTGGCAAACACACCTTCGCTGAGAATGTTGTC
ACGACCTTTCCTACACGTATTGTGAAGCGTGATGACAGGACCGTTTGTGTGGAAGCGACTCTTACAACAG

         11350     11360     11370     11380     11390     11400     11410
         |         |         |         |         |         |         |
GTTCCTGGTGGGGCTCAGAAGGCCGTTCAGGAAGCACCTGCGGCGAAGACAGAGACGAAGAAGATGTCGA
CAAGGACCACCCCGAGTCTTCCGGCAAGTCCTTCGTGGACGCCGCTTCTGTCTCTGCTTCTTCTACAGCT

         11420     11430     11440     11450     11460     11470     11480
         |         |         |         |         |         |         |
AGCTGGATCCTACCAAGGAGGCGTATCCGGGCATTCCGCTCACCACGACCGTGCACAAGGTCATTGAAGA
TCGACCTAGGATGGTTCCTCCGCATAGGCCCGTAAGGCGAGTGGTGCTGGCACGTGTTCCAGTAACTTCT
```

```
         11490     11500     11510     11520     11530     11540     11550
         |         |         |         |         |         |         |
GAAGACGGAGCCTCTGGGAGCGCAGTTCACGGTCGAGACGGATATCTCCCGCAAGGATGTCAACAGCATC
CTTCTGCCTCGGAGACCCTCGCGTCAAGTGCCAGCTCTGCCTATAGAGGGCGTTCCTACAGTTGTCGTAG
OT pks KO/4 /gateway OT pks KO/4 ←

         11560     11570     11580     11590     11600     11610     11620
         |         |         |         |         |         |         |
GCTCAAGGTCACACTGTTGACAGCATTCCCCTCTGCACGCCATCATTCTATGCGGATATTGCGCTTCAGG
CGAGTTCCAGTGTGACAACTGTCGTAAGGGGAGACGTGCGGTAGTAAGATACGCCTATAACGCGAAGTCC

         11630     11640     11650     11660     11670     11680     11690
         |         |         |         |         |         |         |
TTGGCAAATACGCCATGGACCGCATCCGTGCTGGACATCCCGGTGCCGGTGCTATTGATGGAAGGGTCGA
AACCGTTTATGCGGTACCTGGCGTAGGCACGACCTGTAGGGCCACGGCCACGATAACTACCTTCCCAGCT

         11700
         |
TGTTACGGATCGAG
ACAATGCCTAGCTC
```

APPENDIX 3: cDNA SEQUENCES

Primers used for RT-PCR of *mox*^{dot}: Mox-dot 5' end and Mox-dot 3' end
Primers used for RT-PCR of *ord*^{dot}: OT R162A Fwd1 and OT R162A Rev6
Primers used for RT- PCR of *epox*^{dot}: OT R162A Rev4 and Epoxide 3' end

mox^{dot} cDNA

```
         10        20        30        40        50        60        70
         |         |         |         |         |         |         |
TTGTGAGAATGTGGAACATGTGGTTTATGAGAAGAATGAGGATGTTGGTGGAACGTGGCTGGAGAATCGA
AACACTCTTACACCTTGTACACCAAATACTCTTCTTACTCCTACAACCACCTTGCACCGACCTCTTAGCT

         80        90        100       110       120       130       140
         |         |         |         |         |         |         |
TATCCTCGAGCGGGGTGTGATATTCCGAGTCATGCTTATACGTACCAGTTTGCTTTGAACCCTGACTGGC
ATAGGAGCTCGCCCCACACTATAAGGCTCAGTACGAATATGCATGGTCAAACGAAACTTGGGACTGACCG

         150       160       170       180       190       200       210
         |         |         |         |         |         |         |
CGAGATTCTTTTCCTTCGCTCCGGACATCTGGGCGTACTTGAACAAAGTCTGCGAGACGTTCGATTTGAA
GCTCTAAGAAAAGGAAGCGAGGCCTGTAGACCCGCATGAACTTGTTTCAGACGCTCTGCAAGCTAAACTT

         220       230       240       250       260       270       280
         |         |         |         |         |         |         |
GAAGTACATGAGGTTCCACGTGGAGGTTGTGGGGTGCTACTGGCAGGAGCATGCGGGCGAATGGGTGGTC
CTTCATGTACTCCAAGGTGCACCTCCAACACCCCACGATGACCGTCCTCGTACGCCCGCTTACCCACCAG

         290       300       310       320       330       340       350
         |         |         |         |         |         |         |
AAGCTTCGAGAACATCTTCCAAACCACGAGGTCCGGGAGTTCGAGGATAGATGTAATGTGCTGTTGTATG
TTCGAAGCTCTTGTAGAAGGTTTGGTGCTCCAGGCCCTCAAGCTCCTATCTACATTACACGACAACATAC

         360       370       380       390       400       410       420
         |         |         |         |         |         |         |
GCGCTGGCGTGCTGAATAATTTCAAGTTCCCCGATATTCCAGGCCTGCAGGATCGCTTCAAGGGCCGGGT
CGCGACCGCACGACTTATTAAAGTTCAAGGGGCTATAAGGTCCGGACGTCCTAGCGAAGTTCCCGGCCCA

         430       440       450       460       470       480       490
         |         |         |         |         |         |         |
CATTCACACCGCCCGGTGGCCAAAGGACTACAAGGAGGAAGACTGGGCAAAGGAGCGAGTCGCCGTAATT
GTAAGTGTGGCGGGCCACCGGTTTCCTGATGTTCCTCCTTCTGACCCGTTTCCTCGCTCAGCGGCATTAA

         500       510       520       530       540       550       560
         |         |         |         |         |         |         |
GGATCTGGTGCTTCATCCATACAGACTGTTCCAGGCATGCAGCCATACGCAAAACACCTCGACATCTTCG
CCTAGACCACGAAGTAGGTATGTCTGACAAGGTCCGTACGTCGGTATGCGTTTTGTGGAGCTGTAGAAGC

         570
         |
TCCGTACGGGAGT
AGGCATGCCCTCA
```

ord^{dot} cDNA

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
TCACCGCTATCGTCACCACCCTGGGGATGCTCCGCAAACTCCACGGCAGCGCCTACAACGCACCGACGAT
AGTGGCGATAGCAGTGGTGGGACCCCTACGAGGCGTTTGAGGTGCCGTCGCGGATGTTGCGTGGCTGCTA

        80        90       100       110       120       130       140
         |         |         |         |         |         |         |
CCTCCAGCTCCGCAGCGCGTCCCTCAACCCTAAGCTCTCATGTCAGGTACCACGCCTGGTCTACAACATC
GGAGGTCGAGGCGTCGCGCAGGGAGTTGGGATTCGAGAGTACAGTCCATGGTGCGGACCAGATGTTGTAG

       150       160       170       180       190       200       210
         |         |         |         |         |         |         |
GTCTCCTTCTGTCTCCACTACAGCCATCTCGACATCGTCAAAGCCTGCGAACACTACGAGGCCGCTGCAG
CAGAGGAAGACAGAGGTGATGTCGGTAGAGCTGTAGCAGTTTCGGACGCTTGTGATGCTCCGGCGACGTC

       220       230       240       250       260       270       280
         |         |         |         |         |         |         |
CCAAAGGCCTGTTGTCGTATATCTACGTCGACCCACCCACGATCCACGATGCTTTCGGGCCTAATCGCAC
GGTTTCCGGACAACAGCATATAGATGCAGCTGGGTGGGTGCTAGGTGCTACGAAAGCCCGGATTAGCGTG

       290       300       310       320       330       340       350
         |         |         |         |         |         |         |
GGGACACAAGCTAATCTCCTGCAAACCCGACGTCTGCGACAAGCAGGAGACGGCACTGAGCTACGCCGAT
CCCTGTGTTCGATTAGAGGACGTTTGGGCTGCAGACGCTGTTCGTCCTCTGCCGTGACTCGATGCGGCTA

       360       370       380       390       400       410       420
         |         |         |         |         |         |         |
CTCGGTGCTGGATTCGTTGAGATCGCAAGCCGAAAGGAAGACTTCCTTAACCAACCTGTGGGTGTGACGG
GAGCCACGACCTAAGCAACTCTAGCGTTCGGCTTTCCTTCTGAAGGAATTGGTTGGACACCCACACTGCC

       430       440       450       460       470
         |         |         |         |         |
CGACGGGGAAGGCCAAGGAGACGTGGGGAGTGCTGGCGGGATTCTTGTTCGATGGGGCG
GCTGCCCCTTCCGGTTCCTCTGCACCCCTCACGACCGCCCTAAGAACAAGCTACCCCGC
```

epox^{dot} cDNA

```
         10        20        30        40        50        60        70
         |         |         |         |         |         |         |
TCTCGAGGGAGCAGGCCGCGGGTTGTGAAGCTTGCAAGGGATTTCATCTTAATATGATTTTGCTGCCGCC
AGAGCTCCCTCGTCCGGCGCCCAACACTTCGAACGTTCCCTAAAGTAGAATTATACTAAAACGACGGCGG

         80        90        100       110       120       130       140
         |         |         |         |         |         |         |
GCCGGCGAACATGAAAGAGTTGACGTTGGAGGAGGTGGAGAAGAAGGCTATGCCCAACGCCCTGGCGTTC
CGGCCGCTTGTACTTTCTCAACTGCAACCTCCTCCACCTCTTCTTCCGATACGGGTTGCGGGACCGCAAG

         150       160       170       180       190       200       210
         |         |         |         |         |         |         |
CGTCAAAGTGGCATGGCCTATGCCCTGGAGCACGGAACCCGCGGCGGCACGATTGGATTGGCATTGCAGG
GCAGTTTCACCGTACCGGATACGGGACCTCGTGCCTTGGGCGCCGCCGTGCTAACCTAACCGTAACGTCC

         220       230       240       250       260       270       280
         |         |         |         |         |         |         |
CGAGTCCCGTGGCGCTGTTATGTTGGATCGGGGAGAAGATGATGGCCTGGAGTGATTCCTCTTCTCAGCC
GCTCAGGGCACCGCGACAATACAACCTAGCCCCTCTTCTACTACCGGACCTCACTAAGGAGAAGAGTCGG

         290       300       310       320       330       340       350
         |         |         |         |         |         |         |
GTCTCTCGAACAAATCTTGGAGACCGTCTCCCTATACTGGCTCACCGATTCCATCACGCGGGGTCTTTAC
CAGAGAGCTTGTTTAGAACCTCTGGCAGAGGGATATGACCGAGTGGCTAAGGTAGTGCGCCCCAGAAATG

         360       370       380       390       400       410       420
         |         |         |         |         |         |         |
CCCTACCGTCGCTTCGCTTCTGGTAACGAACCCAAAATTAATTTCATCGAGAAGCCGCTGGGGTATTCGT
GGGATGGCAGCGAAGCGAAGACCATTGCTTGGGTTTTAATTAAAGTAGCTCTTCGGCGACCCCATAAGCA

         430       440       450       460       470       480       490
         |         |         |         |         |         |         |
TTTTCCCGAATACGTACCTCCCTTGCCCTGTGAGCTGGGCGAAGACGACGGCGAATTTGGTCCAGTACCG
AAAAGGGCTTATGCATGGAGGGAACGGGACACTCGACCCGCTTCTGCTGCCGCTTAAACCAGGTCATGGC

         500       510
         |         |
CAGACATGAGAGTGGAGGTCAT
GTCTGTACTCTCACCTCCAGTA
```

APPENDIX 4: DISRUPTION VECTOR

5'and 3' flanking regions are shown in black, whilst the *hph* cassette is shown in blue.
The *att*B1 and *att*B2 sites linking the *hph* cassette to the 5'*pks* flank and 3' *pks* flank respectively are indicated in red.
The two *Xba*I sites enclosing the *hph* cassette are indicated by underlining.

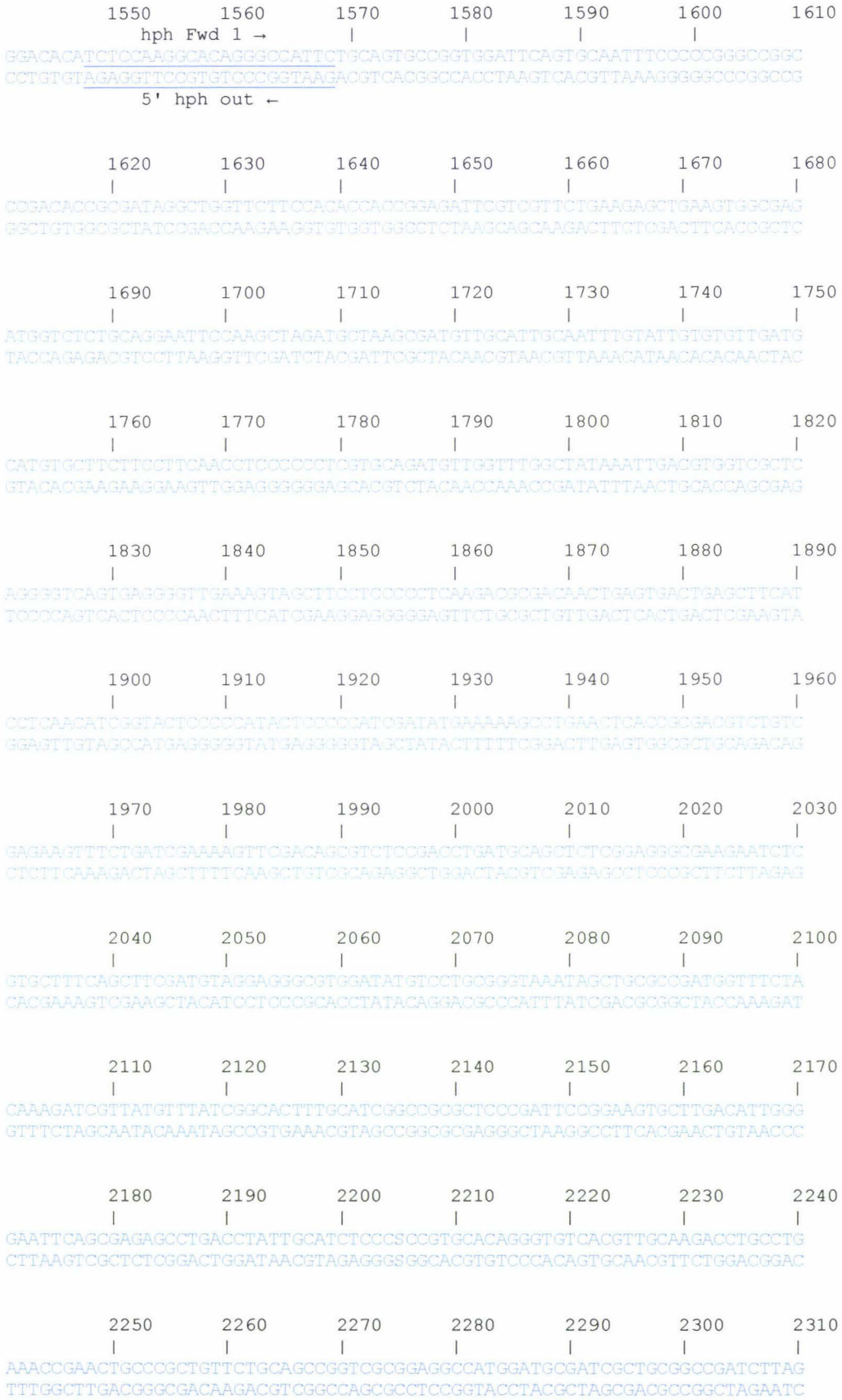
The primers used for checking the vector integrity:
Pks CB2
5' hph out
3' hph out
OT pks KO/4

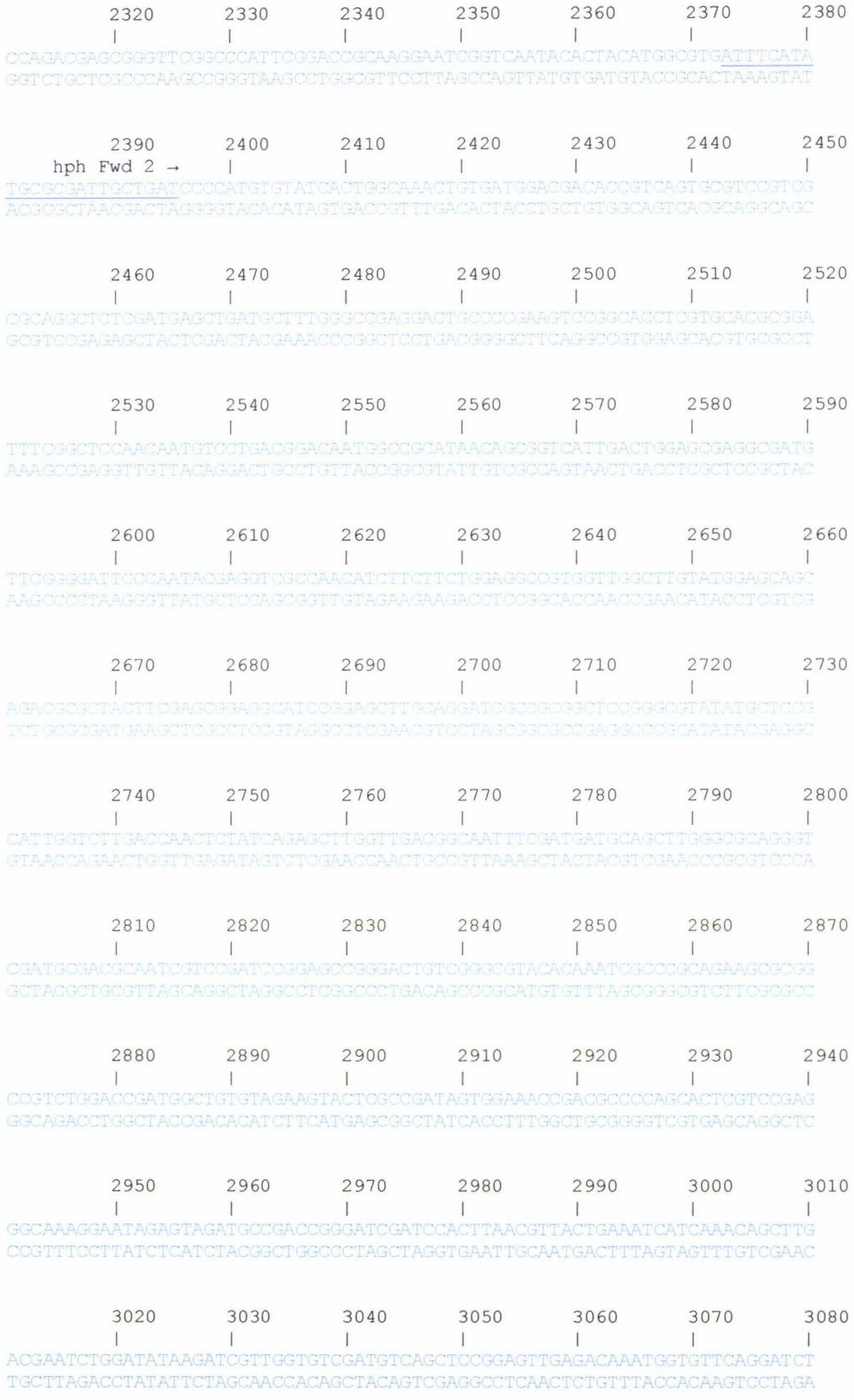
Primers used for sequencing the disruption vector:
Pks CB2
OT pks KO/4
hph Fwd1
hph Fwd2
hph Rev1

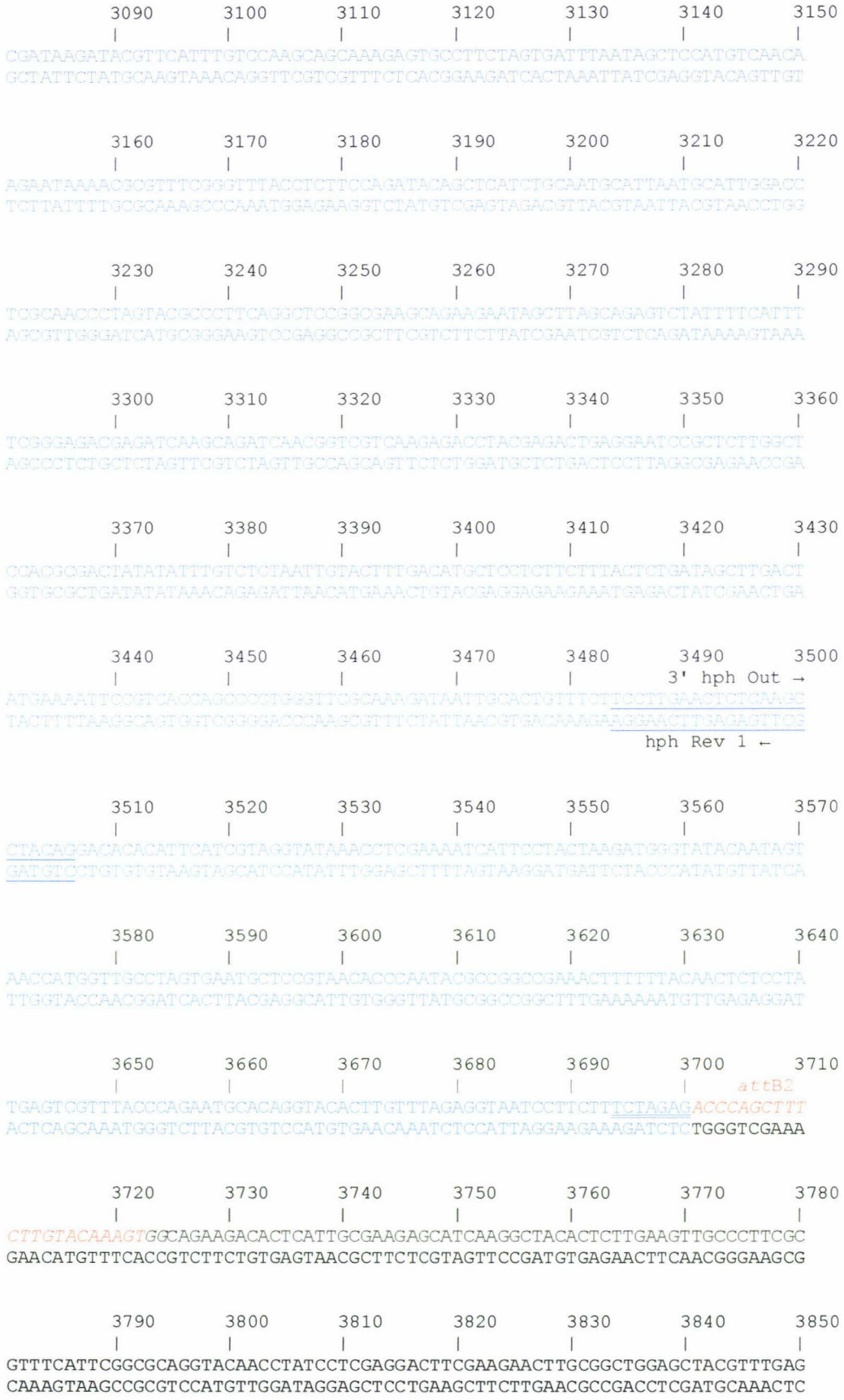
5' flanking region

```
        10         20         30         40         50         60         70
       pks CB2 →   |          |          |          |          |          |
TGAAGAAGTATGTCGCCGGCAAGCTGAACCCTGCCTTCCAGACGGCCCTCAGCTGCATCACACAACTCGG
ACTTCTTCATACAGCGGCCGTTCGACTTGGGACGGAAGGTCTGCCGGGAGTCGACGTAGTGTGTTGAGCC

        80         90         100        110        120        130        140
        |          |          |          |          |          |          |
TGTGTTCATGCGAGAGTTCCATGACTTCACCAAGCCATATCCACGACACGATAGCAGCTACGTGTTGGGT
ACACAAGTACGCTCTCAAGGTACTGAAGTGGTTCGGTATAGGTGCTGTGCTATCGTCGATGCACAACCCA

        150        160        170        180        190        200        210
        |          |          |          |          |          |          |
ATCTGCACCGGCTCGCTTGCCGCTGCGGCCGTCAGCTCCAGCAACTCTTTATCCGAGCTTCTGCCTATTG
TAGACGTGGCCGAGCGAACGGCGACGCCGGCAGTCGAGGTCGTTGAGAAATAGGCTCGAAGACGGATAAC

        220        230        240        250        260        270        280
        |          |          |          |          |          |          |
CCGTGCAAACGGCTTTGATCGCCTTTCGCCTCGGTCTGTGCGTCACAGACATGCGAGATCGTCTCGAAAG
GGCACGTTTGCCGAAACTAGCGGAAAGCGGAGCCAGACACGCAGTGTCTGTACGCTCTAGCAGAGCTTTC

        290        300        310        320        330        340        350
        |          |          |          |          |          |          |
CTCGGAAGAAGACCGCACCCAGCCTTGGTCGGTAGTTCTGTTCGACACAGACGAGCAGACTGTTACCAAG
GAGCCTTCTTCTGGCGTGGGTCGGAACCAGCCATCAAGACAAGCTGTGTCTGCTCGTCTGACAATGGTTC

        360        370        380        390        400        410        420
        |          |          |          |          |          |          |
GCCATCAAAGACTTCTGCACATCCAACGTCCTCCCGAAGACCAAGCAGCCCTGGATTACTTCCGCGTCAT
CGGTAGTTTCTGAAGACGTGTAGGTTGCAGGAGGGCTTCTGGTTCGTCGGGACCTAATGAAGGCGCAGTA

        430        440        450        460        470        480        490
        |          |          |          |          |          |          |
CGAAGACAATCACCATCAGCGGAGCACCACGTGTGCTGAAGAAGTTGTCGCAAGAGCCTGCGCTCAAGGA
GCTTCTGTTAGTGGTAGTCGCCTCGTGGTGCACACGACTTCTTCAACAGCGTTCTCGGACGCGAGTTCCT

        500        510        520        530        540        550        560
        |          |          |          |          |          |          |
CAAAAAGACCAGACAGATCCCGATTTATGTCCCAGCACACAATTCTGCCCTCTTCACCCCCGAAGATGTC
GTTTTTCTGGTCTGTCTAGGGCTAAATACAGGGTCGTGTGTTAAGACGGGAGAAGTGGGGGCTTCTACAG

        570        580        590        600        610        620        630
        |          |          |          |          |          |          |
AAGTCTATCCTTGAGACTACCCCTGTCGACACTTGGAGCAACTACCCTACCAAGCTTCCATTCATCTCGA
TTCAGATAGGAACTCTGATGGGGACAGCTGTGAACCTCGTTGATGGGATGGTTCGAAGGTAAGTAGAGCT

        640        650        660        670        680        690        700
        |          |          |          |          |          |          |
GCGTTTCTGGCAAGATGGCTTGGGCAGACAACTACCTTGCAGTAATCCATCTCGCCCTCAATCAGTGCCT
CGCAAAGACCGTTCTACCGAACCCGTCTGTTGATGGAACGTCATTAGGTAGAGCGGGAGTTAGTCACGGA

        710        720        730        740        750        760        770
        |          |          |          |          |          |          |
GCTCGAGAGCATTGGCTGGGGTAAGGTCGAGACTGAGCTCCCAAGGCTCCTCAAGTCTCGCGGCGCAGAA
CGAGCTCTCGTAACCGACCCCATTCCAGCTCTGACTCGAGGGTTCCGAGGAGTTCAGAGCGCCGCGTCTT
```

```
         780        790        800        810        820        830        840
         |          |          |          |          |          |          |
AATGTGCTCATCACGCCAATCACCACCTCTGCCGACCGTGCTCTGTCGGCTGCACTCAGCCCAACGATCT
TTACACGAGTAGTGCGGTTAGTGGTGGAGACGGCTGGCACGAGACAGCCGACGTGAGTCGGGTTGCTAGA

         850        860        870        880        890        900        910
         |          |          |          |          |          |          |
CCAATATCGAGGTCGAGAAGCCCACAATCAACGAATCTTTCGCCCATAGACCTGGTTCGGGAAAGAGCAA
GGTTATAGCTCCAGCTCTTCGGGTGTTAGTTGCTTAGAAAGCGGGTATCTGGACCAAGCCCTTTCTCGTT

         920        930        940        950        960        970        980
         |          |          |          |          |          |          |
GCTTGCTATCGTCTCAATGTCTGGTCGCTTCCCAGAGGCACAAAGCACCGACGCTTTTTGGGATCTGCTC
CGAACGATAGCAGAGTTACAGACCAGCGAAGGGTCTCCGTGTTTCGTGGCTGCGAAAAACCCTAGACGAG

         990        1000       1010       1020       1030       1040       1050
         |          |          |          |          |          |          |
TACAAGGGGCTTGACGTGGTCAAGGAAGTGCCCAAACGTCGTTGGGACGTCGAGACTCACGTCGACCCAA
ATGTTCCCCGAACTGCACCAGTTCCTTCACGGGTTTGCAGCAACCCTGCAGCTCTGAGTGCAGCTGGGTT

         1060       1070       1080       1090       1100       1110       1120
         |          |          |          |          |          |          |
CTGGGCGTGCTCGCAACAAGGGTGCGACCAAATGGGGCTGCTGGCTCGATTTCGCCGGAGAATTCGACCC
GACCCGCACGAGCGTTGTTCCCACGCTGGTTTACCCCGACGACCGAGCTAAAGCGGCCTCTTAAGCTGGG

         1130       1140       1150       1160       1170       1180       1190
         |          |          |          |          |          |          |
TCGCTTCTTCAGCATCTCACCCAAGGAAGCACCACAAATGGACCCCGCCCAGCGTATGGCTTTGATGTCG
AGCGAAGAAGTCGTAGAGTGGGTTCCTTCGTGGTGTTTACCTGGGGCGGGTCGCATACCGAAACTACAGC

         1200       1210       1220       1230       1240       1250       1260
         |          |          |          |          |          |          |
ACCTGGGAAGCGATGGAACGTGGTGGCATCGTCCCGGATACTACGCCCTCAACACAAAGAAATCGAATTG
TGGACCCTTCGCTACCTTGCACCACCGTAGCAGGGCCTATGATGCGGGAGTTGTGTTTCTTTAGCTTAAC

         1270       1280       1290       1300       1310       1320       1330
         |          |          |          |          |          attB1       |
GTGTCTTCCATGGCGTTACCTCCAACGACTGGATGGAGACCAACACCAAGTTTGTACAAAAAAGCAGGCT
CACAGAAGGTACCGCAATGGAGGTTGCTGACCTACCTCTGGTTGTGGTTCAAACATGTTTTTTCGTCCGA

         1340       1350       1360       1370       1380       1390       1400
         |          |          |          |          |          |          |
TCTAGAGTCGACGTATCGATAAGCTTGATACTAGTTATATTAATGGAAGGGTATATACCACGCGTTGGAC
AGATCTCAGCTGCATAGCTATTCGAACTATGATCAATATAATTACCTTCCCATATATGGTGCGCAACCTG

         1410       1420       1430       1440       1450       1460       1470
         |          |          |          |          |          |          |
CTTGGGACCTGCATTATAGCTTCCCGTTAGGTATAATTACCGTTGTTATAGCAGCCAATCAAGCCACCAC
GAACCCTGGACGTAATATCGAAGGGCAATCCATATTAATGGCAACAATATCGTCGGTTAGTTCGGTGGTG

         1480       1490       1500       1510       1520       1530       1540
         |          |          |          |          |          |          |
GCTTGACCGGGGACGGCGAATCCCCGGGAATTGAAATAAATTGCAATTCAGGTCAATGCGGCCAGCGATT
CGAACTGGCCCCTGCCGCTTAGGGGCCCTTAACTTTATTTAACGTTAAGTCCAGTTACGCCGGTCGCTAA
```


1550 1560 1570 1580 1590 1600 1610
hph Fwd 1 →
GGACACATCTCCAAGGCACAGGGCCATTCTGCAGTGCCGGTGGATTCAGTGCAATTTCCCCCGGGCCGGC
CCTGTGTAGAGGTTCCGTGTCCCGGTAAGACGTCACGGCCACCTAAGTCACGTTAAAGGGGGCCCGGCCG
5' hph out ←
1620 1630 1640 1650 1660 1670 1680
CCGACACCGCGATAGGCTGGTTCTTCCACACCACCGGAGATTCGTCGTTCTGAAGAGCTGAAGTGGCGAG
GGCTGTGGCGCTATCCGACCAAGAAGGTGTGGTGGCCTCTAAGCAGCAAGACTTCTCGACTTCACCGCTC
1690 1700 1710 1720 1730 1740 1750
ATGGTCTCTGCAGGAATTCCAAGCTAGATGCTAAGCGATGTTGCATTGCAATTTGTATTGTGTGTTGATG
TACCAGAGACGTCCTTAAGGTTCGATCTACGATTCGCTACAACGTAACGTTAAACATAACACACAACTAC
1760 1770 1780 1790 1800 1810 1820
CATGTGCTTCTTCCTTCAACCTCCCCCCTCGTGCAGATGTTGGTTTGGCTATAAATTGACGTGGTCGCTC
GTACACGAAGAAGGAAGTTGGAGGGGGGAGCACGTCTACAACCAAACCGATATTTAACTGCACCAGCGAG
1830 1840 1850 1860 1870 1880 1890
AGGGGTCAGTGAGGGGTTGAAAGTAGCTTCCTCCCCCTCAAGACGCGACAACTGAGTGACTGAGCTTCAT
TCCCCAGTCACTCCCCAACTTTCATCGAAGGAGGGGGAGTTCTGCGCTGTTGACTCACTGACTCGAAGTA
1900 1910 1920 1930 1940 1950 1960
CCTCAACATCGGTACTCCCCCATACTCCCCCATCGATATGAAAAAGCCTGAACTCACCGCGACGTCTGTC
GGAGTTGTAGCCATGAGGGGGTATGAGGGGGTAGCTATACTTTTTCGGACTTGAGTGGCGCTGCAGACAG
1970 1980 1990 2000 2010 2020 2030
GAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTC
CTCTTCAAAGACTAGCTTTTCAAGCTTGTCGCAGAGGCTGGACTACGTCGAGAGCCTCCGCTTCTTAGAG
2040 2050 2060 2070 2080 2090 2100
GTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTA
CACGAAAGTCGAAGCTACATCCTCCCGCACCTATACAGGACGCCCATTTATCGACGCGGCTACCAAAGAT
2110 2120 2130 2140 2150 2160 2170
CAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGG
GTTTCTAGCAATACAAATAGCCGTGAAACGTAGCCGGCGCGAGGGCTAAGGCCTTCACGAACTGTAACCC
2180 2190 2200 2210 2220 2230 2240
GAATTCAGCGAGAGCCTGACCTATTGCATCTCCCSCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTG
CTTAAGTCGCTCTCGGACTGGATAACGTAGAGGGSGGCACGTGTCCCACAGTGCAACGTTCTGGACGGAC
2250 2260 2270 2280 2290 2300 2310
AAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAG
TTTGGCTTGACGGGCGACAAGACGTCGGCCAGCGCCTCCGGTACCTACGCTAGCGACGCCGGCTAGAATC

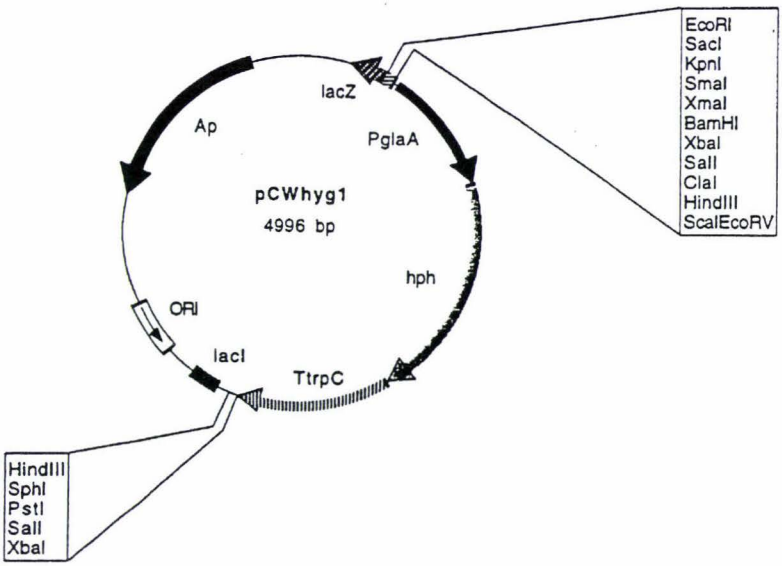

2320 2330 2340 2350 2360 2370 2380
CCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTCATA
GGTCTGCTCGCCCAAGCCGGGTAAGCCTGGCGTTCCTTAGCCAGTTATGTGATGTACCGCACTAAAGTAT
2390 2400 2410 2420 2430 2440 2450
hph Fwd 2 →
TGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCG
ACGCGCTAACGACTAGGGGTACACATAGTGACCGTTTGACACTACCTGCTGTGGCAGTCACGCAGGCAGC
2460 2470 2480 2490 2500 2510 2520
CGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGA
GCGTCCGAGAGCTACTCGACTACGAAACCCGGGCTCCTGACGGGGCTTCAGGCCGTGGAGCACGTGCGCCT
2530 2540 2550 2560 2570 2580 2590
TTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGATG
AAAGCCGAGGTTGTTACAGGACTGCCTGTTACCGGCGTATTGTCGCCAGTAACTGACCTCGCTCCGCTAC
2600 2610 2620 2630 2640 2650 2660
TTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGC
AAGCCCCTAAGGGTTATGCTCCAGCGGTTGTAGAAGAAGACCTCCGGCACCAACCGAACATACCTCGTCG
2670 2680 2690 2700 2710 2720 2730
ACACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCG
TCTGCGCGATGAAGCTCGCCTCCGTAGGGCTCGAACGTCCTAGCGGCGCCGAGGCCCGCATATACGAGGC
2740 2750 2760 2770 2780 2790 2800
CATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCAGGGT
GTAACCAGAACTGGTTGAGATAGTCTCGAACCAACTGCCGTTAAAGCTACTACGTCGAACCCGCGTCCCA
2810 2820 2830 2840 2850 2860 2870
CGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCCCGCAGAAGCGCGG
GCTACGCTGCGTTAGCAGGCTAGGCCTCGGCCCTGACAGCCCGCATGTGTTTAGCGGGCGTCTTCGCGCC
2880 2890 2900 2910 2920 2930 2940
CCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAG
GGCAGACCTGGCTACCGACACATCTTCATGAGCGGCTATCACCTTTGGCTGCGGGGTCGTGAGCAGGCTC
2950 2960 2970 2980 2990 3000 3010
GGCAAAGGAATAGAGTAGATGCCGACCGGGATCGATCCACTTAACGTTACTGAAATCATCAAACAGCTTG
CCGTTTCCTTATCTCATCTACGGCTGGCCCTAGCTAGGTGAATTGCAATGACTTTAGTAGTTTGTCGAAC
3020 3030 3040 3050 3060 3070 3080
ACGAATCTGGATATAAGATCGTTGGTGTCGATGTCAGCTCCGGAGTTGAGACAAATGGTGTTCAGGATCT
TGCTTAGACCTATATTCTAGCAACCACAGCTACAGTCGAGGCCTCAACTCTGTTTACCACAAGTCCTAGA


3090 3100 3110 3120 3130 3140 3150
CGATAAGATACGTTCATTTGTCCAAGCAGCAAAGAGTGCCTTCTAGTGATTTAATAGCTCCATGTCAACA
GCTATTCTATGCAAGTAAACAGGTTCGTCGTTTCTCACGGAAGATCACTAAATTATCGAGGTACAGTTGT
3160 3170 3180 3190 3200 3210 3220
AGAATAAAACGCGTTTCGGGTTTACCTCTTCCAGATACAGCTCATCTGCAATGCATTAATGCATTGGACC
TCTTATTTTGCGCAAAGCCCAAATGGAGAAGGTCTATGTCGAGTAGACGTTACGTAATTACGTAACCTGG
3230 3240 3250 3260 3270 3280 3290
TCGCAACCCTAGTACGCCCTTCAGGCTCCGGCGAAGCAGAAGAATAGCTTAGCAGAGTCTATTTTCATTT
AGCGTTGGGATCATGCGGGAAGTCCGAGGCCGCTTCGTCTTCTTATCGAATCGTCTCAGATAAAAGTAAA
3300 3310 3320 3330 3340 3350 3360
TCGGGAGACGAGATCAAGCAGATCAACGGTCGTCAAGAGACCTACGAGACTGAGGAATCCGCTCTTGGCT
AGCCCTCTGCTCTAGTTCGTCTAGTTGCCAGCAGTTCTCTGGATGCTCTGACTCCTTAGGCGAGAACCGA
3370 3380 3390 3400 3410 3420 3430
CCACGCGACTATATATTTGTCTCTAATTGTACTTTGACATGCTCCTCTTCTTTACTCTGATAGCTTGACT
GGTGCGCTGATATATAAACAGAGATTAACATGAAACTGTACGAGGAGAAGAAATGAGACTATCGAACTGA
3440 3450 3460 3470 3480 3490 3500
3' hph Out →
ATGAAAATTCCGTCAGCCAGCCCCTGGGTTCGCAAAGATAATTGCACTGTTTCTTCCTTGAACTCTCAAGC
TACTTTTAAGGCAGTCGGTCGGGGACCCAAGCGTTTCTATTAACGTGACAAAGAAGGAACTTGAGAGTTCG
hph Rev 1 ←
3510 3520 3530 3540 3550 3560 3570
CTACAGGACACACATTCATCGTAGGTATAAACCTCGAAAATCATTCCTACTAAGATGGGTATACAATAGT
GATGTCCTGTGTGTAAGTAGCATCCATATTTGGAGCTTTTAGTAAGGATGATTCTACCCATATGTTATCA
3580 3590 3600 3610 3620 3630 3640
AACCATGGTTGCCTAGTGAATGCTCCGTAACACCCAATACGCCGGCCGAAACTTTTTTACAACTCTCCTA
TTGGTACCAACGGATCACTTACGAGGCATTGTGGGTTATGCGGCCGGCTTTGAAAAAATGTTGAGAGGAT
3650 3660 3670 3680 3690 3700 3710
attB2
TGAGTCGTTTACCCAGAATGCACAGGTACACTTGTTTAGAGGTAATCCTTCTTTCTAGAGACCCAGCTTT
ACTCAGCAAATGGGTCTTACGTGTCCATGTGAACAAATCTCCATTAGGAAGAAAGATCTCTGGGTCGAAA
3720 3730 3740 3750 3760 3770 3780
CTTGTACAAAGTGGCAGAAGACACTCATTGCGAAGAGCATCAAGGCTACACTCTTGAAGTTGCCCTTCGC
GAACATGTTTCACCGTCTTCTGTGAGTAACGCTTCTCGTAGTTCCGATGTGAGAACTTCAACGGGAAGCG
3790 3800 3810 3820 3830 3840 3850
GTTTCATTCGGCGCAGGTACAACCTATCCTCGAGGACTTCGAAGAACTTGCGGCTGGAGCTACGTTTGAG
CAAAGTAAGCCGCGTCCATGTTGGATAGGAGCTCCTGAAGCTTCTTGAACGCCGACCTCGATGCAAACTC

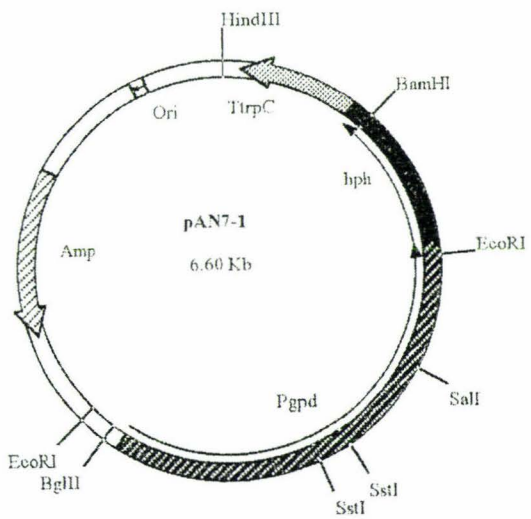
```
         3860       3870       3880       3890       3900       3910       3920
         |          |          |          |          |          |          |
AAGCCCAAGCTTGCGGTCATTTCCCCGCTACTGGGCAGTGTGGTCGAGGACGAGGGAGTCGTTGGACCCA
TTCGGGTTCGAACGCCAGTAAAGGGGCGATGACCCGTCACACCAGCTCCTGCTCCCTCAGCAACCTGGGT

         3930       3940       3950       3960       3970       3980       3990
         |          |          |          |          |          |          |
ACTACCTTGCACGCCACTGCCGTGAGGCGGTCGGAATGGTCAAAGCCCTCGGAGTGGCGAAGGAGAAGGG
TGATGGAACGTGCGGTGACGGCACTCCGCCAGCCTTACCAGTTTCGGGAGCCTCACCGCTTCCTCTTCCC

         4000       4010       4020       4030       4040       4050       4060
         |          |          |          |          |          |          |
TATAATCAACGAGAAGACCATCGTCATTGAGATTGGTCCTAAGCCGCTTCTCTGCGGAATGATCAAGAAC
ATATTAGTTGCTCTTCTGGTAGCAGTAACTCTAACCAGGATTCGGCGAAGAGACGCCTTACTAGTTCTTG

         4070       4080       4090       4100       4110       4120       4130
         |          |          |          |          |          |          |
ATACTCGGCCAGAACATCGTAGCCTTGCCTACGTTGAAGGACAAGGGTCCAGACGTCTGGCAGAACCTCT
TATGAGCCGGTCTTGTAGCATCGGAACGGATGCAACTTCCTGTTCCCAGGTCTGCAGACCGTCTTGGAGA

         4140       4150       4160       4170       4180       4190       4200
         |          |          |          |          |          |          |
CGAACATCTTCACGACGCTCTACACCGGTGGTCTAGACATCAACTGGACTGCCTTCCACGCCCCCTTCGA
GCTTGTAGAAGTGCTGCGAGATGTGGCCACCAGATCTGTAGTTGACCTGACGGAAGGTGCGGGGGAAGCT

         4210       4220       4230       4240       4250       4260       4270
         |          |          |          |          |          |          |
GCCCGCGAAGAAGGTCCTGCAACTTCCTGATTATGGCTGGGATCTCAAGGATTACTTCATCCAGTATGAA
CGGGCGCTTCTTCCAGGACGTTGAAGGACTAATACCGACCCTAGAGTTCCTAATGAAGTAGGTCATACTT

         4280       4290       4300       4310       4320       4330       4340
         |          |          |          |          |          |          |
GGCGATTGGGTTCTGCATCGGCACAAGATCCACTGCAACTGTGCAGATGCTGGAAAGGATGTGCATAACA
CCGCTAACCCAAGACGTAGCCGTGTTCTAGGTGACGTTGACACGTCTACGACCTTTCCTACACGTATTGT

         4350       4360       4370       4380       4390       4400       4410
         |          |          |          |          |          |          |
CTTCGCACTACTGTCCTGGCAAACACACCTTCGCTGAGAATGTTGTCGTTCCTGGTGGGGCTCAGAAGGC
GAAGCGTGATGACAGGACCGTTTGTGTGGAAGCGACTCTTACAACAGCAAGGACCACCCCGAGTCTTCCG

         4420       4430       4440       4450       4460       4470       4480
         |          |          |          |          |          |          |
CGTTCAGGAAGCACCTGCGGCGAAGACAGAGACGAAGAAGATGTCGAAGCTGGATCCTACCAAGGAGGCG
GCAAGTCCTTCGTGGACGCCGCTTCTGTCTCTGCTTCTTCTACAGCTTCGACCTAGGATGGTTCCTCCGC

         4490       4500       4510       4520       4530
         |          |          |          |          |
TATCCGGGCATTCCGCTCACCACGACCGTGCACAAGGTCATTGAAGAGAAGACGGAGCC   3' flanking region
ATAGGCCCGTAAGGCGAGTGGTGCTGGCACGTGTTCCAGTAACTTCTCTTCTGCCTCGG
                                        OT pks KO/4 ←
```

APPENDIX 5: PLASMID MAPS

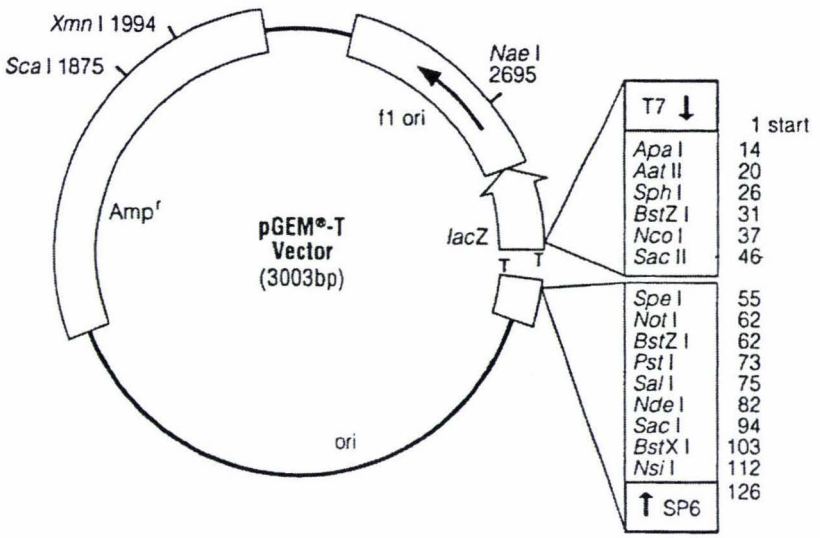
Appendix 5A: pCWhyg (Figure from Seconi, 2001)

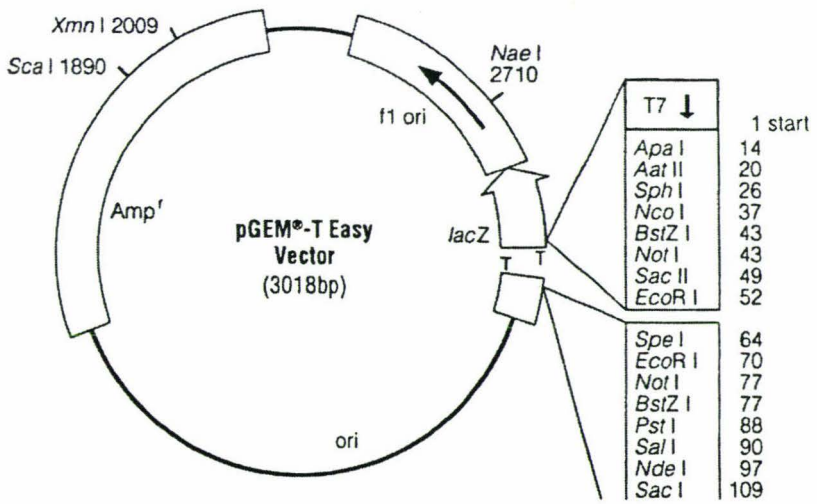


pAN7-1 (Figure From Seconi, 2001)

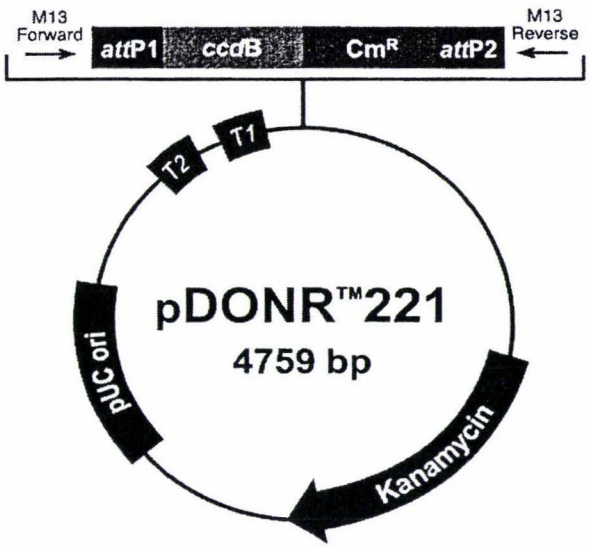


Appendix 5B: pGEM-T and pGEM-T Easy

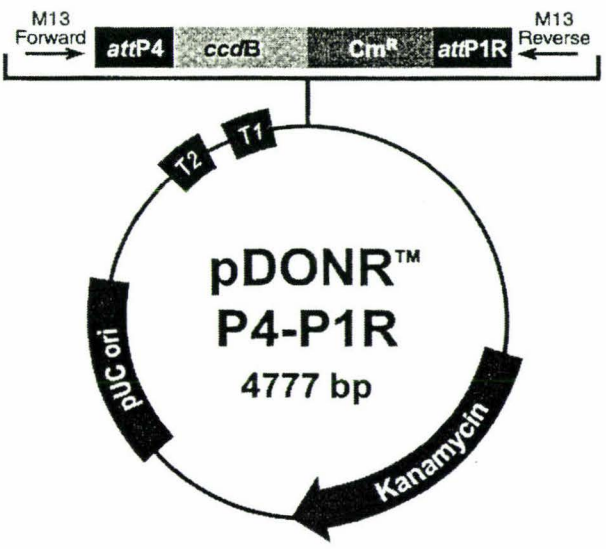




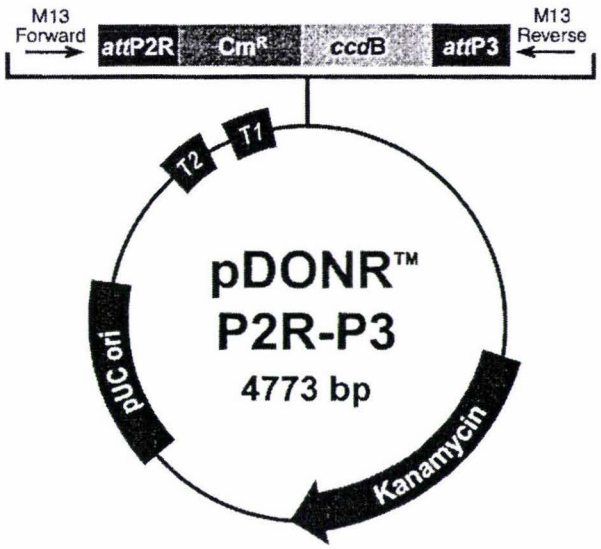
Appendix 5C: pDONR 221 (Figure From Invitrogen Multisite Gateway Three-Fragment Vector Construction Manual)



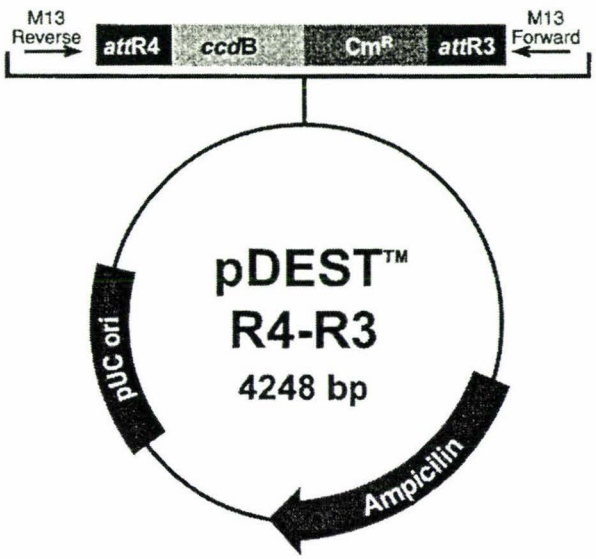
pDONR P4-P1R (Figure From Invitrogen Multisite Gateway Three-Fragment Vector Construction Manual)



Appendix 5D: pDONR P2R-P3 (Figure From Invitrogen Multisite Gateway Three-Fragment Vector Construction Manual)



pDEST R4-R3 (Figure From Invitrogen Multisite Gateway Three-Fragment Vector Construction Manual)



Appendix 5E: pMSGW (Figure From Invitrogen Multisite Gateway Three-Fragment Vector Construction Manual)