Identification and Characterization of Dothistromin Biosynthetic Genes in the Peanut Pathogen *Passalora arachidicola*

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

Yanan Guo

2008

Abstract

Dothistromin (DOTH) is a secondary metabolite produced by the fungal peanut pathogen *Passalora arachidicola* and pine needle pathogen *Dothistroma septosporum*. The chemical structure of DOTH is similar to a precursor of aflatoxin (AF) and sterigmatocystin (ST), which are secondary metabolites produced by *Aspergillus* species.

A size fractionated genomic library was made and 11 putative DOTH genes were identified in *P. arachidicola*. The DOTH genes in *P. arachidicola* were compared to DOTH genes in *D. septosporum* as well as to AF and ST genes in *Aspergillus* species. The DOTH gene products in *P. arachidicola* showed 73 - 96% amino acid identity to DOTH genes in *D. septosporum* and 50 - 69% amino acid identity to AF/ST genes in *Aspergillus*. The DOTH biosynthesis genes in *P. arachidicola* had similar gene organization and direction of transcription to DOTH biosynthesis genes in *D. septosporum* and is similar in that 11 putative DOTH genes are separated into three mini-clusters. This differs from the AF/ST clusters in which 25 AF/ST genes are tightly clustered in a 70 kb region.

Identification of transcription factor binding sites upstream of DOTH genes in *P. arachidicola* and *D. septosporum* suggested similar co-regulation of DOTH gene expression in *P. arachidicola* and *D. septosporum*.

Tandem and inverted repeat sequences were identified in intergenic regions in the *P. arachidicola* DOTH gene cluster, but the distribution of those repeats appears to be random. This suggests that the fragmentation of the DOTH biosynthesis gene cluster is not due to retrotransposon activity or recombination between repeat sequences. The DOTH biosynthesis gene clusters in *P. arachidicola* and *D. septosporum* could be ancestral to AF/ST biosynthesis clusters in *Aspergillus* species.

Acknowledgements

I would like to gratefully thank my supervisor Dr Rosie Bradshaw for her constant encouragement and guidance during the course of my Masters. Thanks for your kindness and patience when I have questions. Thanks for your for helping me with my English. You are a great supervisor and I couldn't ask for more.

To my lab colleagues, thanks for Shuguang for helpful technical support and advice, your willingness to listen and sharing your knowledge make it easier for me to start my project. Thanks for Zhilun and Arne for technical assistance and advice. Thanks for Justin, Naydene, Rebecca. It is a pleasure to work with all of you.

I would also like to thank Kathy Hamilton for helping me arrange my course. Thanks for Dr Kathryn Stowell for giving me advice when I start my Master project. Many thanks to other IMBS staff who gave me help over the duration of my studies.

Finally I would like to thank my parents, although it is far from home, you always been so supportive, encouraging and understanding over the duration of my studies. To my husband, thanks for believing in me and helping me with revision before exams.

Table of Contents

Chapter Two - Materials and Methods

Chapter Three - Results

List of Tables

List of Figures

Abbreviations

Abbreviation	Meaning
AF	aflatoxin
AVF	averufin
AVN	averantin
bp	base pair
cDNA	complementary deoxyribonucleic acid
°C	degree Celsius
DMST	demethylsterigmatocystin
Dnase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DOTH	dothistromin
G	gram
HAVN	5'-hydroxy-averantin
ITS	internal transcribed spacer region
IPTG	isopropyl-β-d-thiogalactoside
kb	kilobase pair
L	litre
M	molar
Mb	megabase
MFS transporter	major facilitator superfamily transporter
ml	milliliter
mM	millimolar
NOR	norsolorinic avid
OAVN	oxoaverantin
OMST	o-methylsterigmatocystin
ORF	open reading frame
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
ST	sterigmatocysin
µl	microlitre
µM	micromolar
µg	microgram
UV	ultraviolet
VAL	versiconal
VER	versicolorin
VHA	versiconal hemiacetal acetate
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter one – Introduction

1.1 *Passalora arachidicola* and early leaf spot disease

1.1.1 Early leaf spot disease

Early spot disease cause by *P. arachidicola* is characterized by brown circular spots surrounded by yellow halo on the upper surface of peanut leaves. Brown dots first appear on the upper surface of peanut leaves. As the dots enlarge to become circular brown spots, a yellow halo develops around each spot (Figure 1.1). Spots can also develop on leaf petioles and plant stems after numerous lesions are formed on leaflets. Without treatment, the disease causes premature leaf drop and up to 75% yield reduction. Even with fungicide treatment, it can still lead to 5% yield reduction (Damicone and Melouk, 1998; Kucharek, 2000).

Figure 1.1 Early leaf spot on peanut


(Damicone and Melouk, 1998)

P. arachidicola remains as stroma or mycelium on peanut residues in the field. After a period of wetness, primary infection occurs on the adaxial surface of lower peanut leaves. Conidia produced by *P. arachidicola* are the main initial inoculum source. The conidia penetrate either through stomata or directly through epidermal cells but do not produce intracellular haustoria. Symptoms develop within 10-14 days at 21°C and more spores are produced. The secondary spread is by conidia dispersed by wind, water splash or insects (Culbreath et al., 2002; Damicone and Melouk, 1998; Kucharek, 2000).

Spray of fungicides and selecting resistant peanut cultivars are the two main methods used to control early spot disease. The fungicides used include chlorothalonil, tebuconazole or azoxystrobin. The spray of fungicides begins approximately 30 to 40 days after planting and every 10 to 14 days up to 2 weeks before harvest date. During an extremely wet season, a shortened spray interval can be applied (Culbreath et al., 2002; Damicone and Melouk, 1998; Nutsugah et al., 2007). Selection of resistant peanut cultivars also increases yield. The selective experiments can be done in greenhouse or field. Ricker (1985) tested 20 genotypes of peanut. Phenotypes such as rate of infection, number of lesions per leaf, lesion diameter leaf defoliation and sporulation of fungi were compared between each genotype and genotype NC3033 had highest overall resistance to early leaf spot disease (Ricker et al., 1985).

1.1.2 *Passalora arachidicola*

1.1.2.1 The fungus

Phylogenetic analysis by comparing the internal transcribed spacers (ITS) and the adjacent 5' end of the 25S rRNA gene sequences of different *Passalora* species showed this anamorphic genus to belong to one clade of the teleomorphic *Mycosphaerella* genus (Goodwin et al., 2001; Stewart et al., 1999). *Passalora arachidicola* (previously known as *Cercospora arachidicola*, teleomorphic name *Mycosphaerella arachidis*) is the fungal pathogen that causes peanut early leaf spot disease.

1.1.2.2 *Passalora arachidicola* produces dothistromin (DOTH) toxin

When Assante screened 61 species of the *Cercospora* genus, she found that 8 *Cercospora* species produce DOTH, a polyketide secondary metabolite distinctively red in colour (Assante et al., 1977). Column chromatography and mass spectral analysis of a liquid culture showed that *P. arachidicola* (previously called *Cercospora arachidicola*) produces DOTH and averufin as secondary metabolites. Subsequent work using solid agar resulted in greatly increased yields of DOTH and

averufin by *P. arachidicola* (Stoessl, 1984; Stoessl and Stothers, 1985). Minor metabolites such as versicolorin B, averantin and nidurufin were also produced by *P. arachidicola* in small amounts (Stoessl and Stothers, 1985).

1.2 Dothistromin and biosynthesis gene clusters

1.2.1 Structure and biological activity of dothistromin

DOTH is a major secondary metabolite produced by *P. arachidicola* and the pine needle pathogen *Dothistroma septosporum*. The structure of DOTH was identified by mass spectrometry and nuclear magnetic resonance (NMR) (Bear et al., 1972). A NMR study using ^{13}C labelled DOTH precursors showed that the DOTH bistetrahydrofurano side chain has a similar structure to the aflatoxin (AF) side chain. Furthermore the DOTH tetrahydro-2-hydroxy- bisfuran moiety is similar to that of AF precursor versicolorin B, although DOTH has different hydroxyl groups in the anthraquinone ring (Shaw et al., 1978) (Figure 1.2). Sterigmatocystin (ST) and AF are secondary metabolites synthesized by fungi *A. nidulans, A. parasiticus* and related *Aspergillus* species. ST is the precursor of AF, so the ST and AF biosynthesis share many steps. AF toxin is highly carcinogenic. Due to the risk to agriculture production and human health, the ST and AF biosynthesis pathways have been extensively investigated. Because of the similarity of DOTH with AF, the DOTH biosynthesis genes were initially identified in *D. septosporum* using aflatoxin genes as hybridisation probes (Bradshaw et al., 2002).

Figure 1.2 Comparison of the structures of DOTH and AF precursor

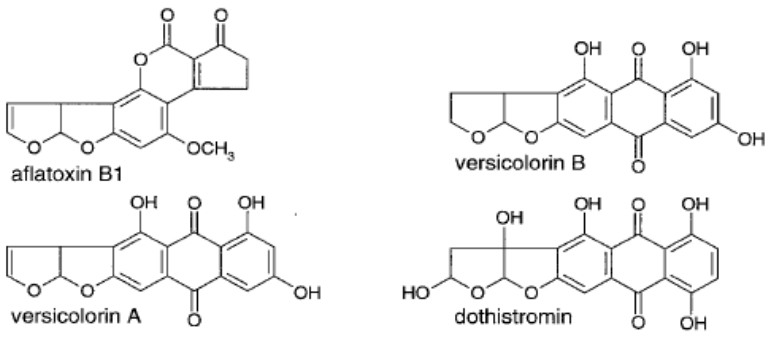


Figure 1.2 Chemical structures of AF B1, versicolorin A, versicolorin B, and DOTH (Bradshaw et al., 2002)

Studies showed that DOTH is a weak mutagen and clastogen. DOTH showed a significant mutagenic activity in an Ames assay and in a screen for specific induced mutations in Chinese hamster fibroblasts (Elliott et al., 1989; Ferguson et al., 1986; McLarin and Ferguson, 1985). The clastogenic activity of DOTH has been demonstrated by observations that DOTH induces a dosage dependent increase in sister-chromatid exchange frequency in human lymphocytes (McLarin and Ferguson, 1985; Skinnider et al., 1989). DOTH is toxic to a wide range of cell types. There is an effective dosage for DOTH to be mutagenic/clastogenic or toxic. At low dosage, DOTH is more likely to cause genetic damage, at high dosage, DOTH is more likely to be toxic to cells before the cells' genetic material has been damaged (Ferguson et al., 1986).

Dothistroma pini was described to be the causal agent of dothistroma red-band needle blight in 1941 by Hulbary (Bradshaw, 2004; Gibson, 1972). A taxonomic revision renamed most *D. pini* as *D. septosporum* (Barnes et al., 2004). DOTH toxin can be isolated from lesions of infected pine needles and from cultures *in vitro* (Bassett et al., 1970). A strong host defence response to invading pathogens contributes to the dothistroma needle blight symptoms (Franich et al., 1986). Benzoic acid is a phytoalexin, produced by the plant, and accumulates at high concentrations surrounding the lesions. The role of benzoic acid is to inhibit Dothistroma, however, it is also toxic to the plant. Purified DOTH is broken down to oxalic acid and CO_2 in the needle tissue, with only 10-20% of the toxin remaining after 24 hours (Bradshaw et al., 2000; Harvey et al., 1976). Because the necrotic lesion continues to expand after most of the DOTH has been degraded, this showed that damage of pine needle is probably mainly due to the plant's defence response rather than the toxicity of DOTH (Franich et al., 1986). Studies showed that in *P. radiata*, even in the absence of DOTH, the defence response to *D. septosporum* has been observed (Hotter, 1997). Further studies showed that DOTH is not required for *D. septosporum* to cause red-band needle blight disease, but may provide an advantage to *D. septosporum* in growth with other pine fungi (Schwelm et al., in

press). All those studies suggested that DOTH is a possible virulence factor rather than a pathogenicity factor.

The identification of *P. arachidicola* as a producer of DOTH suggested that it may have a similar DOTH biosynthesis cluster as that seen in *D. septosporum*. Furthermore, aflatoxins are produced in peanuts as a result of invasion and growth of *A. flavus* and *A. parasiticus*. The dothistromin biosynthesis cluster in *P. arachidicola* may provide evidence of how the dothistromin and aflatoxin gene clusters evolved.

1.2.2 Sterigmatocystin biosynthesis gene cluster

The ST biosynthesis pathway has been studied in detail in the model organism *A. nidulans* and involves at least 15 enzymatic activities. Studies showed that 25 genes clustered within a 60 kb DNA region are involved in ST biosynthesis (Table 1.1). Among those 25 genes, 24 genes ranging in size from 0.6 to 7.2 kb are co-ordinately induced only under ST production conditions and were named as *stc* genes. The *aflR* gene is involved in ST biosynthesis pathway regulation. It encodes a specific zinc binuclear DNA-binding protein and the AflR protein acts as a positive regulator by activating the transcription of other ST biosynthesis pathway genes (Brown et al., 1996b).

1.2.3 Aflatoxin biosynthesis gene clusters

Studies showed that biosynthesis of AF involves acetate→polyketide→norsolorinic acid (NOR)→averantin (AVN)→5'-hydroxy-averanti (HAVN)→oxoaverantin (OAVN)→averufin (AVF)→versiconal hemiacetal acetate (VHA)→versiconal (VAL) → versicolorin B (VERB) → versicolorin A (VERA) → demethylsterigmatocystin (DMST) → sterigmatocystin (ST) → *O*-methylsterigmatocystin (OMST)→AF and there are at least 23 enzymatic reactions involved (Yu et al., 2004a; Yu et al., 2005). To date 25 genes involved in

AF biosynthesis have been identified (Table 1.1). The genes within the AF biosynthesis cluster were initially named based on the substrate converted by the gene product, but were then renamed as *afl* genes. The AF biosynthesis genes are clustered within a 70 kb DNA region, on average, about 2.8 kb of chromosome DNA contains one gene (Yu et al., 2004a; Yu et al., 2004b). A 2 kb DNA region with no identifiable ORF was located at the 5' end of the cluster to mark the end of the AF biosynthesis cluster. A well-defined sugar utilization gene cluster which contains four genes delineated the 3' end of AF biosynthesis cluster (Cleveland and Bhatnagar., 1991). Among those 25 genes, 17 of them (*aflA* to *aflQ*) are involved in the major conversion steps from precursors to AF. This was confirmed by gene disruption experiments. Two genes *aflR* and *aflJ* are involved in AF biosynthesis pathway regulation. *aflR* is a positive regulator, it encodes a specific zinc binuclear DNA-binding protein which activates the transcription of other AF biosynthesis pathway genes (Yu et al., 2004a). *aflJ* is located adjacent to *aflR* gene in the AF biosynthesis cluster. Knockout of *aflJ* resulted in a 5-20 fold reduction of expression of some AF biosynthesis genes such as *aflC, aflD, aflM* and *aflP*, and reduced ability to synthesise some AF intermediates (Meyers et al., 1998; Yu et al., 2004a). Six additional (*aflT* to *aflY*) genes have been identified that are putatively involved in AF biosynthesis (Yu et al., 2004b).

Table 1.1 AF/ST pathway cluster genes and the roles of their gene products

Afl gene	Protein function	Step in pathway	*Stc* gene
aflA (*fas*-2)	fatty acid synthase α subunit	Acetate → polyketide	*stcJ*
aflB (*fas*-1)	fatty acid synthase β subunit	Acetate → polyketide	*stcK*
*afl*C (*pksA*)	polyketide synthase	Acetate → polyketide	*stcA*
aflD (*nor-1*)	ketoreductase (reductase)	NOR → AVN	*stcE*
aflE (*norA*)	NOR reductase /dehydrogenases	NOR → AVN	*stcV*
aflF (*norB*)	dehydrogenases	NOR → AVN	*stcG*
aflG (*avnA*)	Cytochrome P450 monooxygenase	AVN → HAVN	*stcF*
aflH (*adhA*)	alcohol dehydrogenase	HAVN → OAVN and AVF	
aflI (*avfA*)	oxidase	AVF→VHA	*stcO*
aflJ (*estA*)	esterase	VNA →VAL	*stcI*
aflK (*vbs*)	versicolorin B synthase	VAL → VERB	*stcN*
aflL (*verb*)	desaturase	VERB → VERA	
aflM (*ver-1*)	dehydrogenases / ketoreductase	VERA → DMST	*stcU*
aflN (*verA*)	monooxygenase	VERA → DMST	*stcW*
aflO (*omtB*)	*o*-methyltransferase B	DMST → ST	*stcP*
aflP (*omtA*)	*o*-methyltransferase A	ST → OMST	
aflQ (*ordA*)	oxidoreductase	OMST → Aflatoxin	
aflR (*aflR*)	transcription activator	pathway regulator	*aflR*
aflS (*aflJ*)	transcription enhancer	pathway regulator	
aflT (*aflT*)	Transmembrane protein	unknown	
aflU (*cypA*)	P450 monooxygenase	unknown	*stcL/stcS*
aflV (*cypX*)	P450 monooxygenase	unknown	*stcB*
aflW (*moxY*)	monooxygenase	unknown	*stcW*
aflX (*ordB*)	monooxygenase / oxidase	unknown	*stcC*
aflY (*hypA*)	hypothetical protein	VA →demethylsterigmatocystin	

1.2.4 Dothistromin biosynthesis gene cluster

Studies showed that DOTH shares many steps of the biosynthetic pathway with AF (Shaw et al., 1978). By using AF genes as hybridisation probes, genomic clones containing 11 putative DOTH genes have been found in *D. septosporum*. The first DOTH gene cluster contains four genes with proposed functions of ketoreductase (*Ds-dotA*), oxidase (*Ds-dotB*), major facilitator superfamily transporter (*Ds-dotC*), and thioesterase (*Ds-dotD*) (Bradshaw et al., 2002). The *dotA* gene encodes a 263-amino acids sequence, which is 80% identical to the *ver-1* product of *A. parasiticus*. A *Ds-dotA* gene replacement experiment showed that the *Ds-dotA* mutant did not produce DOTH but accumulated versicolorin A, this confirmed that *Ds-dotA* is a DOTH pathway gene and it has an essential role in DOTH biosynthesis in *D. septosporum*. The *Ds-dotB* gene has 24% amino acid identity to the *StcC* sterigmatocystin gene product of *A. nidulans*. The *Ds-dotC* gene has 31.2% amino acid identity to the *AflT* gene in *A. parasiticus*. It is proposed to act as a toxin pump in *D. septosporum*. The *Ds-dotD* gene encodes a 322 amino acid polypetide that has putative thioesterase enzyme activity (Bradshaw et al., 2002).

Further studies of DOTH gene cluster in *D. septosporum* using AF genes as hybridisation probe identified another clone containing a *Ds-pksA* gene and four additional genes (*Ds-cypA, Ds-moxA, Ds-avfA* and *Ds-epoA*) clustered immediately alongside *Ds-pksA* (Bradshaw et al., 2006). The *Ds-pksA* gene encodes a polyketide synthase with 55% amino acid identity to the AF PKS protein. A targeted gene replacement experiment showed that the *Ds-pksA* mutant did not produce DOTH and did not accumulate any known AF precursors, this not only confirmed the essential role of *Ds-pksA* gene in DOTH biosynthesis in *D. septosporum* but also showed that PksA acts at an early stage of DOTH biosynthesis (Bradshaw et al., 2006; Bradshaw and Zhang, 2006; Zhang et al., 2007). *Ds-cypA, Ds- moxA* and *Ds-avfA* are predicted to be orthologs of AF biosynthetic genes. The *Ds-epoA* gene has no homologue in the AF gene cluster, it is located between *Ds-avfA* and *Ds-moxA* genes and predicted to encode an epoxide hydrolase (Bradshaw et al., 2006).

A third clone containing the DOTH gene *Ds-vbsA* and *Ds-hexA* has also been found. The two *DS* genes were separated by two genes not involved in DOTH biosynthesis. *Ds-vbsA* shows almost 60% identity to AF *AflK*. Gene replacement of *Ds-vbsA* produced a mutant strain that had over 20-fold less DOTH than the wild type, it confirmed that *Ds-vbsA* was required for DOTH biosynthesis (Bradshaw et al., 2006; Bradshaw and Zhang, 2006).

The order and orientation of DOTH genes in *D. septosporum* is different from ST/AF gene cluster (Figure 1.3). The *dotA*, *pksA* and *vbsA* genes hybridize to a small 1.3-Mb chromosome, while the ST/AF gene cluster is located on a main chromosome (Bradshaw et al., 2006; Yu et al., 2004b). It is not known if the 1.3- Mb chromosome is conditionally dispensible or not. Only eleven DOTH genes have been identified in *D. septosporum* so far, it is predicted at least ten more DOTH genes remain to be found (Bradshaw et al., 2006; Bradshaw and Zhang, 2006; Zhang et al., 2007). Although no homologue of the transcription factor gene *aflR* was identified in *D. septosporum*, conserved *aflR* binding sites were identified upstream of several predicted DOTH biosynthesis genes (Zhang et al., 2007).

The DOTH biosynthesis genes in *D. septosporum* are very similar to ST/AF biosynthesis cluster genes in *Aspergillus* species. The gene organization in the DOTH cluster suggested that DOTH cluster could be ancestral to the AF cluster (Figure 1.4). The basal AF biosynthesis cluster proposed by Cary and Ehrlich (2006) contains genes that are required for stabilizing the polyketide as an anthraquinone. Further biosynthetic genes were thought to be recruited to the basal cluster, thereby allowing initially formed anthraquinones to become increasingly toxic compounds (Cary and Ehrlich, 2006b).

Figure 1.3 Comparison of dothistromin gene cluster withST/AF gene clusters

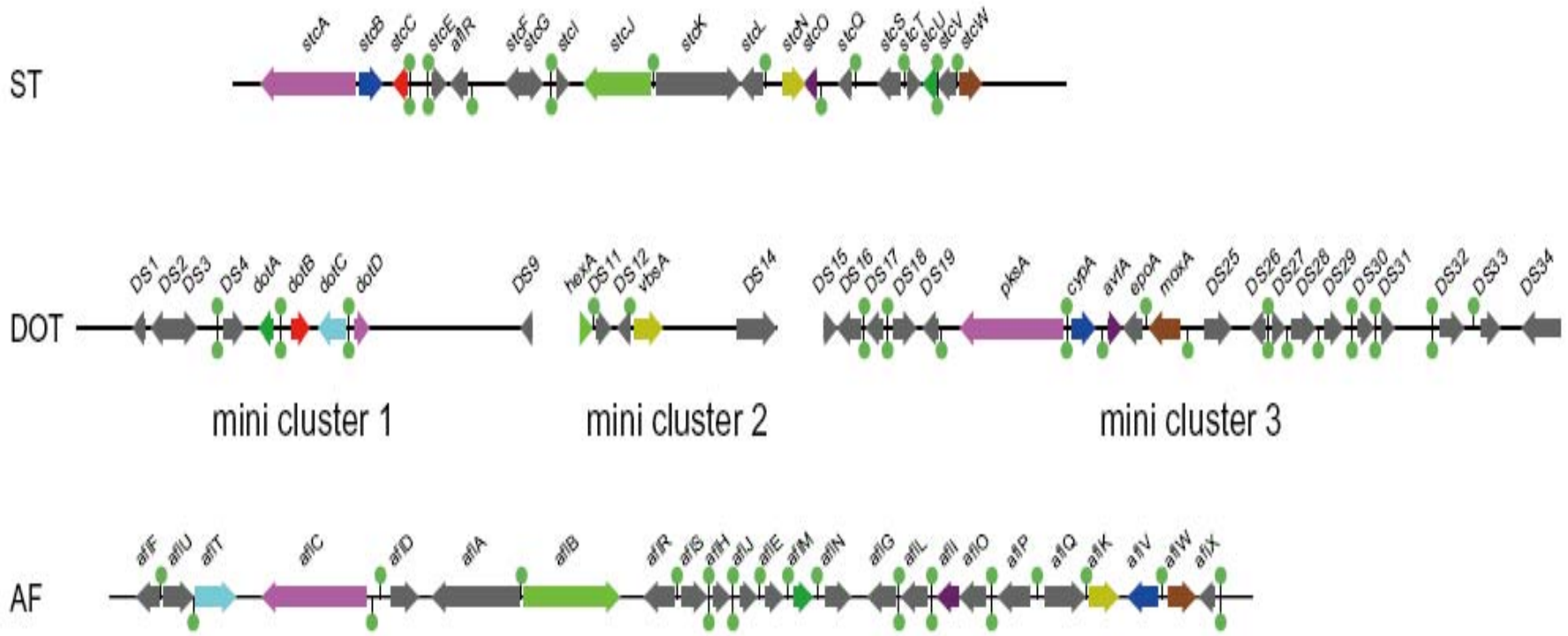


Figure 1.3 Comparison of DOTH gene cluster with ST/AF gene clusters. Large arrows represent direction of transcription of each gene in the cluster. Homologous genes were indicated with same coloured arrows. The green dot indicates the positions of putative *aflR* binding sites (Zhang et al., 2007).

Figure 1.4 Model for steps in the evolution of the ST/AF clusters

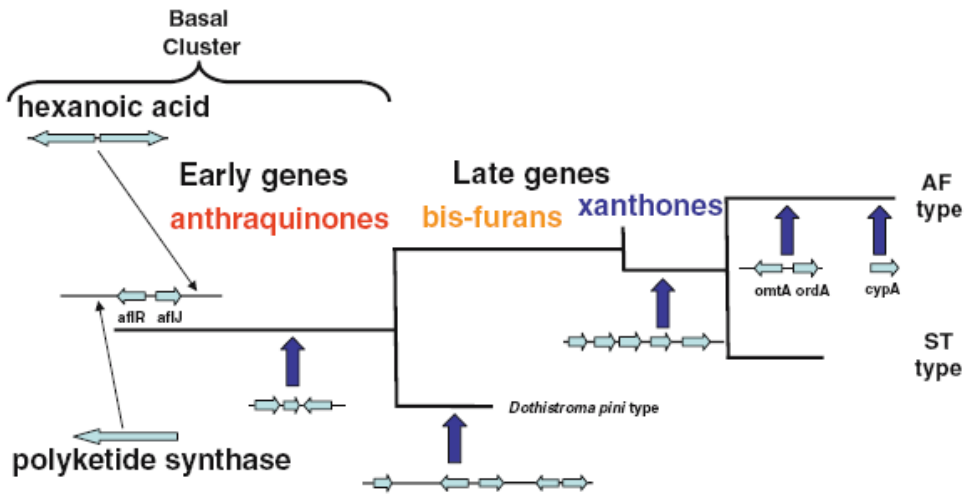


Figure1.4 Model for steps in evolution of ST/AF gene cluster (Cary and Ehrlich, 2006). See text on P9.

1.3 Double-stranded-RNA in fungi

In the course of this project, dsRNA was identified in strain 18667 of *P. arachidicola*. Double-stranded RNAs (dsRNA) are commonly found in fungi, and some are associated with virus-like particles, defined as mycoviruses (Michelmore and Hulbert, 1987). Mycoviruses generally contain protein coat and dsRNA genome (van Diepeningen et al., 2006). Many of these mycoviruses belong to the virus families *Narnaviridae*, *Hypoviridae*, *Partitiviridae* and *Totiviridae*. Transmission of dsRNA into fungi mainly occurs during compatible hyphal anastomoses (Buck, 1986). The mycoviruses are usually located in the cytoplasm of fungi, but sometimes mycovirus dsRNA is associated with the mitochondria (van Diepeningen et al., 2006). Elevated temperature, hyphal tip transfer, UV irradiation, single conidium subculture, cycloheximide treatments were used to eliminate dsRNA infection (Elias and Cotty, 1996; Hunst et al., 1986; Kousik et al., 1994; Schmidt et al., 1983).

Mycovirus infections are widespread in fungi. For example, in *Aspergillus* species, both sexual and asexual , dsRNA have been detected with infection rates up to 13% (van Diepeningen et al., 2006). Two dsRNA molecules, which are 5 kb and 3 kb, were also detected in *Epichloë festucae*. Hybridization experiments with a probe

complementary to the 5 kb dsRNA showed that there was no nucleotide sequence homologue between the 5 kb and 3 kb dsRNA elements (Zabagogeazcoa et al., 1998). The 5 kb dsRNA from *E. festucae* was characterized to belong to the *Totiviridae* family (Romo et al., 2007).

Although infections of fungi by mycoviruses are latent and usually do not affect their hosts, some mycovirus infections do cause morphological and physiological changes. Some of these changes are beneficial for the host, while others more closely resemble diseases. For example, when *Agaricus bisporus* is affected by mycoviruses, the infected isolates are weak and patchy and few fruit bodies are produced compared to uninfected isolates (Shen et al., 1993). The mycovirus infection can also alter enzymatic activities, for example dsRNA infected chestnut blight pathogen *Cryphonectria parasitica* had reduced accumulation of extracellular laccase A (Rigling and Van Alfen, 1993). Mycovirus infection of plant pathogenic fungi can alter the virulence of the pathogen. There are many reports of increased virulence (hypervirulence) or reduced virulence (hypovirulence) in isolates infected with dsRNA mycoviruses (Ahn and Lee, 2001). In *C. parasitica* 7 related hypovirulent strains were used to convert 49 virulent vegetatively compatibility strains to hypovirulence by transmission of dsRNAs using hyphal anastomosis. The vegetative compatibility was important for rapid and successful conversion and stability of the dsRNA in conversions. When more than one sized dsRNA was present, the vegetative compatibility also determined which sized dsRNA was transmitted (Anagnostakis and Day, 1979). In *Nectria radicicola*, a 6 kb dsRNA was responsible for up regulation of fungal virulence. When the 6 kb dsRNA was reintroduced into a dsRNA-cured strain by hyphal anastomosis, it restored the virulence related phenotypes of the cured strain. Sequence analysis of a cDNA clone derived from the 6 kb dsRNA revealed a RNA-dependent RNA polymerase gene. Studies showed that this 6 kb dsRNA regulates virulence through signal transduction pathways. The 6 kb dsRNA containing strain had six fold increases of cAMP-dependent protein kinase activity and decreased calcium-dependent protein kinase activity compared to a 6 kb dsRNA lacking strain

(Ahn and Lee, 2001).

1.4 Aim and objectives

1.4.1 Aim

To characterise the dothistromin gene cluster in the peanut pathogen *P. arachidicola.*

1.4.2 Objectives

-Identify and characterise dothistromin genes in *P. arachidicola*

-Determine arrangement of dothistromin genes in *P. arachidicola*

-Compare gene sequences and synteny with *D. septosporum*

-Compare the DOTH gene cluster in *P. arachidicola* and *D. septosporum* with ST/AF gene clusters in *Aspergillus* species

Chapter Two - Materials and Methods

2.1 Growth and maintenance of cultures

2.1.1 Growth and maintenance of *E. coli* cultures

The *E. coli* cells were grown on LB + ampicillin agar plates (A1.1) or in LB + ampicillin broth (A1.2) at 37°C overnight with shaking. For storage, the LB agar plates were sealed with parafilm and stored at 4°C, and freshly sub-cultured every three months. For long-term storage, *E. coli* cells grown in LB broth were mixed with 50% (v/v) sterile glycerol to reach a final concentration of 15% (v/v) glycerol and stored at -80°C.

2.1.2 Growth, maintenance and harvest of *P. arachidicola* culture and mycelia

To grow *P. arachidicola* on PD agar plates (A1.4), a small piece (4 mm x 4 mm) of *P. arachidicola* mycelium was cut from the *P. arachidicola* colony and placed on the agar plate. The plates were sealed with parafilm and grown at 22°C for two weeks until colony diameters reached approximately 2 cm. For storage, the PD agar plates were sealed with parafilm and stored at 4°C. Every six months, fresh sub-cultures were made to maintain the *P. arachidicola* strain. For long term storage, the mycelia were cut from an actively growing colony and stored in 20% (v/v) sterile glycerol at -80°C.

To grow *P. arachidicola* in PD broth (A1.3), a small piece of *P. arachidicola* mycelium was cut from the *P. arachidicola* colony and ground using a sterile plastic pestle in a microcentrifuge tube. The grinding worked better if mycelium was dry, then 100 µl of MilliQ water were added to ground mycelia. 25 µl of mycelia mixture were added to 20 ml PD broth in a 125 ml flask. All flasks were incubated at 22°C for two weeks with shaking (150 rpm).

The *P. arachidicola* mycelia were harvested by filtration through a funnel covered

with sterile nappy liner. The liquid was collected in a flask and discarded according to the DOTH toxin discard procedure. The mycelia were collected in a 50 ml falcon tube, snap frozen in liquid nitrogen for 10 minutes before freeze drying overnight. The dried mycelia were stored at 4°C.

2.2 DNA extraction, quantification and analysis

2.2.1 Genomic DNA isolation from *P. arachidicola* by CTAB method

The CTAB (hexadecyltrimethylammonium bromide) DNA extraction method used in this research was developed by Doyle and Doyle (Doyle and Doyle, 1987). This method allows a high yield of high molecular weight genomic DNA from fresh freeze dried fungal samples. *P. arachidicola* is a very slow growing fungus and further experimentation requires high molecular weight genomic DNA, therefore this method was chosen.

The freeze-dried mycelia were ground using a sterile mortar and pestle with liquid nitrogen. 500 mg of ground sample was transferred into a sterile microcentrifuge tube. 600 µl CTAB was added to the sample and mixed thoroughly by inversion. The sample was incubated at 37°C for 10 minutes and then at 65°C for 45 minutes with occasional inversion. Removed from the water bath, the sample was cooled to room temperature and 600 µl chloroform was added, mixed by inversion then stored on ice for 2 minutes. After centrifugation at 15110 g (13000 rpm Biofuge) for 2 minutes, the upper aqueous phase was transferred to a new tube and 600 µl isopropanol was added, mixed and stored on ice for 5 minutes. The mixture was centrifuged at 15110 g (13000 rpm Biofuge) for 30 seconds to settle the DNA, the solution was discarded, 600 µl 80% ethanol was added to the precipitated DNA and centrifuged at 15110 g (13000 rpm Biofuge) for 1 minute. The ethanol washing step was repeated at least twice. Then the ethanol was decanted off completely, and the sample was left to air dry. Finally the DNA was resuspended in 50 µl TE buffer.

2.2.2 Purification of *P. arachidicola* genomic DNA

Two purification methods, phenol purification and column purification were used. The column purification method, involves using a commercial kit which was convenient, did not involve hazardous chemicals and can yield high quality purified genomic DNA. But the purified genomic DNA was slightly fragmented (around 50 kb), which is not suitable for 8-cutter Southern blot. So the phenol purification method, which can yield high molecular weight genomic DNA was used when larger fragments of gDNA were required.

2.2.2.1 Column purification

The extracted genomic DNA was purified using QIAquick Gel extraction kit (Qiagen) according to the manufacturer's instruction.

2.2.2.2 Phenol purification

To unpurified genomic DNA sample, an equal volume of phenol and equal volume of chloroform was added, mixed thoroughly by vortex then centrifuged at 15110 g (13000 rpm Biofuge) for 3 minutes. The upper aqueous phase was transferred to a fresh microcentrifuge tube and an equal volume of chloroform was added, mixed thoroughly by vortex and centrifuged at 15110 g (13000 rpm Biofuge) for 1 minute. The DNA was then precipitated by transferring the upper aqueous phase to a new microcentrifuge tube and adding 0.1 volume of 3 M Na acetate and 0.6 volume of isopropanol to the sample. The tube was inverted several times to mix the sample and centrifuged at 15110 g (13000 rpm Biofuge) for 5 minutes to settle the DNA. The supernatant was discarded and an equal volume of 70% ethanol was added to the sample, then centrifuged at 15110 g (13000 rpm Biofuge) for 5 minutes. The ethanol was completely decanted, the sample left to air dry and the DNA resuspended in water. The purified DNA was stored at 4 °C.

2.2.3 Isolation of plasmid DNA

Plasmid DNA was isolated using an Ultra-fast rapid alkaline extraction method

(Cormack and Somssich, 1997). 0.2 ml of overnight *E. coli* culture (section 2.1.1) was transferred into a sterile microcentrifuge tube. 0.2 ml of lysis solution (1% SDS, 0.2 N NaOH) was added to the culture and mixed by inverting the tubes several times. Immediately, 0.2 ml of solution III (3 M potassium acetate, pH 5.50) was added to the tube and mixed gently by inverting the tube. After centrifugation at 15110 g (13000 rpm Biofuge) for 1 minute, the supernatant was transferred into a fresh tube, 0.5 ml isopropanol was added and mixed by inversion. After centrifugation as before, the supernatant was discarded and the pellet left to air dry. 50 µl TE buffer was added to resuspend the pellet. This method can reduce the time of isolating plasmid DNA for further restriction analysis, but the quality of plasmid is not so good for sequence analysis. So to increase the yield and quality of plasmid DNA isolated from *E. coli* cells, QIAprep Spin Miniprep kit (Qiagen) was used according to manufacturer's instructions.

2.2.4 Fluorometric assay to determine DNA concentration

DNA was quantified using a "Hoefer DyNA Quant 200 fluorometer" (Amersham Biosciences). 2 ml of working buffer (A2.7) was used to zero the fluorometer, then 2 µl of standard DNA (100 ng/ml) (A2.5) was added to 2 ml of working buffer in the cuvette for calibration. To determine the sample DNA concentration, 2 µl of sample DNA was added to 2 ml of working buffer, the concentration recorded by the fluorometer as ng/µl.

2.2.5 Agarose gel electrophoresis

An 0.4%-1% (w/v) agarose gel was prepared depending on the DNA sample molecular weight. 1x TBE buffer (A2.3) was added to the gel apparatus, the DNA sample was mixed with gel loading dye (A2.9) and loaded along with an appropriate DNA ladder on the gel. The electrophoresis was carried out at 40-80 volts until the dye moved to 2 cm from the end of the gel. Then the gel was stained with 1 mg/mL ethidium bromide staining buffer (A2.8) for 15 minutes followed by rinsing in MilliQ water for 2 minutes. Finally the gel was visualized and photographed using the Gel

Documentation system (BioRad) and Quantity One 4.4.0 basic software.

2.2.6 Agarose gel purification of size fractionated DNA

After gel electrophoresis, the selected DNA fragment was cut out from the gel using a sterile scalpel under long wave UV light and transferred into a sterile microcentrifuge tube. The DNA was then recovered using a QIAquick gel extraction kit (Qiagen) according to manufacturer's instructions. The concentration and size of recovered DNA sample was measured using Fluorometric assay (section 2.2.4) and gel electrophoresis (section 2.2.5)

2.2.7 Restriction endonuclease digestion of DNA

The selected restriction endonuclease and appropriate digestion buffer was added to a DNA sample in a microcentrifuge tube according to manufacturer's instruction. The volume of the digestion mixture (20 µl to 50 µl) depended on the amount of DNA sample used, the temperature of incubation depended on the selected restriction endonuclease. The incubation time was varied from 1 hour to overnight. The digestion of the DNA sample was checked using gel electrophoresis (section 2.2.5). To increase the performance of the restriction enzyme (for 8-cutter Southern blot), bovine serum albumin (BSA) was added to the restriction digestion mixture in a final concentration of 0.1 mg/ml (Williams et al., 1996).

Table 2.1 Example of *Eco*RI digestion of gDNA for Southern blot and fractionated genomic library. In a microcentrifuge tube, the following were added:

gDNA (150 ng/µl)	13.3µl
EcoRI (10 unit/µl)	2µl
Buffer H (10x)	5µl
MQ water	29.7µl

Mix well and incubated in 37 °C overnight

2.3 Polymerase chain reaction (PCR)

All PCR reactions were set up on ice, in a total volume of 25 µl and were carried out in an Eppendorf Gradient Mastercycler® (Eppendorf, Hamburg, Germany). The sequences of primers used in this project are in Appendix III.

2.3.1 Standard PCR reaction

The 25 µl PCR mixture contained 1×PCR buffer, 0.05 mM dNTP; 1.5 mM $MgSO_4$, 0.4 µM of each primer and 0.02 unit *Taq* DNA polymerase enzyme (Invitrogen). Finally DNA template was added (0.5 ng of plasmid DNA or 5 ng of gDNA). The PCR conditions were: an initialisation step of 94°C for 2 minutes; then 30 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds and elongation at 72°C for 1 min/kb depending on product size; followed by 72°C for 5 minutes. When required the PCR products were purified using QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions and stored at 4°C**.**

2.3.2 Inverse PCR

Inverse PCR was used to determine the flanking sequences of the target DNA in this experiment, when one internal sequence of the target DNA was known. 500 µg of the target DNA was digested using a selected restriction endonuclease (section 2.2.7) then purified using QIAquick PCR purification kit (Qiagen) according to manufacturer's instruction. 5 ng of the purified DNA fragment was self-ligated using DNA T4 ligase (section 2.4.1.1). 1 µl of ligation mixture was used as DNA template and two outward direction primers complementary to the terminal sequences of the target DNA were used for inverse PCR. The PCR conditions were based on the standard PCR reaction (section 2.3.1). The PCR products were checked by gel electrophoresis (section 2.2.5). Selected PCR products were purified and ligated into the selected vector for further experiment.

2.3.3 Touchdown PCR

Touchdown PCR was used to minimise amplification of nonspecific sequences by primers. The first step in touchdown PCR has a high annealing temperature, so it is likely that only the specific region will be amplified. The annealing temperature is decreased in increments for subsequent touchdown PCR steps. The specific regions amplified in the earlier step compete with nonspecific regions, where the primer can bind at lower temperature. The PCR mixture was same as the standard PCR reaction (section 2.3.1).

Table 2.2 Touchdown PCR conditions in this experiment:

Step	cycles	temperature	Time
initialisation	1	94°C	2 minutes
denaturation	3	94°C	30 seconds
annealing		68°C	30 seconds
elongation		72°C	1kb/min depended on product size
denaturation	3	94°C	30 seconds
annealing		64°C	30 seconds
elongation		72°C	1kb/min depended on product size
denaturation	3	94°C	30 seconds
annealing		60°C	30 seconds
elongation		72°C	1kb/min depended on product size
denaturation	3	94°C	30 seconds
annealing		56°C	30 seconds
elongation		72°C	1kb/min depended on product size
denaturation	3	94°C	30 seconds
annealing		52°C	30 seconds
elongation		72°C	1kb/min depended on product size
final step	1	72°C	5minutes

2.3.4 Nested PCR

Nested PCR was used to increase the specificity of amplification. Two sets of primers were used in two runs of PCR. The first run of PCR using the first set of primers had the same PCR mixture and PCR conditions as the standard PCR reaction (section 2.3.1). The second run of PCR was using 1/50 diluted first run PCR product as DNA template and a second set of primers, which only amplifies a target region within the first run product. The PCR conditions were the same as the standard PCR reaction (section 2.3.1).

2.3.5 *E. coli* colony PCR

E. coli colony PCR was used to screen the transformants from a library. A single *E. coli* colony or a group of selected *E. coli* colonies can be used as DNA template for *E. coli* colony PCR. For single colony, a sterile P2 pipette tip was used to pick up cells from one *E. coli* colony by gently touching the surface of the colony on the plate. The cells were transferred into a PCR tube containing 24 µl of PCR mixture same as the standard PCR reaction (section 2.3.1). The initial PCR step was 3 minutes at 96°C to release the cell components, the rest of the steps were same as the standard PCR condition (section 2.3.1).

For pooled screening of multiple colonies, cells from each *E. coli* colony were picked up the same way as single colony from plate and added to 100 µl MQ water in an microcentrifuge tube. A fresh pipette tip was used for each colony. The tube was incubated in boiling water for 30 seconds to release the cell components, centrifuged at 13200 rpm for 1 minute. 1 µl of the supernatant was used as DNA template for PCR. The PCR mixture and conditions were the same as the standard PCR reaction (section 2.3.1).

2.4 Ligation and transformation

2.4.1 Ligation

2.4.1.1 Ligation reaction

All the ligation reactions were carried out using DNA T4 ligase, which is suitable to ligate cohesive ends or blunt ends of DNA. The ligation mixture contains 1x ligation buffer and 1 µl of 3U/µl T4 DNA ligase (Roche). The ATP in ligation buffer may degrade due to repeated freeze, thaw cycles. So ATP may be added to the ligation buffer as required at a final concentration of 10 mM . When ligating a DNA fragment into a vector, a 1:1 insert : vector molar ratio was used unless otherwise stated. The amount of insert DNA and vector used in ligation mixture can be calculated using this formula:

$$\frac{\text{ng of vector x Kb size of insert}}{\text{Kb size of vector}} \quad \text{x} \quad \text{Molar ratio of} \quad \frac{\text{Insert}}{\text{Vector}} \quad = \text{ng of insert}$$

2.4.1.2 Preparation of the vector

The vector plasmid DNA was digested with the selected restriction endonuclease (section 2.2.7). The digested vector plasmid was incubated at 65 °C for 15 minutes to inactivate the restriction endonuclease. Then vector plasmid was mixed with shrimp alkaline phosphatase. For 1 pmol DNA, both 5’ protruding and recessive ends were incubated with 1 unit of phosphatase alkaline shrimp at 37°C for 1 hour. Finally the shrimp alkaline phosphatase was inactivated by incubating at 65°C for 15 minutes. The dephosphorylation of vector was tested by using DNA T_4 ligase.

2.4.1.3 N-butanol purification of ligation mixture

To exclude salt from the mixture after ligation, 1 ml of N-butanol was added to 20 µl ligation mixture, mixed by vortexing, centrifuged at 15110 g (13000 rpm Biofuge) for 5 minutes. The supernatant was discarded, the pellet left to air dry and resuspended in 10 µl of MilliQ water.

2.4.2 Transformation

2.4.2.1 Preparation of *E. coli* competent cells for electroporation

A single purified colony of Top 10 *E. coli* cells (F^- *mcr*A D(*mrr-hsd* RMS-*mcr*-BC) 80 (*lacZ*) DM15 D*lac*X74 *rec*A1 *deo*R *ara*D139 D(*ara-leu*)7697) (Invitrogen) was obtained by streaking using a sterile loop on LB plate. The plate was incubated at 37°C overnight. One single Top 10 *E. coli* colony was inoculated into 5 ml LB culture using a sterile loop. The LB culture was incubated at 37°C overnight with shaking. 50 µl of the above overnight LB culture was inoculated into 500 ml LB culture in a flask and incubated at 37°C with shaking until the OD reading of the LB broth reach mid-log phase, A600 0.5-1.0. The flask was chilled on ice for 20 minutes, and then cells were harvested by centrifugation at 5000 rpm in GSA rotor for 10 minutes at 4°C. The supernatant was removed and cells were resuspended in 1 L ice cold sterile MilliQ water. The cells were harvest by centrifugation as before. The supernatant was discarded and the cells were resuspended in 500 ml ice cold sterile MilliQ water. The cells were harvested by centrifuge as before, then the supernatant was pulled off and the cells were resuspended in 20 ml ice cold sterile10% glycerol. Finally the cells were harvested by centrifuge as before, the supernatant discarded and cells resuspended in 4 ml ice cold sterile 10% glycerol. Aliquots of 100 µl cells were transferred into fresh sterile microcentrifuge tubes, freeze dried in liquid nitrogen and stored at -80°C.

2.4.2.2 Transformation reaction of *E. coli* cells by electroporation

1 µl of 1 ng/µl plasmid (positive control) or 2-5 µl of N-butanol treated ligation mixture was mixed with 50 µl of *E. coli* competent cells in an ice cold cuvette. The cuvette was placed in the Gene Pulser (Bio-Rad), which was set at 25 µF, 2.5 kV and 200 Ω. 1 ml LB culture was immediately added to the cuvette after electroporation. The LB culture was transferred into a sterile microcentrifuge tube and grown at 37°C for 2 hours. 50 µl of transformation culture was grown on LB + ampicillin plate at 37°C overnight. If the vector has blue and white selection system, 40 µl 2% x-gal and 40 µl 20% IPTG was added to LB + ampicillin plate.

2.5 Hybridization of DIG-labelled probes to blots

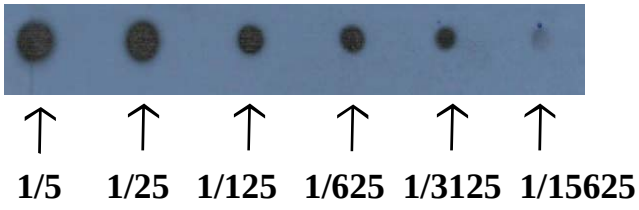
2.5.1 DIG-labelling of the probes (PCR based)

A PCR based method was used to prepare DIG-labelled probes. The 25 µl PCR mixture contained 1×PCR buffer; 1.5 mM $MgSO_4$, 0.4 µM of each primer and 0.02 unit *Taq* DNA polymerase enzyme (Invitrogen). The dNTPs used for DIG-labelling PCR contains 0.05 mM of each dATP, dCTP, dGTP and 0.033 mM dTTP and 0.016 mM Digoxigenin-11-2'-deoxy-uridine-5'-triphosphate, alkali-labile (DIG-11-dUTP, Roche). 1 µl of the above dNTP mix was added to 25 µl PCR reaction mixture. The PCR conditions were the same as the standard PCR programme (section 2.3.2).

2.5.2 Probe concentration determination

To determine the labelled probe concentration, serial dilutions of probe were made (1/5, 1/25, 1/125, 1/625, 1/3125 and 1/15625). 2 µl of each dilution was spotted on a Hybond-N^+ membrane (Amersham). The DNA was fixed to membrane using a UV crosslinker (Ultralum). The hybridization and detection procedure was as described in section 2.5.5 and 2.5.6. Figure 2.1 shows an example of probe concentration detection. The 1/3125 dilution spot can be clearly observed, because the incorporation of probes to the DNA template was not 100%, the actual working probe concentration used was one grade higher than detected probe concentration, which is 1/625 dilution. In the experiment, if the hybridization buffer used was 30 ml, then 48 µl of undiluted probe was added to the hybridization buffer. (30 ml /625= 48 µl)

Figure 2.1 An example of probe concentration detection



2.5.3 Southern blot

2 μg of each restriction endonuclease digested genomic DNA were separated on a 0.8% agarose gel at 40 V overnight at 4°C (section 2.2.5 and 2.2.7). The gel was stained in ethidium bromide for 20 minutes, visualized and photographed using Gel Documentation system (BioRad) and Quantity One 4.4.0 basic software with a ruler alongside. The gel was treated with 0.25 M HCL for 15 minutes for depurination when it contained DNA fragments bigger than 5 kb. Then the gel was incubated in denaturation buffer (A2.11) for 45 minutes followed by neutralization buffer (A2.11) for 1 hour. Finally the gel was incubated in 20 x SSC solution (A2.11) for 20 minutes. A blotting apparatus was constructed as outlined in the DIG application manual for filter hybridization (Roche). Then the gel was transferred to the blotting apparatus to allow transfer of digested genomic DNA to the Hybond-N^+ membrane (Amersham) overnight. The membrane was washed in 2 x SSC for 1 minute and fixed under the UV crosslinker (Ultralum). The crosslinked membrane was allowed to air dry and was ready for hybridization.

2.5.4 Hybridisation of DIG labelled probe

The membranes were prehybridised in DIG Easy Hyb solution (Roche, approximately 10 ml/100 cm^2 of membrane) at 42°C for 2 hours in a hybridization tube with rotation. The probe was incubated for 10 minutes in a boiling water bath and immediately chilled on ice to denature it. The denatured probe was added to the hybridisation buffer and incubated at 42°C overnight with rotation. Because the hybridization has 50% formamide, the hybridization buffer containing the denatured probe can be re-used following denaturation at 68°C water bath for 10 minutes. After overnight incubation, the membrane was washed 2 x 5 minutes in washing solution I (A2.11) at room temperature, then 2 x 15 minutes in washing solution II (A2.11) at 68°C.

2.5.5 Immunological detection

To detect the hybridisation bands, the membrane was rinsed in buffer I (A2.11) (Roche) then incubated in buffer II (buffer I + 1% blocking reagent) (Roche) for 30

minutes. The anti-DIG-Ap conjugate, 1:10000 dilution (Roche) was added to buffer II and incubated for another 30 minutes. Then the membrane was washed 2×15 minutes in buffer I and equilibrated in buffer III (A2.11) (Roche) for 5 minutes. The membrane was placed in an A4 copysafe pocket and CSPD lumigen (Roche) was added to membrane directly. The membrane was incubated at 37°C for 10 minutes to activate the enzyme, then it was exposed to x-ray film (Fuji film) for 20 minutes to 1 hour. Finally the x-ray film was developed using a 100 plus Automatic x-ray Processor (All Pro imaging).

2.5.6 Stripping blot

To re-hybridise the membrane with another probe, the membrane was washed in MilliQ water for 1 minute and incubated in stripping buffer (A2.11) at 37°C for 2 x 20 minutes.

2.5.7 Screen the library for positive clones

2.5.7.1 Colony lifts

50 µl transformation culture from 1 ml of a ligation for a sub-genomic library was plated on a LB + ampicillin plate. Duplicates were made of the primary plates and stored at 4°C. A Hybond-N^+ membrane (Amersham) was placed on one primary plate for 1 to 2 minutes to lift the colonies, the membrane was marked to provide orientation with the same way as the primary plate. The membranes contain colonies that were grown face up on LB + ampicillin plates for 6 hours and then grown on LB + ampicillin + 250 µg/ml chloramphenicol plates at 37°C overnight. The high concentration of chloramphenicol will stop the cells growth but allow the plasmids to increase in copy number.

After incubation, the membranes were placed face up on filter paper containing denaturation buffer (A2.11) for 15 minutes to lyse the cell membranes. Then incubated with filter paper containing neutralization buffer (A2.11) for 15 minutes and 2 x SSC solution for 10 minutes. The membrane was allowed to dry and fixed under a

UV crosslinker. The membrane was incubated in 10 ml 2 x SSC with 2 mg/ml proteinase K solution at 37°C with shaking for 2-4 hours. Finally the membrane was washed in 2 x SSC solution and allowed to air dry and was ready for hybridization.

2.5.7.2 Screen the library for positive clones

The membranes were hybridized with one DIG-labelled probe as described in section 2.5.4. The hybridization spots were detected as described in section 2.5.5. To re-hybridise the membrane with another probe, the membrane was stripped as described in section 2.5.6.

2.6 DNA sequencing

2.6.1 Sequencing

All sequencing reactions were carried out in the Allan Wilson Centre Genome Analysis Centre, The Institute of Molecular Biosciences of Massey University. For plasmid sequencing, 300 ng of purified plasmid DNA and 3.2 pmol primer was mixed with sterile MilliQ water to a final volume of 15 µl. For PCR product sequencing, 2 ng / 100 bp PCR product and 3.2 pmol primer was mixed with sterile MilliQ water to a final volume of 15 µl. New primers for sequencing reactions were designed using Clone Manger SECentral software. The assembling of sequence for each clone was done using ContigExpress (VectorNTI software package). The sequences of primers are shown in Appendix III

2.6.2 Sequence analysis

Sequence analysis was done using the National Center for Biotechnology Information (NCBI) BLAST programs (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) and Vector NTI 7.0 software (Informax, Invitrogen, Frederck, MD). Gene sequence alignment was done using multiple sequence alignment editor Gene Doc software. The tandem repeats were detected using the program Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.basics.submit.html) and the inverted repeats were detected using the program Palindrome (http://sf01.bic.nus.edu.sg/EMBOSS/). The

six-frame translation was done using the program six-pack translation (http://bioinfo.nhri.org.tw/cgi-bin/emboss/sixpack?_pref_hide_optional=0). The GC content was calculated using the program GeeCee (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=geecee) and MacVector software.

Chapter three – Results

3.1 Inverse polymerase chain reaction (PCR)

3.1.1 Background

When one of the target region's internal sequences is known, inverse PCR can be used to identify flanking sequence. Fragments of DOTH biosynthesis genes in *P. arachidicola* had been identified by degenerate PCR (*Pa-cypA* 1.0kb and *Pa-vbsA* 1.4kb) and inverse PCR was carried out in attempts to identify the flanking sequences of these genes.

3.1.2 Results

P. arachidicola gDNA was extracted from isolate 18667 using a CTAB method (Section 2.2.1) and column purified (section 2.2.2.1). The gDNA was at least 35-40 kb in length and was digested with five enzymes (*Pst*I, *Bam*HI, *Sal*I, *Sca*I and *Eco*RI) that either cut or did not cut within the gene (Section 2.2.7).

The purified digested DNA was ligated (section 2.4.1) to form circular DNA and used as DNA templates for inverse PCR (section 2.3.2) using outward facing primers (Figure 3.1). The primers named szdp were designed based on DOTH gene sequences from *D. septosporum*. Because the *Pa-cypA* and *Pa-vbsA* containing fragments had 94% and 95.6% amino acid identity to the corresponding *Ds-cypA* and *Ds-vbsA* regions, it was anticipated that matching *P. arachidicola* regions would be amplified by szdp primers. The PCR products were checked on a gel (Figure 3.2), from which four bands amplified using *Pa-cypA* primers, and three bands using *Pa-vbsA* primers, were selected. Those PCR products were purified and ligated into PGEMT-Easy vectors, and transformed into *E. coli* (section 2.4.2).

Figure 3.1 Enzyme cutting sites on *Pa-cypA* and *Pa-vbsA* fragments and primers used for inverse PCR.

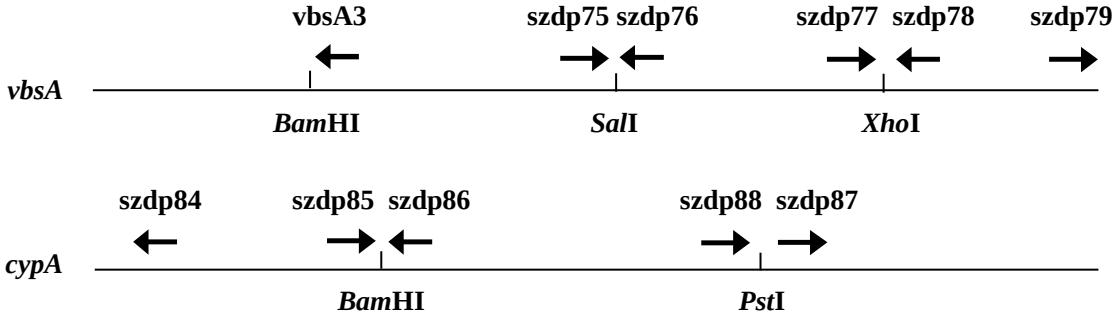
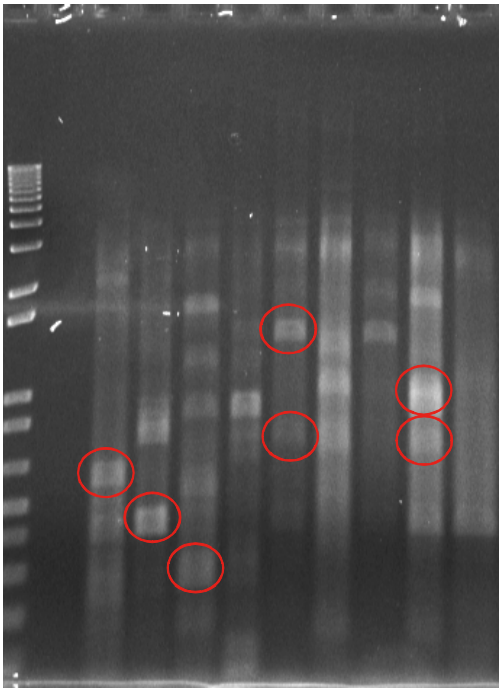


Figure 3.2 Inverse PCR reuslts



Lane	Gene	Enzyme	Primers	selected bands
2	*Pa -vbsA*	*BamHI*	vbsA3 and szdp79	
3	*Pa-cypA*	*Bam* HI	szdp84 and szdp85	700 bp
4	*Pa-cypA*	*Bam* HI	szdp86 and szdp87	500 bp
5	*Pa -vbsA*	*Sal* I	szdp76 and szdp79	400 bp
6	*Pa -vbsA*	*Sal* I	vbsA3 and szdp79	
7	*Pa -vbsA*	*Eco* RI	vbsA3 and szdp79	900 bp, 1.65 kb
8	*Pa-cypA*	*Eco* RI	szdp84 and szdp87	
9	*Pa -vbsA*	*Sca* I	vbsA3 and szdp79	
10	*Pa-cypA*	*Sca* I	szdp84 and szdp87	800 bp, 1,1 kb
11	*Pa-cypA*	*Sca* I	szdp84 and szdp88	

Figure 3.2 Lane 1 is the 1 kb^{+} ladder. Lane 2 to 11 were inverse PCR products amplified using outward primers. The red circles indicate the selected bands which were cloned into the PGEMT-Easy vector.

After blue/white selection of transformants, some white positive colonies were selected and screened by PCR using the same primers used for inverse PCR (Figure 3.3). Five colonies showed the expected band size compared to inverse PCR results (circled in Figure 3.2). They were two *Pa-cypA* (*Bam*HI) positive colonies, which contain a 500 bp insert, a *Pa-cypA* (*Sal*I) positive colony with a 1.1 kb insert, a *Pa-vbsA* (*Eco*RI) positive colony with a 1.65 kb insert and a *Pa-vbsA* (*Sca*I) positive

colony with a 1.1 kb insert. The plasmids of those colonies were extracted and checked by enzyme digestion. All of the plasmids gave a 3.5 kb vector band and the expected size insert bands. The plasmids were sequenced and the sequences were analysed as described in section 2.6. However none of the plasmids had the expected *Pa-cypA* or *Pa-vbsA* sequences.

Figure 3.3 PCR screens of selected transformants

A **B**

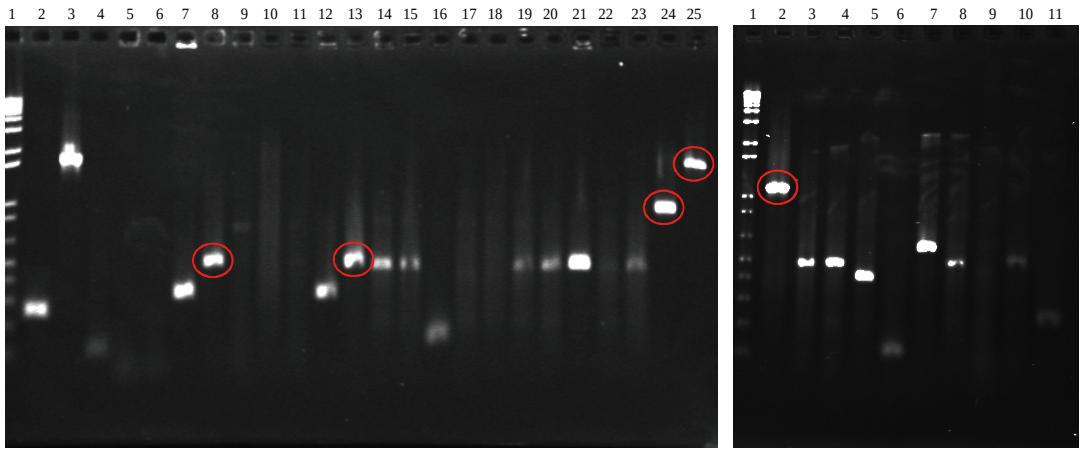


Figure 3.3 (A) Lane 1 is the 1 kb$^+$ ladder.

Lane	Gene	Enzyme	Primers
2-3	*Pa-cypA*	*BamHI*	szdp84 and szdp85
4-13	*Pa-cypA*	*Bam*HI	szdp86 and szdp87
14-23	*Pa-cypA*	*Sca*I	szdp84 and szdp87
24	*Pa-vbsA*	*Sal*I	vbsA3 and szdp75
25	*Pa-vbsA*	*Eco*RI	vbsA3 and szdp79

(B) Lane 1 is the 1 kb$^+$ ladder.

Lane	Gene	Enzyme	Primers
2-11	*Pa-cypA*	*Sca*I	szdp84 and szdp87

3.1.3 Discussion

For inverse PCR, the DNA templates used were circular DNA formed by self-ligation of enzyme digested gDNA. The formation of circular DNA requires low concentration of digested gDNA, if the concentration of digested gDNA is too high, it tends to form concatemers rather than circles (Dugaiczyk et al., 1975). From the enzyme digestion

results (data not shown), the average fragment was about 4 kb, the digested DNA was used in the ligation reaction.

Primers designed based on the *D. septosporum* DOTH sequences were available. Because of the high amino acid identity between the *P. arachidicola Pa-cypA* and *Pa-vbsA* containing fragments and the corresponding regions in *D. septosporum*, the primers were used for inverse PCR to identify *Pa-cypA* and *Pa-vbsA* genes. DNA sequencing results suggested that identification of *Pa-cypA* and *Pa-vbsA* genes using inverse PCR was not successful. One reason could be that the degenerate primers used for inverse PCR were not *Pa-cypA* and *Pa-vbsA* specific primers. The sequence of the DOTH *Pa-vbsA* genes were identified in section 3.5 to 3.7. The sequence of the primers used in inverse PCR were compared to the *P. arachidicola* DOTH genes and some important differences were found. For example, primer vbsA3 had a mismatched nucleotide at the 3' end, so the inverse PCR may have amplified some other regions from the enzyme digested gDNA. Increasing the PCR annealing temperature should reduce the unspecific binding of primers to unwanted regions. Or *P. arachidicola* specific primers could be designed based on the *Pa-cypA* and *Pa-vbsA* containing fragments to carry out inverse PCR. The inverse PCR experiment was not continued because mean while a size fractionated *P. arachidicola* genomic DNA library was made (Results see section 3.3).

3.2 RNA that co-purified with genomic DNA

3.2.1 Background

A broad band of 4 kb that co-purified with *P. arachidicola* strain 18667 genomic DNA was analysed in this section. It was predicted that the band was composed of RNA because double-stranded RNAs (dsRNA) are commonly found in fungi, and some are associated with virus-like particles, defined as mycoviruses (Michelmore and Hulbert, 1987).

3.2.2 Results

The *P. arachidicola* gDNA 18667 strain was extracted using a CTAB method (Section 2.2.1), purified using a column (section 2.2.2.1) and fractionated through a 0.7% agarose gel. In additional to the gDNA, an additional 4 kb band appeared. The 18667 gDNA was treated with DNase and RNase. When treated with DNase, the upper genomic DNA band disappeared. When treated with RNase, the bottom RNA band and the unknown 4 kb band disappeared, showing that the 4 kb band is RNA as predicted (Figure 3.4).

Figure 3.4 The RNA bands

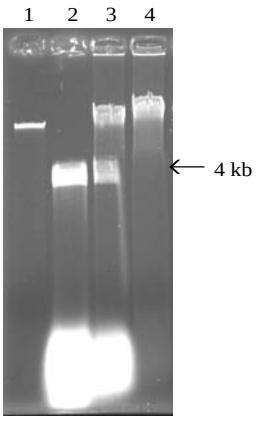


Figure 3.4 Lane 1 is the uncut λ DNA ladder. Lane 2 is 18667 genomic DNA treated with DNase. Lane 3 is the untreated control lane. Lane 4 is the genomic DNA treated with RNase.

In another experiment set up for Southern blotting, the *P. arachidicola* 18667 gDNA was digested with three enzymes (*Eco*RI, *Bam*HI and *Sal*I) without RNase treatment

(Section 2.2.7). The digested gDNA were electrophoresed at a low voltage (40 V), and the broad RNA band was clearly resolved into three RNA bands, ranging in size from 3 kb to 4 kb (Figure 3.5). Since RNA is not cut by restriction enzymes and the patterns on the gel for each digest were identical, the appearance of the three bands was due to increased resolution of the gel electrophoresis.

Figure 3.5 Restriction digestion of the RNA bands

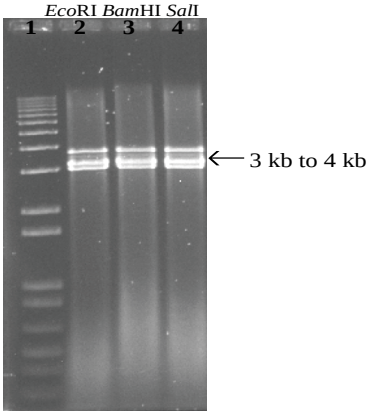


Figure 3.5 Lane 1 is the 1kb+ ladder. 18667 genomic DNA is shown digested with *Eco*RI (lane2), *Bam*HI (lane3) and *Sal*I (lane 4). There are three RNA bands ranging from 3 kb to 4 kb for each lane.

Genomic DNA was extracted from *P. arachidicola* strains K4246 and 32674 to determine if they have the RNA band. Figure 3.6 showed that RNA bands only appeared in the 18667 strain but were missing from the other two strains.

Figure 3.6 The RNA bands in other *P. arachidicola* strains

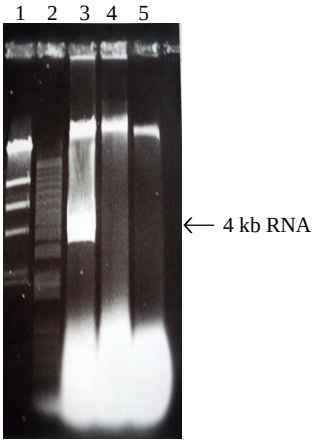
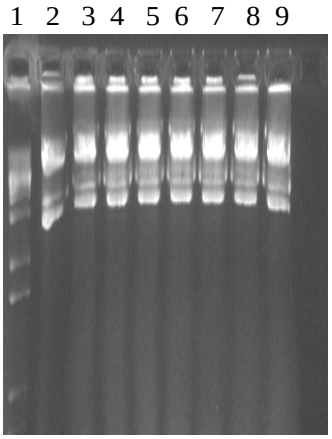


Figure 3.6 Lane 1 is the λ DNA/ *Hind* III ladder. Lane 2 is the 1kb+ ladder. Lane 3 is the 18667 strain which has the RNA bands. Lanes 4 and 5 are gDNA from *P. arachidicola* strains 32674 and K4246 respectively, the RNA bands are missing from those strains.

The stability of the RNA in strain 18667 was detected with different water-bath temperatures, incubation times and in the presence of QG buffer from the QIAquick gel extraction kit (Qiagen) (Figure 3.9). The QG buffer and 50°C water bath were used in this experiment because the Qiagen gel extraction kit could potentially to be used to extract the RNA bands from the gel for further analysis. According to kit instructions, the gel will be treated with QG buffer in 50°C water bath. The RNA bands were quite stable, none of the treatments caused degradation of disrupted the RNA bands. Further analysis of the RNA was not carried out in this project due to time limitations.

Figure 3.7 The stability of the RNA bands



Lane 1 is the 1kb+ ladder			
Lane	Incubation (min)	Temperature (°C)	Buffer
2	-	Room	Water
3	-	Room	QG
4	10	50	Water
5	10	42	Water
6	10	37	Water
7	20	50	Water
8	20	42	Water
9	20	37	Water

3.2.3 Discussion

The RNA bands found in *P. arachidicola* 18667 strain were unexpected. There were at least three different sizes or forms of the RNA present in the strain. Out of the three tested *P. arachidicola* strains, only the 18667 strain contains the RNA. To confirm whether the dsRNA found in *P. arachidicola* strain 18667 was due to mycovirus infection, cDNA could be made using reverse transcriptase and sequenced. RNA-dependent polymerase genes were identified in *Aspergillus* species mycovirus dsRNA (Genbank accession No.EU289896 and EU289897). The cDNA sequence results should be compared to these already known mycovirus dsRNA sequences and a RNA-dependent polymerase gene may be identified.

The RNA was stable, even after overnight restriction enzyme digestion. The stability of RNA would allow it to be extracted using gel extraction kit (Qiagen) which requires incubation with QG buffer at 50°C water bath for 10 minutes.

Studies of one isolate of *Aspergillus flavus* (NRRL5565) that cannot produce AF, showed that NRRL5565 contains a viral genome that has a size expected for mycovirus dsRNA. Exposing NRRL5565 to dsRNA virus synthesis inhibitor allowed it to produce AF, showing that inhibition of AF biosynthesis in *A. flavus* isolate NRRL5565 was due to dsRNA infection (Schmidt et al., 1986). The presence of dsRNA in different isolates of *A. flavus* was detected. Overall an average of 10.9% (10 of 92 isolates) was infected. But there was no correlation of AF production ability to the presence of dsRNA, strains that produce a high level of AF were just as likely to be infected by dsRNA as strains that produce no AF (Elias and Cotty, 1996). In three tested strains of *P. arachidicola*, only strain 18667 had dsRNA infection, but all three strains can produce DOTH. More isolates of *P. arachidicola* collected from culture and field should be tested for dsRNA infection and whether dsRNA infection can affect the production of DOTH in *P. arachidicola*.

3.3 Attempts to identify additional dothistromin genes in *P. arachidicola* using degenerate primers

3.3.1 Background

The DOTH biosynthesis genes in *D. septosporum* and *P. arachidicola* that have been identified so far have homologues to AF biosynthesis genes in *Aspergillus* species. Degenerate primers designed according to highly conserved regions of AF genes can be used to identify further DOTH biosynthesis genes. In *D. septosporum* the DOTH biosynthesis genes *avnA* (homologous to *aflG* in *Aspergillus*), *norA* (homologous to *aflE* in *Aspergillus*) and *verB* (homologous to *aflL* in *Aspergillus*) were successfully identified using this method. However a homologue of the *aflR* gene, which is a positive regulator in ST biosynthesis pathway in *Aspergillus* species, could not be identified in *D. septosporum* by degenerate PCR. The aim of this section of the project was to identify the *Pa-avnA*, *Pa-norA*, *Pa-verb* and possibly *Pa-aflR* genes in *P. arachidicola*, using the same degenerate primers as used for *D. septosporum*.

The degenerate primers were designed by Justine Baker and Zhilun Feng (Massey University). Alignment of gene product sequences from five *Aspergillus* species (*A. parasiticus, A. flavus, A. oryzae, A. nomius* and *A. nidulans*) was used to designed *avnA* degenerate primers. Alignment of gene product sequences from four *Aspergillus* species (*A. parasiticus*, *A. oryzae*, *A. nomius* and *A. nidulans*) was used to design *norA*, *verB* and *aflR* degenerate primers (the primer sequences used are shown in Appendix III).

3.3.2 Results

Touchdown PCR was carried out using degenerate primers (Table 3.1) and *P. arachidicola* 18667 gDNA as DNA template (Section 2.3.3). Only two PCR products had the expected sizes which were a 1.5 kb *Pa-aflR* fragment and a 1.1 kb *Pa-avnA* fragment (Figure 3.8). The touchdown PCRs which produced the right sized fragments were repeated, the fragments were gel-extracted and ligated into pGEMT-

Easy vector in a 1:1 vector : insert ratio (Section 2.2.6 and 2.4.1) and transformed into *E. coli* (Section 2.4.2).

Table 3.1 Degenerate primers used in touchdown PCR

Lane	Gene	Primer name	Expected size
2	***aflR***	**szdp127, szdp128**	**1.5 kb**
3	*avnA*	aflG F1, aflG R1	1.2 kb
4	*avnA*	aflG F1, aflG R2	1.3 kb
5	*avnA*	aflG F2, aflG R1	1.0 kb
6	***avnA***	**aflG F2, aflG R2**	**1.1 kb**
7	*verB*	vbrBF1, verBR3	435 bp
8	*verB*	vbrBF1, verBR3	915bp
9	*verB*	vbrBF1, verBR3	105 bp
10	*verB*	vbrBF1, verBR3	585 bp
11	*norA*	norAF1, norAR3	240bp
12	*norA*	norAF1, norAR4	700 bp
13	*norA*	norAF2, norAR3	240 bp
14	*norA*	norAF2, norAR4	700 bp

Figure 3.8 Touchdown PCR using degenerate primers

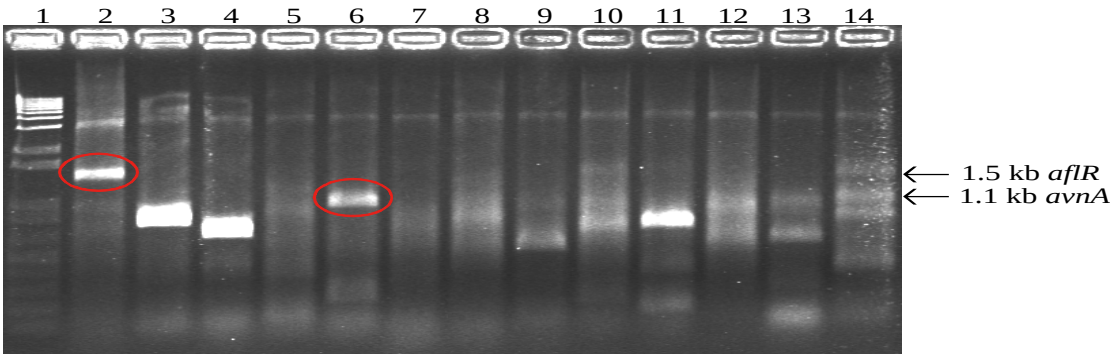


Figure 3.8 Lane 1 is the 1kb+ ladder. Lanes 2 to 14 are touchdown PCR products using degenerate primers. The primers used for each lane are presented in the following table. The red circles indicate the expected sized fragments: 1.5 kb *Pa-aflR* fragment in lane 2 and 1.1 kb *Pa-avnA* fragment in lane 6.

Transformants were checked by *E. coli* colony PCR (Section 2.3.5). From the gel, one white colony contained the 1.1 kb *Pa-avnA* insert. The white colony tested for *Pa-aflR* only showed a 162 bp fragment which was an empty vector. The *E. coli* colony PCR was repeated for another 10 *Pa-aflR* transformants, but no positive colony was found (Figure 3.9). The *Pa-aflR* ligation was repeated with a 1:2 vector :

insert ratio, then transformed into *E. coli* and 10 white colony were checked by PCR, but no positive colony was found (Data not shown).

Figure 3.9 PCR using universal primers to check the transformants

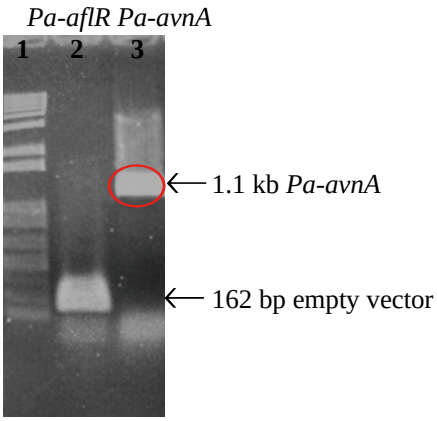


Figure 3.9 Lane 1 is the 1kb+ ladder. Lanes 2 and 3 are products from colony PCR with *E. coli* transformed with *Pa-aflR* and *Pa-avnA* ligations respectively. The red circle indicates the 1.1 kb *Pa-avnA* insert.

The putative *Pa-avnA* plasmid was extracted and sequenced (Section 2.2.3 and section 2.6). The sequence of the 1.1 kb fragment was analysed using BLASTN and BLASTX, the result indicated no similarity to *D. septosporum Ds-avnA* or any *Aspergillus* species *aflG*. The closest BLASTN hits for this 1.1 kb fragments were showed in table 3.2.

Table 3.2 BLASTN result of 1.1 kb fragment

Description	Product	Score	E value	Genebank accession No.
Phaeosphaeria nodorum SN15 mitochondrion, complete genome	tRNA-Leu	110	2e-20	EU053989.1
Mycosphaerella graminicola mitochondrion, complete genome	tRNA-Leu	108	6e-20	EU090238.1
Aspergillus tubingensis strain 0932 mitochondrion, complete genome	tRNA-Leu	62.1	5e-06	DQ217399.1
Aspergillus niger strain N909 mitochondrion complete genome	tRNA-Leu	62.1	5e-06	DQ207726.1

3.3.3 Discussion

The touchdown PCR using degenerate PCR only produced two fragments of expected

sizes: 1.5 kb *Pa-aflR* fragment and expected 1.1kb *Pa-avnA* fragment. The 1.5 kb fragment did not ligate into the vector. To repeat the experiment, several different vector : insert ratios could be tried for ligation, or ATP could be added to the ligation mixture to increase the T4 DNA ligase enzyme activity.

For the 1.1 kb putative *Pa-avnA* fragment, PCR using universal primers identified one positive colony, but the sequencing analysis showed that it was not *avnA* but part of a mitochondrial tRNA-Leu gene. The degenerate primers clearly bound to other places in the DNA template during PCR. Although the degenerate primers successfully identified DOTH biosynthesis genes in *D. septosporum*, may be the *Pa-avnA* gene in *P. arachidicola* is less similar to the *Aspergillus* sequences used to design the *avnA* degenerate primers. Some new degenerate primers could be designed according to the *D. septosporum* sequence to repeat the touchdown PCR, or nested PCR could be done to improve specificity for identifying the *Pa-avnA*, *Pa-norA* and *Pa-verb* genes.

This work was not continued due to the priority on investigating library clones (Section 3.4). Once fragments of *Pa-avnA*, *Pa-norA*, *Pa-verb* and *Pa-aflR* genes have been found, the fragments could be DIG-labelled for Southern blot and the Southern blot results could be used to screen the size fractionated *Eco*RI library.

3.4 A size fractionated *P. arachidicola* genomic DNA library

3.4.1 Background

A size fractionated *P. arachidicola* genomic DNA library was made to identify and sequence the DOTH genes in *P. arachidicola*. Fragments containing putative DOTH genes from *P. arachidicola* had already been found by degenerate PCR in previous work (Shuguang Zhang unpublished): 0.94 kb *Pa-cypA*, 1.35 kb *Pa-vbsA*, 0.7 kb *Pa-dotA* and 1.3 kb *Pa-pks*A. Those fragments were DIG-labelled for Southern hybridization and the results of this hybridization were used to direct the screening of the library.

3.4.2 Results

3.4.2.1 Restriction enzyme digestion of purified *P. arachidicola* 18667 gDNA

The extracted *P. arachidicola* 18667 gDNA was about 35-40kb (Figure 3.10A). PCR using primers ITS4 and ITS5 was carried out to amplify the ribosomal ITS region of purified 18667 genomic DNA to confirm that the DNA was from *P. arachidicola*. This was done because there was a lot of work with the DOTH producer *D. septosporum* at the time of the *P. arachidicola* library was made, and it was essential to rule out contamination. The sequence showed 100% identity to *Mycosphaerella arachidis*, which is the teleomorph of *P. arachidicola*.

Purified gDNA (4.2 μg) was digested with each enzyme (*Eco*RI, *Bam*HI, *Sal*I, or *Sca*I) (Section 2.2.7) and 200 ng of digested gDNA was checked on a 0.7% gel (Figure 3.10 B). The gDNA was not treated with RNase before digestion, so there were three RNA bands ranging from 3 kb to 4 kb presented for each digestion. The gel indicated complete digestion of the gDNA with each restriction enzyme. For each 4 μg of restriction enzyme digested gDNA, 2 μg was used for a Southern blot and the other 2 μg was used for making the size fractionated *P. arachidicola* gDNA library.

Figure 3.10 Purified *P. arachidicola* 18667 gDNA for restriction enzyme digestion

(A) Purified *P. arachidicola* 18667 gDNA **(B) Enzyme digestion of gDNA**

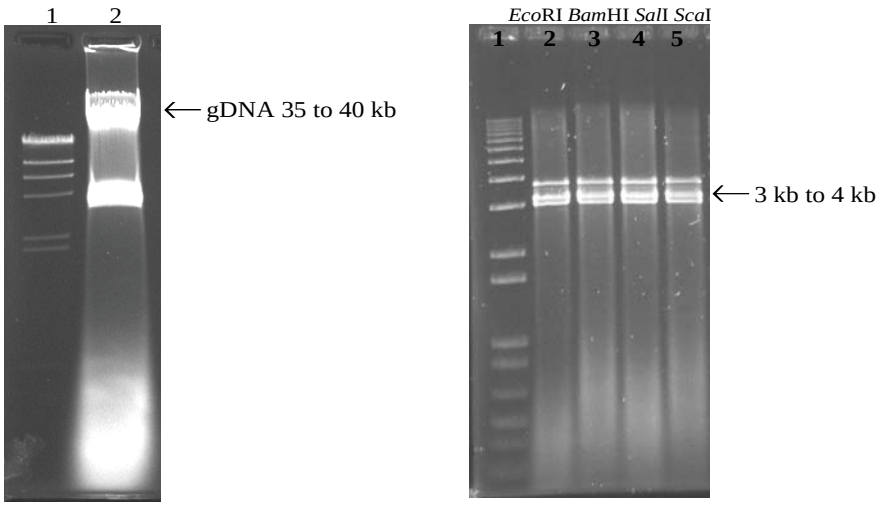


Figure 3.10 (A) Lane 1 is the λ DNA/ *Hind* III ladder. Lane 2 is the purified gDNA . (B) Lane 1 is the 1kb+ ladder. 18667 genomic DNA is shown digested with *Eco*RI (lane2), *Bam*HI (lane3), *Sal*I (lane 4) and *Sca*I (lane5). The three bands ranging from 3 kb to 4 kb are RNA in each lane.

3.4.2.2 Southern hybridization to identify the size of restriction fragments containing DOTH genes

Probes for genes *Pa-cypA*, *Pa-vbsA*, *Pa-dotA* and *Pa-pksA* were prepared by PCR amplification using plasmids containing *Pa-cypA*, *Pa-vbsA*, *Pa-dotA* and *Pa-pksA* fragments and specific primers designed according to the fragments sequences. The four PCR products (0.7 kb *Pa-cypA*, 1.35 kb *Pa-vbsA*, 0.9 kb *Pa-dotA* and 1.3 kb *Pa-pksA*) were extracted from a gel (Section 2.2.6) and used as DNA templates for DIG-labelling using a PCR based method (Section 2.5) (Figure 3.11 A). Because the DIG-labelled dUTPs were incorporated into the PCR products, the DIG-labelled probes had a higher molecular weight compared to the original PCR products (Figure 3.11 B).

Figure 3.11 DIG-labelled probes for Southern blotting

(A) PCR to amplify the templates **(B) DIG-labelled probes**

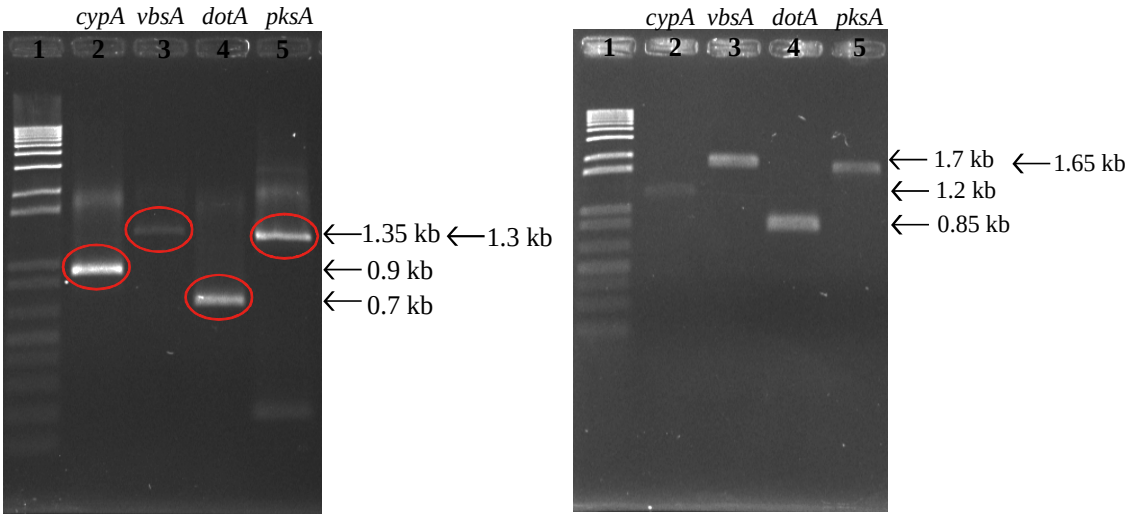


Figure 3.11 (A) Lane 1 is the 1kb+ ladder. Lanes 2 to 5 are *Pa-cypA*, *Pa-vbsA* *Pa-dotA* and *Pa*-pks*A* PCR products. The red circles indicate the PCR products extracted from the gel. (B) Lane 1 is the 1kb+ ladder. Lane 2 to 5 are DIG-labelled *Pa-cypA*, *Pa-vbsA* *Pa-dotA* and *Pa*-pks*A* DNA probes.

The working concentrations of probes used in the Southern hybridisation were detected as described in section 2.5.2. From the x-ray film, the 1/3125 dilution spot could be observed, but 1/625 dilution was used in experiment for each probe (Section 2.5.2, Figure 3.12).

Figure 3.12 Working concentration of probes

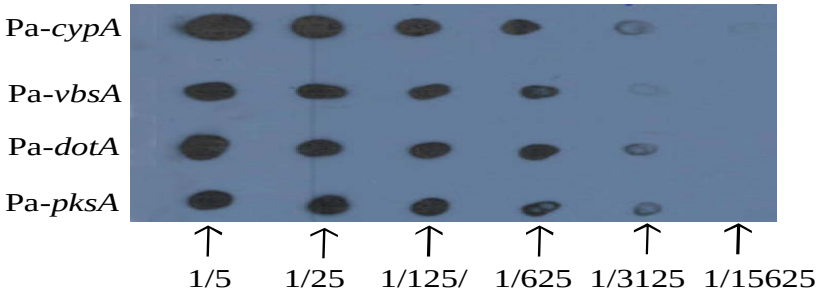


Figure3.12 The working concentration of probes used for Southern hybridization was 1/625 dilution.

The Southern blot was done as described in section 2.5.3. One probe was used at a time for Southern hybridization, then the membrane was stripped and another probe was used (Section 2.5.3 to Section 2.5.6). The restriction enzyme digested fragments that contained the DOTH genes are shown in Figure 3.13. For the *Pa-cypA* gene, the

*Bam*HI fragment was missing. This may be because the fragment that contains the *Pa-cypA* gene was too large and did not transfer from the gel to the membrane. For the *Pa-dotA* gene, the *Eco*RI cut 18667 gDNA had two fragments of 14 kb and 1.2 kb. This is because there is an *Eco*RI cutting site in the *Pa-dotA* probe. For *Pa-pksA* gene, the cut 18667 gDNA had two fragments 3.7 kb and 3.4 kb. This is because there is a *Sal*I cutting site in the *Pa-pksA* probe.

Figure 3.13 Southern blot results

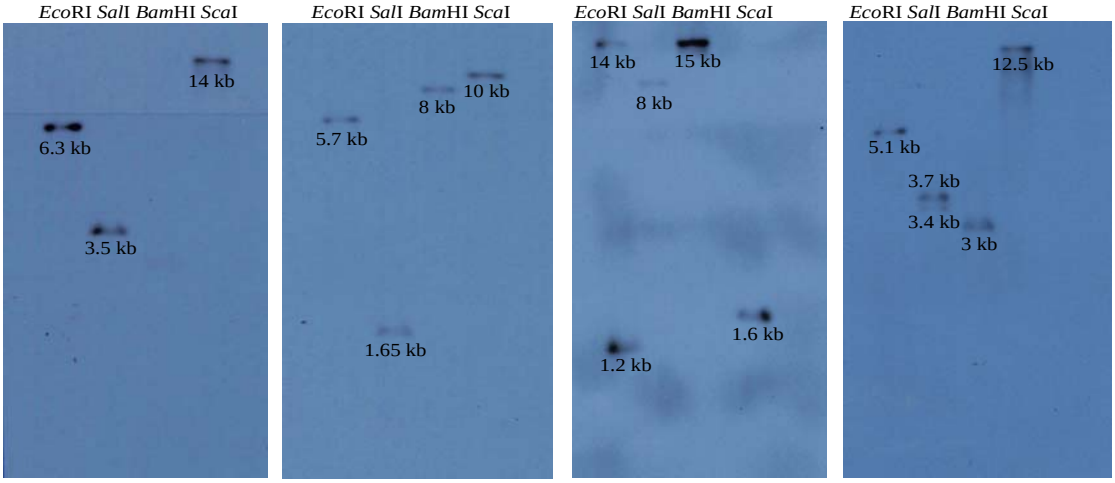


Figure 3.13 Genomic DNA digested with *Eco*RI, *Bam*HI, *Sal*I or *Sca*I was hybridised with *Pa-cypA, Pa-vbsA Pa-dotA or Pa-pksA* DIG-labelled probes.

Restriction digested fragments that contained the doth genes

Probe	Enzyme	Fragments(kb)		Probe	Enzyme	Fragments(kb)
Pa-cypA	*Eco*RI	6.3		*Pa-dotA*	*Eco*RI	14 and 1.2
	*Sal*I	3.5			*Sal*I	8
	*Bam*HI	-			*Bam*HI	15
	*Sca*I	14			*Sca*I	1.6
Pa-vbsA	*Eco*RI	5.7		*Pa-pksA*	*Eco*RI	5.1
	*Sal*I	1.65			*Sal*I	3.7 and 3.4
	*Bam*HI	8			*Bam*HI	3
	*Sca*I	10			*Sca*I	12.5

3.4.2.3 A size fractionated *EcoR*I *P. arachidicola* genomic DNA library

A plasmid library was made, because the plasmid library is easier to prepare and to handle than a cosmid library. When the library was made the first time, few white colonies were obtained. PCR and enzyme digestion results showed that the vector was contaminated. So a fresh vector plasmid was used and checked by enzyme digestion before transformation (data not shown).

The Southern blot results indicated the size of fragments that contained *P. arachidicola* DOTH genes and also suggested there was only a single copy of each gene in the genome. The *Eco*RI cut 18667 genomic DNA produced several large fragments that between them contained all four *P. arachidicola* DOTH genes, so the *Eco*RI cut 18667 genomic DNA was selected to make a size fractionated *P. arachidicola* gDNA library. The 2 μg of *Eco*RI cut 18667 gDNA was purified from a gel according to size (5-11 kb and 11-20 kb) to make insert DNA. The 6.3 kb Pa-*cypA* fragment, 5.7 kb *Pa-vbsA* fragment and 5.1 kb *Pa-pksA* fragment were within 5-11 kb size range and the 14 kb *Pa-dotA* fragment was within the 11-20 kb size range.

The pIC19H vector was digested with *EcoR*I and dephosphorylated by shrimp alkaline phosphatase to prevent self-ligation. The dephosphorylation reaction was checked by DNA T4 ligase (section 2.4.1.2) (Figure 3.14). From the gel, in lane 5, after DNA T4 ligase treatment, some *EcoR*I digested PIC19H vector self-ligated. Comparing lanes 2 and 3, after the vector had been dephosphorylated, the vector could not self-ligate. So the dephosphorylation of *EcoR*I digested PIC19H vector was successful.

Figure 3.14 Checking dephosphorylation of *EcoR*I digested PIC19H vector

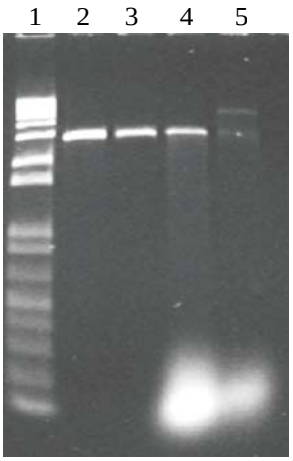


Figure 3.14 Lane 1 is the 1kb+ ladder. Lane 2 is the dephosphorylated *EcoR*I digested PIC19H vector. Lane 3 is the dephosphorylated *EcoR*I digested PIC19H vector with DNA T4 ligase. Lane 4 is the *EcoR*I digested PIC19H vector. Lane 5 is the *EcoR*I digested PIC19H vector with DNA T4 ligase.

The 5-11 kb and 11-20 kb *Eco*RI cut 18667 genomic DNA was inserted into dephosphorylated *EcoR*I digested pIC19H vector with a 1:1 molecular ratio using DNA T4 ligase (section 2.4.1.1). The ligation mixture was transformed into *E. coli* cells and the colonies were plated on LB + ampicillin plates with Xgal (Section 2.4.2).

3.4.2.4 Screen the library for positive clones

About 1000 white colonies were screened for each size fractionated library (about 200 white colonies per plate, 5 plates were screened). The DIG-labelled probes for colony hybridisation were the same DIG-labelled probes as used for the Southern blot. The 5-11 kb library was probed with *Pa-cypA*, *Pa-vbsA* or *Pa-pksA* probes and the 11-20 kb library was probed with the *Pa-dotA* probe. The positive colonies were detected as described in section 2.5.7. An example of colony hybridisation x-ray film using the *Pa-cypA* probe is shown in Figure 3.15. The 7 black spots on x-ray film indicate the positive colonies detected.

Figure 3.15 Colony hybridisation using *Pa-cypA* probe

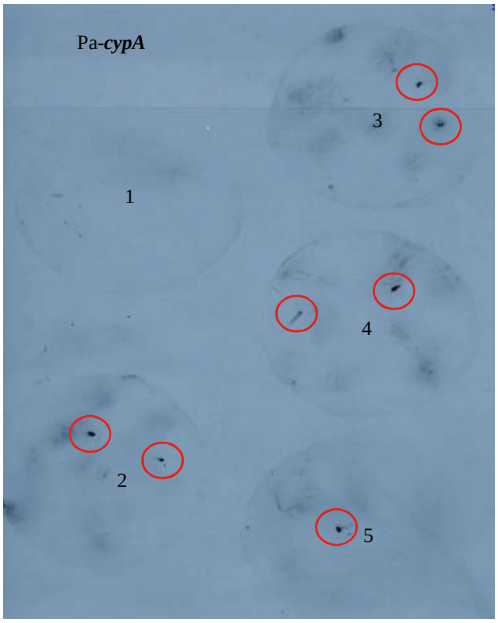


Figure3.15 7 The red circles indicate the 7 positive colonies found for *Pa-cypA* on hybridization membranes.

For a preliminary screen of the positive colonies from the library plates, *E. coli* colony PCR was carried out as described in section 2.3.5. Two positive controls were set up, one was using gDNA and primers used for Southern hybridization, and the other one was using a blue colony from library LB plate and universal M13 forward and M13 reverse primers. The PCR results are shown in Figure 3.16. Three of the colonies had the expected 0.9 kb PCR band for the *Pa-cypA* gene, 2 colonies had the expected 1.35 kb PCR band for *Pa-vbsA* gene, only 1 colony had the expected 0.5 kb PCR band for *Pa-dotA* gene and 3 colonies had the expected 1.3 kb PCR band for *Pa-pksA* gene.

Purified single cell plasmid of positive colony was extracted (section 2.2.3) and digested with *Eco*RI to check the insert fragments (Figure3.17). All of the checked plasmids had a 2.7 kb vector DNA and the expected size of insert DNA, which were 6.3 kb *Pa-cypA* (renamed as plasmid pR285), 5.7 kb *Pa-vbsA* (renamed as plasmid pR283), and 5.1 kb *Pa- pksA* (renamed as plasmid pR284). The *Pa-dotA* plasmid (renamed as plasmid pR286) had a 14 kb expected insert band and a 10 kb unexpected band. The plasmids were sequenced as described in section 2.6.

Figure 3.16 *E. coli* colony PCR to screen the positive colonies

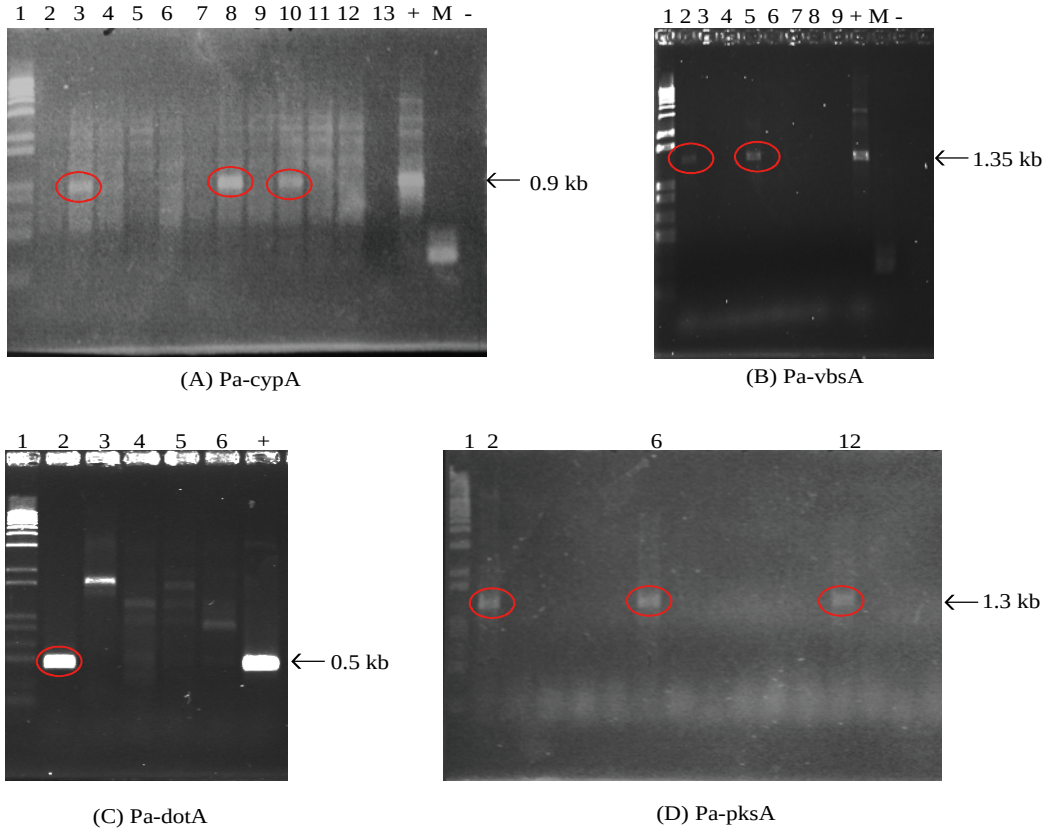


Figure 3.16 For all four photos, Lane 1 is the 1kb+ ladder. Lane + is the gDNA positive control. Lane M is the vector colony positive control. Lane – is the negative control. The red circles indicate the bands with the expected sizes. (A) Lane 3,8 and 10. (B) Lane 2 and 5. (C) Lane 2. (D) Lane 2, 6 and 12.

Figure 3.17 *Eco*RI digestion to check the insert DNA

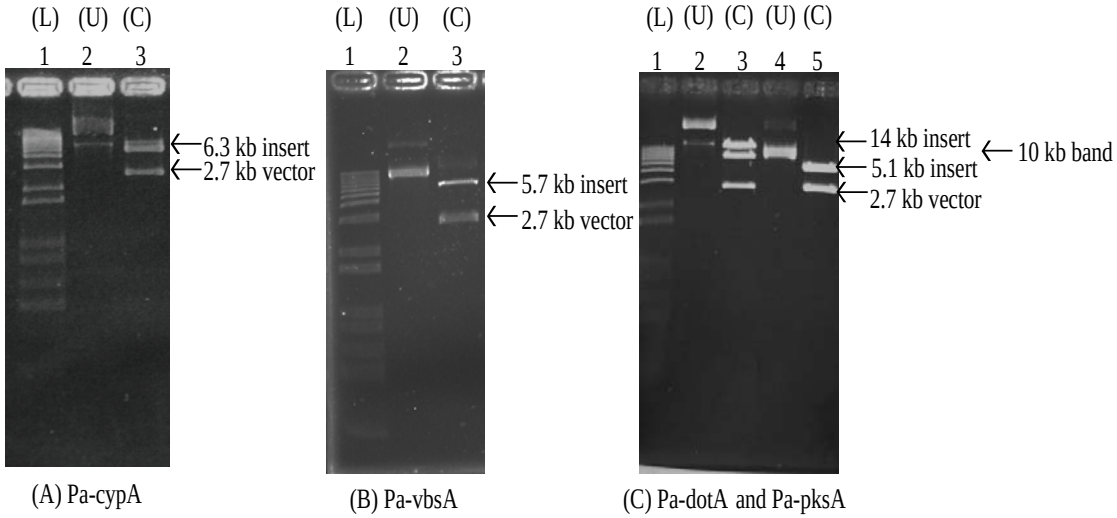


Figure 3.17 For all three photos, Lane 1 (L) is the 1kb+ ladder, (U) is uncut plasmid, (C) is *EcoR*I digested plasmid and the 2.7 kb band was the vector DNA. (A) The *Pa-cypA* plasmid had a 6.3 kb insert. (B) The *Pa-vbsA* plasmid had a 5 .7 kb insert. (C) The *Pa-dotA* plasmid lanes 2 and 3 had a 14 kb insert and a 10 kb band. *Pa- pksA* plasmid lanes 4 and 5 had a 5.1 kb insert.

3.4.3 Discussion

P. arachidicola is a slow growing fungus. In PDA culture, it normally took about two weeks for the fungus to grow. It was important to check the ribosomal ITS sequence of the extracted *P. arachidicola* 18667 to make sure the strain was not contaminated during the growth procedure. There were some difficulties in isolating enough *P. arachidicola* genomic DNA at the beginning, a CTAB based DNA extraction method was used later to increase the yield of genomic DNA.

The 10 kb band that appeared on the *Eco*RI digested *Pa-dotA* plasmid was not expected (Figure 3.15). The *E. coli* containing the *Pa-dotA* plasmid was single-cell purified to make sure it was not a mixed culture. Due to the large size of the insert, it is unlikely (although possible) that the insert is a chimeric clone with a 24 kb clone. To find out what this band is, the 10 kb band could be extracted from the gel and cloned into a vector for sequencing.

Apart from the size fractionated *Eco*RI *P. arachidicola* gDNA library, a size fractionated *Sca*I *P. arachidicola* genomic DNA library was also prepared (5-11 kb and 11-20 kb). Because there was no *Sca*I cutting site in the multiple cloning site of PIC19H, the vector was cut with *Sma*I. Both *Sca*I and *Sma*I digested DNA had blunt end, so the ScaI digested *P. arachidicola* gDNA was inserted into *Sma*I digested and dephosphorylated PIC19H vector. But only 400 white colonies were obtained for each size fractionated *Sca*I library and no positive colonies were found using the same DIG-labelled probes that had been used for screening the *Eco*RI library.

3.5 Genes identified in plasmids pR284 and pR285 (*P. arachidicola pksA* cluster)

3.5.1 Background

The ST/AF biosynthesis gene clusters are well established in *Aspergillus* species. A 70 kb DNA region on chromosome 3 contains 25 clustered genes that are involved in biosynthesis of AF in *A. parasiticus*. 22 ST biosynthesis genes clustered on chromosome 4 in *A. nidulans* have homologues of many of AF genes. The function and regulatory mechanisms of ST biosynthesis genes are similar to those of the AF biosynthesis genes, but the gene order and organization are different between the ST and AF gene clusters (Brown et al., 1996b; Yu et al., 2004a; Yu et al., 2004b).

By using AF genes as hybridisation probes, ten putative DOTH biosynthesis genes were identified in *D. septosporum* (Bradshaw and Zhang, 2006). Different from the AF and ST biosynthesis cluster, the DOTH biosynthesis genes are located in three mini-clusters on a 1.3 Mb chromosome, and genes not related to DOTH biosynthesis are located adjacent to or between the DOTH biosynthesis genes in the mini-cluster (Zhang et al., 2007).

One mini cluster contains *Ds-pksA* and four additional genes (*Ds-cypA, Ds-moxA, Ds-avfA* and *Ds-epoA*) clustered immediately alongside *Ds-pksA*. Gene replacement and complementation experiments showed that *Ds-pksA* is required for DOTH biosynthesis in *D. septosporum* (Bradshaw et al., 2006).

P. arachidicola, a peanut pathogen, also produces DOTH. Genes with high predicted amino acid identity to *Ds-pksA* mini cluster genes have been identified using *D. septosporum* DOTH genes as probes (Section 3.4). Those genes are analysed in this section.

3.5.2 Results

3.5.2.1 Primary sequence analysis of plasmids pR284 and pR285

Plasmids pR284 (containing a 5.1 kb *Pa-pksA* fragment) and pR285 (containing a 6.3 kb *Pa-cypA* fragment) that had been isolated from a genomic library of *P. arachidicola* by hybridization (Section 3.3.2.4) were sequenced by primer walking (Section 2.6.1). Primers used for sequencing included universal primers M13 forward and reverse, as well as primers previously used for amplifying the probes for Southern blot (section 3.3.2.2) and library screening (section 3.3.2.4). New custom primers were designed using a computer program Clone Manger SECentral (Section 2.6.2). The overall sequencing strategy is shown in Figure 3.18.

Sequence assembly for each plasmid was done using ContigExpress. Plasmid pSZ-Pa-pksA-PCR which contains a *Pa-pksA* fragment obtained using degenerate PCR by Dr. S. Zhang (Massey University), overlapped with both pR284 and pR285. Together, a total contiguous sequence of 11.5 kb was obtained. BLAST results suggested five open reading frames (ORFs) clustered within the 11.5 kb fragment. The predicted genes showed high amino acid identity to DOTH biosynthesis gene products from *D. septosporum* and some amino acid identity to AF biosynthesis and ST biosynthesis gene products from *Aspergillus* species. The five ORFs in the 11.5 kb fragment were predicted to be *Pa-pksA* that encodes a polykide synthase, *Pa-cypA* that encodes cytochrome P450 monooxygenase, *Pa-avfA* that encodes averufin oxidase, *Pa-epoA* that encodes epoxide hydrolase and *Pa-moxA* that encodes putative flavin-binding monooxygenase (Table 3.3, Figure 3.18). Southern blot analysis confirmed that there was only one copy of each predicted gene in *P. arachidicola* (Southern blot results section 3.3.2.2).

Figure 3.18 Schematic diagram of the 11.5 kb contig showing five ORFs and positions of primers used for sequencing

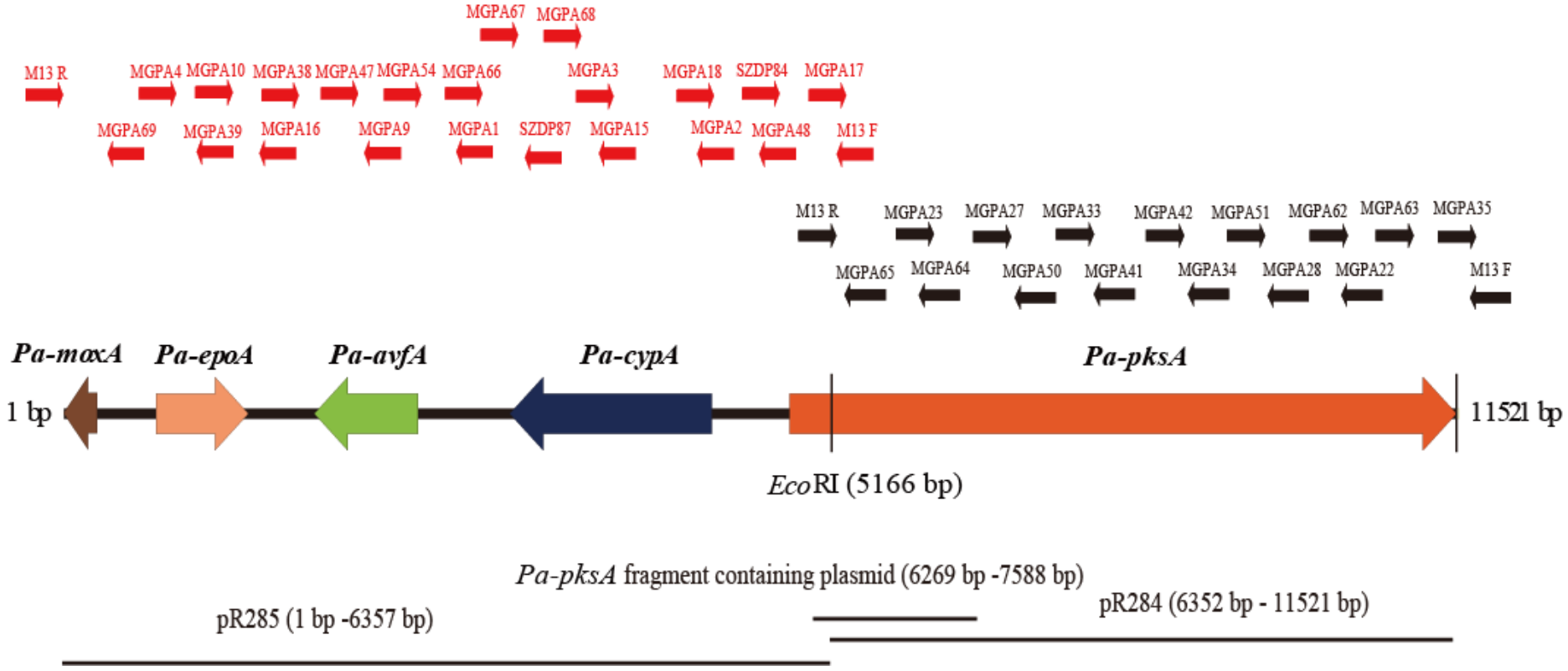


Figure 3.18 The black and red arrows indicate the primers used for sequencing the pR284 and pR285 plasmids. The black lines indicate how existing plasmid pSZ-Pa-pksA-PCR Sequence overlapped with pR284 and pR285. There are five ORFs detected in this 11.5 kb fragment. Large coloured arrows indicate the direction of transcription of each ORF.

Table 3.3 Main BLASTX results of 11.5 kb fragment

P. arachidicola gene position (11.5 kb fragment)	Aligned protein (BLASTX)	Score	E value	Genbank accession number
1 bp - 276 bp (*Pa-moxA*)	putative flavin-binding monooxygenase *Mycosphaerella pini*	84	2.00E-16	AAZ95013.1
	MoxY *Aspergillus nomius*	69.3	9.00E-11	AAS90063.1
762 bp - 1528 bp (*Pa-epoA*)	epoxide hydrolase *Mycosphaerella pini*	48.9	4.00E-10	AAZ95015.1
2074 bp -2934 bp (*Pa-avfA*)	averufin oxidase *Mycosphaerella pini*	410	8.00E-113	AAZ95014.1
	AvfA *Aspergillus flavus*	244	8.00E-63	AAS90043.1
3690 bp - 5346 bp (*Pa-cypA*)	cytochrome P450 monooxygenase *Mycosphaerella pini*	701	0	AAZ95016.1
	CypX *Aspergillus flavus*	493	6.00E-164	AAS90107.1
6007 bp - 11520 bp (*Pa-pksA*)	polyketide synthase *Mycosphaerella pini*	2872	0	AAZ95017.1
	PksA *Aspergillus flavus*	2075	0	AAS89999.1

Note: *Dothistroma septosporum* (teleomorph: *Mycosphaerella pini*)

There is no epoxide hydrolase gene in the AF/ST gene cluster

3.5.2.2 The *P. arachidicola Pa-pksA* gene

The 11.5 kb fragment only contained part of a predicted *Pa-pksA* gene, which had an ORF of 5514 bp and encoded a polypeptide of 1804 amino acids. The amino acid sequence of partial Pa-PksA has 88.7% identity to the corresponding region of Ds-PksA, and more than 50% identity to *Aspergillus* species Pks proteins *stcA* and *aflC*. However Pa-PksA has low amino acid identity to PKS proteins involved in biosynthesis of other metabolites (Table 3.4). Two introns of 50 bp and 52 bp were identified in partial *Pa-pksA* by comparing predicted *Pa-pksA* and *Ds-pksA* amino acid sequences and looking for the presence of 5' GT(AGT)NGTY and 3' YAG fungal consensus intron splice site sequences (Ballance, 1986). The positions and lengths of these introns were similar to the positions and lengths of introns in *Ds-pksA* (Figure 3.19).

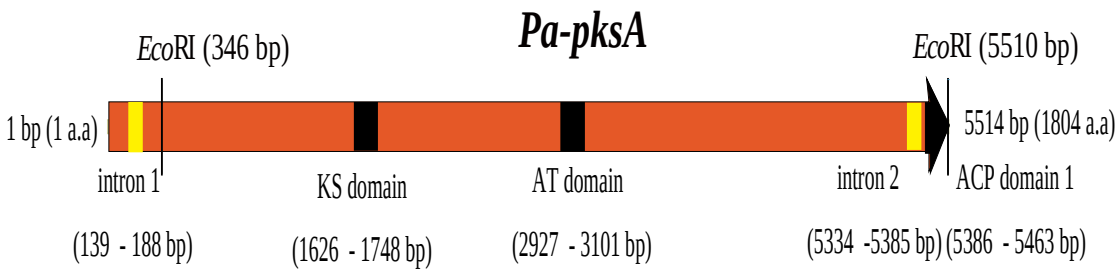
The *Pa-pksA* was predicted to encode a polyketide synthase as it contains one each of characteristic β-ketoacyl synthase (KS), acyl transferase (AT), and acyl carrier protein (ACP) domains within the partial Pa-PksA. These domains were identified by alignment of Pa-PksA with Ds-PksA and other Pks proteins. The orders and the positions of those domains are conserved between Pa-PksA and Ds-PksA (Figure 3.20). *Ds-pksA* gene has three ACP domains in the carboxy terminal followed by a TE domain, whilst the *aflC* gene has two ACP domains followed by a TE domain (Zhang et al., 2007). Because the entire sequence of *Pa-pksA* was not known, it was unclear whether the carboxy terminal of *Pa-pksA* had similar structure to the *Ds-pksA* or *aflC* genes. It is expected that 595 amino acids of *Pa-pksA* remain to be sequenced. The partial *Pa-pksA* gene sequence and alignments are shown in Appendix IV and V.

Table 3.4 Amino acids identity of Pa-PksA to other Pks proteins

Organism	Amino acid identity	Genbank accession number	Gene product function
D. septosporum	88.70%	AAZ95017	DOTH biosynthesis
A. nidulans	50.10%	AAA81586	sterigmatocystin biosynthesis
A. parasiticus	53.10%	AAS66004	AF biosynthesis
A. flavus	53.50%	AAS89999	AF biosynthesis
Gibberella fujikuroi	30%	CAB92399	pigment biosynthesis
Botryotina fuckeliana	29.90%	AAR90249	melanin biosynthesis

Table 3.4 The predicted amino acid identity of Pa-PksA was compared with other PKS proteins, the highest amino acid identity was 88.7% with Ds-PksA, which also produces DOTH. Much lower amino acid identities were seen with PksA proteins involved in pigment and melanin biosynthesis.

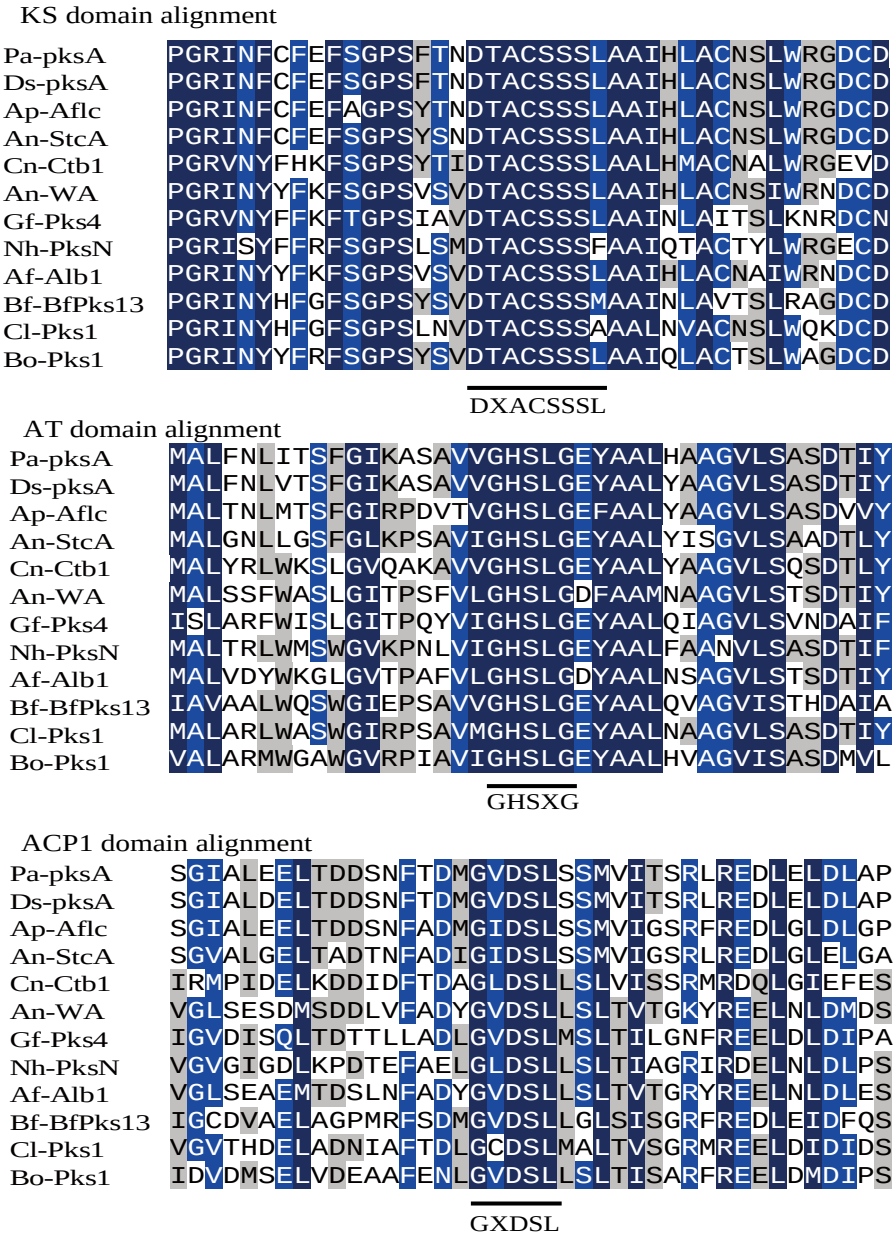
Figure 3.19 Schematic diagram of *Pa-pksA*



Gene	ORF length (bp)	Intron 1	Intron 2
Pa-pksA	5514 (Partial)	139 bp - 188 bp (50 bp)	5334 bp -5385 bp (52 bp)
Ds-pksA	7304	138 bp - 186 bp (49 bp)	5342 bp - 5396 bp (55 bp)

Figure 3.19 The upper diagram shows the partial *Pa-pksA* gene, which has an ORF of 5514 bp and encodes a polypeptide of 1084 amino acids. The yellow boxes indicate the positions of two introns. The positions and length of introns in *Pa-pksA* and *Ds-pksA* are listed in the table below. The black arrow and boxes indicate KS, AT and ACP domains.

Figure 3.20 Alignments of Pa-PksA KS, AT and ACP domains with other PKS proteins



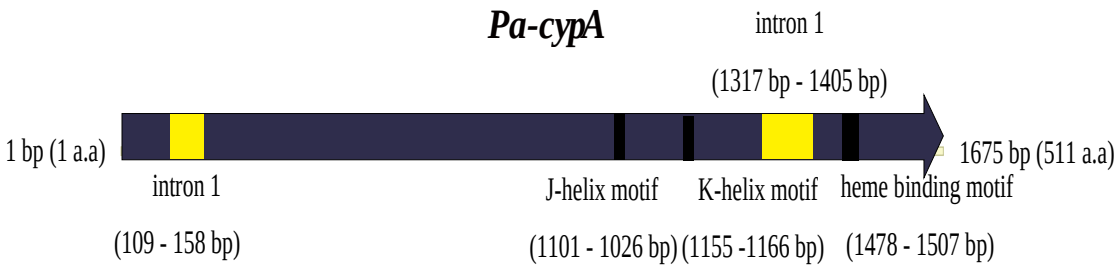
Name	organism	Protein	Genbank No.
Pa-PksA	*P. arachidicola*	PksA	
Ds-PksA	*D. septosporum*	PksA	AAZ95017
Ap-AflC	*A. parasiticus*	AflC	AAS66004
An-StcA	*A. nidulans*	StcA	AAA81586
Cn-Ctb1	*Cercospora nicotianae*	Ctb1	AAT69682
Ao-Pks	*A. ochraceus*	Pks	AAP32477
An-WA	*A. nidulans*	WA	CAA46695
Gf-Pks4	*Gibberella fujikuroi*	Pks4	CAB92399
Nh-PksN	*Nectria haematococca*	PksN	AAS48892
Af-Alb1	*A. fumigatus*	Alb1	AAC39471
Bf-BfPks13	*Botryotina fuckeliana*	BfPks13	AAR90249
Cl-Pks1	*Colletotrichum lagenarium*	Pks1	BAA18956
Bo-Pks1	*Bipolaris oryzae*	Pks1	BAD22832

Figure 3.20 Alignments of Pa-PksA with other fungal Pks proteins. The KS, AT and ACP domains were identified. Amino acid identities are shaded. Core conserved catalytic regions are shown under alignment.

3.5.2.3 The *P. arachidicola Pa-cypA* gene

The *Pa-cypA* had an ORF of 1675 bp that was predicted to encode a polypeptide of 511 amino acids. The amino acid sequence of Pa-CypA had 92.8% identity to Ds-CypA, 60.2% identity to CypX protein in *A. parasiticus* and 59.9% identity to StcB protein in *A. nidulans*, suggesting that Pa-CypA is also a cytochrome P450 monooxygenase. Two introns of 50 bp and 89 bp were identified. The positions of these introns were similar to the *Ds-cypA*, but intron two differs in length (Figure 3.21). Alignment of Pa-CypA with other CypA proteins identified characteristic J-helix, K-helix and heme binding motifs in Pa-CypA (Figure 3.21 and Figure 3.22).

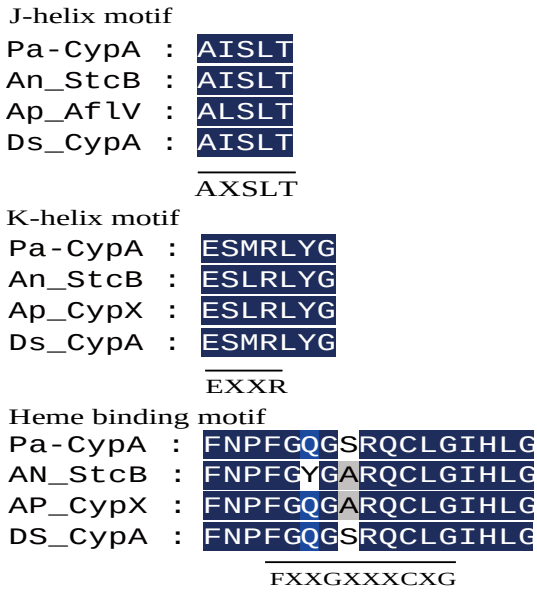
Figure 3.21 Schematic diagram of *Pa-cypA*



Gene	ORF length (bp)	Intron 1	Intron 2
Pa-cypA	1675	109 bp - 158 bp (50 bp)	1317 bp -1405 bp (89 bp)
Ds-cypA	1640	109 bp - 159 bp (51 bp)	1317 bp - 1369 bp (53 bp)

Figure 3.21 The upper diagram shows the *Pa*-cyp*A* gene ORF. The yellow boxes indicate the positions of two introns, the position and length of introns in *Pa-cypA* and *Ds*-cyp*A* are listed in the table below. The black boxes indicate the K-helix, J-helix and heme binding motifs.

Figure 3.22 Characteristic cytochrome P450 monooxygenase motifs in Pa-CypA



Name	organism	Protein	Genbank No.
Pa-CypA	*P. arachidicola*	CypA	
Ds-CypA	*D. septosporum*	CypA	AAZ95016
Ap-AflV	*A. parasiticus*	AflV	AAS66022
An-StcB	*A. nidulans*	StcB	EAA61612

Figure 3.22 Alignments of Pa-CypA with other cytochrome P450 monooxygenase proteins. The J-helix motif, K-helix motif and the heme binding motif were identified. Amino acid identities are shaded. Core conserved catalytic regions are shown under each alignment.

3.5.2.4 The *P. arachidicola Pa-avfA* gene

The *Pa-avfA* had an ORF of 861 bp and is predicted to encode an averufin oxidase of 286 amino acids. The Pa-AvfA has 73.3% amino acids identity to Ds-AvfA protein, 46.9% amino acid identity to AflI and 46.8% amino acid identity to StcO. (Figure 3.23). No intron was founded in the *Pa-avfA* in contrast to *Ds-avfA*, which has an intron at position 60 bp to 123 bp. The Pa-AvfA had similar length to the oxidase protein AflI in *A. parasiticus*, but was 19 and 11 amino acids shorter than the *D. septosporum* AvfA and *A. nidulans* oxidase protein StcO respectively.

Figure 3.23 Schematic diagram of *Pa-avfA*

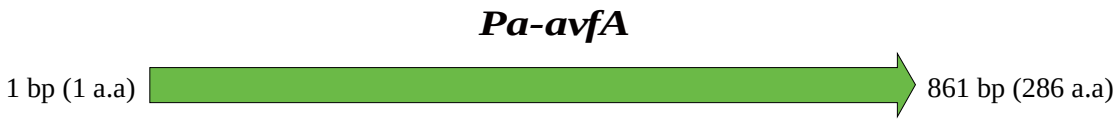
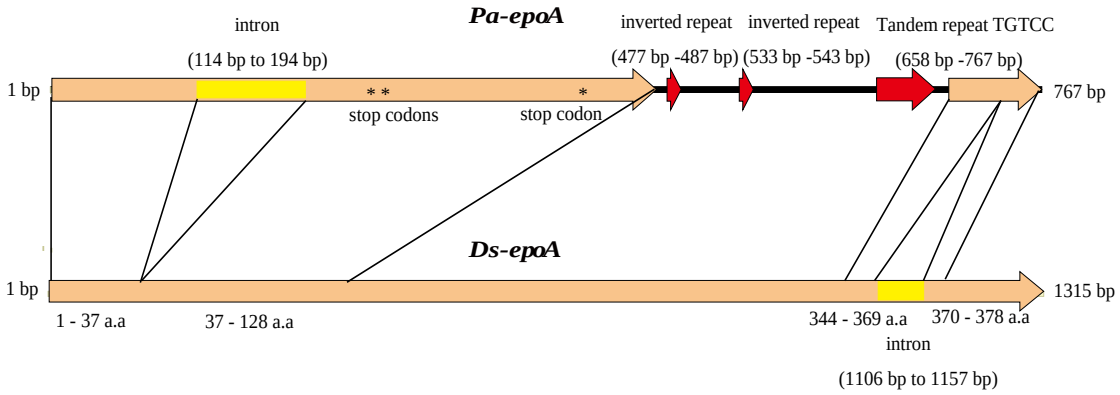


Figure 3.23 A schematic diagram of the *Pa-avfA* gene. There is no intron in the sequences; the gene encodes an averufin oxidase of 286 amino acids.

3.5.2.5 The *P. arachidicola Pa-epoA* gene

BLASTX results showed that amino acid sequence predicted from the *Pa-epoA* gene matched to three parts of the *Ds-epoA* amino acid sequence. The first two matches corresponded to exons flanking an intron of 81 bp at position 114 bp to 194 bp in *Pa-epoA* (1-468 bp). In contrast, *Ds-epoA* does not have an intron in this region but instead has one intron at position 1106 bp to 1157 bp. The third BLASTX match of *Pa-epoA* to *Ds-epoA* was of a short region in the downstream region. *Pa-epoA* sequences from 658 bp to end had a 79% amino acid identity to the *Ds-epoA* 344 a.a to 378 a.a sequences. Alignment of *Pa-epoA* to *Ds-epoA* revealed that 228 a.a of the expected coding region of *Pa-epoA* was missing and instead a repeat rich region was seen (477 bp -686 bp). This region does not match any known hydrolase sequences. Inverted repeat sequences were found at positions 477 - 487 bp and 533 - 543 bp and a tandem repeat also found at position 641 - 686 bp. In *Ds-epoA*, there were three of the highly conserved active site amino acids: Asp^{197}, Thr^{346} and His^{372}. In *Pa-epoA*, only one active site His^{372} was identified. In position 346, a Pro was identified instead of Thr and the region expected to contain Asp was missing. There were three stop codons in this *Pa-epoA* coding region. The original sequencing results were checked, signals were good and clear at the positions of the stop codon regions (Figure 3.24).

Figure 3.24 Comparison of the *Pa-epoA* gene and the *Ds-epoA* gene



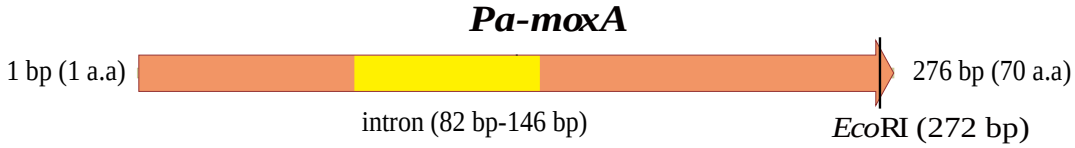
Gene	ORF length (bp)	Intron 1
Pa-epoA	768	114 bp - 194 bp (81 bp)
Ds-epoA	1315	1106 bp - 1157 bp (52 bp)

Figure 3.24 The yellow boxes indicate the positions of introns. The position and length of introns in *Pa-epoA* and *Ds*-epo*A* are listed in the table below. The black lines indicate the regions of *Pa-epoA* matched to *Ds-epoA*. The red arrows indicate the positions of inverted repeats and tandem repeats. The black stars indicate the positions of stop codons.

3.5.2.6 The *P. arachidicola Pa-moxA* gene

A partial *Pa-moxA* gene was identified at the opposite end of the 11.5 kb fragment to *Pa-pksA*. The partial *Pa-moxA* had an ORF of 276 bp with one intron of 65 bp and encoded a polypeptide of 70 amino acids. The position of the intron in *Pa-moxA* was slightly different from the first intron in *Ds-moxA* and was 10 bp longer (Figure 3.25). The first 70 amino acids of Pa-MoxA encoded by this fragment had 70.4% identity to the corresponding region of Ds-MoxA, but only 10.8% and 25.3% identities to those of AflW and StcW in *Aspergillus* species respectively

Figure 3.25 Schematic diagram of *Pa-moxA* gene



Gene	ORF length (bp)	Intron 1
Pa-moxA	276	82 bp - 164 bp (65 bp)
Ds-moxA	2163	98 bp - 152 bp (55 bp)

Figure 3.25 Schematic diagram of the partial *Pa-moxA* gene with ORF of 276 bp. The yellow box indicates the position of intron, the position and length of introns in *Pa-moxA* and *Ds*-mox*A* are listed in the table below.

3.5.2.7 The dothistromin biosynthesis genes in *P. arachidicola* (*pksA* cluster)

The DOTH biosynthesis genes identified in the 11.5 kb fragment from *P. arachidicola* were compared to the corresponding region of the DOTH biosynthesis cluster in *D. septosporum* and to AF and ST biosynthesis genes in *Aspergillus* species. Four of the *P. arachidicola* genes *Pa-pksA*, *Pa-cy*pA, *Pa-avfA* and *Pa-epoA* had the same orientations as the homologous *D. septosporum* DOTH biosynthesis genes. The *Pa-moxA* gene had opposite gene orientation to *Ds-moxA*.

Figure 3.26 Comparison of *D. septosporum* and *P. arachidicola* dothistromin *pksA* biosynthesis gene clusters

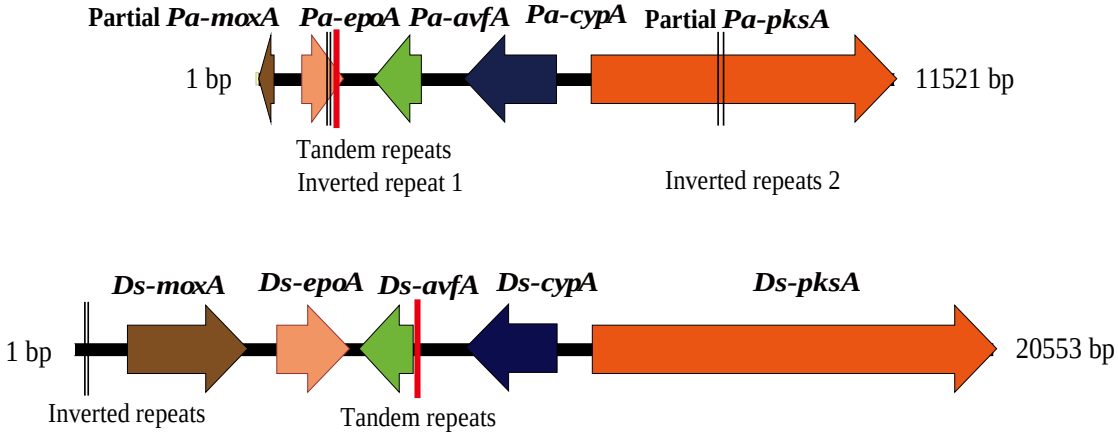


Figure 3.26 Schematic diagrams of partial *D. septosporum* and *P. arachidicola* DOTH biosynthesis gene clusters. The order and orientation of each gene is presented with colour arrows. The red vertical lines show the positions of tandem repeats. Black vertical lines show positions of inverted repeats.

The intergenic regions between the DOTH biosynthesis genes in *P. arachidicola* and *D. septosporum* were compared (Table 3.5). The intergenic regions of *cypA* and *pksA* between *P. arachidicola* and *D. septosporum* had similar length and highest nucleotide sequence identity of 70.6%. The intergenic regions of *epoA* and *avfA* between *P. arachidicola* and *D. septosporum* varied considerably in size and had the lowest nucleotide identity of only 15.9%. A tandem repeat (TGTCC)n was found inside the *Pa-epoA* ORF region, this was different in position and sequence to a *D. septosporum* cluster tandem repeat (CAGTGC)n at the end of the *Ds-avfA* gene. The *P. arachidicola* cluster also had two inverted repeat sequences within the ORFs of *Pa-epoA* and *Pa-pksA* that were not seen in the corresponding regions of *D.*

septosporum (Figure 3.24, Table 3.6 A, B).

Table 3.5 Intergenic regions of *P. arachidicola* and *D. septosporum* dothistromin genes

Region	Position	Length	GC content	Nucleotide identity
Pa moxA to *epoA*	227 bp - 761 bp	485 bp	48%	53%
Ds moxA to *epoA*	29485 bp - 30013 bp	529 bp	51%	
Pa epoA to *avfA*	1530 bp - 2074 bp	545 bp	55%	15.9%
Ds epoA to *avfA*	31329 bp - 33418 bp	176 bp	45%	
Pa avfA to *cypA*	2936 bp - 3690 bp	755 bp	54%	43.5%
Ds avfA to *cypA*	32475 bp - 33418 bp	944 bp	54%	
Pa cypA to *pksA*	5366 bp - 6007 bp	642 bp	54%	70.6%
Ds cypA to *pksA*	35095 bp - 35717 bp	659 bp	54%	

Table 3.6 (A) Inverted repeats

Inverted repeat	Location	Repeat length	Sequence
Ds inverted repeat	26553-26562 26613-26604	10	ctcacctgct |||||||||| gagtggacga
Pa inverted repeat 1	1239-1249 1305-1295	11	gggggggggg |||||||||| cccccccccc
Pa inverted repeat 2	8294-8303 8405-8396	10	tccctcctca |||||||||| agggaggagt

Table 3.6 (B) Tandem repeats

Tandem repeat	Location	Copy No.	Consensus pattern
Ds	32559-32650 bp	15 of 6 bp	CAGTGC
Pa	1403-1448 bp	9 of 5 bp	TGTCC

3.5.3 Discussion

The partial *Pa-pksA* identified in this study showed high amino acid identity to *D. septosporum Ds-pksA*, *A. parasiticus aflC* and *A. nidulans stcA* genes which encoded polykide synthase. Functional inactivation of *stcA* blocked the production of ST and all ST intermediates in *A. nidulans*. Further studies showed that *stcA* was involved in

early steps of ST biosynthesis, which was converting hexanoyl CoA and seven malonates to NOR (Brown et al., 1996b; Yu and Leonard, 1995). Disruption of *aflC* produced mutants unable to produce AF in *A. parasiticus* (Trail et al., 1995). Disruption of *Ds-pksA* gene in *D. septosporum* was unable to produce DOTH, but feeding the *Ds-pksA* mutant with NOR or VERA (precursors in AF biosynthesis pathway) resulted in production of DOTH. Those experiments showed that *Ds-pksA* functions at an early stage in DOTH biosynthesis with VERA acting as a precursor (Bradshaw et al., 2006; Bradshaw and Zhang, 2006). *Pa-pksA*, which also encodes a polykide synthase, may be involved in production of NOR from hexanoyl CoA and malonates in DOTH biosynthesis in *P. arachidicola*.

The *Pa-avfA* gene showed high amino acid identity to *D. septosporum Ds-avfA*, *A. parasiticus aflI* (previously called *avfA*) and *A. nidulans stcO* genes which encode averufin oxidase. Gene complementation experiments showed that an averufin-accumulating, non-aflatoxigenic mutant strain of *A. parasiticus* can be rescued by an equivalent gene from *A. flavus* to produce AF, So *aflI* is involved in converting AVF to VHA (Yu et al., 2000). Further studies showed that two genes *aflV* and *aflW* are also involved in converting of AVF to VHA in *A. parasiticus* (discussed later). The *Pa-avfA* predicted to encode averufin oxidase may have a similar function in DOTH biosynthesis.

The *Pa-cypA* gene was predicted to encode a cytochrome P450 monooxygenase. It showed high amino acid identity to *D. septosporum Ds*-cyp*A*, *A. parasiticus aflV* and *A. nidulans stcB* genes. Mutants with disruption of *stcB* or *aflV* cannot produce AF and resulted in accumulation of AVR in *A. nidulans* and *A. parasiticus* respectively. Feeding experiments showed that the *aflV* mutant was able to convert HVN and VHA to AF but failed to convert OAVN or AVR to AF. So *aflV* is involved in converting AVR to HVN in AF biosynthesis in *A. parasiticus* (Wen et al., 2005). *Pa-cypA* may be involved in converting AVR to HVN in DOTH biosynthesis in *P. arachidicola*.

The partial *Pa-moxA* identified in this study showed high amino acid identity to the *D. septosporum Ds-moxA*, *A. parasiticus aflW* and *A. nidulans stcW* genes which encoded flavin-binding monooxygenase. Disruption of *aflW* blocked AF biosynthesis and resulted in accumulation of HVN in *A. parasiticus*. Feeding experiments showed that, the *aflW* mutant was able to convert VHA to AF but cannot convert OAVN, AVR or HVN to AF. So *aflW* is involved in converting HVN to VHA in AF biosynthesis in *A. parasiticus* (Wen et al., 2005). *Pa-moxA* may be involved in converting HVN to VHN in DOTH biosynthesis in *P. arachidicola*.

Only single stranded sequence of the *Pa-epoA* region was obtained. Clear second-strand sequence could not be obtained even though the primer sequence was checked and the plasmid template was purified. There was a G-C inverted repeat identified in the *Pa-epoA* coding region (Table 3.6 A). A secondary loop structure could be formed here, leading to unsuccessful sequencing of the double stranded *Pa-epoA* coding region. Increased temperature could be used to disrupt the secondary structure in future sequencing reactions.

No epoxide hydrolase gene is found in ST/AF gene clusters in *Aspergillus* species. There is a *Ds-epoA* identified in *D. septosporum* just downstream of the *Ds-avfA* gene. Although the *Ds-epoA* gene was expressed at an early growth in coordination with the *Ds-pksA* and *Ds-dotA* DOTH genes (Schwelm et al., 2008), the actual function of *Ds-epoA* in DOTH biosynthesis is not clear. Studies by Hongping Jin at Massey university showed that deletion of *Ds-epoA* by homologous recombination produced a mutant strain with similar growth rate, sporulation rate and dothistromin biosynthesis to wild type strain. This suggested that *Ds-epoA* is not involved in early DOTH biosynthesis steps. Later studies by Townsend suggested that in late step of DOTH biosynthesis, an epoxide intermediate is present by rearrangement of VERA to form an A ring edge epoxide. This step requires an epoxide hydrolase, so *Ds-epoA* may have a role in this late step (Henry and Townsend, 2005). Like the DOTH biosynthesis gene cluster in *D. septosporum*, there was a *Pa-epoA* gene, which is

predicted to encode an epoxide hydrolase located between the *Pa-avfA* and *Pa-moxA* genes. The presence of three stop codons in the coding region (the signals are clear at the positions of stop codon) and the deletion of 228 a.a sequence suggested that *Pa-epoA* gene seems to be truncated and fragmented in *P. arachidicola*. The lack of two of the three expected conserved amino acids also suggests that *Pa-epoA* may be not functional in the biosynthesis of DOTH in *P. arachidicola*. Since *Ds-epoA* may have a role in late step of DOTH biosynthesis, but the *Pa-epoA* identified so far may be non-functional, there may be more than one copy of *Pa-epoA* in *P. arachidicola*. A Southern blot could be used to check the copy number of *Pa-epoA*.

The arrangement of DOTH biosynthesis genes in *D. septosporum* and ST/AF biosynthesis genes in *Aspergillus* species are quite different. It seems that the genes involved in DOTH biosynthesis in *P. arachidicola* are arranged in a similar way to the DOTH biosynthesis cluster in *D. septosporum*, except that the *Pa-moxA* and *Ds-moxA* gene which are each adjacent to *epoA* have opposite gene directions in the cluster. Although the sequenced region containing the genes had high predicted amino acid identities to Ds DOTH, the analysis is not complete yet, new clones that cover the end of the 11.5 kb fragment need to be identified.

To further investigate the DOTH biosynthesis genes in *P. arachidicola*, a size fragmented *Sca*I genomic library could be made to pull out more positive clones containing the rest of the DOTH biosynthesis genes region. Gene replacement and complementation experiments should be carried out to confirm whether the *Pa*-genes identified in this study are involved in DOTH biosynthesis.

3.6 Genes identified in plasmids pR283 (*P. arachidicola vbsA* cluster)

3.6.1 Background

One of the mini clusters involved in DOTH biosynthesis in *D. septosporum* contains the genes *Ds-vbsA* and *Ds-hexA*. *Ds-vbsA*, which encodes a versicolorin B synthase is a homologue to *stcN* in *A. nidulans* and *aflK* in *A. parasiticus*. *Ds-hexA*, which encodes a fatty acid synthase, is a homologue to *stcJ* in *A. nidulans* and *aflA* in *A. parasiticus*. Gene replacement and complementation experiments showed that *Ds-vbsA* is required for DOTH biosynthesis in *D. septosporum* (Zhang et al., 2007). Three other genes, which have no homologues in AF and ST biosynthesis clusters, were also identified in the mini cluster. *Ds11* and *Ds12* are between *Ds-vbsA* and *Ds-hexA*. *Ds14*, which encodes a potassium channel, is downstream of *Ds-vbsA* (Zhang et al., 2007). Genes with high amino acid identity to *Ds-vbsA* mini cluster genes have been identified in *P. arachidicola* using *D. septosporum* DOTH genes as probes (Section 3.4). Those genes are analysed in this section.

3.6.2 Results

3.6.2.1 Primary sequence analysis of plasmids pR283

The overall primer walking (Section 2.6.1) sequencing strategy of plasmid pR283 is shown in figure 3.27. A total contiguous sequence of 5.7 kb was obtained. BLASTX results suggested four ORFs clustered within the 5.7 kb fragment (Table 3.7). Two predicted genes *Pa-vbsA* and *Pa-hexA* showed high amino acid identity to DOTH biosynthesis gene products from *D. septosporum* and some amino acid identity to AF biosynthesis and ST biosynthesis gene products from *Aspergillus* species. *Pa-vbsA* was predicted to encode a versicolorin B synthase. *Pa-hexA* was predicted to encode a fatty acid synthase. The other two ORFs *Pa11* and *Pa12* showed high predicted amino acid identity to those of non-DOTH genes *Ds11* and *Ds12* from *D. septosporum* (Figure 3.27). Thus it seem that in *P. arachidicola*, as in *D. septosporum*, two putative DOTH genes are separated by unrelated genes.

Figure 3.27 Schematic diagram of the 5.7 kb *P. arachidicola* contig showing four ORFs and positions of primers used for sequencing

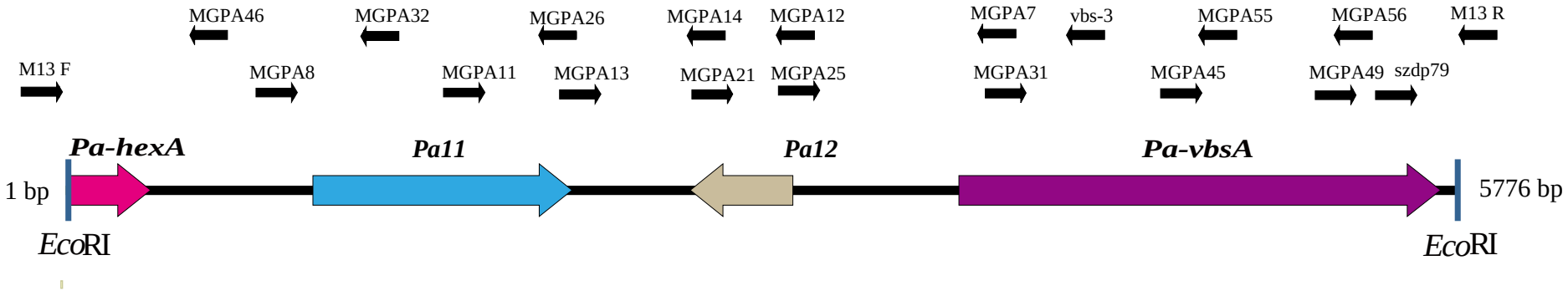


Figure 3.27 The black arrows indicate the primers used for sequencing the plasmid pR283. There are four ORFs detected in this 5.7 kb fragment. Large colour arrows indicate the direction of transcription of each ORF.

Table 3.7 Main BLASTX results of 5.7 kb fragment

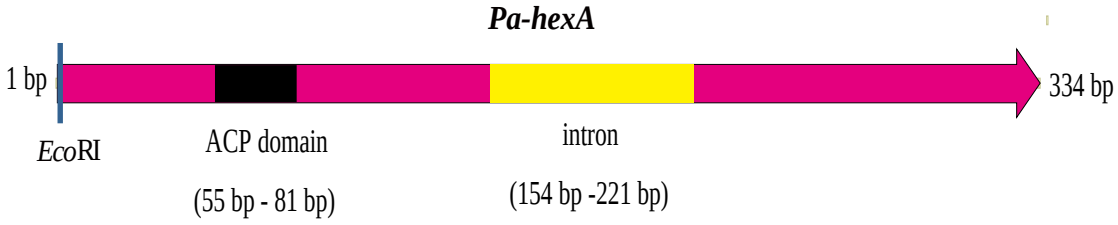
P. arachidicola gene position (5.7 kb fragment)	Aligned protein (BLASTX)	Score	E value	Genbank accession number
1 bp to 344 bp (*Pa-hexA*)	*Ds-hexA* fatty acid synthase *Mycosphaerella pini*	92	2e-30	ABU23831.1
	hexA Aspergillus nomius	73.2	5e-18	AAS90049.1
1020 bp - 2096 bp *(Pa11*)	*Ds11* hypothetical protein *Mycosphaerella pini*	483	1e-133	ABU23832.1
2583 bp- 3008 bp (*Pa12*)	*Ds12* hypothetical protein *Mycosphaerella pini*	286	2e-74	ABU23833.1
3704 bp -5698 bp (*Pa-vbsA*)	*Ds-vbsA* versicolorin B synthase *Mycosphaerella pini*	1189	0	ABO72541.1
	vbsA versicolorin B synthase *Aspergillus parasiticus*	911	0	AAC49318.1

Note: *Dothistroma septosporum* (teleomorph: *Mycosphaerella pini*)

3.6.2.2 The *P. arachidicola Pa-hexA* gene

The 5.7 kb fragment only contained part of predicted a *Pa-hexA* gene, with an ORF of 344 bp and encoding a polypeptide of 92 amino acids. BLASTX results showed a match to the C-terminus (236 - 325 aa) of Ds-HexA. The partial amino acid sequence of Pa-HexA had 89.5%, 59.6% and 63.6% amino acid identity to the corresponding regions of Ds-HexA in *D. septosporum*, StcJ in *A. nidulans* and AflA in *A. parasiticus* respctively. One intron of 68 bp was identified in partial *Pa-hexA* (as in section 3.5.2.2) and had similar position and length to the intron in *Ds-hexA* (Figure 3.28). *Pa-hexA* was predicted to encode a fatty acid synthase. One ACP domain was identified in the C-terminal by alignment of Pa-HexA with other HexA proteins (Figure 3.29). The partial *Pa-hexA* gene sequence and alignments are shown in appendices XIV and XV

Figure 3.28 Schematic diagram of *Pa-hexA*

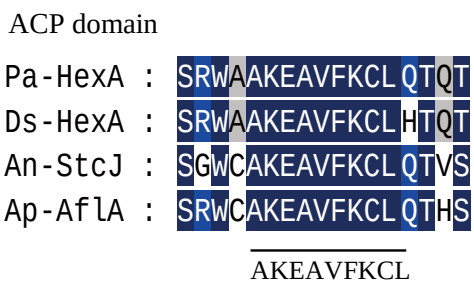


Gene	ORF length (bp)	Intron
Pa-hexA	344 (Partial)	154 bp - 221 bp (68 bp)
Ds-hexA	1053	865 bp - 937 bp (73 bp)

Note: The 344 bp of partial *Pa-hexA* corresponded to the 710 to 1053 bp of *Ds-hexA* ORF sequence.

Figure 3.28 The upper diagram shows the *Pa-hexA* ORF. The yellow box indicates the position of intron. The black box indicates the ACP domain. The position and length of introns in *Pa-hexA* and *Ds-hexA* are listed in the table below.

Figure 3.29 Alignments of Pa-HexA ACP domain with other ACP domains



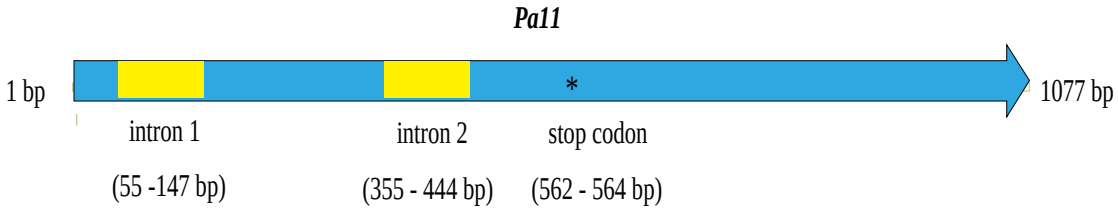
Name	Organism	Protein	Genbank No.
Pa-HexA	*P. arachidicola*	HexA	
Ds-HexA	*D. septosporum*	HexA	ABU23831.1
An-StcJ	*A. nidulans*	StcJ	AAC49198.1
Ap-AflA	*A. parasiticus*	AflA	AAS66002.1

Figure 3.29 One ACP domain was identified in Pa-HexA. Amino acid identities are shaded. The core conserved catalytic region is shown under the alignment.

3.6.2.3 The *P. arachidicola Pa11* gene

Downstream of *Pa-hexA*, an ORF of 1077 bp, which encoded a polypeptide of 298 amino acids was identified. BLASTX results showed 86.2% amino acid identity to *Ds11* in *D. septosporum* and 29.7% amino acid identity to a hypothetical protein in *A. nidulans*. This ORF was predicted to encode a hypothetical protein PA11. Two introns of 93 bp and 90 bp were identified in *Pa11*. The position and length of the first intron in *Pa11* was similar to the first intron in Ds11. The position of second intron in *Pa11* was similar to the second intron in Ds11, but the length was 11 bp longer. There is a stop codon in the coding region of *Pa11*. The original sequencing results were checked, signals were good at the position of the stop codon region (Figure 3.30). The full *Pa11* gene sequence and alignments are shown in appendices XVI and XVII

Figure 3.30 Schematic diagram of *Pa11*



Gene	ORF length (bp)	Intron 1	Intron 2
Pa11	1077	55 bp - 147 bp (93 bp)	355 bp -444 bp (90 bp)
Ds11	1070	55 bp - 148 bp (94 bp)	356 bp - 434 bp (79 bp)

Figure 3.30 The diagram shows the *Pa11* ORF. The yellow boxes indicate the positions of two introns. The star indicates the position of stop codon in the coding region of *Pa11*. The positions and length of introns in *Pa11* and *Ds11* are listed in the table below.

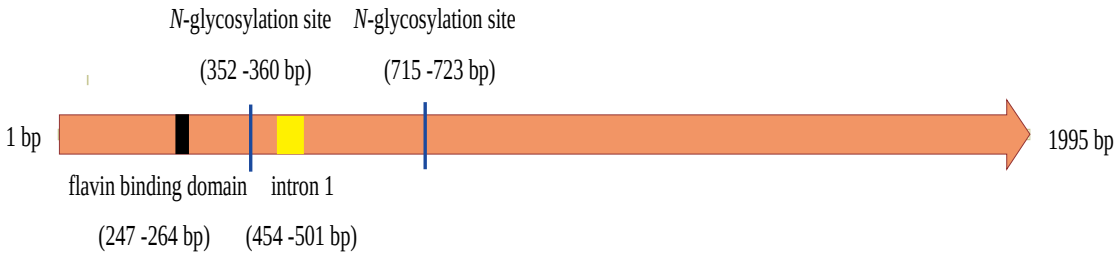
3.6.2.4 The *P. arachidicola Pa12* gene

The 5.7 kb fragment contained *Pa12*, which had an ORF of 426 bp, predicted to encode a polypeptide of 142 amino acids. The amino acid sequence of Pa12 was short compared to *Ds12* (266 amino acids). Pa12 had 87.3% amino acid identity to Ds12 (amino acids 125 to 266) in *D. septosporum* and amino acid identity of 32.9% and 42.8% to hypothetical proteins in *A. fumigatus* and *A. oryzae*. Pa12 is predicted to encode a hypothetical protein. No intron was found in *Pa12*. The full *Pa12* gene sequence and alignments are shown in appendices XVIII and XIX

3.6.2.5 The *P. arachidicola Pa-vbsA* gene

The last ORF found in the 5.7 kb fragment was *Pa-vbsA*, which was 1995 bp long and predicted to encode a polypeptide of 649 amino acids. The Pa-VbsA had 92.1% amino acids identity to Ds-VbsA, 69.6% to AflK and 66.4% to StcN. One intron of 48 bp was found in *Pa-vbsA* with similar position and length to the intron in *Ds-vbsA* (figure 3.31). Pa-VbsA was predicted to encode a versicolorin B synthase. Alignment of Pa-VbsA with other VbsA proteins identified a characteristic flavin binding domain and two potential sites of *N*-glycosylation in Pa-VbsA (Figure 3.32). The full *Pa-vbsA* gene sequence and alignments are shown in appendices XX and XXI.

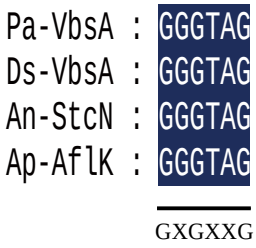
Figure 3.31 Schematic diagram of *Pa-vbsA*



Gene	ORF length	Intron 1
Pa-vbsA	2003 bp	454 bp-501 bp (48 bp)
Ds-vbsA	1991 bp	451 bp- 497 bp (47 bp)

Figure 3.31 The upper diagram shows the *Pa-vbsA* ORF. The yellow boxes indicate the positions of introns. The black box indicates the flavin binding domain. The blue vertical lines indicate the positions of *N*-glycosylation sites. The position and length of introns in *Pa-vbsA* and *Ds-vbsA* were listed in the table below.

Figure 3.32 Alignments of Pa-VbsA flavin binding domain with other flavin binding domains



Name	organism	Protein	Genbank No.
Pa-VbsA	*P. arachidicola*	VbsA	
Ds-VbsA	*D. septosporum*	VbsA	ABO72541.2
Ap-AflK	*A. parasiticus*	AflK	AAS66012.1
An-StcN	*A. nidulans*	StcN	EAA61600

Figure 3.32 One flavin binding domain was identified in Pa-VbsA. Amino acid identities are shaded. Core conserved catalytic region is shown under alignment.

3.6.2.6 The dothistromin biosynthesis genes in *P. arachidicola* (*vbsA* cluster)

The DOTH biosynthesis genes *Pa-hexA* and *Pa-vbsA* identified in the 5.7 kb fragment from *P. arachidicola* were separated by non-DOTH biosynthesis genes *Pa11* and *Pa12*. The organization and direction of those genes in the 5.7 kb fragment from *P. arachidicola* matched the corresponding region of the DOTH biosynthesis cluster in *D. septosporum* (Figure 3.33).

Figure 3.33 Comparison of *D. septosporum* and *P. arachidicola* dothistromin *vbsA* biosynthesis gene clusters

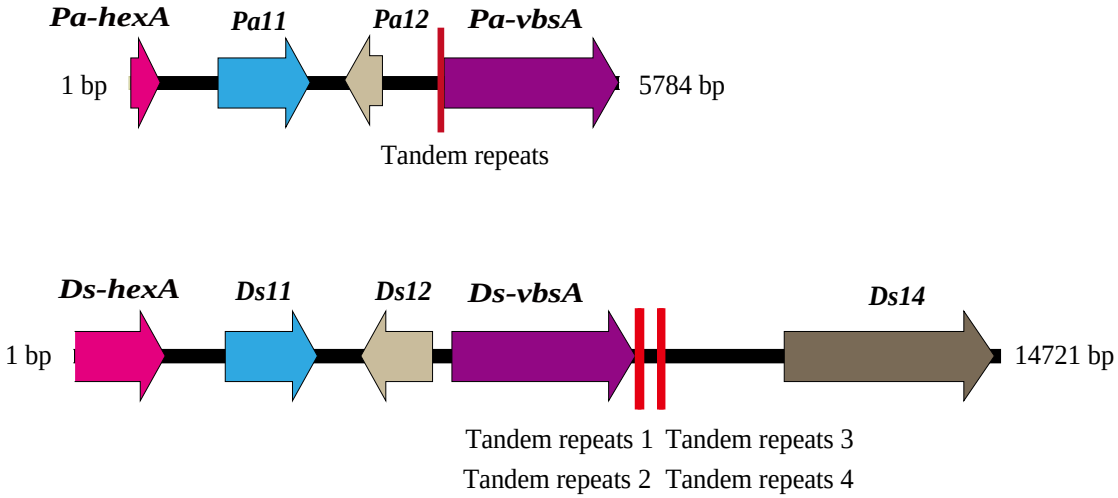


Figure 3.33 schematic diagrams of partial *D. septosporum* and *P. arachidicola* DOTH biosynthesis gene clusters. The order and orientation of each gene is presented with coloured arrows. The red vertical lines show the positions of tandem repeats.

The intergenic regions in the 5.7 kb fragment in *P. arachidicola* and the corresponding region in *D. septosporum* were compared. The intergenic regions of *hexA* and *Pa11* between *P. arachidicola* and *D. septosporum* had high nucleotide identity of 63.5%. The intergenic regions between *11* and *12* were different in size: 900 bp in *P. arachidicola* but only 503 bp in *D. septosporum* and were dissimilar in sequence (Table 3.8). Tandem repeats sequence (CCCTCTCATCC)n were found in the intergenic region between *Pa12* and *Pa-vbsA*, in contrast to *D. septosporum*, where four tandem repeats were found between the intergenic region of *Ds-vbsA* and *Ds14* (Table 3.9). The region downstream of *Pa-vbsA* has not yet been identified and sequenced.

Table 3.8 Intergenic regions of *P. arachidicola* and *D. septosporum* dothistromin genes

Region	Position	Length	GC content	Nucleotide identity
Pa-hexA to *Pa11*	227 bp - 761 bp	675 bp	55%	64%
Ds-hexA to *Ds11*	29485 bp - 30013 bp	695 bp	53%	
Pa11 to *Pa12*	1530 bp - 2074 bp	901 bp	55%	28.8%
Ds11 to *Ds12*	31329 bp - 33418 bp	504 bp	48%	
Pa12 to *Pa-vbsA*	2936 bp - 3690 bp	266 bp	56%	53.2%
Ds12 to *Ds-vbsA*	32475 bp - 33418 bp	322 bp	57%	

Table 3.9 Tandem repeats

Tandem repeat	Location	Copy No.	Consensus pattern	Note
Pa	3651 -3683 bp	2.9	CCCTCTCATCC	
Ds1	6505 -6554 bp	2.8	GTAGCCCTCGTGGAGAGCC	overlapping
Ds2	6547 -6644 bp	5.4	TGGTGGAGCACCGCCCTT	
Ds3	6842 -6949 bp	3	GTAATCGTCACGGTCGTAGTCTGAACGTCGCTCCAT	overlapping
Ds4	6867 -6974 bp	3	ACGTCGCTCCAAGTGACGGTCTTGGTGGTAGTCTGG	

Ds repeats sequences (Zhang et al., 2007)

3.6.3 Discussion

The partial *Pa-hexA* identified in the 5.7 kb fragment showed high amino acid identity to *D. septosporum Ds-hexA, A. parasiticus aflA* and *A. nidulans stcJ* genes which encoded fatty acid synthase. In *A. nidulans*, two fatty acid synthase genes *stcJ* and *stcK* together with a polyketide synthase *stcA* (Section 3.5.2.2 and 3.5.3) are required for the early steps in ST biosynthesis. Disruption of *stcJ* and *stcK* produced mutant strains that had similar morphology to wild-type strains but failed to produce ST. Feeding experiments showed that adding C6 straight-chain fatty acid hexanoic acid to *stcJ* and *stcK* mutant strains restored the ST production, but the ST level was approximately 20 fold less than the wild type strain. Addition of hexanoic acid to *stcA* mutant strains failed to produce ST (Brown et al., 1996a). In *A. parasiticus*, two fatty acid synthase *aflA* and *aflB* were required to mediate transfer of the synthetic C6 primer to PKS to initiate the AF biosynthesis. The protein products AflA, AflB and AflC needed to be physically associated for this transformation (Watanabe et al., 1996). *Pa-hexA*, which also encodes a fatty acid synthase in *P. arachidicola* may work together with *Pa-pksA* in the biosynthesis of NOR, a precursor for DOTH.

Pa11 and Pa12 had high amino acid identity to Ds11 and Ds12 in *D. septosporum*. *Ds11* and *Ds12* had no similarities to any AF or ST biosynthesis genes in *Aspergillus* species. The functions of Ds11 and Ds12 are not known, it was predicted that they are not involved in biosynthesis of DOTH (Zhang et al., 2007). A premature stop codon found in the coding region of *Pa11* also suggests no involvement in DOTH biosynthesis in *P. arachidicola*.

Pa-vbsA identified in the 5.7 kb fragment showed high amino acid identity to *D. septosporum Ds-vbsA, A. parasiticus aflK* and *A. nidulans stcN* genes which encode versicolorin B synthase. In *D. septosporum*, *Ds-vbsA* was expressed at an early growth stage (Schwelm et al., 2008), a *Ds-vbsA* replacement mutant strain produced at least 20-fold less DOTH than the wild type, and confirmed that *Ds-vbsA* was required for DOTH biosynthesis (Zhang et al., 2007). In *A. parasiticus*, *aflK* is

involved in conversion of VHA to VERB (McGuire et al., 1996). The *aflK* protein product acts as a dimer, and has three *N*-glycosylation sites. Silva and Townsend suggested that glycosidation of the protein could be important for dimer formation and catalytic activity. AflK also has some amino acid identity to flavin-dependent oxidases and dehydrogenases. It is suggested that AflK may have an oxidative activity in AF biosynthesis (Silva et al., 1996; Silva and Townsend, 1996). Pa-VbsA had a high amino acid identity to Ds-VbsA, it may also be involved in converting VHA to VERB in DOTH biosynthesis in *P. arachidicola*. The identification of two N-glycosylation sites and the flavin binding domain in Pa-VbsA suggest that it may act as a dimer and may harbour an oxidative activity in DOTH biosynthesis.

The genes identified in the 5.7 kb fragment showed similar gene organization to one of the DOTH biosynthesis mini-clusters in *D. septosporum*. Two possible DOTH biosynthesis genes *Pa-hexA* and *Pa-vbsA* were separated by two non-DOTH biosynthesis genes *Pa11* and *Pa12*. The *Pa-hexA – Pa11* intergenic region had higher similarity in length and sequence to its corresponding *D. septosporum* region than the other two other intergenic regions had. In *D. septosporum* a Ds14, which encoded a potassium channel was downstream of *Ds-vbsA*. It is predicted that there may be a similar gene located downstream of the *Pa-vbsA*.

3.7 Genes identified in plasmids pR286 (*P. arachidicola dotA* cluster)

3.7.1 Background

In *D. septosporum*, a *Ds-dotA* gene, which encodes a ketoreductase, had been confirmed to be involved in DOTH biosynthesis by gene replacement. Another three genes *Ds-dotB*, *Ds-dotC* and *Ds-dotD* are clustered alongside *Ds-dotA*: *Ds-dotB*, which encodes an oxidase, *Ds-dotC*, which encodes a toxin pump protein, and *Ds-dotD*, which encodes a thioesterase (Bradshaw et al., 2002; Zhang et al., 2007).

Genes with high amino acid identity to *Ds-dotA* mini cluster genes were identified in *P. arachidicola* using *D. septosporum* DOTH genes as probes (Section 3.4). Those genes are analysed in this section.

3.7.2 Results

3.7.2.1 Primary sequence analysis of plasmid pR286

Plasmid pR286 was sequenced by primer walking, and the sequences assembled as described previously (Section 3.5.2.1). The estimated length of plasmid pR286 was 14 kb. Due to time limitations, only 9.3 kb of this plasmid was sequenced, of which only single strand sequence of the last 2.4 kb was obtained (Figure 3.34).

Four predicted genes *Pa-dotA*, *Pa-dotB*, *Pa-dotC* and *Pa-dotD* showed high predicted amino acid identity to DOTH biosynthesis gene products from *D. septosporum* and some amino acid identity to AF biosynthesis and ST biosynthesis gene products (Table 3.10). *Pa-dotA* was predicted to encode a ketoreductase, *Pa-dotB* an oxidase, *Pa-dotC* a toxin pump and *Pa-dotD* a thioesterase. A fifth ORF showed high amino acid identity to major facilitator superfamily (MFS) transporters in *A. fumigatus* and *A. clavatus* and has been named *Pa-mfs* (Figure 3.34).

Figure 3.34 Schematic diagram of the 9.3 kb *P. arachidicola* contig showing four ORFs and positions of primers used for sequencing

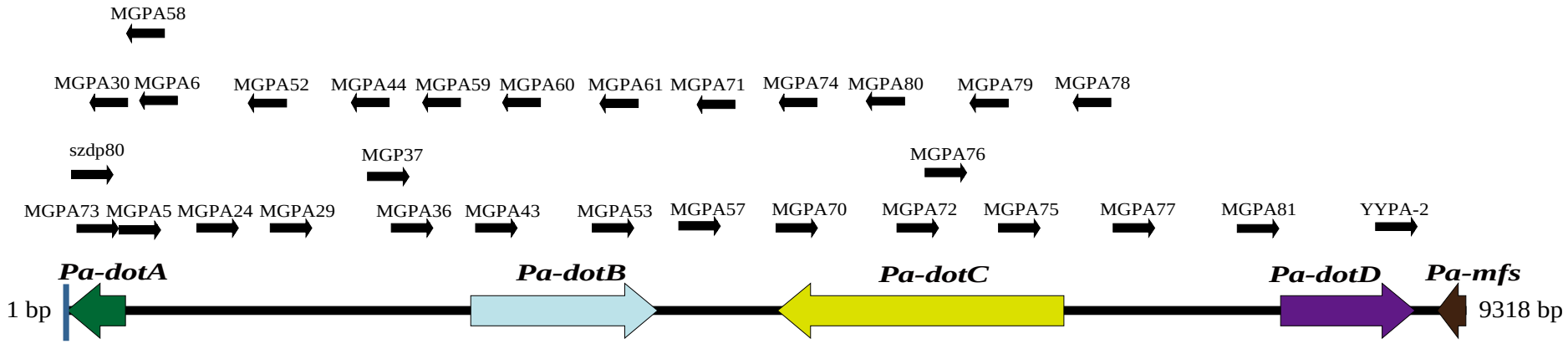


Figure 3.34 The black arrows indicate the primers used for sequencing the plasmid pR286. Large colour arrows indicate the direction of transcription of each ORF.

Table 3.10 Main BLASTX results of 5.7 kb fragment (best matches only)

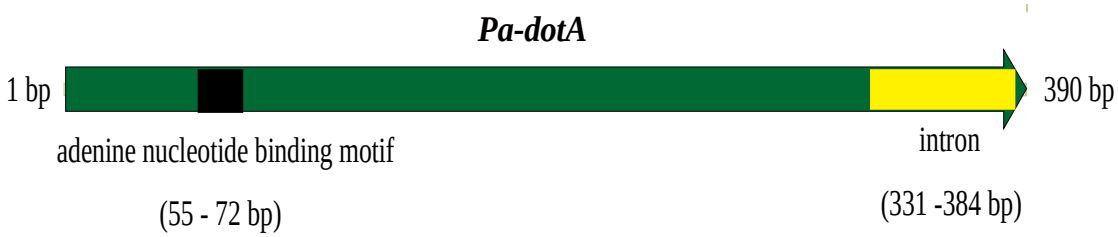
P. arachidicola gene position (9.3 kb fragment)	Aligned protein (BLASTX)	Score	E value	Genbank accession number
1 bp to 390 bp (*Pa-dotA*)	*Ds-dotA* ketoreductase *Mycosphaerella pini*	184	1e-43	AAL87045
2694 bp to 3935 bp (*Pa-dotB*)	*Ds-dotB* putative oxidase *Mycosphaerella pini*	737	0	AAL87046
4722 bp to 6627 bp (*Pa-dotC*)	*Ds-dotC* putative DOTH transporter *Mycosphaerella pini*	398	0	AAL87047
8086 bp to 8982 bp (*Pa-dotD*)	*Ds-dotD* putative thioesterase *Mycosphaerella pini*	487	7e-135	AAL87048
9129 bp to 9318 bp (*Pa-mfs*)	MFS transporter *Aspergillus fumigatus*	73.6	5e-12	XP749930
	MFS transporter, putative *Aspergillus clavatus*	72	1e-11	XP001269961

Note: *Dothistroma septosporum* (teleomorph: *Mycosphaerella pini*)

3.7.2.2 The *P. arachidicola Pa-dotA* gene

The 9.3 kb fragment only contained part of the predicted *Pa-dotA* gene, with an ORF of 390 bp encoding a polypeptide of 112 amino acids. The amino acid sequence of partial Pa-DotA had 96.4%, 68.1% and 67.2% amino acid identity to the corresponding regions of Ds-DotA in *D. septosporum*, StcU in *A. nidulans* and AflM in *A. parasiticus*. Because of the high amino acid identity between Pa-DotA and Ds-DotA, the Pa-DotA was predicted to encode a ketoreductase with a full polypeptide of 263 amino acids. One intron of 54 bp was identified, with the same position as the first intron in *Ds-dotA*, but 12 bp shorter (Figure 3.35). One adenine nucleotide binding motif was identified in Pa-DotA (Figure 3.36), the position was conserved between Pa-DotA and Ds-DotA. The full *Pa-dotA* gene sequence and alignments are shown in appendices XXII and XXIII.

Figure 3.35 Schematic diagram of *Pa-dotA*



Gene	ORF length (bp)	Intron 1	Intron 2
Pa-dotA	390 (Partial)	331 bp - 384 bp (54 bp)	
Ds-dotA	916	331 bp - 396 bp (66 bp)	586 bp - 643 bp (58 bp)

Figure 3.35 The upper diagram shows the *Pa-dotA* ORF. The yellow box indicates the position of intron, the position and length of introns in *Pa-dotA* and *Ds-dotA* were listed in the table below. The black box indicates the adenine nucleotide binding motif.

Figure 3.36 Alignments of Pa-DotA adenine binding motif with other adenine binding motifs

Adenine nucleotide binding motif

Name	Organism Protein	Protein	Genbank No.
Pa-DotA	*P. arachidicola*	Pa-DotA	
Ds-DotA	*D. septosporum*	Ds-DotA	AAL87045
An-StcU	*A.nidulans*	StcU	AAC49205.1
Ap-AflM	*A. parasiticus*	AflM	AAS66014.1

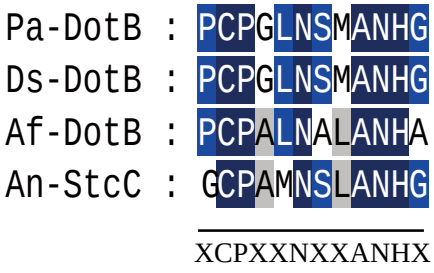
Figure 3.36 One adenine nucleotide binding motif was identified in Pa-DotA. Amino acid identities are shaded. The core conserved catalytic region is shown under the alignment.

3.7.2.3 The *P. arachidicola Pa-dotB* gene

The *Pa-dotB* had an ORF of 1242 bp and encoded an oxidase of 414 amino acids. The Pa-DotB had 84.3% amino acid identity to Ds-DotB in *D. septosporum* and 32.3% identity to an oxidase in *A. fumigatus* but only 15.6 % identity to StcC in *A. nidulans*. No introns are present in *Pa-dotB* or *Ds-dotB.* A putative heme-binding domain was identified in Pa-DotB (Figure 3.37). The full *Pa-dotB* gene sequence and alignments are shown in appendices XXIV and XXV.

Figure 3.37 Alignments of Pa-DotB heme-binding domain with other oxidase proteins

Heme-binding domain



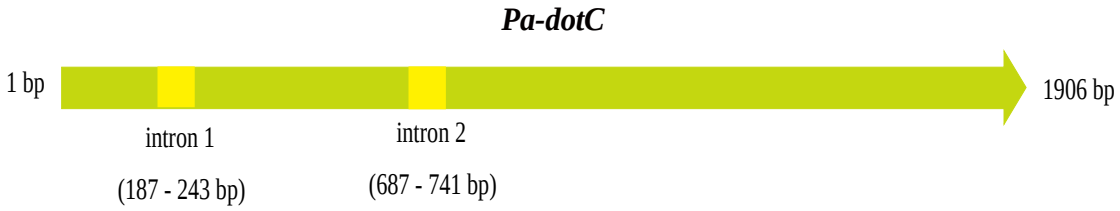
Name	Organism Protein	Protein	Genbank No.
Pa-DotB	*P. arachidicola*	Pa-DotB	
Ds-DotB	*D. septosporum*	Ds-DotB	AAL87046
Af-DotB	*A.fumigatus*	Af-DotB	EDP50907
An-StcC	*A.nidulans*	StcC	AAC49193.1

Figure 3.37 One heme-binding domain was identified in Pa-DotB. Amino acid identities are shaded. Core conserved catalytic regions shown under the alignment.

3.7.2.4 The *P. arachidicola Pa-dotC* gene

Pa-dotC had an ORF of 1906 bp and encoded a polypeptide of 598 amino acids. Pa-DotC had 83.7% amino acid identity to Ds-DotC but only 15.8% amino acid identity to another MFS transporter Ds28 in *D. septosporum*. It also had a low amino acid identity of 25.8% to AflT in *A. parasiticus*. Two introns of 57 bp and 55 bp were found in *Pa-dotC*, in similar positions to introns in Ds-dotC, but the first intron in *Pa-dotC* was 19 bp longer than that in *Ds-dotC* (Figure 3.38). The full *Pa-dotB* gene sequence and alignments are shown in appendices XXVI and XXVII.

Figure 3.38 Schematic diagram of *Pa-dotC*



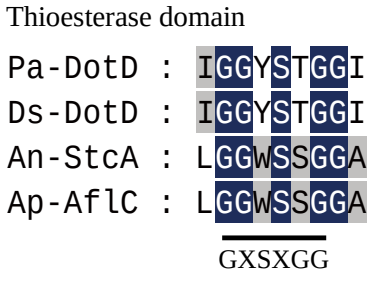
Gene	ORF length (bp)	Intron 1	Intron 2
Pa-dotC	1906	187 bp - 243 bp (57 bp)	687 bp - 741 bp (55bp)
Ds-dotC	1904	184 bp - 221 bp (38 bp)	680 bp - 736 bp (57 bp)

Figure 3.38 The upper diagram shows the *Pa-dotC* ORF. The yellow boxes indicate the positions of introns, the position and length of introns in *Pa-dotC* and *Ds-dotC* are listed in the table below.

3.7.2.5 The *P. arachidicola Pa-dotD* gene

The *Pa-dotD* had an ORF of 981 bp and encoded a polypeptide of 327 amino acids. Pa-DotD had 74.6% amino acid identity to the thioesterase Ds-DotD and 29.9% and 28.1% amino acid identity to the corresponding C-terminal TE domain of the polyketide synthase AflC and StcA in *Aspergillus* species. A conserved active site region of the thioesterase domain was identified (Figure 3.39). No intron was found in *Pa-dotD*. The full *Pa-dotD* gene sequence and alignments are shown in appendices XXVIII and XXIX.

Figure 3.39 Alignment of Pa-DotD with other thioesterase domains



Name	organism	Protein	Genbank No.
Pa-DotD	*P. arachidicola*	Pa-DotD	
Ds-DotD	*D. septosporum*	Pa-DotD	AAL87048.1
An-StcA	*A. nidulans*	StcA	AAA81586
Ap-AflC	*A. parasiticus*	AflC	AAS66004

Figure 3.39 One TE domain was identified in Pa-DotD. Amino acid identities are shaded. Core conserved catalytic regions references are shown under each alignment with the active amino acid underlined.

3.7.2.6 The *P. arachidicola* Partial *Pa-mfs* gene

Pa-mfs was downstream of *Pa-dotD*, it had an partial ORF of 190 bp and encoded a polypeptide of 63 amino acids. The Pa-Mfs had 54% and 52% amino acid identity to the corresponding region of MFS transporters in *A. fumigatus* and *A. clavatus*.

3.7.2.7 Arrangement of dothistromin biosynthesis genes in *P. arachidicola* (*dotA* cluster)

The arrangement of predicted DOTH biosynthesis genes identified in the 9.3 kb fragment from *P. arachidicola* was compared to the corresponding region of the DOTH biosynthesis cluster in *D. septosporum*. Four of the *P. arachidicola* genes *Pa-dotA*, *Pa-dotB*, *Pa-dotC* and *Pa-dotD* had the same orientations as the homologous *D. septosporum* DOTH biosynthesis genes. Downstream of *Pa-dotD*, a *Pa-mfs* predicted to encode a MFS transporter protein was identified in *P. arachidicola*. No homologous gene is present in the corresponding region in *D. septosporum* (Figure 3.40).

Figure 3.40 Comparison of *D. septosporum* and *P. arachidicola* dothistromin *dotA* biosynthesis gene cluster

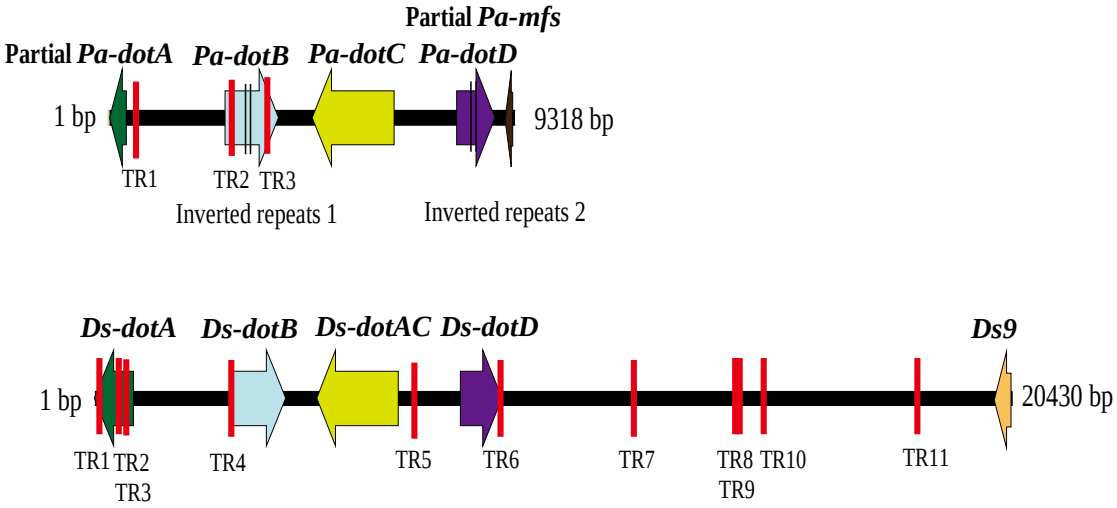


Figure 3.40 Schematic diagrams of partial *D. septosporum* and *P. arachidicola* DOTH biosynthesis gene clusters. The order and orientation of each gene is presented with coloured arrows. Black vertical lines show positions of inverted repeats. The red vertical lines show the positions of tandem repeats. The sequences and positions of repeats are shown in the tables 3.12 and 3.13.

Intergenic regions between the DOTH biosynthesis genes in *P. arachidicola* and *D. septosporum* were compared (Table 3.11). The intergenic regions of *dotA-B, dotB-C* and *dotC -D* each had similar length and nucleotide sequence identity of 45% to 59% between the two species. Because the presence of *Pa-mfs* in *P. arachidicola*, the intergenic region downstream of *Pa-dotD* was significantly different from *D. septosporum*, where there is a 10 kb region between Ds-dotD and Ds9 with no gene present. The *P. arachidicola* cluster had two inverted repeat sequences within the coding region of *Pa-dotB* and *Pa-dotD* that were not seen in *D. septosporum* (Figure 3.40 and Table 3.12). Three tandem repeat sequences were identified between Pa-*dotA* and *Pa-dotB* and within the ORF of *Pa-dotB*. The positions and sequences of those tandem repeats were different to the *D. septosporum* cluster tandem repeats (Figure 3.40 and Table 3.13).

Table 3.11 Intergenic regions of *P. arachidicola* and *D. septosporum* dothistromin genes

Region	Position	Length	GC content	Nucleotide identity
Pa-dotA to *Pa-dotB*	391 bp - 2693 bp	2303 bp	52%	59%
Ds-dotA to *Ds-dotB*	13057 bp - 15361 bp	2035 bp	53%	
Pa-dotB to *Pa-dotC*	3936 bp - 4721 bp	786 bp	47%	48.5%
Ds-dotB to *Ds-dotC*	16607 bp - 17326 bp	720 bp	50%	
Pa-dotC to *Pa-dotD*	6628 bp - 8085 bp	1458 bp	52%	47%
Ds-dotC to *Ds-dotD*	19231 bp - 20681 bp	1378 bp	53%	
Pa-dotD to Pa-mfs	8983 bp - 9128 bp	146 bp	51%	0.8%
Ds-dotD to Ds9	21651 bp - 32345 bp	10695 bp	53%	

Table 3.12 Inverted repeats

Inverted repeat	Location	Repeat length	Sequence
Pa inverted repeat 1	3230-3240 3317-3307	11	5’gaacctcttca 3’ | | | | | | | | | | | 3’ cttggagaagt 5’
Pa inverted repeat 2	8492-8502 8570-8560	11	5’ aagcgctggat 3’ | | | | | | | | | | | 3’ ttcgcgaccta 5’

Table 3.13 Tandem repeats

Tandem repeat	Location	Copy No.	Consensus pattern
Pa TR1	723-738 bp	2.3	TCGGGAT
Pa TR2	2784-2801 bp	2	TATCAGAAG
Pa TR3	3596-3613 bp	2	GTATTCCAA
Ds TR1	12493-12545 bp	2.2	GATGGCGCCCTTGGAAACACACTA
Ds TR2	12805-12852 bp	2.5	CTCACCGACCTGGTCAAT
Ds TR3	12856-12896 bp	2.7	GGCCTTGATGGCGAT
Ds TR4	15285-15324 bp	2.2	AGCTTCACAGAATCCGCAA
Ds TR5	19785-19836 bp	2	ACGAGCTCAGAGGAACCCATCTGGC
Ds TR6	21517-21583 bp	2.1	GGAGATAGTTGTCGAGATTGAGGA
Ds TR7	24681-24730 bp	2.3	TACGAGAGCTTTCTTCTCGT
Ds TR8	26904-26961 bp	3.3	TGGTCGGCTGTTGATCT
Ds TR9	26779-27030 bp	2	GGTCGGCTGCTGACCACCACCGCCACCATAGGTCTCGGTCGGCTGCTCATATTCGTCATCTGTGCA CGTAGACGATAACGTGGTGGATGTAGTAATAGATGAAGACCATGTTGAAGAAGCTAAAGT
Ds TR10	27377-27412 bp	2	CTGGTTACATTACGGTGA
Ds TR11	30533-30608 bp	2.6	GCCACGGTGTGCAGAAGACCACACTCATA

Ds repeat sequences (Zhang et al., 2007)

3.7.3 Discussion

The partial *Pa-dotA* identified in the 9.3 kb fragment had a high amino acid identity to *D. septosporum Ds-dotA*, *A. parasiticus aflM* and *A. nidulans stcU* genes which encode a ketoreductase. A gene *ver-1* (renamed as *aflM*) was originally isolated by genetic complementation of an *A. parasiticus* strain CS10 that was unable to convert VERA to DMST. Transformation of a fragment containing *ver-1* to CS10 resulted in production of wild-type level AF, so *ver-1* was shown to be involved in conversion of VERA to DMST (Skory et al., 1992). By using the *ver-1* gene as heterologous probe, the *A. nidulans* gene *verA* (renamed as *stcU*) was isolated from a genomic library. Disruption of *stcU* confirmed that VerA was involved in conversion of VERA to DMST in ST biosynthesis (Keller et al., 1994).

Later Keller showed that another gene *stcS* in *A. nidulans*, homologous to *aflN* in *A. parasiticus*, encoded a monooxygenase also involved in conversion of VERA to DMST (Keller et al., 1995). Henry and Townsend (2005) showed that conversion of VERA to DMST involves two rounds of oxidation and one reduction step. First AflN disrupts the A-ring of VERA, followed by AflM mediated reductive step to yield a 5,8-dihydroxyanthraquinone product, AflN again carries out a Baeyer-Villiger cleavage to allow decarboxylation for DMST formation.

In *D. septosporum*, targeted gene replacement of *Ds-dotA* produced a mutant strain unable to produce DOTH that accumulated VERA (Bradshaw et al., 2002). Although only partial *Pa-dotA* was identified in *P. arachidicola*, the high amino acid identity of *Pa-dotA* to *Ds-dotA* and the identification of the adenine binding motif strongly suggested that *Pa-dotA* encodes a ketoreductase. Like its homologue, Pa-DotA is predicted to be involved in modification of ring structure in the VERA to DMST conversion step in DOTH biosynthesis.

Pa-dotB is predicted to encode an oxidase with high amino acid identity to *Ds-dotB* in *D. septosporum* and some amino acid identity to *stcC* in *A. nidulans*, but no

homologue was identified in *A. parasiticus*. The function of *stcC* in ST biosynthesis is not clear. It was suggested that *Ds-dotB* may be involved in VERB or VERA formation, which is the divergence point of AF and DOTH pathways (Bradshaw et al., 2002). *Pa-dotB* may also be involved in VERB or VERA formation in *P. arachidicola*.

The *Pa-dotC* gene was predicted to encode a MFS transporter with amino acid identity to *Ds-dotC* in *D. septosporum* and *aflT* in *A. parasiticus*. *Pa-mfs,* which was also predicted to encode a MFS transporter, had no homologue in *D. septosporum*. In *A. parasiticus* deletion of *aflT*, the MFS transporter, produced a mutant strain that had similar AF production and toxin secretion to the wild-type strain. So *aflT* was not required for AF production or toxin secretion, other transporters may be required or there may be some redundancy of function. Unlike other AF biosynthesis genes, *aflT* was not regulated by the AF pathway-specific activator AflR and the co-activator AflJ but instead by the Fad-dependent G-protein signalling pathway (Chang et al., 2004).

Studies by Feng at Massey University showed that although a *Ds-dotC* knockout strain had much lower levels of DOTH compared to the wild-type strain, it still had detectable DOTH secreted out of and accumulated in *D. septosporum*. Although *Ds-dotC* is not essential for DOTH biosynthesis, it may be involved in secretion of the toxin and reduced secretion in the mutant may have a feedback effect inhibiting DOTH biosynthesis (Feng, 2007). The secretion of DOTH out of the fungi could involve more than one MFS transporter, for example *Pa-dotC* and *Pa-mfs*. A possible AflR-like binding site was identified upstream of *Ds-dotC*, although no homologue of *aflR* has yet been identified in *D. septosporum*. Three possible AflR conserved binding sites (5'-TCGN$_5$CGR-3') were identified upstream of *Pa-dotC* (1594-1604 bp, 1681-1691 bp and 2232-2242 bp).

The *Pa-dotD* gene had high amino acid identity to *Ds-dotD*, a thioesterase in *D.*

septosporum. It also had some amino acid identity to the TE domain of AflC and StcA, which encode a polyketide synthase in *Aspergillus* species (Section 3.5.2.2). The TE domain of AflC and StcA is involved in releasing the polyketide product from the polyketide synthase complex. The function of *Ds-dotD* in DOTH biosynthesis is not clear (Bradshaw et al., 2002). The *Pa-dotD* may be involved in an early step in DOTH biosynthesis in *P. arachidicola*, working together with *Pa-pksA* in NOR formation. Alternatively the *dotD* genes might be fragments of an ancestral *pks* gene cleaved during genome evolution.

The four genes Pa-*dotA*, *Pa-dotB*, *Pa-dotC* and *Pa-dotD* showed similar gene organization to the *Ds-dotA – Ds-dotD* mini-cluster in *D. septosporum*. However in *D. septosporum*, there is a 10 kb gap downstream of *Ds-dotD* in which, no ORF was found. The presence of the fifth gene *Pa-mfs* just 146 bp downstream of *Pa-dotD* in *P. arachidicola* was thus not expected. Although only part of *Pa-mfs* was identified, the whole gene was predicted to encode a MFS transporter of 510 amino acids.

3.8 DOTH biosynthesis gene cluster in *P. arachidicola*

3.8.1 Background

The AF and ST biosynthesis genes are closely clustered in *Aspergillus* species (Brown et al., 1996b; Yu et al., 2004a). In contrast DOTH genes are in "fragmented clusters" in *D. septosporum.* The DOTH biosynthesis genes that have been identified so far are fragmented into three mini-clusters, located on a 1.3 Mb chromosome and many are separated by non-DOTH biosynthesis genes (Zhang et al., 2007).

This study suggests that the DOTH biosynthesis gene cluster in *P. arachidicola* is similar to that in *D. septosporum*. Comparative analysis of the DOTH biosynthesis gene cluster in *P. arachidicola* will be discussed in this section.

3.8.2 Results

3.8.2.1 DOTH biosynthesis gene cluster in *P. arachidicola*

Characterization of the genes identified in three *P. arachidicola* DOTH mini-clusters are summarized in Table 3.14. Most of the genes identified in *P. arachidicola* had higher amino acid identity to the homologous genes in *D. septosporum* than homologues in *Aspergillus* species, except for *Pa-mfs*, which had 14.3% amino acid identity to *aflT*, but 12.9% to *Ds-dotC*. The DOTH genes in *D. septosporum* are more similar to AF/ST genes than the DOTH genes in *P. arachidicola*. *D. septosporum* DOTH genes had 53.7% and 52% mean amino acid identity to AF and ST genes respectively, while *P. arachidicola* DOTH genes had 44.1% and 46.4% mean amino acid identity to AF and ST genes. The mean nucleotide identity of DOTH gene homologues (67.6%) was higher than the mean nucleotide identity of the corresponding gene gaps (48.8%) between *P. arachidicola* and *D. septosporum* (Figure 3.41). The GC content of each gene cluster in *P. arachidicola* is shown in figure 3.42. The GC content is about 40 – 60% for most of the regions but it is lower than 40% for the intergenic region of *Pa-epoA – Pa-avfA*, which had low nucleotide identity of 15.9% to the corresponding region of *D. septosporum* DOTH cluster. The

highest GC content was about 70% at the beginning of the *Pa-cypA* coding region, which had a high nucleotide identity of 82.3% to the corresponding region of *D. septosporum* DOTH cluster (Figure 3.42).

Although the sequences of flanking regions of each mini-cluster in *P. arachidicola* are not known, the DOTH genes identified in *P. arachidicola* had similar gene organization to that in *D. septosporum* but different from the AF or ST biosynthesis gene clusters in *Aspergillus* species (Figure 3.43). The DOTH biosynthesis gene clusters in *P. arachidicola* and *D. septosporum* had two main differences to each other. First there was a MFS transporter *Pa-mfs* identified downstream of *Pa-dotD* in mini-cluster 1 that was not seen in *D. septosporum*. The sequence of *Pa-dotD – Pa-mfs* (146 bp) had 54.3% nucleotide identity to a 161 bp region of the 10605 bp *Ds-dotD – Ds9* intergenic sequence that, intriguingly, was not following the *Ds-dotD* immediately, but was 1971 bp away from the end of *Ds-dotD*. (Figure 3.44). Second the adjacent *moxA* and *epoA* genes are divergently transcribed in *P. arachidicola* but transcribed in the same direction in *D. septosporum*. The DOTH biosynthesis pathway was predicted to be similar to that of *D. septosporum* (Figure 3.45).

Figure 3.44 Compare the intergenic region between *Pa-dotD-Pa-mfs* and *Ds-dotD –Ds9*

Figure 3.44 The order and orientation of each gene is presented with colour arrows. The black lines indicate the region of *Pa-dotD –Pa-mfs i*ntergenic region matched to *Ds-DotD-Ds9* intergenic region (1971-2137 bp).

3.8.2.2 Regulatory motifs of DOTH biosynthesis genes in *P. arachidicola*

Although no homologue of the AF regulatory genes *aflR* and *aflJ* were found in *P. arachidicola*, consensus AflR binding motifs (5'-TCGN$_5$CGR-3') were found upstream and in the coding region of several predicted DOTH biosynthesis genes. The positions of those AflR binding sites were compared between *P. arachidicola* and *D. septosporum* (Figure 3.46). In mini cluster 1, the AflR binding sites were distributed evenly with 2 AflR binding sites within 650 bp upstream of *Pa-dotA*, *Pa-dotB* and *Pa-dotD* in *P. arachidicola*. But in *D. septosporum* the AflR binding sites are located closer to *Ds-dotA* and *Ds-dotD* than to *Ds-dotB* and *Ds-dotC* (Figure 3.46) (Schwelm, 2007). The identification of AflR binding sites within 450 bp upstream of *Pa-dotC* and in the *Pa-dotC* coding region were not expected. In mini-cluster 2, AflR binding sites were identified upstream of both *Pa-vbsA* and *Ds-vbsA*, although the spacing was different in the two species. In mini cluster 3, *Pa-cypA* and *Pa-pksA* each have AflR binding sites in the gap between these two genes. The positions of the 3 AflR binding sites in *Pa-pksA* were similar to 3 out of the 6 AflR binding sites in the *Ds-cypA* - *Ds-pksA* gap. In both *D. septosporum* and *P. arachidicola*, AflR sites were absent from upstream regions of *avfA* homologues but present in the coding region and upstream of *epoA*. The positions of the *Pa-epoA* AflR binding sites were similar to those in *Ds-epoA*, but 4 AflR binding sites upstream of *Ds-epoA* were missing in *Pa-epoA* upstream region (Figure 3.46).

Table 3.14 Characterization of genes identified in *P. arachidicola* DOTH mini-clusters

Putative function (name)	*D. septosporum* DOTH cluster		*A. parasiticus* AF cluster			*A. nidulans* ST cluster		
	homologue	aa % identity to *Pa* homologue	homologue	aa % identity to *Pa* homologue	aa% identity to *Ds* homologue	homologue	aa % identity to *Pa* homologue	aa% identity to *Ds* homologue
Hydroxyversicolorone monooxygenase (*Pa-moxA*)	*Ds-moxA*	70.4%	*aflW*	10.8%	55.1%	*stcW*	25.3%	59%
Epoxide hydrolase (*Pa-epoA*)	*Ds-epoA*	22.5%	-	-		-	-	-
Oxidase (*Pa-avfA*)	*Ds-avfA*	73.3%	*aflI*	46.9%	47.8%	*stcO*	46.8%	43.7%
Averufin monooxygenase (*Pa-cypA*)	*Ds-cypA*	92.8%	*aflV*	60.2%	59.3%	*stcB*	59.9%	59.8%
Polyketide synthase (*Pa-pksA*)	*Ds-pksA*	88.7%	*aflC*	53%	54.8%	*stcA*	50%	57%
Fatty acid synthase (*Pa-hexA*)	*Ds-hexA*	89.5%	*aflA*	63.6%	48.8%	*stcJ*	59.6%	41.3%
Unknown (*Pa11*)	*Ds11*	86.2%	-	-	-	-	-	-
Unknown (*Pa12*)	*Ds12*	87.3%	-	-	-	-	-	-
Versicolorin B synthase (*Pa-vbsA*)	*Ds-vbsA*	92.1%	*aflK*	69.6%	72%	*stcN*	66.4%	69.1%
Ketoreductase (*Pa-dotA*)	*Ds-dotA*	96.4%	*aflM*	67.2%	80.2%	*stcU*	68.1%	79.1%
Oxidase (*Pa-dotB*)	*Ds-dotB*	84.3%	-	-	-	*stcC*	15.6%	24%
Toxin pump (*Pa-dotC*)	*Ds-dotC*	83.7%	*aflT*	25.8%	31.2%	-	-	-
Thioesterase (*Pa-dotD*)	*Ds-dotD*	74.6%	*aflC*	29.9%	34.8%	*stcA*	28.1%	37.9%
MFS transporter (*Pa-mfs*)	*Ds-dotC*	12.9%	*aflT*	14.3%	-	-	-	–

Figure 3.41 Amino acid and nucleotide identity of genes and gene gaps between *P. arachidicola* and *D. septosporum*

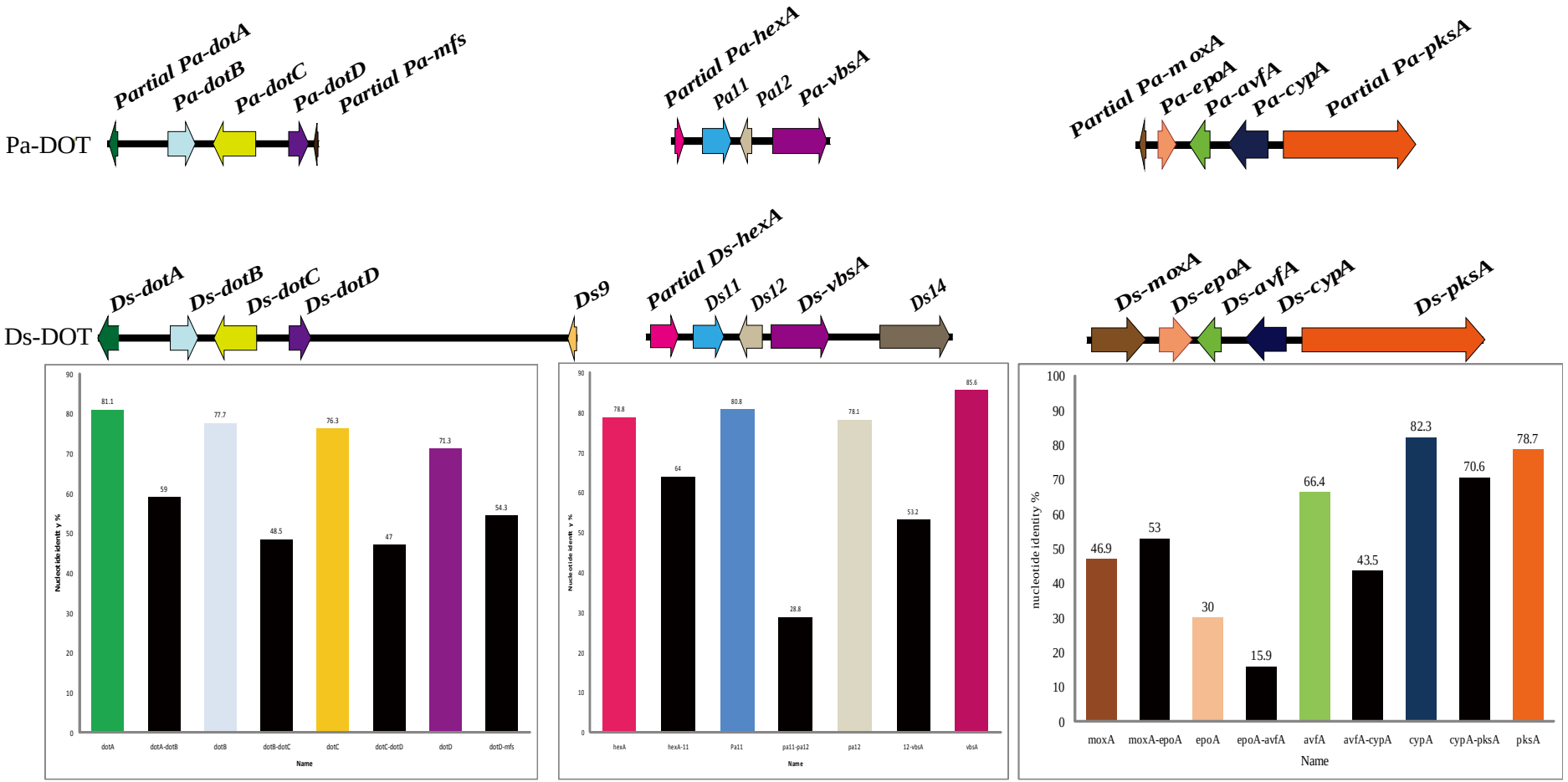


Figure 3.41 The order and orientation of each gene is presented with colour arrows in three mini-clusters. The bottom chart showed the % of nucleotide identity of genes and gene gaps between *P. arachidicola* and *D. septosporum*

Figure 3.42 The GC content of three DOTH gene clusters in *P. arachidicola*

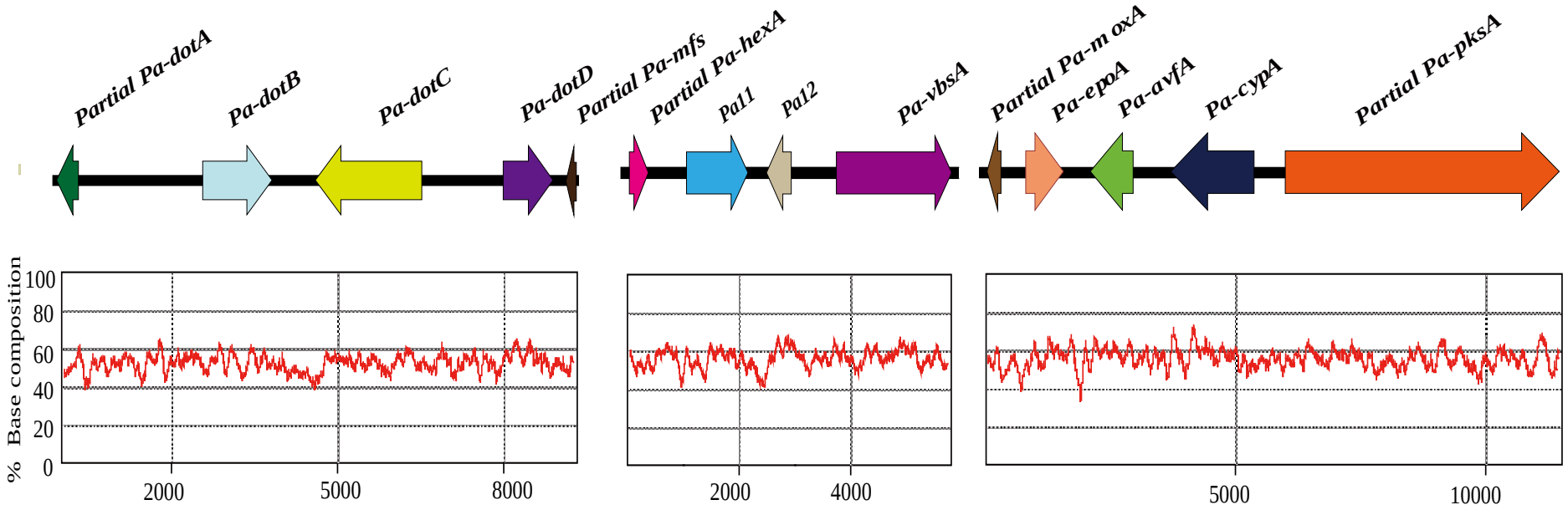


Figure 3.42 The order and orientation of each gene is presented with colour arrows in three mini-clusters. The GC content of DOTH gene cluster are shown below. The Y-axis is the % of GC bases, the X-axis is the length (bp) of the cluster.

Note: The GC content for each cluster was calculated by R. Bradshaw (Massey University) using the program MacVector

Figure 3.43 Comparison of the DOTH, AF and ST biosynthesis clusters in *P. arachidicola*, *D. septosporum* and *Aspergillus* species

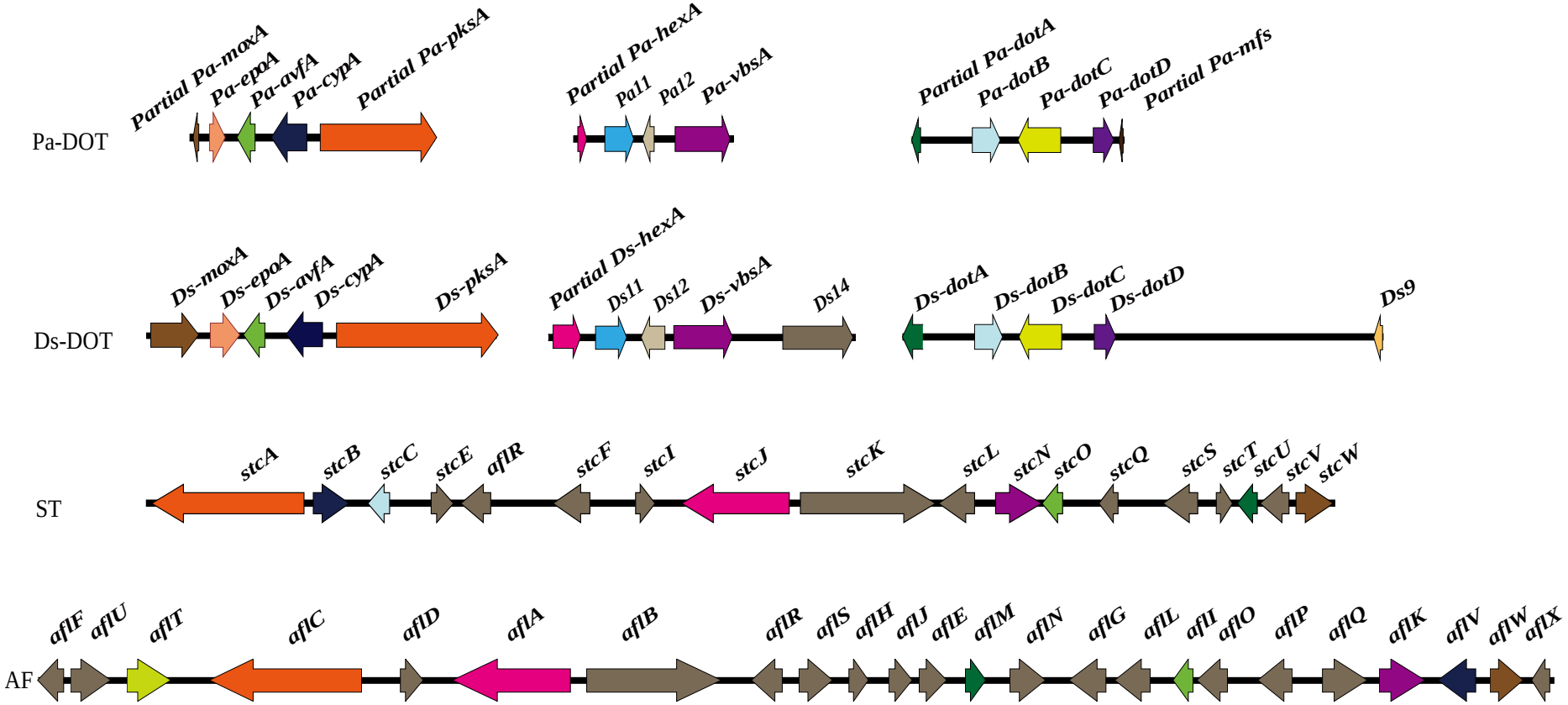


Figure 3.43 The DOTH biosynthesis gene cluster in *P. arachidicola* was compared with the DOTH biosynthesis gene cluster in *D. septosporum*, the AF cluster in *A. parasiticus* and ST cluster in *A. nidulans*. The direction of each arrow indicates the direction of transcription of each gene. Homologous genes are presented with same colour. Gray arrows indicates genes for which no homologues have yet been identified in *P. arachidicola*.

Figure 3.45 Proposed DOTH biosynthesis pathway in *P. arachidicola* compared to the AF biosynthesis pathway in *A. parasiticus*

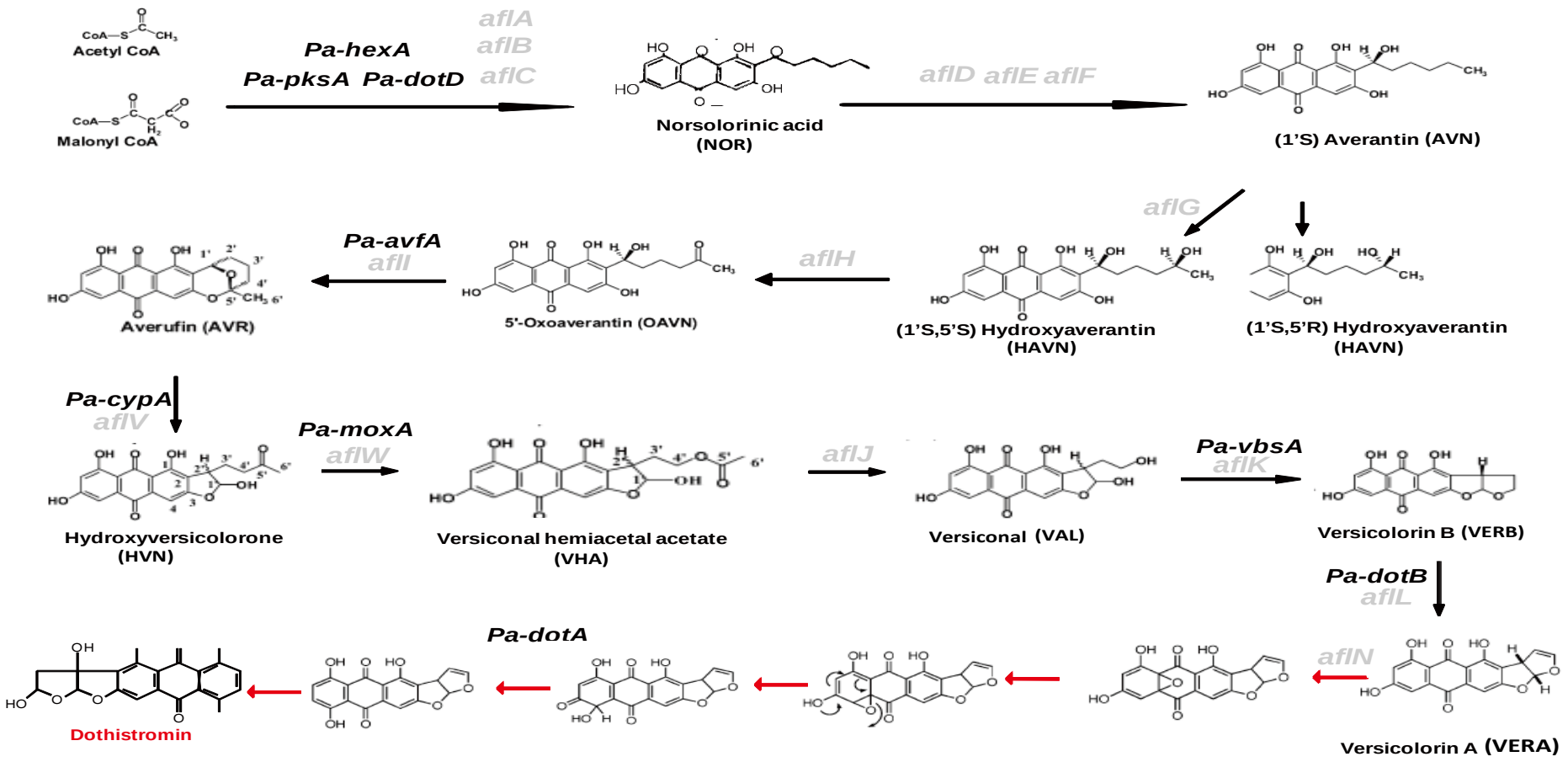
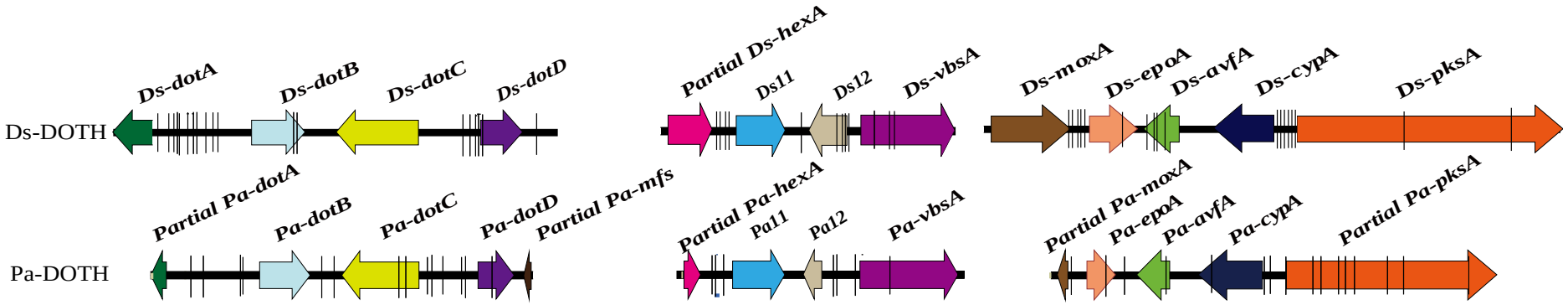


Figure 3.45 The biosynthesis steps of DOTH and AF are identical until VERA. The red arrows represent proposed steps unique for DOTH biosynthesis. The DOTH genes predicted to be involved in each step are showed in black letters; the AF genes in grey letters (Henry and Townsend, 2005; Schwelm, 2007; Wen et al., 2005).

Figure 3.46 AflR binding sites of DOTH genes in *P. arachidicola* and *D. septosporum*



mini-cluster 1

AflR	position
1	*Pa-dotA* - *Pa-dotB* gap (723-733 bp)
2	*Pa-dotA* - *Pa-dotB* gap (1029-1039 bp)
3	*Pa-dotA* - *Pa-dotB* gap (2221-2231 bp)
4	*Pa-dotA* - *Pa-dotB* gap (2372-2382 bp)
5	*Pa-dotB* - *Pa-dotC* gap (4285-4295 bp)
6	*Pa-dotB* - *Pa-dotC* gap (4592-4602 bp)
7	*Pa-dotC* coding region (6077-6087 bp)
8	*Pa-dotC* coding region (6320-6330 bp)
9	*Pa-dotC* - *Pa-dotD* gap (6790-6800 bp)
10	*Pa-dotC* - *Pa-dotD* gap (6846-6856 bp)
11	*Pa-dotC* - *Pa-dotD* gap (7077-7087 bp)
12	*Pa-dotC* - *Pa-dotD* gap (7628-7638 bp)
13	*Pa-dotC* - *Pa-dotD* gap (7715-7725 bp)
14	*Pa-mfs* coding region (8712-8722 bp)

mini-cluster 2

AflR	position
1	*Pa-hexA* - *Pa11* gap (671-681 bp)
2	*Pa-hexA* - *Pa11* gap (693-703 bp)
3	*Pa-hexA* - *Pa11* gap (831-841 bp)
4	*Pa12* - *Pa-vbsA* gap (3398-3408 bp)
5	*Pa12* - *Pa-vbsA* gap (3405-3415 bp)
6	*Pa12* - *Pa-vbsA* gap (3517-3527 bp)
7	*Pa-vbsA* coding region (4294-4302 bp)

mini-cluster 3

AflR	position
1	*Pa-moxA* - *Pa-epoA* gap (517-527 bp)
2	*Pa-epoA* coding region (1317-1326 bp)
3	*Pa-epoA* - *Pa-avfA* gap (1862-1871 bp)
4	*Pa-avfA* coding region (2846-2855 bp)
5	*Pa-cypA* coding region (3927-3936 bp)
6	*Pa-cypA* - *Pa-pksA* gap (5468-5477 bp)
7	*Pa-cypA* - *Pa-pksA* gap (5527-5536 bp)
8	*Pa-cypA* - *Pa-pksA* gap (5995-6004 bp)
9	*Pa-pksA* coding region (6453-6462 bp)
10	*Pa-pksA* coding region (6496-6505 bp)
11	*Pa-pksA* coding region (7147-7156 bp)
12	*Pa-pksA* coding region (7381-7390 bp)
13	*Pa-pksA* coding region (7589-7598 bp)
14	*Pa-pksA* coding region (8562-8571 bp)
15	*Pa-pksA* coding region (8868-8877 bp)
16	*Pa-pksA* coding region (11229-11238 bp)

Figure 3.46 AflR binding sites of DOTH genes in *P. arachidicola* and *D.septosporum*. The order and orientation of each gene is presented with colour arrows. The vertical black lines indicate the positions of putative AflR binding sites. The positions of putative AflR binding sites in each cluster in *P. arachidicola* are shown in the table below.

3.8.3 Discussion

The AF and ST biosynthesis genes are regulated by AflR and its co-regulator AflJ in *Aspergillus* species. The AflR also regulates genes not involved in AF biosynthesis (Price et al., 2006). The ST biosynthesis genes are also regulated by a global secondary metabolite regulator LaeA in a location specific manner. Deletion of *laeA* produced a mutant strain with down-regulated ST gene expression, but expression of genes flanking ST genes was not affected. Furthermore, an extra copy of AflR placed outside of ST cluster could remediate ST gene expression in a *laeA* deletion strain (Bok et al., 2006). Although no *aflR* or *aflJ* homologues have yet been identified in either *D. septosporum* or *P. arachidicola*, AflR binding motifs were found upstream and in the coding region of several DOTH biosynthesis genes. It is possible that DOTH biosynthesis genes in *P. arachidicola* are regulated by genes similar to *aflR* or *aflJ* and that these genes will be found in future studies.

Nine DOTH genes (*Ds-dotA* to *Ds-dotD*, *Ds-vbsA*, *Ds-pksA*, *Ds-cypA*, *Ds-avfA* and *Ds-moxA*) from *D. septosporum* were tested for co-expression. Most of the genes showed high co-regulation with other genes, except for *Ds-vbsA*, *Ds-dotB* and *Ds-dotC*. *Ds-vbsA* had no significant co-regulation with *Ds-dotA* and *Ds-avfA* but some co-regulation to other tested DOTH genes. *Ds-dotB* only had co-regulation with *Ds-avfA*. Although AflR binding sites were identified in the upstream region of *Ds-vbsA* and *Ds-dotB*, the AflR binding sites were quite far away from the ORF of these two genes (Schwelm, 2007). In *P. arachidicola*, both *Pa-dotB* and *Pa-vbsA* had AflR binding sites close to the ORF. The involvement of *Pa-vbsA* in DOTH biosynthesis was confirmed by a complementation experiment, in which *Pa-vbsA* was able to rescue a dothistromin deficent *Ds-vbsA* knockout mutant strain to reproduce DOTH (Zhang unpublished). In *D. septosporum*, *Ds-dotC* showed no co-regulation with any DOTH genes tested is constitutively expressed. No AflR binding sites close to the *Ds-dotC* coding region were identified (Schwelm, 2007). In *A. parasiticus*, *dotC* homologous *aflT* also does not have AflR binding sites in the regulatory region (Cary and Ehrlich, 2006) and *aflT* showed different expression pattern to other AF

genes (Chang et al., 2004). However in *P. arachidicola*, AflR binding sites were identified upstream of *Pa-dotC*, it is of interest to know the co-regulation of *Pa-dotC* with other Pa-DOTH genes.

The DOTH biosynthesis gene cluster in *P. arachidicola* was similar to the DOTH biosynthesis cluster in *D. septosporum* in several respects. The DOTH biosynthesis genes are fragmented into three mini-clusters. Homologous genes had about 73% to 96% amino acid identity between the species, and most of the *Pa-DOTH* genes had similar gene order and direction of transcription to *Ds-DOTH* genes. The DOTH biosynthesis gene cluster in *P. arachidicola* has three main differences to the DOTH biosynthesis cluster in *D. septosporum* including genes *epoA*, *moxA* and *mfs*. DOTH biosynthesis genes in *P. arachidicola* had about 50 % to 69% amino acid identity to homologous AF/ST biosynthesis genes. Like in *D. septosporum*, the DOTH biosynthesis gene cluster in *P. arachidicola* was quite different to the AF or ST biosynthesis gene clusters, both in terms of clustering and gene order. Genes of both AF and ST pathways are clustered and the homologous genes have similar structure and function in the pathway. But the gene order and arrangement are remarkably different between AF and ST clusters. *A. flavus*, *A. parasiticus* and *A. nomius* are three common species that produce AF, the AF genes and cluster organization are virtually identical in these species. The AF biosynthesis genes in *A. ochraceoroseus*, which produces both AF and ST, were studied. Southern blot hybridization using several *A. flavus*, *A. parasiticus* and *A. nidulans* ST/AF gene probes showed that *A. ochraceoroseus* DNA only hybridized weakly to AF *aflL* gene and to ST *aflR* and *stcL* genes (Klich et al., 2005). Further studies showed that AF biosynthesis genes in *A. ochraceoroseus* are more similar to the ST biosynthesis cluster in *A. nidulans* than the AF cluster in *A. parasiticus* in aspects of gene order and direction of transcription (Cary et al., 2005).

Phylogenetic evidence suggested that the AF/ST biosynthesis gene clusters were formed by gene duplication and gene recruitment followed by purifying selection

(Ehrlich et al., 2005). Cary and Ehrlich (2006) proposed a model for AF gene cluster evolution in which a basal cluster which contained the *pksA* genes, genes involved in stabilizing the nascent polyketide (*nor-1*), regulatory genes (*aflR, alfJ*) and genes for production of medium-chain fatty acids (*hexA, hexB*) was established first. Genes encoding modifying enzymes were recruited to the basal cluster later. The DOTH biosynthesis gene cluster could be ancestral to the AF/ST biosynthesis clusters (Cary and Ehrlich, 2006). This evolution model was supported by Carbone et al (2007) who showed a high level of AF gene duplication and seven putative gene modules consisting of genes with conserved gene order, direction of transcription and similar pattern of copy number across species. There authors speculate that the genes in these modules were duplicated together in groups in *Aspergillus* species. One gene module identified was *alfC*/*aflW*, which contains homologues of *Pa-pksA* and *Pa-moxA* (Carbone et al., 2007).

Is the DOTH biosynthesis cluster in *P. arachidicola* a closer ancestor to AF/ST biosynthesis clusters than the DOTH cluster in *D. septosporum*? The closer amino acid identity of predicted *D. septosporum* gene products to AF/ST (Table 3.14) suggested *D. septosporum* is more closely related than *P. arachidicola*. However some features of the *P. arachidicola* DOTH gene organisation suggests a closer similarity of *P. arachidicola* to AF/ST clusters. First, *Pa-moxA* had different transcription direction to *Ds-moxA*, but it has the same transcription direction to *aflW* and *stcW* in the AF/ST biosynthesis cluster relative to *pksA* (*stcA*/*aflC*) (Figure 3.43). The opposite direction of *Ds-moxA* in *D. septosporum* could be a result of gene rearrangement. Second, studies showed that the closest ORF at the distal end (closest to chromosomal end) of the ST cluster was predicted to encode a MFS transporter (Genbank accession no.AN7826) in *A. flavus* and a MFS transporter was identified within 5 kb downstream of AF cluster in *A. flavus* strain AF13 (Cary and Ehrlich, 2006; Ehrlich et al., 2005). There was a predicted MFS transporter *Pa-mfs* downstream of *Pa-dotD* which was not seen in the *D. septosporum* cluster. The entire sequence of *Pa-mfs* is not yet known, however Pa-Mfs had 12.7%, 12.9% and 14.3%

amino acid identity to the corresponding regions of homologous genes St-Mfs, Ds-DotC and AflT respectively, which is much lower than other homologues involved in DOTH, ST and AF biosynthesis clusters. A *Pa-mfs* knock out mutant needs to be constructed to test whether *Pa-mfs* has a role in DOTH biosynthesis or if it acts as toxin pump like *Pa-dotC* is predicated to do in *P. arachidicola*.

Fragmented secondary metabolite biosynthesis gene clusters are seen in other fungal species. In the grass endophyte *Neotyphodium lolii*, the lolitrem biosynthesis gene cluster is fragmented into three mini-clusters, with lolitrem biosynthesis genes separated by large blocks of AT-rich sequences and Type I retrotransposon sequences were identified in mini-cluster 1 (Young et al., 2006). The fragmented gene cluster organization showed evidence of retrotransposon activity. The retrotransposon relics and AT rich regions were not identified in the DOTH biosynthesis gene clusters in either *P. arachidicola* or *D. septosporum*. The nucleotide identity of each homologous gene is higher than the average nucleotide identity of each intergenic region between *D. septosporum* and *P. arachidicola*. Several tandem and inverted repeat sequences were identified in DOTH gene cluster intergenic regions in *P. arachidicola*, but the distribution of those repeats appears to be random when compared to *D. septosporum*. All those suggested that the fragmentation of the DOTH biosynthesis gene cluster may not due to retrotransposon activity or recombination between repeat sequence.

Chapter 4 Conclusions and future work

4.1 The DOTH biosynthesis gene clusters in *P. arachidicola*

Although it is not known if the three regions containing DOTH biosynthetic genes are clustered in *P. arachidicola*, they are each similar to mini-clusters found in the fragmented DOTH cluster in *D. septosporum*.

Southern blots using DIG-labelled *P. arachidicola* probes suggested that there is only one copy of *Pa-pksA*, *Pa-dotA*, *Pa-vbsA* and *Pa-cypA* genes in the whole genome. Based on Southern blot results that showed the size of fragments that contained the DOTH genes in *P. arachidicola*, *Eco*RI 5-11 kb and 11-20 kb size-fractionated genomic libraries were made and screened. Four positive clones were found, two of which overlapped to make three sequences of 11.5 kb, 5.7 kb and 9.3 kb were obtained. Sequence analysis identified 11 putative DOTH biosynthesis genes separated into three mini-clusters, similar to the DOTH biosynthesis gene cluster in *D. septosporum*. Most of the predicted DOTH biosynthesis genes in *P. arachidicola* had high amino acid identity to homologous DOTH genes in *D. septosporum*, and were therefore predicted to have the same functions. Homologous genes had about 73% to 96% amino acid identity between the species, and most of the *Pa-DOTH* genes had similar gene order and direction of transcription to *Ds-DOTH* genes. No regulatory genes were identified in either *P. arachidicola* or *D. septosporum*, but conserved *aflR* binding sites were identified in the upstream and coding region of several DOTH genes in two species. The DOTH genes in *P. arachidicola* may have a similar co-expression pattern to DOTH genes in *D. septosporum*.

There are three different points between the DOTH biosynthesis gene clusters in *P. arachidicola* and *D. septosporum*. Firstly *Pa-epoA* may not functional due to a premature stop codon in the coding region. Secondly the *Pa-moxA* and *Ds-moxA* had different transcription orientations. Thirdly a MFS transporter, *Pa-mfs*, was found in *P.*

arachidicola that is not seen in *D. septosporum*. This was additional to the *Pa-dotC* MFS gene upstream of *Pa-mfs* in mini-cluster 3. The amino acid identity between *Pa-dotC* and *Pa-mfs* was quite low, only 12.9%, while *Pa-mfs* had 54% amino acid identity to the corresponding region of a MFS transporter in *A. fumigatus*. The DOTH biosynthesis genes in *P. arachidicola* had different gene organization and direction of transcription to homologous AF or ST biosynthesis genes in *Aspergillus* species. Several tandem and inverted repeat sequences were identified in DOTH gene cluster intergenic regions in *P. arachidicola,* but the distribution of those repeats appears to be random, suggesting that the fragmentation of the DOTH biosynthesis gene cluster may not due to retrotransposon activity or recombination between repeat sequences. The DOTH biosynthesis gene cluster in *P. arachidicola* was predicted to be ancestral to the AF/ST biosynthesis clusters. Understanding the DOTH biosynthesis gene cluster in *P. arachidicola* could provide insights into the evolution of AF/ST biosynthesis cluster

Studies showed that DOTH is not required for *D. septosporum* to cause Dothistroma needle blight of pines (Schwelm, 2007). The growth of other pine-needle inhabitants were inhibited by DOTH in *vitro*, so DOTH may provide an advantage to *D. septosporum* in growth competition with other fungi. It may play a role in competition of *D. septosporum* with other fungi in its ecological niche (Schwelm et al., in press). So the DOTH produced by *P. arachidicola* may have a similar role.

4.2 Future works

4.2.1 Identification of other DOTH biosynthesis genes in *P. arachidicola*

The three mini-clusters identified in *P. arachidicola* so far are not complete. It is of interest to know the regions flanking the DOTH biosynthesis genes. First, according to Southern blot results and sequence analysis of the three mini-clusters, 7-13 kb and 13-20 kb size-fractionated *Sca*I genomic libraries should contain positive clones that cover more regions of mini-clusters 1 and 2, so further efforts could be made to obtain and sequence these. Second, a genome walking strategy could be used to identify

flanking sequences adjacent to known regions. This PCR based method involves gene-specific primer (GSP) and a tailing step with a 5'-RACE abridged anchor primer (AAP) (Leoni et al., 2008). Third, because there is high amino acid identity of DOTH genes between *P. arachidicola* and *D. septosporum*, new degenerate primers can be designed according to *D. septosporum* gene sequences and used for PCR to identify fragments containing DOTH genes in *P. arachidicola*.

4.2.2 Identification of DOTH gene functions in *P. arachidicola*

The *Pa-vbsA* and *Pa-dotA* genes from *P. arachidicola* were recently shown to be involved in DOTH biosynthesis by complementation experiments in this lab (unpublished data). The *Pa-vbsA* and *Pa-dotA* genes were able to rescue *D. septosporum Ds-vbsA* and *Ds-dotA* mutant strains and allowed them to reproduce DOTH. A *Ds-pksA* knockout strain is also available that cannot produce DOTH. It would be interesting to see if *Pa-pksA* is able to rescue the *Ds-pksA* mutant strain. *D. septosporum* knockout strains were obtained by transformation of the *Ds*-gene containing vector to *D. septosporum* wild type strain using a protoplast mediated method. Because there are some difficulties in making *P. arachidicola* protoplast, *D. septosporum* knockout strains were used for complementation experiments. An alternative transformation system such as Agrobacterium mediated transformation would be a possibility for making *P. arachidicola* knockout strains.

Another gene of interest was *Pa-mfs* which did not have a homologue in *D. septosporum*. If a transformation system could be developed, a *Pa-mfs* knockout mutant could be obtained by gene replacement. The DOTH biosynthesis and toxin secretion can be monitored by ELISA assay. It is also of interest to compare the effect of *Pa-dotC* or *Pa-mfs* mutant strain for DOTH biosynthesis and toxin secretion. This may help to understand the relation of these two MFS transporters.

Based on similarities between *P. arachidicola* and *D. septosporum* gene clusters seen so far, it is expected that further analysis will reveal a fragmented gene cluster in *P.*

arachidicola, and that the genes will have similar functions to their homologues. There are plans in our laboratory to sequence the genome of *D. septosporum*. This is expected to reveal other DOTH genes and their locations. From this, further studies of the *P. arachidicola* DOTH gene cluster can be made that will lead to a more complete understanding of the evolutionary origins of these gene clusters.

References

Ahn, I.P., and Lee, Y.H. (2001). A viral double-stranded RNA up regulates the fungal virulence of *Nectria radicicola*. Molecular Plant-Microbe Interactions *14*, 496-507.

Anagnostakis, S.L., and Day, P.R. (1979). Hypovirulence conversion in *Endothia parasitica*. Phytopathology *69*, 1226-1229.

Assante, G., Locci, R., Camarda, L., Merlini, L., and Nasini, G. (1977). Screening of the genus *Cercospora* for secondary metabolites. Phytochemistry *16*, 243-247.

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

Barnes, I., Crous, P.W., Wingfield, B.D., and Wingfield, M.J. (2004). Multigene phylogenies reveal that red band needle bight of *Pinus* is caused by two distinct species of *Dothistroma*, *D. septosporum* and *D. pini*. Studies in Mycology *50*, 551-565.

Bassett, C., Buchanan, M., Gallagher, R.T., and Hodges, R.L. (1970). A toxic difuroanthraquinone from *Dothistroma pini*. Chemistry and Industry *26*, 1659-1660.

Bear, C.A., Waters, J.M., and Waters, T.N. (1972). Crystal structure and absolute configuration of a derivative of dothistromin, a fungal toxin implicated in pine-needle blight. Journal of the Chemical Society, Perkin Transactions II *2*, 2375-2378.

Bok, J.W., Noordermeer, D., Kale, S.P., and Keller, N.P. (2006). Secondary metabolic gene cluster silencing in *Aspergillus nidulans*. Molecular Microbiology *61*, 1636-1645.

Bradshaw, R.E. (2004). Dothistroma (red-band) needle blight of pines and the dothistromin toxin: a review. Forest Pathology *34*, 163-185.

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.

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., Jin, H.P., Morgan, B.S., Schwelm, A., Teddy, O.R., Young, C.A., and Zhang, S.G. (2006). A polyketide synthase gene required for biosynthesis of the aflatoxin-like toxin, dothistromin. Mycopathologia *161*, 283-294.

Bradshaw, R.E., and Zhang, S.G. (2006). Biosynthesis of dothistromin. Mycopathologia *162*, 201-213.

Brown, D.W., Adams, T.H., and Keller, N.P. (1996a). *Aspergillus* has distinct fatty acid synthases for primary and secondary metabolism. Microbiology *93*, 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*. Microbiology *93*, 1418-1422.

Buck, K.W. (1986). Fungal virology - an overview (CRC Press. Boca Raton, FL).

Carbone, I., Ramirez-prado, J., Jakobek, J.L., and Horn, B.W. (2007). Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster. Evolutionary Biology *7*, 1471-2148.

Cary, J.W., and Ehrlich, K.C. (2006). Aflatoxigenicity in *Aspergillus*: molecular genetics, phylogenetic relationships and evolutionary implications. Mycopathologia *162*, 167-177.

Cary, J.W., Klich, M.A., and Beltz, S.B. (2005). Characterization of aflatoxin-producing fungi outside of *Aspergillus* section *Flavi*. Mycologia *97*, 425-432.

Chang, P.K., Yu, J., and Yu, J.H. (2004). *aflT*, a MFS transporter-encoding gene located in the aflatoxin gene cluster, does not have a significant role in aflatoxin secretion. Fungal Genetics and Biology *41*, 911-920.

Cleveland, T.E., and Bhatnagar., D. (1991). Molecular regulation of aflatoxin biosynthesis (Pennington Center Nutrition Series, LSU Press, Baton Rouge, La.).

Cormack, R.S., and Somssich, I.E. (1997). Ultra-fast rapid alkaline extraction (UFX). Technical tips online *1*. http://mysite.verizon.net/vze1y721/cvpdf/tto_2.pdf. July, 2006.

Culbreath, A.K., Brenneman, T.B., and Kemerait, R.C. (2002). Management of early leaf spot of peanut with pyraclostrobin as affected by rate and spray interval. Plant Management Network. http://www.plantmanagementnetwork.org/pub/php/research/pyraclostrobin/. 2008.

Damicone, J.P., and Melouk, H.A. (1998). Foliar disease of peanut. http://pods.dasnr.okstate.edu/docushare/dsweb/Get/Document-2319/EPP-7655web.pdf. 2007.

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin *19*, 11-15.

Dugaiczyk, A., Boyer, H.W., and Goodman, H.M. (1975). Ligation of *Eco*RI endonuclease-generated DNA fragments into linear and circular structure. Journal of Molecular Biology *96*, 171-184.

Ehrlich, K.C., Yu, J., and Cotty, P.J. (2005). Aflatoxin biosynthesis gene clusters and flanking regions. Journal of Applied Microbiology *99*, 518-527.

Elias, K.S., and Cotty, P.J. (1996). Incidence and stability of infection by double stranded RNA genetic elements in *Aspergillus* section *flavi* and effects on aflatoxigenicity. Canadian Journal of Botany *74*, 716-725.

Elliott, G.S., Marson, R.W., Ferry, D.G., and Edwards, I.R. (1989). Dothistromin risk assessment for forest workers. New Zealand Journal of Forest Science *19*, 163-170.

Feng, Z.L. (2007). Further studies of dothistromin toxin genes in the fungal forest pathogen *Dothistroma septosporum*. In Institute of Molecular Biosciences (MSc, Massey University), pp. 149.

Ferguson, L.R., Parslow, M.I., and McLarin, J.A. (1986). Chromosome damage by dothistromin in human peripheral blood lymphocyte cultures: a comparison with aflatoxin B1. Mutation Research *170*, 47-53.

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 *Pinus radiata* families to *Dothistroma pini*. Physiological and Molecular Plant Pathology *26*, 267-286.

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

Goodwin, S.B., Dunkle, L.D., and Zismann, V.L. (2001). Phylogenetic analysis of *Cercospora* and *Mycosphaerella* based on the internal transcribed spacer region of ribosomal DNA. Ecology and Population Biology *91*, 648-658.

Harvey, A.M., Batt, R.D., and Pritchard, G.G. (1976). Inhibition of RNA synthesis in *Chlorella pyrenoidosa* and *Bacillus megaterium* by the pine-blight toxin, dothistromin. Journal of General Microbiology *96*, 268-276.

Henry, K.M., and Townsend, C.A. (2005). Ordering the reductive and cytochrome P450 oxidative steps in demethylsterigmatocysin formation yields general insights into the biosynthesis of aflatoxin and related fungal metabolites. Journal of the American Chemical Society *127*, 3724-3733.

Hotter, G.S. (1997). Elicitor-induced oxidative burst and phenylpropanoid metabolism in *Pinus radiata* cell suspension cultures. Australian Journal of Plant Physiology *24*, 797-804.

Hunst, P.L., Latterell, F.M., and Rossi, A.E. (1986). Variation in double-stranded RNA from isolates of *Pyricularia oryzae*. Phytopathology *76*, 674-678.

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

Keller, N.P., Segner, S., Bhatnagar, D., and Adams, T.H. (1995). *stcS*, a putative P-450 monooxygenase, is required for the conversion of versicolorin A to sterigmatocystin in *Aspergillus nidulans*. Applied and Environmental Microbiology *61*, 3628-3632.

Klich, M.A., Mullaney, E.J., Daly, C.B., and Cary, J.W. (2005). Molecular and physiological aspects of aflatoxin and sterigmatocystin biosynthesis by *Aspergillus tamarii* and *A. ochraceoroseus* Applied and Environmental Microbiology *53*, 605-609.

Kousik, C.S., Snow, J.P., and Valverde, R.A. (1994). Comparison of double-stranded RNA components and virulence among isolates of *Rhizoctonia solani* AG-1 IA and AG-IB. Phytopathology *84*, 44-49.

Kucharek, T. (2000). Peanut leaf spot and rust. Plant Pathology Fact Sheet, 5. http://plantpath.ifas.ufl.edu/takextpub/FactSheets/pp0005.pdf. 2007

Leoni, C., Gallerani, R., and Ceci, L.R. (2008). A genome walking strategy for the identification of eukaryotic nucleotide sequences adjacent to know regions. Biotechniques *44*, 229-235.

McGuire, S.M., Silva, J.C., Casillas, E.G., and Townsend, C.A. (1996). Purification and characterization of versicolorin B synthase from *Aspergillus parasiticus*. Catalysis of the stereodifferentiating cyclization in aflatoxin biosynthesis essential to DNA interaction. Biochemistry *35*, 11470-11486.

McLarin, J.A., and Ferguson, L.R. (1985). Mutagenesis and clastogenesis induced by dothistromin in V79 chinese-hamster fibroblasts. New Zealand Medical Journal *98*, 962.

Meyers, D.M., Brian, G.O., Du, W.L., Bhatnagar, D., and Payne, G.A. (1998). Characterization of *aflJ*, a gene required for conversion of pathway intermediates to aflatoxin. Applied and Environmental Microbiology *64*, 3713–3717.

Michelmore, R.W., and Hulbert, S.H. (1987). Molecular markers for genetic analysis of phytopathogenic fungi. Annual Review of Phytopathology *25*, 383-404.

Nutsugah, S.K., Abudulai, M., Oti-Boateng, C., Brandenburg, R.L., and Jordan, D.L. (2007). Management of leaf spot disease of peanut with fungicides and local detergents in Ghana. Plant Pathology Journal *6*, 248-253.

Price, M.S., Yu, J., Nierman, M.C., Kim, H.S., Pritchard, B., Jacobus, C.A., Bhatnagar, D., Cleveland, T.E., and Payne, G.A. (2006). The aflatoxin pathway regulator AflR induces gene transcription inside and outside of the aflatoxin biosynthetic cluster. FEMS Microbiology Letters *255*, 275-279.

Ricker, M.D., Beute, M.K., and Campbell, C.L. (1985). Components of resistance in peanut to *Cercospora arachidicola*. The American Phytopathological Society: Plant Disease *69*, 1059-1064.

Rigling, D., and Van Alfen, N.K. (1993). Extra and intracellular laccase of the chestnut blight fungus, *Cryphonectria parasitica*. Applied and Environmental Microbiology *59*, 3634-3639.

Romo, M., Leuchtmann, A., Garcia, B., and Zabagogeazcoa, I. (2007). A totivirus infecting the mutualistic fungal endophyte *Epichloe festucae*. Virus Research *124*, 38-43.

Schmidt, F.R., Davis, N.D., Diener, U.L., and Lemeke, P.A. (1983). Cyclohexamide induction of aflatoxin synthesis in a non-toxigenic strain of *Aspergillus flavus*. Biochemistry *1*, 794-795.

Schmidt, F.R., Davis, N.D., Diener, U.L., and Lemeke, P.A. (1986). Viral influences on aflatoxin formation by *Aspergillus flavus*. Applied Microbiology and Biotechnology *24*, 248-252.

Schwelm, A. (2007). Investigations of dothistromin gene expression in *Dothitroma septosporum* and the putative role of dothistromin toxin. In Institute of Molecular Biosciences (PhD, Massey University), pp. 140.

Schwelm, A., Barron, N.J., Baker, J., Dick, M., Long, P.G., Zhang, S., and Bradshaw, R.E. (in press). Dothistromin toxin is not required for dothistroma needle blight in *Pinus radiata*. Plant Pathology.

Schwelm, A., Barron, N.J., Zhang, S., and Bradshaw, R.E. (2008). Early expression of aflatoxin-like dothistromin genes in the forest pathogen *Dothistroma septosporum*. Mycological Research *112*, 138-146.

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

Shen, X.R., Chen, M.J., Shen, J.Y., Pan, Y.J., and Gong, Z.X. (1993). The detection of singlestranded RNA in an isometric virus-like particle from Shiitake mushroom Annals of Applied Biology *122*, 271-277.

Silva, J.C., Minto, R.E., Barry III, C.E., Holland, K.A., and Townsend, C.A. (1996). Isolation and characterization of the versicolorin B synthase gene from *Aspergillus parasiticus*. The Journal of Biological Chemistry *271*, 13600-13608.

Silva, J.C., and Townsend, C.A. (1996). Heterologous expression, isolation, and characterization of versicolorin B synthase from *Aspergillus parasiticus*. The Journal of Biological Chemistry *272*, 804-813.

Skinnider, L., Stoessl, A., and Wang, J. (1989). Increased frequency of sister-chromatid exchange induced by dothistromin in CHO cells and human lymphocytes. Mutation Research *222*, 167-170.

Skory, C.D., Chang, P.K., Cary, J., and Linz, J.E. (1992). Isolation and characterization 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.

Stewart, E.L., Liu, Z., Crous, P.W., and Szabo, L.J. (1999). Phylogenetic relationships among some cercosporoid anamorphs of *Mycosphaerella* based on rDNA sequence analysis. Mycological Research *103*, 1491-1499.

Stoessl, A. (1984). Dothistromin as a metabolite of *Cercospora arachidicola*. Mycopathologia *86*, 165-168.

Stoessl, A., and Stothers, J.B. (1985). Minor anthraquinoid metabolites of *Cercospora arachidicola*. Canadian Journal of Chemistry *63*, 1258-1261.

Trail, F., Mahanti, N., Rarick, M., Mehigh, R., Liang, S.H., Zhou, R., and Linz, J.E. (1995). Physical and transcriptional map of an aflatoxin gene cluster in *Aspergillus parasiticus* and functional disruption of a gene involved early in the aflatoxin pathway. Applied and Environmental Microbiology *61*, 2665-2673.

Van Diepeningen, A.D., Debets, A.J.M., and Hoekstra, R.F. (2006). Dynamics of dsRNA mycoviruses in black *Aspergillus* populations. Fungal Genetics and Biology *43*, 446-452.

Watanabe, C.M., Wilson, D., Linz, J.E., and Townsend, C.A. (1996). Demonstration of the catalytic roles and evidence for the physical association of type I fatty acid synthases and a polyketide synthase in the biosynthesis of aflatoxin B1. Chemistry and Biology *3*, 463-469.

Wen, Y., Hatabayashi, H., Arai, H., Kitamoto, H.K., and Yabe, K. (2005). Function of the *cypX* and *moxY* genes in aflatoxin biosynthesis in *Aspergillus parasiticus*. Applied and Environmental Microbiology *71*, 3192-3198.

Williams, R., Kline, M., and Smith, R. (1996). BSA and restriction enzyme digestions. Promega Notes Magazine *59*, 46.

Young, C.A., Felitti, S., Shields, K., Spangenberg, G., Johnson, R.D., Bryan, G.T., Saikia, S., and Scott, B. (2006). A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte *Neotyphodium lolii*. Fungal Genetics and Biology *43*, 679-693.

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

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

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

Yu, J.J., Bhatnagar, D., and Cleveland, T.E. (2004b). Completed sequence of aflatoxin pathway gene cluster in *Aspergillus parasiticus*. FEBS Letters *564*, 126-130.

Yu, J.J., Cleveland, T.E., Nierman, W.C., and Bennett, J.W. (2005). *Aspergillus flavus* genomics: gateway to human and animal health, food safety, and crop resistance to diseases. Revista Iberoamericana De Micologia *22*, 194-202.

Zabagogeazcoa, I., Benito, E.P., Ciudad, A.G., Crado, B.G., and Eslava, A.P. (1998). Double-stranded RNA and virus-like particles in the grass endophyte *Epichloe festucae*. Mycologia *102*, 914-918.

Zhang, S.G., Schwelm, A., Jin, H.P., Collins, L.J., and Bradshaw, R.E. (2007). A fragmented aflatoxin-like gene cluster in the forest pathogen *Dothistroma septosporum*. Fungal Genetics and Biology. *44*, 1342-1354.

Appendix I Media

All media were prepared using MilliQ water and sterilized by autocleaving at 121°C for 15 minutes. Media were cooled to approximately 50°C before addition of antibiotics

A1.1 Luria Broth Media (LB)

Tryptone (Becton, Dickison and company), 10 (g/L); NaCl 5 (g/L); Yeast extract (Becton, Dickison and company), 5 (g/L)

A1.2 Luria Broth Agar (LB A gar)

Tryptone (Becton, Dickison and company), 10 (g/L); NaCl 5 (g/L); Yeast extract (Becton, Dickison and company), 5 (g/L); Agar, 15 (g/L)

A 1.3 Potato Dextrose Media (PD)

Potato dextrose agar (Becton, Dickison and company) 24 (g/L)

A 1.4 Potato Dextrose Agar (PDA)

Potato dextrose agar (Merck) 39 (g/L)

Appendix II Buffers and solutions

A2.1 Hexadecyltrimethylammonium bromide (CTAB)

2% CTAB (Sigma); 1% PVP 40 (Sigma); 1.4M NaCl (Merk); 20mM EDTA (Sigma) 0.1M Tris/HCl pH8 (Invitrogen). Make volume up to 40 ml with MilliQ water. Heat to 60°C to dissolve the CTAB and PVP.

A2.2 10 x TNE buffer

12.11 g/L Tris (Invitrogen); 3.72 g/L EDTA (Sigma); 116.89 g/L NaCl (Merck). Dissolve in 800 ml MilliQ water, adjust pH to 7.4 using concentrated HCl (BDH), then make volume up to 100 ml with MilliQ water.

A2.3 10 x TBE buffer

108 g/L Tris (Invitrogen); 9.3 g/L EDTA (Sigma); 55 g/L Boric acid (Univar). Dissolve in 800 ml MilliQ water, adjust pH to 8.2 using concentrated HCl (BDH), then make volume up to 100 ml with MilliQ water.

A2.4 1 x TE buffer

10 mM Tris (Invitrogen); 1 mM EDTA (Sigma); pH8.5

A2.5 Fluorometer DNA standard

100 ng/µl Calf thymus DNA (Amersham Biosciences) in 1 x TNE

A.2.6 Hoechst dye stock solution

10 ml of MilliQ water was added to 10 mg of Hoechst H33258 (Amersham Biosciences).

A2.7 Fluorometer working solution

5 µl Hoechst H33258 stock solution; 5 ml 10 x TNE; 45 ml MilliQ water

A2.8 Ethidium bromide staining solution

1 mg/mL ethidium bromide in MilliQ water

A2.9 Gel loading dye

20% (w/v) Sucrose (BDH); 5 mM EDTA Na2.H_2O (DBH); 1% (w/v) SDS (DBH); 0.2% (w/v) bromophenol blue (J.T. Baker Chemical Co); 0.2% (w/v) xylene cyanol (Sigma)

A2.10 Plasmid DNA extraction solution

Lysis solution

1% (w/v) SDS (Sigma); 0.2 M NaOH (BDH)

Solution III

3 M potassium acetate, pH 5.50

A2.11 Sourth Blot and hybridization solution

Denaturing Solution

500 mM NaOH (BDH); 500 mM NaCl (BDH)

Neutralising Solution

500 mM Tris pH 7.4; 0.5M NaCl (BDH)

20 x SSC Solution

3 M NaCl (BDH); 0.3 mM Na citrate

Low stringency Washing Solution

2 x SSC; 1% (w/v) SDS (Sigma)

High stringency Washing Solution

0.5 x SSC; 1% (w/v) SDS (Sigma)

Buffer I

100 mM Tris pH7.5; 150 mM NaCl (BDH)

Buffer II

buffer I + 1% blocking reagent (Roche)

Buffer III

100 mM Tris-HCl pH9.5; 100 mM NaCl (BDH)

Stripping Solution

200 mM NaOH (BDH); 0.1% (w/v)SDS (Sigma)

Appendix III Primers used in this project

Primers used for inverse PCR (section 3.1)

Gene	Enzyme	Primers
Pa-vbsA	*Sal*I	vbsA3 and szdp75 or szdp76 and szdp79
Pa-vbsA	*Bam*HI	vbsA3 and szdp79
Pa-vbsA	*Sca*I	vbsA3 and szdp79
Pa-vbsA	*Eco*RI	vbsA3 and szdp79
Pa-cypA	*Bam*HI	szdp84 and szdp85 or szdp86 and szdp87
Pa-cypA	*Pst*I	szdp84 and szdp88
Pa-cypA	*Sca*I	szdp84 and szdp87
Pa-cypA	*Eco*RI	szdp84 and szdp87

Degenerate primers (section 3.3)

Gene	Primer name	Primer nucleotide sequence (5'-3')
aflR	szdp127	CAGGCGCGCYMBYRYWGYYT
aflR	szdp128	ACGTAGCCAYCCYSYGMRYR
avnA	aflG F1	GARTTYCCNAARGAYCCNATGCA
avnA	aflG F2	TGGTTYAAYTGGGCNACNTTYGA
avnA	aflG R1	GCNARCCARTCNGTRTCYTTCAT
avnA	aflG R2	GGYTTYTTRTCCCANARDATCCA
verB	vbrBF1	CACCARAARTAYGGNGAYACNGT
verB	vbrBF2	TACTTYAAYATGGCNATHTTYGA
verB	verBR3	TTATGAATNGCRTCNACCCANGG
verB	verBR4	TGGCCTGGNACNGCNGGRTACAT
norA	norAF1	GGCAACTTCATHGAYRYNGCNAA
norA	norAF2	GATGTNGCNAAYTTYTAYCARGG
norA	norAR3	GTAGTRAARTCCCACATRTGNACRTA
norA	norAR4	CCGATCACIGGRAA IACRTAIGG

Primers for sequencing

Name	Sequence	Gene	Organism
MGPA1	CCTTCCACCATTCTTCTCTG	cypA	*P. arachidicola*
MGPA2	TCGGGATGGATTCACTTGAC	cypA	*P. arachidicola*
MGPA3	TCGATGGATGCGACTTCCTC	cypA	*P. arachidicola*
MGPA4	CAATCAACATGGCCTGCTAC	cypA	*P. arachidicola*
MGPA5	CGTCCAGATCCAGTGACAAG	ver-1	*P. arachidicola*
MGPA6	AGCTGGTTTAGCGGGTGCTC	ver-1	*P. arachidicola*

MGPA7	GAAAGAGGTCGACGGCTGTG	vbsA	*P. arachidicola*
MGPA8	CATCCCAGAATAGCCATGTG	vbsA	*P. arachidicola*
MGPA9	ACTGCACCATCCGCATCATC	cypA	*P. arachidicola*
MGPA10	AACGAGGAATGGACGCTCAG	cypA	*P. arachidicola*
MGPA11	CAACGTAAGCTCCACCTCAC	vbsA	*P. arachidicola*
MGPA12	CGCCAACTCCTGCACCAATG	vbsA	*P. arachidicola*
MGPA13	TGGTATGGGTATGTGCAGAG	vbsA	*P. arachidicola*
MGPA14	GGCAGGTAGGGTTGAGGTAG	vbsA	*P. arachidicola*
MGPA15	GGACACTGATGGCCAACGAG	cypA	*P. arachidicola*
MGPA16	TTGGCGGTGGACACAACAAG	cypA	*P. arachidicola*
MGPA17	ACTGCCTTCCTGCAACAGTC	cypA	*P. arachidicola*
MGPA18	ACCTCCTTTGGACCGATACG	cypA	*P. arachidicola*
MGPA19	GGTCGTTCCATGTTGTGTTC	moxA	*P. arachidicola*
MGPA20	AATAGCGTGCTGCGCATTCC	moxA	*P. arachidicola*
MGPA21	ACACCACCAACCTACAAGTC	vbsA	*P. arachidicola*
MGPA22	GATCTCCCTTCGAGTCTTTG	pksA	*P. arachidicola*
MGPA23	CAGCTATCTCGCCGTCATGC	pksA	*P. arachidicola*
MGPA24	AGCCACTTCTGCGTCGTACC	ver-1	*P. arachidicola*
MGPA25	GGGACATGTTGCTGGTTGTG	vbsA	*P. arachidicola*
MGPA26	CACAAACCTCGGCTTCATAC	vbsA	*P. arachidicola*
MGPA27	ATTGTGCCCGACACGACACC	pksA	*P. arachidicola*
MGPA28	ACGTTCGTCCGGAGAAGTTG	pksA	*P. arachidicola*
MGPA29	TAATTGGGCGCCTTTGATCG	ver-1	*P. arachidicola*
MGPA30	GCTGTCCTTCCACGGATCTG	ver-1	*P. arachidicola*
MGPA31	GGGCTCGATGCTCAATACTC	vbsA	*P. arachidicola*
MGPA32	GCTGCGACTGCTTGTGGTTC	vbsA	*P. arachidicola*
MGPA33	CTGTCTTTGCACCGCACGAG	pksA	*P. arachidicola*
MGPA34	TCAACGTGGGCAGCGTTGTC	pksA	*P. arachidicola*
MGPA35	TCGACAGCTTGAGCTCTATG	pksA	*P. arachidicola*
MGPA36	ACGCATCTGTGCGGTATTTC	ver-1	*P. arachidicola*
MGPA37	ATAGCTCGCTACCAGGTTTG	ver-1	*P. arachidicola*
MGPA38	CGATCCTCGAACTCCCAAAC	cypA	*P. arachidicola*
MGPA39	CCCATGAATGGTAGGGAAAG	cypA	*P. arachidicola*
MGPA40	GGAACCAACAGCCAGTAGAC	moxA	*P. arachidicola*
MGPA41	TATTCGCAGCTCGCTGTATC	pksA	*P. arachidicola*
MGPA42	GCACCACAACTGGTAGATTC	pksA	*P. arachidicola*
MGPA43	TGCGACATCCATCGACTATC	ver-1	*P. arachidicola*
MGPA44	TAGCGAGCGGACATCAAAGG	ver-1	*P. arachidicola*
MGPA45	CAAGGCTCTCGATGCATTTG	vbsA	*P. arachidicola*
MGPA46	TGGTGCCGTGGTCATCCTAC	vbsA	*P. arachidicola*
MGPA47	TGCTGTTGCTGTGGCTGTTG	cypA	*P. arachidicola*
MGPA48	CGGTCAGCAGATGTAAGAAG	cypA	*P. arachidicola*
MGPA49	ACTGGCCGACCTACTCTTTC	vbsA	*P. arachidicola*
MGPA50	ACTCACACCCTCACCATGAC	pksA	*P. arachidicola*

MGPA51	CATCGAGATCGGCCCTAAGC	pksA	*P. arachidicola*
MGPA52	AAGTGTCGTGGCCTCATAAC	ver-1	*P. arachidicola*
MGPA53	GGGTCGAACTTGAAGTACAC	ver-1	*P. arachidicola*
MGPA54	CAAAGCTTGCGCCAAGATCC	cypA	*P. arachidicola*
MGPA55	CGAGGGTGCGGATGTAGATG	vbsA	*P. arachidicola*
MGPA56	GACGTAGCCACGGCTGAAAG	vbsA	*P. arachidicola*
MGPA57	TTCTCCCATGTCGTGACTTC	ver-1	*P. arachidicola*
MGPA58	GTACTGGCGAGCAACCAAAG	ver-1	*P. arachidicola*
MGPA59	TAGCTTCGCGCCTTGTCTGC	ver-1	*P. arachidicola*
MGPA60	GAAGCGGTTGTGGGTGTTAC	ver-1	*P. arachidicola*
MGPA61	AATTGAGCGGTGCCAGCATC	ver-1	*P. arachidicola*
MGPA62	GCGCCATTGATGGTCGTGTC	pksA	*P. arachidicola*
MGPA63	AGCCGCTCCTCAAGGAGAAG	pksA	*P. arachidicola*
MGPA64	CCATCTGTGGTGCTTCCTTC	pksA	*P. arachidicola*
MGPA65	GTCTCCAGAATTGCCTTGAC	pksA	*P. arachidicola*
MGPA66	TGAGCCTGTGGCTCCTGTTG	cypA	*P. arachidicola*
MGPA67	TTGAGACCCGGAGTCCGATG	cypA	*P. arachidicola*
MGPA68	TTGGCCGAACGGGTTGAAGG	cypA	*P. arachidicola*
MGPA69	GTACTGTCTGCGCACAACTC	cypA	*P. arachidicola*
MGPA70	TCCTCTCCTGCTCAGCTTTG	ver-1	*P. arachidicola*
MGPA71	CATGCCGATGGTGTAGATAG	ver-1	*P. arachidicola*
MGPA72	GTTGCGGTACTTGAACAGTC	ver-1	*P. arachidicola*
MGPA73	CAGCACGAACTCTTTCGATG	ver-1	*P. arachidicola*
MGPA74	AACCGTCTTCAACAGCTAGG	ver-1	*P. arachidicola*
MGPA75	TTGCGGCAAGGAGCATAGGC	ver-1	*P. arachidicola*
MGPA76	GCCGAAGAGGAACAACAAAG	ver-1	*P. arachidicola*
MGPA77	CCTGCGAGGCTAGTAGAAAG	ver-1	*P. arachidicola*
MGPA78	TAGCCTCGCAGGACCCAATG	ver-1	*P. arachidicola*
MGPA79	ATCGCCGAACACTTCCACTC	ver-1	*P. arachidicola*
MGPA80	TCCTCTTCGGCCTCCAATAC	ver-1	*P. arachidicola*
MGPA81	CGCAAGCCAAGCTACTTCTG	ver-1	*P. arachidicola*
MGPA82	GATCGAGACTCGATGGAATG	ver-1	*P. arachidicola*

Appendix IV Nucleotide and amino acid sequence of *Pa-pksA*

```
      M  A  P  S  N  A  T  R  V  L  V  F  G  D  Q  T  Y  D  F  V
    1 atggcaccctccaacgcaaccagagttctggtctttggggaccagacctacgacttcgtg  60
      P  R  L  R  E  L  F  Q  V  K  D  N  P  I  L  T  A  F  L  Q
   61 cctaggttgcgagaactgttccaggtcaaggacaacccgatcctgactgccttcctgcaa 120
      Q  S  H  Y  V  V
  121 cagtcacattatgtggtgtaggtggttgccactgaccaagcatggaaccgttactgacgc 180
            R  A  Q  M  I  Q  A  L  P  P  A  E  H  K  A  A  R  T
  181 tagctagtcgggcacaaatgatccaggctctgcctccagcggagcacaaggcagcccgaa 240
        F  D  L  A  D  M  L  K  K  Y  V  A  G  K  L  S  P  A  F  Q
  241 ccttcgacttggcagacatgctgaagaagtatgtcgctggcaagttgagccctgccttcc 300
        T  A  L  S  C  I  T  Q  L  G  V  F  M  R  E  F  H  D  F  T
  301 agaccgccctcagctgcattacacaactgggtgtcttcatgcgagaattccacgacttta 360
        K  P  Y  P  R  H  D  S  S  Y  V  L  G  I  C  T  G  S  L  A
  361 ccaagccgtatccacgacacgatagcagctatgtgctgggtatttgcaccggctctctcg 420
        A  A  A  V  S  S  S  S  S  L  S  E  L  L  P  I  A  V  Q  T
  421 ccgccgcagcagtcagctccagcagctcgttgtccgagctcctgcccatcgctgtccaaa 480
        A  L  I  A  F  R  L  G  L  C  V  T  D  M  R  D  R  L  E  S
  481 cagccttgatcgcctttcgcctcggcctgtgcgtcacagacatgcgcgatcgtcttgaaa 540
        A  T  D  D  R  T  E  P  W  S  V  V  L  F  D  T  D  E  Q  T
  541 gcgccacagacgaccgcacagagccttggtcggtagttctcttcgacacagacgagcaga 600
        A  T  K  A  I  K  D  F  C  T  E  N  V  L  P  K  T  K  Q  P
  601 ccgctaccaaggccatcaaggacttctgcaccgagaacgtccttccaaagacgaagcagc 660
        W  I  T  S  A  S  S  K  T  I  T  I  S  G  A  P  R  V  L  K
  661 cttggatcacttccgcctcgtcgaagaccattactatcagcggcgctcctcgcgtgttga 720
        H  L  T  Q  E  P  A  L  K  D  K  K  T  R  Q  I  P  I  Y  V
  721 agcacttgacacaagagcctgcactcaaggacaagaagacccgacagatcccaatctacg 780
        P  A  H  N  S  A  L  F  T  P  E  D  V  K  A  I  L  E  T  T
  781 ttccagcgcacaactcggcgctctttacacctgaagatgtcaaggcaattctggagacga 840
        P  V  E  V  W  S  K  F  P  T  K  I  P  F  I  S  S  V  S  G
  841 ccccagtcgaagtctggagcaaattccccaccaagattcctttcatctccagcgtctctg 900
        K  L  A  W  A  N  S  Y  L  A  V  M  Q  L  A  L  N  Q  C  L
  901 gcaagctggcctgggcaaacagctatctcgccgtcatgcagctggcgctcaaccaatgcc 960
        L  E  P  V  G  W  S  H  V  E  T  E  F  P  R  L  L  K  S  R
  961 tcctggagcccgtcggctggagccacgtcgaaacagagttccctcgacttctgaagtctc 1020
        G  A  H  N  V  L  I  T  P  I  T  T  S  A  D  R  A  L  S  A
 1021 gaggcgctcacaacgtcctcattaccccaatcacgacttccgccgatcgtgccttgtccg 1080
        A  L  S  S  T  I  S  N  I  E  V  E  K  P  S  T  A  E  T  V
 1081 ctgctctcagctcgaccatctccaacatcgaagtggagaagccttctacagctgagactg 1140
        V  H  R  P  G  E  G  K  S  K  L  A  I  V  S  M  S  G  R  F
 1141 tcgtccaccgtcctggcgaagggaagagcaaactcgctatcgtctccatgtccggccgct 1200
        P  D  S  Q  S  T  Q  A  F  W  D  L  L  Y  K  G  L  D  V  V
```

```
1201 tccccgattcccagagcacacaagcattctgggacttgctgtataagggtctcgatgttg 1260
       K  E  V  P  K  R  R  W  D  T  D  T  H  V  D  P  T  G  R  A
1261 tcaaggaagttcccaagcgccgctgggatactgacacgcacgtcgaccccactggtcgtg 1320
       R  N  K  G  A  T  R  W  G  C  W  L  D  F  A  G  E  F  D  P
1321 ctcgcaacaagggtgcgacaagatggggctgctggctcgactttgccggcgagttcgatc 1380
       R  F  F  S  I  S  P  K  E  A  P  Q  M  D  P  A  Q  R  M  A
1381 ctcgcttcttcagtatttcgccgaaggaagcaccacagatggatcccgcccagcgcatgg 1440
       L  M  S  T  Y  E  A  M  E  R  G  G  I  V  P  D  T  T  P  S
1441 ctttgatgtctacatacgaggccatggagcgtggtggtattgtgcccgacacgacacctt 1500
       T  Q  R  N  R  V  G  V  F  H  G  V  T  S  N  D  W  M  E  T
1501 caacgcaacggaaccgcgtcggtgtcttccacggtgttacatccaacgactggatggaga 1560
       N  T  L  A  Q  K  H  R  H  S  Y  F  I  T  G  G  N  R  G  F
1561 ccaacactctagcacagaagcatcgacattcgtacttcatcacgggtggaaacagaggtt 1620
       I  P  G  R  I  N  F  C  F  E  F  S  G  P  S  F  T  N  D  T
1621 tcatcccaggccgtatcaacttctgcttcgagttctctgggccctctttcaccaacgaca 1680
       A  C  S  S  S  L  A  A  I  H  L  A  C  N  S  L  W  R  G  D
1681 cggcttgctcgtcctccctcgcggcgattcacctggcttgtaactcactctggagaggag 1740
       C  D  T  A  V  A  G  G  T  N  M  I  F  T  P  D  G  H  A  G
1741 attgcgatactgctgtagcaggcggaacgaatatgattttcacacctgatggacatgctg 1800
       L  D  K  G  F  F  L  S  R  T  G  N  C  K  P  F  D  D  K  A
1801 gtctggacaagggattcttcttgagcaggactggtaattgcaagcctttcgatgataaag 1860
       D  G  Y  C  R  A  E  G  V  G  T  V  M  I  K  R  L  E  D  A
1861 cggatggatactgtcgtgctgaaggtgttggcaccgtcatgatcaagaggctggaagatg 1920
       L  A  D  G  D  P  I  L  G  T  I  L  D  A  K  T  N  H  S  A
1921 ctcttgcagatggcgatcctatccttggcacgattctggatgcgaagacgaatcattcgg 1980
       M  S  D  S  M  T  R  P  F  V  P  A  Q  I  D  N  M  E  A  C
1981 ctatgtcggattctatgacgaggccgtttgtgccggcgcagattgacaacatggaagctt 2040
       L  S  T  A  G  V  D  P  T  S  L  D  Y  I  E  M  H  G  T  G
2041 gtctcagcaccgctggcgtagacccgacatctctcgattacatcgagatgcacggtactg 2100
       T  Q  V  G  D  A  V  E  M  E  S  V  L  S  V  F  A  P  H  E
2101 ggacgcaagtcggtgatgccgtggagatggagagtgttctctctgtctttgcaccgcacg 2160
       K  F  R  T  K  E  Q  P  L  Y  V  G  S  A  K  A  N  I  G  H
2161 agaagtttaggaccaaggagcagccgctttatgttgggtctgcgaaagccaacattggtc 2220
       G  E  G  V  S  G  V  T  S  L  I  K  V  L  L  M  L  Q  N  N
2221 atggtgagggtgtgagtggtgttacgagtctcatcaaagtgttgttgatgctgcagaata 2280
       T  I  P  P  H  C  G  I  K  P  G  S  K  I  N  H  N  Y  P  D
2281 acaccatccctcctcactgcggcatcaagcctggcagcaagatcaaccacaattacccag 2340
       L  A  A  R  N  V  H  I  A  F  E  P  K  P  F  L  R  R  E  G
2341 accttgcggcgagaaatgtgcacatcgcattcgagccgaaaccattcttgaggagggaag 2400
       K  L  R  R  V  L  I  N  N  F  S  A  A  G  G  N  T  A  L  L
2401 gcaagttgaggagagtgttgattaacaatttctctgccgcgggtggcaataccgctctcc 2460
       I  E  D  A  P  D  R  A  P  I  Q  T  K  D  P  R  T  T  Q  T
2461 tcatcgaagatgcgccagaccgcgcaccaatccagaccaaggatcccaggacgacgcaga 2520
       V  T  V  S  G  H  V  G  K  S  L  S  N  N  V  A  N  L  L  S
```

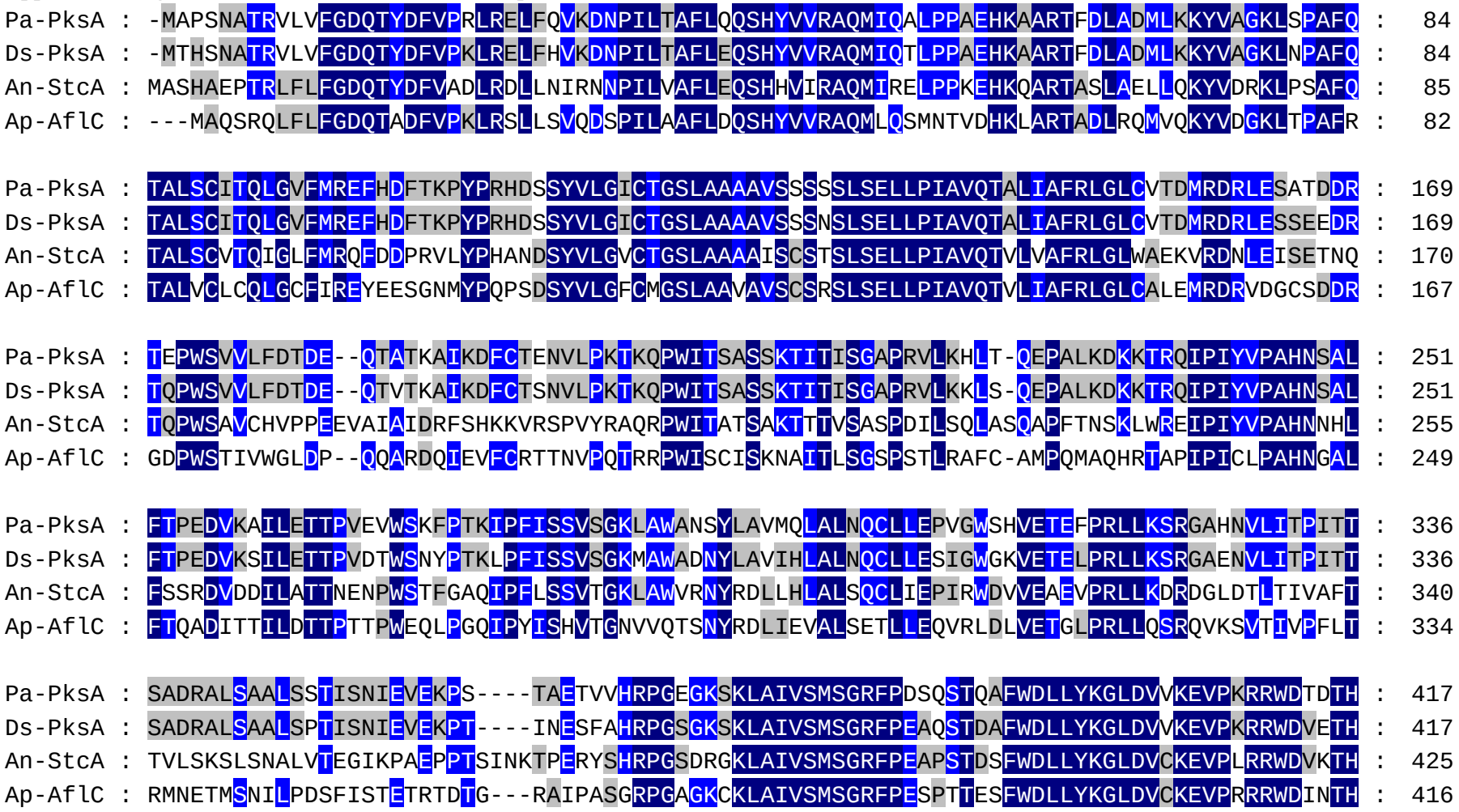
```
2521 ctgtcacagtctctggacacgttggaaagagcttgtcgaacaacgtcgccaacctcctct 2580
       H  L  K  S  N  P  T  I  D  L  P  H  L  S  Y  T  T  T  A  R
2581 cgcatctcaagtcgaacccaacaatcgacctcccccacctctcctacaccacaacagccc 2640
       R  W  H  H  L  H  R  V  A  V  S  G  T  S  I  A  E  I  T  Q
2641 gccgttggcatcatctccatcgcgtcgccgtctccggcacctcaatcgccgagatcacgc 2700
       K  L  E  K  A  V  E  N  K  D  G  V  N  R  P  K  A  K  P  G
2701 agaagctcgaaaaagccgtcgagaacaaagacggcgtcaacagacccaaggccaaacctg 2760
       V  F  F  A  F  T  G  Q  G  S  Q  Y  L  G  M  G  K  Q  L  Y
2761 gcgtcttcttcgccttcacaggccaaggatcacagtatcttggcatgggcaagcaactct 2820
       D  A  Y  P  K  F  K  F  E  L  Q  R  Y  S  Q  L  A  V  S  H
2821 acgacgcgtaccccaaattcaagtttgaacttcagcgatattcgcagctcgctgtatcgc 2880
       G  F  P  S  F  L  H  I  F  S  E  T  K  G  D  V  E  Q  N  L
2881 atgggttcccgtccttcctgcacatcttcagcgagacgaagggagatgtcgagcagaatc 2940
       P  V  V  V  Q  L  A  I  T  C  L  Q  M  A  L  F  N  L  I  T
2941 taccagttgtggtgcagctcgctattacctgtctgcaaatggcactcttcaacctcatta 3000
       S  F  G  I  K  A  S  A  V  V  G  H  S  L  G  E  Y  A  A  L
3001 cttcctttgggatcaaggcctctgctgtagttgggcacagtctgggcgagtacgctgcgc 3060
       H  A  A  G  V  L  S  A  S  D  T  I  Y  L  V  G  K  R  A  E
3061 ttcacgccgctggcgtgctttctgcgagtgacacgatctacctcgtcggcaagagggcgg 3120
       L  L  Q  E  R  C  Q  R  G  T  H  A  M  L  A  C  K  A  S  E
3121 agctcttgcaagagcgctgtcagcgcggcacgcatgcgatgcttgcgtgcaaggcgagtg 3180
       W  S  L  A  G  I  T  T  R  K  E  V  E  V  A  C  V  N  G  P
3181 agtggtccctcgcgggaatcaccacacgcaaggaagtggaggtcgcctgtgtcaatggac 3240
       E  D  T  V  L  S  G  T  V  E  E  I  V  E  V  Q  K  T  L  S
3241 cggaagatacggttctttcagggacggttgaggaaatcgtagaggtgcagaagactctca 3300
       G  K  G  I  K  A  T  T  L  K  L  P  F  A  F  H  S  A  Q  V
3301 gcggtaagggcatcaaggccacgacgctgaagttaccatttgcattccattctgcgcagg 3360
       Q  P  I  L  E  D  F  E  R  L  A  S  G  A  T  F  E  K  P  K
3361 tgcagcctattcttgaggacttcgagcgtttggctagcggagctacgttcgagaagccga 3420
       M  A  V  L  S  P  L  L  G  S  L  V  E  D  E  G  I  I  G  P
3421 agatggctgtgctgtctccgctgcttgggagtttggttgaggatgagggcatcatcggcc 3480
       A  Y  L  A  R  H  C  R  E  A  V  G  M  V  K  A  L  E  A  A
3481 ctgcctatctcgcacgccactgtcgcgaggcagttggcatggtcaaggctctcgaggcag 3540
       K  E  K  G  T  I  N  D  K  T  I  V  I  E  I  G  P  K  P  L
3541 caaaggagaagggcacaatcaacgacaagaccatcgtcatcgagatcggccctaagccac 3600
       L  S  G  M  I  K  N  I  L  G  Q  S  M  T  T  L  P  T  L  K
3601 tcctcagcggcatgatcaagaatatccttggacagagcatgacaacgctgcccacgttga 3660
       E  K  G  P  D  V  W  T  N  L  S  N  I  F  S  T  L  Y  T  G
3661 aggaaaagggccccgatgtctggacgaacctttcgaacattttctcaacactctacactg 3720
       G  L  D  I  N  W  T  A  F  H  A  P  I  E  G  V  K  K  V  I
3721 gcggtctagacataaactggactgcattccacgctcccatcgagggtgtgaagaaggtca 3780
       Q  L  P  D  Y  A  W  D  L  K  D  Y  F  I  Q  Y  E  G  D  W
3781 ttcaactcccggactacgcctgggatctgaaagattacttcatccagtatgagggtgatt 3840
       V  L  H  R  H  K  I  H  C  N  C  A  D  I  G  K  D  V  H  D
```

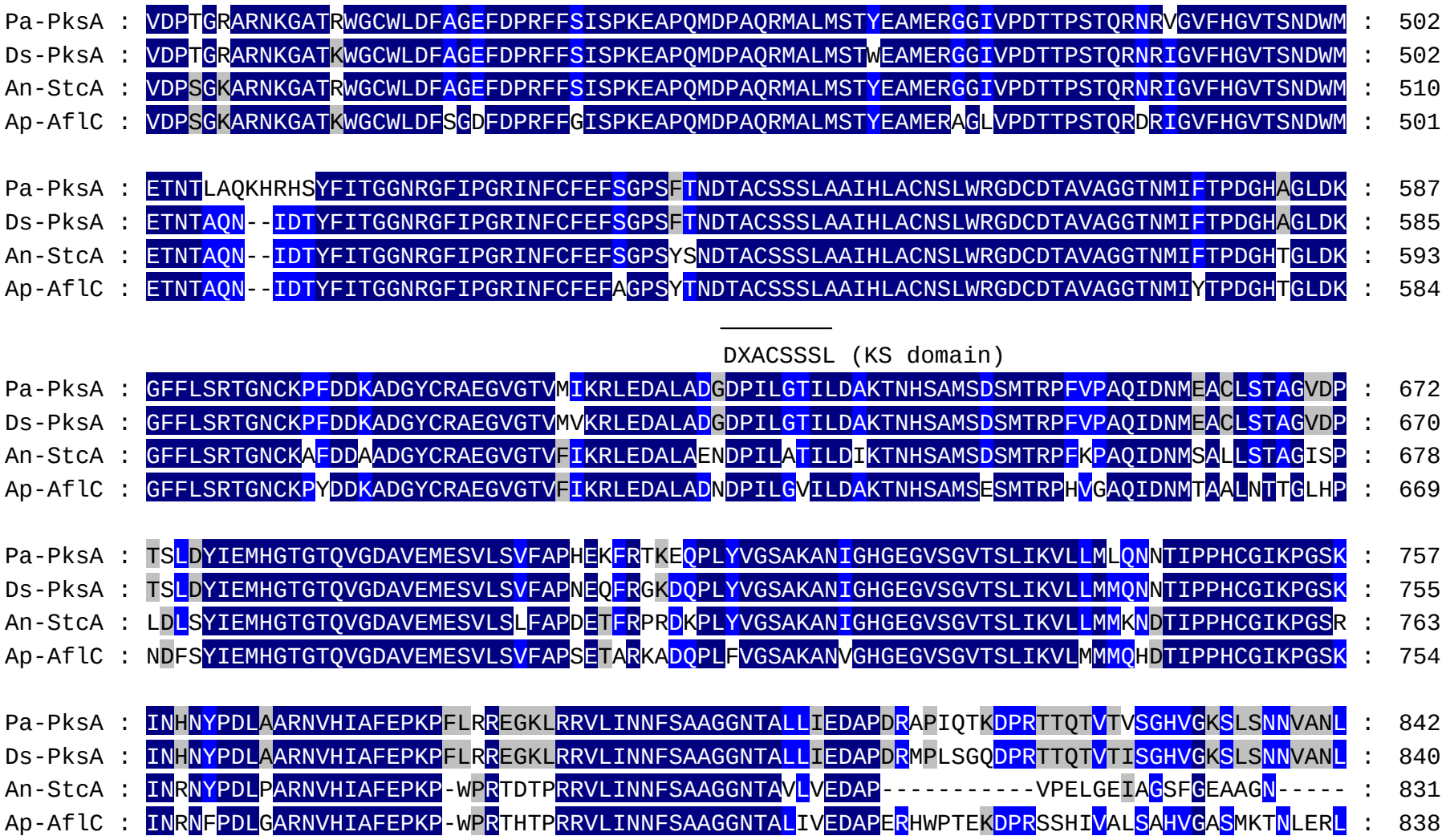
3841 gggttctgcaccgacacaagattcattgtaattgtgcggatattgggaaggatgtgcatg 3900
T S H Y C P G K H T F V E N V V V P G G
3901 atacttcgcactactgtcccgggaagcatactttcgtggagaatgttgttgttcctggtg 3960
G P Q A I K A V P E K P A K K M S K L D
3961 gagggccgcaggccatcaaggctgttccggagaagcctgcgaagaagatgtcgaagttgg 4020
P T K E A Y P S I P L T T T I H K V I E
4021 atcctactaaggaggcgtacccgagtatccctctgacgacgacgatacacaaggtcattg 4080
E K T E P L G A Q F T V E T D I S R K D
4081 aggagaagacggagccgctgggtgcacaatttacggttgagactgatatttcccgtaagg 4140
V N S I A Q G H T V D D I P L C T P S F
4141 atgtcaacagcattgcacagggacacactgtcgatgacattccgctctgcacaccttcct 4200
Y A D I A L Q V G K Y A M D R I R A G H
4201 tctacgcagacatcgcgctacaagttggaaagtacgccatggaccgtatacgcgccggac 4260
P G A G A I D G R V D V T D L V V D K A
4261 accctggcgccggcgccattgatggtcgtgtcgatgtcacagacctcgttgtcgacaagg 4320
L I P H G K A P Q L L R T N V T M S W P
4321 ctctcatcccacatggcaaagctccgcaacttctccggacgaacgtcaccatgtcctggc 4380
P K M A A T T R S A K V T F K T Y T A D
4381 ccccgaagatggcagccaccaccaggtccgcgaaagtcacgttcaagacatatacagctg 4440
G K L D T D H A Y C T V R F T T D A Q Q
4441 acggaaagcttgatacggaccacgcctattgcaccgtccgcttcacaaccgatgcgcagc 4500
K S L Q K K V P E Y Q A A I A S L R A R
4501 agaaatccctgcagaagaaggtacccgagtaccaagccgcgattgcatcgctccgagctc 4560
V K K G E L V H Y N T K S G Y K L M S S
4561 gcgtcaagaagggcgaacttgtccactacaacaccaagtccggctacaagctcatgtctt 4620
M A H F H P D Y K L L N N L I L N E A E
4621 ccatggcccactttcacccagactacaagctcctcaacaacctcatcctgaacgaagcgg 4680
N E A V S V M N F S T C T D A G T Y A A
4681 agaacgaggcagtcagcgttatgaacttctctacctgcaccgatgcagggacttacgccg 4740
H P A Y I D A I T Q V G G F A M N A K D
4741 cacacccggcatacatcgacgccatcacgcaggtgggaggcttcgcaatgaatgccaagg 4800
D T D I D K E V Y V N H G W E S F Q V Y
4801 atgacaccgatatcgacaaggaggtgtatgtcaaccatggatgggagtctttccaggtgt 4860
K P L L K E K Q Y V V Y S K M V K D S K
4861 ataagccgctcctcaaggagaagcagtatgttgtatactccaagatggtcaaagactcga 4920
G D L V H G D T I V L D G D E V V A F F
4921 agggagatctcgtccatggcgacactattgtccttgatggagatgaggtcgttgcgttct 4980
R G L S L R S V P R K A L R A V L Q S A
4981 ttaggggtttgtcactgagaagtgtgcccaggaaagcgcttcgtgcggttctccagagtg 5040
M D K G I R Q R G G K P G A A K G T A A
5041 cgatggataaggggatccgacagagaggtgggaagccgggtgctgcgaagggtactgctg 5100
A P V A K K P A A P V A T A P V K A A P
5101 cagcgcccgtcgcgaagaagcctgctgccccggttgccactgctccagtgaaagctgcgc 5160
V A A A P S S S K P A P P E G P K A A A

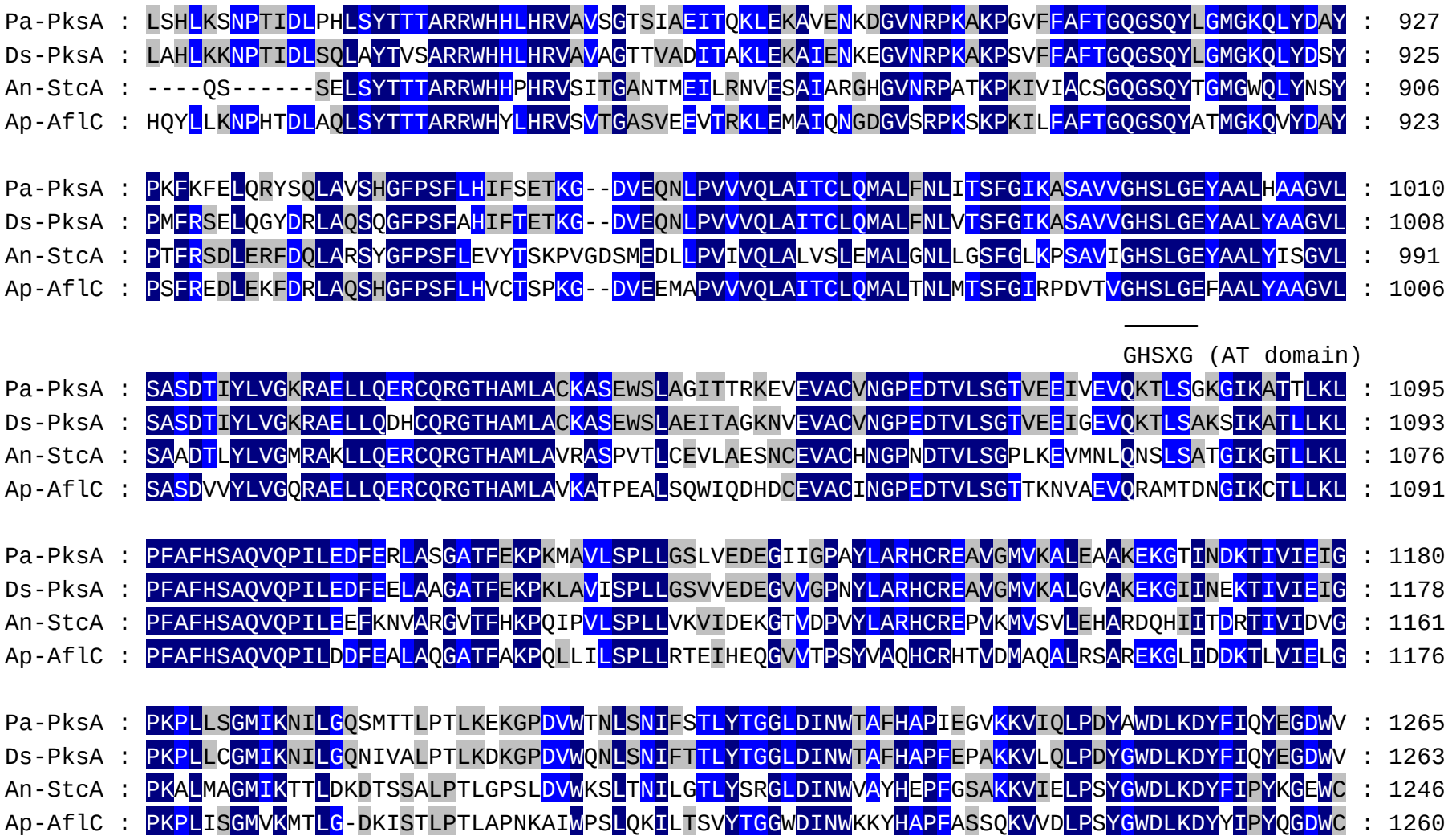
```
5161 cggtcgctgctgcaccatcgtcatcgaagcctgcacctcctgaaggcccgaaagctgcgg 5220
       S  K  V  V  A  K  A  D  S  G  K  I  D  E  A  L  K  I  I  S
5221 cttcgaaggtcgtggcgaaggcggatagtggtaagatcgacgaggcgttgaaaatcattt 5280
       E  E  S  G  I  A  L  E  E  L  T  D  D  S  N  F  T
5281 cggaggagagtggtattgcgttggaggagttgacggatgatagtaacttcactgtgagtc 5340
                                                    D  M  G  V  D
5341 gctccctctccactcaatgatattgaggatcctaacacttcgcaggatatgggtgtcgac 5400
     S  L  S  S  M  V  I  T  S  R  L  R  E  D  L  E  L  D  L  A
5401 agcttgagctctatggtgatcacctcccgtctccgcgaggatctagaactggaccttgcc 5460
     P  E  F  A  L  F  A  D  C  P  T  V  A  S  L  R  E  F
5461 cctgagtttgcgctcttcgctgactgccctaccgtcgccagcttgagggaattc       5514
```

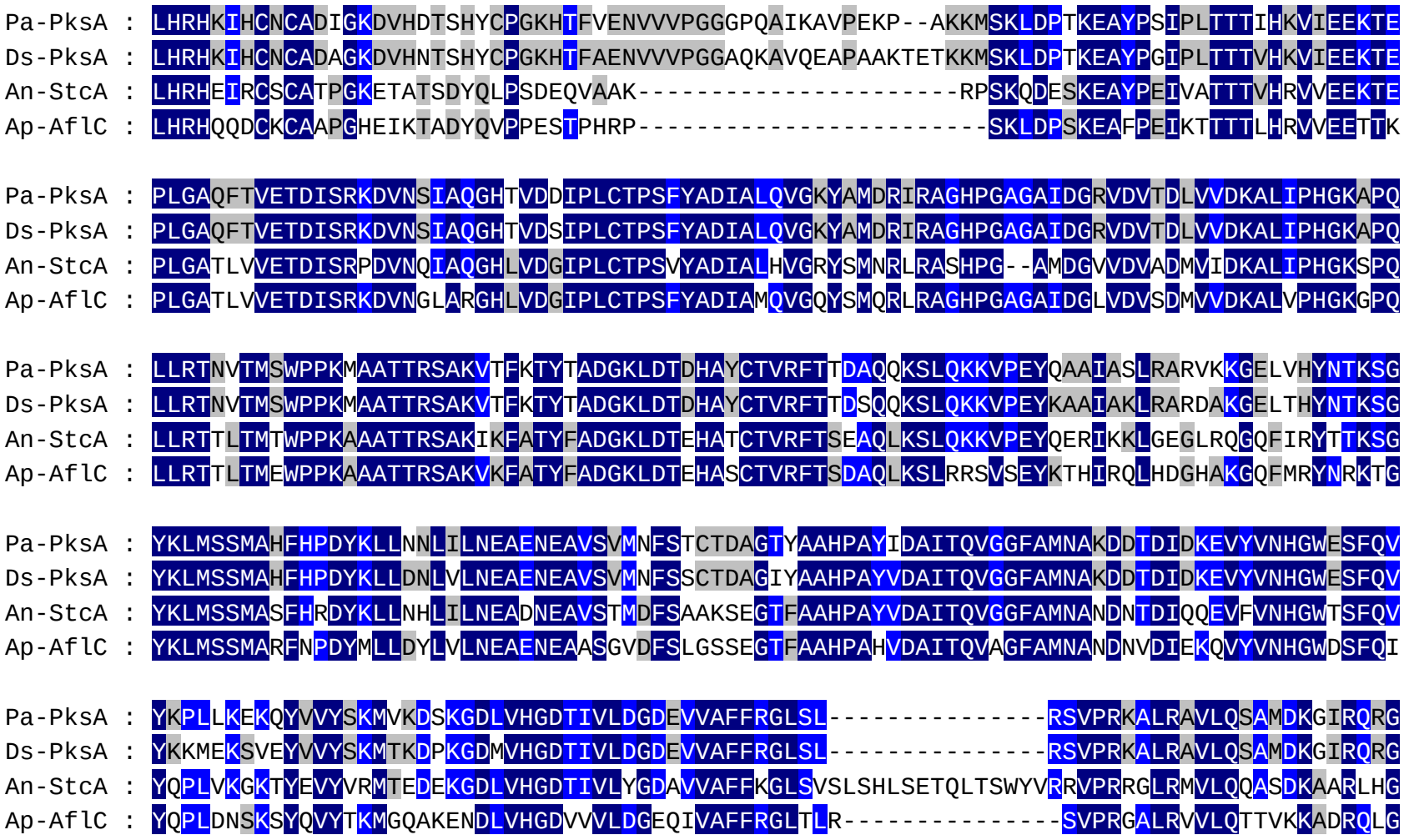
Appendix IV The partial ORF and amino acid sequence of *Pa-pksA*. The red nucleotide sequence indicates the positions of two introns. The blue amino acid sequences indicates the positions of the KS, AT and ACP domain regions, which were identified by alignment of Pa-PksA with Ds-PksA and other Pks proteins as shown in Figure 3.20.

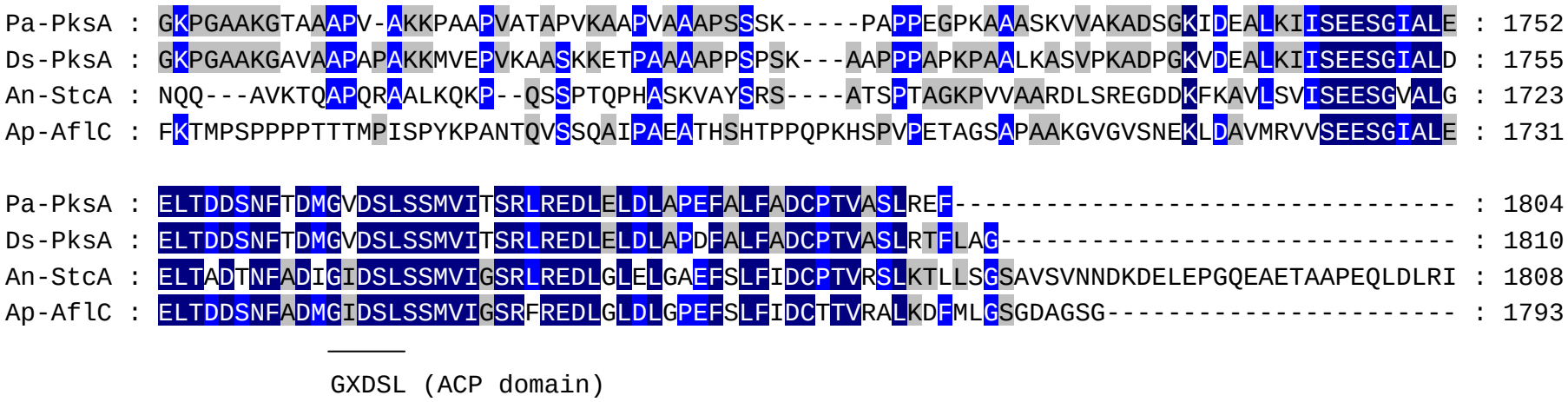
Appendix V Alignment of Pa-PksA with other PKSA proteins




Pa-PksA : VDPTGRARNKGATRWGCWLDFAGEFDPRFFSISPKEAPQMDPAQRMALMSTYEAMERGGIVPDTTPSTQRNRVGVFHGVTSNDWM : 502
Ds-PksA : VDPTGRARNKGATKWGCWLDFAGEFDPRFFSISPKEAPQMDPAQRMALMSTWEAMERGGIVPDTTPSTQRNRIGVFHGVTSNDWM : 502
An-StcA : VDPSGKARNKGATRWGCWLDFAGEFDPRFFSISPKEAPQMDPAQRMALMSTYEAMERGGIVPDTTPSTQRNRIGVFHGVTSNDWM : 510
Ap-AflC : VDPSGKARNKGATKWGCWLDFSGDFDPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRDRIGVFHGVTSNDWM : 501
Pa-PksA : ETNTLAQKHRHSYFITGGNRGFIPGRINFCFEFSGPSFTNDTACSSSLAAIHLACNSLWRGDCDTAVAGGTNMIFTPDGHAGLDK : 587
Ds-PksA : ETNTAQN--IDTYFITGGNRGFIPGRINFCFEFSGPSFTNDTACSSSLAAIHLACNSLWRGDCDTAVAGGTNMIFTPDGHAGLDK : 585
An-StcA : ETNTAQN--IDTYFITGGNRGFIPGRINFCFEFSGPSYSNDTACSSSLAAIHLACNSLWRGDCDTAVAGGTNMIFTPDGHTGLDK : 593
Ap-AflC : ETNTAQN--IDTYFITGGNRGFIPGRINFCFEFAGPSYTNDTACSSSLAAIHLACNSLWRGDCDTAVAGGTNMIYTPDGHTGLDK : 584
DXACSSSL (KS domain)
Pa-PksA : GFFLSRTGNCKPFDDKADGYCRAEGVGTVMIKRLEDALADGDPILGTILDAKTNHSAMSDSMTRPFVPAQIDNMEACLSTAGVDP : 672
Ds-PksA : GFFLSRTGNCKPFDDKADGYCRAEGVGTVMVKRLEDALADGDPILGTILDAKTNHSAMSDSMTRPFVPAQIDNMEACLSTAGVDP : 670
An-StcA : GFFLSRTGNCKAFDDAADGYCRAEGVGTVFIKRLEDALAENDPILATILDIKTNHSAMSDSMTRPFKPAQIDNMSALLSTAGISP : 678
Ap-AflC : GFFLSRTGNCKPYDDKADGYCRAEGVGTVFIKRLEDALADNDPILGVILDAKTNHSAMSESMTRPHVGAQIDNMTAALNTTGLHP : 669
Pa-PksA : TSLDYIEMHGTGTQVGDAVEMESVLSVFAPHEKFRTKEQPLYVGSAKANIGHGEGVSGVTSLIKVLLMLQNNTIPPHCGIKPGSK : 757
Ds-PksA : TSLDYIEMHGTGTQVGDAVEMESVLSVFAPNEQFRGKDQPLYVGSAKANIGHGEGVSGVTSLIKVLLMMQNNTIPPHCGIKPGSK : 755
An-StcA : LDLSYIEMHGTGTQVGDAVEMESVLSLFAPDETFRPRDKPLYVGSAKANIGHGEGVSGVTSLIKVLLMMKNDTIPPHCGIKPGSR : 763
Ap-AflC : NDFSYIEMHGTGTQVGDAVEMESVLSVFAPSETARKADQPLFVGSAKANVGHGEGVSGVTSLIKVLMMMQHDTIPPHCGIKPGSK : 754
Pa-PksA : INHNYPDLAARNVHIAFEPKPFLRREGKLRRVLINNFSAAGGNTALLIEDAPDRAPIQTKDPRTTQTVTVSGHVGKSLSNNVANL : 842
Ds-PksA : INHNYPDLAARNVHIAFEPKPFLRREGKLRRVLINNFSAAGGNTALLIEDAPDRMPLSGQDPRTTQTVTISGHVGKSLSNNVANL : 840
An-StcA : INRNYPDLPARNVHIAFEPKP-WPRTDTPRRVLINNFSAAGGNTAVLVEDAP-----------VPELGEIAGSFGEAAGN----- : 831
Ap-AflC : INRNFPDLGARNVHIAFEPKP-WPRTHTPRRVLINNFSAAGGNTALIVEDAPERHWPTEKDPRSSHIVALSAHVGASMKTNLERL : 838


Pa-PksA : LSHLKSNPTIDLPHLSYTTTARRWHHLHRVAVSGTSIAEITQKLEKAVENKDGVNRPKAKPGVFFAFTGQGSQYLGMGKQLYDAY : 927
Ds-PksA : LAHLKKNPTIDLSQLAYTVSARRWHHLHRVAVAGTTVADITAKLEKAIENKEGVNRPKAKPSVFFAFTGQGSQYLGMGKQLYDSY : 925
An-StcA : ----QS------SELSYTTTARRWHHPHRVSITGANTMEILRNVESAIARGHGVNRPATKPKIVIACSGQGSQYTGMGWQLYNSY : 906
Ap-AflC : HQYLLKNPHTDLAQLSYTTTARRWHYLHRVSVTGASVEEVTRKLEMAIQNGDGVSRPKSKPKILFAFTGQGSQYATMGKQVYDAY : 923
Pa-PksA : PKFKFELQRYSQLAVSHGFPSFLHIFSETKG--DVEQNLPVVVQLAITCLQMALFNLITSFGIKASAVVGHSLGEYAALHAAGVL : 1010
Ds-PksA : PMFRSELQGYDRLAQSQGFPSFAHIFTETKG--DVEQNLPVVVQLAITCLQMALFNLVTSFGIKASAVVGHSLGEYAALYAAGVL : 1008
An-StcA : PTFRSDLERFDQLARSYGFPSFLEVYTSKPVGDSMEDLLPVIVQLALVSLEMALGNLLGSFGLKPSAVIGHSLGEYAALYISGVL : 991
Ap-AflC : PSFREDLEKFDRLAQSHGFPSFLHVCTSPKG--DVEEMAPVVVQLAITCLQMALTNLMTSFGIRPDVTVGHSLGEFAALYAAGVL : 1006
GHSXG (AT domain)
Pa-PksA : SASDTIYLVGKRAELLQERCQRGTHAMLACKASEWSLAGITTRKEVEVACVNGPEDTVLSGTVEEIVEVQKTLSGKGIKATTLKL : 1095
Ds-PksA : SASDTIYLVGKRAELLQDHCQRGTHAMLACKASEWSLAEITAGKNVEVACVNGPEDTVLSGTVEEIGEVQKTLSAKSIKATLLKL : 1093
An-StcA : SAADTLYLVGMRAKLLQERCQRGTHAMLAVRASPVTLCEVLAESNCEVACHNGPNDTVLSGPLKEVMNLQNSLSATGIKGTLLKL : 1076
Ap-AflC : SASDVVYLVGQRAELLQERCQRGTHAMLAVKATPEALSQWIQDHDCEVACINGPEDTVLSGTTKNVAEVQRAMTDNGIKCTLLKL : 1091
Pa-PksA : PFAFHSAQVQPILEDFERLASGATFEKPKMAVLSPLLGSLVEDEGIIGPAYLARHCREAVGMVKALEAAKEKGTINDKTIVIEIG : 1180
Ds-PksA : PFAFHSAQVQPILEDFEELAAGATFEKPKLAVISPLLGSVVEDEGVVGPNYLARHCREAVGMVKALGVAKEKGIINEKTIVIEIG : 1178
An-StcA : PFAFHSAQVQPILEEFKNVARGVTFHKPQIPVLSPLLVKVIDEKGTVDPVYLARHCREPVKMVSVLEHARDQHIITDRTIVIDVG : 1161
Ap-AflC : PFAFHSAQVQPILDDFEALAQGATFAKPQLLILSPLLRTEIHEQGVVTPSYVAQHCRHTVDMAQALRSAREKGLIDDKTLVIELG : 1176
Pa-PksA : PKPLLSGMIKNILGQSMTTLPTLKEKGPDVWTNLSNIFSTLYTGGLDINWTAFHAPIEGVKKVIQLPDYAWDLKDYFIQYEGDWV : 1265
Ds-PksA : PKPLLCGMIKNILGQNIVALPTLKDKGPDVWQNLSNIFTTLYTGGLDINWTAFHAPFEPAKKVLQLPDYGWDLKDYFIQYEGDWV : 1263
An-StcA : PKALMAGMIKTTLDKDTSSALPTLGPSLDVWKSLTNILGTLYSRGLDINWVAYHEPFGSAKKVIELPSYGWDLKDYFIPYKGEWC : 1246
Ap-AflC : PKPLISGMVKMTLG-DKISTLPTLAPNKAIWPSLQKILTSVYTGGWDINWKKYHAPFASSQKVVDLPSYGWDLKDYYIPYQGDWC : 1260


Pa-PksA : LHRHKIHCNCADIGKDVHDTSHYCPGKHTFVENVVVPGGGPQAIKAVPEKP--AKKMSKLDPTKEAYPSIPLTTTIHKVIEEKTE : 1348
Ds-PksA : LHRHKIHCNCADAGKDVHNTSHYCPGKHTFAENVVVPGGAQKAVQEAPAAKTETKKMSKLDPTKEAYPGIPLTTTVHKVIEEKTE : 1348
An-StcA : LHRHEIRCSCATPGKETATSDYQLPSDEQVAAK----------RPSKQDESKEAYPEIVATTTVHRVVEEKTE : 1309
Ap-AflC : LHRHQQDCKCAAPGHEIKTADYQVPPESTPHRP----------SKLDPSKEAFPEIKTTTTLHRVVEETTK : 1321
Pa-PksA : PLGAQFTVETDISRKDVNSIAQGHTVDDIPLCTPSFYADIALQVGKYAMDRIRAGHPGAGAIDGRVDVTDLVVDKALIPHGKAPQ : 1433
Ds-PksA : PLGAQFTVETDISRKDVNSIAQGHTVDSIPLCTPSFYADIALQVGKYAMDRIRAGHPGAGAIDGRVDVTDLVVDKALIPHGKAPQ : 1433
An-StcA : PLGATLVVETDISRPDVNQIAQGHLVDGIPLCTPSVYADIALHVGRYSMNRLRASHPG--AMDGVVDVADMVIDKALIPHGKSPQ : 1392
Ap-AflC : PLGATLVVETDISRKDVNGLARGHLVDGIPLCTPSFYADIAMQVGQYSMQRLRAGHPGAGAIDGLVDVSDMVVDKALVPHGKGPQ : 1406
Pa-PksA : LLRTNVTMSWPPKMAATTRSAKVTFKTYTADGKLDTDHAYCTVRFTTDAQQKSLQKKVPEYQAAIASLRARVKKGELVHYNTKSG : 1518
Ds-PksA : LLRTNVTMSWPPKMAATTRSAKVTFKTYTADGKLDTDHAYCTVRFTTDSQQKSLQKKVPEYKAAIAKLRARDAKGELTHYNTKSG : 1518
An-StcA : LLRTTLTMTWPPKAAATTRSAKIKFATYFADGKLDTEHATCTVRFTSEAQLKSLQKKVPEYQERIKKLGEGLRQGQFIRYTTKSG : 1477
Ap-AflC : LLRTTLTMEWPPKAAATTRSAKVKFATYFADGKLDTEHASCTVRFTSDAQLKSLRRSVSEYKTHIRQLHDGHAKGQFMRYNRKTG : 1491
Pa-PksA : YKLMSSMAHFHPDYKLLNNLILNEAENEAVSVMNFSTCTDAGTYAAHPAYIDAITQVGGFAMNAKDDTDIDKEVYVNHGWESFQV : 1603
Ds-PksA : YKLMSSMAHFHPDYKLLDNLVLNEAENEAVSVMNFSSCTDAGIYAAHPAYVDAITQVGGFAMNAKDDTDIDKEVYVNHGWESFQV : 1603
An-StcA : YKLMSSMASFHRDYKLLNHLILNEADNEAVSTMDFSAAKSEGTFAAHPAYVDAITQVGGFAMNANDNTDIQQEVFVNHGWTSFQV : 1562
Ap-AflC : YKLMSSMARFNPDYMLLDYLVLNEAENEAASGVDFSLGSSEGTFAAHPAHVDAITQVAGFAMNANDNVDIEKQVYVNHGWDSFQI : 1576
Pa-PksA : YKPLLKEKQYVVYSKMVKDSKGDLVHGDTIVLDGDEVVAFFRGLSL----------RSVPRKALRAVLQSAMDKGIRQRG : 1673
Ds-PksA : YKKMEKSVEYVVYSKMTKDPKGDMVHGDTIVLDGDEVVAFFRGLSL----------RSVPRKALRAVLQSAMDKGIRQRG : 1673
An-StcA : YQPLVKGKTYEVYVRMTEDEKGDLVHGDTIVLYGDAVVAFFKGLSVSLSHLSETQLTSWYVRRVPRRGLRMVLQQASDKAARLHG : 1647
Ap-AflC : YQPLDNSKSYQVYTKMGQAKENDLVHGDVVVLDGEQIVAFFRGLTLR----------SVPRGALRVVLQTTVKKADRQLG : 1646



Name	organism	Protein	Genbank No.
Pa-PksA	*P. arachidicola*	PksA	
Ds-PksA	*D. septosporum*	PksA	AAZ95017
Ap-AflC	*A. parasiticus*	AflC	AAS66004
An-StcA	*A. nidulans*	StcA	AAA81586

Appendix V Alignment of Pa-PksA with other PKSA proteins. The black lines indicate the core regions of the KS, AT and ACP domains.

Appendix VI Nucleotide and amino acid sequence of *Pa-cypA*

```
        M  A  G  E  L  Y  K  W  I  M  D  T  T  A  G  A  P  L  P  F
      1 atggcgggagagctttacaagtggatcatggacaccactgctggtgctccattgccattc 60
        S  L  A  L  V  V  G  A  F  I  L  Y  N  I  V  S
     61 tcgctggcgcttgttgtcggtgcctttattctgtataacattgtttccgtgagtggaagc 120
                                          I  I  T  T  A  Y  F  S
    121 cgaccaggtgttgggaacgtcgctaacagatattgcagatcatcacgacagcatacttct 180
        P  L  S  A  I  P  G  P  W  Y  A  K  L  T  D  A  R  L  T  Y
    181 ctccgctcagcgcgattcccggaccgtggtacgcgaagttgacagacgctcggttgacgt 240
        S  V  F  A  G  N  R  I  Y  Y  V  D  S  L  H  R  K  Y  G  P
    241 actcggtcttcgctggtaatcgtatctattatgtcgattcattacaccgcaagtatggcc 300
        M  V  R  I  G  P  K  E  V  D  V  A  D  P  A  A  A  R  E  V
    301 ctatggtgcgtatcggtccaaaggaggtcgatgtggcagatccagcggctgcacgagagg 360
        H  R  M  G  T  V  F  T  K  A  P  F  Y  R  L  L  S  P  G  P
    361 tccatcggatgggcacagtgttcaccaaagcgccgttctaccgtctgctgtctcccggtc 420
        V  D  N  I  F  N  F  R  D  Q  K  K  H  S  Q  R  R  K  L  Y
    421 cagttgacaacatcttcaacttccgcgaccagaagaagcacagccagcgccgcaagctgt 480
        A  K  G  F  T  L  V  E  L  R  R  N  W  E  G  T  I  N  Q  T
    481 acgccaagggcttcactcttgtcgagctgagaagaaactgggaaggcacgatcaaccaga 540
        I  R  M  A  V  E  K  M  K  E  E  A  G  N  G  N  T  E  L  M
    541 cgatccgcatggccgtggagaagatgaaggaagaggcgggcaatggcaacaccgaactca 600
        G  W  W  T  L  M  A  N  E  V  V  C  R  L  T  F  N  G  G  H
    601 tggggtggtggacactgatggccaacgaggtcgtgtgccgcctgacgttcaacgggggcc 660
        G  T  V  E  K  G  I  K  D  P  F  V  L  M  L  E  K  R  K  G
    661 acggaactgtcgagaagggcatcaaggatccattcgtcctcatgctggagaagcggaagg 720
        D  L  A  H  L  L  K  M  F  V  P  P  L  Y  Y  V  G  R  V  L
    721 gcgatcttgcacatctcctgaaaatgtttgtccctcctctctactacgtcggccgtgtgc 780
        G  R  V  N  T  R  M  N  D  I  F  Y  S  Q  E  K  M  F  Q  A
    781 tcggaagggtcaacacgcgcatgaacgacatcttctactcgcaggagaagatgttccagg 840
        G  A  G  V  V  K  S  A  R  Q  D  K  E  A  G  E  F  N  Q  N
    841 cgggggctggtgtggtgaagagcgcccgtcaggataaagaggctggagagttcaaccaga 900
        L  F  A  K  A  L  Q  E  G  E  G  D  A  A  T  L  T  D  T  D
    901 acctgttcgccaaggcgctgcaggagggcgaaggcgatgctgcgactctgactgacacgg 960
        I  I  T  D  A  G  A  L  L  L  A  G  S  D  P  T  A  I  S  L
    961 acatcatcacggacgctggtgctctgttgctggcgggctctgacccgacggccatctcgt 1020
        T  F  L  I  Y  L  V  L  S  R  P  E  L  Q  T  Q  L  E  E  E
   1021 tgacgttcttgatctatctcgtcctgagcagaccggagctccagacgcagctggaggagg 1080
        V  A  S  I  D  G  E  V  T  D  T  A  C  E  G  L  P  L  M  N
   1081 aagtcgcatccatcgatggagaggtgactgacacggcatgtgaaggcctgccgctcatga 1140
        A  V  I  D  E  S  M  R  L  Y  G  A  A  P  G  C  L  P  R  S
   1141 acgccgtcatcgacgagagcatgcggctgtacggtgctgccccggctgtcttcctcgga 1200
        P  P  A  G  G  A  K  L  G  G  Y  Y  I  P  A  G  T  V  V  D
   1201 gtccgccggcgggaggggccaagcttggcggctactacatccccgcgggcaccgtggtcg 1260
        T  Q  N  Y  T  L  H  T  D  A  V  T  W  K  D  A  Q  T
```

1261 acacgcagaactatacgctacacactgacgctgtcacgtggaaggacgcgcaaacgtgag 1320

1321 tcacatgtgtcctgtctgtttttgcaggtgtttgcaggagttgtgtcgagatgttgtgct 1380

F D H T R F L P E N R L

1381 gacgtgaatatcactgtattccagattcgaccacacgcgcttcctgcccgaaaaccgatt 1440

A F S E R Q K M A F N P F G Q G S R Q C

1441 ggccttctccgagcggcagaagatggccttcaacccgttcggccaaggttcgagacagtg 1500

L G I H L G K L E M R L A V A H F F R E

1501 tttgggtatccatctgggtaagttggagatgcgcctcgccgtcgctcacttcttccgcga 1560

L R G V T L A K S A T P A S M T V V D S

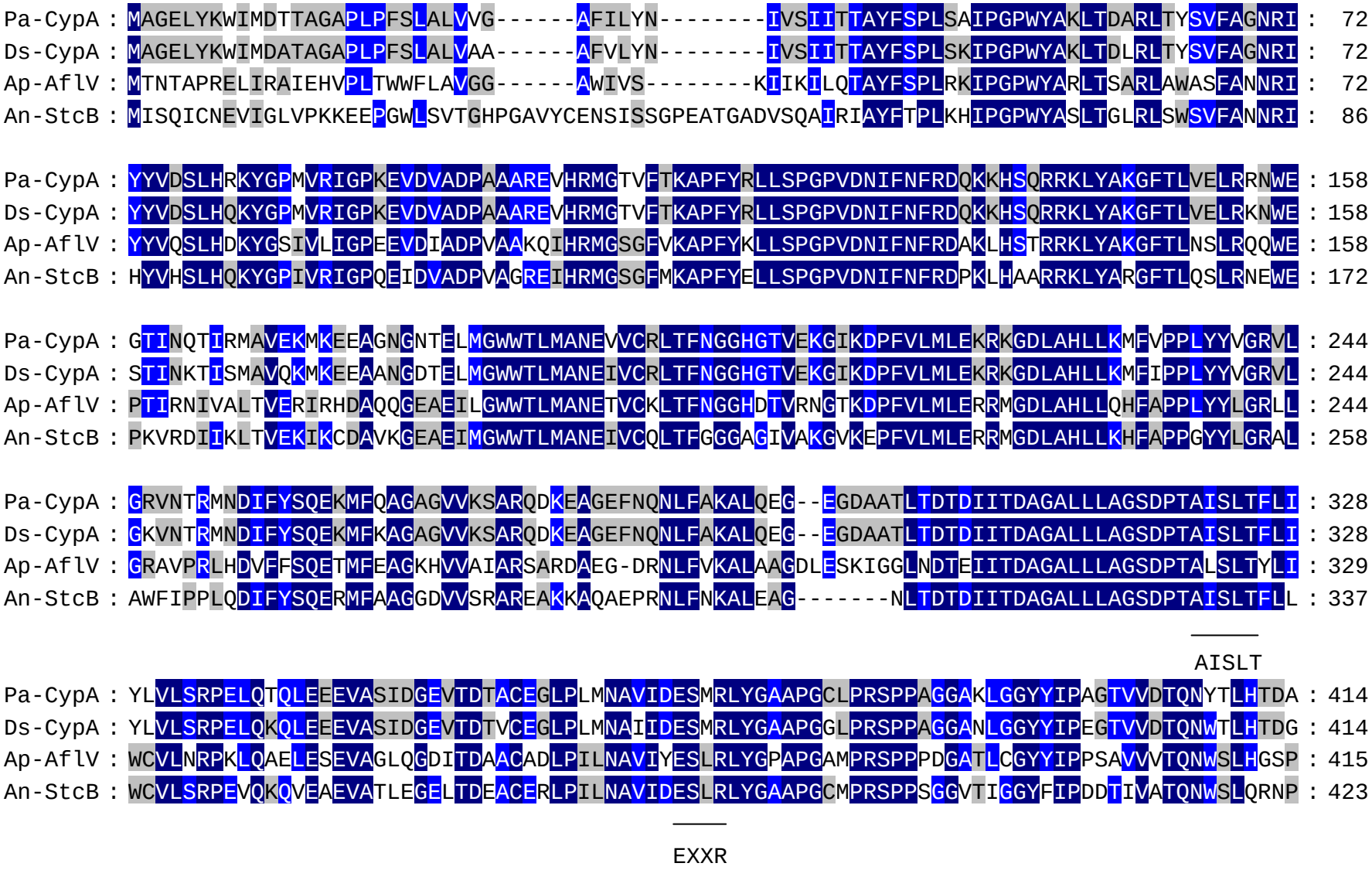
1561 actgcggggcgtgacgctggccaagtcggcgacacccgcgagcatgaccgtcgtcgacag 1620

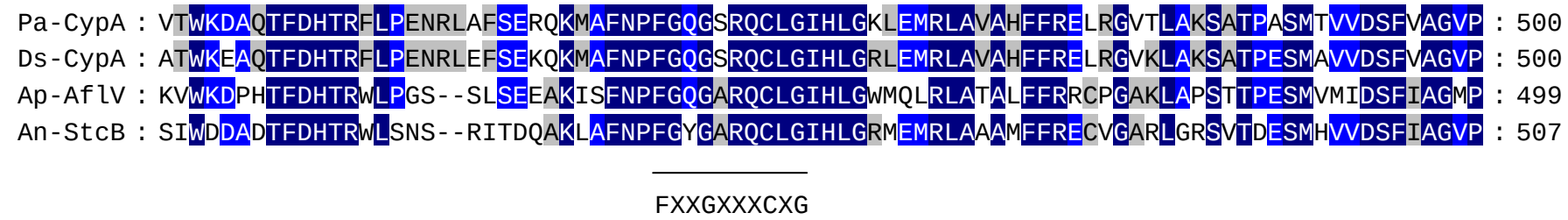
F V A G V P R D R R C E V T L A R

1621 tttcgttgccggcgtgccccgggaccgacggtgcgaggtgacattagcacgg

Appendix VI The ORF and amino acid sequence of *Pa-cypA*. The red nucleotide sequence indicates the positions of two introns. The blue amino acid sequences indicates the positions of the J-helix motif, K-helix motif and the heme binding motif, which were identified by alignment of Pa-CypA with other CypA proteins.

Appendix VII Alignment of Pa-CypA with other CypA proteins





Name	organism	Protein	Genbank No.
Pa-CypA	*P. arachidicola*	CypA	
Ds-CypA	*D. septosporum*	CypA	AAZ95016
Ap-AflV	*A. parasiticus*	AflV	AAS66022
An-StcB	*A. nidulans*	StcB	EAA61612

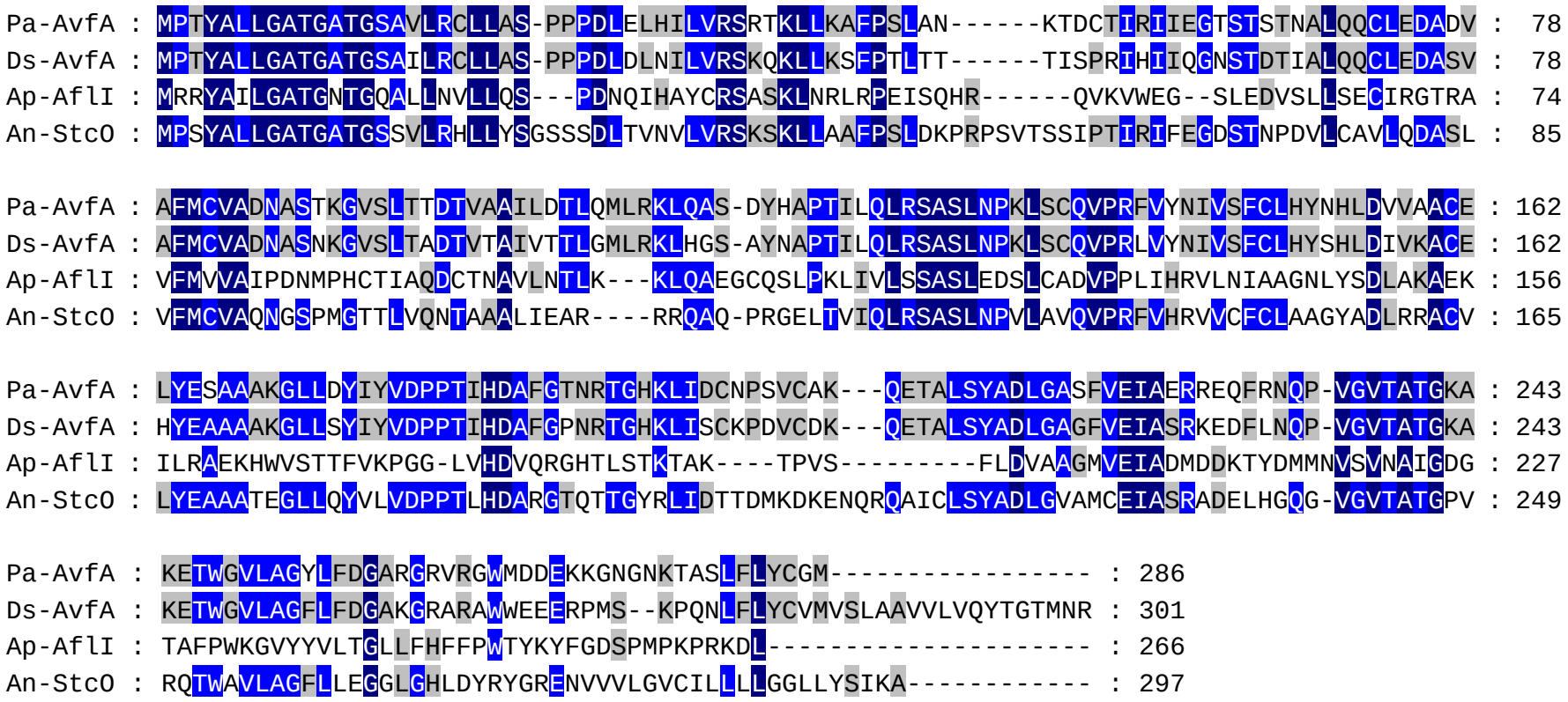
Appendix VII Alignment of Pa-CypA with other CypA proteins. The black lines indicate the positions of J-helix, K-helix and heme binding motifs.

Appendix VIII Nucleotide and amino acid sequence of *Pa-avfA*

```
        M  P  T  Y  A  L  L  G  A  T  G  A  T  G  S  A  V  L  R  C
    1 atgcctacctacgctctactgggggcaacaggagccacaggctcagcagtcctgcgctgc  60
        L  L  A  S  P  P  P  D  L  E  L  H  I  L  V  R  S  R  T  K
   61 ttgctggcttcaccacccccggacctcgaactccacatcctcgtccgctcgagaaccaag 120
        L  L  K  A  F  P  S  L  A  N  K  T  D  C  T  I  R  I  I  E
  121 ctcctcaaagccttcccaagccttgcaaacaaaaccgactgcaccatccgcatcatcgaa 180
        G  T  S  T  S  T  N  A  L  Q  Q  C  L  E  D  A  D  V  A  F
  181 ggcacctcaacctccaccaatgccctccaacaatgcctcgaagatgccgatgtagcctttt 240
        M  C  V  A  D  N  A  S  T  K  G  V  S  L  T  T  D  T  V  A
  241 atgtgcgtcgccgataacgcctccaccaaaggcgtctccctcacaacagacaccgtcgcc 300
        A  I  L  D  T  L  Q  M  L  R  K  L  Q  A  S  D  Y  H  A  P
  301 gccatactggacactctccaaatgctgcgcaagctgcaagccagcgactaccacgcgcca 360
        T  I  L  Q  L  R  S  A  S  L  N  P  K  L  S  C  Q  V  P  R
  361 acaatcctccagctccgcagcgcatcgctcaaccccaagctgagctgccaagtcccccgc 420
        F  V  Y  N  I  V  S  F  C  L  H  Y  N  H  L  D  V  V  A  A
  421 ttcgtctacaatatcgtctccttctgcttgcactacaaccacctcgacgtcgtcgcagcc 480
        C  E  L  Y  E  S  A  A  A  K  G  L  L  D  Y  I  Y  V  D  P
  481 tgtgagctctacgaatcagccgccgccaaaggcctcctcgactacatctacgtcgaccca 540
        P  T  I  H  D  A  F  G  T  N  R  T  G  H  K  L  I  D  C  N
  541 cccaccatccacgacgcgttcggcacaaaccgcacgggacacaagctgatcgactgcaac 600
        P  S  V  C  A  K  Q  E  T  A  L  S  Y  A  D  L  G  A  S  F
  601 ccttcagtctgcgcgaagcaagaaacggcactgagctacgcggatcttggcgcaagcttt 660
        V  E  I  A  E  R  R  E  Q  F  R  N  Q  P  V  G  V  T  A  T
  661 gtcgaaatcgccgagaggcgggaacaatttcggaatcagcctgtcggcgtgaccgccacg 720
        G  K  A  K  E  T  W  G  V  L  A  G  Y  L  F  D  G  A  R  G
  721 gggaaagccaaggagacgtggggcgtcctggcggggtacctttcgacggtgcaagggga 780
        R  V  R  G  W  M  D  D  E  K  K  G  N  G  N  K  T  A  S  L
  781 agggtgagaggctggatggatgacgagaagaagggcaatgggaacaagactgcgagcttg 840
        F  L  Y  C  G  M
  841 tttctctactgtggcatg
```

Appendix VIII The ORF and amino acid sequence of *Pa-avfA*. No intron was identified in *Pa-avfA*.

Appendix IX Alignment of Pa-AvfA with other oxidase proteins



Name	organism	Protein	Genbank No.
Pa-AvfA	*P. arachidicola*	AvfA	
Ds-AvfA	*D. septosporum*	AvfA	AAZ95014
Ap-AflI	*A. parasiticus*	AflI	AAS66024
An-StcO	*A. nidulans*	StcO	XP681080

Appendix IX Alignment of Pa-AvfA with other oxidase proteins. No specific motif was identified in Pa-AvfA.

Appendix X Nucleotide and amino acid sequence of *Pa-epoA*

The first matched sequence

```
        M  A  C  Y  T  T  F  P  A  N  A  T  L  K  P  T  P  S  K  V
    1 ATGGCCTGCTACACCACATTCCCAGCAAACGCCACGCTCAAACCTACGCCCTCCAAAGTC 60
        D  I  P  D  S  K  L  Q  Q  L  Q  I  R  L  S  P  T  G
   61 GACATCCCAGACTCCAAACTTCAACAGCTTCAAATCCGTCTCTCACCCACCGGGTCCATA 120
  121 CCTTACTCTTCCACTCTACAAAAACCCGACTTCAACCAACTGAAGGGTATCCACCCTCCA 180
                    P  T  N  S  N  P  K  A  G  T  Q  Y  G  I  R
  181 CGCAGGCGCCGCAGACCCACCAACTCCAACCCAAAGGCCGGAACGCAATATGGCATCCGC 240
        R  A  *  L  L  N  A  K  A  Q  *  Q  T  T  S  T  W  T  P  S
  241 CGCGCCTAGCTCCTCAACGCCAAAGCACAATGACAAACCACTTCCACCTGGACACCCTCC 300
        E  A  K  P  R  P  L  A  T  I  R  N  P  S  S  K  K  A  T  S
  301 GAAGCAAAACCTCGACCGCTAGCCACAATCCGAAATCCCAGTAGTAAAAAGGCCACCAGC 360
        S  E  E  L  K  F  H  S  I  A  F  F  S  N  E  E  W  T  L  S
  361 AGCGAAGAGCTCAAATTCCACTCCATCGCATTCTTCAGCAACGAGGAATGGACGCTCAGC 420
        H  *  H  S  P  L  S  F  A  H  F  L  P  N  L  D
  421 CACTAGCATTCCCCCTTAAGCTTCGCACATTTCCTCCCAAACCTCGACTGAGTGTTGGGG 480
```

The second matched sequence

```
          P  C  P  A  L  S  C  P  M  T  C  A  A  T  T  S  D  A  V  Q
  661 TGCCCTGCCCTGCCCTGTCCTGTCCTATGACTTGTGCCGCTACAACGTCGGATGCGGTGC 720
          Y  R  R  H  E  S  G  G  H  F  A  Y  W  E  R
  721 AGTACAGACGGCATGAGAGCGGAGGACATTTTGCATATTGGGAACGGA             768
```

Appendix X The partial ORF and amino acid sequence of *Pa-epoA* that predicated fom BLASTX results matched to three parts of the *Ds-epoA* amino acid sequence. Three stop codons (Black star) were identified in the first coding region. There are 228 a.a sequence of *Pa-epoA* was missing from *Pa-epoA* compared to *Ds-epoA*.

Appendix XI Alignment of Pa-EpoA with Ds-EpoA protein

First matched a.a sequence:

```
Pa-EpoA : MACYTTFPANATLKPTPSKVDIPDSKLQQLQ--IRLSPTGPTNSN :  43
Ds-EpoA : MEGYTTLPSTATLKPSPFTVSISESKLQTLQDLIRLSPIGPADYN :  45

Pa-EpoA : ---PKAGTQYGIRRA-LLNAKAQ-QTTSTWTPSEAKPRPLATIRN :  83
Ds-EpoA : NSSPSTGSKYGIRRDWLINAKKQWEDNFSWRTFEKKLKKYPQYTV :  90

Pa-EpoA : PSSKKATSSEELKFHSIAFFS---NEEWTLSH-HSPLSFAHFLPN : 124
Ds-EpoA : PVK--GESGETIEIHFIALFSQRQDARPLAFYHGWPSSPFDFLPI : 133

Pa-EpoA : LD-VLG
Ds-EpoA : LDLLTN
```

Second matched a.a sequence:

```
Pa-EpoA : PCPALSCPMTCAATTSDAVQYRRHESGGHFAYWER------- : 378
Ds-EpoA : PNTYLPCPVSWAKTTANLVQYRRHESGGHFAPWERPRELLED : 385
          ↓                                ↓
          Thr346                           His372
```

Appendix XI Alignment of Pa-EpoA with Ds-EpoA protein. BLASTX results showed that Pa-EpoA matched to three parts of the Ds-EpoA amino acid sequence. The black arrows indicates the position of active sites Thr[346] and His[372].

Appendix XII Nucleotide and amino acid sequence of *Pa-moxA*

```
     M  A  P  F  L  D  G  H  G  Q  A  G  F  E  A  P  P  G  Y  T
   1 atggccccgttccttgatggccacggccaagctggcttcgaggcaccaccaggctacaca 60
     P  T  K  R  H  Q  R
  61 cccaccaagcgccaccagaggtaatcccttcagcatggaagacgtcttccatcccaagat 120
                          N  Q  Q  P  V  D  Y  S  T  P  G  S
 121 agccaagactgatctctcccaacaggaaccaacagccagtagactattcaaccccaggct 180
       T  G  Y  N  I  P  Q  N  T  T  W  N  D  P  S  N  R  K  I  R
 181 cgactggctacaacatcccccagaacacaacatggaacgaccccagcaaccgcaagatcc 240
       V  L  T  I  G  A  G  I  S  G  I
 241 gcgtgctcacaattggagctggaatttctggaattc
```

Appendix XII The partial ORF and amino acid sequence of *Pa-moxA*. The red nucleotide sequence indicates the position of intron.

Appendix XIII Alignment of Pa-MoxA with other monooxygenase proteins

```
Pa-MoxA : MAPFLDGHGQAGFEAPPGYTPTKRHQRNQQPVDYSTPGSTGYNIP : 45
Ds-MoxA : MAPFLSAHGESASSSSSSPTPSRHTRN-QHVDYSTPGSTGYNIP : 44
Ap-AflW : ---------------------------------------------- : -
An-StcW : -------------------------MTVHYVHEGPEPQESRYSIP : 20

Pa-MoxA : QNTTWNDPSNRKIRVLTIGAGISGI-------------------- : 70
Ds-MoxA : QNTTWNAPSNRKIRVLTIGAGISGIL------------------- : 70
Ap-AflW : -----MDPANRPLRVVTIGTGISGILMAYQIQKQCPNVEHVLYEK : 40
An-StcW : QHTTWMDPNNRRLRVITIGAGFSGILMAYQIQKQCANIEHVVYEK : 65
```

Name	organism	Protein	Genbank No.
Pa-AvfA	*P. arachidicola*	MoxA	
Ds-AvfA	*D. septosporum*	MoxA	AAZ95013
Ap-AflW	*A. parasiticus*	AflW	AAS66023
An-StcW	*A. nidulans*	StcW	EAA61592

Appendix XIII Alignment of Pa-MoxA with other monooxygenase proteins. No specific motif was identified in Pa-MoxA.

Appendix XIV Nucleotide and amino acid sequence of partial *Pa-hexA*

```
          E  F  A  R  Q  S  L  D  Q  K  M  A  F  A  S  R  W  A  A  K
    1 gaattcgctcgacagagtctggaccagaagatggcgtttgccagtcgatgggccgcgaag  60
      E  A  V  F  K  C  L  Q  T  Q  T  K  G  A  G  A  A  M  K  D
   61 gaggctgtgttcaagtgtctgcagacccagacgaaaggtgctggggctgcgatgaaggat 120
      I  E  I  V  K  S  A  D  A  P  K
  121 attgagattgtgaagagtgctgatgcgcccaaagtgaaggtacggtgatgtcgtttacct 180
                                        C  I  R  A  A  R  K
  181 gtctctgaacgaatgctgacgagtggtagttgcacaatgactgcatccgagcggctcgca 240
        A  G  L  E  D  I  Q  L  S  I  S  H  G  E  D  C  L  I  A  V
  241 aagccgggttagaggacattcagctgagtatcagtcatggggaggactgtttgattgctg 300
        A  I  G  I  A  E  N  R  P  A  R  Y  T  L
  301 ttgctattggtatcgctgaaaaccgtccagcgaggtacacgttg               344
```

Appendix XIV The partial ORF and amino acid sequence of *Pa-hexA*. The red nucleotide sequence indicates the position of intron. The blue amino acid sequences indicates the position of the ACP domain.

Appendix XV Alignment of Pa-HexA with other HexA proteins

```
Pa-hexA : --EFARQSLDQKMAFASRWAAKEAVFKCLQTQTKGAGAAMKDIEI :  43
Ds-hexA : --AFAKQSLDQKMAFASRWAAKEAVFKCLHTQTKGAGAAMKDIEI :  43
An-stcJ : QSLQLQSHRSFRSAVASGWCAKEAVFKCLQTVSKGAGAAMSEIEI :  45
Ap-aflA : --DWAEKSADVRAAYASRWCAKEAVFKCLQTHSQGAGAAMKEIEI :  43
                                  ----------
                                  AKEAVFKCL
Pa-hexA : VKSADAPKCI-------RAARKAGLEDIQLSISHGEDCLIAVAIG :  81
Ds-hexA : VKSDNAPQG--------QAGRKAGLEDIQLSISHGEDCLIAVAIG :  80
An-stcJ : VRVQGAPSVLHG--DALAAAQKAGLDNIQLSLSYGDDCVVAVALG :  88
Ap-aflA : EHGGNGAPKVKLRGAAQTAARQRGLEGVQLSISYGDDAVIAVALG :  88

Pa-hexA : IAENRPARYTL :  92
Ds-hexA : IAGNGPAKYTL :  91
An-stcJ : VRKWCLWPLAS :  99
Ap-aflA : LMSGAS----- :  94
```

Name	organism	Protein	Genbank No.
Pa-HexA	*P. arachidicola*	HexA	
Ds-HexA	*D. septosporum*	HexA	ABU23831.1
An-StcJ	*A. nidulans*	StcJ	AAC49198.1
Ap-AflA	*A. parasiticus*	AflA	AAS66002.1

Appendix XV Alignment of Partial Pa-HexA with corresponding regions of other HexA proteins. The black line indicates the position of the ACP domain.

Appendix XVI Nucleotide and amino acid sequence of *Pa11*

```
      M  A  P  Q  L  R  D  P  R  N  A  K  R  T  Q  P  M  Q
    1 atggcgccccagctgagagatccgagaaacgccaagcggacgcagccgatgcaggtcagt  60

   61 cctcacgcgtgaccttccccttcctccctttgacaacaacatcgataccaccacaatgtt 120
                              V  L  C  L  G  M  S  R  S  G  T
  121 gggtacgatcgttgacagcatgcatagGtcctctgcttgggcatgtctcgctcaggaaca 180
      E  S  L  C  K  A  L  E  I  L  G  I  P  T  Y  H  G  W  R  S
  181 gaatctttgtgcaaagccctcgagattctcggcatcccgacttaccatggatggaggtcg 240
      M  E  N  H  K  Q  S  Q  L  W  V  E  A  M  Q  A  K  Y  E  N
  241 atggagaaccacaagcagtcgcagctgtgggtagaagcgatgcaggccaagtatgagaac 300
      K  G  K  V  W  K  R  E  D  F  D  Q  I  L  G  D  Y  S
  301 aaaggcaaagtgtggaagagagaggacttcgatcagattctgggtgattacagcgtaagt 360

  361 ctcgttccactggacaatcatggcagcattacaataccttcctatttcaagggcaacat 420
                           A  C  T  D  F  P  S  A  A  Y  P  V
  421 atctaaccccacctcgcccaccaggcctgcaccgacttcccctcggccgcctaccccgtc 480
      E  L  I  E  A  Y  P  E  A  K  V  I  L  N  Q  R  K  L  H  L
  481 gaactcatcgaagcctacccggaagccaaagtaatcctcaaccaacgtaagctccacctc 540
      T  S  H  L  P  F  I  *  Q  S  P  A  G  P  F  D  V  W  Y  K
  541 acctcccacctccccttcatctgacaatcccccgcaggccccttcgacgtctggtacaaa 600
      S  T  L  N  T  A  N  A  L  R  E  S  W  L  Y  W  V  L  M  N
  601 tccacactaaacaccgccaacgccctgcgcgaatcctggctctactgggtcctcatgaac 660
      L  D  S  E  A  R  A  V  Y  R  C  W  W  Y  I  N  D  H  Y  F
  661 ctcgactccgaagcccgcgccgtgtaccgctgctggtggtacatcaacgatcactacttc 720
      R  H  N  I  R  R  N  A  F  D  I  Y  K  E  H  A  A  L  V  R
  721 cgccacaacatccgtcgcaacgcctttgacatctacaaagagcatgccgctttggtccgg 780
      G  A  A  K  P  A  D  A  S  G  N  F  L  E  Y  D  V  T  Q  G
  781 ggagcggccaaacccgctgatgcctctgggaacttcttggagtatgatgttacgcaaggg 840
      W  G  P  L  C  A  F  L  G  A  K  V  P  D  V  E  F  P  M  G
  841 tgggggccgttgtgtgcgtttttgggggcgaaggtgcccgatgtggagttcccgatgggc 900
      N  V  A  E  E  F  H  K  R  V  E  G  S  M  K  P  R  F  V  R
  901 aatgtggcggaggaatttcataagagggtggaggggagtatgaagccgaggtttgtgagg 960
      S  V  R  N  L  V  I  M  V  G  S  V  V  G  L  G  W  Y  G  Y
  961 agtgtgaggaatttggtgattatggtggggagtgtggtggggttggggtggtatgggtat 1020
      V  Q  R  E  A  V  V  A  A  V  R  E  V  V  R  P  L  L  K
 1021 gtgcagagggaggctgtggtggcggctgtaagggaggttgtgaggccgttgttgaag    1077
```

Appendix XVI The ORF and amino acid sequence of *Pa11*. The red nucleotide sequence indicates the positions of two introns. The green star indicates the stop in the coding region.

Appendix XVII Alignment of Pa11 with Ds11

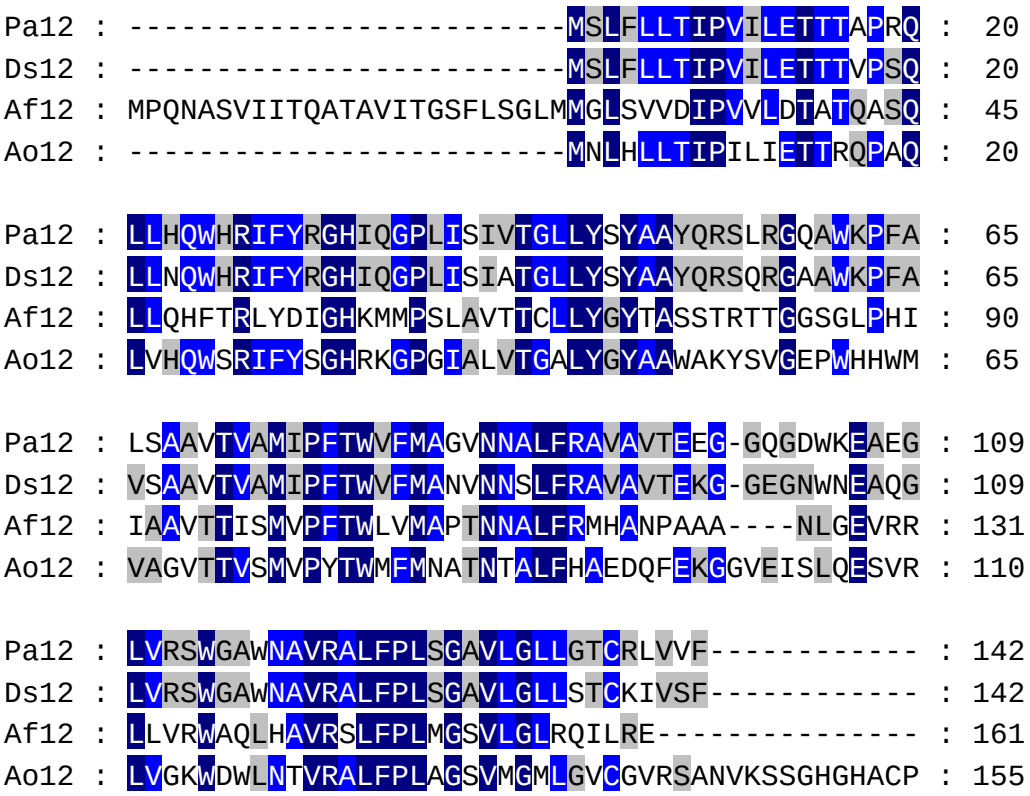
Appendix XVII Alignment of Pa11 with Ds11. The black arrow indicates the position of stop codon in *Pa11* coding region.

Appendix XVIII Nucleotide and amino acid sequence of *Pa12*

```
        M  S  L  F  L  L  T  I  P  V  I  L  E  T  T  T  A  P  R  Q
    1 atgtcgctcttcctcctaacgatccctgtcatcctcgagaccaccaccgccccgcgccaa 60
        L  L  H  Q  W  H  R  I  F  Y  R  G  H  I  Q  G  P  L  I  S
   61 ctcctgcaccaatggcaccggatcttctaccgcgggcacatccagggtccgctcatctcc 120
        I  V  T  G  L  L  Y  S  Y  A  A  Y  Q  R  S  L  R  G  Q  A
  121 atcgtaaccggcctcttgtactcctacgccgcctaccagcgctcgctccgcggccaggcg 180
        W  K  P  F  A  L  S  A  A  V  T  V  A  M  I  P  F  T  W  V
  181 tggaaaccctttgcgctctcagcggctgttacggtcgcgatgatccctttacgtgggtg 240
        F  M  A  G  V  N  N  A  L  F  R  A  V  A  V  T  E  E  G  G
  241 tttatggccggtgtgaataatgcgctgttcagggcggtggcggtcacggaggagggtggg 300
        Q  G  D  W  K  E  A  E  G  L  V  R  S  W  G  A  W  N  A  V
  301 cagggggattggaaggaggcggaggggttggtgaggagttgggggcatggaatgcggtt 360
        R  A  L  F  P  L  S  G  A  V  L  G  L  L  G  T  C  R  L  V
  361 cgagcgttgtttccactttctggggcggttttggggttgcttgggacttgtaggttggtg 420
        V  F
  421 gtgttt
```

Appendix XVIII The ORF and amino acid sequence of *Pa12*.

Appendix XIX Alignment of Pa12 with other hypothetical proteins



Name	organism	Protein	Genbank No.
Pa12	*P. arachidicola*	Pa12	
Ds12	*D. septosporum*	Ds12	AUB23833.1
Af12	*A. fumigatus*	hypothetical protein	XP751386
Ao12	*A. oryzae*	hypothetical protein	XP001821512

Appendix XIX Alignment of Pa12 with other hypothetical proteins. For Ds12, only the amino acid sequence from 125 to 266 was used for alignment.

Appendix XX Nucleotide and amino acid sequence of *Pa-vbsA*

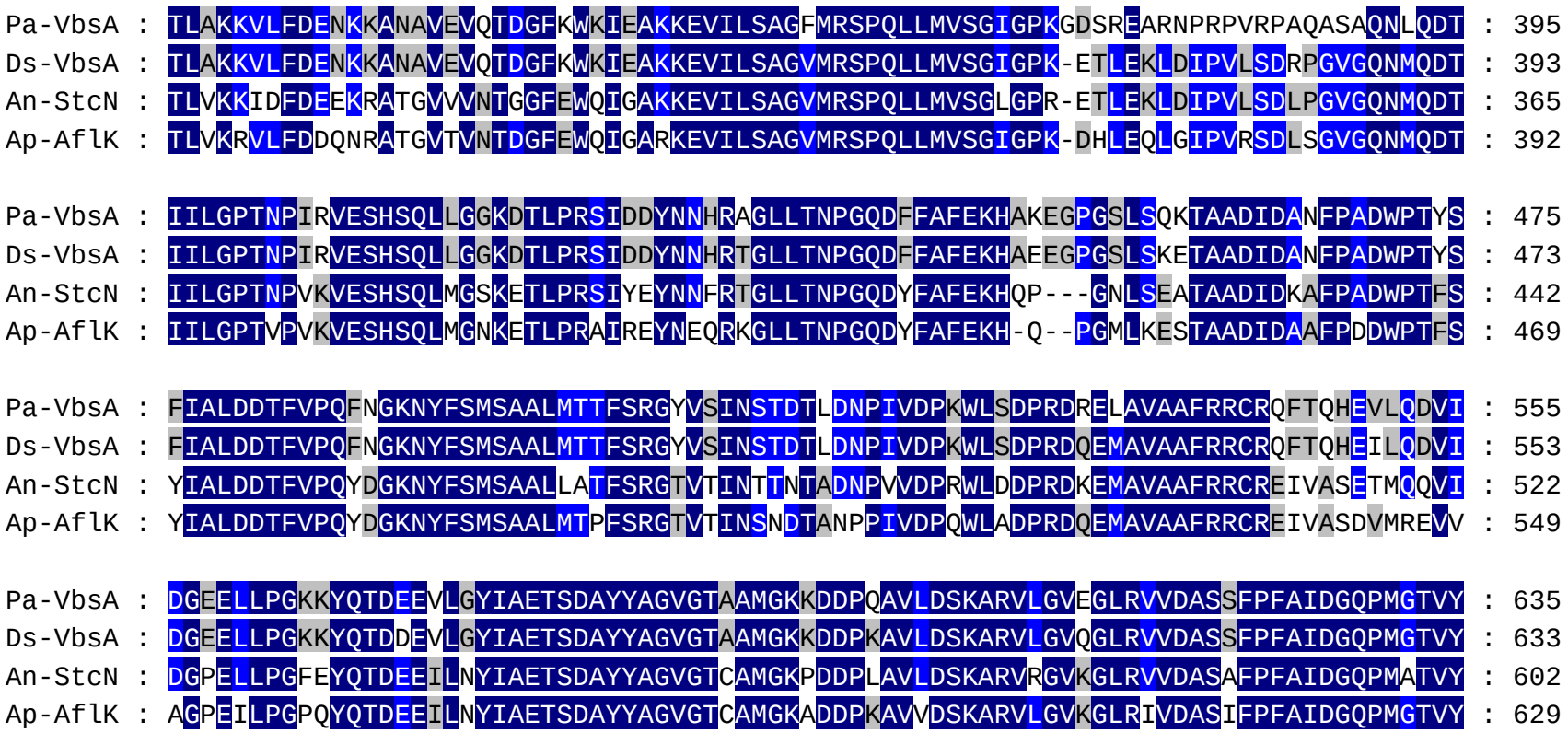
```
      M  A  L  S  T  L  L  S  A  A  A  M  P  V  A  G  L  F  A  L
    1 atggcgctcagcactctcctctccgcggcggcgatgcccgtcgctggccttttgctctt 60
      A  Q  P  S  T  S  F  F  D  Q  I  A  V  L  G  S  M  L  N  T
   61 gcacagccgtcgacctctttctttgatcagattgctgtcctgggctcgatgctcaatact 120
      P  S  V  G  T  Y  D  A  Q  A  A  Q  Y  D  G  R  I  Q  G  R
  121 cccagcgtgggcacatacgacgcacaggcggcgcagtacgatggtcgtatccaaggccgt 180
      D  L  L  S  S  H  F  G  P  I  G  W  P  H  Q  S  F  D  Y  V
  181 gacttgctcagctcacacttcggcccgattggctggcctcatcagagctttgactatgtc 240
      I  V  G  G  G  T  A  G  L  A  M  A  K  R  L  S  E  G  E  G
  241 attgttggcggtggtacagcaggtttggccatggccaagagattgtctgagggagagggc 300
      N  S  V  A  L  I  E  A  G  G  F  Y  E  V  D  A  G  N  A  T
```

```
301 aactcggtcgctctcattgaggctggtggtttctatgaggtggatgctgggaatgctacc 360
    E  V  P  M  Y  L  F  N  Y  F  F  D  N  G  Y  M  K  N  P  L
361 gaggtgccgatgtacctcttcaattacttcttcgacaatggctacatgaagaacccgctg 420
    F  D  W  Y  Q  Y  T  E  P  Q  E
421 ttcgactggtaccagtacaccgaaccacaagaggtcaggatccgacgacctacttaatgc 480
                       G  L  H  N  R  E  M  F  Y  M  Q  G  K
481 ttccagtgctgacacagacaggtctccacaatcgtgaaatgttctacatgcagggcaag 540
    T  L  G  G  S  T  A  R  G  A  M  L  Y  H  R  G  S  K  G  A
541 acgctcggaggcagcactgctcgtggcgctatgctgtaccaccgtggctcgaagggcgcc 600
    Y  Q  K  W  A  D  Q  V  G  D  D  S  Y  T  W  E  K  W  L  P
601 taccagaagtgggcagaccaagtcggtgatgactcgtacacctgggagaagtggctgcca 660
    H  F  Q  R  G  I  K  F  S  G  P  N  T  N  P  R  P  A  N  A
661 catttccagaggggcatcaaattctctggtccaaacaccaacccgcgccctgcgaacgcc 720
    T  A  V  N  D  D  T  A  W  S  A  T  G  G  P  V  H  V  A  Y
721 acggctgtcaacgatgacacggcgtggtctgcaaccggtggtccagtccacgttgcgtat 780
    P  Y  L  T  N  A  I  S  S  W  V  D  K  A  L  D  A  F  G  F
781 ccatacctcaccaacgccatctcttcctgggtcgacaaggctctcgatgcatttggcttc 840
    T  N  V  E  G  F  S  N  G  V  L  L  G  K  S  Y  I  T  H  T
841 accaatgtcgagggtttctccaacggcgtgcttctcggcaagtcgtacattacccacacc 900
    I  N  P  F  T  R  R  R  E  T  A  S  S  S  Y  L  R  E  A  L
901 attaacccattcactcgccgccgtgagactgcttcctcatcctacctccgtgaggcactc 960
    V  E  S  N  N  L  N  I  Y  I  R  T  L  A  K  K  V  L  F  D
961 gtcgagagcaacaacctcaacatctacatccgcaccctcgccaagaaggtcctcttcgac 1020
    E  N  K  K  A  N  A  V  E  V  Q  T  D  G  F  K  W  K  I  E
1021 gagaacaagaaggccaacgccgtcgaagtccagaccgacggcttcaagtggaagatcgag 1080
    A  K  K  E  V  I  L  S  A  G  F  M  R  S  P  Q  L  L  M  V
1081 gccaagaaggaagtcatcctctccgcagggttcatgcgctccccccaactcctcatggtc 1140
    S  G  I  G  P  K  G  D  S  R  E  A  R  N  P  R  P  V  R  P
1141 agcggcattggccccaaaggagactctcgagaagctcgaaatccccgtcctgtccgaccc 1200
    A  Q  A  S  A  Q  N  L  Q  D  T  I  I  L  G  P  T  N  P  I
1201 gcccaggcgtcggcccaaaacctccaagacaccatcatcctcggcccaaccaacccaatc 1260
    R  V  E  S  H  S  Q  L  L  G  G  K  D  T  L  P  R  S  I  D
1261 cgggtcgaatcccactcccaactcctcggcggaaaggacaccctcccccgctccattgat 1320
    D  Y  N  N  H  R  A  G  L  L  T  N  P  G  Q  D  F  F  A  F
1321 gactataacaaccaccgcgccggcctccttaccaacccaggccaagacttcttcgccttt 1380
    E  K  H  A  K  E  G  P  G  S  L  S  Q  K  T  A  A  D  I  D
1381 gagaagcacgccaaggaaggacctggctcgctctcccagaagacggcagcggacatcgac 1440
    A  N  F  P  A  D  W  P  T  Y  S  F  I  A  L  D  D  T  F  V
1441 gccaacttcccagccgactggccgacctactctttcatcgccctcgacgatactttcgtc 1500
    P  Q  F  N  G  K  N  Y  F  S  M  S  A  A  L  M  T  T  F  S
1501 ccacagttcaacggaaagaactatttttccatgtccgccgctctcatgacgactttcagc 1560
    R  G  Y  V  S  I  N  S  T  D  T  L  D  N  P  I  V  D  P  K
1561 cgtggctacgtcagtatcaacagcactgataccctcgataaccctatcgttgaccccaaa 1620
    W  L  S  D  P  R  D  R  E  L  A  V  A  A  F  R  R  C  R  Q
```

1621 tggctctccgatccaagggatcgagagttggctgttgcggcgttccgtcgttgcagacaa 1680
F T Q H E V L Q D V I D G E E L L P G K
1681 ttcacgcagcatgaggtgctccaggatgtcattgatggcgaagaactcctgccgggtaag 1740
K Y Q T D E E V L G Y I A E T S D A Y Y
1741 aagtaccagaccgacgaagaagttctgggctacattgccgagacgagtgatgcgtactac 1800
A G V G T A A M G K K D D P Q A V L D S
1801 gccggagtcggtactgccgccatgggcaagaaggatgacccccaggctgtcttggactcg 1860
K A R V L G V E G L R V V D A S S F P F
1861 aaggcgagagtgctgggtgtggagggtttgagagttgttgatgcgtcttcattcccgttc 1920
A I D G Q P M G T V Y A L A E K I A A D
1921 gcgattgatgggcagcctatgggcactgtttatgctctggcagagaagatcgctgccgat 1980
I L A G N
1981 attctggctggtaac

Appendix XX The ORF and amino acid sequence of *Pa-vbsA*. The red nucleotide sequence indicated the position of intron. The blue amino acid sequences indicated the position of the heme-binding domain and *N*-glycosylation sites

Appendix XXI Alignment of Pa-VbsA with other VbsA proteins



Name	organism	Protein	Genbank No.
Pa-VbsA	*P. arachidicola*	Pa-VbsA	
Ds-VbsA	*D. septosporum*	Ds-VbsA	ABO72541.2
An-StcN	*A. nidulans*	StcN	EAA61600
Ap-AflK	*A. parasiticus*	AflK	AAS66012.1

Appendix XXI Alignment of Pa-VbsA with other VbsA proteins. The black lines indicate the positions of flavin binding domain and two possible positions for *N*-glycosylation.

Appendix XXII Nucleotide and amino acid sequence of partial *Pa-dotA*

```
        M  S  A  D  N  F  R  L  D  G  K  V  A  L  V  T  G  S  G  R
    1 atgtctgccgacaacttccgcctcgacggcaaggtcgcccttgtcactggatctggacgc  60
        G  I  G  A  A  I  A  I  E  F  G  K  R  G  A  N  V  V  V  N
   61 ggtatcggtgccgccatcgccattgaatttggcaagcgtggtgcaaatgtcgtcgtgaac 120
        Y  S  R  A  V  D  E  A  N  K  V  V  E  T  I  I  A  N  G  T
  121 tactcgcgagctgtagacgaggccaacaaggtcgttgagacaatcattgcgaacggaacc 180
        K  A  I  A  I  K  A  D  V  G  E  I  E  Q  V  A  K  M  M  D
  181 aaggctattgctatcaaggccgatgttggtgagatcgaacaggttgcgaagatgatggac 240
        Q  A  V  E  H  F  G  Q  L  D  I  V  S  S  N  A  G  L  V  S
  241 caggctgttgagcactttggtcaactcgatattgtctcgtcgaatgctggtcttgtttca 300
        F  G  H  L  K  D  V  T  G  D
  301 ttcgggcacttgaaggatgtcactggtgatgtatgccttgacctcttctcatcgaaagag 360
                                    E  F
  361 ttcgtgctgaccatgagctaataggaattc
```

Appendix XXII The partial ORF and amino acid sequence of *Pa-dotA*. The red nucleotide sequences indicate the position of intron. The blue amino acid sequences indicate the position of the adenine nucleotide binding motif.

Appendix XXIII Alignment of Pa-DotA with other ketoreductase proteins

```
Pa-DotA : -MSADNFRLDGKVALVTGSGRGIGAAIAIEFGKRGANVVVNYSRA :  44
Ds-DotA : -MSVDNFRLDGKVALVTGSGRGIGAAIAIELGKRGANVVVNYSRA :  44
An-StcU : MSSSDNYRLDGKVALVTGAGRGIGAAIAVALGQRGAKVVVNYANS :  45
Ap-AflM : --MSDNHRLDGKVALVTGAGRGIGAAIAVALGERGAKVVVNYAHS :  43
                                  ______
                                  GXGIGX
Pa-DotA : VDEANKVVETIIANGTKAIAIKADVGEIEQVAKMMDQAVEHFGQL :  89
Ds-DotA : VAEANKVVETIIANGTKAIAIKADVGEIDQVAKMMDQAVEHFGQL :  89
An-StcU : REAAEKVVDEIKSNAQTAISIQADVGDPDAVTKLMDQAVEHFGYL :  90
Ap-AflM : REAAEKVVEQIKANGTDAIAIQADVGDPEATAKLMAETVRHFGYL :  88

Pa-DotA : DIVSSNAGLVSFGHLKDVTGDEF---- : 112
Ds-DotA : DIVSSNAGLVSFGHLKDVTGDEF---- : 112
An-StcU : DIVSSNAGIVSFGHVKDVTPDEF---- : 113
Ap-AflM : DIVSSNAGIVSFGHLKDVTPEEFDRVF : 115
```

Name	Organism Protein	Protein	Genbank No.
Pa-DotA	*P. arachidicola*	DotA	
Ds-DotA	*D. septosporum*	DotA	AAL87045
An-StcU	*A. nidulans*	StcU	AAC49205.1
Ap-AflM	*A. parasiticus*	AflM	AAS660141

Appendix XV Alignment of partial Pa-DotA with corresponding regions of other ketoreductase proteins. The black line indicates the position of the adenine nucleotide binding motif.

Appendix XXIV Nucleotide and amino acid sequence of *Pa-dotB*

```
        M  Y  S  S  A  A  L  V  L  T  C  L  A  S  T  A  V  A  Y  P
      1 atgtactcctctgcagcactcgttttgacctgcctggcatccacggcagtggcatatcca 60
        A  L  D  Q  L  A  T  S  I  D  Y  Q  K  Y  Q  K  E  E  K  R
     61 gccttggaccagcttgcgacatccatcgactatcagaagtatcagaaggaagagaagcgc 120
        Q  T  L  G  F  D  A  A  S  Q  K  V  S  T  T  G  D  H  A  W
    121 cagactctcggcttcgacgctgcttctcagaaggtcagcacgacaggcgatcatgcctgg 180
        Q  A  P  G  P  N  D  F  R  G  P  C  P  G  L  N  S  M  A  N
    181 caggcaccaggaccgaacgactttcgcggaccatgcccaggactgaacagcatggcgaac 240
        H  G  Y  I  P  R  N  G  Y  T  S  N  T  Q  V  I  A  A  M  K
    241 catggatatatcccacgcaatgggtacacttcaaacacgcaagtgattgctgccatgaag 300
        D  V  F  N  I  S  P  E  F  G  G  F  L  T  I  L  G  S  A  M
    301 gatgttttcaacatctctcccgagtttggtggcttcctgacaattcttggctcggccatg 360
        G  G  D  G  L  G  F  S  I  G  G  P  P  P  A  S  L  L  P  A
    361 ggcggcgatggtctcggcttttccattggtggacctcccccagcatcgcttttgccggca 420
        T  G  L  V  G  R  P  Q  G  M  S  N  T  H  N  R  F  E  S  D
    421 actggtcttgtcggtagaccacagggtatgagtaacacccacaaccgcttcgaaagcgac 480
        Q  S  I  T  R  D  D  L  Y  Q  T  G  N  A  V  T  L  N  M  N
    481 cagagcataacgcgcgacgacctgtaccagactggcaacgctgtaactttgaacatgaac 540
        L  F  K  D  L  L  N  S  P  L  P  R  G  W  Y  D  I  D  V  L
    541 ctcttcaaggatcttttgaacagtccactgccaagaggttggtacgacatcgatgtgctc 600
        G  N  H  Q  V  K  R  F  Q  Y  S  K  A  N  N  P  Y  F  F  K
    601 ggaaatcatcaagtgaagaggttccaatactccaaggcaaacaacccatatttcttcaag 660
        G  L  N  T  A  F  I  P  E  A  T  S  A  L  V  A  Y  L  F  S
    661 ggtctcaacacggctttcatccccgaagcgacttcagcactcgtcgcatacctcttctcc 720
        N  H  T  A  E  C  P  T  G  C  L  D  A  A  G  L  K  S  F  Y
    721 aaccacaccgccgaatgcccaactggatgtctcgacgctgctggcctcaagagcttctac 780
        G  V  T  G  S  G  S  N  L  K  Y  T  P  G  T  E  R  I  P  D
    781 ggcgttaccggctccgggtcgaacttgaagtacactccaggaaccgagcgcattcccgac 840
        N  W  Y  K  Y  P  I  G  Y  G  V  A  N  V  F  A  D  M  V  T
    841 aattggtacaaatatccaataggctatggcgtcgccaacgttttcgcggacatggtcaca 900
        V  Y  S  K  Y  S  N  Q  A  A  F  G  G  N  T  G  T  V  N  S
    901 gtgtattccaagtattccaatcaagccgcgtttggcggcaacaccgggactgtcaacagc 960
        F  V  G  L  D  V  S  N  I  T  G  G  A  Y  N  T  A  T  L  L
    961 ttcgtcggcttggacgtctcgaacattactggaggcgcgtacaacacggcgactcttctg 1020
        Q  G  N  N  L  G  C  F  L  F  L  G  M  Q  F  F  M  P  D  L
   1021 caaggcaacaatctcgggtgcttcttgttccttggcatgcagttttcatgcctgatttg 1080
        I  T  Q  G  G  V  L  G  D  V  A  G  V  V  S  D  L  T  G  S
   1081 atcacacaaggtggtgtcctgggcgatgtggcaggcgttgtgtccgatttgacaggaagc 1140
        I  T  S  M  L  A  P  L  N  C  P  K  L  S  S  I  D  K  S  A
   1141 attacctcgatgctggcaccgctcaattgtcccaagttgtcgagcattgataagagcgct 1200
        F  S  I  Y  P  G  W  N  N  G  R  P  R  K
   1201 ttctccatctatcccggctggaacaatggcagacctaggaaa
```

Appendix XXIV The ORF and amino acid sequence of *Pa-dotB*.

Appendix XXV Alignment of Pa-DotB with other oxidase proteins

```
Pa-DotB : MYSSAALVLTCLASTAVAYPA-----LDQLATSIDYQKYQKEEKRQTLGFDAASQKVST-----------TGDHAWQAPGPNDF :  68
Ds-DotB : MHFFSAIVLTCLASTAVAYPA-----LEQAASSAEFKEYQKQEKRQTLGFDAASQIVST-----------TGDHAWQAPGANDI :  68
Af-DotB : MKATISLVILGLGSLASGFPDRMGGSAHMNCPYAALKKGEDAELGKRFLFDSLTKPIDVGGGENGAERATVTGVHAFQPPRKGDQ :  85
An-StcC : ------MLLKSIQNIVCGLVP----------------------TFFLFGSAAAELDF---------------EQWHPAGLGDL :  40

Pa-DotB : RGPCPGLNSMANHGYIPRNGY-TSNTQVIAAMKDVFNISPEFGGFLTILGSAMGGDGL----GFSIGGPPPAS---LLPATGLVG : 145
Ds-DotB : RGPCPGLNSMANHGYIPRNGY-TSDAQIIAAMQAVFNISPDFGGFLTVLGSAMGGDGL----GFSIGGPPSAS---LLTATGLVG : 145
Af-DotB : RGPCPALNALANHAYIPRSGV-VSVCGPSLARTFVYGMGVDLATILALMGLVWTGNPLALVPSFSIGGRDPGVNNLLNNLGGLLG : 169
An-StcC : RCGCPAMNSLANHGFINHNGSNITVNEVIPLMQEVFHLSEELATIVTGLAVLSADDPA--------S-------------G--I : 101

                XCPXXNXXANHX
Pa-DotB : RPQGMSNTHNRFESDQSITRDDLYQTGNAVTLNMNLFKDLLNSPLPRGWYDIDVLGNHQVKRFQYSKANNPYFFKGLNTAFIPEA : 230
Ds-DotB : KPQGMSNTHNRFESDQSITRDDLYQTGNDVTLNMNFFQDLLNSSLPKGWYDIDVLGNHAVKRFQYSVANNPYFFKGLNTAFIPEA : 230
Af-DotB : EPQGLIGSHNFIEADSSNTRDDLYVTGNNYALNMDKFLEWYNMS-TDGTFSMDLMAERAKTRFEQSIQTNPEFYYGPVTGFIARN : 253
An-StcC : FNLDMLNRHNIFEHDASLTRKDFYLGGDGHTIDQPTLDEFLSYFDGKEWIDLNDAAAARYARVLDSREKNPSFLYQDQQLITSYG : 186

Pa-DotB : TSALV-AYLFSNHTAECPTGCLDAAGLKSFYGVTGSGSNLKYTPGTERIPDNWYKYPIGYGVANVFADMVTVYSKYSNQAAFGGN : 314
Ds-DotB : TSALV-TYLFANHSAACPAGCLDATNLKSFYSVTGSGSTLKYTPGHERIPDNWYKYPVGYGVANVFADMVTVYSKYSNQAAFGGN : 314
Af-DotB : AGYLFPARFFRNYSSENPEGVLTKEIVRNFYGIYGEEGNLTYREGWERIPENWHKTPLDYGLVQFNIDLVDWVLKHPELASIGGN : 338
An-StcC : ETIKYFRTMVDPRSNKTSAEFVRILFTEERLPVRKGGSAREKRSVGSRWPAMSFSWRCAPQRSSLACRSTSVRLQ-SRPLTRCHG : 270

Pa-DotB : TGTVNSFVGLDVSNITGGAYNTATLLQGNNLGCFLFLGMQFFMPDLITQGGVLGDVAGVVSDLTGSITSMLAPLNCPKLSSIDKS : 399
Ds-DotB : TGTVNSFTGLDVANITGGVYNAETLLQGNNLGCFLFNGMEFFMPDLISNGGVIGDVSGVVSSLTGTITSLLAPFNCPKLSGIDKK : 399
Af-DotB : TGTVNSFTGVNLADVTGGVLNGTTLLEGNNLLCFVFEVLKFASPN--ALAGLYKTLAVPLDLVNRIIAVPLLDMSCPAFKDMQMG : 421
An-StcC : SGLPSGLPRTTRVSVRGISLSLSGGLRRRPFRFVLDRHLSF---------------------------------------- : 311

Pa-DotB : AFSIYPGWNNGRPRK------- : 414
Ds-DotB : AFAIYPGWNDGKPRK------- : 414
Af-DotB : GRPLWDAIKDDFPGAMKSGGAL : 443
An-StcC : ---------------------- :   -
```

Appendix XXIV Black line indicates the heme-binding domain.

Name	Organism Protein	Protein	Genbank No.
Pa-DotB	*P. arachidicola*	Pa-DotB	
Ds-DotB	*D. septosporum*	Ds-DotB	AAL87046
Af-DotB	*A.fumigatus*	Af-DotB	EDP50907
An-StcC	*A.nidulans*	StcC	AAC49193.1

Appendix XXVI Nucleotide and amino acid sequence of *Pa-dotC*

```
        M  S  Q  D  H  T  K  A  E  D  L  S  E  K  E  Q  H  S  P  S
      1 atgtctcaagaccatactaaagcggaggatctgtccgagaaggagcagcattcgccttcg 60
        R  S  D  S  G  S  Q  H  D  V  V  A  N  T  A  E  E  E  S  S
     61 aggagtgatagtgggagtcagcatgatgttgttgccaatacggcggaggaggagagttct 120
        D  D  M  G  A  L  D  G  K  P  K  S  L  I  A  I  V  M  I  A
    121 gatgatatgggagcgcttgacggcaagccaaagagcttaattgcgatcgttatgattgcg 180
        L  S
    181 ctatcggtgagtcctctcgtcaatactctaggggatgatggcggcatgctgacacattcg 240
           L  A  V  F  L  S  A  L  D  T  T  I  V  T  V  A  L  P  A
    241 tagctagcagtcttcttgtcagcgcttgacacgaccatcgtcactgtggcactcccagcg 300
        I  A  E  H  F  H  S  T  A  S  Y  T  W  V  G  S  A  Y  L  L
    301 atcgccgaacacttccactcgaccgcatcgtatacctgggttggatccgcatacctactc 360
        A  N  A  A  S  T  P  I  W  G  K  L  A  D  I  F  G  R  K  P
    361 gccaatgccgcatcgacaccaatctggggcaagctggccgacatcttcggacgcaagcct 420
        M  L  L  A  A  N  A  L  F  M  I  G  S  L  I  C  G  L  S  I
    421 atgctccttgccgcaaacgcacttttcatgatcggatccctcatttgtggtctgtcgata 480
        N  V  G  M  L  V  T  A  R  A  I  Q  G  A  A  G  G  G  L  L
    481 aatgtcggcatgcttgtcaccgcccgtgccattcaaggtgccgcaggcggtggtttgttg 540
        T  L  V  D  T  I  I  G  D  L  F  S  L  R  T  R  G  T  Y  L
    541 acgctggtcgatacgatcattggcgatctgttctcgttgaggaccagaggaacatacctg 600
        G  M  I  G  G  V  W  A  I  A  C  A  L  G  P  I  I  G  G  A
    601 ggcatgattggtggtgtctgggccattgcttgtgctttgggtcctataattggtggtgct 660
        F  T  S  G  V  T  W  R  W
    661 ttcacatctggagtgacctggagatggtgagtcgtatgaacatcggttttttagagaatc 720
                             C  F  Y  I  N  L  P  I  D  G  I  A  F
    721 agattgctgatacgcgggcaggtgtttctacatcaacctgccgatcgacggcatcgcgtt 780
         F  I  I  L  F  F  L  K  L  K  T  P  K  T  P  L  I  E  G  F
    781 cttcatcatccttttcttttgaagcttaagactcccaagacgcctctaatcgaaggctt 840
         K  A  I  D  W  A  G  S  F  F  V  S  I  T  A  N  S  T  R  L
    841 caaggccatcgactgggctggcagcttctttgtgagtattaccgcaaacagtaccaggct 900
         R  T  L  T  M  L  Q  I  I  G  G  T  L  L  F  L  F  G  L  Q
    901 tcggacactgactatgttgcagatcatcggcggcactttgttgttcctcttcggcctcca 960
         Y  G  G  Q  T  F  P  W  D  S  A  T  V  I  C  L  L  V  F  G
    961 atacggcggccagacattcccatgggattctgcgactgtcatctgcctgttggtctttgg 1020
         I  V  C  V  V  I  F  G  F  V  E  W  K  V  A  K  Y  P  I  I
   1021 catcgtctgtgtcgtcatcttcggcttcgtcgagtggaaggtcgccaaatatcctatcat 1080
         P  L  R  L  F  K  Y  R  N  N  C  G  A  L  L  V  A  F  F  H
   1081 cccgcttcgactgttcaagtaccgcaacaactgcggtgctcttctggtggccttcttcca 1140
         S  L  V  F  I  S  A  F  Y  Y  L  P  L  Y  F  Q  A  V  K  G
   1141 ctcattagtctttatctcggcgttctactacctaccactatacttccaggccgtcaaggg 1200
         A  S  P  I  L  A  G  V  Y  I  L  P  A  V  L  M  T  G  V  S
   1201 agcatccccaatcttggctggtgtgtacatcctgccagctgtgcttatgacgggtgtcag 1260
         A  A  A  T  G  A  F  I  G  N  T  G  N  Y  L  I  P  M  Y  F
   1261 cgctgcagctaccggagcctttatcggaaacacaggcaactacctcatcccaatgtactt 1320
         G  M  T  T  M  V  L  G  Y  G  L  L  I  N  F  Q  A  S  S  G
   1321 tggaatgaccacgatggtgctcggatatggccttttgatcaacttccaggccagctctgg 1380
         W  A  K  L  I  I  Y  Q  L  L  V  G  I  G  N  G  P  N  F  Q
   1381 ctgggcgaagctcattatctaccagcttcttgttggtatcggcaatgggcccaacttcca 1440
         A  P  L  V  A  L  Q  T  K  I  K  Q  S  D  I  A  T  G  T  A
```

```
1441 ggctccgttggttgcgctccagaccaagatcaagcaaagcgatattgctactggcactgc 1500
      T   F   N   F   V   R   N   I   A   T   A   V   S   V   V   V   G   Q   V   V
1501 cacattcaacttcgtccgcaacatcgccactgcagtcagtgtagtcgtcggccaagtcgt 1560
      Y   A   N   Q   L   S   G   M   K   N   R   L   Q   Q   L   G   S   A   A   G
1561 ctatgcgaatcagctcagcggcatgaagaaccgtcttcaacagctaggttcagctgcagg 1620
      L   I   A   S   G   D   A   G   A   N   V   E   M   I   Q   D   L   P   R   D
1621 acttatcgcctctggagatgccggagccaacgtggagatgatccaagatcttcccaggga 1680
      Q   R   S   I   A   R   S   A   I   A   D   A   L   S   P   M   W   I   M   Y
1681 ccaaagatcgattgctcggtcagccattgcggatgctctctcgccgatgtggatcatgta 1740
      T   A   F   A   A   A   G   L   I   C   I   L   L   V   S   K   T   E   L   T
1741 taccgcctttgccgcagcaggtctaatttgcatcctcctggtcagcaaaactgaactcac 1800
      T   T   H   E   V   T   E   V   G   L   E   A   Q   K   R   A   E   A   E   R
1801 caccacgcacgaggtgacggaggttggtctcgaagcgcaaaagagggctgaagctgagcg 1860
      K   A   E   Q   E   R   K   D   V   E   K   A   N   K   S
1861 caaagctgagcaggagaggaaggatgttgagaaggcgaataagtcc
```

Appendix XXVI The ORF and amino acid sequence of *Pa-dotC*. The red nucleotide sequence indicated the positions of introns. The blue amino acid sequences indicated the position of the adenine nucleotide binding motif.

Appendix XXVII Alignment of Pa-DotC with other MFS transporter proteins

```
Pa-DotC  MSQDHTKAEDLSEKEQHSPSRSDSGSQHDVVANTAEEESSDDMGALDGKPKSLIAIVMIALS-----LAVFLSALDTTIVTVALPAIAEHFHS-TASYTWVGSAYLLA : 102
Ds-DotC  MSEDHTKADNLSEKDPHSPERSDSSSHEDAHAREEEE-SSDDDGALDGKPASLIAIVMIALSLIGLQLAVFLSALDTTIVTVALPAISAHFNS-TAAYTWVGSAYLLA : 106
Ap-AflT  ------------------------------------MLIDEAAEASSHISGMKLYLIVLS---LLLAVFCVALDNTILSVAIPRITDEFHR-LNDIGWYASAYLLT :  66
Ds28     -----MNSTNSSAEVLKSTNDESSLARYGVHLEGNEVHWNDDASDHPRNWRPSTKYYSAIVIS---WLELYMTGISSAGVSAADTAREEYHMSRTLAYFAFVSIYLLG : 100

Pa-DotC  NAASTPIWGKLADIFGRKPMLLAANALFMIGSLICGLSINVGMLVTARAIQGAAGGGLLTLVDTIIGDLFSLRTRGTYLGMIGGVWAIACALGPIIGGAFTSGVTWRW : 210
Ds-DotC  NAASTPIWGKLADIFGRKPMLLLANALFMIGSLVCALSINVGMLITARAIQGAAGGGLLTLVDTIIGDLFSLRTRGTYLGMIGGVWAIACALGPIVGGAFTSSVTWRW : 214
Ap-AflT  TCAFQLLYGKLYALFSTKWVFLVALCIFEVGSLICGVAPSSVVLIVGRAIAGVGSSGIFTGALVTIAHIVPLAKRPVYMGLLGGMYGIASVAGPLLGGAFTNEVTWRW : 174
Ds28     QTVGGIFLAPISETFGRRTIYIIATSMFCVFSVITAAVPSVAGVYVGRWFQGIAAAIPAIVAFGNFQDMFDARMRIGGVFGYTMSGFIGLSMGPVYAAYITERCGWRW : 208

Pa-DotC  CFYINLPIDGIAFFIILFFLKLKTPKTPLIEGFKAIDWAGSFFVSITA-NSTRLRTLTMLQIIGGTLLFLFGLQYGGQTFPWDSATVICLLVFGIVCVVIFGFVEWKV : 317
Ds-DotC  CFYINLPIDGLAFGIIFFFLKLKTPKTPILEGFAAIDWAGSFFVCIRR-YSSTASRLTFSQIIGGTLMFLFGLQYGGITFPWDSATVICLLVFGVVCIVLFGLVEWKF : 321
Ap-AflT  CFYINLPVGGVTAVVILFLLRIPKSADLRTH-------G-AWEMLKG---LDPLGTIVFTPSI---ICVLLALQWGGVDYAWSNGRIIALFVLFGVLLITFIIIQVLM : 268
Ds28     VFYISAIASAISVVLSYGVKESDAGQLLQAKVKAISEETGRTDLTAGAGSSQDFSIKTFAR-----N--DLLRPL--IFLTTEPIVLFCAILCAIAFGLLYGLTAGLT : 307

Pa-DotC  AKYPIIPLRLFKYRNNCGALLVAFFHSLVFISAFYYLPLYFQAVKGASPILAGVYILPAVLMTGVSAAATGAFIGNTGNYLIPMYFGMTTMVLGYGLLINFQASSGWA : 425
Ds-DotC  ARFPIIPLRLFQYRNNCGALLVAFFHSFVFTSAFYYLPLYFQAVKGATPILAGVYILPAVLSTGVSAAATGAFIGNTGNYLIPMYFGMSMMILGYGLLINFDAGSGWA : 429
Ap-AflT  KDKATVPIKVASQRSVACASVFVFFIGASMFVMIYYVPIWFQAIRNQSPVQAGIDSIALILANTAGAIISGAVTNKTGHYAPWFIVSSVIMSIGAGCLTLFTVDIAQS : 376
Ds28     VAYTDPPFSNTFNEVSSSLSFIAILIGILLDVIPRFYDDHLYRKCHKN-------NIRIVPETKIRSFALACPLFAIGLWIFAWTVPPKVTTVPWPVSMIGLICIGFA : 408

Pa-DotC  KLIIYQLLVGIGNGPNFQAPLVALQTKIKQSDIATGTATFNFVRNIATAVSVVVGQVVYANQLSGMKNRLQQLGSAAGLIASGDAGANVEMIQDLPRDQRSIARSAIA : 533
Ds-DotC  KLIIYQLIAGIGNGPNFQAPLVALQTKIKQSDIATGTATFNFVRNIATAISVVAGQVLYQNQLKKMTSTLQQLGPAASLIAAGDAGANTQAINALPTPQRDLARSAIA : 537
Ap-AflT  KWIGFLFLYGIGVGFGFQQGAVAVQAVLPMAQVPIGTALIWFVQMLGGALFTSVAQNIFSTHLAENLANLQLPGLDPEAIVGAGATGFRQLVQPEYMDQVLVAYNAAL : 484
Ds28     TTDFSYVLFGYVTDSYGEYAASAVSALSTTRTIAAAVFPLFAYQMFSGLGTNIAATILAAVATLFAFTPILFLKYGHALRHKSKTAGNDEDCLEEENSHLELDEKCTD : 516

Pa-DotC  DALSPMWIMYTAFAAAGLICILLVSKTELTTTHEVTEVGLEAQKRAEAERKAEQERKDVEKANKS : 598
Ds-DotC  DALSPMWIMYTAFAAAGLFCILLVSKTELTTTHEVTEVGLEAQKKAEAERKAERQAKDLEKAQKS : 602
Ap-AflT  LDVFQVALICSCLSILGAVGIEWRSVKQNR---------------------------------- : 514:
Ds28     SEGTATDDAV-------------------------------------------------- - : 526
```

Name	Organism Protein	Protein	Genbank No.
Pa-DotC	*P. arachidicola*	DotC	
Ds-DotC	*D. septosporum*	DotC	AAL87047
Ap-AflT	*A.parasiticus*	AflT	ABS57485.1
Ds-28	*D. septosporum*		AAS66020.1

Appendix XXVII Alignment of Pa-DotC with other MFS transporter proteins

Appendix XXVIII Nucleotide and amino acid sequence of *Pa-dotD*

```
        M  S  T  T  T  T  A  P  A  P  T  T  P  S  K  P  S  H  P  A
      1 atgtccaccacaaccacagcaccagcacccacaacgccctccaaaccatcgcatccagcc 60
        L  A  T  D  I  F  S  P  P  T  F  D  E  K  Q  W  K  R  A  L
     61 ctggcaacggacatcttctcccctccaaccttcgacgaaaaacaatggaaacgcgccctt 120
        I  A  A  D  H  A  L  K  K  E  K  S  A  L  C  R  T  S  C  P
    121 atcgcagcagaccacgccctcaaaaaggaaaagtcagccctctgcagaacctcctgccca 180
        T  T  S  S  H  P  P  A  T  S  V  L  L  Q  G  N  P  R  T  A
    181 acgacctcatcccacccgccagcaacatccgtgctcctccaaggaaaccctcgcaccgcc 240
        A  R  N  L  F  L  F  P  D  G  S  G  S  A  A  S  Y  T  Q  L
    241 gccagaaacctcttcctctttcccgacggctccggttcagcagcatcctacacccaactc 300
        S  L  I  S  S  D  L  A  V  Y  G  L  N  C  P  Y  L  K  S  P
    301 tcgctcatctcttccgatctggcagtctacggcctcaattgcccttacctcaaatcaccc 360
        E  D  W  K  C  G  P  Q  E  L  T  P  L  F  I  S  E  I  Q  R
    361 gaagactggaaatgcggtccccaagagttaacaccgctgttcatatccgagatccagcgc 420
        R  Q  A  S  G  P  Y  Y  I  G  G  Y  S  T  G  G  I  A  A  F
    421 cgccaggcttccggaccctattatatcggcggctactccaccggcggaatcgcagcgttc 480
        D  A  A  Q  A  L  D  R  M  A  E  R  V  E  R  L  I  L  I  D
    481 gacgccgcacaagcgctggatagaatggcggagagagtcgagaggttgattctcatcgac 540
        A  P  C  P  I  H  I  Q  R  L  P  G  R  L  M  E  Y  L  K  G
    541 gcgccctgtccaatccacatccagcgcttacctggaagattaatggagtacctgaaagga 600
        V  H  A  F  N  S  R  G  R  A  P  P  R  C  V  F  E  H  F  E
    601 gtccacgccttcaattcccgcggccgcgcaccaccacgctgtgtcttcgaacatttcgaa 660
        A  N  T  K  N  L  Q  K  Y  R  P  R  P  C  E  A  Y  R  E  P
    661 gcgaataccaagaacctccagaaataccgaccaaggccttgcgaagcatatcgcgagccg 720
        S  T  H  I  I  Y  A  R  H  G  V  C  E  S  L  E  E  G  V  P
    721 tcaacacacattatctatgcccggcatggagtctgcgagagtcttgaagagggcgtcccg 780
        Q  M  E  I  F  E  D  D  P  K  E  M  Q  W  I  M  C  A  R  R
    781 cagatggaaatctttgaggacgatccgaaagagatgcagtggattatgtgtgcgaggaga 840
        D  F  G  P  M  G  W  E  K  L  L  N  D  E  D  I  H  V  Q  V
    841 gattttggtccgatggggtgggagaagttattgaatgatgaggatattcatgtgcaggtt 900
        V  E  A  A  N  H  F  T  M  M  K  G  E  A  A  G  R  L  A  G
    901 gtggaggcggcgaatcattttacgatgatgaaggggaggctgcgggaaggttggcaggg 960
        C  I  E  R  A  I  G
    961 tgtattgagagggcgattgga
```

Appendix XXVIII The ORF and amino acid sequence of *Pa-dotD*.

Appendix XXIX Alignment of Pa-DotD with Ds-DotD and C-terminal of Pks proteins

```
Pa-DotD : MSTTTTAPAPTTPSKPSHPALATDIFSPPTFDEKQWKRALIAADH :  45
Ds-DotD : MSAAVTSAASIGVPKAPHPALNMDI---PALDEKQVKRALIAADH :  42
An-StcA : -LGMSLDSSFNLFEEVPTVARLQEFFGTTSGSTTGSSGSGSSEDE :  44
Ap-AflC : ------------------PGINPET----DWSSSASDSIFASEDH :  23

Pa-DotD : ALKKEKSALCRTSCPT--TSSHPPATSVLLQGNPRTAARNLFLFP :  88
Ds-DotD : DSKKKLKRSRSSQSP----STHPPATSVLLQGNPRQATKNLFLFP :  83
An-StcA : TDSIPSTPEAYTTADTR-VPECRPTTSVVLQGLPQMAKQILFMLP :  88
Ap-AflC : GHSSESGADTGSPPALDLKPYCRPSTSVVLQGLPMVARKTLFMLP :  68

Pa-DotD : DGSGSAASYTQLSLISSDLAVYGLNCPYLKSPEDWKCGPQELTPL : 133
Ds-DotD : DGSGAAASYTHLTLISRDLAVYGLNCPYLRSPQDWKCGPQDLTPL : 128
An-StcA : DGGGSASSYLTIPRLHADVAIVGLNCPYARDPENMNCTHQSMIQS : 133
Ap-AflC : DGGGSAFSYASLPRLKSDTAVVGLNCPYARDPENMNCTHGAMIES : 113

Pa-DotD : FISEIQRRQASGPYYIGGYSTGGIAAFDAAQALDRMAERVERLIL : 178
Ds-DotD : FISEIQRRQPSGPYYIGGYSTGGIAAFDAAQALDKLGEKVERLIL : 173
An-StcA : FCNEIKRRQPEGPYHLGGWSSGGAFAYVTAEALINAGNEVHSLII : 178
Ap-AflC : FCNEIRRRQPRGPYHLGGWSSGGAFAYVVAEALVNQGEEVHSLII : 158
                                  ______
                                  GXSXGG
Pa-DotD : IDAPCPIHIQRLPGRLMEYLKGVHAFNSR-------GRAPPRCVF : 216
Ds-DotD : IDSPCPIHIQRLPSRLMDYLKRVHAFNSR-------GRPPPAWVF : 211
An-StcA : IDAPVPQVMEKLPTSFYEYCNNLGLFSNQPGGTTDGTAQPPPYLI : 223
Ap-AflC : IDAPIPQAMEQLPRAFYEHCNSIGLFATQPGASPDGSTEPPSYLI : 203

Pa-DotD : EHFEANTKNLQKYRPRPCEAYREPSTHIIYARHGVCESLEEGVPQ : 261
Ds-DotD : EHFEANTTNLQKYRPRPFEAYKEPRTHIIYARQGVCESFEAGVPQ : 256
An-StcA : PHFQATVDVMLDYRVAPLKTNRMPKVGIIWASETVMDEDNAP--- : 265
Ap-AflC : PHFTAVVDVMLDYKLAPLHARRMPKVGIVWAADTVMDERDAP--- : 245

Pa-DotD : MEIFEDDPKEMQWIMCARRDFGPMGWEKLLNDEDIHVQVVEAANH : 306
Ds-DotD : MEILEEDPKEMKWIMCARSDFGPLGWEKLLNEEEIFVEIVEGANH : 301
An-StcA : ------KMKGMHFMVQKRWDFGPDGWDVVCPGAVFDILRAEGANH : 304
Ap-AflC : ------KMKGMHFMIQKRTEFGPDGWDTIMPGASFDIVRADGANH : 284
```

```
Pa-DotD : FTMMKGEAAGRLAGCIERAIG : 327
Ds-DotD : FGMMRGDAAERLAGCIGRAVA : 322
An-StcA : LR------------------- : 306
Ap-AflC : FTLMQKEHVSIISDLIDRVMA : 305
```

Name	organism	Protein	Genbank No.
Pa-DotD	*P. arachidicola*	Pa-DotD	
Ds-DotD	*D. septosporum*	Pa-DotD	AAL87048.1
An-StcA	*A. nidulans*	StcA	AAA81586
Ap-AflC	*A. parasiticus*	AflC	AAS66004

Appendix XXIX Alignment of Pa-DotD with Ds-DotD and C-terminal of Pks proteins. The black line indicates the position of thioesterase domain in Pa-DotD

Appendix XXX Nucleotide and amino acid sequence of partial *Pa-mfs*

```
      T  G  H  S  V  V  L  A  Y  L  I  L  F  Q  F  G  G  T  L  F
    1 gactggtcacagtgtggttctggcatacctcatcctattccagttcggcggcactcttt 60
      I  R  T  K  L  A  R  E  N  K  K  R  R  N  G  E  Q  D  H  L
   61 catccggaccaagctggcgcgggagaacaagaaacgcagaaatggcgagcaagatcattt 120
      L  D  G  K  T  E  D  E  I  V  V  A  G  D  K  R  P  D  F  M
  121 gctcgatgggaagacggaagatgagattgtagttgctggtgataagaggcctgacttcat 180
      Y  T  L
  181 gtacactctc
```

Appendix XXX Nucleotide and amino acid sequence of partial *Pa-mfs*.

Appendix XXXI Alignment of Pa-Mfs with other Mfs proteins

```
Pa-Mfs : TGHSVVLAYLILFQFGGTLFIRTKLARENKKRRNGEQDHLLDGKT : 45
Ac-Mfs : PGHGTVLAYLVLFLFGGSVLQYLLLRRENRKRLNRERDHWVQGLD : 45
Af-Mfs : PGHGTVLAYLVLFQFGGSLVQYLLLRLENRKRLRGDRDHWIQGLD : 45

Pa-Mfs : EDEIVVAGDKRPDFMYTL : 63
Ac-Mfs : QSQIELLGDQRPDFIYTL : 63
Af-Mfs : RSQIELLGDKRPDFIYTL : 63
```

Name	Organism Protein	Protein	Genbank No.
Pa-Mfs	*P. arachidicola*	Mfs	
Ds-Mfs	*D. septosporum*	Mfs	XP001269961
Af-Mfs	*A. fumigatus*	Mfs	XP749930

Appendix XXXI Alignment of Pa-Mfs with other Mfs proteins