INHERITANCE OF PISATIN DEMETHYLASE IN THE *FUSARIA*

By

Nicholas A. Milani

A Thesis Submitted to the Faculty of the

DIVISION OF PLANT PATHOLOGY
SCHOOL OF PLANT SCIENCES

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2010

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Nicholas Milani

APPROVAL BY THESIS DIRECTOR
This thesis has been approved on the date shown below:

Hans VanEtten — 2-18-2010
Dr. Hans VanEtten — Date
Professor of Plant Sciences

ACKNOWLEDGEMENTS

I wish to deeply thank my major advisor, Dr. Hans VanEtten, for all of his support, which I could not have done without. I also wish to thank Dr. Betsy Arnold for assisting me with most of the analyses. Her help was indispensable. I also wish to thank Dr. Steve Rounsley for his great help in providing me data to analyze for this thesis.

Aside from my committee, mentioned above, I wish to thank Daniel Lawrence who helped me with phylogenetic analyses and has become my best friend and has always stuck by my side and supported me.

DEDICATION

I dedicate this to my mother and father, who have always supported me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Host specificity in plant pathogens is dictated by the presence of certain genes that enable the plant pathogen to circumvent host defenses leading to pathogenicity. Upon recognition of the pathogen, plants initiate defense responses that can include the production of antimicrobial compounds such as phytoalexins. *Nectria haematococca* mating population VI (MPVI) is a filamentous ascomycete that contains a cluster of genes known as the pea pathogenicity (PEP) cluster in which the pisatin demethylase (*PDA)* gene resides. PDA is responsible for the detoxification of the phytoalexin, pisatin, produced by the pea plant (*Pisum sativum*). Thus, PDA allows the fungus to colonize and become a pea pathogen. It has been proposed that the evolutionary origin of *PDA* and the PEP cluster is explained by a horizontal gene transfer (HGT) event. The observations supporting this hypothesis include the location of the PEP cluster on a conditionally dispensable (CD) chromosome, the phylogenetically discontinuous distribution of the cluster among closely related species, and bias in GC content and codon usage. This study used a three-pronged approach to test the hypothesis that *PDA* and the PEP cluster were inherited via HGT. Percent identities of PDA were measured and compared against values from nine putative orthologs that were highly conserved, and two beta-tubulin introns, across *F. oxysporum* f.sp *pisi*, *Neocosmospora boniensis* and *N. haematococca* MPVI. The GC content of *PDA* and the PEP cluster was analyzed and compared to those orthologs as well. Finally, phylogenetic analyses based on the orthologous

sequences and *PDA* was carried out to determine the evolutionary inheritance of *PDA*. The results of phylogenetic and percent identity analyses support vertical inheritance. However, GC content analysis does not rule out the possibility that *PDA*, the PEP cluster, or even the entire CD chromosome could have been the result of an ancient HGT event.

I. INTRODUCTION

In response to stress and pathogens, plants activate a number of defense responses, one of them being the production of phytoalexins (***Agrios***, ***2005***). The first phytoalexin isolated was pisatin (Figure 1), a low molecular weight, toxic compound that disrupts the plasmalemma of invading microorganisms and plant cells alike (***Smith,1978***; ***Bailey and Mansfield, 1982***). Pisatin initially was found in the garden pea (*Pisum sativum* L.) (***Cruickshank and Perrin, 1960; 1962***) and later shown to be produced by select members of the plant family Fabaceae such as sweet pea (*Lathyrus oderatus*) (***Robeson and Harborne, 1980***; ***Bailey and Mansfield, 1982***).

Phytoalexins have been documented as non-specific resistance compounds that are produced *de novo* in response to the invasion of plant tissues by microbial organisms (***Cruickshank and Perrin, 1960***). The mode-of-action of all phytoalexins, including pisatin, is far from being understood. However the effect of phytoalexins on cells has been observed with both light and electron microscopes (***Smith, 1976***). It is known that because phytoalexins are lipophilic compounds (***Bailey and Mansfield, 1982***), they are capable of integrating into the lipid-based plasmalemma of any cell in a non-specific manner, thus compromising cellular structural integrity and resulting in cell lysis (***Smith, 1976; 1978***). Once inside a cell, phytoalexins cause additional detrimental effects, such as inhibition of spore germination, inhibition of germ-tube elongation, inhibition of radial mycelial growth, decrease of mycelia biomass, granulation of the

cytoplasm, disruption of cell contents, and cessation of cytoplasmic streaming (***Smith, 1976***). The detrimental effects mentioned above cannot be fully explained yet as there is still a huge amount of knowledge that remains to be acquired (***Bailey and Mansfield, 1982***), but the one thing agreed upon by those who study phytoalexins, is that the mode-of-action of these plant antimicrobial compounds is attributed to their ability to integrate into the plasma membrane and cause structural damage to the cells' plasmalemma (***Shiraishi et al., 1975; Bailey and Mansfield, 1982***). Pisatin effects fit the generalizations of the phytoalexins mentioned above, as its mode-of-action is thought to be similar to that of phaseollin, another pterocarpanoid phytoalexin made by bean plants (***VanEtten and Bateman, 1971***; ***Shiraishi et al., 1975***).

In order for pathogens to overcome harmful plant defense compounds such as pisatin, the pathogens have evolved several tolerance mechanisms against their toxic effects that allow them to colonize their host (***VanEtten, et al., 2001***). For example, a few fungi in the phylum Ascomycota have evolved a way to detoxify and efflux pisatin from their cells. These fungi are effective colonizers and pathogens of pisatin-producing plants such as pea (*P. sativum*). In fact, the enzymatic degradation of pisatin constitutes the first report of phytoalexin degradation (***Uehara, 1964***). In this study, the pea pathogens *Ascochyta pisi* and *Fusarium oxysporum* were demonstrated to degrade pisatin to a non-toxic compound (***Uehara, 1964***). Fungi with the ability to detoxify pisatin contain PDA (pisatin demethylating activity), which catalyzes a one-step demethylation

reaction that converts pisatin into a much less toxic compound, 3,6a-dihydroxy-8,9-methylenedioxypterocarpan (DMDP) (***Delserone et al., 1999; George and VanEtten, 2001***) (Figure 1).

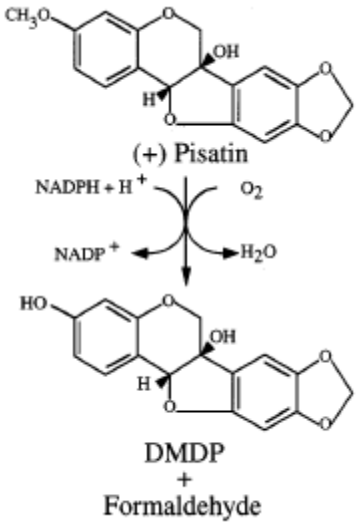

Figure 1. Demethylation of pisatin into 3,6a-dihydroxy-8,9-methylenedioxypterocarpan (DMDP). This reaction assists the fungi in their ability to colonize and become pathogenic on the plant. **(*Mathews et al., 1983*)**

To date, other pea pathogens such as *Colletotrichum pisi*, *Mycosphaerella pinodes*, *Nectria haematococca* Mating Population VI (MPVI), *Neocosmospora boniensis*, and *Phoma pinodella* have been shown to convert pisatin to DMDP (Figure 1) (***VanEtten and Stein, 1978; Delserone et al, 1999; George and VanEtten, 2001; Temporini and VanEtten, 2004***). In the fungus *N. haematococca* MPVI (ana. *Fusarium solani;* Hypocreales, Sordariomycetes, Ascomycota), *PDA* encodes a cytochrome P450 monooxygenase enzyme (***Mathews and VanEtten, 1983***) that due to its evolutionary divergence from other known P450s constitutes the first member of a new cytochrome P450 family, termed CYP57A (***Maloney and VanEtten, 1994***). Further work has shown that this CYP57A gene is present in *N. boniensis* (no known anamorph) (***Temporini and VanEtten, 2004***) and *F. oxysporum* f.sp. *pisi* (no known teleomorph) (***Delserone et al, 1999***).

The discovery that PDA is a cytochrome P450 monooxygenase enzyme (P450, CYP) (***Mathews and VanEtten, 1983***;***George et al., 1998***) made it possible to screen other fungi for similar PDA activity. By studying the inactivation of CYP450s via carbon monoxide and antibodies to cytochromone P450 reductase (***George and. VanEtten, 2001***), it was possible to determine that other pea pathogens such as *P. medicaginis*, *M. pinodes*, and *A. pisi* have a cytochrome P450 monooxygenase that demethylates pisatin. Furthermore, induction specificities, substrate specificities, and low Kms for pisatin as a substrate all suggest that pisatin was the natural substrate for these P450

monooxygenases (**George *and VanEtten, 2001***). These observations are consistent with the hypothesis that pea pathogens have evolved a similar biochemical mechanism for detoxifying the phytoalexin from their host plant. However, the *PDA* genes encoding the cytochrome P450s in *P. medicaginis*, *M. pinodes*, and *A. pisi* are not detected in Southern blot analyses using *PDA* probes from *N. haematococca* MPVI (***Delserone et al., 1999)*** suggesting that they are not highly similar to the cloned *PDA* genes from *N. haematococca* MPVI, *N. boniensis* and *F. oxysporum* f.sp. *pisi.*

The Pea Pathogenicity (PEP) gene cluster.

The most extensive studies examining the evolutionary origin of *PDA* have been on *N. haematococca* MPVI. Results of these studies suggest that a horizontal gene transfer (HGT) event might have accounted for the presences of *PDA* in this fungus. The *PDA* gene resides in a cluster of six genes, all of which are expressed during infection of pea (***Han et al., 2001; Liu et al., 2003***). Four of these genes, termed *Pea Pathogenicity (PEP)* genes (*PEP1*, *PEP2*, *PEP5* and *PDA1*), function independently as virulence determinants when transformed into an isolate of *N. haematococca* MPVI that lacks this cluster (***Han et al, 2001; Liu et al, 2003***). This is consistent with the idea that gene clusters frequently enhance fitness in fungi and often increase virulence or success on a host plant (***Van der Does and Rep, 2007***). *PEP5* has homology to a major facilitator transporter and *PEP4* has sequence similarity to transposases, however, there is no significantly similar sequence to known genes for the other genes present in the cluster (***Han et***

al., 2001; Liu et al., 2003). In addition to *PEP4*, there are four other regions in this cluster that have sequence similarities to transposases. Moreover, the *PEP* genes have a different GC content and codon usage than other *N. haematococca* MPVI genes (***Han et al., 2001; Liu et al., 2003***). These features, a cluster of pathogenicity genes, presence of transposons, a different GC content and codon usage are all observed in what were originally called "pathogenicity islands", now referred to as genomic islands, in prokaryotic organisms (***Hacker & Kaper, 2000; Ochman et al., 2000***). These features were some of the earliest evidence for HGT in prokaryotic organisms. Therefore, the evolutionary origin of the PEP cluster is hypothesized to be linked to a HGT event. Additionally, it is hypothesized that gene clusters in fungi help to facilitate HGT by allowing a recipient organism to acquire a complete biosynthetic pathway cluster in a single horizontal transfer (***Walton, 2000***).

Phylogenetically discontinuous distribution of PEP genes.

Strains of *N. haematococca* MPVI can exist as a pathogen of plants and animals and as a root-associated or free-living saprobe. However, individual members of the species are restricted to specific habitats. All genotypes that are pathogenic on pea have a *PEP* cluster, while isolates that are not pathogenic on pea lack an intact cluster (***Temporini and VanEtten, 2002***). Thus, the *PEP* cluster has a discontinuous distribution within the species. As indicated above, *PDA* genes in the CYP57A family have been demonstrated in another *Fusarium* species in addition to *N. haematococca* MPVI, *F. oxysporum* f.sp. *pisi* (***Delserone***

et al. 1999). A phylogenetic analysis of the *PEP* genes was undertaken to determine the frequency of this cluster in the Fusaria and identified *N. boniensis*, but few others, as containing all of the *PEP* genes (***Temporini, and VanEtten, 2004***) (Figure 2). The discontinuous distribution of the *PEP* genes was demonstrated by amplifying 28s rDNA from various Fusaria species and building a phylogeny, along with Southern blot analyses to show presence or absence of individual *PEP* genes. Additional studies demonstrated that *N. boniensis* was also a pea pathogen (***Temporini and VanEtten, 2004***). Thus, there is a phylogenetically discontinuous distribution of *PDA* and *PEP* genes within a species and among closely related species, with only the isolates pathogenic on pea carrying all the *PEP* genes involved in pathogenicity on pea. The linkage of a group of genes to a particular phenotype and a discontinuous distribution of these genes within a species and between species are two characteristics of HGT in prokaryotic microorganisms (***Lawrence, 2005; Ochman et al., 2000***).

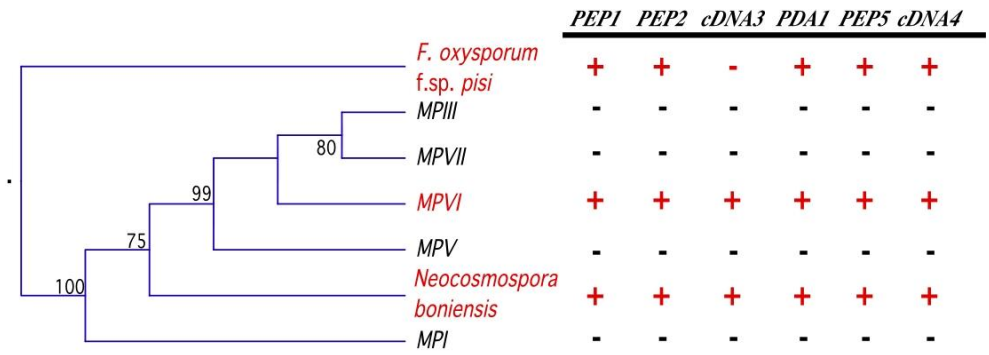

Figure 2. 28s phylogeny of Fusaria. Discontinuous distribution of *PEP* genes among the Fusaria (MP = mating population of *N. haematococca*). The phylogeny of the three taxa shows that the PEP cluster is limited to a few fungal species, and creates a discontinuous distribution. (Adapted from ***Temporini and VanEtten, 2004***)

Dispensable chromosomes in fungi.

The PEP cluster resides on a dispensable portion of the genome of *N. haematococca*, a supernumerary chromosome called a conditionally dispensable (CD) chromosome (***Han et al., 2001***). The CD chromosomes of *N. haematococca* MPVI have been compared to bacterial plasmids, which can be lost and gained, with no growth effects, but when they are present they allow for niche expansion (***Miao et al., 1991;VanEtten et al., 1994***). CD chromosomes contain genes that benefit the fungus by allowing it to grow in some habitats, but are not essential for growth in culture or some natural habitats (***Covert, 1998***).

Such chromosomes have also been observed in *Alternaria alternata* which has a 1.05 Mb-CD chromosome that harbors host specificity genes (***Hatta et al., 2002***). Each of seven pathotypes of *A. alternata* contains this chromosome, and the genes for a distinct host specific toxin responsible for each of the seven pathotypes are on the CD chromosome. Due to the presence of these genes for these toxins on the CD chromosomes, *A. alternata* can be a pathogen on Japanese pear, apple, tomato, strawberry, rough lemon, tangerine, and tobacco (***Akamatsu et al. 1999***).

In comparison, *N. haematococca* MPVI contains two distinct CD chromosomes that confer pathogenicity: one for pea (*Pisum sativum*) containing the PEP cluster, and one for chickpea (*Cicer arietinum*), containing the gene for detoxifying the chickpea phytoalexins (medicarpin and maackiain) (***Han et al., 2001;*** **Covert et al. 1996**). Additionally, CD chromosomes appear to be

transferable among strains of the same fungal species (***Hatta et al., 2002***), and perhaps may have been gained by HGT from a nonsexual compatible donor species (***Khaldi et al., 2008***). Thus, the presence of genes on a dispensable portion of the genome is evocative of genes obtained by HGT (***Waterhouse and Russell, 2006***). Since the PEP cluster resides on a dispensable portion of the genome it suggests a possible horizontal transfer of *PEP* genes or the entire CD chromosome (***Han et al., 2001***).

Thesis research objectives.

The dispensability of the *PEP* genes, their discontinuous phylogenetic distribution, the difference in their GC content and codon usage and their presence in a cluster in *N. haematococca* MPVI supports the hypothesis of a foreign origin of these genes (***Temporini and VanEtten, 2004***). In addition, the biochemical studies that indicate the other more distantly related pea pathogens (*P. medicaginis*, *M. pinodes*, and *A. pisi)* lack CYP57A *PDA* homologs but have paralogous cytochrome P450s with PDA activity may indicate that these pea pathogens have convergently evolved their pisatin-detoxifying ability by gene recruitment. Thus, a working hypothesis is that the CYP57A *PDA* homologs in *N. haematococca* MPVI, N. *boniensis*, and *F. oxysporum* f.sp. *pisi* were obtained by HGT from an unknown donor and the *PDA* genes in *P. medicaginis*, *M. pinodes*, and *A. pisi* were obtained by gene recruitment. However, there are alternative explanations. For example, codon bias differences may be due to endogenous heterogeneity within individual fungal genomes. The presence of the *PDA* and

PEP genes in *N. boniensis* and *F. oxysporum* f .sp. *pisi,* and their absence in the intervening *Nectria* lineages that are not pathogens of pea could be due to the loss of these vertically inherited genes in these lineages, thus leading to non-pea-pathogenic fungi.

The research in this thesis was directed to determine whether the *PDA* genes and the *PEP* cluster were acquired by HGT. A three-pronged approach was taken to test the method of inheritance of *PDA* and the *PEP* cluster. First, a comparison of the percent of identities between nine orthologs, two beta-tubulin introns and *PDA* was made at the nucleotide and amino acid levels (when applicable) among *F. oxysporum* f.sp. *pisi, N. boniensis,* and *N. haematococca* MPVI. If HGT in fact occurred, it was expected that the percent identities of *PDA* would be higher than that of the orthologs, as it was likely a relatively more recent acquisition. Secondly, GC content was analyzed and compared for *PDA* and nine orthologs, as well as comparison of the GC content of the entire *PEP* cluster to the genome. A difference in GC content could be a reflection of the donor organism's GC content in the case of an HGT event. Thirdly, a phylogenetic approach was taken to demonstrate whether or not the *PDA* genealogy is concordant with the genealogies of the orthologous genes, which should produce phylogenies of the three taxa that match known species relationships. If HGT was involved one would expect that ortholog genealogies would be discordant with the *PDA* topology, displaying a different method of inheritance of *PDA* compared to the orthologs.

II. MATERIALS AND METHODS

Fungal isolates and culture conditions

N. boniensis NRRL 22470 culture was obtained originally from Kerry O'Donnell (***Temporini and VanEtten, 2004***), and I acquired it from H. VanEtten (University of Arizona, Tucson, AZ). The isolate was transferred to potato dextrose agar (Difco Laboratories, Sparks, MD) in a 100 mm Petri plate and grown at room temperature for two weeks.

For DNA extraction, a mycelial plug was taken from the culture and placed into a 250 ml- flask containing 125 ml of potato dextrose broth (Difco Laboratories, Sparks, MD). Broth cultures were incubated at 37°C at 100 rpm for two weeks on a rotary shaker (VWR). The mycelial mass was then placed in miracloth (CalBiochem) and rinsed to eliminate the medium. The rinsed mycelium was lyophilized and used for total genomic DNA extraction following the procedure of Arnold and Lutzoni (2007). DNA was stored at -40°C.

Selection of orthologs

At the beginning of this study, genome sequence data for *F. oxysporum* f.sp. *pisi* was not yet available. Thus, *F. oxysporum* f.sp. *lycopersici* was used as a proxy to carry out a reciprocal blast against *N. haematococca* MPVI to identify putative orthologs. This was done by aligning these sequences in MEGA VERSION 4 (***Tamura et al., 2007***) and BLASTing the two sequences against each other in NCBI. The selected sequences to carry out the analyses are shown

in Table 1. The sequence data for *N. haematococca* MPVI was obtained from the Joint Genome Institute (JGI) site (http://genome.jgipsf.org/Necha2/Necha2.home.html) and for *F. oxysporum* f. sp. *lycopersici* from the Broad Institute of MIT and Harvard (http://www.broadinstitute.org/annotation/genome/fusarium_group/MultiHome.html) (Table 2).

Nine highly conserved putative orthologs, including the EF1-alpha gene and the two beta-tubulin gene introns were used for the comparisons of percent identities between these 11 sequences and *PDA* across the following taxa: *F. oxysporum* f.sp *pisi*, *N. boniensis* and *N. haematococca* MPVI. The EF1-alpha gene and the beta-tubulin introns were included because these are sequences that are traditionally used to establish phylogenetic relationship. EF1-alpha has been used in phylogenetic tree building extensively, and has been evaluated for its usefulness in revealing expected evolutionary relationships. With the exception of a few groups, all eukaryotic groups are monophyletic and produce expected phylogenetic species relationships when using EF1-alpha alone (***Roger et al., 1999***). The use of introns is thought to be very useful in evaluating closely related species because they evolve more rapidly than exons and show resolution when dealing with closely related taxa (***Ning and Runsheng, 1999***). Because *F. oxysporum* f.sp *pisi*, *N. boniensis* and *N. haematococca* MPVI are closely related taxa, the beta-tubulin introns were included in the phylogenetic analyses carried out during this study. As mentioned earlier, at the beginning of this project, the

only taxa of the three species for which the genomic sequence was available was *N. haematococca* MPVI. However, the genomic sequence of *F. oxysporum* f.sp. *lycopersici* was available and it was used as a proxy to find the eight putative orthologs to use in this study. The complete protein genome of *F. oxysporum f.sp lycopersici* was used to run a reciprocal BLAST against *N. haematococca's* protein genome. An excel file was the output from this, with a list of 8,000 putative orthologs, that were arranged so the genes with the highest identities were at the top of the list. The orthologs were selected from this list by starting at the top (most conserved) and moving down the list approximately every 400th gene, to get a total of eight orthologs (Table 1). The sequences of the eight orthologous genes identified in the reciprocal blast of *F. oxysporum* f.sp. *lycopersici,* the EF1-alpha gene and the two beta-tubulin gene introns were retrieved from JGI site for *N. haematococca* MPVI (http://genome.jgipsf.org/Necha2/Necha2.home.html). These orthologous sequences for *F. oxysporum* f.sp. *pisi*, were given by Dr. Lijun Ma at The Broad Institute (MIT and Harvard, 320 Cambridge, MA 02141) and they were obtained by PCR amplification for *N. boniensis*. *N. haematococca* was used as a sole sequence to design primers (Table 1) for *N. boniensis* as *N. boniensis* and *N. haematococca* are very closely related (***O'Donnell, 2000***).

Table 1. Orthologs selected for comparisons among *Nectria haematococca* MPVI, *Neocosmospora boniensis* and *Fusarium oxysporum* f.sp. *pisi.* Also shown are primers used to amplify the *N. boniensis* orthologs.

Sequence ID	Sequence Description	Primer name	Primer Sequence (5'-3')	Annealing Temperature	PCR product size
1	Serine/Threonine Phosphatase	Nb1F	TCTGGAAGACCTTTACCGATTGC	60ºC	967 bp
		Nb1R	TCGTTATCAAGTTCGTTCGCATAGAGGC		
2	Hypothetical. Similar to glutathione-independent formaldehyde dehydrogenase	Nb2F	ATGAAGGCTGTCAACTAYCARGG	54ºC	955 bp
		Nb2R	CGRTCAACCTCCTCTCCRYCGTT		
3	Peroxisomal co-A synthetase	Nb3F	TGGCATTGACTCTCCAAACCG	60ºC	752 bp
		Nb3R	TGTCTGCTTGATGTTGTCTGGTAGG		
4	Conserved hypothetical protein: Glycolate oxidase	Nb4F	TAGGGCAGAGAGGCTTGATGGAAC	59.4ºC	760 bp
		Nb4R	TGAGGTAGAGTTGCTGGAAGACG		
5	Conserved hypothetical protein: inorganic phosphate transporter	Nb5F	TATTCYRTACGCCAGAACCGCGG	55ºC	828 bp
		Nb5R	CCTGRCCGATSGGCTGCATCATG		
6	Polyphosphoinositide phosphatase	Nb6F	GCTCCCATTTCCTGACATTGC	58.6ºC	764 bp
		Nb6R	TCTGGCTTGAACTTGGCTGG		
7	Conserved hypothetical protein: Peroxidase Oxygenase	Nb7F	CAAGCAGGACCACGACCTTTTC	59.4ºC	799 bp
		Nb7R	AAGTGACAGCATCAGACAGCAC		
8	Conserved hypothetical protein: Sec5 subunit of exocyst complex	Nb8F	AGGGAGTATGGTGTCCTGGGTATC	59.4ºC	1017 bp
		Nb8R	AGAAGGTGAGTTTGGCGTTGG		
9	Elongation factor 1 alpha	EF12F	CCTATTTTTCGCAAGGTCGACG	60.1ºC	777 bp
		EF12R	CCCCAATCAAAAGCCGAGTG		
10	Beta-tubulin intron 1	Bt2a*	GGTAACCAAATCGGTGCTGCTTTC	55ºC	288 bp
		Bt2b*	ACCCTCAGTGTAGTGACCCTTGGC		
11	Beta-tubulin intron 2	Bt121*	CCACCTGTCTCCGTTTCCCCG	55ºC	347 bp
		Btt22*	TCTGGATGTTGTTGGGAATCC		

*Primers retrieved (http://www.lutzonilab.net/primers/page246.shtml)

Table 2. Loci and transcript identification number from the Broad Institute and JGI site to retrieve sequences of interest used in the analysis of *F. oxysporum* f.sp *pisi*, *F. oxyporum* f.sp *lycopersici* and *N. haematococca*.

Taxon	*F. oxysporum* f.sp. *pisi* / *F. oxysporum* f.sp. *lycopersici*	*N. haematococca*
Source	**Broad Institute (locus ID)**	**JGI (transcript ID)**
Gene 1	FOXG_02183	60399
Gene 2	FOXG_11415	42983
Gene 3	FOXG_05037	69658
Gene 4	FOXG_11181	30206
Gene 5	FOXG_00090	103059
Gene 6	FOXG_11341	42947
Gene 7	FOXG_10645	704898
Gene 8	FOXG_06281	911349
EF1-alpha	FOXG_03515	59329
Beta-tubulin	FOXG_06228	66759

The *PDA* sequence acquisition sources and accession numbers corresponding to *N. haematococca* MPVI and *F. oxyporum* f.sp *pisi* are shown in Table 3 below.

PCR amplification, sequencing, sequence editing.

PCR primers targeting eight putative orthologous genes, elongation factor-1 alpha (EF1-alpha), and two beta-tubulin introns were designed to amplify these sequences from *N. boniensis* (Table 1). *N. haematococca* MPVI genomic sequence data are available at JGI (http://genome.jgi-psf.org/Necha2/Necha2.home.html). All primers except the beta-tubulin introns primers were designed using MacVector v. 10.6 using *N. haematococca* MPVI as the sole sequence. For beta-tubulin introns, primers (Table 1) were retrieved from the Lutzoni Lab (François Lutzoni, Professor at the Department of Biology, Duke University, Box 90338, Durham, North Carolina, USA 27708; http://www.lutzonilab.net/primers/page246.shtml).

PCR amplifications were performed in reactions containing 2 µl of template DNA at ~ 100 µg/ml, 12.5 µl of 2X Master Mix (Fermentas Inc., Glen Burnie, Maryland), 8.5 µl of sterile nuclease free water in a final volume of 25 µl. PCR cycling conditions were the same except varying the annealing temperatures as indicated in table 1. The PCR program was set to the following conditions: 96°C for 4 min, followed by 40 cycles of 94°C for 30 secs, specific annealing

temperature for 30 secs (Table 1), 72°C for 1 min and a final extension of 72° for 5 min.

PCR products were analyzed by agarose gel electrophoresis, loading DNA product with 6x loading dye (Fermentas). PCR products were visualized under UV light following staining with ethidium bromide. Sequencing of PCR products were conducted in Applied Biosystems 3730 DNA Analyzer at the Genomic Analysis and Technology Core (GATC) facility at the University of Arizona. To create contigs from forward and reverse amplified sequences chromatograms obtained from GATC were edited in MacVector v. 10.6 using the assembler package phred phrap (***Ewing and Green, 1998, Ewing et al. 1998***).

Table 3. *PDA* sequence sources and accession number

Source	**NCBI**	**NCBI**	**VanEtten Laboratory**
Taxon	***F.oxysporum* f.sp. *pisi***	***N. haematococca***	***N. boniensis***
PDA accession number	AYY87143	X73145.1	N/A

Sequence alignments

Nucleotide and amino acid sequences were aligned for each gene in the program MEGA v. 4.0 (***Tamura et al., 2004***) using the CLUSTALW algorithm. Each alignment included the following taxa: *F. oxysporum* f.sp. *pisi*, *F. oxysporum* f.sp. *lycopersici*, N. *haematococca*, and *N. boniensis.* Ambiguous regions of the sequence were edited manually.

Percent identity analysis

The percent identity between each pairwise taxon comparison (*F. oxysporum* f.sp. *pisi* vs. *N. haematococca*, *F. oxysporum* f.sp. *pisi* vs. *N. Boniensis*, and *N. haematococca* vs. *N. boniensis)* was found using NCBI's BLAST 2 sequence site. The percent identity for all comparisons were examined based on nucleotide and amino acid sequences. In all but one analysis a student's t-test was used to test for significance between the orthologs and *PDA*, because the distribution of data was normal. The other test for significance used in one instance was a Wilcoxon signed-rank test because the data was not normally distributed. Ultimately a p-value was found to see if *PDA* was significantly different from the orthologs.

GC content analysis

The GC content of *PDA* and the eleven orthologous sequences was acquired with an online calculator (http://mbcf.dfci.harvard.edu/docs/oligocalc.html). *PDA*'s GC content was compared to the orthologous genes' GC content for the three taxa *(i.e. F. oxysporum* f.sp. *pisi* versus *N. haematococca*, *F. oxysporum* f.sp. *pisi* versus *N. boniensis*, and *N. haematococca* versus *N. boniensis*). For the PEP cluster, which is approximately 25 kb, the GC content of the cluster was compared against all other 25kb-blocks in a sliding window analysis across the CD chromosome. Two Perl scripts were written to perform this analysis. Scripts were written and provided by Dr. Steve Rounsley. One script segmented the genome into 25 kb-blocks, while the other script recorded the GC content. Statistical analyses of the GC contents were executed using JMP v 8.0 statistical software (SAS). T-tests were used for the PEP cluster analysis (comparison of PEP cluster against the rest of the chromosome it resides on, in 25kb blocks) and for *PDA* analyses (*PDA* against the orthologs), with the exception of one comparison of *PDA* against the orthologs where a Wilcoxon signed rank test was used because of the non-normal distribution of the ortholog data.

Phylogenetic analysis

For Parsimony analysis, each multiple sequence alignment was imported into PAUP* v 4.0 (***Swofford, 2003***) and a heuristic search was executed with

1000 random sequence additions with the tree-bisection-reconstruction (TBR) algorithm. Strict consensus trees were generated from all equally most parsimonious trees. Topology support was evaluated by 1000 parsimony bootstrap replicates with the above setting for each.

For Bayesian analysis, Mr. Modeltest v 2.3 (***Nylander, 2004***) was used to infer a substitution model of nucleotide evolution for each alignment. The best model of evolution was incorporated into Mr.Bayes V 3.1.1, (***Huelsenbeck and Ronquist 2001***). For each dataset, analyses consisted of two independent runs, with four chains each, for 1 million generations, sampling every 100^{th} generation. The first 250,000 trees of the Bayesian analyses were considered 'burn-in', and were removed; the remainder were used to generate majority rule consensus trees in PAUP* v 4.0 (***Swofford, 2003***).

III. RESULTS

Percent Identity

A comparison of the identities for the nine orthologs, the two introns of beta-tubulin, and *PDA* was obtained by aligning these sequences and BLASTing the two sequences against each other. The analysis was based on *F. oxysporum* f.sp. *pisi* vs. *N. boniensis, F. oxysporum* f.sp. *pisi* vs. *Nectria haematococca* MPVI, and *N. boniensis* vs. *N. haematococca* MPVI. All comparisons were made between the *PDA* genes at the nucleotide level to the 11 orthologs, plus the beta-tubulin introns, and at the amino acid level for the 9 orthologs. Statistical analysis consisted of a t-test for all comparisons with the exception of the amino acid comparison between *N. boniensis* and *N. haematococca* (Figure 4, comparison C), where a Wilcoxon signed-rank test was used..

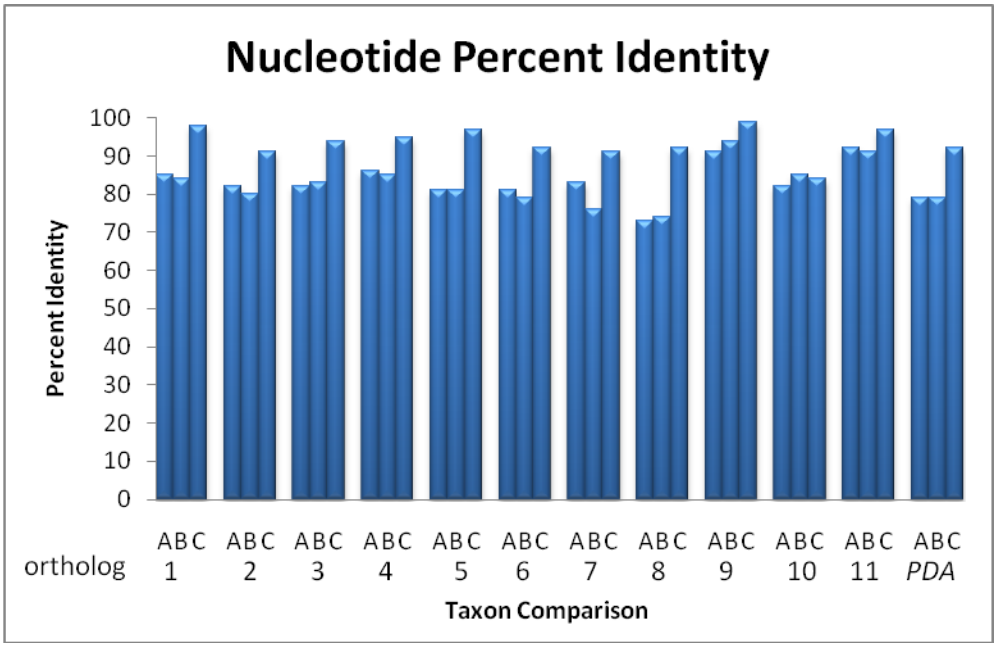

Figure 3. Nucleotide percent identity comparison among three taxa for 12 orthologous genes; Comparison A= *F. oxysporum* f.sp. *pisi* vs. *N. boniensis*, B= *F. oxysporum* f.sp. *pisi* vs. *N. haematococca*, C= *N. boniensis* vs. *N. haematococca*.

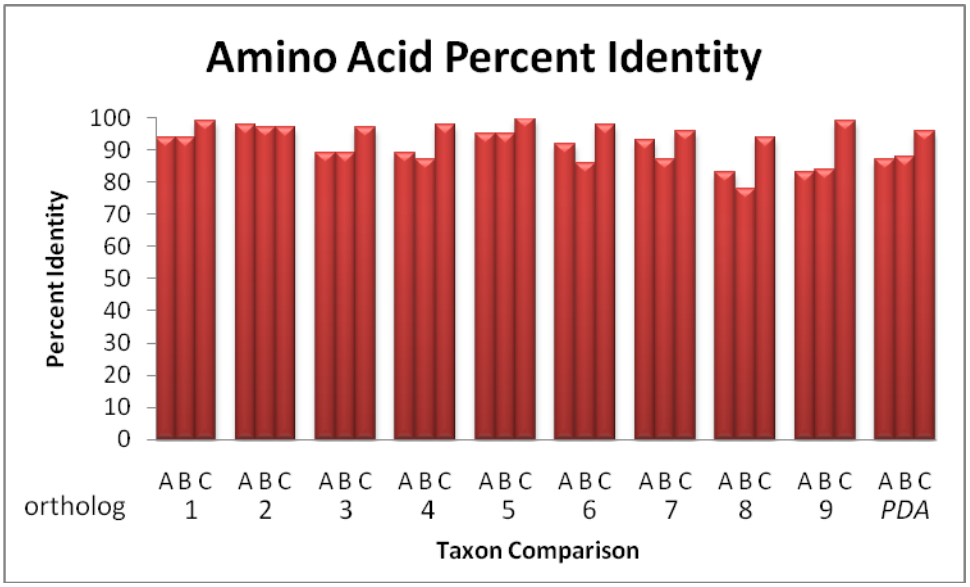

Figure 4. Amino acid percent identity comparison among three taxa for 12 orthologous genes; Comparison A= *F. oxysporum* f.sp. *pisi* vs. *N. boniensis*, B= *F. oxysporum* f.sp. *pisi* vs. *N. haematococca*, C= *N. boniensis* vs. *N. haematococca*

Illustrated in Figure 3, comparison A, At the nucleotide level, the observed percent identity of the *PDA* genes, including beta-tubulin introns, for *F. oxysporum* f.sp. *pisi* and *N. boniensis* (79%) was significantly lower than the hypothesized mean based on comparison of 11 orthologous sequences (mean = 83.5%, standard deviation = 5.2, 95% CI = 79.9-86.9; t = 2.8607, DF= 10, p = 0.0169). The observed percent identity of the *PDA* genes to orthologous sequences, with beta-tubulin introns excluded, for *F. oxysporum* f.sp. *pisi* and *N. boniensis* (79%) was not significantly different than the hypothesized mean based on comparison of the 9 orthologous genes (mean = 82.66%, standard deviation = 4.8218, 95% CI = 78.96-86.37; t = 2.2813, DF= 8, p = 0.0520).

For the amino acid comparison (Figure 4, comparison A), the observed percent identity of orthologous sequences (with Beta-tubulin introns excluded) for *F. oxysporum* f.sp. *pisi* and *N. boniensis* (87%) was not significantly different than the hypothesized mean based on comparison of 9 orthologous genes (mean=90.33%, standard deviation = 5.7899, 95% CI = 85.88-94.78; t = 1.7277, DF= 8, p = 0.1223).

Result: Percent identity shows that for *F. oxysporum* f.sp. *pisi* vs. *N. boniensis: PDA* was less similar among taxa than were orthologs at the nucleotide level, not at the amino acid level..

Illustrated in Figure 3, comparison B, at the nucleotide level, the observed percent identity of orthologous sequences, with beta-tubulin introns included, for *F. oxysporum* f.sp. *pisi* and *N. haematococca* (79%) was not significantly different than the hypothesized mean based on comparison of 11 orthologous sequences (mean=82.91%, standard deviation = 5.9406, 95% CI = 78.92-86.90; t = 2.1824, DF= 10, p = 0.0540). The observed percent identity of orthologous sequences, with beta-tubulin introns excluded, for *F. oxysporum* f.sp. *pisi* and *N. boniensis* (79%) was not significantly different than the hypothesized mean based on comparison of 9 orthologous genes (mean = 81.77%, standard deviation = 5.8261, 95% CI = 77.29-86.26; t = 1.4303, DF= 8, p = 0.1905).

At the amino acid level (Figure 4, comparison B), the observed percent identity of orthologous sequences (with beta-tubulin introns excluded) for *F. oxysporum* f.sp. *pisi* and *N. haematococca* MPVI (88%) was not significantly different than the hypothesized mean based on comparison of 9 orthologous genes (mean=88.55%, standard deviation = 5.98, 95% CI = 83.96-93.15; t = 0.2786 DF= 8, p = 0.7876).

Results: Percent identity shows that for *N. haematococca* MPVI vs. *F. oxysporum* f.sp *pisi: PDA* genes are not significantly different than the orthologs.

As illustrated in Figure 3, comparison C, the observed percent identity of orthologous sequences, with beta-tubulin introns included, for *N. boniensis* and *N. haematococca* (92%) at the nucleotide level was not significantly different than the hypothesized mean based on comparison of 11 orthologous sequences (mean=93.64%, standard deviation = 4.2958, 95% CI = 90.75-96.52; t =1.2634 , DF= 10, p = 0.2351). Consistent with these results, the observed percent identity of orthologous sequences, with beta-tubulin introns excluded, for *N. boniensis* and *N. haematococca* (92%) was not significantly different than the hypothesized mean based on comparison of 9 orthologous genes (mean=94,33%, standard deviation = 3.082, 95% CI = 91.96-96.70; t = 2.271, DF= 8, p = 0.0528).

At the amino acid level (Figure 4, comparison C), excluding the beta-tubulin introns, indicated that the observed percent identity of orthologous sequences for *N. boniensis* and *N. haematococca* (92%) was significantly lower than the hypothesized mean based on comparison of 9 orthologous genes (mean=97.55%, standard deviation = 1.8104, 95% CI = 96.16-98.95; t = 2.577, DF= 8, p = 0.0327*). A summary of the results for the comparison of the three species is shown in Table 4.

Result: Percent identity shows that for *N. boniensis* vs. *N. haematococca: PDA* was less similar among taxa than were orthologs at the amino acid level, not at the nucleotide level.

Table 4. Summary of percent identity analysis comparing *PDA* to orthologs and beta-tubulin introns. The percent identity of *PDA* differs from orthologous sequences; *PDA* was compared to 9 orthologs (and two beta-tubulin introns for the nucleotide analysis), sequence identities are listed in table 1.

Percent Identity Analysis Parameters			p-value
Fo vs. Nb	nucleotide level	with beta-tubulin intron	0.0169*
Fo vs. Nh			0.054
Nb vs. Nh			0.2351
Fo vs. Nb		without beta-tubulin intron	0.052
Fo vs. Nh			0.1905
Nb vs. Nh			0.0528
Fo vs. Nb	amino acid level	without beta-tubulin intron	0.1223
Fo vs. Nh			0.7876
Nb vs. Nh			0.0327*

*For both instances of significant difference between *PDA* and the orthologs, *PDA* was less identical than the orthologs.

GC Content comparisons

The comparison between the GC content of the nine orthologs, genes 1-8 and *EF1*-alpha, indicated that *PDA* GC content (48.3%) in *F. oxysporum f.sp pisi* was significantly lower than the GC content of the orthologous genes in this species *(*mean = 52.58%, standard deviation = 2.3745, 95% CI = 50.76-54.41; t = 5.4185, DF = 8, p= 0.0006*). Additionally, the same comparison resulted in finding that the *PDA* GC content (51.8%) in *N. boniensis* was significantly lower than the GC content of the orthologous genes *(*mean=56.17%, standard deviation = 2.3731, 95% CI = 54.35-58.00; t = 5.5341, DF = 8, p = 0.0006*). Finally, the same comparison in *N. haematococca* MPVI indicated that the *PDA* GC content (51.9%) was significantly lower than the GC content of the orthologous genes in this fungus (mean = 54.3, standard deviation = 1.8384, 95% CI = 52.88-55.71, DF= 8, p = 0.0044).

Table 5. Summary of GC content results comparing *PDA* to orthologs

Taxon	p-value
Fusarium oxysporum f.sp. *pisi*	0.0078*
Neocosmospora boniensis	0.0001*
Nectria haematococca	0.0039*

* For all three instances of significant difference between *PDA* and orthologs, *PDA* had a lower GC content than the orthologs.

GC Content analysis in the PEP cluster

A comparison of the GC content of the 25 kb PEP cluster (50.5%) to all other 25 kb-blocks on the 1.6 Mb chromosome where the *PEP* cluster resides revealed that the *PEP* cluster has a significantly higher GC content (mean = .492, standard deviation = 0.039, 95% CI =.4898-.4944; t = -10.890, p = 0.0001). There were 1,138 25 kb-blocks that had an average GC content of 49.2%, ranging from 33.9% GC to 54.3%. This difference is probably not biologically significant, as the mean for the data set was 0.492 and the value 0.505 was used to test the mean, which is a small but numerically significant difference. A t-test was performed to test the 25kb PEP cluster region against the other 1,138 25kb regions, to find the significance of the PEP cluster's GC content compared to the orthologs.

Phylogenetic analysis

As a third strategy to examine the inheritance of *PDA*, a phylogenetic approach was taken to demonstrate whether or not the *PDA* genealogy is concordant with the genealogies of the orthologous genes. Phylogenies for each of the genes were constructed using both Parsimony and Bayesian approaches. The topologies of all gene genealogies (Figure 9-19) generated from the orthologous sequences were in agreement and had the same topology as the of *PDA* tree (Figure 20), supporting vertical inheritance.

Genes 1-11

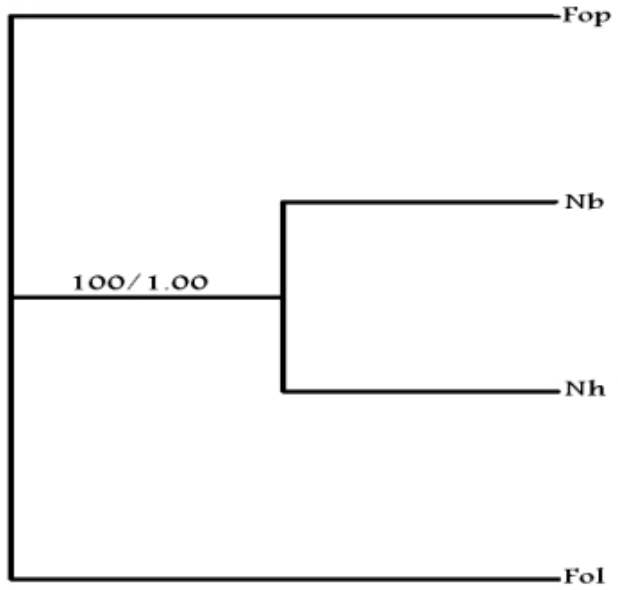

Figure 5. Genealogy of Genes 1-11, (Table 1), inferred from Parsimony and Bayesian analysis. All genealogies had identical topologies and support. *Fusarium oxysporum* f.sp. *pisi* is abbreviated as Fop, *Neocosmospora boniensis* is abbreviated as Nb, *Nectria haematococca* MPVI is abbreviated as Nh , and Fol represents *Fusarium oxysporum* f.sp. *lycopersici.* Value at internal node represents parsimony bootstrap (left of slash) and Bayesian posterior probability (right of slash).

PDA

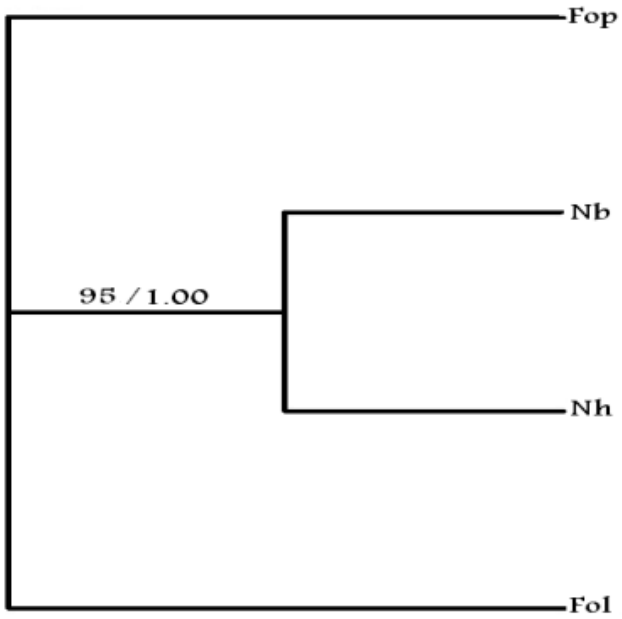

Figure 6. Genealogy of *PDA* inferred from Parsimony and Bayesian analysis. Taxon abbreviations same as Figure 5.

Phylogenetic statistics

Genealogy of Gene 1, serine/threonine phosphatase (Table 1), inferred from Parsimony and Bayesian analysis. Alignment was based on 459 nucleotides; 52 characters were parsimony informative, 40 were parsimony uninformative, and 367 were constant. Tree length is 103 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of Gene 2, Conserved hypothetical protein similar to glutathione-independent formaldehyde dehydrogenase (Table 1), inferred from Parsimony and Bayesian analysis. Alignment was based on 513 nucleotides; 72 characters were parsimony informative, 42 were parsimony uninformative, and 399 were constant. Tree length is 121 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of Gene 3, peroxisomal co-A synthetase (Table1), inferred from Parsimony and Bayesian analysis. Alignment was based on 474 nucleotides; 66 characters were parsimony informative, 27 were parsimony uninformative, and 381 were constant. Tree length is 97 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of Gene 4, Conserved hypothetical protein: glycolate oxidase (Table1), inferred from Parsimony and Bayesian analysis. Alignment was based on 396 nucleotides; 46 characters were parsimony informative, 16 were parsimony uninformative, and 334 were constant. Tree length is 65 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of Gene 5, conserved hypothetical protein: inorganic phosphate transporter (Table1), inferred from Parsimony and Bayesian analysis. Alignment was based on 437 nucleotides; 65 characters were parsimony informative, 12 were parsimony uninformative, and 360 were constant. Tree length is 81 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00

Genealogy of Gene 6, polyphosphoinositide phosphatase (Table1), inferred from parsimony and Bayesian analysis. Alignment was based on 621 nucleotides; 87 characters were parsimony informative, 45 were parsimony uninformative, and 489 were constant. Tree length is 144 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of Gene 7, conserved hypothetical protein: peroxidase oxygenase, (Table 1), inferred from Parsimony and Bayesian analysis. Alignment based on 708 nucleotides; 91 characters were parsimony informative, 57 were parsimony uninformative, and 560 were constant. Tree length is 159 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of Gene 8, Conserved hypothetical protein: sec5 subunit of exocyst complex (Table1), inferred from parsimony and Bayesian analysis. Alignment was based on 891 nucleotides; 188 characters were Parsimony informative, 70 were parsimony uninformative, and 633 were constant. Tree length is 272 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of translation elongation factor 1-alpha (EF1-α), (Table1), inferred from parsimony and Bayesian analysis. Alignment was based on 183 nucleotides; 8 characters were parsimony informative, 7 were parsimony uninformative, and 168 were constant. Tree length is 15 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00

Genealogy of beta-tubulin intron 1 (Table1), inferred from Parsimony and Bayesian analysis. Alignment was based on 252 nucleotides; 17 characters were parsimony informative, 7 were parsimony uninformative, and 228 were constant. Tree length is 25 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00.

Genealogy of beta-tubulin intron 2, (Table1), inferred from parsimony and Bayesian analysis. Alignment was based on 682 nucleotides; 58 characters were parsimony informative, 5 were parsimony uninformative, and 619 were constant. Tree length is 64 steps, consistency index (CI) = 1.00, retention index (RI) = 1.00

Genealogy of *PDA*, inferred from Parsimony and Bayesian analysis. Alignment was based on 979 nucleotides; 56 characters were parsimony informative, 463 were parsimony uninformative, and 460 were constant. Tree length is 617 steps, consistency index (CI) = 0.956, retention index (RI) = 0.518.

IV. DISCUSSION

The discontinuous distribution of *PDA* and the other *PEP* genes in *N. haematococca* MPVI, *N. boniensis* and in *F. oxysporum* f.sp. *pisi*, along with the observed GC content and codon bias of *PDA* and other *PEP* genes, and the fact that the PEP cluster resides on a dispensable portion of the genome (***Temporini and VanEtten, 2004***), raised the question whether these genes had been acquired via HGT. This study combined analyses of percent identity, GC content and phylogenetics to further test the evolutionary inheritance of *PDA* and the PEP cluster.

The percent identity analysis revealed that there were only two cases where *PDA* was significantly different from the orthologs. In both cases, *PDA* had a lower percent identity than the orthologs, which is the opposite of what is predicted if *PDA* had been recently acquired by HGT. One of the well-supported events of HGT reported in fungi is the acquisition of *ToxA* in *Pyrenophora tritici-repentis* from *Stagonospora nodorum.* In this case, the high percent identity (98-100% at nucleotide level) of the genes between *S. nodorum* and *P. tritici-repentis* provided evidence consistent with a recent HGT event (***Friesen, 2006***). With *PDA* being less identical than the orthologs, the results of my percent identity analysis support vertical inheritance. However, the results discussed in this thesis may be limited by how the orthologous sequences were selected for comparison. The reciprocal BLAST between *F. oxysporum f.sp lycopersici* and *N. haematococca's* identified a list of 8,000 putative orthologs, which were arranged so the genes with the highest identities were at the top of the list. The eight orthologs selected from this list started with the most conserved and moving down the list approximately

every 400th gene, to get a total of eight orthologs. Thus the eight orthologs selected were from the subset of the 2,800 most similar orthologs and could have provided a bias to the study as they are all highly conserved among the analyzed taxa. Since these taxa are so closely related, and *PDA* is a non-essential gene, therefore potentially subjected to less selective pressure and more diversification, selecting orthologs with such level of conservation could account for *PDA* showing a lower level of conservation. To prevent this bias in the future, a future study should include a comparison of orthologs that are more representative of the genome and represent a wider range of levels of conservation.

Because GC contents within individual prokaryotic genes have been shown to be very uniform, GC content is often an excellent identifier of possible horizontal gene transfer (***Bohlin and Skjerve, 2009***). The GC content analysis showed that *PDA* is statistically different from the orthologs and, in each case, *PDA* had a statistically lower GC content (Table 5). The biological significance of a lower GC content is that the gene or sequence is aberrant when compared to other vertically inherited genes in the taxon, which is a signal that inheritance of *PDA* might not have occurred vertically. The comparison of the GC content of the entire PEP cluster to the orthologs indicated that the PEP cluster has a statistically higher GC content than the orthologs. The mean of the PEP cluster is 50.5% compared to 49.2% for the other 25 kb regions compared. Although a difference of 1.3% sounds like a small difference, it was found to be statistically significant, and could be a remaining signal of an HGT event.

The phylogenetic analysis showed the expected phylogenetic relationship of the three taxa, with *N. boniensis* and *N. haematococca* being closely related and *F.*

oxysporum f.sp. *pisi* more distantly related (Figure 2). The primer design using *N. haematococca* as the sole sequence was extremely effective. This is because *N. boniensis* and *N. haematococca* are very closely related as shown by this study, and that done by O'Donnell (2000), stating that the genus *Neocosmospora* is deeply nested inside the *F. solani* species complex. Each of the 11 sequence genealogies showed the same topology, all with 100% Parsimony bootstraps and 1.00 Bayesian posterior probabilities. *PDA* showed 95% bootstraps with a 1.00 posterior probability, with the same topology as the ortholog sequence genealogies. This analysis supports vertical inheritance of *PDA*, as this study found the expected species phylogenies in the gene genealogies, and they were all congruent with the *PDA* genealogy. The manner in which the orthologs were selected should not influence the results of this study because each ortholog should be, by definition, vertically inherited, so each genealogy will have the same topology.

In *M. grisea* the avirulence gene *ACE1* that encodes for a hybrid PKS-NRPS (***Bohnert et al. 2004***) belongs to a cluster of 15 genes involved in secondary metabolism (***Collemare et al. 2008***). Clusters similar to the *ACE1* cluster in *M. grisea* appear to be present in other ascomycetes (ie. *S. nodurum, Aspergillus clavatus, Chaetomium globosum*) (***Khaldi et al. 2008***). A phylogenetic analysis was used in order to examine the evolution history of the *ACE1*-like clusters in several species. In that study, the incongruent phylogenetic trees between orthologous genes and *ACE1* was indicative of the horizontal inheritance of *ACE1* from a relative of *M. grisea* into an ancestor of *A. clavatus* (***Khaldi et al, 2008***). For our study, we predicted that if the *PDA* gene genealogy produces phylogenies of the three taxa that match the known species

relationships. This would argue against a horizontal acquisition of *PDA*, and support vertical inheritance.

Conclusion

In summary, the percent identity analysis of *PDA* in comparison to the chosen orthologs supports vertical inheritance, which is corroborated by the phylogenetic data. The GC content analysis of *PDA* does not indicate the method of inheritance of *PDA* or the PEP cluster, however it might be a signal of a more ancient horizontal transfer event.

Genes associated with genomic islands, or pathogenicity islands, in fungi are often located in subtelomeric regions of chromosomes, which could account for genes in this region to having aberrant GC contents and codon biases (***Farman, 2007***). This is supported by some of the hypothesis on the origin of the CD chromosome. These supernumerary chromosomes are often compared to "B" chromosomes and some propose B chromosomes originate from the main chromosomes ('A' chromosomes), via interspecific hybridizing (***Jones and Houben, 2003***). If a subtelomeric portion of a main chromosome and its centromere was the only portion of a main chromosome that had gone through such a hybridization to yield a CD chromosome, that would explain why *PDA* and other PEP cluster genes would have a skewed GC content and resulting codon bias. Subtelomeric regions are naturally GC poor and would lead to biases that could falsely appear to be signals of HGT.

I propose that the *PEP* genes were likely present in a common ancestor of the Sordariomycetes, either as a cluster, or as separate genes that were acquired by HGT or

brought together to form the cluster following numerous sexual recombinations, which brought them together by chance. After the cluster was formed, selective pressures could have been responsible for maintaining that cluster, in this case those selective pressures could have been provided by the presence of pea plants as hosts. A clustering of genes may eliminate crossing over during meiosis because the genes in the cluster are tightly packed together; thus eliminating the possibility of disrupting the genes in the cluster, although this hypothesis has not yet been assessed for significance (***Keller and Hohn, 1997***). Other theories on the maintenance of a cluster are for co-regulation by chromatin remodeling, or an epistatic selection maintaining a tight linkage of the genes (***Khaldi, 2008***). This could result in maintaining their collective function and ensuring co-regulation of gene expression (***Keller and Hohn, 1997***). It has long been believed that eukaryotic genes are not co-regulated as prokaryotic genes are in an operon. However, more and more evidence has come to light demonstrating that genes in eukaryotes that are proximal to each other are in fact regulated similarly (***Michalak, 2008***). For example, an experiment was done where two transgenes were inserted adjacently in tobacco and they had similar expression patterns, even though they each had very different promoters (***Mickalak, 2008***). Also, a genome-wide study including organisms such as *Drosophila*, nematodes, mice, humans, and *Arabidopsis*, has shown that genes with similar expression patterns are typically in clusters (***Michalak, 2008***). The leading theory behind this co-expression in eukaryotes is that genomes are compartmentalized into chromatin domains, such that if a gene is in close proximity to other genes, and the chromatin becomes euchromatic, then the whole chromatin region would be

transcriptionally active due to the change in chromatin conformation (***Michalak, 2008***). So, in eukaryotes it seems that gene proximity leads to co-regulation based on chromatin structure and how genes in a cluster affect each others' expression is by being on the same region of chromatin. Natural selection acts to keep these clusters together, to eliminate the above-mentioned problems associated with breaking up a cluster, or losing a portion of one, so that their collective function and regulation is maintained. Another hypothesis for the maintenance of a cluster is the 'selfish operon' theory, which is linked to the theory that a clustering enables the genes to be more easily horizontally transferred (***Lawrence, 1999***).

Rather than being a result of HGT, the discontinuous distribution of the *PDA* and PEP cluster could be explained by the ancestors of these extant taxa either being near or far from pea plants. The pea plant environment is the only known selective pressure that maintains the cluster's necessity. Those lineages in close proximity to pea plants would be expected to maintain the *PEP* genes, while those fungal lineages that were not associated with pea plants could have lost the *PEP* genes because there would be no selective pressures to maintain them.

As previously stated, the likeliest explanation for the origin of *PDA* and PEP genes is that the cluster was vertically inherited, and some lineages maintained the cluster because they co-habitated with pea and pisatin and others lost the cluster in the absence of natural selection. The fact that *PDA* and the PEP cluster reside on a dispensable portion of the genome, the CD chromosome, and has a biased GC content, along with a discontinuous distribution of the genes, does suggest a HGT origin because this data is

consistent with known means to identify an HGT event (***Zaneveld et al., 2008***). Currently, the most frequent methods to identify an HGT event are through phylogenetic and compositional approaches (***Zaneveld et al., 2008***). Thus, I conclude that *PDA* has been inherited vertically for some time, as shown by the phylogenetic analysis. The GC content bias may be a remaining signal that the PEP cluster or even the entire CD chromosome went through HGT in the progenitor to these Fusaria, followed by subsequent vertical inheritance of these genes. It is also possible that another taxon passed the CD chromosome to these Sordariomycetes, as these three species were diverging, or soon after. Even though, it is likely that an HGT event occurred, the data provided in this study points towards vertical inheritance for the PEP cluster/CD chromosome possibly preceded by an ancient HGT event.

REFERENCES

Agrios, G.N. 2005. Plant Pathology 5th edition. Burlington, MA. Elsevier Academic Press. 922 pgs.

Arnold A.E., and Lutzoni F. 2007. Diversity and host range of foliar fungal endophytes: Are tropical leaves biodiversity hotspots? Ecology 88:541–549.

Akmatsu, H., Masatoki, T., Kodama, M., Johnson, R., Otani, H., and Kohmoto, K. 1999. Molecular karyotypes for *Alternaria* plant pathogens known to produce host-specific toxins. Current Genetics 35: 647–656.

Bailey, J.A. and Mansfield, J.W. 1982. Phytoalexins. London, U.K. Blackie and Son Ltd. 334 pgs.

Bohlin, J. and Skjerve, E. 2009. Examination of genome homogeneity in prokaryotes using genomic signatures. PLoS ONE 4(12): e8113. doi:10.1371/journal.pone.0008113.

Bohnert H.U., Fudal I., Dioh W., Tharreau D., Notteghem J.L., and Lebrun M.H. 2004. A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. Plant Cell 16:2499-2513.

Coleman *et al.* 2009. The Genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. PLOS Genetics 5:1-14.

Collemare, J., Pianfetti, M., Houlle, A.E., Morin, D. Camborde, L., Gagey, M.J., Barbisan, C., Fudal, I., Lebrun, M.H, and Böhnert H.U. 2008. *Magnaporthe grisea* avirulence gene *ACE1* belongs to an infection-specific gene cluster involved in secondary metabolism. New Phytologist 179:196-208.

Covert, S. F., J. Enkerli, V. P. W.Miao, and VanEtten H.D. 1996 A gene for maackiain detoxification from a dispensable chromosome of *Nectria haematococca*. Mol. and Gen. Genetics 251:397-406.

Covert, S.F. 1998. Supernumerary chromosomes in filamentous fungi. Curr. Genet. 33:311-319

Cuomo, C.A., Guldener, U., Xu, J., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L., Baker, S.E., Rep, M., et al. 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317: 1400–1402.

Cruickshank, I. A.M., and Perrin, D. R.1960. Isolation of a phytoalexin from *Pisum Sativum* L. Nature 4739:799-800.

Cruickshank, I. A.M., and Perrin, D.R. 1961. Studies on phytoalexins III. The isolation assay, and general properties of a phytoalexins from *Pisum Sativum* L. Austral. J. Biol. Sci. 336-348.

Delserone, L.M., McCluskey, K.D., Matthews, E., and VanEtten, H.D. 1999. Pisatin demethylation by fungal pathogens and nonpathogens of pea: association with pisatin tolerance and virulence. Physiological and Molecular Plant Pathology 55: 317-326.

Farman, M.L., 2007. Telomeres in the rice blast fungus *Magnaporthe oryzae*: the world of the end as we know it. Fems Microbiology Letters 273: 125-132.

Feuillet, C., and Keller, B. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. Proc. Natl. Acad. Sci. 96: 8265–8270.

Friesen, T.L., Stukenbrock, E.H., Liu, Z., Meinhardt, S., Ling, H., Faris, J.D., Rasmussen, J.B., Solomon, P.S., McDonald, B.A., and Oliver, R. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. Nature Genetics 38: 953–956.

George H.L., Hirschi K.D., and VanEtten H.D. 1998. Biochemical properties of the products of cytochrome P450 genes (PDA) encoding pisatin demethylase activity in *Nectria haematococca*. Archives of Microbiology 17:147-154.

George H.L. and VanEtten H.D. 2001. Characterization of pisatin-inducible cytochrome P450s in fungal pathogens of pea that detoxify the pea phytoalexin pisatin. Fungal Genetics and Biology 33:37-48.

Hacker, J. and Kaper, J.B. 2000. Pathogenicity islands and the evolution of microbes. Ann. Rev. of Microbiology. 54: 641-679.

Han, Yinong, 2001. Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectria haematococca*. The Plant Journal 25:305-314.

Hatta, R., Ito, K., Hosaki, Y., Tanaka, T. Tanaka, A., Yamamoto, M., Akimitsu, K., and Tsuge, T. 2002. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. Genetics 161: 59–70.

Jones, N., and Houben, A. 2003. B chromosomes in plants: escapees from the A chromosome genome? Trends in Plant Science 8:417-423.

Keller, N.P., Hohn, T.M. 1997. Metabolic pathway gene clusters in filamentous fungi. Fungal Genetics and Biology 21;1;17-29.

Khaldi, N., Collemare, J., Lebrun, M.C., and Wolfe, K.H. 2008. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. Genome Biology. 9: R18.1–R18.10.

Lawrence, J.G. 2005. Common themes in the genome strategies of pathogens. Current Opinion Genes and Development 15:584-588

Liu, X., Inlow, M., and VanEtten, H.D. 2003. Expression profiles of pea pathogenicity (PEP) genes in vivo and in vitro, characterization of the flanking regions of the PEP cluster and evidence that the PEP cluster region resulted from horizontal gene transfer in the fungal pathogen *Nectria haematococca*. Current Genetics 44: 95–103.

Maloney, A.P. and VanEtten, H.D. 1994. A gene from the fungal plant pathogen *Nectria haematococca* that encodes the phytoalexin-detoxyfying enzyme pisatin demethylase defines a new cytochrome P450 family. Molecular and General Genetics 243:506-514.

Matthews, D.E. and VanEtten, H.D. 1983. Detoxification of the phytoalexin pisatin by a fungal cytochrome P-450. Archives Biochemistry and Biophysics 224:494-505.

Miao, V.P.W. and VanEtten, H.D. 1992. Genetic analysis of the role of phytoalexin detoxification in virulence of the fungus *Nectria haematococca* on chickpea (*Cicer arietinum*). Applied and Environmental Microbiology 58: 809-814

Michalak, P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics 91:243-248

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. Trends in Genetics 17:589-596.

Ning, W. and Runsheng, C. 1999. Comparison between phylogeny of introns and exons in primates. Chinese Science Bulletin 44:21 1940-1946.

O'Donnell, K. 2000. Molecular phylogeny of the *Nectria haematococca-Fusarium solani* species complex. Mycologia 92: 919–938.

Ochman, H., Lawrence, J. G. & Groisman, E. A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature. 405:299–304

Robeson, D. J. and Harborne, J. B. 1980. A chemical dichotomy in phytoalexin induction within the tribe Vicieae of the Leguminosae. Phytochemistry 19: 2359-2365.

Rodriguez-Carres, M., White, G., Tsuchiya, D., Taga, M., and VanEtten, H.D. 2008. The supernumerary chromosome of *Nectria haematococca* that carries pea-pathogenicity-related genes also carries a trait for pea rhizosphere competitiveness. Applied and Environmental Microbiology 74: 3849–3856.

Roger, A.J., Sandblom, O., Doolittle, W.F., and Philippe, H. 1999. An evaluation of Elongation Factor 1-alpha as a phylogenetic marker for eukaryotes. Mol. Biol. Evol. 16:218–233.

Shiraishi, T., Oku, H., Isono, M., and Ouchi S. 1975. The injurious effect of pisatin on the plasma membrane of pea. Plant and Cell Physiology 16:939-942

Smith, D.A., 1976. Some effects of the phytoalexin, kievitone, on the vegetative growth of *Aphanomyces euteiches*, *Rhizoctonia solani* and *Fusarium solani* f.sp. *phaseoli*. Physiol. Plant Pathol. 9: 45-55.

Smith, D.A., 1978. Observations on the fungitoxicity of the phytoalexin, kievitone. Phytopathology 58: 81-87.

Tamura, K., Dudley J., Nei M., and Kumar, S. 2007. *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596–1599.

Temporini, E.D. and VanEtten, H.D., 2002. Distribution of the pea pathogenicity (PEP) genes in the fungus *Nectria haematococca* mating population VI. Curr Genet. 41: 107-114.

Temporini, E.D. and VanEtten, H.D. 2004. An analysis of the phylogenetic distribution of the pea pathogenicity (*PEP*) genes of *Nectria haematococca* MPVI supports the hypothesis of their origin by horizontal transfer and uncovers a potentially new pathogen of garden pea: *Neocosmospora boniensis*. Current Genetics 46: 29-36.

Uehara, K. 1964. Relationship between host specificity of pathogen and phytoalexins. Annals of the Phytopathological Society of Japan 29:103-110.

VanEtten, H.D. and Bateman, D.F. 1971. Studies on the mode of action of the phytoalexin phaseollin. Phytopathology 61:1363-1372

VanEtten, H.D. and Stein J.I. 1978. Differential response of *Fusarium solani* isolates to pisatin and phaseolin. Phytopathology 68:1276-1283.

VanEtten, H.D., Funnell-Baerg D., Wasmann C., McCluskey K. 1994. Location of pathogenicity genes on dispensible chromosomes in *Nectria haematococca* MPVI. Ant Van Leeuwenhoek 65: 263-267.

VanEtten, H.D., Temporini, E.D., and Wasmann, C. 2001 Phytoalexin (and phytoanticipin) tolerance as a virulence trait: why is it not required by all pathogens? Physiological and Molecular Plant Pathology 59: 83-93.

Van Der Does, H.C. and Rep, M. 2007. Virulence genes and the evolution of host specificity in plant-pathogenic fungi. Molecular Plant-Microbe Interactions 20: 1175–1182.

Walton, J.D. 2000. Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: a hypothesis. Fungal Genetics and Biology 30: 167–171.

Ward, Todd J. 2002. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic Fusarium. PNAS 99: 9278-9283.

Waterhouse, J.C., Russell, R.R.B. 2006. Dispensable genes and foreign DNA in *Streptococcus mutans*. Microbiology 152: 1777-1789.

Zaneveld, J.R., Nemergut, D.R., and Knight, R. 2008. Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. Microbiology
154:1-15.