

EVOLUTION OF NONRIBOSOMAL PEPTIDE SYNTHETASE PROTEINS
INVOLVED IN SECONDARY METABOLISM IN FUNGI

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Kathryn E. Bushley

August 2009

© 2009 Kathryn E. Bushley

EVOLUTION OF NONRIBOSOMAL PEPTIDE SYNTHETASE PROTEINS
INVOLVED IN SECONDARY METABOLISM IN FUNGI

Kathryn E. Bushley, Ph. D.

Cornell University 2009

Nonribosomal peptide synthetases (NRPSs) are multimodular enzymes which biosynthesize peptides (NRPs) independently of ribosomes. Three core domains (adenylation (A), thiolation (T), condensation (C)) comprise a functional module for NRP biosynthesis. Although NRPSs produce a diversity of bioactive compounds, little is known about the evolutionary relationships of genes encoding NRPSs and the mechanisms by which they evolve. The objectives of this research were to perform phylogenomic analyses to identify major NRPS subclasses and determine evolutionary relationships and to elucidate fine-scale evolutionary mechanisms giving rise to the diverse NRPS domain structures in fungi.

Chapter 2 is a published manuscript on ferrichrome synthetases tracking the evolution of domain architectures of these relatively conserved enzymes across fungi. Results supported the hypothesis that ferrichrome synthetases evolved by tandem duplication of complete modules (A-T-C) (single or double units) and loss of single A domains or complete A-T-C modules. A mechanism for evolution of iterative biosynthesis is proposed. Protein modeling of the A domain substrate binding pockets refined characterization of key residues involved in substrate specificity, by identifying novel sites.

Chapter 3 reports a fungal kingdom-wide phylogenomic study of NRPSs, with the objective of identifying subclasses. Nine were identified which fell into two major

groups. One consisted of primarily mono/bi-modular NRPSs with conserved domain architectures which group with bacterial NRPSs and whose products are associated with conserved metabolic roles. The other consisted of primarily multimodular and exclusively fungal NRPSs with variable domain architectures whose products perform niche-specific functions. All groups of NRPSs were much more common in Eucaryotes than in any other fungal taxonomic group. Although NRPSs are discontinuously distributed across fungal taxa, little evidence was found for horizontal gene transfer from bacteria to fungi.

Overall, this study showed that both tandem duplication and loss, as well as recombination and rearrangement, of modular units (either complete A-T-C modules or single A domains) are mechanisms by which NRPSs and their chemical products evolve. Phylogenomic analysis identified subgroups of NRPSs possibly reflecting common function and suggested an older evolutionary origin of several mono/bimodular groups while multimodular fungal NRPSs are more recently derived and highly expanded in Eucaryotes.

BIOGRAPHICAL SKETCH

Kathryn Bushley was born in Seattle, Washington on August 25, 1968. She attended Oberlin College and obtained a B.A. degree with a double major in Biology and Anthropology in 1991. After working and dancing for several years in the Seattle area and deciding not to pursue a professional career in dance, she returned to graduate school for a Master's in Environmental Management at Duke University. It was during her master's degree at Duke where she first fell in love with fungi while working on a masters thesis investigating turnover of mycorrhizal fungal root systems and taking a mycology course with Dr. Rytas Vilgalys. After completing her master's degree in 1997, she worked for her advisor, Dr. Janet MacFall, for several years at the Duke Medical Center investigating how mycorrhizal fungi bind aluminum in soil using innovative techniques of magnetic resonance imaging to noninvasively measure changes in root volume. She entered the PhD program in Plant Pathology at Cornell University in fall of 2001 to study mycology and joined the Turgeon lab in 2002.

I would like to dedicate this work to my great grandmother Lillian Vogle
whose adventurous spirit inspires me to explore the boundaries of the unknown and go
where no woman has gone before

ACKNOWLEDGMENTS

Many people have contributed in both large and small ways to the fruition of this work. I am grateful first and foremost for being blessed with exceptional parents whose love and support sustain me in all my endeavors. I would like to thank my committee members, B. Gillian Turgeon, Jeff Doyle, and Donna Gibson for their support, enthusiasm, and patience throughout the PhD process. I am also grateful to Oberlin College, my undergraduate institution, for teaching me to think critically and independently. Thanks to Miguel Carvahlo, Jean Bonasera, and fellow labmembers Shunwen Lu and Patrik Inderbitzin, for providing training and mentoring in molecular biology and phylogenetic techniques during the early stages of my PhD. Special thanks to Genevieve DeClerck and Paul Stodghill for their patience in initiating me into the art of computer programming and for troubleshooting more than one incomprehensible error message. I am also especially grateful to Daniel Ripoll, who performed protein structural modelling of NRPS AMP domains and contributed significantly to insights into substrate specificity of fungal ferrichrome siderophore synthetases discussed in Chapter 2 as well as other staff members of the Cornell Computation Biology Service Unit, Conrad Schoch, Adam Siepel, and Dave Schneider for providing both advice and resources for computational and phylogenetic analyses. I would also like to thank others who have provided advice along the way (Scott Kroken, Henk DeBakker, Kevin Nixon, and Ning Zhang) and the Plant Pathology administrative staff, especially Carol Fisher, for superb administrative support and assistance in formatting the thesis.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
List of Tables	xi
List of Figures	xiii
List of Appendices	xiv
1. Chapter 1: General Introduction	1
1.1 Modular Proteins in Secondary Metabolism	1
1.2 NRPS Biosynthesis	4
1.2.1 Adenylation (A) Domain	6
1.2.2 Thiolation (T) Domain	8
1.2.3 Condensation (C) Domain	9
1.2.4 Termination Domains	10
1.2.5 Decorating Domains	12
1.3 Evolutionary Origins of NRPS and PKS Synthetases	13
1.3.1 Relationship Between NRPSs, PKSs and Primary Metabolism	13
1.3.2 Discovery of NRPSs and Related AMP Adenylating Enzymes	14
1.4 Mechanisms of Evolution of Modular Proteins	16
1.4.1 Models of Gene Family Evolution	16
1.4.2 Evolution of Repeated Units in Proteins	17
1.5 Fungal NRPS: Evolution and Functional Groups	20

1.5.1	Mechanisms Leading to the Discontinuous Distribution of Secondary Metabolite Genes in Fungi: Gene Clusters, Horizontal Gene Transfer, and Duplication and Differential Loss (DDL)	20
1.5.2	Known Functional Classes of Fungal NRPSs	25
1.5.2.1	Conserved Homologs of ChNPS6, ChNPS4, ChNPS10, and ChNPS12	26
1.5.2.2	Siderophore Synthetases	29
1.5.2.3	ACV Synthetases	29
1.5.2.4	Cyclosporin Synthetases	30
1.5.2.5	Cyclic Depsipeptide Synthetases: Enniatin and Related Compounds	32
1.5.2.6	Ergot Alkaloid Synthetases	33
1.5.2.7	Peramine Synthetase	37
1.5.2.8	Peptaibols	37
1.5.2.9	Diketopiperazines and ETP toxins	38
1.5.2.10	Dothideomycete Host-Selective Toxins	41
1.5.2.11	Fungal PKS:NRPS Hybrids	43
1.5.2.12	NRPS:PKS Hybrids	46
1.6	Objectives	46
	References – Chapter 1	47
2.	Chapter 2: Module Evolution and Substrate Specificity of Fungal Nonribosomal Peptide Synthetases Involved in Siderophore Biosynthesis. Published manuscript.	87
2.1	Abstract	87

2.2	Background	88
2.3	Materials and Methods	92
2.3.1	Genomes Surveyed for Ferrichrome-Associated Nonribosomal Peptide Synthetases	92
2.3.2	Annotation of Candidate Ferrichrome Synthetases	94
2.3.3	Phylogenetic Analyses	95
2.3.3.1	Complete set of A domains	95
2.3.3.2	Individual Lineage Analysis	96
2.3.4	Substrate Specificity	97
2.3.4.1	Structural Modeling	97
2.3.4.2	Evolutionary Approaches to Identify Specificity Residues	99
2.4	Results	100
2.4.1	Distribution of Ferrichrome Synthetases in Fungi	100
2.4.2	Domain Architecture of Ferrichrome Synthetases	101
2.4.3	Two Distinct Lineages of Ferrichrome Synthetases	103
2.4.4	Additional Duplications Within the NPS1/SidC Lineage	109
2.4.5	<i>S. pombe</i> sib1	110
2.4.6	Putative Ferrichrome Synthetases in the SidE Clade	110
2.4.7	Individual Lineage Analysis	111
2.4.8	Adenylation Domain Substrate Choice	115
2.4.8.1	Structural Modeling	115
2.4.8.2	Evolutionary Approaches to Identification of Specificity Residues	122
2.5	Discussion	124
2.5.1	Distinct Lineages of Ferrichrome Synthetases	124

2.5.2	Evolution of Domain Architecture	125
2.5.3	Domain Architecture and Mechanism of Biosynthesis	129
2.5.4	Substrate Specificity	131
2.6	Conclusions	133
	Appendices – Chapter 2	135
	References – Chapter 2	151
3.	Chapter 3: Identification, Distribution, and Lineage Specific Expansions of NRPS and NRPS-like Synthetase Subfamilies in Fungi Manuscript submitted	158
3.1	Abstract	158
3.2	Background	159
3.3	Results and Discussion	163
3.3.1	Identification and Domain Structure of Candidate NRPSs	163
3.3.2	Phylogenomic Analysis and Subfamily Identification	164
3.3.3	Relationships Between Fungal and Bacterial NRPSs: Horizontal Transfer or Vertical Transmission and Massive Loss?	172
3.3.4	Distribution of NRPS Subfamilies Across Fungal Taxonomic Groups	174
3.3.5	Lineage Specific Expansions and Contractions	177
3.3.6	Subfamily Distribution	178
3.3.7	Hypothesized Origins Based on Taxonomic Distribution	180
3.3.8	Mono- and Bi- Modular NRPS Subfamilies	182
3.3.9	Multimodular NRPS Subfamilies	188
3.3.9.1	Diversity Within the EAS Subfamily	189

3.3.9.2	Evolutionary Mechanisms Giving Rise to Multimodular NRPSs	193
3.3.10	Stability of NRPS Gene Copy Number and Domain Architectures Across Subfamilies	200
3.3.11	Chain Termination Mechanisms	202
3.4	Conclusions	203
3.5	Materials and Methods	204
3.5.1	Identification of Putative NRPSs in Fungal Genomes	204
3.5.2	Annotation of Domain Architectures	205
3.5.3	Phylogenomic Analyses	206
3.5.4	Subfamily Identification and Modelling	208
3.5.5	Distribution of NRPS Subfamilies Across Fungal Taxonomic Groups	208
3.5.6	Lineage Specific Expansions and Variation in Birth-Death Rates	208
	Appendices – Chapter 3	211
	References – Chapter 3	250

LIST OF FIGURES

Figure 1.1	NRPSs in <i>Cochliobolus heterostrophus</i>	3
Figure 2.1	Ferrichrome Structure	90
Figure 2.2	Six modular architectures for ferrichrome synthetase NRPSs	102
Figure 2.3	Maximum Likelihood Phylogeny of A domains	104
Figure 2.4	Schematic representation of phylogenetic relationships among A and among C domains within each lineage.	107
Figure 2.5	Diagrammatic depiction of separate NPS2 (A) and NPS1/SidC	111
Figure 2.6	Evaluation of <i>C. heterostrophus</i> NPS2 and <i>A. nidulans</i> SidC with the PDH algorithm	114
Figure 2.7	3D modeling of selected NRPS AMP binding domains	118
Figure 2.8	Models for evolution of a hexamodular ancestral ferrichrome synthetase gene	127
Figure 3.1	Cartoons of tree topologies showing major NRPS subfamilies	165
Figure 3.2	ML phylogenetic tree (PhyML, WAG plus gamma) from the reduced A domain dataset	168
Figure 3.3	Lineage specific expansions and contractions in number of NRPS genes per genome	178
Figure 3.4	Hypothesized origins of major fungal NRPS subfamilies based on the oldest member of each subfamily	181
Figure 3.5	Conserved domain architectures for mono-bimodular NRPS subfamilies	184
Figure 3.6	Phylogeny of the ChNPS11/ETP/ChNPS12 subclade	186
Figure 3.7	Phylogenetic analysis of the Euascomycete subclade	191

Figure 3.8	Modular organization of Peptaibol synthetases and proposed evolution by tandem duplication	195
Figure 3.9	Phylogenetic groupings and modular organization of ChNPS1 and ChNPS3 showing recombinant structure of these NRPSs	198
Figure 3.10	Number and range of NRPSs and A domains for each subfamily	201

LIST OF TABLES

Table 1.1	Core Motifs in NRPS domains	7
Table 2.1	Fungal genomes and number of ferrichrome synthetases identified	100
Table 2.2	Key positions in AMP domain binding pocket identified by structural modeling	117
Table 2.3	Residues showing evidence of functional divergence	123
Table 3.1	Numbers of NRPSs per subfamily across fungal taxonomic groups	175

LIST OF APPENDICES

Appendix 2.1	135
Appendix 2.2	137
Appendix 2.3	138
Appendix 2.4	140
Appendix 3.1	211
Appendix 3.2	212
Appendix 3.3	225
Appendix 3.4	226
Appendix 3.5	228
Appendix 3.6	229
Appendix 3.7	239
Appendix 3.8	241
Appendix 3.9	243
Appendix 3.10	244
Appendix 3.11	245
Appendix 3.12	246
Appendix 3.13	248
Appendix 3.14	249

CHAPTER 1

GENERAL INTRODUCTION

1.1 Modular Proteins in Secondary Metabolism

Low molecular weight peptide and polyketide natural products are produced by nonribosomal peptide synthetases (NRPSs) and polyketide synthetases (PKSs), respectively. NRPSs have been found previously only in bacteria and fungi while PKSs have been documented in bacteria, fungi, plants [1] and more recently in a few animal species [2]. Both NRPSs and PKSs are large multidomain enzyme complexes. NRPSs and bacterial PKSs are often organized into repeated units known as modules. For NRPSs, a module is defined as a portion of the protein responsible for incorporation of one substrate molecule [3]. A set of three core domains comprise a functional module: 1) an adenylation (A- AMP) domain which activates and adenylates a substrate molecule with ATP, 2) a thiolation (T- THIOL) domain which binds the substrate to a phosphopantetheine group via a thioester bond, and a condensation (C-CON) domain which joins two adjacent substrates via a condensation reaction [4]. For PKSs, three core domains comprise a functional module: 1) an acyltransferase (AT) domain which primes and attaches the substrate to the 2) acyl carrier (ACP) domain which catalyzes the transfer of the growing polyketide acyl chain to the 3) ketosynthase domain (KS) domain active site which performs a condensation reaction between two substrates. Both NRPSs and PKSs accomplish chain elongation in a similar fashion utilizing one domain to recognize and activate the substrate (A for NRPS and AT for PKS) for bonding to an acyl carrier domain (T for NRPS and ACP for PKS) with a long sidearm which transfers the substrate to a final domain responsible for joining of two substrate molecules (C for NRPS and KS

for PKS). In NRPS biosynthesis, the C domain usually forms an amide bond between NRPS substrates [4, 5], although other types of chemical bond formation such as C-O esters have been observed [6]. In PKS biosynthesis, a Claisen condensation reaction between two carbon substrates, usually acetyl CoA and malonyl-CoA, forms a carbon-carbon bond in the β -keto chain [1, 7].

NRPSs and bacterial PKSs can consist of a single modular unit (monomodular) or tandem repeats of modular units (multimodular). The suite of 13 NRPS-encoding genes (*NPS*), plus one pseudogene, found in the Dothideomycete fungus, *Cochliobolus heterostrophus* demonstrates the diversity of domain architectures found even within a single species (Figure 1.1). The modular structure of the protein encoded by each gene is unique, except for duplicated copies of *NPS12*. In addition to mono- and multi-modular NRPSs, a hybrid protein consisting of an incomplete NRPS module (A-T; ChNPS7) followed by a PKS module (KS-AT-DH-KR-T-D; PKS24) is also present. PKS;NRPS hybrid proteins, the reverse of the hybrid in *C. heterostrophus*, which consist of an N-terminal PKS and a C-terminal single NRPS module have been identified in other fungi and bacteria [8-13].

While it was originally proposed that there is a one to one correspondence between NRPS modules and substrates in the peptide product such that the chemical composition of the metabolite produced by an NRPS can be predicted based on the order and specificity of A domains (termed the “colinearity” rule) [14], it is now clear that many NRPSs do not conform to the colinearity rule. Instead, it has been proposed that NRPSs can be classified into three types based on mechanism of biosynthesis of the corresponding metabolites: 1) Linear (type A), 2) Iterative (type B), and 3) Nonlinear (type C) [4]. Linear systems conform to the colinearity rule and show a one-to-one correspondence between modular organization and product. An example

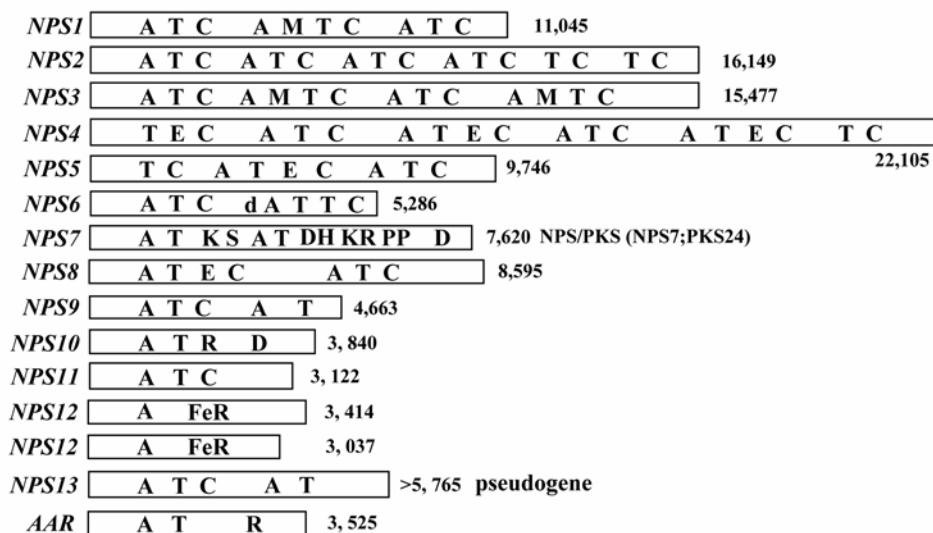


Figure 1.1: Diagram of 12 NRPSs plus one AAR, one NRPS;PKS hybrid (NPS7;PKS24), and one pseudogene (NPS13) found in the Dothideomycete *C. heterostrophus*. Annotation of domain architectures shows that with the exception of the duplicated copy of ChNPS12, NRPSs in *C. heterostrophus* have unique domain architectures. Domain abbreviations: Adenylation (A), Thiolation (T), Condensation (C), Dehydrogenase (D), Epimerization (E), Methylation (M), Thioester reductase (R), Beta-ketosynthase (KS), Acyl Transferase (AT), Dehydratase (DH), Ketoreductase (KR), and Ferric transmembrane reductase (FeR). Length of each gene in bp is shown to the right.

from fungi is the eleven module *Tolypocladium inflatum* NRPS, SimA which biosynthesizes cyclosporin, a cyclic peptide with eleven substrates [15]. Iterative systems are exemplified by Esyn1, a *Fusarium equiseti* bimodular NRPS which synthesizes the hexapeptide product, enniatin, via cyclization of dipeptide units by three iterative rounds of synthesis [3, 16]. Nonlinear systems include those in which two or more separate NRPSs are involved in synthesizing a single peptide product. While nonlinear systems are quite common in bacteria (eg. vibriobactin) [17], a single example is currently known from fungi. Synthesis of ergot alkaloids by *Claviceps sp.* involves two separate NRPSs, a monomodular LPS2 which activates D-lysergic acid

and transfers it *in trans* to the trimodular LPS1 which adds L-alanine, L-phenylalanine, and L-proline to complete synthesis of ergotamine [18, 19]. The ferrichrome synthetases discussed in Chapter 2, however, show a mixture of linear and iterative biosynthetic mechanisms within the same protein, suggesting that strict classification into these three types may not adequately describe the diversity of strategies utilized by NRPSs. Thus, it has become clear that the modular domain architecture of an NRPS may not always be predictive of its chemical product.

Similarly, PKS biosynthetic mechanisms have been classified into several types. Type I PKSs include those that, like animal fatty acid synthases (FAS), contain all domains for chain extension of the polyketide product in a single protein. Type I PKSs can be either modular (using multiple modules for chain extension) or iterative (reusing a single module for chain extension). Bacterial Type I PKSs are usually modular while fungal Type I PKSs are typically iterative [7, 20]. Type II PKSs include those that, like bacterial fatty acid systems, encode the domains needed for chain extension on separate proteins [7]. Type III PKSs, or chalcone synthases, were previously thought to be restricted to plants but have also been found in a number of bacteria [7]. Unlike Type I and Type II PKSs, these systems do not utilize acyl carrier domains but instead have thioester domains which form an acyl CoA thioester bond to bind substrates to the enzyme [20-24].

1.2 NRPS Biosynthesis

NRPS biosynthesis shows similarities to some aspects of ribosomal synthesis such as charging of amino acid substrates by acyl adenylation with ATP

(accomplished by aa-tRNA in ribosomal synthesis and the A domain in NRPS biosynthesis) and subsequent transfer to a carrier or carrier domain (tRNA in ribosomal synthesis and T in NRPS biosynthesis). However, NRPS synthesis differs significantly from ribosomal synthesis in many respects [3]. Ribosomal peptide synthesis involves two proofreading steps, 1) hydrolysis of an incorrectly activated amino acid by aa-tRNA synthetase, and 2) complementary base-pairing of tRNA and mRNA. In contrast, NRPSs lack proofreading ability and have been shown to tolerate relaxed substrate specificity [3]. For a number of NRPSs, it has been demonstrated that A domains will preferentially incorporate a particular substrate but are also able to incorporate other substrates depending on their relative concentrations, thus resulting in a diversity of products from a single NRPS [25-27].

Ribosomal protein synthesis is restricted to 20 amino acid (L) substrates while NRPS peptide synthesis can involve hundreds of different substrates, thus allowing for far greater diversity of products than could be accomplished by ribosomal synthesis [3, 28]. The substrates that NRPSs are known to utilize include the 20 amino acid (L) substrates of ribosomal synthesis as well as their D-isomers, δ -(L- α -aminoadipic acid) utilized in penicillin biosynthesis [29], α -amino butyric acids (L- α -butyric acid and (4R)-4[(E)-2-butenyl]-4-methyl-L-threonine) in cyclosporin A biosynthesis [15], hydroxy acids such as dihydroxybenzoate incorporated into the bacterial siderophores enterobactin and myxochelin A [30, 31], modified amino acids such as ornithines, carboxy acids, and acetate or propionate units [3]. Lipid and sugar groups may also be attached to produce lipopeptides and glycopeptides, respectively.

1.2.1 Adenylation Domains

The A domain of NRPSs (~550 aa in length) plays the primary role in recognizing and activating substrates by adenylation with ATP [3]. The A domain contains a number of conserved motifs (A1- A10) (Table 1.1) [32]. The A3 motif contains a highly conserved Ser/Thr/Gly rich motif that is shared by all members of the AMP binding superfamily (PF00501) of adenyating enzymes and functions in binding ATP [33].

The first crystal structure of an NRPS A domain [gramicidin S-synthetase A (1AMU)] revealed 10 residues in direct contact with the Phe substrate [35]. Phylogenetic analysis of the corresponding residues identified from an alignment of primarily bacterial and a few fungal A domains revealed clusters of A domains predicted to code for the same substrate which also shared similar residues in these 10 AA positions [36]. This finding led to the proposal of a 10AA ‘code’ for substrate specificity of amino acid activating A domains which is based on the amino acids found at these 10 residues (corresponding to the 1AMU positions 235, 236, 239, 278, 299, 301, 322, 330, 331, and 517) [36, 37]. Relatively few studies have investigated these residues via site-directed mutagenesis but the few experiments that have altered 10 AA code positions have resulted in a change in substrate incorporation [38, 39][36]. However, it has been shown that A-domains with distinct ‘codes’ may bind the same substrate and it has also been suggested that the code may not be applicable to smaller substrates [40]. The carboxy acid activating NRPSs such as the 2',3'-dihydroxybenzoic acid activating domain of DhbE from *Bacillus subtilis*, for example, shows a different set of residues involved in substrate recognition from those of amino acid activating domains [41]. Schwecke et al. [42] found 3 additional residues involved in binding N⁵-acyl-N⁵-hydroxy-L-ornithine (AHO) in the

Table 1.1: Consensus sequences for conserved core motifs of NRPS domains^a

Domain	Core Motif	Consensus Sequence
Adenylation (A)	A1	L(TS)Y _x EL
	A2	LKAG _x AYL(VL)P(LI)D
	A3 ^b	LA _{Yxx} YTSG(ST)TG _x PKG
	A4	FD _x S
	A5	N _x YGPTE
	A6	GEL _x l _x G _x G(VL)ARGYL
	A7	Y(RK)TGDL
	A8	GR _x D _x QVKIRG _x RIELGEIE
	A9	LP _x YP(IV)P
	A10	NGK(VL)DR
Thiolation (T)	T	D _x FF _{xx} LGG(HD)S(LI)
Condensation (C) ^c	C1	S _x AQ _x R(LM)(WY) _x L
	C2	RHE _x LRT _x F
	C3 (His)	MHH _x lSDG(WV)S
	C4	Y _x D(FY)AVW
	C5	(IV)G _x FVNT(QL)(~) _x R
	C6	(HN)QD(YV)PFE
	C7	RD _x SRNPL
Epimerization (E) ^c	E1	PIQ _x WF
	E2 (His)	HH _x lSDG(WV)S
	E3 (race A)	D _x LL _x A _x G
	E4 (race B)	EGHGRE
	E5 (race C)	RTVGWRT _{xx} TP(YV)PFE
	E6	P _{xx} G _x GYG
	E7	FNYLG(QR)
N-Methylation (M)	M1 (SAM)	VL(DE)G _x G _x G
	M2	NEL _{Sx} YRY _x AV VEx _{Sx} ARQ _x G _x LD
Thioesterase (TE)	Te	G(HY)S _x G
Reductase (R)	R1	V(LF)(LV)TG(AV)(TN)G(YF)LG
	R2	V _{xxx} VRA
	R3	GDL
	R4	VYPY _{xx} LR _x (PL)NV _{xx} T
	R5	GY _{xx} SKW _{xx} E
	R6	RPG
	R7	LE _{xx} (VI)GFL _{xx} P
Heterocyclization (Cyc) ^c	Z1	FPL(TS) _{xx} Q _x AY _{xx} GR
	Z2	RH _x (IM)L(PAL) _x (ND)G _x Q
	Z3	(DNR) _{xxxx} D _{xx} S
	Z3	(LI)P _{xx} (PAL) _x (LPF)P
	Z4	(TS)(PA) _{xxx} (LAF) _{xxxxxx} (IVT)L _{xx} W
	Z5	(GA)(DQN)FT
	Z6	P(IV)VF(TA)S _x L
Z7	QV _x (LI)D _x (QH) _{xxxxxxxxxx} W(DYF)	

^a Compiled from Konz and Marahiel [32]. ^b The A3 motif contains a highly conserved Ser/Thr/Gly rich motif shared by all members of the AMP binding superfamily (PF00501) of adenyating enzymes. ^c The Condensation (C), Epimerization (E), and Heterocyclization (Cyc) domains are evolutionarily related and share several similar core motifs [34]. The C3 and E2 domains share a histidine rich (HIS) motif.

Schizosaccharomyces pombe ferrichrome synthetase Sib1. Since the 10AA code is based primarily on bacterial sequences, its applicability to fungal A domains remains unclear. In reviewing the available fungal sequences for which substrates can be reliably assigned, Walton et al. [43] concluded that the 10AA code is of limited utility in predicting specificity for fungal A domains. The applicability of the 10AA code to predicting substrate specificity of fungal A domains of various ferrichrome synthetases is examined in Chapter 2.

1.2.2 Thiolation Domain

T domains (~80-100 aa in length) of NRPSs, also known as Peptidyl Carrier Protein (PCP) domain, belong to the larger class of ACP domains found in Fatty Acid Synthases (FAS), PKSs, and a number of other proteins [44] [3]. All ACP domains have a relatively conserved structure consisting of a four-helix bundle [3]. T domains also have a conserved core motif (Table 1.1) [32] containing an invariant serine residue to which a 4'PP cofactor is attached posttranslationally by a 4-phosphopantetheinyl transferase (PPTase) [3, 45]. The T domain attaches the activated acyl adenylated substrate to its 4-phosphopantetheine (PP) cofactor via a thioester bond and then acts as a flexible arm to carry the substrate to a C domain for peptide bond formation [3]. T domains show differences in sequence and structure depending on their location within the modular enzyme and helix 2 has been implicated as having a role in mediating interactions with other protein domains [3].

PPTases responsible for attaching the 4' PP cofactor are found in a wide variety of organisms including bacteria, fungi, plants, and animals [46]. In *Saccharomyces cerevisiae*, the PPTase Lys5 is involved in posttranslational modification of α -amino adipate reductase, an enzyme responsible for lysine synthesis

and which shows homology to NRPSs (discussed below) [47]. A single Lys5 homolog, Lys7, has been found in the fission yeast, *Schizosaccharomyces pombe* [48], and therefore likely interacts with both the α -aminoadipate reductase and the single NRPS (*sib1*) identified in *S. pombe* [42, 49]. Similarly, the *npgA/cfwA* PPTase from *A. nidulans* plays a role in NRPS mediated biosynthesis of penicillin [50] and knockouts have pleiotrophic effects on development [51] and pigmentation [52].

1.2.3. Condensation Domain

The C domains (~450 aa) are responsible for forming the peptide bond between two substrates via a condensation reaction resulting from the nucleophilic attack of the amino group of the downstream substrate (donor) on the carboxyl group of the upstream substrate (acceptor) [3, 5]. The crystal structure of the VibH amide synthetase producing vibriobactin in *Vibrio cholerae* shows similarities to NRPS C domains and revealed a structure consisting of two $\alpha\beta\alpha$ sandwiches with two entryways to the active site, one for the electrophile and the other for the nucleophile [3, 53, 54]. Similarly, two faces, a C-face or donor site where the nucleophile enters and an N-face or acceptor site for the electrophile have been identified in NRPS C domains [3]. The acceptor site has been shown to be able to discriminate between different nucleophiles based on stereochemistry as well as chemical features of amino acid side chains, thus demonstrating a role for the C domain in selectively accepting substrates from the downstream A domain (ie. substrate selectivity) [55-59]. A number of conserved motifs have also been characterized for C domains (Table 1.1) [32].

Various subgroups of C domain have recently been delineated by phylogenetic analysis [34]: 1) $^L C_L$ which catalyzes condensation between two L-amino acids, 2)

^DC_L which catalyzes condensation between a D-amino acid and an L-amino acid, 3) a starter C domain, 4) the related cyclization (Cyc) domain which creates heterocyclic oxazoline or thiazoline rings by cyclization of cysteine, serine, or threonine residues [60], 5) the closely related epimerization (E) domains which convert L-amino acids to a D configuration, and 6) dual E/C domains which catalyze both conversion from L to D configuration and subsequent peptide bond formation [34]. The starter C domain, which is the first C domain in lipopeptide synthetases and other NRPSs such as EntE which incorporate β -hydroxy acids, catalyzes condensation of a lipid of β -hydroxy-carboxylic acid to the substrate of the first A domain [34].

1.2.4. Termination Domains

Termination of chain elongation and release of the peptide product is accomplished by a variety of mechanisms. In bacteria, a thioester (TE) domain which accepts the peptide chain from an adjacent T domain and forms an acyl-O-TE-enzyme intermediate [61] which then undergoes nucleophilic attack either by one of the amino acids from the peptide chain to release a cyclic product or by a water molecule to form a linear product [3]. TE domains have also been shown to catalyze lactonization (attachment of the C-terminal carboxyl group to the hydroxyl group of an N-terminal β -hydroxy fatty acid in lipopeptides such as surfactin) as well as oligomerization of subunits for iterative NRPSs such as gramicidin and enterobactin synthetases [3]. The crystal structure of the surfactin (Srf)-TE domain shows similarities to serine esterases and lipases, members of the α - β hydrolase superfamily [3, 62, 63]. Release by TE domains is much less common among fungal NRPSs but has been observed for ACV synthetases [64] which are hypothesized to be of bacterial origin [29]. TE domains of ACV synthetases are atypical of other TE domains found in NRPSs in that they are

directly associated with an epimerization domain [64].

A number of alternative mechanisms for release of the peptide chain have been characterized for fungal NRPSs. In cyclosporin synthesis, for example, a specialized terminal C domain catalyzes amide bond formation between the amino group of the first peptide and the carboxyl group of the final peptide in the chain to accomplish head to tail cyclization [61]. This mechanism for C domain cyclization differs from TE mediated cyclization in lacking an acyl-O-C intermediate and is instead accomplished by direct nucleophilic attack on the thioester bond [61]. This cyclization mechanism has also been proposed for a number of other fungal NRPSs [32] including Enniatin synthetases [65], the related cyclooctadepsipeptide synthetases PF1022A [25], and HC-toxin synthetase HTS1 [66]. Both Enniatin and PF1022A are iterative NRPSs and it is hypothesized that the final C domain functions in a manner analogous to the final TE domain of the bacterial iterative NRPS synthesizing enterobactin by tethering and cyclizing the oligomers produced by successive rounds of synthesis [61].

Another mechanism, which has been demonstrated for the yeast α -amino adipate reductase Lys2, involves a terminal NAD(P)-dependent reductase (R) domain, which catalyzes a two-step reduction reaction involving 1) formation of an aldehyde by the NADPH/NADH dependent domain and 2) subsequent hydride transfer and reduction of the thioester bond linking the activated substrate to the T domain, thus resulting in release of the α -amino adipate 6-semialdehyde product with a reduced C-terminal carboxyl group [61] [47, 67]. A number of fungal NRPS synthetases, notably those making peptaibols, contain a reduced C-terminal carboxyl group and various other fungal NRPSs including *Aspergillus nidulans* EAA595380 and *Gibberella zeae* EAA75314 have a C-terminal reductase domain [67]. Alternatively, the aldehyde formed in the first step can be transaminated to form a terminal amide as has been

observed in the bacterial NRPS Mx1 which produces Saframycin [67] [61, 68]. The reductase domain involved in these reactions shows similarities to nucleoside-diphosphate-sugar epimerases, flavonols, reductase/cinnamoyl-CoA reductase, NAD dependent epimerases, and other NADPH dependent enzymes [67]. Yet another mechanism proposed for chain termination is the formation of a diketopiperazine ring through a cyclization reaction which has been demonstrated for the ergot alkaloids [69].

1.2.5. Decorating Domains

A number of other domains involved in modification of substrates after incorporation by the A domain are found in NRPSs. The epimerization (E) domain, which catalyzes the conversion of an amino acid substrate from the L to the D configuration [70], and the cyclization domain (Cyc), which catalyzes formation of heterocyclic ring structures from cysteine, serine, and threonine, are as discussed above, both closely related to C domains [34]. A number of conserved sequence motifs have been identified for both E and Cyc domains (Table 1.1) [34]. C-methyltransferase and N-methyltransferase (M) domains, which show similarity to both S-adenosyl-L-methionine (SAM)-dependent methyltransferases and DNA methyltransferases, catalyze transfer of a methyl group from an S adenosylmethionine to the α -amino of the amino acid substrate. Methylation (M) domains were first found in Enniatin synthetase where they form an internal part of the A domain between the A8 and A9 motifs [71] and later in cyclosporin synthetase [72]. An additional domain, termed the communication (COM) domain, has recently been shown to play a role in mediating protein-protein interactions and may facilitate crosstalk between different NRPS proteins [73].

1.3 Evolutionary Origins of NRPS and PKS Synthetases

1.3.1 Relationship of NRPSs and PKSs with Primary Metabolism

Some metabolic pathways associated with primary metabolism, particularly fatty acid biosynthesis via fatty acid synthases (FAS) show similarities to NRPS and PKS synthesis and suggest a common evolutionary origin of these three protein classes. FASs are also large mega-enzyme complexes composed of multiple interacting protein domains and all three classes of protein utilize acyl-activated substrates and an acyl carrier domain to transfer their substrates to a target molecule [44]. FASs and PKSs are most closely related and share a number of protein domains: ketosynthase (KS), acyl transferase (AT), acyl carrier protein (ACP), ketoreductase (KR), dehydratase (DH), and enoyl-reductase (ER). The first three domains (KS, AT, and ACP) are required for both FAS and PKS biosynthesis. KR, DH, and ER are also essential for FAS synthesis but are optional for PKS biosynthesis. Similar to iterative type I PKS systems, FASs also assemble their 16 carbon chain product via iterative use of the core set of domains. Perhaps the best evidence for a close relationship among these three classes of multimodular proteins is the presence of hybrid proteins in nature. An increasing number of hybrid PKS;NRPS or NRPS;PKS systems have been identified [74-76] [13]. The protein synthesizing mycosubtilin (MycB) contains a mixture of NRPS, FAS, and aminotransferase domains [77]. Hybrid PKS;FAS systems have also been identified in the slime mold *Dictyostelium discoideum* [78]. Freestanding FASs have also been shown to have a direct role in the synthesis of some PKS products including alfatoxin in *A. parasiticus* [79] and Sterigmatocystin in *Aspergillus fumigatus* [80] among others. Some PKSs are also known to produce fatty acid products [81].

1.3.2 Discovery of NRPSs and Related AMP Adenylating Enzymes

Lipmann was the first to recognize that cyclic peptides such as Gramicidin and Tyrocidine contain unusual D-amino acids [82] and are produced independent of ribosomes on large protein templates resembling fatty acid synthases [83]. Lipmann and others also discovered that an ATP driven mechanism was involved in substrate activation [84, 85]. It was recognized as early as the 1950's that adenylation with ATP to form an acyl-AMP adenylate intermediate occurs as the first reaction in many fundamental metabolic processes involving utilization of compounds with a carboxyl (COOH) group, including protein synthesis via aminoacyl tRNA synthetase (atTRSs), activation of acetate to form Acetyl Coenzyme A (Acetyl CoA), long chain fatty acid synthesis, oxidation of molecular oxygen by luciferases, and synthesis of benzoic, pantoic, biotin, and lipoic acids [86]. An increasing number of enzymes have been identified since which share this mechanism. Many of these proteins, including NRPSs, are classified within the AMP superfamily PF00501 (<http://pfam.sanger.ac.uk/>) all members of which adenylate substrates via ATP and are characterized by a Ser/Thr/Gly-rich P-loop like motif containing a conserved Pro-Lys-Gly triplet [(T,S)(S,G)G(T,S)(T,E)G(L,X)PK(G,-)] which is involved in binding AMP [87-89]. However, not all AMP adenylating enzymes belong to the AMP superfamily. Notably, aa-tRNA synthetases are structurally unrelated to the A domains of NRPSs and other AMP superfamily enzymes [33, 90].

Other members of the AMP superfamily include aryl activating enzymes such as DhbE [41], Bile acid-inducible operon [91], bacterial siderophore synthetases (EntF, EntE) [92], microbial 4-chlorobenzoate dehalogenase involved in degradation of halogenated hydrocarbons [89, 93], the plant defense compound 4-coumarate CoA

ligase [94, 95], fatty acid CoA ligases [96], acetyl CoA synthetase [97] and related enzymes [98], CPS1 and other acyl-CoA ligases [99], α -aminoadipate reductase (AAR) involved in lysine synthesis in fungi [47, 100], α -aminoadipate semi-aldehyde dehydrogenase (AAS) involved in lysine degradation in metazoans [101], the Ebony protein from *Drosophila melanogaster* [102], and D-alanine conjugating enzymes involved in bacterial cell wall biosynthesis [102] among others. The angR protein, a transcriptional activator which regulates response to Fe^{2+} also shows homology to these adenylating enzymes although it is not currently clear that it catalyzes an adenylation reaction. [103] While the evolutionary origins of this family of enzymes is unclear, similarity of 4-chlorobenzoate dehalogenase to enoyl-CoA hydratases/isomerases suggests that at least this member of the family may have evolved from the *B*-oxidation pathway of fatty acid degradation [89].

Many members of the AMP superfamily accomplish their enzymatic processes through two half-reactions: 1) adenylation of a substrate molecule with AMP to create an activated intermediate and 2) the subsequent transfer of this intermediate to a target molecule, usually either Coenzyme A (CoA) or a thiol acyl carrier domain [28, 33, 86, 97]. Firefly luciferase from the firefly *Photinus pyralis* was the first enzyme of this family to be characterized structurally and shows closest structural similarities to acyl-CoA ligases involved in a number of metabolic reactions involving adenylation and acyl transfer of CoA to a target molecule and second to NRPSs from both bacteria and fungi (Conti, 1998). The structure of this enzyme revealed two separate subunits, a small C-terminal and large N-terminal unit, separated by a cleft which is lined with a set of conserved motifs including the P-loop like motif as well as two other motifs 340 [YFWGASW]-x-[TSA]-E 344 and 420 [STA]-[GRK]-D 422 which show similarities to the A5 and A7 conserved motifs of NRPS A domains (Table 1.1). It was proposed that binding of the substrate induces a conformational change in the enzyme which

closes the cleft to create a tight binding pocket which excludes water and allows for the efficient oxidation of molecular oxygen [33]. Several other adenylating enzymes including acetyl CoA synthetase [97] and 4-chlorobenzoate dehalogenase [104] have been shown to undergo a similar conformational change upon substrate binding. However, this mechanism has not been explicitly shown for NRPSs and a different mechanism involving only slight structural movements has been demonstrated for the structure of the aryl acid activating AMP domain of DhbE synthetase [41]. No studies have clearly demonstrated which members of the AMP superfamily are most closely related to NRPSs. Structural and phylogenetic analyses suggest that acetyl CoA synthetase (1pg3) [97] is the closest structure to the NRPS phenylalanine activating domain of Gramicidin (1AMU) [28] (D.R. Ripoll, K.E. Bushley, and B.G. Turgeon, unpublished).

1.4 Mechanisms of Evolution of Modular Proteins

1.4.1 Models for Gene Family Evolution: Birth and Death, Divergence, and Concerted Evolution

Three basic models have been proposed for the patterns of evolution within a gene family: 1) Divergence, 2) Concerted Evolution, and 3) Birth-and-Death [105]. Divergent evolution occurs when orthologous copies of a gene in different taxa or duplicated genes within a single genome diverge by sequence evolution. In concerted evolution, gene conversion acts to homogenize differences between gene copies [105]. The birth and death model of evolution for multigene families postulates that new genes are created by duplication with some copies persisting while others are lost or degenerate into pseudogenes [106, 107]. Many classes of rapidly evolving genes in

other organisms including components of the animal immune system [107-109], olfactory and chemosensory genes [110, 111], and plant resistance genes [112] are thought to evolve by a birth-and-death process. However, even conserved genes such as histones and ubiquitins have been shown to evolve by a birth-and-death process followed by purifying selection [113, 114]. The Birth-and-Death model is likely the best model to explain the disjunct distribution of secondary metabolite genes observed in fungi, as neither of the other two models can account as fully for the heterogeneous distribution of these genes across taxa.

1.4.2 Evolution of Repeated Units in Proteins

The recognition and characterization of internal repeats within proteins and the processes involved in their generation dates back to the work of McLachlan beginning in the 1970's [115]. Internally repeated units within proteins vary in size and can range from a few nucleotides, to short amino acid motifs, to supersecondary structural elements, to large protein domains such as those found in NRPSs and PKSs [116, 117]. Those containing large domain repeats are a type of multidomain protein generally termed multimodular proteins. Various definitions for what constitutes a module have been proposed, including a segment of homology found in diverse proteins [118]. In the context of NRPSs and PKSs, a module has been defined as a repeated unit of protein domains responsible for a single catalytic reaction, such as the A-T-C repeat responsible for incorporation of a single substrate [3].

While the divergence, concerted evolution, and birth-and-death models of evolution were initially conceived to consider whole, individual proteins as the unit of evolution, they can also operate on the evolutionary unit of repeats within a single protein. Recently, a number of studies have documented concerted evolution

operating among tandem repeats within a variety of proteins including immunomodulating cell-surface proteins [119], sea urchin matrix proteins [120], abalone sperm lysine [121], and fungal self/non-self recognition (HET) proteins [122], among others. All of these studies, however, involve short amino acid repeats and not large protein domains as found in NRPS and PKS proteins. Sequence divergence, as well as birth-and-death processes, are also viable models for evolution of repeated units within proteins although less work has been done to characterize these.

Protein domains are considered a fundamental unit of protein evolution and domain rearrangements in multidomain proteins are thought to be important in the evolution of novel functions and organismal complexity [123, 124]. The processes giving rise to different domain architectures of multidomain proteins can be classified into three main types: 1) domain(s) exchange or recombination (ie. domain shuffling), 2) domain(s) indel/deletion analogous to indels in sequence evolution which can occur either internally or at the N- or C- terminus of a protein, and 3) domain(s) repetition or duplication [125]. In a comprehensive analysis of bacterial multidomain proteins, Pasek et al. (2006) found that insertions at the N-or-C termini of genes were the most frequent events, and most likely occurred by gene fusion events [125]. In a comparison of proteins from all three kingdoms of life (Archaea, Eubacteria, and Eukaryotes) [126] found that gene fusion was nearly four times as likely to occur as gene fission and that both fusion and fission are relatively rare events that generally occur only once in the evolution of a gene family and are propagated by duplication [126] or horizontal transfer [127]. A related process termed circular permutation has been proposed as a possible mechanism for the generation of domain diversity in NRPS and other multimodular proteins [128, 129]. Circular permutation is a complex process involving both domain duplication, deletion, and the evolution of new start and stop codons such that the N-terminal domain is transferred to the C-terminus of

the protein. A process of circular permutation has been demonstrated for a number of protein families including DNA methyltransferases (Jeltsch, 199, Bujnicki, 2002), swaposins [130], bovine trypsin inhibitor, glucosyltransferases, and glucosidases [128].

Domain repetition (duplication) and domain exchange/recombination (domain shuffling) are likely the most common mechanisms operating to give rise to the diversity of multidomain proteins [131]. The majority of protein domains appear to be related by duplication and protein domain families have been shown to follow a power law distribution showing a few very large families and many smaller families [131]. A correlation has also been observed between domain frequency and the tendency to recombine and form new domain combinations [131]. Protein domains have been shown to be mobile units and shuffling of these domains may have been a major force in the evolution of complex metazoans during the Cambrian explosion [123, 132]. However, it is also clear that only a fraction of all possible domain combinations occur in nature [133] and domain shuffling may be a less frequent than is commonly assumed [134]. The observation that in some classes of vertebrate proteins, domains are flanked by introns that share the same phase led to the exon-shuffling hypothesis which proposes that in ancestral genomes, small units corresponding to present day exons and/or supersecondary structures were assembled into multidomain proteins and shuffled between multidomain proteins by recombination in intervening introns [135]. However, recent evidence has shown that only a few classes of multidomain protein are flanked by in-phase introns [136, 137] and that many exon boundaries do not correspond to supersecondary structures [138].

Interestingly, a similar hypothesis has been proposed for modular evolution of PKSs. The “Linker Hypothesis” proposes that the short linker regions between individual domains (interdomain linker) and between modules (intermodule linker) are

the regions where domain shuffling can occur [139]. One study of PKSs in *Streptomyces avermitilis* has documented a case of recombination within an intradomain linker region between the AT and DH domains of a *PKS* gene and found greater incongruence in gene genealogies of the KS domains involved in substrate specificity than in any other domain [140]. There is also evidence for recombination and/or domain shuffling in NRPS systems. Studies of the microcystin (*mcy*) PKS;NRPS hybrid gene have documented a mosaic structure of genes and discordant gene genealogies indicative of both intragenic and intergenic recombination [141] and demonstrate a higher density of recombination breakpoints within A domains and T domains with little evidence for recombination in C domains [142]. The available evidence suggests that in nature, recombination and domain shuffling may occur more frequently among domains with a role in substrate specificity (A and KS in NRPS and PKS systems respectively), thus providing a rapid mechanism for evolution of new chemical compounds [140]. Domain shuffling and modular evolution has also been demonstrated in a number of other types of proteins including the bHLH transcription factors [143].

1.5 Fungal NRPSs: Evolution and Functional Classes of NRPs

1.5.1 Mechanisms Leading to the Discontinuous Distribution of Secondary Metabolite Genes in Fungi: Gene Clusters, Horizontal Gene Transfer, and Duplication and Differential Loss (DDL).

Most genes involved in secondary metabolite biosynthesis show a highly disjunct or discontinuous distribution across fungi and thus, closely related chemical products may be produced by highly divergent taxa with little or no evidence for these

compounds in intervening taxa. Two primary hypotheses have been proposed to explain this pattern of discontinuous distribution: duplication and differential loss (DDL) and horizontal gene transfer (HGT) of complete clusters of genes involved in production of a given metabolite. The DDL hypothesis is essentially the birth-and-death model of gene family evolution.

The identification of the genetic pathways for fungal secondary metabolites during the 1980's and 1990's revealed that genes involved in production of a given secondary metabolite are often clustered in the genome [144]. Subsequent work showed coregulated gene expression of genes within the cluster [145-147]. Coregulated gene clusters also appear in other eukaryotes and hypotheses regarding their origins and maintenance include: 1) the selfish operon hypothesis [148], 2) epistatic selection [149, 150], and 3) The Fisher Model which proposed that linkage arrangements which confer a selective advantage will be selected for in a population (an "orthotopic linkage system") [151, 152]. The selfish operon hypothesis was originally proposed for prokaryotic systems where transfer of operons coding for complete metabolic pathways confers a direct selective advantage to the receiving organism [153]. Walton [148], however, argues that clustering of fungal secondary metabolite pathways also allows transfer of complete or nearly complete metabolite pathways by horizontal transfer. Plausible evidence for horizontal transfer of fungal secondary metabolite clusters has been documented in a number of recent studies [154-156].

In fungi, certain types of metabolic pathways show greater evidence for clustering than others. A study of clustered genes in *S. cerevisiae*, a fungus lacking secondary metabolite production, suggests that clusters containing genes involved in a single pathway are rare in the genome and that, if present, generally fall into categories related to carbon utilization, siderophore utilization, vitamin synthesis, aryl-

sulfate utilization, and allantoin and nitrogen utilization [154, 157]. Other types of genes that show evidence for clustering in fungi include those involved in other nutrient utilization pathways [144] and pathogenicity genes [158]. Together with secondary metabolite pathways, these constitute “dispensable” metabolic pathways, or those not essential for growth or required for growth only under a specific set of conditions [144].

The evolution of dispensable gene clusters is clearly complex and may involve more than one mechanism. The seminal study of formation of the pathway for allantoin degradation in *S. cerevisiae* demonstrated that six of eight genes in this metabolic pathway were recruited from disperse genomic locations into the cluster [159]. Formation of the biotin prototrophy cluster in *S. cerevisiae* presents an even more complex scenario involving both gene duplication and horizontal transfer of individual genes, but not complete pathways, from bacteria [157]. Like ability to synthesize many secondary metabolites, ability to synthesize biotin is found only in some fungal species and notably some species contain a partial pathway which is able to synthesize biotin from intermediates at several stages in the pathway, suggesting that loss of pathway function may be quite common [157]. In fact, this study demonstrated that the eukaryotic biotin synthesis pathway found in most other fungi has been lost from all hemiascomycetes and replaced with a novel cluster which evolved through a combination of horizontal transfer of individual pathway genes from bacteria and recruitment of one copy of duplicated genes in *S. cerevisiae* [157].

Gene duplication and recruitment of duplicate genes no longer essential for primary metabolism provides a reasonable explanation for the origins of secondary metabolite clusters. The biotin pathway in yeast contains several paralogous genes that were likely recruited by neofunctionalization [157]. Formation of the PKS cluster for aflatoxin also suggests that several gene duplicates were recruited into the cluster

from dispersed genomic locations [160]. The study of ETP toxin clusters suggests that cluster genes share closest relationships with paralogous genes in filamentous fungi and supports the hypothesis of recruitment of cluster genes from duplicate copies elsewhere in the genome. Differences among gene content of ETP like clusters in different *Aspergillus* species also shows specific instances of differential recruitment in the replacement of a dipeptidase J cluster gene found in *A. fumigatus* with a related paralog in *A. flavus*, *A. oryzae*, and *A. clavatus* as well as occurrence of a unique gene in the *A. fumigatus* and *N. fischeri* clusters that does not occur in other *Aspergillus* species [156].

Another hypothesis is that clustering can be explained by horizontal transfer of complete operons of these pathways from bacteria to fungi [144]. However, relatively few secondary metabolite genes are known to be shared between bacteria and fungi and genes in fungal secondary metabolite clusters show GC content, introns, and codon bias characteristic of other fungal, not bacterial genes. The most convincing cases for transfer of secondary metabolite genes from bacteria to fungi remains genes encoding the ACV synthetases and the *C. heterostrophus* NPS;PKS hybrid gene, *NPS7;PKS24*, for which both the *PKS* [161] and the *NRPS* [162] portions group phylogenetically with bacterial genes [161] as discussed in Chapter 3. Horizontal transfer of individual pathway genes that have not been transferred as a complete operon has been implicated in the origins of the biotin prototrophy cluster in yeast [157], as described above, and in the transfer of beta-glucuronidase from bacteria to fungi [163].

Horizontal transfer of clusters between fungi has also been invoked as a mechanism to explain the discontinuous distribution of secondary metabolite genes within fungi. Horizontal gene transfer between fungi has been argued for pathogenicity gene clusters including the pea pathogenicity (PEP) gene cluster in

Nectria hematococca which resides on a dispensible chromosome [164], the virulence gene, ToxA, from *Stagonospora nodorum* to *Pyrenophora tritici-repentis* [165], a segment of 14 genes between *Cryptococcus neoformans* species [166], the ACE1 gene cluster in *Magnaporthe oryzae* [155], and the NRT2 high-affinity nitrate transporter from a basidiomycete to an ascomycete [154], and others [167].

Several studies have shown fairly convincing evidence for the horizontal transfer of clusters or partial clusters of secondary metabolite genes [155, 156, 168]. Both the incongruence of phylogenies of individual genes within a cluster with an accepted species phylogeny and the disjunct distribution of different subtypes of clusters even among closely related taxa provide evidence for a HGT scenario. However, studies to date on the evolution of complete clusters demonstrate a core set of genes common to all clusters and conclude that all ACE1 type clusters [155] and all ETP type clusters [156], most likely derive from a single ancestral cluster that was already assembled from a core set of genes in the ancestor of ascomycetes. These data argue against the frequent independent evolution of clusters in distinct taxa and instead suggest a pattern of frequent loss of gene clusters. Unlike other co-expressed gene clusters in eukaryotes, however, fungal secondary metabolite clusters often show evidence of extensive internal duplications and recombination, giving rise to a diversity of gene order and content among different cluster subtypes, [155, 169], although some clusters and duplicated regions of clusters do show relatively conserved gene order [156].

Many authors have suggested that secondary metabolite genes tend to be located in subtelomeric regions containing transposons and repetitive DNA which may contribute to the rapid evolution, rearrangement, and/or loss of clusters or parts of clusters [170-173]. While data are not available for all secondary metabolite clusters and clearly some types of NRPSs do not fall within subtelomeric regions [174], a

convincing case can be made that at least rapidly evolving and discontinuously distributed secondary metabolite genes do show a subtelomeric bias. A recent study of *Aspergillus fumigatus* has identified large subtelomeric genomic islands which contain the majority of lineage specific genes, including those involved in secondary metabolism [170]. These authors propose that these regions may function as genetic “dumps” for inactive genes and as factories for synthesis of novel genes by recombination. In the aflatoxin biosynthesis gene cluster, for example, recombination within telomeric regions has been implicated in loss of the entire aflatoxin cluster in non-aflatoxin producing strains [172]. Interestingly, in the protist, *Plasmodium falciparum*, a subtelomeric family of the related adenylating enzyme, acyl-CoA synthetase, has been identified which shows evidence for extensive duplication, gene conversion, recombination, and selection which is unusual for a putative “housekeeping” gene [175].

1.5.2. Known Functional/Chemical Classes of Fungal NRPSs and Their Distribution Across Fungi

Although fungi produce a diversity of NRPS products, the function and mode of action is known for only a few and the potential for discovery of new and useful products remains immense. Products of known function and mode of action are generally those with easily detectable phenotypes which affect human health and welfare. These include toxins functioning in plant or animal pathogenesis or those with demonstrated antimicrobial, anticancer, or antiviral properties. However, the importance of these compounds for fungi themselves in their natural habitats remains largely unknown.

The level of conservation of different classes of NRPSs varies widely and may

provide some clues as to function. Despite the fact that many NRPSs do show a highly discontinuous distribution across fungal taxa, some classes of NRPSs are relatively conserved, at least among filamentous ascomycetes, with most species containing at least one representative. These conserved classes correspond to homologs of *C. heterostrophus* NPS2 (intracellular siderophore/sexual reproduction), NPS6 (extracellular siderophore/oxidative stress), NPS4 (control of hydrophobicity of the conidial cell wall), NPS10 (oxidative stress, morphological development), and NPS12 (no known phenotype). Recent investigation of function suggests that these classes of NRPSs are conserved because their products perform fundamental roles for fungal cells, including scavenging and sequestering of iron, control of cell surface properties and cell wall development, both sexual and asexual reproduction, and defense against oxidative stress. At the other end of the spectrum are genes encoding NRPSs producing host-selective toxins, such as HC-toxin of *Cochliobolus carbonum* or AM-toxin of *Alternaria alternata* apple pathotype, which are produced only by a single race or pathotype within a species [66, 176]. The NRPSs producing these compounds are thus highly lineage specific and discontinuously distributed and clearly play important roles in defining ecological niche by allowing pathogenesis on a particular host species. Other lineage specific NRPSs may also play niche-specific roles by serving as antimicrobials in competition between organisms or performing other functions which affect fitness.

1.5.2.1 Conserved Homologs of ChNPS6, ChNPS4, ChNPS10, and ChNPS12

As discussed above, homologs of these four synthetases in *C. heterostrophus* tend to be relatively conserved across euascomycete fungi. ChNPS6 is the most

conserved, with a representative gene from all eukaryotes sequenced to date [162] (with the exception of the recently sequenced *C. purpurea* (C. Shardl, unpublished). The domain structure of ChNPS6 is also conserved, consisting of one complete A-T-C module followed by a module with a degenerate A domain (dA-T-C). However, ChNPS6 does show evidence for an ancestral duplication event in fungi as two copies are present in the genome of *Nectria hematococca* (S. Kroken, unpublished). Interestingly, data on chemical products of homologs of ChNPS6 suggest that the two phylogenetic groupings may correspond to distinct chemical products. *C. heterostrophus* ChNPS6 as well as homologs in *Alternaria brassicicola* [177], *Neurospora crassa* [178-181], and *Magnaporthe oryzae* [182] produce coprogen or a modified coprogen (N^o-dimethylcopenin in *A. brassicicola*), while members of the paralogous clade including *Fusarium graminearum* [177] and *Aspergillus* species produce triacetylfusarinine C [183]. Thus, while the gene and domain architecture are conserved, there are subtle differences in the chemical product likely due to differences in specificity of the A domain. Strains with deletions of ChNPS6 show decreased virulence on the corn host [162, 177], increased sensitivity to oxidative stress [162, 177], high-salinity, basic pH, and iron depletion, reduced asexual sporulation, and reduced pigmentation on minimal medium likely due to reduced accumulation of DHN-melanin [49].

Homologs of ChNPS10 also show increased sensitivity to oxidative stress as well as defects in development such as increased growth of aerial hyphae and irregular colony formation [49]. A homolog of the monomodular ChNPS10 (MAA1) has been characterized in the related Dothidiomycete, *Leptosphaeria maculans*. Both of these proteins show an unusual domain architecture with respect to NRPSs, consisting of an incomplete module (A-T domains) followed by a NAD(P)H-dependent reductase domain with closest hits to NADP(H) thioester reductase (R) domain (IPR010080)

followed by a dehydrogenase domain showing closest hits to short chain dehydrogenases (IPR002198) [162, 184] and is discussed further in Chapter 3. However, deletion strains of *L. maculans* MAA1 did not show any phenotype related to development or virulence [184]. A homolog of ChNPS10 is not present in all euascomycete taxa.

ChNPS4, another NRPS which has homologs in many euascomycete taxa, also showed morphological defects when deleted. The *nps4* mutant showed decreased hydrophobicity of the cell wall surface [49]. Analysis of the homolog in *A. brassicicola*, AbNPS2, showed that it was expressed exclusively during conidial development and that deletion strains also show decreased hydrophobicity of the conidial cell walls as well as a number of other phenotypes including abnormal morphology of the conidial cell wall, decreased spore production, decreased germination rates especially in older spores, and increase in lipid bodies [185]. Inoculation of Brassica plants with older spores (>14 days) also resulted in significantly decreased lesion size [185]. Kim et. al. [185] hypothesize that the product of AbNPS2 may either serve as a component of the conidial cell wall or function as a regulator or signal for cell wall development.

Homologs of ChNPS12 have been characterized to date only in the Dothideomycetes *C. heterostrophus* and *A. brassicicola*. Both the *C. heterostrophus* and *A. brassicicola* NPS12 homologs also show an unusual domain organization consisting of an A-T followed by a transmembrane domain showing closest similarity to ferric reductase transmembrane domain (IPR013130) (Discussed in Chapter 3). Preliminary functional characterization of the metabolite products of these genes shows they may have a role in maintaining ROS homeostasis (Lawrence, unpublished).

1.5.2.2 Siderophore Synthetases

In addition to producing extracellular siderophores such as ChNPS6, NRPSs also synthesize intracellular siderophores that sequester reactive Fe within cells [49, 186, 187]. The corresponding products are conserved in all euascomycetes and some basidiomycetes. Even so, variability in domain architectures is apparent due to duplication and loss of both complete A-T-C modules and individual A domains (Chapter 2). Known functions of the products of these NRPSs include roles in iron homeostasis, oxidative stress resistance, and both asexual [188-190] and sexual spore development [186, 191]. The chemical products and evolutionary history of ferrichrome synthetases are discussed extensively in Chapter 2.

1.5.2.3 ACV Synthetases

β -lactam Antibiotics. Among the best known NRPS products are the β -lactam antibiotics including penicillin (produced by *A. nidulans* and *P. chrysogenum*) and cephalosporin (produced by *C. acremonium*) although similar compounds have been isolated from other *Penicillium* species [192, 193] and more recently from the marine fungus *Kallichroma tethys* [194]. The NRPSs producing antimicrobial β -lactam antibiotics are generally called ACV-synthetases as they assemble a linear tripeptide (ACV) from three substrates: α -amino adipic acid, L-cysteine, and L-valine [18]. A large collection of modifying enzymes including isopenicillin-N-synthase (IPNS), transacetylases, and epimerases further transform the initial NRP tripeptide to either the cephalosporin or penicillin product [43].

The evolutionary origin of the ACV synthetases has been debated extensively. Claims of horizontal gene transfer from bacteria to fungi have been made based on

GC-content, a higher than expected sequence similarity between bacterial and fungal genes, the clustering of fungal genes in the pathway and lack of introns in many of the fungal genes, the relatively narrow distribution of species in which ACV synthetases are found, and the observation that transcription factors regulating β -lactam genes in fungi are wide-domain factors likely recruited from other sources [29, 195-202]. However, several aspects of the bacterial and fungal pathways differ and require explanation in order to entirely rule out the possibility of vertical transmission. Bacterial genes are all transcribed in the same direction whereas fungal genes are transcribed in divergent directions [196], the epimerization of isopenicillin-N to penicillin is catalyzed by different enzymes in bacteria and fungi [203], and the hydrophobic class of penicillins are known only from fungi [196].

1.5.2.4 Cyclosporin Synthetases

Cyclosporin A. The immunosuppressant Cyclosporin A (CsA; Sandimmune®), another well known NRP synthesized by the NRPS SimA and is a member of a group of cyclic undecapeptides produced by *T. inflatum* [15, 18]. The SimA gene, while smaller than those for peptaibol synthetases, is one of the largest open reading frames known (45.8kb), and encodes an NRPS that incorporates 11 substrates into the cyclic peptide Cyclosporin. In addition to amino acid substrates L-valine, L-leucine, L-alanine, and glycine, Cyclosporins contain three non-proteinogenic substrates: 1) 2-aminobutyric acid, 2) (4R)-4-[(E)-s-butenyl]-4-methyl-L-threonine [204], and D-alanine [18].

CsA functions as an immunosuppressant via interaction with a signal transduction pathway initiated by the serine-threonine specific protein phosphatase, calcineurin, which is conserved across eukaryotes and regulated by calmodulin in

response to intracellular Ca^{2+} concentrations [51, 205]. CsA first forms a complex with the immunophilin cyclophilin A (CyPA) which then binds and inhibits calcineurin [51, 205]. As a conserved target, calcineurin has been shown to play fundamental roles in cell differentiation and morphology in both mammals [206] and other fungi [207]. Mutants lacking calcineurin in both *Cryptococcus neoformans* [207] and *Neurospora crassa* [208] are deficient in mating, particularly in hyphal elongation and heterokaryon viability, as well as in filamentous growth. Calcineurin mutants in both *S. pombe* and *S. cerevisiae* also show a deficiency in mating [209, 210]. CsA plays a similar role in stunting hyphal elongation during haploid fruiting in *C. neoformans* [207]. In *A. fumigatus*, calcineurin mutants are also affected in hyphal growth, production of conidia, and adhesion of hyphae to host tissue, all of which likely contribute to decreased virulence in a murine host [206]. The *C. neoformans* mutants show a similar decrease in virulence [211]. In *C. albicans*, calcineurin mediates survival under membrane stress [210]. Given these effects on fungal fitness, the hypothesis that cyclosporin compounds evolved as toxins involved in competitive interactions between microorganisms seems plausible [51]. Cyclosporin synthetase was also one of the first NRPSs for which S-adenosyl methionine-dependent methyltransferases were recognized and described as important modifying domains in NRPS systems, functioning in methylating and demethylating substrates to give rise to a wide diversity of cyclosporin analogs [15, 212].

Cyclosporins are made by a number of different fungal species. *T. inflatum* produces 25 different analogs (Cyclosporins A-I and K-Z) [213, 214] and various other *Tolypocladium* species produce Cyclosporins [215-217]. Many fungi produce a consistent profile of these analogs, with Cyclosporin A-D being the major metabolites and cyclosporin E-F the minor metabolites [215]. These include a number of other groups of the *Hypocreaceae* (*Nectria*, *Neocosmospora*, *Trichoderma viride*) and

various *Fusarium*, *Tolypocladium*, *Isaria*, *Acremonium*, *Verticillium*, and *Chaunopycnis* species [215]. All of these fungi, except *Chaunopycnis*, share a common ecology, being parasitic on fungi or animals (nematodes, rotifers, and insects) [215], suggesting that Cyclosporins could also have evolved for parasitism on fungi or insects. They clearly have shown toxicity [218] and immunosuppressant activity in insects [219]. A number of other species outside of the *Hypocreaceae* produce only a single variety of cyclosporin, including *Acremonium luzulae* [220] (Cyclosporin C), *Cylindrotrichum*, *Leptostroma*, and others [215].

1.5.2.5 Cyclic Depsipeptide Synthases: Enniatin and Related Compounds

Enniatins are cyclohexadepsipeptides produced by a number of *Fusarium* species. Like Cyclosporin, they contain N-methylation domains which methylate the peptide product during NRPS biosynthesis [16, 221]. Enniatins are also closely related to a number of depsipeptide compounds from insect pathogenic fungi including Beauvericin [222] and Bassianolide [223] which are antiinsecticidal [222-224]. PF1022A is a related cyclic depsipeptide compound [25] although the sequence of the encoding NRPS remains under patent protection [225]. Cyclodepsipeptides contain repeated units of one 2-hydroxycarboxylic acid and one amino acid. In the case of Enniatin, this two modular unit is composed of one D-2-hydroxyisovaleric acid and one amino acid (either valine, leucine, or isoleucine) [16, 226]. Enniatin synthetases have an unusual domain architecture (C-A-T-C-A-T-T-C) with two adjacent thiolation domains and a C domain on both the N and C terminal ends of the protein [65, 226]. It has been shown that ESYN1 functions as a monomer, suggesting that it produces all three two unit monomers iteratively [227]. While iterative

biosynthesis followed by oligomerization and cyclization appears to be quite common in bacterial systems including synthetases for Enterobactin [30], Bacillibactin [228], Gramicidin S [229], and Surfactin [230], iterative systems in fungi have been less well characterized. Known examples include biosynthesis of Enniatin by ESYN1 and siderophores by *Schizosaccharomyces pombe* Sib1 ferrichrome synthetase [42] and other ferrichrome synthetases in fungi [42] (Bushley, Chapter 2). However, presumably iterative systems also operate in the synthesis of other cyclic depsipeptides.

In the case of both Gramicidin [231] and Enterobactin [232, 233] biosynthesis, monomer chains resulting from each round of synthesis are transferred to the C-terminal TE domain where they are held until the next round of synthesis has been completed. The TE domain then both oligomerizes the monomer subunits and releases the final peptide by cyclization. [233, 234]. It has been suggested that the final T-C domain repeat on ESYN1 performs the same function, holding monomeric units on the extra T domain and oligomerizing them with the final C domain [4]. These domains, the TE in bacterial systems and T-C repeat in fungal NRPSs, may also control the number of iterative cycles and the nature of chemical bonds which join monomers [4]. However, the actual mechanism controlling iterative biosynthesis is not currently understood.

1.5.2.6 Ergot Alkaloid Synthetases

Ergot Alkaloids. Other well known NRP products from fungi include the ergot alkaloids. Ergot alkaloids can be divided into four main types: 1) the clavines and elymoclavines, 2) D-lysergic acids, 3) D-lysergic acid derivatives such as D-lysergic amides, and 4) ergopeptines. All ergot alkaloids contain a tetracyclic ring

known as ergoline and are derived from the prenylated tryptophan precursor dimethylallyltryptophan (DMAT) [235-237]. Clavines and D-lysergic acids lack amide sidechains and thus do not require an NRPS for synthesis. The more complex D-lysergic acid amides and ergopeptines require NRPSs to attach a short peptide sidechain [236, 237]. The principle products known from *Aspergillus* species, the fumigaclavines, do not require an NRPS for biosynthesis.

Four NPSs (*cpps1-cpps4*) encode the NRPSs LPS1-LPS4 respectively in *Claviceps purpurea* [18]. LPS2 contains a single adenylation domain responsible for activation of D-lysergic acid and supplies this substrate *in trans* to LPS1 which adds L-alanine, L-phenylalanine, and L-proline to produce ergotamine [18]. As mentioned previously, this coordinated synthesis represents one of the only known examples of nonlinear synthesis in fungi [4]. LSP3 is a monomodular NRPS with unknown product while LPS4 is a tetramodular NRPS like LPS1 but contains different amino acid residues in the 10AA code positions within the first and second adenylation domains, suggesting it may add Val-Leu/Ile-Pro to produce ergocryptine [18].

Ergot alkaloids synthetases are prevalent in two families of fungi, Clavicipitaceae and Eurotiaceae. Producers within the Clavicipitaceae are generally grass endophytes including *Claviceps sp.*, *Epichl e festuca* and its asexual *Neotyphodium* anamorphs [238, 239] and *Balansia sp.* [240]. Within the Eurotiaceae, *A. fumigatus*, as well as a number of other *Aspergillus* species including *A. flavus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *A. oryzae*, *Aspergillus tamari*, and *Aspergillus versicolor* produce both clavine and ergopeptine alkaloids [240]. A wide variety of *Penicillium* species have been shown to produce ergot alkaloids, primarily clavines [240]. Ergopeptines (Ergocryptine), however, have been isolated from a number of other ascomycetes (*Botrytis fabae*, *Curvularia lunata*, *Hypomyces aurantius*, and *Sepedonium sp.* [240]. Ergot alkaloids have also been reported from

higher plants, particularly the plant family Convolvulaceae [240, 241]. Recently, however, a number of unknown Clavicipitalean endophyte species have been isolated from Convolvulaceae and are likely responsible for alkaloid production [242, 243].

Various clavines, lysergic acids, and cyclopiazonic acids, whose synthesis does not require an NRPS, have been isolated from ascomycetes (*Geotrichum candidum*, *Hypomyces aurantius*, *Sepedonium sp.*), as well as from basidiomycetes (*Corticium caeruleum*, *Lenzites trabeae*, *Pellucularia filamentosa*) and zygomycetes (*Phycomyces blakesleana*, *Mucor hiemalis*, *Rhizopus arrhizus*, *Rhizopus Nigricans*). These data suggest that the base of the pathway for ergot alkaloid biosynthesis, which does not depend on NRPS biosynthesis, is present in a wide variety of fungi. Although NRPSs completing the pathway to ergot alkaloid synthesis are discontinuously distributed in fungi, a recent study comparing the cluster of genes for ergot alkaloids biosynthesis in *A. fumigatus* with those of *C. purpurea* concluded that these clusters have a common genetic origin [244]. In analyzing the differences in ergot alkaloid production among closely related *Claviceps* species, Lorenz et al. [245] demonstrated that lack of alkaloid production is correlated with loss of *NPS* genes, specifically the *NPSs lpsB* and *lpsC*. The ancestor of all plant-associated *Claviceps sp.* likely possessed genes encoding NRPSs for ergopeptine synthesis which have been lost in some species [245, 246].

Ergot alkaloids are best known for their toxic and psychoactive effects in humans and animals. The drug LSD, otherwise known as D-lysergic acid, is notorious for its psychoactive hallucinatory effects and also has potential use in the treatment of a number of psychiatric disorders. Ergotism, caused by ingestion of alkaloid containing sclerotia of *Claviceps* species infecting rye and various grains, was first described as a disease in the 1800's but has seen numerous outbreaks throughout history and is implicated in such notorious historical events as the Salem Witch Trials

[235]. Ergotism, also known as “St. Anthony’s fire”, is characterized by hallucinations, convulsions, delusions, and gangrene. The toxic effects of ergot alkaloids in animals have also been extensively studied due to the ingestion of these compounds by livestock feeding on tall fescue, a common agricultural grass, infected with the Clavicipitalean endophyte *Neotyphodium coenophialum*. Symptoms in horses and other livestock include increased body temperature, decreased milk production, and reproductive problems [247].

Many of the toxic and psychoactive effects of ergot alkaloids can be attributed to the structural similarity of their tetracyclic ring ergoline to neurotransmitters like noradrenaline, dopamine, and serotonin [236]. Affinity for a particular neurotransmitter is influenced by the sidechain attached to the C-8 group of D-lysergic acid. Ergotamines have vasoconstrictive effects due to their greater affinity for adrenergic receptors and a synthetic analog of ergotamine, dihydroergotamine, is used in the treatment of migraine [236, 248, 249]. Ergovaline activates 5HT_{2A} (serotonin) and also causes constriction of blood vessels [250]. Bromocryptine has affinity for dopamine receptors and has been used in the treatment of Parkinson’s disease as have several other synthetic derivatives of ergolines [236, 251]. Ergotoxine, inhibits release of the peptide hormone prolactin [236]. The effects of ergot alkaloids on the immune systems of animals and humans may be mediated by changes in prolactin levels although other mechanisms such as interactions with dopamine, serotonin, or α -adrenoreceptors, inhibition of signaling pathways, or direct interaction with DNA may also operate [252].

These diverse effects of ergot alkaloids on animals may provide clues as to the evolutionary origins and function of these compounds for the producing fungi. The most commonly cited role of ergot alkaloids in the grass-endophyte symbiosis is providing anti-herbivore and anti-insect protection to the plant [238, 253]. Indeed,

ergopeptines seem to deter feeding of insects [254, 255], soil invertebrates [256], and mammals [257, 258]. Given recent phylogenetic evidence that the closest relatives of Clavicipitalean grass endophytes are *Hypocrella* and *Metarhizium*, both genera containing primarily insect pathogenic fungi, it seems likely that grass pathogens evolved from an animal pathogen ancestor via an interkingdom host jump [259]. Interestingly, clavine ergot alkaloids have been shown to have a role in conidiation, a process important for pathogenicity, in the human pathogen *A. fumigatus*. The bioactive and toxic effects of ergot alkaloids in animals can best be explained by the hypothesis that these compounds first arose with a role in animal (insect) pathogenesis.

1.5.2.7 Peramine Synthetase

Another NRPS found in Clavicipitalean fungi, PerA, is one of the only NRPSs for which an ecological role in symbiosis has been defined [18]. The product of PerA, peramine, a pyrrolopyrazine insect deterrent, confers a direct advantage to the plant in the symbiotic interaction between the *Epichloë/Neotyphodium* endophytes and their grass host [260].

1.5.2.8 Peptaibols

Peptaibols are linear peptides that have antimicrobial properties against both bacteria and fungi. Their name, which derives from the words PEPTide AIB AlcohOLS, accurately describes some of their key features. They include a high proportion (ranging from 14-56%) of AIB (α -aminoisobutyric acid), a type of α - α dialkylated amino acid, as well as isovaline substrates, have an acylated group on the

N-terminal end, and begin with an alcohol group on the C-terminal end [261-263]. They have been classified into three types, Long (including 18-20 residues), Short (including 11-16 residues), and Lipopeptaibols which have a fatty acid on the N-terminal end [263]. Their mode of action involves modification of lipid membranes, either by functioning as surfactants to destroy membrane integrity and cause leakage [264] or through formation of ion channels [265-267]. Peptaibols may also compromise the ability of membrane-associated proteins to synthesize cell walls and act synergistically with cell wall degrading enzymes produced by *Trichoderma* species to inhibit cell wall growth and thus pathogen growth [268, 269]. A number of *Trichoderma* species producing peptaibols are currently used as biocontrol agents against a variety of pathogens. More recently, peptaibols have been shown to also affect plants, having auxin-like activity and also eliciting an induced defense response [270].

Peptaibols are currently known only from soil inhabiting fungi and appear to be fairly lineage specific, occurring primarily in Hypocrealean anamorphs (*Trichoderma*, *Hypocrea*, *Clonostachys*, *Emericellopsis*, *Apiocrea*, *Sepedonium*, *Acremonium*, *Mycogone*, *Stilbella*, *Gliocladium*, and *Cephalosporium*) although one has also been isolated from *Verticimonosporium*, a pezizomycete [263]. The ecology of fungi from which peptaibols have been isolated as well as their antimicrobial properties are suggestive of antibiosis in the competition for space and resources which are both likely to be scarce in soil environments.

1.5.3.9 Diketopiperazines and ETP toxins

Diketopiperazines are a diverse groups of compounds found in both bacteria and fungi that are characterized by two amino acid substrates cyclized to form a

diketopiperazine ring [18]. Fungal diketopiperazines include tremorgenic fumitremorgines from *Aspergillus* (diketopiperazine fumitremorgin B) [18] and *Penicillium* species and Epipolythiodioxopiperazine (ETP) toxins. ETPs also contain an internal disulphide bridge in the diketopiperazine ring that is responsible for the toxicity of these compounds [271, 272]. The better known ETP toxins include Gliotoxin from *A. fumigatus* [271] and Sirodesmin PL produced by *Leptosphaeria maculans* [273, 274] but 14 different compounds have been isolated from a diverse group of fungi including four classes of euascomycetes (Dothideomycetes, Eurotiomycetes, Laconoromycetes, and Sordariomycetes) [275] as well as two basidiomycetes (*Stereum hirsutum* [276] and *Hyalodendron* sp [277]) lichens (scabrosin ester ETP)[278], and the pathogenic yeast *Candida albicans* [275, 279]. ETPs are especially common in Eurotiomycetes, including various *Aspergillus* and *Chaetomium* species but distribution of ETP toxins is highly discontinuous. In many cases, closely related species may not produce the same compound while distantly related species may produce remarkably similar compounds [275]. However, as mentioned above, examination of complete ETP clusters suggests that they all derive from a common core ancestral cluster [156].

ETPs are toxic to fungi, bacteria and viruses [280] and clearly are involved in virulence of the producing fungi to both animal and plant hosts, causing apoptotic [281, 282] and necrotic cell death, direct cytotoxic effects [283], and immunosuppressive effects [284, 285]. Gliotoxin is the best characterized functionally of the ETP toxins because it has potential as an antiviral agent [286] as it inhibits reverse transcriptase [287] and it has potential also as an anticancer agent [288, 289]. There are two proposed mechanisms for Gliotoxin toxicity: 1) inhibition of protein activity through formation of bonds between the disulphide bridge of Gliotoxin and thiol residues in proteins, and 2) generation of reactive oxygen species

through redox cycling between the reduced (dithiol) to the oxidized (disulfide) form of Gliotoxin [275, 290]. Gliotoxin inhibits a number of cellular proteins which may explain some of its virulence effects. Inhibition of NF- κ B, a transcription factor controlling expression of cytokines involved in the inflammatory immune response, may be responsible for the immunosuppressive effects of gliotoxin observed in animals [285]. Gliotoxin also causes apoptosis, interacting with adenine nucleotide transporter (ANT), an important gatekeeper of apoptosis via a thiol redox-dependent mechanism [291, 292].

A similar redox dependent interaction with a plasma membrane calcium channel causes calcium influx which may result in further oxidative stress [293]. Thus, it is clear that thiol-disulfide exchanges play an important role in mediating interactions with protein targets [294] and that these are in turn dependent on redox status of the disulfide bridge [290]. ETP toxins cycle through the reduced (dithiol) and oxidized (disulfide) forms of the disulfide bridge and are capable of generating reactive oxygen species (ROS) by this mechanism [275]. While ROS may play a direct role in toxicity for some ETP compounds such as sporodesmin [295-297], it cannot explain all toxic effects of gliotoxin as oxidative stress does not appear to play a role in gliotoxin mediated apoptosis [298].

The role of Sirodesmin PL, produced by the Dothideomycete *Leptosphaeria maculans*, in the development of blackleg disease of *Brassica napus* (canola) is more equivocal [299] although Sirodesmin PL mutants have shown decreased virulence on stems of canola [300]. In studies on the mode of action of Sirodesmin PL, Rouxel et. al. [274] found that, like the antiviral activity of gliotoxin which inhibits viral reverse transcriptase, Sirodesmin PL also inhibits RNA replication [274]. Interestingly, like the protective effect of zinc on Sporodesmin toxicity [301], addition of the metals Zn, Cd, and Hg ameliorates this effect [274, 302]. In cells, Zn²⁺ is essential for proper

functioning of RNA and DNA polymerases. Rouxel et al. (1988) [274] suggest that the inhibition of RNA synthesis by Sirodesmin PL may be caused by the toxin interacting with and binding intracellular zinc rather than the widely held hypothesis that this effect is due to the interaction of the sulphide bridge with a sulfate group in RNA polymerase [294, 303].

1.5.3.10 Dothideomycete Host-Selective Toxins: HC-toxin, AM-Toxin, and Victorin

Dothideomycetes, including *C. heterostrophus*, are prolific producers of secondary metabolites synthesized by NRPSs and PKSs. Many of these compounds are highly lineage specific, found only within specific races of a single species. They are termed host-specific or host-selective toxins because they are responsible for development of disease symptoms on specific host plants [304-306]. Among the most well known of these are HC-Toxin [66], AM-toxin [176], and likely Victorin produced by *Cochliobolus carbonum*, *Alternaria alternata*, and *Cochliobolus victoriae* respectively.

HC-toxin is produced by Race 1 of *Cochliobolus carbonum* [66] and allows colonization of corn with double recessive genotypes for the Hm1 and Hm2 loci coding for HC-toxin reductase, an enzyme that destroys HC-toxin activity [307-309]. HC-toxin is a cyclic tetrapeptide composed of D-Pro-L-Ala-D-ala-L-Aeo where Aeo stands for the unusual non-amino acid substrate 2-amino-9,10-epoxi-8-oxodecanoic acid [309]. The mode of action for HC-toxin is inhibition of deacetylation of histones H3 and H4 [310, 311]. Reversible histone acetylation controls many biological processes that involve chromatin, including regulation of gene expression, cancer, circadian rhythm, developmental processes, and pathogenesis. A number of related

histone deacetylase (HDAC) inhibitors, including Trapoxin and Trichostatin have been shown to be promising anticancer agents as they appear to reverse oncogene transformed cells [312-315]. Six other compounds structurally and biochemically related to HC-Toxin are known from fungi. These metabolites are extremely discontinuously distributed, being present in the related Dothideomycete *Alternaria brassicicola*, two *Fusarium* species, and other filamentous ascomycetes from diverse ecological niches ranging from plant pathogens, to saprobes, to nematode pathogens [309]. However, no homologs of the gene encoding HTS1, the HC-Toxin synthetase, have been identified yet in other fungi.

A. alternata is known to produce nine host-selective toxins which restrict host range of various pathotypes on specific host species [316]. As is typical of all genes encoding host-selective toxins, AM-toxin synthetase has a restricted distribution and is found only in the apple pathotype of *A. alternata* which causes Alternaria blotch on susceptible apple cultivars [176]. AM-toxin is a cyclic peptide composed of four substrates [317] and has two main sites of action in the cell: 1) the chloroplasts [318] and 2) the plasma membrane-cell wall interface [319].

Victorin is another cyclic peptide host-selective toxin [320] produced by *C. victoriae* and responsible for the development of Victoria Blight on Oats [321]. Victorin also causes an apoptotic response and cell death [322, 323]. However, the genetic locus (Tox3) responsible for its synthesis remains unidentified. Susceptibility to Victorin maps to the oat *Vb* locus which is inseparable from the Pc-2 locus involved in resistance to crown rust of oats caused by *Puccinia coronata* [324]. Recently, *Arabidopsis* mutant lines susceptible to Victorin were reported [325, 326]. In *Arabidopsis*, susceptibility is conferred by mutations in the LOV1 gene encoding an NBS-LRR protein required for both defense responses and programmed cell death (PCD) [326, 327]. This is the first indication that a host-selective peptide toxin may,

like an avr protein, interact with a plant resistance gene product [326]. However, other recent work has suggested that the Pc locus in *Sorghum bicolor*, responsible for resistance to the phytotoxic peptide producing fungus *Periconia circinata*, may also encode an NB-LRR gene [328].

1.5.3.11 Fungal PKS;NRPS Hybrids

Fungal PKS:NRPS hybrid synthetases are known from a number of eucaryote taxa. The first fungal PKS;NRPS gene to be cloned (FusA) encodes a hybrid PKS;NRPS synthetase (FUSS) responsible for production of Fusarin C, an acyl tetramic acid compound in both *Fusarium verticillioides* and *Fusarium venenatum* [329]. Tetramic acids include compounds containing the tetramic acid (2,4-pyrrolidinedione) ring [330]. They have been isolated from organisms as diverse as slime molds [331], fungi, and marine sponges [329]. Reported bioactivities of tetramic acids include antibiotic, cytotoxic, antiviral, antitumor, antifungal, and antibacterial properties [330]. The related fungal tetramic acid, Equisetin, produced by the hybrid synthetases eqiS in *Fusarium heterosporum* shows broad cytotoxicity and also inhibits HIV-1 integrase [332]. Both FUSS and eqiS have similar domain structures consisting of a monomodular PKS component including domains KS-AT-DH-M-ER-KR-AC followed by a monomodular NRPS unit (C-A-T). Unlike bacterial PKS;NRPS hybrids which may include multimodular PKS components, all characterized fungal PKS;NRPS hybrids have a single monomodular iterative Type I PKS. Due to the presence of an extremely large and unusual intron (546 bp) and degeneration of core NADPH binding motifs, the (ER) domain is hypothesized to be inactive [329]. Both FUSS and eqiS show close similarity to the Lovastatin nonaketide synthase (LNKS) which also consists of a single iterative Type I PKS with

a degenerate ER domain followed by a single C domain [161, 329, 332, 333]. As the C domain in LNKS also lacks an essential histidine in the core 3 motif (HHxxxDG) [334] (Table 1.1), it is likely nonfunctional but it is interesting to speculate that LNKS originated from a PKS;NRPS that has lost its NRPS component [161, 335]. Similar PKSs with a truncated NRPS module are also found in bacteria such as the MlcA gene for Compactin biosynthesis in *Penicillium citrinum* [336] [335]. Previous phylogenetic analyses based on the KS and AT domains suggest that the fungal PKS;NRPSs form a monophyletic group separate from other PKS enzymes in fungi [161, 335, 337] which includes several other PKSs which either lack an NRPS component (BfPKS4, BfPKS6, ChPKS16, ncu08399) or have a single C domain at the C-terminus (LNKS, Syn7, Syn6, MlcA) [335]. Other fungal PKS;NRPS hybrid synthetases with domain structures identical to FUSS and eqiS have previously been identified in the genomes of *Botrytis fuckeliana* [161], *Gibberella moniliformis* [161], *Fusarium graminearum*, *Magnaporthe oryzae* [335], *A. nidulans* [9], and *A. fumigatus* [8]. *Aspergillus* species produce a number of known compounds related to tetramic acids. An *A. fumigatus* hybrid gene (afu8g00540, Pso) has been shown to be responsible for synthesis of the tetramic acid Pseurotin A. The single PKS;NRPS in *A. nidulans* (ApdA) produces related compounds Aspyridones A and B which contain a pyridone moiety [9]. Pyridones and pyridine alkaloids often form yellow pigments and have been isolated primarily from insect pathogens including Militarionone D from *Paecilomyces militaris* [338], Tenellin from *Beauveria bassiana*, and Farinosone A from *Paecilomyces farinosus* [339]. Other known tetramic acid like compounds likely produced by PKS;NRPSs but for which genetic loci have not been identified include Pramanicin produced by *Stagonospora* sp., Zopfiellamide produced by *Zopfiella latipes* [329], Tenuazonic acid [340] and Pyrichalasin H [341] produced by *M. oryzae*.

Among euascomycetes, *M. oryzae* contains an unusually large number (nine)

of PKS;NRPS hybrids. Six of these are PKS:NRPSs with a complete NRPS C-terminal module (*ACE1*, MGG03810, MGG03818, MGG09589, *SYN6*, and *SYN8*) [335], two are LNKS-like (*Syn6* and *Syn7*) with a truncated C-terminal NRPS module, [335] and one (*Syn9*) is incomplete in sequence [337]. The *ACE1* gene is novel in that it is the only known PKS or NRPS secondary metabolite that has been shown to function in avirulence signalling with the plant host. Mutant strains lacking the *ACE1* gene are virulent to rice while the wild-type strains are avirulent on the same rice cultivar containing the cognate resistant gene Pi33 [335]. In experiments to characterize the nature of the interaction suggest that the secondary metabolite produced by *ACE1*, not the NRPS protein itself, is involved in avirulence signaling [335]. Functional analyses have shown that *ACE1* NRPS is localized and the corresponding gene is expressed in the cytoplasm of the appressorium [12]. Expression studies have shown that out of the nine potential PKS;NRPS hybrid genes in *M. oryzae*, four appear to be expressed during infection. *SYN2* and *SYN8* are expressed in appressoria, although at lower levels than *ACE1*, suggesting a role in penetration or early colonization [337]. *SYN6* is expressed in hyphae colonizing leaves and thus likely plays a role in the colonization phase [337]. However, knockout mutants of each of these genes did not show reduced pathogenicity. Thus, they may not be essential for infection although it is also possible that they play redundant roles [337].

The distribution of PKS:NRPSs is highly discontinuous and there does not seem to be a clear pattern with respect to lifestyle (eg. pathogens v.s. saprobes) [337]. Some fungi, including *C. heterostrophus* [162], lack a representative .

1.5.2.12 NRPS;PKS Hybrids

While NRPS;PKS hybrids are common in bacteria [13], very few have been identified in fungi. To date, the only reported NRPS;PKS hybrid is *ChNPS7;PKS24* from *C. heterostrophus*. However, a putative homolog of this gene has been identified in *Chaetomium globosum* (Bushley, unpublished). As discussed above, some have argued that ChNPS7 was horizontally transferred from bacteria to fungi as *ChNPS7;PKS24* falls within a large clade of bacterial sequences in phylogenies of both the PKS KS domain [2, 161] and the NRPS A domain [162] (Bushley, Chapter 3). The product of ChNPS7 is currently unknown and available data do not show a phenotype predictive of function [162].

1.6 Objectives

The overall objectives of this research were twofold. The first objective, addressed in Chapter 2, was to dissect the fine-scale evolutionary mechanisms by which NRPSs generate the incredible diversity of both domain architectures and chemical products observed in fungi. The second objective, addressed in Chapter 3 was to characterize the broad-scale distribution and evolutionary relationships of NRPSs across fungi with the goal of identifying subgroups by phylogenetic analysis of unknown *NPS* genes with those of known function.

REFERENCES

1. Hopwood DA, Khosla C: **Genes for polyketide secondary metabolic pathways in microorganisms and plants.** *Ciba Foundation Symposia* 1992, **171**:88-112.
2. Castoe TA, Stephens T, Noonan BP, Calestani C: **A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs.** *Gene* 2007, **392**(1-2):47-58.
3. Finking R, Marahiel MA: **Biosynthesis of nonribosomal peptides.** *Annual Review of Microbiology* 2004, **58**:453-488.
4. Mootz HD, Schwarzer, Dirk, and Mohamed A. Marahiel: **Ways of assembling complex natural products on modular nonribosomal peptide synthetases.** *ChemBioChem* 2002, **3**:490-504.
5. Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA: **Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain.** *Journal of Biological Chemistry* 1998, **273**(35):22773-22781.
6. Zaleta-Rivera K, Xu CP, Yu FG, Butchko RAE, Proctor RH, Hidalgo-Lara ME, Raza A, Dussault PH, Du LC: **A bidomain nonribosomal peptide synthetase encoded by *FUM14* catalyzes the formation of tricarballic esters in the biosynthesis of fumonisins.** *Biochemistry* 2006, **45**(8):2561-2569.
7. Hopwood DA: **Genetic contributions to understanding polyketide synthases.** *Chemical Reviews* 1997, **97**(7):2465-2497.
8. Maiya S, Grundmann A, Li X, Li SM, Turner G: **Identification of a hybrid PKS/NRPS required for pseurotin A biosynthesis in the human pathogen *Aspergillus fumigatus*.** *ChemBiochem* 2007, **8**(14):1736-1743.

9. Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C: **Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans***. *Nature Chemical Biology* 2007, **3**(4):213-217.
10. Brendel N, Partida-Martinez LP, Scherlach K, Hertweck C: **A cryptic PKS-NRPS gene locus in the plant commensal *Pseudomonas fluorescens* Pf-5 codes for the biosynthesis of an antimetabolic rhizoxin complex**. *Organic & Biomolecular Chemistry* 2007, **5**(14):2211-2213.
11. Rees DO, Bushby N, Cox RJ, Harding JR, Simpson TJ, Willis CL: **Synthesis of [1,2-C-13(2), N-15]-L-homoserine and its incorporation by the PKS-NRPS system of *Fusarium moniliforme* into the mycotoxin fusarin C**. *Chembiochem* 2007, **8**(1):46-50.
12. Collemare J, Pianfetti M, Houle AE, Morin D, Camborde L, Gagey MJ, Barbisan C, Fudal I, Lebrun MH, Boehnert HU: ***Magnaporthe grisea* avirulence gene *ACE1* belongs to an infection-specific gene cluster involved in secondary metabolism**. *New Phytologist* 2008, **179**(1):196-208.
13. Du LH, Sanchez C, Shen B: **Hybrid peptide-polyketide natural products: Biosynthesis and prospects toward engineering novel molecules**. *Metabolic Engineering* 2001, **3**(1):78-95.
14. Grunewald J, Marahiel MA: **Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides**. *Microbiology and Molecular Biology Reviews* 2006, **70**(1):121-146.
15. Weber G, Schorgendorfer K, Schneiderscherzer E, Leitner E: **The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame**. *Current Genetics* 1994, **26**(2):120-125.
16. Haese A, Schubert M, Herrmann M, Zocher R: **Molecular characterization of the Enniatin synthetase gene encoding a multifunctional enzyme catalyzing N-methyldepsipeptide formation in *Fusarium-scirpi***. *Molecular Microbiology* 1993, **7**(6):905-914.

17. Keating TA, Marshall CG, Walsh CT: **Reconstitution and characterization of the *Vibrio cholerae* vibriobactin synthetase from VibB, VibE, VibF, and VibH.** *Biochemistry* 2000, **39**(50):15522-15530.
18. Hoffmeister D, Keller NP: **Natural products of filamentous fungi: enzymes, genes, and their regulation.** *Natural Product Reports* 2007, **24**(2):393-416.
19. Riederer B, Han M, Keller U: **D-lysergyl peptide synthetase from the ergot fungus *Claviceps purpurea*.** *Journal of Biological Chemistry* 1996, **271**(44):27524-27530.
20. Moss SJ, Martin CJ, Wilkinson B: **Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity.** *Natural Product Reports* 2004, **21**(5):575-593.
21. Austin MB, Izumikawa M, Bowman ME, Udway DW, Ferrer JL, Moore BS, Noel JP: **Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates.** *Journal of Biological Chemistry* 2004, **279**(43):45162-45174.
22. Austin MB, Noel AJP: **The chalcone synthase superfamily of type III polyketide synthases.** *Natural Product Reports* 2003, **20**(1):79-110.
23. Jez JM, Ferrer JL, Bowman ME, Austin MB, Schroder J, Dixon RA, Noel JP: **Structure and mechanism of chalcone synthase-like polyketide synthases.** *Journal of Industrial Microbiology & Biotechnology* 2001, **27**(6):393-398.
24. Jez JM, Austin MB, Ferrer JL, Bowman ME, Schroder J, Noel JP: **Structural control of polyketide formation in plant-specific polyketide synthases.** *Chemistry & Biology* 2000, **7**(12):919-930.
25. Weckwerth W, Miyamoto K, Inuma K, Krause M, Glinski M, Storm T, Bonse G, Kleinkauf H, Zocher R: **Biosynthesis of PF1022A and related cyclooctadepsipeptides.** *Journal of Biological Chemistry* 2000, **275**(23):17909-17915.

26. Wiest A, Grzegorski D, Xu BW, Goulard C, Rebuffat S, Ebbole DJ, Bodo B, Kenerley C: **Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase.** *Journal of Biological Chemistry* 2002, **277**(23):20862-20868.

27. Rebuffat S, Goulard C, Bodo B: **Antibiotic peptides from *Trichoderma-harzianum* - harzianins Hc, proline-rich 14-residue peptaibols.** *Journal of the Chemical Society-Perkin Transactions I* 1995(14):1849-1855.

28. Conti E, Stachelhaus, T., Marahiel, MA, and Brick, P.: **Structural basis for the activation of phenylalanine in the nonribosomal biosynthesis of gramicidin S** *EMBOJ* 1997, **16**:4174-4183.

29. Martin JF: **alpha-amino adipyl-cysteinyl-valine synthetases in beta-lactam producing organisms - From Abraham's discoveries to novel concepts of non-ribosomal peptide synthesis.** *Journal of Antibiotics* 2000, **53**(10):1008-1021.

30. Gehring AM, Mori I, Walsh CT: **Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF.** *Biochemistry* 1998, **37**(8):2648-2659.

31. Silakowski B, Kunze B, Nordsiek G, Blocker H, Hofle G, Muller R: **The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15.** *European Journal of Biochemistry* 2000, **267**(21):6476-6485.

32. Konz D, Marahiel MA: **How do peptide synthetases generate structural diversity?** *Chemistry & Biology* 1999, **6**(2):R39-R48.

33. Conti E, Franks NP, Brick P: **Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes.** *Structure* 1996, **4**(3):287-298.

34. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH: **Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution.** *BMC Evolutionary Biology* 2007, **7**:78.

35. Conti E, Stachelhaus T, Marahiel MA, Brick P: **Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of Gramicidin S.** *Embo Journal* 1997, **16**(14):4174-4183.
36. Stachelhaus T, Mootz HD, Marahiel MA: **The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.** *Chemistry and Biology (London)* 1999, **6**(8):493-505.
37. Challis GL, Ravel J, Townsend CA: **Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.** *Chemistry and Biology (London)* 2000, **7**(3):211-224.
38. Eppelmann K, Stachelhaus T, Marahiel MA: **Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics.** *Biochemistry* 2002, **41**(30):9718-9726.
39. Stevens BW, Lilien RH, Georgiev I, Donald BR, Anderson AC: **Redesigning the PheA domain of Gramicidin synthetase leads to a new understanding of the enzyme's mechanism and selectivity.** *Biochemistry* 2006, **45**(51):15495-15504.
40. Lautru S, Challis GL: **Substrate recognition by nonribosomal peptide synthetase multi-enzymes.** *Microbiology (Reading)* 2004, **150**(Part 6):1629-1636.
41. May JJ, Kessler N, Marahiel MA, Stubbs MT: **Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(19):12120-12125.
42. Schwecke T, Gottlin, K., Durek, P, Duenas, I, Kaufer, N, Zock-Emmenthal, S., Staub, E., Neuhof, T, Dieckmann, R., and von Dohren, H.: **Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi.** *ChemBioChem* 2006, **7**(4):1-12.

43. Walton JD, Panaccione, Daniel G., and Hallen, Heather, E.: **Peptide synthesis without ribosomes**. In: *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine*. Edited by Lange JSTaL. New York, New York: Kluwer Academic/Plenum Publishers; 2004.
44. Mercer AC, Burkart MD: **The ubiquitous carrier protein - a window to metabolite biosynthesis**. *Natural Product Reports* 2007, **24**(4):750-773.
45. Stein T, Vater J, Kruft V, Otto A, WittmannLiebold B, Franke P, Panico M, McDowell R, Morris HR: **The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates**. *Journal of Biological Chemistry* 1996, **271**(26):15428-15435.
46. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT: **A new enzyme superfamily - The phosphopantetheinyl transferases**. *Chemistry & Biology* 1996, **3**(11):923-936.
47. Ehmann DE, Gehring AM, Walsh CT: **Lysine biosynthesis in *Saccharomyces cerevisiae*: Mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5**. *Biochemistry* 1999, **38**(19):6171-6177.
48. Mootz HD, Schorgendorfer K, Marahiel MA: **Functional characterization of 4'-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of *Saccharomyces cerevisiae* lys5**. *Fems Microbiology Letters* 2002, **213**(1):51-57.
49. Oide S: **Functional characterization of nonribosomal peptide synthetases in the filamentous ascomycete phytopathogen *Cochliobolus heterostrophus***. PhD. Ithaca, NY: Cornell University; 2007.
50. Keszenman-Pereyra D, Lawrence S, Twfieg ME, Price J, Turner G: **The *npgA/cfwA* gene encodes a putative 4'-phosphopantetheinyl transferase which is essential for penicillin biosynthesis in *Aspergillus nidulans***. *Current Genetics* 2003, **43**(3):186-190.

51. Aguirre J, Ortiz R, Clutterbuck J, Tapia R, Cardenas M: **vegA and cfwA define two new developmental genes in *Aspergillus nidulans***. *Fungal Genetics Newsletter* 1993, **40a**:68.
52. Han YJ, Han DM: **Isolation and characterization of null pigment mutant in *Aspergillus nidulans***. *Korean Journal of Genetics* 1993, **15**:1-10.
53. Keating TA, Marshall CG, Walsh CT, Keating AE: **The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains**. *Nature Structural Biology* 2002, **9**(7):522-526.
54. Keating TA, Marshall CG, Walsh CT: **Vibriobactin biosynthesis in *Vibrio cholerae*: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains**. *Biochemistry* 2000, **39**(50):15513-15521.
55. Belshaw PJ, Walsh CT, Stachelhaus T: **Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis**. *Science* 1999, **284**(5413):486-489.
56. Doekel S, Marahiel MA: **Dipeptide formation on engineered hybrid peptide synthetases**. *Chemistry & Biology* 2000, **7**(6):373-384.
57. Mootz HD, Schwarzer D, Marahiel MA: **Construction of hybrid peptide synthetases by module and domain fusions**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(11):5848-5853.
58. Stein DB, Linne U, Hahn M, Marahiel MA: **Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis**. *Chembiochem* 2006, **7**(11):1807-1814.
59. Ehmman DE, Trauger JW, Stachelhaus T, Walsh CT: **Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases**. *Chemistry & Biology* 2000, **7**(10):765-772.

60. Konz D, Klens A, Schorgendorfer K, Marahiel MA: **The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases.** *Chemistry & Biology* 1997, **4**(12):927-937.
61. Keating TA, Ehmann DE, Kohli RM, Marshall CG, Trauger JW, Walsh CT: **Chain termination steps in nonribosomal peptide synthetase assembly lines: Directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis.** *Chembiochem* 2001, **2**(2):99-107.
62. Koglin A, Loehr F, Mofid M, Mittag T, Schaefer B, Marahiel M, Doetsch V, Bernhard F: **Three-dimensional structure and molecular interactions of the Surfactin Thioesterase type II from *Bacillus subtilis*.** *Faseb Journal* 2004, **18**(8):C280-C280.
63. Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT: **Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE.** *Structure* 2002, **10**(3):301-310.
64. Kallow W, Kennedy J, Arezi B, Turner G, von Dohren H: **Thioesterase domain of delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase: Alteration of stereospecificity by site-directed mutagenesis.** *Journal of Molecular Biology* 2000, **297**(2):395-408.
65. Pieper R, Haese A, Schroder W, Zocher R: **Arrangement of catalytic sites in the multifunctional enzyme Enniatin synthetase.** *European Journal of Biochemistry* 1995, **230**(1):119-126.
66. Scotterraig JS, Panaccione DG, Pocard JA, Walton JD: **The cyclic peptide synthetase catalyzing HC-Toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame.** *Journal of Biological Chemistry* 1992, **267**(36):26044-26049.
67. von Dohren H: **Biochemistry and general genetics of nonribosomal peptidic synthetases in fungi.** In: *Molecular Biotechnology of Fungal Beta-Lactam Antibiotics and Related Peptide Synthetases.* vol. 88; 2004: 217-264.

68. Billman-Jacobe H, McConville MJ, Haites RE, Kovacevic S, Coppel RL: **Identification of a peptide synthetase involved in the biosynthesis of glycopeptidolipids of *Mycobacterium smegmatis***. *Molecular Microbiology* 1999, **33**(6):1244-1253.
69. Walzel B, Riederer B, Keller U: **Mechanism of alkaloid cyclopeptide synthesis in the ergot fungus *Claviceps purpurea***. *Chemistry & Biology* 1997, **4**(3):223-230.
70. Pfeifer E, Pavelavrancic M, Vondohren H, Kleinkauf H: **Characterization of Tyrocidine synthetase-1 (Ty1) - Requirement of posttranslational modification for peptide biosynthesis**. *Biochemistry* 1995, **34**(22):7450-7459.
71. Billich A, Zocher R: **N-methyltransferase function of the multifunctional enzyme Enniatin synthetase**. *Biochemistry* 1987, **26**(25):8417-8423.
72. Zocher R, Madry N, Peeters H, Kleinkauf H: **Biosynthesis of Cyclosporin-A**. *Phytochemistry* 1984, **23**(3):549-551.
73. Hahn M, Stachelhaus T: **Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(44):15585-15590.
74. Silakowski B, Kunze B, Muller R: **Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca***. *Gene* 2001, **275**(2):233-240.
75. Huang GZ, Zhang LH, Birch RG: **Multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans***. *Microbiology-Sgm* 2001, **147**:631-642.
76. Silakowski B, Nordsiek G, Kunze B, Blocker H, Muller R: **Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15**. *Chemistry & Biology* 2001, **8**(1):59-69.

77. Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F, Reinhardt R, Schmidt M, Ullrich C *et al*: **The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: A multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(23):13294-13299.
78. Austin MB, Saito T, Bowman ME, Haydock S, Kato A, Moore BS, Kay RR, Noel JP: **Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid - type III polyketide synthase.** *Nature Chemical Biology* 2006, **2**(9):494-502.
79. Mahanti N, Bhatnagar D, Cary JW, Joubran J, Linz JE: **Structure and function of fas-1A, a gene encoding a putative fatty acid synthetase directly involved in aflatoxin biosynthesis in *Aspergillus parasiticus*.** *Applied and Environmental Microbiology* 1996, **62**(1):191-195.
80. Brown DW, Yu JH, Kelkar HS, Fernandes M, Nesbitt TC, Keller NP, Adams TH, Leonard TJ: **Twenty-five coregulated transcripts define a sterigmatocystin gene cluster in *Aspergillus nidulans*.** *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(4):1418-1422.
81. Metz JG, Roessler P, Facciotti D, Levering C, Dittrich F, Lassner M, Valentine R, Lardizabal K, Domergue F, Yamada A *et al*: **Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes.** *Science* 2001, **293**(5528):290-293.
82. Lipmann F, Hotchkiss RD, Dubos RJ: **The occurrence of d-amino acids in Gramicidin and Tyrocidine.** *Journal of Biological Chemistry* 1941, **141**(1):163-169.
83. Lipmann F, Gevers W, Kleinkau.H, Roskoski R: **Polypeptide synthesis on protein templates - Enzymatic synthesis of Gramicidin-S and Tyrocidine.** *Advances in Enzymology and Related Areas of Molecular Biology* 1971, **35**:1-34.

84. Gevers W, Kleinkau.H, Lipmann F: **Peptidyl transfers in Gramicidin S biosynthesis from enzyme-bound thioester intermediates.** *Proceedings of the National Academy of Sciences of the United States of America* 1969, **63**(4):1335-1342.
85. Gevers W, Kleinkau.H, Lipmann F: **Activation of amino acids for biosynthesis of Gramicidin S.** *Proceedings of the National Academy of Sciences of the United States of America* 1968, **60**(1):269-276.
86. McElroy WD, Deluca M, Travis J: **Molecular uniformity in biological catalyses.** *Science* 1967, **157**(3785):150-160.
87. Smith DJ, Earl AJ, Turner G: **The multifunctional peptide synthetase performing the 1st Step of Penicillin biosynthesis in *Penicillium-Chrysogenum* Is a 421-073 dalton protein similar to *Bacillus-Brevis* peptide antibiotic synthetases.** *Embo Journal* 1990, **9**(9):2743-2750.
88. Schroder J: **Protein-sequence homology between plant 4-coumarate - Coa ligase and firefly luciferase.** *Nucleic Acids Research* 1989, **17**(1):460-460.
89. Babbitt PC, Kenyon GL, Martin BM, Charest H, Slyvestre M, Scholten JD, Chang KH, Liang PH, Dunawaymariano D: **Ancestry of the 4-chlorobenzoate dehalogenase - Analysis of amino-acid-sequence identities among families of acyl-adenyl ligases, enoyl-CoA hydratases isomerases, and acyl-CoA thioesterases.** *Biochemistry* 1992, **31**(24):5594-5604.
90. Weber T, Marahiel MA: **Exploring the domain structure of modular nonribosomal peptide synthetases.** *Structure* 2001, **9**(1):R3-R9.
91. Mallonee DH, White WB, Hylemon PB: **Cloning and sequencing of a bile acid-inducible operon from eubacterium sp. strain-Vpi-12708.** *Journal of Bacteriology* 1990, **172**(12):7011-7019.
92. Rusnak F, Sakaitani M, Drueckhammer D, Reichert J, Walsh CT: **Biosynthesis of the *Escherichia-coli* siderophore Enterobactin - Sequence of the EntF gene, expression and purification of EntF, and analysis of covalent phosphopantetheine.** *Biochemistry* 1991, **30**(11):2916-2927.

93. Scholten JD, Chang KH, Babbitt PC, Charest H, Sylvestre M, Dunawaymariano D: **Novel enzymatic hydrolytic dehalogenation of a chlorinated aromatic.** *Science* 1991, **253**(5016):182-185.
94. Knobloch KH, Hahlbrock K: **4-coumarate - CoA ligase from cell-suspension cultures of *Petroselinum-hortense*-Hoffm - Partial-purification, substrate-specificity, and further properties.** *Archives of Biochemistry and Biophysics* 1977, **184**(1):237-248.
95. Zhao Y, Kung SD, Dube SK: **Nucleotide-sequence of rice 4-coumarate-CoA ligase gene, 4-Cl.1.** *Nucleic Acids Research* 1990, **18**(20):6144-6144.
96. Suzuki H, Yamamoto T: **The long-chain acyl-CoA synthetase - Structure and regulation.** *Annals of the New York Academy of Sciences* 1990, **598**:560-560.
97. Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC: **The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A.** *Biochemistry* 2003, **42**(10):2866-2873.
98. Shockey JM, Fulda MS, Browse J: **Arabidopsis contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases.** *Plant Physiology* 2003, **132**(2):1065-1076.
99. Lu SW, Kroken S, Lee BN, Robbertse B, Churchill ACL, Yoder OC, Turgeon BG: **A novel class of gene controlling virulence in plant pathogenic ascomycete fungi.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(10):5980-5985.
100. Sinha AK, Bhattach.Jk: **Lysine biosynthesis in *Saccharomyces* - conversion of alpha-aminoadipate into alpha-aminoadipic delta-semialdehyde.** *Biochemical Journal* 1971, **125**(3):743-&.
101. Kasahara T, Kato T: **A new redox-cofactor vitamin for mammals.** *Nature* 2003, **422**(6934):832-832.

102. Richardt A, Kemme T, Wagner S, Schwarzer D, Marahiel MA, Hovemann BT: **Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in *Drosophila*.** *Journal of Biological Chemistry* 2003, **278**(42):41160-41166.
103. Farrell DH, Mikesell P, Actis LA, Crosa JH: **A regulatory gene, AngR, of the iron uptake System of *Vibrio-anguillarum* - Similarity with phage-P22 Cro and regulation by iron.** *Gene* 1990, **86**(1):45-51.
104. Gulick AM, Lu XF, Dunaway-Mariano D: **Crystal structure of 4-chlorobenzoate : CoA ligase/synthetase in the unliganded and aryl substrate-bound states.** *Biochemistry* 2004, **43**(27):8670-8679.
105. Nei M, Rooney AP: **Concerted and birth-and-death evolution of multigene families.** *Annual Review of Genetics* 2005, **39**:121-152.
106. Nei M, Hughes, A.L.: **Balanced polymorphism and evolution by the birth-and-death process in the MHC loci.** In: *11th Histocompatibility Workshop and Conference: 1992*: Oxford Univ. Press; 1992.
107. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system.** *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(15):7799-7806.
108. Hao L, Nei M: **Genomic organization and evolutionary analysis of Ly49 genes encoding the rodent natural killer cell receptors: rapid evolution by repeated gene duplication.** *Immunogenetics* 2004, **56**(5):343-354.
109. Ota T, Nei M: **Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V-H gene family.** *Molecular Biology and Evolution* 1994, **11**(3):469-482.
110. Nozawa M, Nei M: **Evolutionary dynamics of olfactory receptor genes in *Drosophila* species.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(17):7122-7127.

111. Niimura Y, Nei M: **Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates.** *Journal of Human Genetics* 2006, **51**(6):505-517.
112. Michelmore RW, Meyers BC: **Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process.** *Genome Research* 1998, **8**(11):1113-1130.
113. Piontkivska H, Rooney AP, Nei M: **Purifying selection and birth-and-death evolution in the histone H4 gene family.** *Molecular Biology and Evolution* 2002, **19**(5):689-697.
114. Nei M, Rogozin IB, Piontkivska H: **Purifying selection and birth-and-death evolution in the ubiquitin gene family.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(20):10866-10871.
115. McLachlan AD: **Repeating sequences and gene duplication in proteins.** *Journal of Molecular Biology* 1972, **64**(2):417-437.
116. Barney BM: **Classification of proteins based on minimal modular repeats: Lessons from nature in protein design.** *Journal of Proteome Research* 2006, **5**(3):473-482.
117. Soding J, Lupas AN: **More than the sum of their parts: on the evolution of proteins from peptides.** *Bioessays* 2003, **25**(9):837-846.
118. Riley M, Labedan B: **Protein evolution viewed through *Escherichia coli* protein sequences: Introducing the notion of a structural segment of homology, the module.** *Journal of Molecular Biology* 1997, **268**(5):857-868.
119. Johannesson H, Townsend JP, Hung CY, Cole GT, Taylor JW: **Concerted evolution in the repeats of an immunomodulating cell surface protein, SOWgp, of the human pathogenic fungi *Coccidioides immitis* and *C. posadasii*.** *Genetics* 2005, **171**(1):109-117.

120. Meeds T, Lockard E, Livingston BT: **Special evolutionary properties of genes encoding a protein with a simple amino acid repeat.** *Journal of Molecular Evolution* 2001, **53**(3):180-190.
121. Galindo BE, Vacquier VD, Swanson WJ: **Positive selection in the egg receptor for abalone sperm lysin.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(8):4639-4643.
122. Paoletti M, Saupe, S.J., Clave, C.: **Genesis of a fungal non-self recognition repertoire.** *PloS ONE* 2007, **2**:e283.
123. Tordai H, Nagy A, Farkas K, Banyai L, Patthy L: **Modules, multidomain proteins and organismic complexity.** *Febs Journal* 2005, **272**(19):5064-5078.
124. Patthy L: **Modular assembly of genes and the evolution of new functions.** *Genetica* 2003, **118**(2-3):217-231.
125. Pasek S, Risler JL, Brezellec P: **Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins.** *Bioinformatics* 2006, **22**(12):1418-1423.
126. Kummerfeld SK, Teichmann SA: **Relative rates of gene fusion and fission in multi-domain proteins.** *Trends in Genetics* 2005, **21**(1):25-30.
127. Yanai I, Wolf YI, and Koonin EV: **Evolution of gene fusions: horizontal transfer versus independent events.** *Genome Biology* 2002, **3**(5):0024.0021-0024.0023.
128. Weiner J, Bornberg-Bauer E: **Evolution of circular permutations in multidomain proteins.** *Molecular Biology and Evolution* 2006, **23**(4):734-743.
129. Weiner J, Beaussart F, Bornberg-Bauer E: **Domain deletions and substitutions in the modular protein evolution.** *Febs Journal* 2006, **273**(9):2037-2047.

130. Ponting CP, Russell RB: **Swaposins - circular permutations within genes encoding Saposin homologs.** *Trends in Biochemical Sciences* 1995, **20**(5):179-180.
131. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA: **Structure, function and evolution of multidomain proteins.** *Current Opinion in Structural Biology* 2004, **14**(2):208-216.
132. Patthy L: **Genome evolution and the evolution of exon-shuffling - a review.** *Gene* 1999, **238**(1):103-114.
133. Vogel C, Teichmann SA, Pereira-Leal J: **The relationship between domain duplication and recombination.** *Journal of Molecular Biology* 2005, **346**(1):355-365.
134. Fliess A, Motro B, Unger R: **Swaps in protein sequences.** *Proteins-Structure Function and Genetics* 2002, **48**(2):377-387.
135. Patthy L: **Exon shuffling and other ways of module exchange.** *Matrix Biology* 1996, **15**(5):301-310.
136. De Souza SJ, Long M, Kleln RJ, Roy S, Lin S, Gilbert W: **Toward a resolution of the introns early/late debate: Only phase zero introns are correlated with the structure of ancient proteins.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(9):5094-5099.
137. Long MY, De Souza SJ, Rosenberg C, Gilbert WE: **Relationship between "proto-splice sites" and intron phases: Evidence from dicodon analysis.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(1):219-223.
138. Stoltzfus A, Spencer DF, Zuker M, Logsdon JM, Doolittle WF: **Testing the exon theory of genes - the evidence from protein-structure.** *Science* 1994, **265**(5169):202-207.

139. Gokhale RS, Tsuji SY, Cane DE, Khosla C: **Dissecting and exploiting intermodular communication in polyketide synthases.** *Science* 1999, **284**(5413):482-485.
140. Jenke-Kodama H, Borner T, Dittmann E: **Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*.** *Plos Computational Biology* 2006, **2**(10):1210-1218.
141. Tanabe Y, Kaya K, Watanabe MM: **Evidence for recombination in the microcystin synthetase (mcy) genes of toxic cyanobacteria *Microcystiis spp.*** *Journal of Molecular Evolution* 2004, **58**(6):633-641.
142. Fewer DP, Rouhiainen L, Jokela J, Wahlsten M, Laakso K, Wang H, Sivonen K: **Recurrent adenylation domain replacement in the microcystin synthetase gene cluster.** *Bmc Evolutionary Biology* 2007, **7**.
143. Morgenstern B, Atchley WR: **Evolution of bHLH transcription factors: Modular evolution by domain shuffling?** *Molecular Biology and Evolution* 1999, **16**(12):1654-1663.
144. Keller NP, Hohn TM: **Metabolic pathway gene clusters in filamentous fungi.** *Fungal Genetics and Biology* 1997, **21**(1):17-29.
145. Shwab EK, Bok JW, Tribus M, Galehr J, Graessle S, Keller NP: **Histone deacetylase activity regulates chemical diversity in *Aspergillus*.** *Eukaryotic Cell* 2007, **6**(9):1656-1664.
146. Bok JW, Keller NP: **LaeA, a regulator of secondary metabolism in *Aspergillus spp.*** *Eukaryotic Cell* 2004, **3**(2):527-535.
147. Gardiner DM, Howlett BJ: **Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*.** *Fems Microbiology Letters* 2005, **248**(2):241-248.

148. Walton JD: **Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: An hypothesis.** *Fungal Genetics and Biology* 2000, **30**(3):167-171.
149. Pepper JW: **The evolution of evolvability in genetic linkage patterns.** *Biosystems* 2003, **69**(2-3):115-126.
150. Miyashita NT, Aguade M, Langley CH: **Linkage Disequilibrium in the White Locus Region of *Drosophila-melanogaster*.** *Genetical Research* 1993, **62**(2):101-109.
151. Stahl FW, Murray NE: **The evolution of gene clusters and genetic circularity in microorganisms.** *Genetics* 1966, **53**:569-576.
152. Fisher RA: **The genetical theory of natural selection.** Oxford: Clarendon Press; 1930.
153. Lawrence J: **Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes.** *Current Opinion in Genetics & Development* 1999, **9**(6):642-648.
154. Slot JC, Hallstrom KN, Matheny PB, Hibbett DS: **Diversification of NRT2 and the origin of its fungal homolog.** *Molecular Biology and Evolution* 2007, **24**(8):1731-1743.
155. Khaldi N, Collemare J, Lebrun MH, Wolfe KH: **Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi.** *Genome Biology* 2008, **9**(1):R18.
156. Patron NJ, Waller RF, Cozijnsen AJ, Straney DC, Gardiner DM, Nierman WC, Howlett BJ: **Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes.** *BMC Evolutionary Biology* 2007, **7**:174.

157. Hall C, Dietrich FS: **The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering.** *Genetics* 2007, **177**(4):2293-2307.
158. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O *et al*: **Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*.** *Nature* 2006, **444**(7115):97-101.
159. Wong S, Wolfe KH: **Birth of a metabolic gene cluster in yeast by adaptive gene relocation.** *Nature Genetics* 2005, **37**(7):777-782.
160. Cary JW, Ehrlich KC: **Aflatoxigenicity in *Aspergillus*: molecular genetics, phylogenetic relationships and evolutionary implications.** *Mycopathologia* 2006, **162**(3):167-177.
161. Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG: **Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(26):15670-15675.
162. Lee BN, Kroken S, Chou DYT, Robbertse B, Yoder OC, Turgeon BG: **Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress.** *Eukaryotic Cell* 2005, **4**(3):545-555.
163. Wenzl P, Wong L, Kwang-Won K, Jefferson RA: **A functional screen identifies lateral transfer of beta-glucuronidase (*gus*) from bacteria to fungi.** *Molecular Biology and Evolution* 2005, **22**(2):308-316.
164. Temporini ED, VanEtten HD: **An analysis of the phylogenetic distribution of the pea pathogenicity 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* 2004, **46**(1):29-36.

165. Friesen TL, Stukenbrock EH, Liu ZH, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP: **Emergence of a new disease as a result of interspecific virulence gene transfer.** *Nature Genetics* 2006, **38**(8):953-956.
166. Kavanaugh LA, Fraser JA, Dietrich FS: **Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment.** *Molecular Biology and Evolution* 2006, **23**(10):1879-1890.
167. Rosewich UL, Kistler HC: **Role of horizontal gene transfer in the evolution of fungi.** *Annual Review of Phytopathology* 2000, **38**:325-+.
168. Khaldi N, and Wolfe K: **Elusive origins of the extra genes in *Aspergillus oryzae*.** *Plos One* 2008, **3**(8):e3036 doi:3010.1371.
169. Carbone I, Ramirez-Prado JH, Jakobek JL, Horn BW: **Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster.** *BMC Evolutionary Biology* 2007, **7**:111.
170. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A *et al*: **Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*.** *Plos Genetics* 2008, **4**(4).
171. Carbone I, Jakobek JL, Ramirez-Prado JH, Horn BW: **Recombination, balancing selection and adaptive evolution in the aflatoxin gene cluster of *Aspergillus parasiticus*.** *Molecular Ecology* 2007, **16**(20):4401-4417.
172. Chang PK, Horn BW, Dorner JW: **Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic *Aspergillus flavus* isolates.** *Fungal Genetics and Biology* 2005, **42**(11):914-923.
173. Cuomo CA, Gueldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M *et al*: **The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization.** *Science* 2007, **317**(5843):1400-1402.

174. Turgeon BG, and Bushley KE: **Secondary Metabolism**. 2009. *In Cellular and Molecular Biology of Filamentous Fungi*. Eds. Borkovich, K. and Ebbole, D. ASM Press. In Press.
175. Bethke LL, Zilversmit M, Nielsen K, Daily J, Volkman SK, Ndiaye D, Lozovsky ER, Hartl DL, Wirth DF: **Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of *Plasmodium falciparum***. *Molecular and Biochemical Parasitology* 2006, **150**(1):10-24.
176. Johnson RD, Johnson L, Itoh Y, Kodama M, Otani H, Kahmoto K: **Cloning and characterization of a cyclic peptide synthetase gene from *Alternaria alternata* apple pathotype whose product is involved in AM-toxin synthesis and pathogenicity**. *Molecular Plant-Microbe Interactions* 2000, **13**(7):742-753.
177. Oide S, Moeder W, Krasnoff S, Gibson D, Haas H, Yoshioka K, Turgeon BG: **NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes**. *Plant Cell* 2006, **18**(10):2836-2853.
178. Carrano CJ, Bailey CT, Bonadies JA: **Transport-properties of N-acyl derivatives of the coprogen and ferrichrysin classes of siderophores in *Neurospora-Crassa***. *Archives of Microbiology* 1986, **146**(1):41-45.
179. Chung TDY, Matzanke BF, Winkelmann G, Raymond KN: **Coordination chemistry of microbial iron transport compounds .33. inhibitory effect of the partially resolved coordination isomers of chromic desferricoprogen on coprogen uptake in *Neurospora-Crassa***. *Journal of Bacteriology* 1986, **165**(1):283-287.
180. Ernst J, Winkelma.G: **Metabolic Products of Microorganisms .135. Uptake of Iron by *Neurospora-Crassa* .4. Iron Transport Properties of Semisynthetic Coprogen Derivatives**. *Archives of Microbiology* 1974, **100**(3):271-282.

181. Winkelma.G, Barnekow A, Ilgner D, Zahner H: **Metabolic products of microorganisms .120. Uptake of iron by *Neurospora-crassa* .2. Regulation of biosynthesis of sideramines and inhibition of iron transport by metal analogs of coprogen.** *Archiv Fur Mikrobiologie* 1973, **92**(4):285-300.
182. Hof C, Eisfeld K, Antelo L, Foster AJ, Anke H: **Siderophore synthesis in *Magnaporthe grisea* is essential for vegetative growth, conidiation and resistance to oxidative stress.** *Fungal Genetics and Biology* 2009, **46**(4):321-332.
183. Dieckmann H, Krezdorn E: **Metabolic products of microorganisms. 150. Ferricrocin, triacetylfusigen and other sideramines from fungi of genus *Aspergillus*, group *Fumigatus*.** *Arch Microbiol* 1975, **106**:191-194.
184. Idnurm A, Taylor JL, Pedras MSC, Howlett BJ: **Small scale functional genomics of the blackleg fungus, *Leptosphaeria maculans*: analysis of a 38 kb region.** *Australasian Plant Pathology* 2003, **32**(4):511-519.
185. Kim KH, Cho Y, La Rota M, Cramer RA, Lawrence CB: **Functional analysis of the *Alternaria brassicicola* non-ribosomal peptide synthetase gene AbNPS2 reveals a role in conidial cell wall construction.** *Molecular Plant Pathology* 2007, **8**(1):23-39.
186. Oide S, Krasnoff SB, Gibson DM, Turgeon BG: **Intracellular siderophores are essential for ascomycete sexual development in heterothallic *Cochliobolus heterostrophus* and homothallic *Gibberella zeae*.** *Eukaryotic Cell* 2007, **6**(8):1339-1353.
187. Eisendle M, Oberegger H, Zadra I, Haas H: **The siderophore system is essential for viability of *Aspergillus nidulans*: functional analysis of two genes encoding L-ornithine N-5-monooxygenase (sidA) and a non-ribosomal peptide synthetase (sidC).** *Molecular Microbiology* 2003, **49**(2):359-375.
188. Horowitz NG, Charlang, G., Horn, G. and Williams, N. : **Isolation and identification of the conidial germination factor of *Neurospora crassa*.** *J Bacteriology* 1976, **127**:135-140.

189. Matzanke BF: **Iron Storage in Fungi**. In: *Metal Ions in Fungi*. Edited by Winklemann GaW, Dennis, vol. 11. New York, New York: Marcel Dekker Inc; 1994: 179-214.
190. Matzanke BF, Bill, E., Trautwein, A., and Winklemann, G.: **Role of Siderophores in Iron Storage in Spores of *Neurospora crassa* and *Aspergillus ochraceus***. *Journal of Bacteriology* 1987, **169**(12):5873-5876.
191. Eisendle M, Schrettl M, Kragl C, Muller D, Illmer P, Haas H: **The intracellular siderophore ferricrocin is involved in iron storage, oxidative-stress resistance, germination, and sexual development in *Aspergillus nidulans***. *Eukaryotic Cell* 2006, **5**(10):1596-1603.
192. Laich F, Fierro F, Martin JF: **Production of penicillin by fungi growing on food products: Identification of a complete penicillin gene cluster in *Penicillium griseofulvum* and a truncated cluster in *Penicillium verrucosum***. *Applied and Environmental Microbiology* 2002, **68**(3):1211-1219.
193. Laich F, Fierro F, Cardoza RE, Martin JF: **Organization of the gene cluster for biosynthesis of penicillin in *Penicillium nalgiovense* and antibiotic production in cured dry sausages**. *Applied and Environmental Microbiology* 1999, **65**(3):1236-1240.
194. Kim CF, Lee SKY, Price J, Jack RW, Turner G, Kong RYC: **Cloning and expression analysis of the pcbAB-pcbC beta-lactam genes in the marine fungus *Kallichroma tethys***. *Applied and Environmental Microbiology* 2003, **69**(2):1308-1314.
195. Aharonowitz Y, Cohen G, Martin JF: **Penicillin and cephalosporin biosynthetic genes - Structure, organization, regulation, and evolution**. *Annual Review of Microbiology* 1992, **46**:461-495.
196. Brakhage AA, Al-Abdallah Q, Tuncher A, Sprote P: **Evolution of beta-lactam biosynthesis genes and recruitment of trans-acting factors**. *Phytochemistry* 2005, **66**(11):1200-1210.

197. Liras P, Martin JF: **Gene clusters for beta-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate?** *International Microbiology* 2006, **9**(1):9-19.
198. Buades C, Moya A: **Phylogenetic analysis of the isopenicillin-N-synthetase horizontal gene transfer.** *Journal of Molecular Evolution* 1996, **42**(5):537-542.
199. Landan G, Cohen G, Aharonowitz Y, Shuali Y, Graur D, Shiffman D: **Evolution of isopenicillin-N synthase genes may have involved horizontal gene-transfer.** *Molecular Biology and Evolution* 1990, **7**(5):399-406.
200. Penalva MA, Moya A, Dopazo J, Ramon D: **Sequences of isopenicillin-N synthetase genes suggest horizontal gene-Transfer from prokaryotes to eukaryotes.** *Proceedings of the Royal Society of London Series B-Biological Sciences* 1990, **241**(1302):164-169.
201. Miller JR, Ingolia TD: **Cloning and characterization of beta-lactam biosynthetic genes.** *Molecular Microbiology* 1989, **3**(5):689-695.
202. Ramon D, Carramolino L, Patino C, Sanchez F, Penalva MA: **Cloning and characterization of the isopenicillin-N synthetase gene mediating the formation of the beta-lactam ring in *Aspergillus-nidulans*.** *Gene* 1987, **57**(2-3):171-181.
203. Ullan RV, Casqueiro J, Banuelos O, Fernandez FJ, Gutierrez S, Martin JF: **A novel epimerization system in fungal secondary metabolism involved in the conversion of isopenicillin N into penicillin N in *Acremonium chrysogenum*.** *Journal of Biological Chemistry* 2002, **277**(48):46216-46225.
204. Offenzeller M, Santer G, Totschnig K, Su Z, Moser H, Traber R, SchneiderScherzer E: **Biosynthesis of the unusual amino acid (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine of cyclosporin A: Enzymatic analysis of the reaction sequence including identification of the methylation precursor in a polyketide pathway.** *Biochemistry* 1996, **35**(25):8401-8412.

205. Martinez-Martinez S, Redondo JM: **Inhibitors of the calcineurin/NFAT pathway**. *Current Medicinal Chemistry* 2004, **11**(8):997-1007.
206. Ferreira A, Kincaid R, Kosik KS: **Calcineurin is associated with the cytoskeleton of cultured neurons and has a role in the acquisition of polarity**. *Molecular Biology of the Cell* 1993, **4**(12):1225-1238.
207. Cruz MC, Fox DS, Heitman J: **Calcineurin is required for hyphal elongation during mating and haploid fruiting in *Cryptococcus neoformans***. *Embo Journal* 2001, **20**(5):1020-1032.
208. Prokisch H, Yarden O, Dieminger M, Tropschug M, Barthelmess IB: **Impairment of calcineurin function in *Neurospora crassa* reveals its essential role in hyphal growth, morphology and maintenance of the apical Ca²⁺ gradient**. *Molecular & General Genetics* 1997, **256**(2):104-114.
209. Yoshida T, Toda T, Yanagida M: **A Calcineurin-Like Gene Ppb1(+) in Fission Yeast - Mutant Defects in Cytokinesis, Cell Polarity, Mating and Spindle Pole Body Positioning**. *Journal of Cell Science* 1994, **107**:1725-1735.
210. Cruz MC, Goldstein AL, Blankenship JR, Del Poeta M, Davis D, Cardenas ME, Perfect JR, McCusker JH, Heitman J: **Calcineurin is essential for survival during membrane stress in *Candida albicans***. *Embo Journal* 2002, **21**(4):546-559.
211. Fox DS, Cruz MC, Sia RAL, Ke HM, Cox GM, Cardenas ME, Heitman J: **Calcineurin regulatory subunit is essential for virulence and mediates interactions with FKBP12-FK506 in *Cryptococcus neoformans***. *Molecular Microbiology* 2001, **39**(4):835-849.
212. Velkov T, Lawen A: **Mapping and molecular modeling of S-adenosyl-L-methionine binding sites in N-methyltransferase domains of the multifunctional polypeptide Cyclosporin synthetase**. *Journal of Biological Chemistry* 2003, **278**(2):1137-1148.

213. Traber R, Hofmann H, Loosli HR, Ponelle M, Vonwartburg A: **Novel Cyclosporins from *Tolypocladium-inflatum* - the Cyclosporins-K-Z.** *Helvetica Chimica Acta* 1987, **70**(1):13-36.
214. Vonwartburg A, Traber R: **Chemistry of the Natural Cyclosporine Metabolites.** *Progress in Allergy* 1986, **38**:28-45.
215. Traber R, Dreyfuss MM: **Occurrence of cyclosporins and cyclosporin-like peptolides in fungi.** *Journal of Industrial Microbiology & Biotechnology* 1996, **17**(5-6):397-401.
216. Dreyfuss M, Harri E, Hofmann H, Kobel H, Pache W, Tschertter H: **Cyclosporin-a and C New Metabolites from *Trichoderma polysporum* (Link Ex Pers) Rifai.** *European Journal of Applied Microbiology* 1976, **3**(2):125-133.
217. Sedmera P, Havlicek V, Jegorov A, Segre AL: **Cyclosporine-D Hydroperoxide, a New Metabolite of *Tolypocladium-terricola*.** *Tetrahedron Letters* 1995, **36**(38):6953-6956.
218. Weiser J: **Acute toxicity of conidia of *Tolypocladium* fungi to larvae of *Culex-sitiens*.** *Acta Entomologica Bohemoslovaca* 1991, **88**(6):367-369.
219. Vilcinskas A, Jegorov A, Landa Z, Gotz P, Matha V: **Effects of Beauverolide L and Cyclosporin A on humoral and cellular immune response of the greater wax moth, *Galleria mellonella*.** *Comparative Biochemistry and Physiology C-Pharmacology Toxicology & Endocrinology* 1999, **122**(1):83-92.
220. Moussaif M, Jacques P, Schaarwachter P, Budzikiewicz H, Thonart P: **Cyclosporin C is the main antifungal compound produced by *Acremonium luzulae*.** *Applied and Environmental Microbiology* 1997, **63**(5):1739-1743.
221. Zocher R, Haese A: **Mechanism and molecular-structure of the multifunctional enzyme Enniatin synthetase.** *Abstracts of Papers of the American Chemical Society* 1992, **203**:18-BTEC.

222. Xu YQ, Orozco R, Wijeratne EMK, Gunatilaka AAL, Stock SP, Molnar I: **Biosynthesis of the cyclooligomer depsipeptide Beauvericin, a virulence factor of the entomopathogenic fungus *Beauveria bassiana***. *Chemistry & Biology* 2008, **15**(9):898-907.
223. Xu YQ, Rozco R, Wijeratne EMK, Espinosa-Artiles P, Gunatilaka AAL, Stock SP, Molnar I: **Biosynthesis of the cyclooligomer depsipeptide Bassianolide, an insecticidal virulence factor of *Beauveria bassiana***. *Fungal Genetics and Biology* 2009, **46**(5):353-364.
224. Grove JF, Pople M: **The insecticidal activity of Beauvericin and the Enniatin complex**. *Mycopathologia* 1980, **70**(2):103-105.
225. Mido N, Okakura K., Miyamoto, K. Watanabe, M. Yanai, K. Yasutake, T. Aihara, S., Futamura, T., Kleinkauf, H., Murakami, T.: **Cyclic depsipeptide synthetase and its gene and mass production system of cyclic depsipeptide**. Patent # WO 0118179-A1.
226. Haese A, Pieper R, Vonostrowski T, Zocher R: **Bacterial expression of catalytically active fragments of the multifunctional enzyme Enniatin synthetase**. *Journal of Molecular Biology* 1994, **243**(1):116-122.
227. Glinski M, Urbanke C, Hornbogen T, Zocher R: **Enniatin synthetase is a monomer with extended structure: evidence for an intramolecular reaction mechanism**. *Archives of Microbiology* 2002, **178**(4):267-273.
228. May JJ, Wendrich TM, Marahiel MA: **The dhb operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin**. *Journal of Biological Chemistry* 2001, **276**(10):7209-7217.
229. Kratzschmar J, Krause M, Marahiel MA: **Gramicidin-S biosynthesis operon containing the structural genes GrsA and GrsB has an open reading frame encoding a protein homologous to fatty-acid thioesterases**. *Journal of Bacteriology* 1989, **171**(10):5422-5429.

230. Tseng CC, Bruner SD, Kohli RM, Marahiel MA, Walsh CT, Sieber SA: **Characterization of the Surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase.** *Biochemistry* 2002, **41**(45):13350-13359.
231. Hoyer KM, Mahlert C, Marahiel MA: **The iterative Gramicidin S thioesterase catalyzes peptide ligation and cyclization.** *Chemistry & Biology* 2007, **14**(1):13-22.
232. Sieber SA, Linne U, Hillson NJ, Roche E, Walsh CT, Marahiel MA: **Evidence for a monomeric structure of nonribosomal peptide synthetases.** *Chemistry & Biology* 2002, **9**(9):997-1008.
233. Shaw-Reid CA, Kelleher NL, Losey HC, Gehring AM, Berg C, Walsh CT: **Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E-coli* EntF catalyzes both elongation and cyclolactonization.** *Chemistry & Biology* 1999, **6**(6):385-400.
234. Kohli RM, Trauger JW, Schwarzer D, Marahiel MA, Walsh CT: **Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases.** *Biochemistry* 2001, **40**(24):7099-7108.
235. Isaka M, Kittakoop P, Thebtaranonth Y: **Secondary metabolites of clavicipitalean fungi.** In: *Clavicipitalean Fungi: Evolutionary Biology, Chemistry, Biocontrol, and Cultural Impacts*. Edited by James F. White Jr. CWB, Nigel L. Hywel Jones, Joseph W. Spatafora. New York, New York: Marcel Dekker, Inc.; 2003.
236. Tudzynski P, Correia T, Keller U: **Biotechnology and genetics of ergot alkaloids.** *Applied Microbiology and Biotechnology* 2001, **57**(5-6):593-605.
237. Panaccione DG: **Origins and significance of ergot alkaloid diversity in fungi.** *Fems Microbiology Letters* 2005, **251**(1):9-17.

238. Clay K, Schardl C: **Evolutionary origins and ecological consequences of endophyte symbiosis with grasses.** *American Naturalist* 2002, **160**:S99-S127.
239. Panaccione DG, Tapper BA, Lane GA, Davies E, Fraser K: **Biochemical outcome of blocking the ergot alkaloid pathway of a grass endophyte.** *Journal of Agricultural and Food Chemistry* 2003, **51**(22):6429-6437.
240. Kozlovsky AG: **Producers of ergot alkaloids out of *Claviceps* genus.** In: *Ergot: The Genus Claviceps*. Edited by Kren VaC, L. Amsterdam, The Netherlands: Harwood Academic Publishers; 1999: 479-499.
241. Hofmann A, Tschertter H: **Die Wirkstoffe der mexikanischen Zauberdroge "Ololiuqui".** *Planta Med* 1960, **9**:354-367.
242. Ahimsa-Muller MA, Markert A, Hellwig S, Knoop V, Steiner U, Drewke C, Leistner E: **Clavicipitaceous fungi associated with ergoline alkaloid-containing Convolvulaceae.** *Journal of Natural Products* 2007, **70**(12):1955-1960.
243. Kucht S, Gross J, Hussein Y, Grothe T, Keller U, Basar S, Konig WA, Steiner U, Leistner E: **Elimination of ergoline alkaloids following treatment of *Ipomoea asarifolia* (Convolvulaceae) with fungicides.** *Planta* 2004, **219**(4):619-625.
244. Coyle CM, Panaccione DG: **An ergot alkaloid biosynthesis gene and clustered hypothetical genes from *Aspergillus fumigatus*.** *Applied and Environmental Microbiology* 2005, **71**(6):3112-3118.
245. Lorenz N, Wilson EV, Machado C, Schardl CL, Tudzynski P: **Comparison of ergot alkaloid biosynthesis gene clusters in *Claviceps* species indicates loss of late pathway steps in evolution of *C-fusiformis*.** *Applied and Environmental Microbiology* 2007, **73**(22):7185-7191.
246. Coyle CM, Goetz KE, Panaccione DG: **Clustered genes common to both *Aspergillus fumigatus* and ergot fungi control early steps within the ergot alkaloid pathway.** *Phytopathology* 2008, **98**(6):S214-S214.

247. Cross DL: **Ergot alkaloid toxicity**. In: *Clavicipitalean Fungi*. Edited by White JFJ, Bacon CW, Hywel-Jones NL, Spatafora JW. New York, New York: Marcel Dekker, Inc.; 2003: 475-494.
248. Hofmann A: **Historical view on ergot alkaloids**. *Pharmacology* 1978, **16**:1-11.
249. Hofmann A: **Die Mutterkornalkaloide**. Stuttgart: Enke; 1964.
250. Dyer DC: **Evidence that ergovaline acts on serotonin receptors**. *Life Sciences* 1993, **53**(14):PL223-PL228.
251. Houghton PJ, Howes MJ: **Natural products and derivatives affecting neurotransmission relevant to Alzheimer's and Parkinson's disease**. *Neurosignals* 2005, **14**(1-2):6-22.
252. Fiserova AaP, M.: **Role of ergot alkaloids in the immune system**. In: *The Genus Claviceps*. Edited by Kren V. CL. Amsterdam: Harwood; 1999: 451-467.
253. Bush LP, Wilkinson HH, Schardl CL: **Bioprotective alkaloids of grass-fungal endophyte symbioses**. *Plant Physiology* 1997, **114**(1):1-7.
254. Clay K, Cheplick GP: **Effect of ergot alkaloids from fungal endophyte-infected grasses on fall armyworm (*Spodoptera-frugiperda*)**. *Journal of Chemical Ecology* 1989, **15**(1):169-182.
255. Ball OJP, Miles CO, Prestidge RA: **Ergopeptine alkaloids and *Neotyphodium lolii*-mediated resistance in perennial ryegrass against adult *Heteronychus arator* (Coleoptera: Scarabaeidae)**. *Journal of Economic Entomology* 1997, **90**(5):1382-1391.
256. Davidson AW, Potter DA: **Response of plant-feeding, predatory, and soil-inhabiting invertebrates to *Acremonium* endophyte and nitrogen-fertilization in tall fescue turf**. *Journal of Economic Entomology* 1995, **88**(2):367-379.

257. Panaccione DG, Cipoletti JR, Sedlock AB, Blemings KP, Schardl CL, Machado C, Seidel GE: **Effects of ergot alkaloids on food preference and satiety in rabbits, as assessed with gene-knockout endophytes in perennial ryegrass (*Lolium perenne*)**. *Journal of Agricultural and Food Chemistry* 2006, **54**(13):4582-4587.
258. Bazely DR, Vicari M, Emmerich S, Filip L, Lin D, Inman A: **Interactions between herbivores and endophyte-infected *Festuca rubra* from the Scottish islands of St. Kilda, Benbecula and Rum**. *Journal of Applied Ecology* 1997, **34**(4):847-860.
259. Spatafora JW, Sung GH, Sung JM, Hywel-Jones NL, White JF: **Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes**. *Molecular Ecology* 2007, **16**(8):1701-1711.
260. Tanaka A, Tapper BA, Popay A, Parker EJ, Scott B: **A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory**. *Molecular Microbiology* 2005, **57**(4):1036-1050.
261. Benedetti E, Bavoso A, Diblasio B, Pavone V, Pedone C, Toniolo C, Bonora GM: **Peptaibol antibiotics - a study on the helical Structure of the 2-9 Sequence of Emerimicin-III and Emerimicin-IV**. *Proceedings of the National Academy of Sciences of the United States of America-Physical Sciences* 1982, **79**(24):7951-7954.
262. Bruckner H, Graf H: **Paracelsin, a peptide antibiotic containing alpha-aminoisobutyric-acid, isolated from *Trichoderma-reesei* Simmons .A**. *Experientia* 1983, **39**(5):528-530.
263. Kastin AJ: **Handbook of biologically active peptides**. Boston, MA: Academic Press; 2006.
264. Bechinger B: **Structure and function of membrane-lytic peptides**. *Critical Reviews in Plant Sciences* 2004, **23**(3):271-292.

265. Sansom MSP: **Structure and function of channel-forming peptaibols.** *Quarterly Reviews of Biophysics* 1993, **26**(4):365-421.
266. Duval D, Riddell FG, Rebuffat S, Platzer N, Bodo B: **Ionophoric activity of the antibiotic peptaibol Trichorzin PA VI: a Na-23- and Cl-35-NMR study.** *Biochimica Et Biophysica Acta-Biomembranes* 1998, **1372**(2):370-378.
267. Beven L, Duval D, Rebuffat S, Riddell FG, Bodo B, Wroblewski H: **Membrane permeabilisation and antimycoplasmic activity of the 18-residue peptaibols, Trichorzins PA.** *Biochimica Et Biophysica Acta-Biomembranes* 1998, **1372**(1):78-90.
268. Lorito M, Farkas V, Rebuffat S, Bodo B, Kubicek CP: **Cell wall synthesis is a major target of mycoparasitic antagonism by *Trichoderma harzianum*.** *Journal of Bacteriology* 1996, **178**(21):6382-6385.
269. Lorito M, Woo SL, Dambrosio M, Harman GE, Hayes CK, Kubicek CP, Scala F: **Synergistic interaction between cell wall degrading enzymes and membrane affecting compounds.** *Molecular Plant-Microbe Interactions* 1996, **9**(3):206-213.
270. Vinale F, Sivasithamparam K, Ghisalberti EL, Marra R, Barbetti MJ, Li H, Woo SL, Lorito M: **A novel role for *Trichoderma* secondary metabolites in the interactions with plants.** *Physiological and Molecular Plant Pathology* 2008, **72**(1-3):80-86.
271. Mullbacher A, Waring P, Eichner RD: **Identification of an agent in cultures of *Aspergillus fumigatus* displaying anti-phagocytic and immunomodulating activity in-vitro.** *Journal of General Microbiology* 1985, **131**(MAY):1251-1258.
272. Mullbacher A, Waring P, Tiwari-Paini U, Eichner RD: **Structural relationship of epipolythiodioxopiperazines and their immunomodulating activity.** *Molecular Immunology* 1986, **23**:231-236.

273. Gardiner DM, Cozijnsen AJ, Wilson LM, Pedras MSC, Howlett BJ: **The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans***. *Molecular Microbiology* 2004, **53**(5):1307-1318.
274. Rouxel T, Chupeau Y, Fritz R, Kollmann A, Bousquet JF: **Biological effects of sirodesmin-PL, a phytotoxin produced by *Leptosphaeria maculans***. *Plant Science* 1988, **57**(1):45-53.
275. Gardiner DM, Waring P, Howlett BJ: **The epipolythiodioxopiperazine (ETP) class of fungal toxins: distribution, mode of action, functions and biosynthesis**. *Microbiology-Sgm* 2005, **151**:1021-1032.
276. Kleinwachter P, Dahse HM, Luhmann U, Schlegel B, Dornberger K: **Epicorazine C, an antimicrobial metabolite from *Stereum hirsutum* HKI 0195**. *Journal of Antibiotics* 2001, **54**(6):521-525.
277. Stillwell MA, Magasi LP, Strunz GM: **Production, isolation, and antimicrobial activity of Hyalodendrin, a new antibiotic produced by a species of *Hyalodendron***. *Canadian Journal of Microbiology* 1974, **20**(5):759-&.
278. Moerman KL, Chai CLL, Waring P: **Evidence that the lichen-derived scabrosin esters target mitochondrial ATP synthase in P388D1 cells**. *Toxicology and Applied Pharmacology* 2003, **190**(3):232-240.
279. Shah DT, Larsen B: **Clinical isolates of yeast produce a Gliotoxin-like substance**. *Mycopathologia* 1991, **116**(3):203-208.
280. Waring P, Beaver J: **Gliotoxin and related epipolythiodioxopiperazines**. *General Pharmacology* 1996, **27**(8):1311-1316.
281. Waring P, Eichner RD, Mullbacher A, Sjaarda A: **Gliotoxin induces apoptosis in macrophages unrelated to its antiphagocytic properties**. *Journal of Biological Chemistry* 1988, **263**(34):18493-18499.

282. Kweon YO, Paik YH, Schnabl B, Qian T, Lemasters JJ, Brenner DA: **Gliotoxin-mediated apoptosis of activated human hepatic stellate cells.** *Journal of Hepatology* 2003, **39**(1):38-46.
283. Grovel O, Pouchus YF, Verbist JF: **Accumulation of gliotoxin, a cytotoxic mycotoxin from *Aspergillus fumigatus*, in blue mussel (*Mytilus edulis*).** *Toxicon* 2003, **42**(3):297-300.
284. Sutton P, Newcombe NR, Waring P, Mullbacher A: **In-vivo immunosuppressive activity of gliotoxin, a metabolite produced by human pathogenic fungi.** *Infection and Immunity* 1994, **62**(4):1192-1198.
285. Pahl HL, Krauss B, SchulzeOsthoff K, Decker T, Traenckner EBM, Vogt M, Myers C, Parks T, Warring P, Muhlbacher A *et al*: **The immunosuppressive fungal metabolite gliotoxin specifically inhibits transcription factor NF-kappa B.** *Journal of Experimental Medicine* 1996, **183**(4):1829-1840.
286. Rightsel WA, Ehrlich J, Dixon GJ, Miller FA, Schneide.Hg, Graf PR, Bartz QR, Sloan BJ: **Antiviral activity of Gliotoxin and Gliotoxin acetate.** *Nature* 1964, **204**(496):1333-1334.
287. Declercq E, Billiau A, Ottenheim HCJ, Herscheid JDM: **Anti-reverse transcriptase activity of Gliotoxin analogs.** *Biochemical Pharmacology* 1978, **27**(5):635-639.
288. Vigushin DM, Mirsaidi N, Brooke G, Sun C, Pace P, Inman L, Moody CJ, Coombes RC: **Gliotoxin is a dual inhibitor of farnesyltransferase and geranylgeranyltransferase I with antitumor activity against breast cancer in vivo.** *Medical Oncology* 2004, **21**(1):21-30.
289. Williams DE, Bombuwala K, Lobkovsky E, de Silva ED, Karunaratne V, Allen TM, Clardy J, Andersen RJ: **Ambewelamides A and B, antineoplastic epidithiapiperazinediones isolated from the lichen *Usnea sp.*** *Tetrahedron Letters* 1998, **39**(52):9579-9582.

290. Fox EM, Howlett BJ: **Biosynthetic gene clusters for epipolythiodioxopiperazines in filamentous fungi.** *Mycological Research* 2008, **112**:162-169.
291. Tucker SJ, Orr JG, Marek CJ, Haughton EL, Elrick LJ, Trim JE, Halestrap AP, Wright MC: **The role of the adenine nucleotide translocator (ANT) in apoptosis in response to Gliotoxin.** *Toxicology* 2005, **213**(3):251-251.
292. Orr JG, Leel V, Cameron GA, Marek CJ, Haughton EL, Elrick LJ, Trim JE, Hawksworth GM, Halestrap AP, Wright MC: **Mechanism of action of the antifibrogenic compound gliotoxin in rat liver cells.** *Hepatology* 2004, **40**(1):232-242.
293. Hurne AM, Chai CLL, Moerman K, Waring P: **Influx of calcium through a redox-sensitive plasma membrane channel in thymocytes causes early necrotic cell death induced by the epipolythiodioxopiperazine toxins.** *Journal of Biological Chemistry* 2002, **277**(35):31631-31638.
294. Trown PW, Bilello JA: **Mechanism of action of Gliotoxin - Elimination of activity by sulfhydryl compounds.** *Antimicrobial Agents and Chemotherapy* 1972, **2**(4):261-266.
295. Munday R: **Studies on the mechanism of toxicity of the mycotoxin, Sporidesmin .5. Generation of hydroxyl radical by Sporidesmin.** *Journal of Applied Toxicology* 1987, **7**(1):17-22.
296. Munday R: **Studies on the mechanism of toxicity of the mycotoxin Sporidesmin .2. Evidence for intracellular generation of superoxide radical from Sporidesmin.** *Journal of Applied Toxicology* 1984, **4**(4):176-181.
297. Munday R: **Studies on the mechanism of toxicity of the mycotoxin, Sporidesmin .1. Generation of superoxide radical by Sporidesmin.** *Chemico-Biological Interactions* 1982, **41**(3):361-374.

298. Waring P, Mamchak A, Khan T, Sjaarda A, Sutton P: **DNA-synthesis precedes Gliotoxin-induced apoptosis**. *Cell Death and Differentiation* 1995, **2**(3):201-210.
299. Pedras MSC, Seguinswartz G: **The Blackleg fungus - Phytotoxins and phytoalexins**. *Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie* 1992, **14**(1):67-75.
300. Elliott CE, Gardiner DM, Thomas G, Cozijnsen A, De Wouw AV, Howlett BJ: **Production of the toxin Sirodesmin PL by *Leptosphaeria maculans* during infection of *Brassica napus***. *Molecular Plant Pathology* 2007, **8**(6):791-802.
301. Towers NR: **Effect of zinc on toxicity of mycotoxin Sporidesmin to rat**. *Life Sciences* 1977, **20**(3):413-417.
302. Rouxel T, Kollmann A, Bousquet JF: **Zinc suppresses Sirodesmin PL toxicity and protects *Brassica napus* plants against the Blackleg disease caused by *Leptosphaeria-maculans***. *Plant Science* 1990, **68**(1):77-86.
303. Miller PA, Milstrey KP, Trown PW: **Specific inhibition of viral ribonucleic acid replication by Gliotoxin**. *Science* 1968, **159**(3813):431-&.
304. Walton JD: **Host-selective toxins: Agents of compatibility**. *Plant Cell* 1996, **8**(10):1723-1733.
305. Walton JD, Panaccione DG: **Host-selective toxins and disease specificity - Perspectives and progress**. *Annual Review of Phytopathology* 1993, **31**:275-303.
306. Wolpert TJ, Dunkle LD, Ciuffetti LM: **Host-selective toxins and avirulence determinants: What's in a name?** *Annual Review of Phytopathology* 2002, **40**:251-285.
307. Meeley RB, Walton JD: **Enzymatic detoxification of HC-Toxin, the host-selective cyclic peptide from *Cochliobolus-carbonum***. *Plant Physiology* 1991, **97**(3):1080-1086.

308. Meeley RB, Walton JD: **Molecular biology and biochemistry of Hm1, a maize gene for fungal resistance.** In: *Advances in Molecular Genetics of Plant-Microbe Interactions*. Edited by Nest EW, Verma, D.P.S. Dordrecht, Netherlands: Kluwver Academic Publishers; 1993: 463-467.
309. Walton JD: **HC-Toxin.** *Phytochemistry* 2006, **67**(14):1406-1413.
310. Ransom RF, Walton JD: **Histone hyperacetylation in maize in response to treatment with HC-Toxin or infection by the filamentous fungus *Cochliobolus carbonum*.** *Plant Physiology* 1997, **115**(3):1021-1027.
311. Brosch G, Ramsom R, Lechner T, Walton JD, Loidl P: **Inhibition of maize histone deacetylases by HC Toxin, the host-selective toxin of *Cochliobolus-carbonum*.** *Plant Cell* 1995, **7**(11):1941-1950.
312. Furumai R, Komatsu Y, Nishino N, Khochbin S, Yoshida M, Horinouchi S: **Potent histone deacetylase inhibitors built from trichostatin A and cyclic tetrapeptide antibiotics including trapoxin.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(1):87-92.
313. Kijima M, Yoshida M, Sugita K, Horinouchi S, Beppu T: **Trapoxin, an antitumor cyclic tetrapeptide, is an irreversible inhibitor of mammalian histone deacetylase.** *Journal of Biological Chemistry* 1993, **268**(30):22429-22435.
314. Yoshida H, Sugita K: **A novel tetracyclic peptide, Trapoxin, induces phenotypic change from transformed to normal in sis-oncogene-transformed nih3t3 cells.** *Japanese Journal of Cancer Research* 1992, **83**(4):324-328.
315. Itazaki H, Nagashima K, Sugita K, Yoshida H, Kawamura Y, Yasuda Y, Matsumoto K, Ishii K, Uotani N, Nakai H *et al*: **Isolation and structural elucidation of new cyclotetrapeptides, Trapoxin-A and Trapoxin-B, Having detransformation activities as antitumor agents.** *Journal of Antibiotics* 1990, **43**(12):1524-1532.

316. Otani H, Kohmoto K, Kodama M: **Alternaria toxins and their effects on host plants**. *Canadian Journal of Botany-Revue Canadienne de Botanique* 1995, **73**:S453-S458.
317. Okuno T, Ishita Y, Sawai K, Matsumoto T: **Characterization of Alternariolide - Host-specific toxin produced by *Alternaria-mali*-Roberts**. *Chemistry Letters* 1974(6):635-638.
318. Park P, Nishimura S, Kohmoto K, Otani H, Tsujimoto K: **2 Action sites of AM-Toxin-I produced by apple pathotype of *Alternaria-alternata* in host-cells - An ultrastructural study**. *Canadian Journal of Botany-Revue Canadienne De Botanique* 1981, **59**(3):301-310.
319. Park P, Tsuda M, Hayashi Y, Ueno T: **Effect of a host-specific toxin (AM-Toxin-I) produced by *Alternaria-mali*, an apple pathogen, on ultrastructure of plasma-membrane of cells in apple and Japanese pear leaves**. *Canadian Journal of Botany-Revue Canadienne De Botanique* 1977, **55**(18):2383-2393.
320. Wolpert TJ, Macko V, Acklin W, Jaun B, Seibl J, Meili J, Arigoni D: **Structure of Victorin-C, the major host-selective toxin from *Cochliobolus-victoriae***. *Experientia* 1985, **41**(12):1524-1529.
321. Wolpert TJ, Navarre DA, Lorang JM, Moore DL: **Molecular-interactions of Victorin and oats**. *Canadian Journal of Botany-Revue Canadienne De Botanique* 1995, **73**:S475-S482.
322. Sweat T, Wolpert T: **Characterization of Victorin-induced cell death in *Arabidopsis thaliana***. *Phytopathology* 2005, **95**(6):S101-S101.
323. Navarre DA, Wolpert TJ: **Victorin induction of an apoptotic/senescence-like response in oats**. *Plant Cell* 1999, **11**(2):237-249.
324. Litzenger SC: **Nature of susceptibility to *Helminthosporium victoriae* and resistance to *Puccinia coronata* in Victoria oats**. *Phytopathology* 1949, **39**:300-318.

325. Lorang JM, Carkaci-Salli N, Wolpert TJ: **Identification and characterization of Victorin sensitivity in *Arabidopsis thaliana***. *Molecular Plant-Microbe Interactions* 2004, **17**(6):577-582.
326. Lorang JM, Sweat TA, Wolpert TJ: **Plant disease susceptibility conferred by a "resistance" gene**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(37):14861-14866.
327. Sweat TA, Lorang JM, Bakker EG, Wolpert TJ: **Characterization of natural and induced variation in the LOV1 gene, a CC-NB-LRR gene conferring Victorin sensitivity and disease susceptibility in *Arabidopsis***. *Molecular Plant-Microbe Interactions* 2008, **21**(1):7-19.
328. Nagy ED, Lee TC, Ramakrishna W, Xu ZJ, Klein PE, SanMiguel P, Cheng CP, Li JL, Devos KM, Schertz K *et al*: **Fine mapping of the Pc locus of *Sorghum bicolor*, a gene controlling the reaction to a fungal pathogen and its host-selective toxin**. *Theoretical and Applied Genetics* 2007, **114**(6):961-970.
329. Song ZS, Cox RJ, Lazarus CM, Simpson TJ: **Fusarin C biosynthesis in *Fusarium moniliforme* and *Fusarium venenatum***. *Chembiochem* 2004, **5**(9):1196-1203.
330. Royles BJL: **Naturally-occurring tetramic acids - Structure, isolation, and synthesis**. *Chemical Reviews* 1995, **95**(6):1981-2001.
331. Casser I, Steffan B, Steglich W: **Fungal pigments .52. The chemistry of the plasmodial pigments of the slime-mold *Fuligo-septica* (Myxomycetes)**. *Angewandte Chemie-International Edition in English* 1987, **26**(6):586-587.
332. Sims JW, Fillmore JP, Warner DD, Schmidt EW: **Equisetin biosynthesis in *Fusarium heterosporum***. *Chemical Communications* 2005(2):186-188.
333. Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC, Hutchinson CR: **Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis**. *Science* 1999, **284**(5418):1368-1372.

334. Marahiel MA, Stachelhaus T, Mootz HD: **Modular peptide synthetases involved in nonribosomal peptide synthesis**. *Chemical Reviews* 1997, **97**(7):2651-2673.
335. Bohnert HU, Fudal I, Dioh W, Tharreau D, Notteghem JL, Lebrun MH: **A putative polyketide synthase peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice**. *Plant Cell* 2004, **16**(9):2499-2513.
336. Abe Y, Suzuki T, Ono C, Iwamoto K, Hosobuchi M, Yoshikawa H: **Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in *Penicillium citrinum***. *Molecular Genetics and Genomics* 2002, **267**(5):636-646.
337. Collemare J, Billard A, Boehnert HU, Lebrun MH: **Biosynthesis of secondary metabolites in the rice blast fungus *Magnaporthe grisea*: the role of hybrid PKS-NRPS in pathogenicity**. *Mycological Research* 2008, **112**:207-215.
338. Schmidt K, Riese U, Li ZZ, Hamburger M: **Novel tetramic acids and pyridone alkaloids, militarinones B, C, and D, from the insect pathogenic fungus *Paecilomyces militaris***. *Journal of Natural Products* 2003, **66**(3):378-383.
339. Cheng YX, Schneider B, Riese U, Schubert B, Li ZZ, Hamburger M: **Farinosones A-C, neurotrophic alkaloidal metabolites from the entomogenous deuteromycete *Paecilomyces farinosus***. *Journal of Natural Products* 2004, **67**(11):1854-1858.
340. Lebrun MH, Dutfoy F, Gaudemer F, Kunesch G, Gaudemer A: **Detection and quantification of the fungal phytotoxin tenuazonic acid produced by *Pyricularia-oryzae***. *Phytochemistry* 1990, **29**(12):3777-3783.
341. Tsurushima T, Don LD, Kawashima K, Murakami J, Nakayashiki H, Tosa Y, Mayama S: **Pyrichalasin H production and pathogenicity of *Digitaria*-specific isolates of *Pyricularia grisea***. *Molecular Plant Pathology* 2005, **6**(6):605-613.

CHAPTER 2

MODULE EVOLUTION AND SUBSTRATE SPECIFICITY OF FUNGAL NONRIBOSOMAL PEPTIDE SYNTHETASES INVOLVED IN SIDEROPHORE BIOSYNTHESIS^a

2.1 Abstract

Background: Most filamentous ascomycete fungi produce high affinity iron chelators called siderophores, biosynthesized nonribosomally by multimodular adenyating enzymes called nonribosomal peptide synthetases (NRPSs). While genes encoding the majority of NRPSs are intermittently distributed across the fungal kingdom, those encoding ferrichrome synthetase NRPSs, responsible for biosynthesis of ferrichrome siderophores, are conserved, which offers an opportunity to trace their evolution and the genesis of their multimodular domain architecture. Furthermore, since the chemistry of many ferrichromes is known, the biochemical and structural ‘rules’ guiding NRPS substrate choice can be addressed using protein structural modeling and evolutionary approaches.

Results: A search of forty-nine complete fungal genome sequences revealed that, with the exception of *Schizosaccharomyces pombe*, none of the yeast, chytrid, or zygomycete genomes contained a candidate ferrichrome synthetase. In contrast, all filamentous ascomycetes queried contained at least one, while presence and numbers in basidiomycetes varied. Genes encoding ferrichrome synthetases were monophyletic when analyzed with other NRPSs.

^aReprinted from published article: Bushley, K.E. and Turgeon, B.G. Module evolution and substrate specificity of fungal nonribosomal peptide synthetases involved in siderophore biosynthesis. BMC Evolutionary Biology, 2008, 8:328.

Phylogenetic analyses provided support for an ancestral duplication event resulting in two main lineages. They also supported the proposed hypothesis that ferrichrome synthetases derive from an ancestral hexamodular gene, likely created by tandem duplication of complete NRPS modules. Recurrent losses of individual domains or complete modules from this ancestral gene best explain the diversity of extant domain architectures observed. Key residues and regions in the adenylation domain pocket involved in substrate choice and for binding the amino and carboxy termini of the substrate were identified.

Conclusion: Iron-chelating ferrichrome synthetases appear restricted to fission yeast, filamentous ascomycetes, and basidiomycetes and fall into two main lineages. Phylogenetic analyses suggest that loss of domains or modules led to evolution of iterative biosynthetic mechanisms that allow flexibility in biosynthesis of the ferrichrome product. The 10 amino acid NRPS code, proposed earlier, failed when we tried to infer substrate preference. Instead, our analyses point to several regions of the binding pocket important in substrate choice and suggest that two positions of the code are involved in substrate anchoring, not substrate choice.

2.2 Background

Most filamentous ascomycete fungi produce high affinity iron chelator siderophores for scavenging environmental iron and for cellular sequestration of reactive iron [1]. All known fungal siderophores are synthesized by nonribosomal peptide synthetases (NRPSs) [2], large, usually multimodular enzymes that catalyze peptide bond formation independent of ribosomes. NRPS modules consist of three core domains, ordered 5' A-T-C 3': 1) an adenylation (A) domain responsible for

recognizing and activating a substrate molecule *via* adenylation with ATP, 2) a thiolation (T) domain which binds the substrate to the NRPS protein and 3) a condensation (C) domain which joins two substrates through a condensation reaction.

Although the number of NRPSs encoded by individual filamentous fungi varies from 0 to > 20, most of these and their corresponding metabolites are not conserved across the fungal kingdom, making it difficult to trace the evolutionary history of the corresponding genes. Various evolutionary processes may account for this. The observation that A-T-C modules from a single NRPS often group together as a monophyletic clade suggests tandem duplication of modules as a possible mechanism by which multimodular NRPSs arise [3]. It is clear, however, that other mechanisms such as recombination and gene conversion also operate [4]. Ferrichrome synthetases, which biosynthesize ferrichromes, fungal hydroxamate siderophores that function primarily in intracellular iron storage, are among the most conserved NRPS, offering an opportunity to trace the evolutionary history of the corresponding genes across fungi.

The chemical products of ferrichrome synthetases have been characterized for at least one member of the majority of Ascomycete and Basidiomycete orders [5, 6]. This class of siderophore includes compounds such as ferricrocin, ferrichrome, ferrichrome A, ferrichrome C, and malonichrome. Most ferrichrome siderophores are cyclic hexapeptides (Figure 2.1), with the exceptions of tetraglycylferrichrome, a cyclic heptapeptide, and desdiserylglycerylferrirhodin (DDF) a linear tripeptide of ornithine residues [7]. The chemical structure of ferrichromes is also conserved, consisting of six substrate molecules: a core heme-binding unit consisting of three N^5 -acyl- N^5 -hydroxy-L-ornithines (AHO) and a ring of three amino acids (Figure 2.1). One amino acid is always a glycine, while the remaining two amino acids can be alanine, serine, or glycine [5, 7]. Ferrichrome has three glycines, ferrichrome A has

two serines and one glycine, ferrichrome C and malonichrome have two glycines and one alanine, and ferricrocin has two glycines and one serine [7]. Acyl groups attached to AHO substrates can also vary (Figure 2.1).

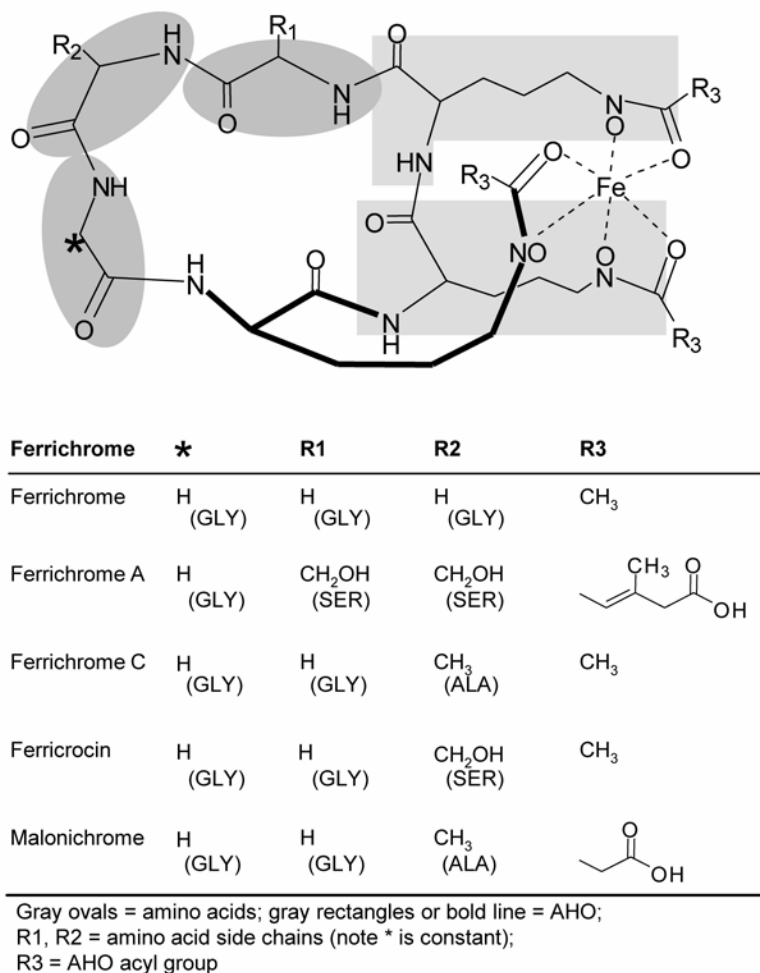


Figure 2.1. Ferrichrome structure. Chemical structure of five different ferrichromes and the corresponding amino acid and AHO acyl group constituents.

Substrate specificity of NRPSs is believed to be mediated by the A domain [8-10] although some studies have suggested a role for the C domain in selective acceptance of substrates from the A domain [8, 11]. A 10 amino acid (AA) NRPS substrate specificity “code” consisting of single, nonadjacent amino acid residues in

the A domain has been proposed, based primarily on examination of bacterial NRPS A domains [9, 12]. Few of these have been tested experimentally and the extent to which this code is applicable to fungal NRPS A domains remains unknown [13]. Since the chemical structure and composition of siderophores produced by fungal ferrichrome synthetases is largely conserved, phylogenetic and structural analyses of these proteins provide an opportunity to correlate protein structure and candidate specificity residues of the A domains with known chemical products.

Ferrichrome siderophores perform key functions in fungal cells. Early work on *Neurospora crassa* suggested that ferricrocin aids in asexual spore germination by storing iron reserves within spores [14, 15]. This role in asexual development has been confirmed for ferrichrome-type siderophores of other fungal species such as *Penicillium chrysogenum* [16] and *Aspergillus nidulans* [17]. In contrast, *Cochliobolus heterostrophus* and *Fusarium graminearum* intracellular siderophores have a major role in sexual spore development, but no obvious role in asexual development [18]. A role in sexual development has also been described for intracellular siderophores of *A. nidulans* [19]. Intracellular siderophores are thought to buffer against reactive oxygen species (ROS) generated by the Haber-Weiss-Fenton reaction in the presence of unbound iron, by sequestering cellular free iron [16]. Indeed, *A. nidulans* mutants lacking ability to produce intracellular siderophores show increased levels of intracellular free iron [17] and a corresponding increase in sensitivity to ROS [19]. *C. heterostrophus* mutants lacking ability to make intracellular siderophores, however, are like wild-type (WT) strains in terms of sensitivity to ROS, although mutants lacking extracellular siderophores do show increased sensitivity to ROS [20].

These subtle functional differences observed between intracellular ferrichrome synthetase mutants of *C. heterostrophus* and *A. nidulans*, as well as the presence of

two or more copies of the genes encoding ferrichrome synthetases in some fungal species suggested the hypothesis that more than one lineage of *NPS* genes may be responsible for intracellular siderophore biosynthesis in fungi. In this study, we sought to: 1) identify homologs of *C. heterostrophus* and *A. nidulans* ferrichrome synthetases in a phylogenetically representative sample of fungal genomes, 2) address the hypothesis of two distinct lineages of ferrichrome synthetases 3) analyze the structural evolution of enzymatic domains encoded by these genes by phylogenetic analysis, and 4) investigate key positions in A domains that may be involved in substrate specificity.

2.3 Materials and Methods

2.3.1 Genomes Surveyed for Ferrichrome-Associated Nonribosomal Peptide Synthetases

Candidate homologs of *C. heterostrophus* NPS2 [3, 18] and *A. nidulans* SidC [19] were identified through blastp and tblastn searches using individual A domains from both NPS2 and SidC proteins as a query set. Fungal genome datasets interrogated included those at the Broad Institute (<http://www.broad.mit.edu/>) (*A. nidulans*, *Aspergillus terreus*, *Batrachomyces dendrobatis*, *Botrytis cinerea*, *Candida albicans*, *Candida guilliermondii*, *Candida lusitanae*, *Chaetomium globosum*, *Coccidioides immitis*, *Coprinus cinereus*, *Cryptococcus neoformans*, *F. graminearum*, *Histoplasma capsulatum*, *Magnaporthe grisea*, *N. crassa*, *Rhizopus oryzae*, *Sclerotinia sclerotiorum*, *Stagonospora nodorum*, *Uncinocarpus reesii*, and *Ustilago maydis*), the Sanger Institute (*Schizosaccharomyces pombe*, *Aspergillus fumigatus*), the Joint Genome Institute (<http://www.jgi.doe.gov/>) (*Laccaria bicolor*,

Aspergillus niger, *Trichoderma reesii*, *Phanerochaete chrysosporium*, and *Phycomyces blakesleeanus*), the DOGAN database at the NITE institute (<http://www.bio.nite.go.jp/ngac/e/rib40-e.html>) (*Aspergillus oryzae*), and the raw genome sequence of *Alternaria brassicicola*, available at Washington University (<http://www.genome.wustl.edu/genome>). The all fungal blast portal at the Saccharomyces Genome Database (<http://seq.yeastgenome.org/cgi-bin/blast-fungal.pl>) was used to survey the *Saccharomyces cerevisiae* genome and those of a number of other wild yeast species (*Saccharomyces bayanus*, *Saccharomyces castellii*, *Saccharomyces kluyveri*, *Saccharomyces kudriavzevii*, *Saccharomyces mikatae*, *Saccharomyces paradoxicus*, *Saccharomyces servizzii*, *Saccharomyces unisporus*, *Ashbya gossypii*, *Candida glabrata*, *Candida parapsilopsis*, *Candida tropicalis*, *Kluveromyces delphensis*, *Kluveromyces lactis*, *Kluveromyces marxianus*, *Kluveromyces thermotolerans*, *Kluveromyces waltii*, *Lodderomyces elongisporus*, and *Yarrowia lypolitica*).

All hits with an e value less than e^{-10} were extracted and an initial phylogenetic analysis used to identify a putative set of ferrichrome NRPSs. The individual A domains of all candidate ferrichrome synthetase NRPSs were aligned with Tcoffee and a phylogeny constructed using the WAG model plus gamma with 100 bootstrap replicates in PhyML [21]. A domains of 12 additional NRPSs found in *C. heterostrophus*, representative of the diverse clades of fungal NRPSs [3], as well as the top bacterial hit (NCBI Accession YP_049592) to both NPS2 and SidC, were used as outgroups in this initial analysis and in further analyses of the complete dataset [3]. A monophyletic clade with bootstrap support >85 % containing all known ferrichrome synthetase NRPSs was identified and all members of this clade were considered in further analyses (see Appendix 2.1). Two additional known ferrichrome siderophores,

one from *Aureobasidium pullulans* (AAD00581) [6] and one from *Omphalotus olearius* (fso1, AAX49356) [22] were included. Several NRPSs identified previously as putative siderophore metabolite producers (designated the SidE clade) [23], which fell in a clade just outside the major clade of known ferrichrome synthetases, were also included.

2.3.2 Annotation of Candidate Ferrichrome Synthetases

Candidate ferrichrome synthetases were annotated by 1) using the candidate NRPS proteins as queries against the PFAM database and 2) utilizing NRPS specific HMM models built using HMMER [24] from a larger dataset of fungal NRPS A and C domains (KE Bushley and BG Turgeon, submitted manuscript, Chapter 3). Discrepancies between the two methods and with published domain architectures were resolved by manual inspection and adjustment. Individual A domains were extracted using a customized Perl script (available upon request) and the limits of the A domain were defined as in Lee et al [3], spanning from ~33 residues upstream of the A1 core motif to three residues downstream of the A10 core motif [12].

Several proteins identified appeared to be incomplete or incorrectly annotated in the databases. The gene corresponding to *B. cinerea* BC1G15494 (see Appendix 2.1) is on the end of supercontig 180; we assumed it is incomplete, as it encodes only a single A-T-C module. We reannotated the genes corresponding to HCAG07428 and HCAG07429 as a single gene. The sequence corresponding to *H. capsulatum* HCAG07428 spanning the first C and second A domains is of low quality; the second A domain and the second and sixth C domains are missing from our analyses. Similarly, *U. reesii* UREG00890 and UREG00891 appear to correspond to a single gene. *C. cinerea* CC1G04210 is unusual in that it contains only a single A-T-C module followed by a T-C repeat. Inspection of sequences flanking this gene did not

reveal additional A, T, or C domains.

2.3.3 Phylogenetic Analyses

2.3.3.1 Complete Set of A Domains

A domain protein sequences were aligned to the crystal structure of the A domain of Gramicidin synthetase (GrsA) [25] using 3D-Coffee with the Blosum 62 substitution matrix and default gap opening and extension parameters [26]. Because the alignment of these highly divergent proteins contained regions of ambiguous alignment, we performed a sensitivity analysis to assess the effect of the alignment on the final phylogeny obtained. Starting with the final manually adjusted alignment of A domains, we created and analyzed three different alignment, using maximum likelihood (ML): 1) an alignment retaining the majority of divergent regions, 2) a semi-conservative alignment omitting the most divergent regions (i.e., those with more than 70% gaps per column in the alignment), and 3) a highly conservative alignment with all divergent regions with more than 50% gaps per column removed. The WAG substitution matrix with rate variation described by a gamma distribution with 4 rate categories was identified as the best protein substitution model for this dataset according to the AIC criterion using ProTest [27]. ML analyses using the WAG model plus gamma in PhyML showed that the three alignments produced identical topologies for the major clades with only slight differences in groupings of taxa within each clade (available upon request). We used the semi conservative alignment for all further analyses. Phylogenetic analyses were conducted with PhyML using the WAG amino acid substitution model and gamma distribution with 4 rate categories and estimated

alpha parameter and 500 bootstrap replicates [21] and with Mr. Bayes using 5 million MCMC generations sampled every 100 generations with a mixed AA prior [28].

The program Genetree [29] was used to reconcile the ML tree to a species tree (see Appendix 2.2) to infer a history of A domain duplications using both duplication and loss as the optimality criterion. The species tree was based on three recent phylogenetic studies of the fungal kingdom [30-32]. These studies agree on placement of all taxa included in this study except the Dothideomycetes whose placement remains unstable. In different types of analyses they have grouped with Eurotiomycetes [31], as more closely related to Sordariomycetes and Leotiomycetes [31], or as basal to all three of these classes [30, 31]. We chose to place the Dothideomycetes as sister to other filamentous ascomycetes in the subphylum Pezizomycotina as they are placed in this position in phylogenies with larger taxon sampling [30] and this placement agrees with another recent phylogenomic study [33] (see Appendix 2.2). *A. pullulans* was shown to have diverged earlier than our other sampled Dothideomycete taxa in a recent class wide phylogeny of Dothideomycetes and is thus placed at the base of the Dothideomycete clade [34].

2.3.3.2 Individual Lineage Analyses

To analyze mechanisms of evolution of the genes encoding ferrichrome synthetase proteins in more detail, those enzymes grouping with *C. heterostrophus* NPS2 and those grouping with *A. nidulans* SidC in phylogenetic analyses of the complete A domain dataset (see above) were examined separately. For each group, A and C domains were extracted using the Perl script described above. T domains were excluded, as they are significantly shorter (66 amino acids versus 300 amino acids) and resulted in highly unresolved phylogenies. The limits of the A domain were

defined as described above while the C domain was delimited according to the PFAM model (PFAM00668)(www.sanger.ac.uk/Software/Pfam/) and extends from four residues before the C1 motif to four residues after the C5 motif. Each domain was aligned separately with TCOFFEE using default parameters and phylogenetic analyses were conducted with PhyML and Mr. Bayes using the same parameters described above for the larger dataset. We used A and C domains from the first complete A-T-C module of the SidE group as an outgroup as this module grouped directly outside the major clade of ferrichrome synthetases in both the ML and Bayesian trees while the second module grouped consistently with other types of fungal NRPSs represented by the other *C. heterostrophus* NRPSs.

As the majority of NRPS genes are multimodular, tandem duplication represents a plausible hypothesis for the generation of a multimodular gene from a single A-T-C unit. To evaluate this hypothesis, we constructed phylogenies in PhyML of a representative ferrichrome synthetase from each lineage, i.e., *C. heterostrophus* NPS2 and *A. nidulans* SidC for the NPS2 and NPS1/SidC lineages, respectively. These trees were evaluated using the Possible Duplication History (PDH) algorithm developed to determine if a phylogeny is consistent with a history of tandem duplication [35].

2.3.4 Substrate Specificity

2.3.4.1 Structural Modeling

Three-dimensional models of A domains were generated by using template-based modeling techniques. Blast searches [36, 37] of the Protein Data Bank (PDB) database [38] (www.rcsb.org/pdb/home/home.do), using a subset of A domain

sequences from *C. heterostrophus* NPS2 (AAX09984), *F. graminearum* NPS2 (FG05372), *F. graminearum* NPS1 (FG11026), *A. nidulans* SidC (AN0607), *U. maydis* sid2 (UM05165), *U. maydis* fer3 (UM01434), and *S. pombe* Sib1 (CAB72227) as queries, indicated a high level of similarity with the phenylalanine activating A domain of the NRPS for gramicidin (GrsA), PDB code: 1AMU; [25]. Using the Combinatorial Extension method [39] and the 1AMU_A (ie., monomer A of 1AMU) structure as input, other structurally similar proteins with associated crystal structures were identified. The structures of the monomers of 1AMU_A, 1PG3_A, 1ULT_A, 1LC_I, 1T5D_X and 1MD9_A were superimposed and a structural alignment of these was produced manually with the help of graphic tools included in the commercial programs ICM (MOLSOFT Inc) and DS-Modeling (Accelrys Inc.). The objective of having a structural alignment of multiple proteins is to better define the regions of the fold that are conserved and understand where structural variability can occur.

The subset of our NRPS A domain sequences (described above) were selected for structural modeling and added to the structural alignment. The alignment was corrected manually by adjusting the positions of insertions and deletions that were incompatible with the secondary-structure elements observed in the 3-dimensional (3D) structures of the templates. All residues forming the walls of the binding pocket for the Phe substrate in 1AMU_A as well as residues that bind the adenosine monophosphate AMP moiety were identified. In addition, residues aligned with the 10 amino acid positions (10AA code) predicted to be involved in substrate specificity in the GrsA sequence [9], as well as three additional residues identified by Schwecke et al. [6] to be important in binding the AHO substrate (13AA code) were identified in the structural alignment. The Cartesian coordinates of the template structures were retrieved from the PDB [38], and the final multiple alignment of the experimental and template structures were used as input data for MODELLER [40-43]. During the

process of model generation, MODELLER minimizes the violations of distance and dihedral-angle restraints derived from the templates. For each sequence a set of 3D models were generated and those that best satisfied the set of restraints were kept. More than one template structure was used during the model generation process in order to assess the variability of the different regions of the A domain structures.

2.3.4.2 Evolutionary Approaches to Identify Specificity Residues

We utilized several amino acid based methods to detect residues with a potential role in specificity. These included the specificity-determining positions (SDPpred) algorithm [44] and server (<http://math.genebee.msu.ru/~psn/>) and Type I and Type II functional divergence, two likelihood based methods in the DIVERGE 2 package to detect functional residues [45, 46]. Type I functional divergence detects changes in evolutionary rates between clusters indicative of changes in constraint or selective pressure, while both the SDP algorithm and Type II functional divergence aim to identify residues that are conserved within a cluster but show a change in amino acid properties between clusters. For these analyses, we used the major groups identified in our ML analysis of all A domains as individual clusters. The second A domain of *S. pombe* *sib1* and the third A domain of *O. olearius* *fso1* were omitted because both are highly divergent from other A domains and likely degenerate as they lack several core functional motifs [6, 9]. The Dothideomycete module 3 A domain was grouped with the cluster for the second *A. nidulans* SidC A domain, as all methods used require clusters of greater than three taxa and our data suggested that all of these domains code for the same amino acid.

2.4. Results

2.4.1 Distribution of Ferrichrome Synthetases in Fungi

With the exception of *S. pombe* none of the yeast, chytrid, or zygomycete genomes surveyed contained a candidate ferrichrome synthetase NRPS. In contrast, all filamentous ascomycetes contained at least one and many had two (Table 2.1).

Table 2.1. Fungal genomes and number of ferrichrome synthetases identified

Species	Number of Ferrichrome NRPSs	Species	Number of Ferrichrome NRPSs
Hemiascomycetes		Ascomycetes	
<i>Ashbya gossypii</i>	0	<i>Alternaria brassicicola</i>	1
<i>Candida albicans</i>	0	<i>Aspergillus fumigatus</i>	1
<i>Candida glabrata</i>	0	<i>Aspergillus nidulans</i>	1
<i>Candida parapsilopsis</i>	0	<i>Aspergillus niger</i>	1
<i>Candida tropicalis</i>	0	<i>Aspergillus oryzae</i>	1
<i>Kluyveromyces delphensis</i>	0	<i>Aspergillus terreus</i>	1
<i>Kluyveromyces lactis</i>	0	<i>Botrytis cinerea</i>	3 ^a
<i>Kluyveromyces marxianus</i>	0	<i>Chaetomium globosum</i>	2
<i>Kluyveromyces thermotolerans</i>	0	<i>Coccidioides immitis</i>	1
<i>Kluyveromyces waltii</i>	0	<i>Fusarium graminearum</i>	2
<i>Lodderomyces elongisporus</i>	0	<i>Histoplasma capsulatum</i>	1 ^{a, b}
<i>Saccharomyces bayanus</i>	0	<i>Magnaporthe grisea</i>	1
<i>Saccharomyces castellii</i>	0	<i>Neurospora crassa</i>	1
<i>Saccharomyces cerevisiae</i>	0	<i>Sclerotinia sclerotiorum</i>	2
<i>Saccharomyces kluyveri</i>	0	<i>Stagonospora nodorum</i>	1
<i>Saccharomyces kudriavzevii</i>	0	<i>Trichoderma reesii</i>	1
<i>Saccharomyces mikatae</i>	0	<i>Uncinocarpus reesii</i>	1 ^b
<i>Saccharomyces paradoxus</i>	0		
<i>Saccharomyces servazzii</i>	0	<u>Schizosaccharomycetes</u>	
<i>Saccharomyces unisporus</i>	0	<i>Schizosaccharomyces pombe</i>	1
<i>Yarrowia lypolitica</i>	0		
		<u>Basidiomycetes</u>	
<u>Chytridiomycota</u>		<i>Coprinus cinerea</i>	1
<i>Batrachochytrium dendrobatis</i>	0	<i>Cryptococcus neoformans</i>	0
		<i>Laccaria bicolor</i>	0
<u>Zygomycota</u>		<i>Phanaerochaete chrysosporium</i>	0
<i>Phycomyces blakesleeanus</i>	0	<i>Ustilago maydis</i>	2
<i>Rhizopus oryzae</i>	0		

^a The genes, BC1G15494 and HCAG07428/HCAG07429 are partial (see text).

^b HCAG07428 and HCAG07429 and UREG00890 and UREG00891 reannotated as single genes.

B. cinerea appears to have three. For the five basidiomycete genomes examined, two known NRPSs (*sid2* and *fer3*) were found in *U. maydis*, one undescribed ferrichrome synthetase was identified in *C. cinerea* while *P. chrysosporium*, *L. bicolor*, and *C. neoformans* lacked genes encoding these enzymes. As noted earlier, the ferrichrome synthetase *fso1* is known from the basidiomycete *O. olearius* [22].

2.4.2 Domain Architecture of Ferrichrome Synthetases

Ferrichrome NRPSs show a diversity of domain architectures (Figure 2.2). These have been designated ‘types’ [6] and we use this terminology here. We found six types, including five previously identified. All are modular (except Type VI), consisting of three to four complete A-T-C modules usually followed by a T-C repeat. *C. heterostrophus* NPS2, as described previously [3, 20], has four complete A-T-C modules and a terminal T-C repeat (Type V). This structure is conserved in NPS2 homologs from the other Dothideomycetes examined (*A. brassicicola* and *S. nodorum*). In contrast, most other ferrichrome synthetases examined (Types I – IV) have only three complete A-T-C modules and a terminal T-C repeat. *U. maydis* *sid2* (Type I) is an exception, with a single terminal T-C unit. *S. pombe* *sib1* (Type III) is the only representative of its class; the second complete module has a degenerate A domain in which many of the signature motifs are missing [6] and an internal T-C unit after the first complete A-T-C module. Similarly, all Type IV NPS2 homologs (e.g., *F. graminearum* NPS2) have an internal T-C after the second complete A-T-C module. The only representative of Type VI, *C. cinerea* CC1G04210, has a single A-T-C module followed by a T-C repeat.

SidE proteins, suggested by Cramer et al [23] to be putative ferrichrome synthetases have a different domain organization from known ferrichrome synthetases.

They consist of only two complete modules and an additional N-terminal C domain (5' C-A-T-C-A-T-C3'), except for *A. fumigatus* Afu3g03350 and Afu3g15270 which lack the N-terminal C domain (5' A-T-C-A-T-C3').

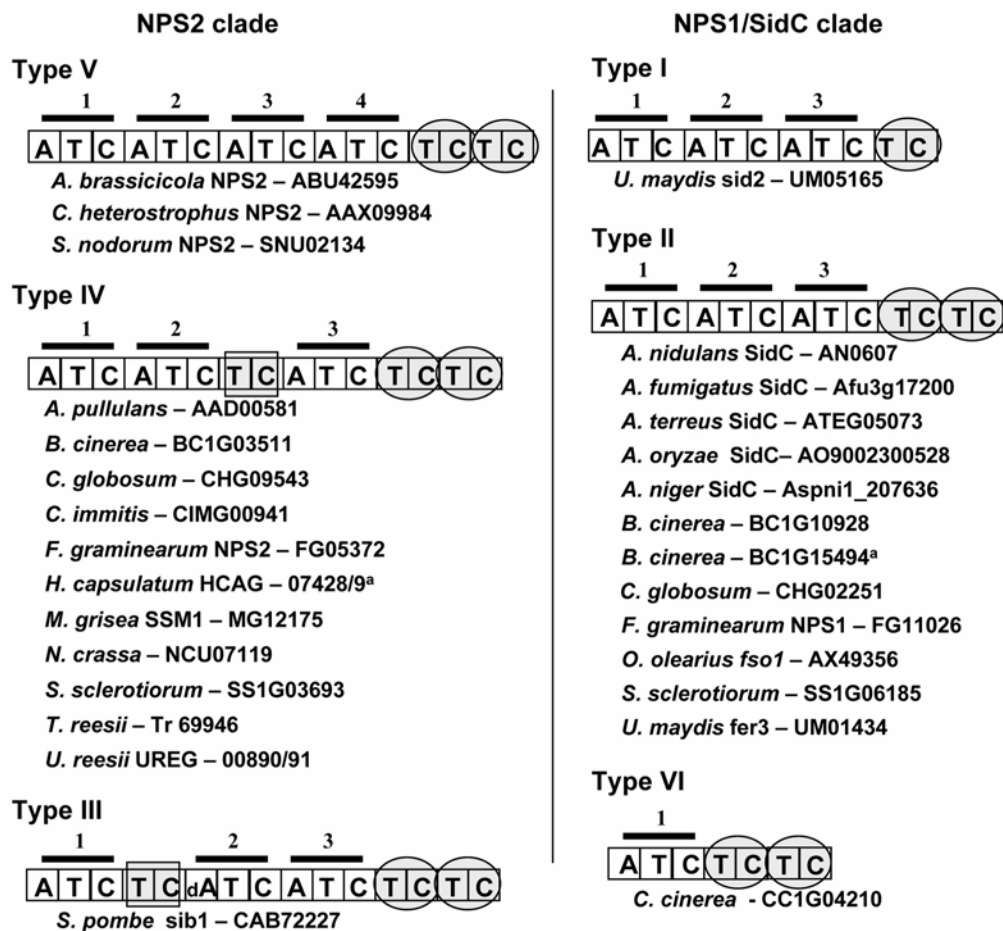


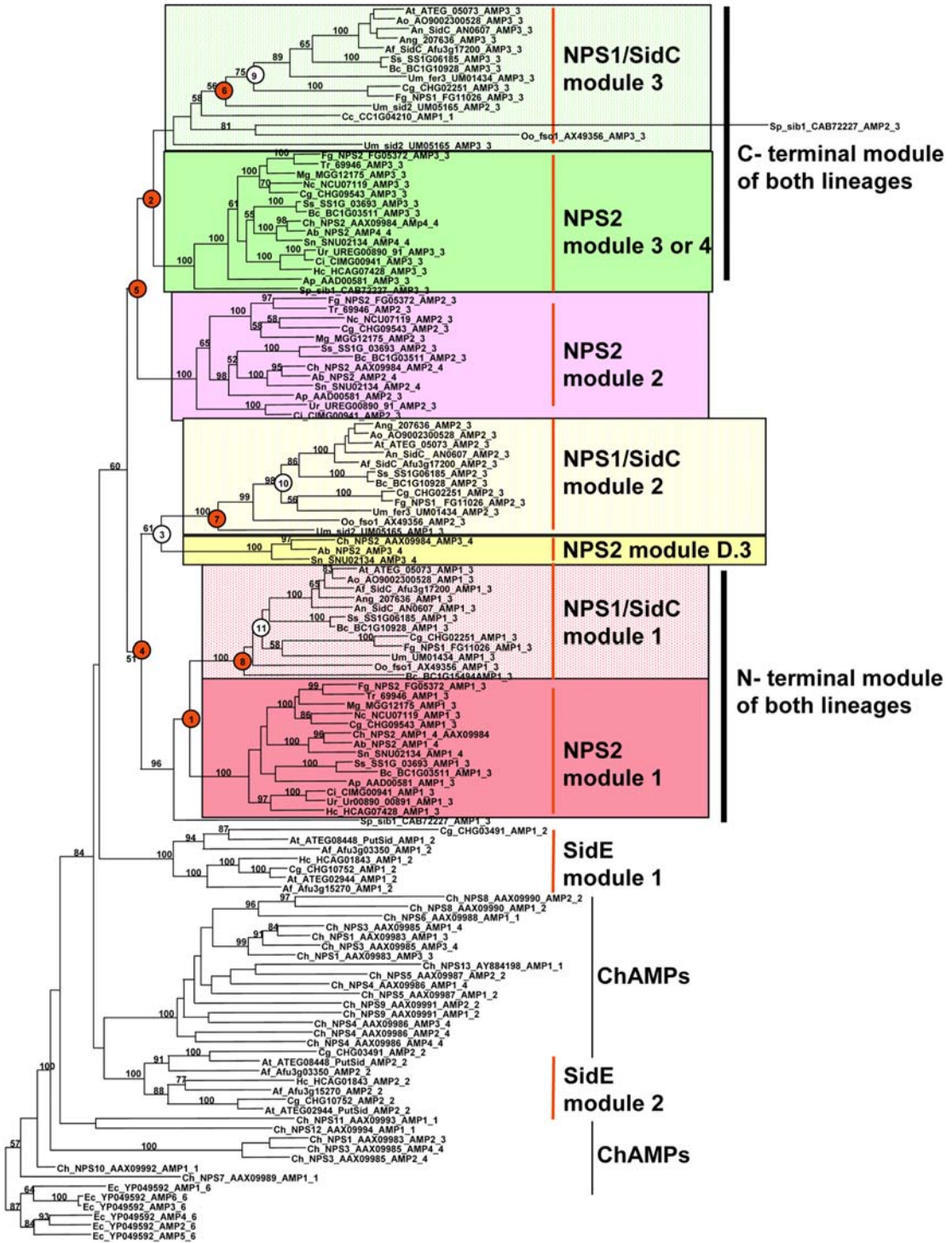
Figure 2.2 Six modular architectures for ferrichrome synthetase NRPSs. Types III, IV, and V are in the NPS2 lineage while Types I, II, and VI are in the NPS1/SidC lineage. A: adenylation domain, T; thiolation domain, C; condensation domain. dA; degenerate A domain. Bars above boxes indicate complete modules. Circles indicate incomplete modules and/or a T-C unit. Superscript ‘a’ indicates partial gene. Thus, although at least one representative of each Type (except Type VI) has been shown to produce the conserved ferrichrome siderophore compound consisting of six substrates (three amino acids and three AHO units) (Figure 2.1), the domain architectures of the ferrichrome synthetases responsible for their biosynthesis vary considerably.

2.4.3 Two Distinct Lineages of Ferrichrome Synthetases

Both methods of phylogenetic analysis of A domains from the complete dataset showed a history of domain duplications that supports the hypothesis of at least two separate lineages of fungal ferrichrome synthetases (Figure 2.3, see Appendix 2.3). For all A domains, we find two clades whose members correspond to homologs of *C. heterostrophus* NPS2 or to *A. nidulans* SidC. For convenience, we call the lineage represented by *C. heterostrophus* and *F. graminearum* NPS2 (Types V and IV, respectively, Figure 2.3), the NPS2 lineage. The other lineage, represented by *A. nidulans* SidC, *U. maydis* fer3, *F. graminearum* NPS1, *U. maydis* sid2 and *C. cinerea* CC1G04120 (Types I, II and VI, Figure 2.2), we call the NPS1/SidC lineage. Some species, e.g., *F. graminearum*, *B. cinerea*, *C. globosum*, *S. sclerotiorum* have representatives in both lineages. Others, such as *U. maydis* and *B. cinerea*, have more than one representative within the NPS1/SidC lineage.

The reconciliation analysis clearly identified duplication nodes giving rise to the first (N-terminal, node 1, red boxes) and final (third or fourth) (C-terminal, node 2, green boxes) A domains of both lineages (Figure 2.3). This analysis also provides support for a relationship at node 3 between the third A domain of NPS2 Type V of the Dothideomycetes (D.3) and the second A domains of NPS1/SidC Type II (Figure 2.3, yellow boxes). ML and Bayesian phylogenetic methods support the duplication at node 1, giving rise to the N-terminal A domains of both lineages (red boxes), with high Bayesian posterior probability (pp = 1.00) but low ML bootstrap support (bs < 50%) (Figure 2.3, see Appendix 2.3). The duplication at node 2, giving rise to the C-terminal A domains of members of both lineages (Figure 2.3, green boxes), is weakly supported by both types of phylogenetic analysis (bs < 50%), pp = .74) (Figure 2.3, see Appendix 2.3). For the internal modules, both ML and Bayesian analyses group

Figure 2.3. Maximum likelihood tree of all AMP domains examined in this study demonstrating two separate lineages of ferrichrome synthase NRPSs. N-terminal A domains of both lineages group together and C-terminal domains of both lineages group together (thick vertical bars). NPS2, module 2 groups with the C-terminal modules, while NPS1/SidC module 2 and Dothideomycete NPS2 module D.3 group with the N-terminal modules. Numbered nodes indicate duplications inferred from the reconciliation analysis. White circles indicate a duplication inferred due to incongruence of the gene tree with the species tree (see Appendix 2.2), while red circles indicate a duplication inferred due to the presence of two copies of a gene in the same species. Bootstrap support values greater than 50% are reported above branches. Note that the A domains of SidE module 1 group as directly sister to all ferrichrome synthetase A domains examined here, while A domains of SidE module 2 group with A domains of other types of *C. heterostrophus* NRPSs. For species and protein Accession numbers see Appendix 2.1. Nomenclature: e.g., Ch_NPS2_AAX09984 AMP3_4 indicates *C. heterostrophus*, protein accession number AAX09984, AMP module 3 of a total of 4 (see Figure 2.2). For Bayesian analysis, see Appendix 2.3.



the third A domain (D.3) of NPS2 Type V and the second A domain of the NPS1/SidC lineage together (yellow boxes), supporting a duplication at node 3 inferred by the reconciliation analysis (Figure 2.3, see Appendix 2.3). The Bayesian analysis provides higher support (pp = 1.00) for this relationship than does the ML analysis (bs = 61%). These clades (yellow boxes) group with the N-terminal modules of both lineages (Figure 2.3, red boxes), with higher Bayesian (pp = 1.00) than ML (bs = 51%) support; a duplication at node 4 was inferred by the reconciliation analysis. Finally, the module 2 A domains of NPS2 Types IV and V (pink boxes) group together and with the C-terminal modules of both lineages (Figure 2.3, green boxes), however with weak support (bs < 50% and pp = .74). The reconciliation analysis identified a duplication at node 5 corresponding to this relationship (Figure 2.3).

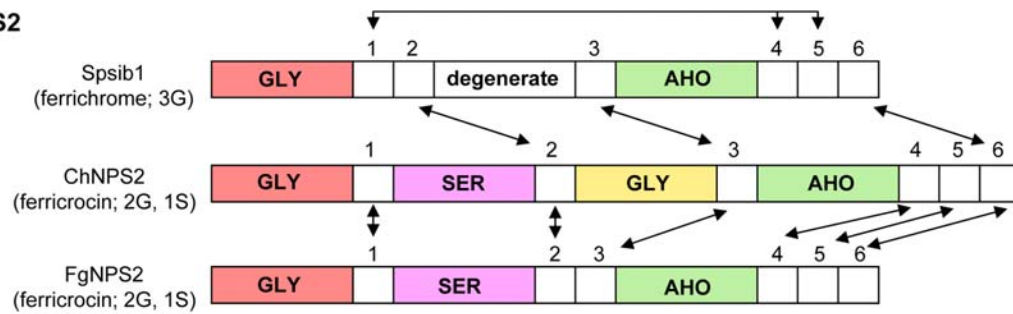
The phylogenetic relationships of A domains are mapped by color to representative ferrichrome synthetases in Figure 2.4 (color corresponds to clades identified in Figure 2.3). These data clearly show that the N-terminal and C-terminal A domains of each lineage are related by duplication (Figure 2.4). Similarly, the third A domain of the Dothideomycete Type V (D.3) proteins appears related to the second A domain of the NPS1/SidC lineage by duplication (yellow). The second module of Dothideomycete Type V, which is the only type of ferrichrome synthetase consisting of four complete A-T-C modules (Figure 2.2), does not have an obvious counterpart in other ferrichrome synthetases (Figure 2.4, pink).

2.4.4 Additional Duplications Within the NPS1/SidC Lineage

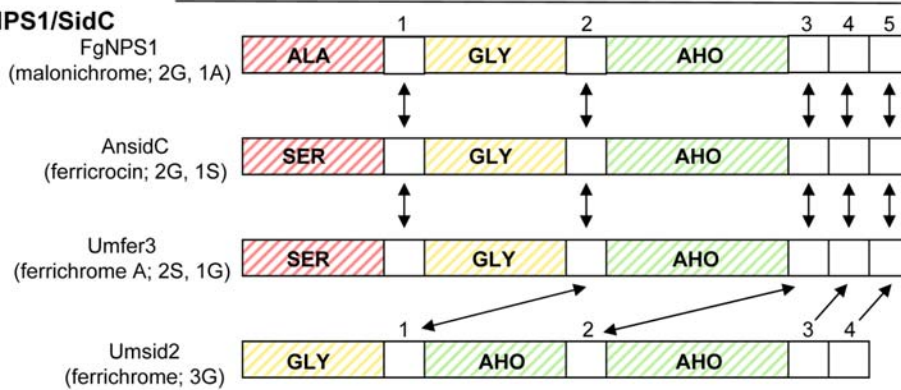
There is evidence for further duplications within the NPS1/SidC lineage. The reconciliation analysis identified duplication nodes at 6, 7, and 8 (Figure 2.3) due to the presence of two representatives from the NPS1/SidC lineage in both *U. maydis*

Figure 2.4. Schematic representation of phylogenetic relationships among A and among C domains within each lineage. A domain relationships for each lineage and between lineages are color coded as in Figure 2.3 and Appendix 2.3. C domain relationships are indicated by arrows for each lineage. The NPS2 lineage relationships are indicated in the top half of figure and the NPS1/SidC lineage relationships in the bottom half of figure. Scheme is based on phylogenetic analyses of A (Figure 2.3, see Appendixs 2.3 and 2.4A , 2.4C) and C (see Appendix 2.4B, 2.4D) domains. Spsib1, ChNPS2, FgNPS2, FgNPS1, AnSidC, Umfer3, and Umsid2 are representative of architectural Types I-V (Figure 2.2). Also mapped on the A domains are predicted substrates adenylated by each domain, based on structural modeling (Table 2.1, Figure 2.7). SER = serine, GLY = glycine, ALA = alanine, AHO = N⁵-acyl-N⁵-hydroxy-L-ornithine. Within the NPS2 lineage, ChNPS2 and FgNPS2 C domain analyses clearly indicate that C2 domains are related, as are C3 domains. Thus the difference in protein architecture in this region is presence/absence of an A domain between C2 and C3. A similar argument can be made for the difference in protein structure between C1 and C2 C domains of Spsib1 vs those of ChNPS2 and FgNPS2. For the NPS1/SidC lineage, A and C domain analyses of FgNPS1, AnSidC, and Umfer3 clearly indicate that there is a one to one relationship for all A and all C domains. Examination of Umsid2, however, indicates that Umsid2 module 1 A domain is related to the module 2 A domains of the other members of this group, while Umsid2 modules 2 and 3 A domains are related to the C-terminal module of the other members of this group. Umsid2 appears to lack the N-terminal A domain of other NPS1/SidC members, since the C domain from module 1 is related to the C domains of module 2 of the rest of the lineage. Similarly the C domains from Umsid2 module 2, 3, 4 are related to the C domains of modules 3, 4, 5 of the rest of the NPS1/SidC lineage.

NPS2



NPS1/SidC



2.4.4 Additional Duplications Within the NPS1/SidC Lineage

There is evidence for further duplications within the NPS1/SidC lineage. The reconciliation analysis identified duplication nodes at 6, 7, and 8 (Figure 2.3) due to the presence of two representatives from the NPS1/SidC lineage in both *U. maydis* [UM01434/fer3 (Type II) and sid2 (Type I)] and *B. cinerea* [BC1G10928 and BC1G15494 (Type II)] (Figure 2.3). Duplication nodes were also identified due to the incongruence of *F. graminearum* FG11026 (NPS1) and *C. cinerea* CHGG02251 with the species phylogeny at nodes 9, 10, and 11 where these two NRPSs group with or outside of basidiomycete *U. maydis* fer3 rather than with other ascomycete NRPSs (Figure 2.3). Thus, the data provide support for one and possibly two additional bifurcations within the NPS1/SidC lineage. The placement of certain NPS1/SidC lineage genes is ambiguous. Type VI *C. cinerea* CC1G04210 has a single A domain which groups consistently with the third A domain of *U. maydis* sid2 (Figure 2.3, see Appendix 2.3). The other basidiomycete gene, *O. olearius* fso1, tends to group with other Type II NPS1/SidC proteins. In both analyses, the first and second modules of fso1 group at the base of the clades containing the corresponding modules of the NPS1/SidC Type II proteins, usually with *U. maydis* fer3 (Figure 2.3, see Appendix 2.3). The third fso1 A domain is highly diverged and contains degenerate core motifs and its placement varies (Figure 2.3, see Appendix 2.3). The single A-domain of incomplete *B. cinerea* BC1G15494 tends to group at the base of the clade containing the first A domain of all Type II NPS1/SidC proteins (Figure 2.3, see Appendix 2.3), however, in both ML and Bayesian analyses (Figure 2.3, see Appendix 2.3), it shows incongruence with the species phylogeny by grouping outside of basidiomycete NRPSs in this clade.

2.4.5 *S. pombe* sib1

The relationship of Type III *S. pombe* sib1 to other ferrichrome synthetases is ambiguous. In both the ML and Bayesian analyses, the first A domain of sib1 groups as sister to the first A domains of both the NPS2 and NPS1/SidC lineages (Figure 2.3, see Appendix 2.3) with fairly high support (bs = 96 % and .89 pp), suggesting an ancestral relationship of this sib1 A domain and the first A domains of both lineages. However, the sib1 module 3 A domain groups with the A domains of NPS2 terminal modules 3 or 4, in both trees (Figure 2.3, see Appendix 2.3), with strong support (bs = 100 % and pp = 1.00). The sib1 module 2 A domain groups with the module 3 A domain of the NPS2 lineage (Type V) with high support in the Bayesian analysis (pp = 1.00) (see Appendix 2.3). In the ML tree, however, it groups with the N-terminal A domain of the NPS1/SidC lineage (Figure 2.3), but without bootstrap support. As discussed above, this second A domain is highly diverged, lacks several core A domain motifs [9], and as suggested by Schwecke [6], is likely nonfunctional. As sib1 most consistently groups with homologs of *C. heterostrophus* NPS2, we placed it in the NPS2 lineage (Figure 2.3).

2.4.6 Putative Ferrichrome Synthetases in the SidE Clade

The SidE proteins, identified as putative ferrichrome synthetases [23], group as sister to all other known ferrichrome synthetases (Figure 2.3, see Appendix 2.3). The A domains of the first and second modules of these proteins however, are not monophyletic. In the ML and Bayesian analyses, SidE module one A domain groups as sister to known ferrichrome synthetases while the SidE module two A domain groups with other (non-ferrichrome synthetase) NRPSs from *C. heterostrophus*. Thus,

these results suggest that only the first module of the SidE proteins is clearly related to other known ferrichrome siderophore NRPSs.

2.4.7 Individual Lineage Analysis

The backbones of the A and C tree topologies for each lineage, rooted with the first module of the SidE clade, are shown in Figures. 2.5A and 2.5B. Within each lineage, all A and all C domains fall into well-supported monophyletic clades (see Appendix 2.4A-D).

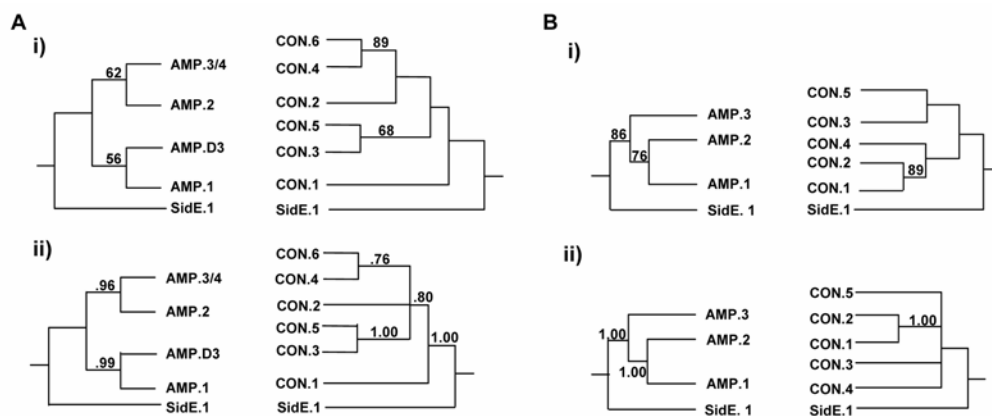


Figure 2.5. Diagrammatic depiction of separate NPS2 (A) and NPS1/SidC (B) lineage AMP and CON domain trees. (i) and (ii) are ML and Bayesian analyses, respectively. **A.** Relationships among A and among C domains in the NPS2 lineage. As demonstrated in the full A domain dataset analyses (Figure 2.3, see Appendix 2.3), both NPS2 lineage A analyses support a relationship between C-terminal modules 3 or 4 and module 2, and a relationship between N-terminal module 1 and Dothideomycete module D.3. For the C trees, both analyses support a relationship between C4, and C6 (bs = 89% and pp = .76) and between C3 and C5 (bs = 68% and pp = 1.00). C2 groups with C4 and 6 in the ML analysis and with C3-6 in the Bayesian analysis but without support in either case. In both trees, C1 is ancestral, but without support. **B.** Relationships among A and among C domains in the NPS1/SidC lineage. As demonstrated in the full A domain analyses (Figure 2.3, see Appendix 2.3), both NPS1/SidC lineage A domain analyses support a relationship between N-terminal module 1 and module 2, and indicate C-terminal module 3 is ancestral. Similarly, the ML and Bayesian trees support a close relationship between the C domains of modules 1 and 2.

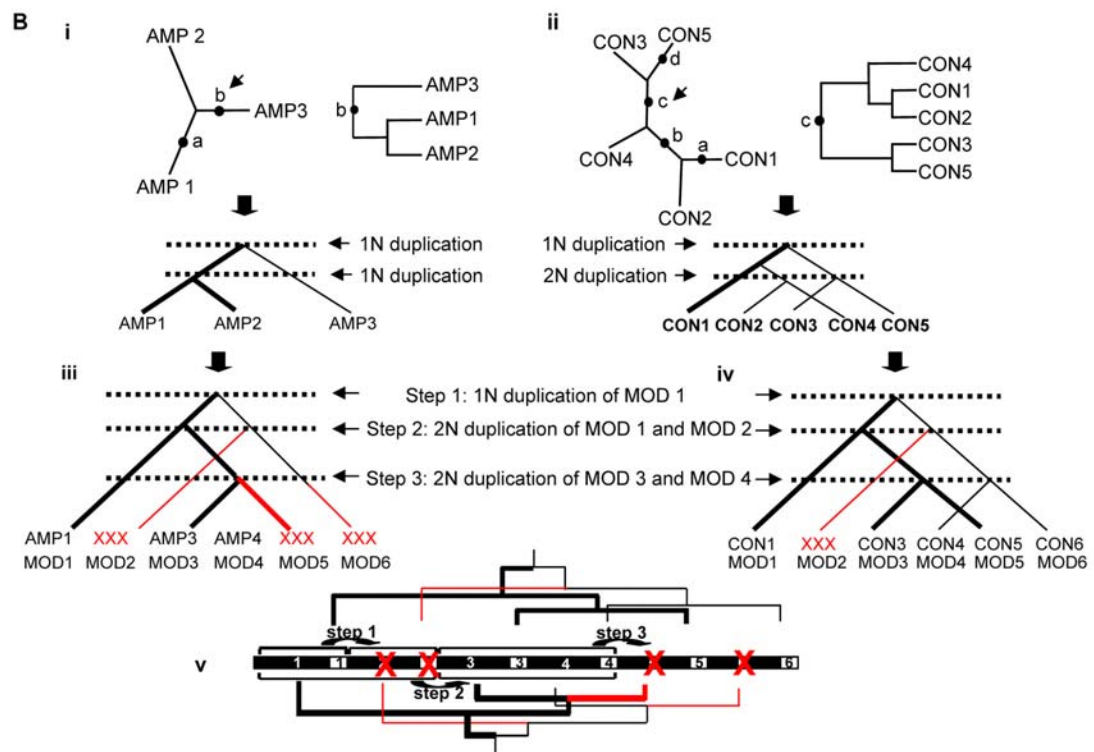
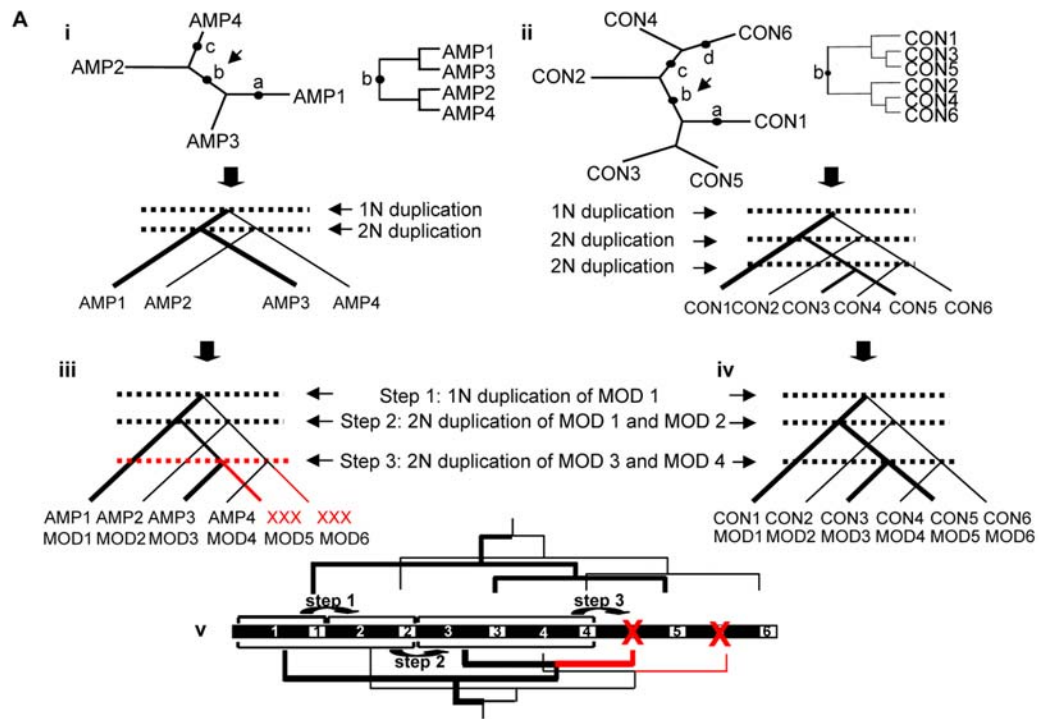
A domain relationships are consistent with those of the full A dataset (compare Appendix 2.4A and 2.4C with Figure 2.3). The first through the sixth C domain of all proteins group together as separate clades for all members of the NPS2 (except *S. pombe* sib1) and the NPS1/SidC lineages (Figure 2.5, see Appendix 2.4B and 2.4D). C domain relationships among representative ferrichrome synthetases are shown in Figure 2.4 (arrows).

For the NPS2 lineage (Figure 2.5A), both A domain tree topologies (ML and Bayesian) support a close relationship between module one A domains of all types (I, IV, V) and the A domain of Dothideomycete Type V module 3 (D.3) (bs = 56% and pp = .99) (Figure 2.5A, see Appendix 2.4Ai-ii). A close relationship is also supported between module 2 A domains of Types IV and V and the terminal module A domains of all types (bs = 62%, and pp = .96) (Figure 2.5A, see Appendix 2.4Ai-ii). The ML and Bayesian analysis of the C domains (Figure 2.5A, see Appendix 2.4Bi-ii) support a close relationship between modules 4 and 6 C domains and between module 3 and 5 C domains (bs = 89% and pp = 0.76, bs = 68% and pp = 1.00, respectively).

The unrooted ML phylogenies of the A and C domains of *C. heterostrophus* NPS2 are shown in Figures 2.6 Ai and Aii. When the C tree is rooted at position b (Figure 2.6Aii) and evaluated with the PDH algorithm [35], the resulting phylogeny is a duplication tree that implies an associated partially ordered duplication history (Figure 2.6Aii). All trees with four taxa are true duplication trees, thus evaluation of the A domains with the PDH algorithm is trivial. However, the duplication tree resulting from rooting the A domain phylogeny at b implies a partially ordered duplication history which also infers a duplication between modules 1 and 3 and between modules 2 and 4, consistent with duplications predicted for C domains (Figures 2.6Ai, Aii).

For the NPS1/SidC lineage, the A domain phylogenies show a strong

Figure 2.6. Evaluation of *C. heterostrophus* NPS2 and *A. nidulans* SidC with the PDH algorithm (possible duplication history). **A.** i) Unrooted maximum likelihood phylogeny of *C. heterostrophus* NPS2 A domains, the duplication tree resulting from rooting the phylogeny at position b (top) and inferred partially ordered duplication history (below). ii) Unrooted maximum likelihood phylogeny of *C. heterostrophus* NPS2 C domains, the duplication tree resulting from rooting the phylogeny at position c, and partially ordered duplication history (bottom). iii) and iv) Representation of the series of three tandem duplication events suggested by the partially ordered duplication trees of C domains. Bold and thin lines indicate relationships among modules 1, 3, and 5 and among modules 2, 4 and 6 respectively. If one infers loss of AMP5 and AMP6, relationships among A domains are consistent with the series of three tandem duplication events inferred from the C domain partially ordered duplication history: Step 1) duplication of A module 1, Step 2) duplication of A modules 1 and 2, and Step 3) duplication of A modules 3 and 4. v) Relationships among A and among C domains in partially ordered duplication histories mapped to the domain architecture with predicted domain losses shown in red. **B.** i) Unrooted maximum likelihood phylogeny of *A. nidulans* SidC A domains, duplication tree rooted at position b (top) and inferred partially ordered duplication history (bottom). ii) Unrooted maximum likelihood phylogeny of *A. nidulans* SidC C domains, duplication tree rooted at position c (top) and inferred partially ordered duplication history (bottom). iii) and iv) Representation of the series of three tandem duplication events suggested by the partially ordered duplication trees. Bold and thin lines as in A above. Relationships among *A. nidulans* SidC A domains are consistent with the series of tandem duplication events predicted by relationships among the *C. heterostrophus* NPS2 C domains if losses of AMP2, AMP5, and AMP6 are invoked (iii). Relationships among SidC C domains are also consistent with a series of three tandem duplication events if loss of CON2 is invoked (iv). v) Relationships from partially ordered duplication histories mapped to the domain architecture with predicted domain losses shown in red.



relationship between A domains of modules 1 and 2 (bs = 76% and pp = 1.0) (Figure 2.5B, see Appendix 2.4Ci-ii). Both the ML and Bayesian trees for the C domains also support a strong relationship between modules 1 and 2 (Figure 2.5B, see Appendix 2.4Di-ii). The ML tree also groups C domains 1, 2 and 4 together and C domains 3 and 5 together, although there is poor bootstrap support for these relationships. The Bayesian tree was unresolved with respect to the remaining C domains. The relationships of A domains in the phylogeny of the complete dataset (Figure 2.3, see Appendix 2.3) suggest that the second A domain of the NPS1/SidC lineage corresponds to the third A domain (D.3) of the NPS2 lineage (Figures 2.2, 2.3 and 2.5). Thus, the NPS1/SidC lineage analyses also support a relationship between A domains corresponding to the first and third modules of the NPS2 lineage.

The unrooted ML phylogenies of A and C domains from *A. nidulans* SidC are shown in Figures 2.6Bi, Bii. When the tree of SidC C domains is rooted at position c (Figures 2.6Bii), and evaluated with the PDH algorithm [35], the resulting tree is a duplication tree which implies the partially ordered duplication history shown in Figures 2.6Bii. Similarly, the SidC A domains are duplication trees with an associated partially ordered duplication history (Figures 2.6Bi) that is also consistent with the duplication history predicted for SidC domains.

2.4.8 Adenylation Domain Substrate Choice

2.4.8.1 Structural Modeling

The experimental structure of Gramicidin GrsA [25] bound to its substrate, phenylalanine (1AMU_A), identified a number of residues that may be relevant for substrate specificity. In the GrsA structure, the binding pocket is formed by residues

at the interface between five β -strands (strand 1; D224 to F229, strand 2; T275 to P280, strand 3; Q296 to A301, strand 4; V317 to Y323 and strand 5; A332 to V336) of a β -sheet, two α -helices (helix 1; D203 to S217 and helix 2; D235 to L245) and at some of the loop regions connecting these secondary structure elements (Figure 2.7 A-C). In addition, a loop (S514 to K517) protruding from a small domain of the protein covers the entrance to the active site region (Figure 2.7B-C). A number of sites with the potential to be in direct contact with the substrate, as well as those lining the cavity in such a way that the side chain could affect the size of the binding pocket, were investigated in this work for a possible role in substrate specificity (Table 2.1). These key residue positions are 229, 230, 240, 243, 280, 320, and 326, plus those in the 10AA 'code' (235, 236, 239, 278, 299, 301, 322, 330, 331, and 517)(Figure 2.7C, Table 2.1). Position 229 was reported previously as part of the 13AA code predicted for the substrate AHO [6], but the additional residues we examined that are not in the 10AA code have not been implicated previously in substrate binding.

Two sites of key importance for binding amino acid substrates correspond to D235 and K517. In the GrsA structure, the carboxyl group of D235 interacts electrostatically with the amino group of the substrate residue (phenylalanine), providing one of the anchoring points for the substrate in the binding cavity, while K517 protrudes from a small domain (involving residues D430 to F530) that sits close to both the substrate as well as to the AMP binding pocket (Figure 2.7B) [9, 25]. Positively charged K517 appears to act as a gatekeeper, lying at the entrance of the active site cavity and projecting its NH_3 group toward the carboxyl group of the phenylalanine substrate [9, 25]. D235 and K517 are conserved across all A domains we examined and thus, though clearly important for substrate binding, should not be considered as residues involved in distinguishing among amino acid substrates (Table 2.1).

Table 2.2. Key positions in AMP domain identified by structural modeling

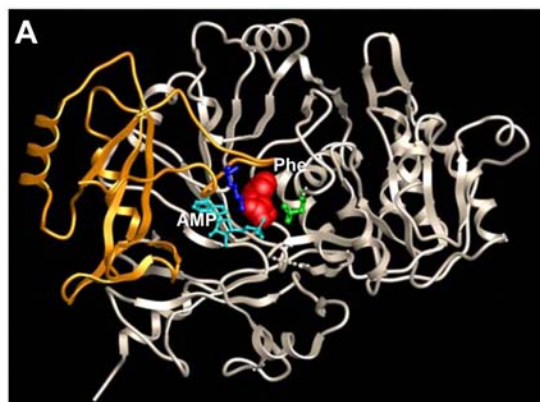
AMP domain ^a	Position ^b																Prediction	
	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	5	
	2	3	3	3	3	4	4	7	8	9	0	2	2	2	3	3	1	
	9	0	5	6	9	0	3	8	0	9	1	0	2	6	0	1	7	
IAMU_A	F	A	D	A	W	E	M	T	P	I	A	I	A	T	I	C	K	Phe
Spsib1 AMP1	F	A	D	V	F	E	G	E	T	I	I	V	A	T	I	H	K	G
ChNPS2 AMP1	F	A	D	V	F	E	F	E	T	L	I	W	M	T	I	H	K	G
FgNPS2 AMP1	F	A	D	V	F	E	F	E	T	L	I	W	M	T	I	H	K	G
FgNPS1 AMP2	L	S	D	V	Q	D	Y	H	T	T	I	Y	T	A	V	V	K	G
AnsidC AMP2	F	S	D	V	Q	D	Y	H	T	T	I	F	T	A	V	V	K	G
Umfer3 AMP2	F	S	D	V	Q	D	W	H	T	T	I	Y	T	A	V	V	K	G
ChNPS2 AMP3	Y	A	D	M	Y	D	L	D	T	Y	I	V	S	T	F	C	K	G
Umsid2 AMP1	Y	S	D	L	M	D	Y	L	T	I	G	L	L	A	L	I	K	G
ChNPS2 AMP2	A	C	D	V	F	E	F	S	T	V	A	Y	G	S	N	I	K	S
FgNPS2 AMP2	A	C	D	V	F	E	Y	S	T	V	A	W	G	S	N	I	K	S
AnsidC AMP1	F	A	D	P	M	E	V	M	T	W	M	V	A	T	I	N	K	S
Umfer3 AMP1	F	A	D	P	M	E	V	M	T	W	M	A	A	T	V	N	K	S
FgNPS1 AMP1	G	A	D	I	F	E	W	N	T	M	G	F	G	T	I	Y	K	A
Spsib1 AMP2	T	A	D	C	C	W	G	I	T	Y	Y	I	A	L	I	C	K	Degenerate
Spsib1 AMP3	F	A	D	V	L	E	F	D	T	I	G	Y	F	T	I	G	K	AHO
ChNPS2 AMP4	F	A	D	V	L	E	W	D	T	I	G	Y	G	T	I	G	K	AHO
FgNPS2 AMP3	F	A	D	V	L	E	W	D	T	I	G	Y	A	T	I	G	K	AHO
FgNPS1 AMP3	L	T	D	P	T	Q	V	G	V	T	G	F	F	T	I	G	K	AHO
AnsidC AMP3	Q	A	D	P	L	E	F	S	V	T	G	V	A	T	I	G	K	AHO
Umfer3 AMP3	L	A	D	V	S	Q	M	S	V	G	G	L	A	T	I	M	K	AHO
Umsid2 AMP2	R	S	D	V	L	E	L	C	V	I	G	L	A	S	I	G	K	AHO
Umsid2 AMP3	L	A	D	V	I	E	M	D	P	M	G	I	A	T	I	G	K	AHO

^a AMP domains in bold within blocks have highly similar residue sets.

^b Positions in bold correspond to the proposed 10 AA code. Position 229, in bold italics, corresponds to one of three additional positions (226, 229, 276) predicted by Schwecke et al. [6] to bind AHO. All other sites were identified in this study. Residues D and K at positions 235 and 517 in bold indicate residue conservation.

AHO and amino acid substrate assignments for A domains are shown in Table 2.2 and Figure 2.4. A domains of all terminal modules were predicted to code for AHO based on a larger binding pocket size with one or two negatively charged residues or a few polar residues (Table 2.2, Figure 2.7E, compare with Figure 2.7D). Besides these features, there was no clear pattern of residues lining the cavity, except for similarity among Spsib1, ChNPS2 and FgNPS2 terminal A domain residues (Table 2.2).

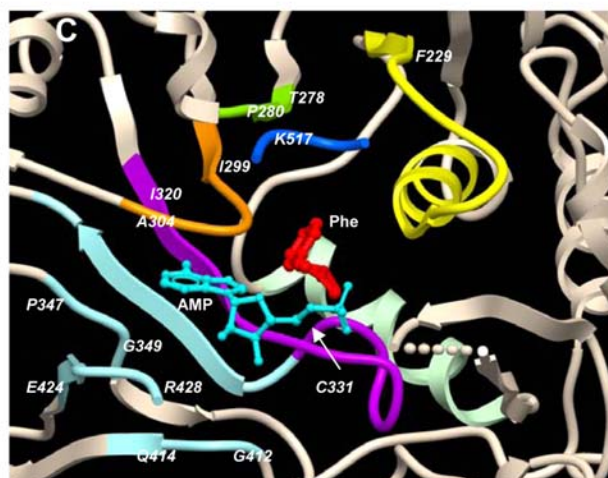
Figure 2.7. 3D modeling of selected NRPS AMP binding domains. **A.** Ribbon representation of the structure of the activated domain of Gramicidin synthetase (PDB code: 1AMU) bound to its Phe substrate (shown as a CPK model; red) and adenosine monophosphate (AMP; shown as “*ball & stick*” representation of the heavy atoms; light-blue). The large domain (gray ribbon), contains the substrate and AMP binding pockets. A second smaller domain (orange), involving residues D430 to F530, sits at the entrance of these pockets. “*Ball & stick*” representations of residues D235 and K517 are shown in green and blue, respectively. **B.** View of the GrsA binding pockets for Phe and AMP showing the positions of the conserved residues F234 (yellow), D235 (green), and K517 (blue). D235 and K517 are in contact with the amino and carboxyl end groups, respectively, of the Phe substrate. **C.** Alternative view of GrsA highlighting all the fragments of the sequence that determine the binding pockets for Phe and AMP. The amino acid composition of those fragments is listed to the right. The color convention for the residues is as follows: red and orange indicate those residues lining the substrate cavity, with residues in red making contact with the substrate Phe in the experimental structure; blue and light blue indicate residues lining the AMP binding site, with residues in blue making contact with AMP in the experimental structure. **D.** Slice through the substrate binding site of a 3D model of ChNPS2 module 3. The central cavity is packed with large residues that produce a shallow pocket. A *ball & stick* representation of a bound GLY residue is also shown to help assess the size of the cavity (compare to Figure 2.7E). **E.** Slice through the substrate binding site of a 3D model of ChNPS2 module 4. The central cavity is lined with small residues that leading to a deep pocket. A *ball & stick* representation of a bound AHO is also shown to help assess the size of the cavity (compare to Figure 2.7D).



1AMU



1AMU



1AMU binding pocket

Fragments defining the substrate binding site

Fragment 1: F229-M243

FASISFDASVWEMFM

Fragment 2: T278-P280

ITLPP

Fragment 3: I299-A304

ITAGSA

Fragment 4: I320-C331

INAYGPTETTIC

Fragments defining the AMP binding site

Fragment 5: P347-G349

PIG

Fragment 6: G412-Q414

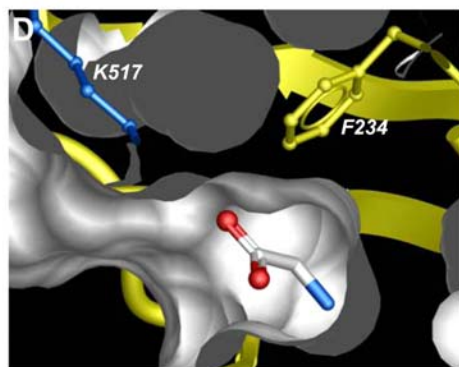
GDQ

Fragment 7: E424-R428

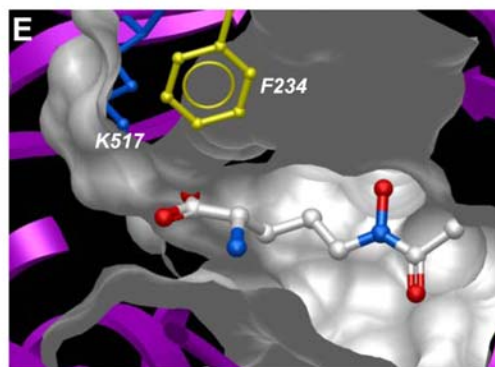
EYLGR

Fragment 8: N515-K517

NGK



GLY bound to ChNPS2 module 3



AHO bound to ChNPS2 module 4

Assignment of the remaining A domains was even more difficult. We found that the consensus 10AA codes for SER, ALA, and ORN identified by Stachelhaus et al. [9] were not represented in the A domains of ferrichrome synthetases we examined and thus we could not simply infer specificity. Initially, to search for patterns representative of A domains binding SER, ALA, GLY, and ORN, structural alignments of A domains predicted [47] [9] to bind these substrates were created (see Appendix 2.5). The small number of fungal and bacterial domains confirmed to be associated with known substrates makes comparing key fungal positions to the bacterial code positions problematic. We found, however, that bacterial A domain 10AA ‘codes’ for the same substrate appeared more conserved than fungal ones. The fungal A domains were either too variable or too few for us to deduce a consensus ‘code’ (see Appendix 2.5). We did not find any consistent pattern associated with A domains coding for ALA, GLY, or ORN. For SER, however, we found that the majority of sequences share a histidine (HIS) residue at position 278 that our 3D-models suggest is projecting from the top of the binding pocket (Table 2.2). A domains from FgNPS1, AnsidC, and Umfer3 module 2, have HIS at 278, and their cavities are quite hydrophilic and lined by similar sets of residues (Table 2.2). We initially considered these modules as the domains most likely to bind SER. We also found that A domains from Spsib1, ChNPS2, and FgNPS2 module 1 share highly similar binding pockets (Table 2.2), with a HIS at position 331 whose side chain may occupy the center of the cavity (i.e., similar to H278 in our structural alignment) but projecting from the bottom of the pocket), making them, by analogy, also probable candidates to bind SER. The chemistry, however, indicates that Spsib1 produces ferrichrome which contains three glycines and no serine (Figure 2.4). Therefore, we infer that the A domain of the Spsib1 module 1 must bind GLY, since it is the only non degenerate A domain, other than the terminal A domain which we predict binds

AHO (Figure 2.1, 2.4). Due to the high similarity of the residues forming the AMP cavity of ChNPS2 and FgNPS2 module 1 to those in Spsib1 module 1 (Table 2.2), we predict these two domains are also likely to bind GLY. By default, module 2 of FgNPS2 is predicted to bind SER (Figure 2.4). Based on similarities to the FgNPS2 module 2 binding pocket, ChNPS2 module 2 is predicted to bind SER also (Table 2.2). Finally, ChNPS2 module 3, which 3D models show has a very crowded and small binding pocket is expected to bind to GLY (Table 2.2, Figure 2.4, Figure 2.7D).

AnSidC has been shown to produce ferricrocin [17, 48], which contains two glycines and one serine, while FgNPS1 produces malonichrome containing two glycines and a single alanine (G. Adam, BG Turgeon, unpublished) and Umfer3 makes ferrichrome composed of three glycines [7]. As noted in Table 2.2, key residues in the binding pockets of the second A domains of FgNPS1, AnSidC, and Umfer3 are highly similar to each other and should likely code for a residue that is common between ferricrocin and malonichrome (i.e., GLY). By default, we infer that module 1 of AnSidC and Umfer3 bind SER (Table 2.2) while module 1 of FgNPS1 binds ALA. 3D modeling shows that the center of these binding pockets are likely filled by many hydrophobic residues. In the case of module 1 of AnSidC and Umfer3, the characteristics of the binding pockets (i.e., highly hydrophobic) do not seem very compatible with binding a hydrophilic residue such as SER. However, an asparagine residue at position 331 in both modules may be able to provide a hydrogen-bond partner to “dock” the side chain of the SER substrate. Lastly, 3D models of Umsid2 module 1, indicate that the binding region must be filled with many hydrophobic residues (Table 2.2) leading to a very shallow pocket, likely to be selective for GLY.

Thus, we found that the 10AA code failed when we tried to *infer* the specificity of the sequences we examined. Instead, A domains predicted to code for the same substrate [e.g., ChNPS2 AMP1 (GLY) and AnSidC AMP2 (GLY)] had widely

divergent ‘codes’ (Table 2.2, see Appendix 2.5) and appeared to diverge according to our A domain phylogeny (e.g., ‘codes’ for GLY, SER, or ORN are conserved among members of the NPS2 and SidC lineages but differ between the two lineages) (Table 2.2, Figure 2.3, see Appendix 2.3). It is noteworthy that, even when protein structural modeling is brought to bear on the issue of key residues ‘coding’ for substrate specificity, no simple rule was found to be applicable to all sequences considered in this study. While it was possible to infer the size and some properties that characterize the binding pockets, highly divergent residue arrangements appear to bind the same substrate (Table 2.2, see Appendix 2.5).

2.4.8.2 Evolutionary Approaches to Identification of Specificity Residues

The SDP, Type I and Type II functional divergence analyses identified, with high probability, a number of positions indicating either a shift in amino acid properties between clusters (SDP and Type II) or a shift in evolutionary rate between clusters reflective of changes in evolutionary constraint or selective pressure (Type I) (Table 2.3). For Type I analyses, all comparisons of paralogous clusters showed θ_I values significant at $p \leq .05$ while for Type II analyses, only comparisons between NPS2 AMP1 and NPS2 AMP 4 ($\theta_{II} = .224 \pm .113$) and between NPS2 AMP2 and NPS2 AMP4 ($\theta_{II} = .283 \pm .113$) were significant at $p \leq .05$. Several positions received high support from all three methods including positions 252, 278, 301, 322, and 331. Several of the positions identified by structural modeling (230, 239, 243, 278, 299, 301, 320, 322, 326, 330, and 331) also received support from at least one method (Table 2.2, Table 2.3). Clusters of significant residues map to the first and second α -helices and to β -strands 2-4, as well as to fragments 1-4 identified by structural modeling as lining the 1AMU_A binding pocket and connecting these key structural

Table 3. Residues showing evidence of functional divergence in SDP and DIVERGE2 analyses.

Position 1AMU_A	Feature	Fragments (Fig. 7)	SDP	Type II	Type I	Position 1AMU_A	Feature	Fragments (Fig. 7)	SDP	Type II	Type I	
200				.95		270					.96	
201				.78		271						
202						272						
203	α-Helix 1 D203-S217				.95	273						
204				.85	.95	274			5.41		.75	
205						.90	275			.95		
206				6.56	.95		276	β-Strand 2 T275-P280		6.26	.88	
207					.80	277			6.44			
208				.77	.87	278			x fragment 2	9.43	.90	.87
209						279			T278-P280			.97
210						280		x				
211			5.5			281						
212					.94	282			7.2			
213					.94	283			6.54		.93	
214			10.7	.95		284					.92	
215						285						
216						286						
217						287						
218						288						
219						289			6.68			
220						290						
221						291						
222						292						
223						293						
224	β-Strand 1 D224-F229					294						
225						295						
226				9.16			296	β-Strand 3 Q296-A301			.81	
227							297					.80
228						298					.98	
229		x fragment 1				299			x fragment 3	7.88	.75	
230		x F229-M243			.90	300		I299-A304		.84		
231			6.76	.74		301		x	10.82	.95	.74	
232			8.92	.75		302				.75		
233				.95		303						
234						304			5.94		.94	
235	α-Helix 2 D235-L245	x				305						
236			x			306			6.71	.95		
237						.90	307				.97	
238							308			.70	.81	
239		x	8.59		.95	309				.77		
240		x				310			7.02		.98	
241			6.00	.95		311						
242						312					.97	
243		x	5.64		.88	313			8.31			
244			5.58		.94	314						
245						315						
246					.95	316						
247				.87	.99	317	β-Strand 4 V317-Y323					
248					.85	318						
249					.98	319						
250					.98	320			x fragment 4			.78
251					.99	321		I320-C331	6.46	.76		
252			5.99	.85	.93	322		x	8.25	.90	.90	
253						323						
254				.70	.93	324						
255						325						
256				.73	C	326		x	5.98			
257					.97	327				.80		
258						328						
259			5.47			329				.95		
260						330		x		.95	.71	
261						331		x	12.46	.80	.94	
262						332	β-Strand 5 A332-V336				.90	
263						333						
264						334						
265						335						
266						336						
267												
268						517		x				
269												

Left to right columns: 1) positions in 1AMU_A, bold are sites corresponding to the 10 or 13 AA code. 2) Loops and strands in 1AMU_A (Fig. 7A). 3) Fragments defining the substrate binding site; 'x' indicates key sites identified by structural modeling (Fig. 7C, Table 2). 4) Sites identified using the SDP algorithm showing significant Z-scores. 5), 6) Sites identified using tests for Type II and Type I functional divergence, respectively. The highest posterior probability for sites above a .70 cutoff for any of the pair-wise comparisons with a significant Θ_I and Θ_{II} value are shown. All amino acid changes for Type II divergence are radical, indicating a change in amino acid properties; the single exception is indicated with 'C'.

features (Table 2.3; Figure 2.7C). Two exceptions to this pattern map to region 246-257 which is on β -strand near the surface of the protein (therefore not located close to the substrate binding site) and 1AMU_A containing both the substrate and AMP-binding pockets. Thus, residues predicted to be involved in functional divergence point to many of the same key regions of the binding pocket predicted by structural modeling to have a potential role in substrate specificity.

2.5 Discussion

2.5.1 Distinct Lineages of Ferrichrome Synthetases

Our phylogenetic analyses support the hypothesis that fungal ferrichrome synthetases fall into two distinct lineages corresponding to homologs of *C. heterostrophus* NPS2 and *A. nidulans* SidC. Some fungi contain representatives of both lineages while others lack a ferrichrome synthetase altogether. Significantly, ferrichrome NRPSs were not detected in any yeast species sampled (except the fission yeast, *S. pombe*), or in the zygomycetes *R. oryzae* and *P. blakesleeanus*, the ectomycorrhizal fungus *L. bicolor* or the chytrid *B. dendrobatitis*. While absence of a gene must be interpreted with caution, as genome sequences may be incomplete, the lack of the NPS1/SidC lineage in all Dothideomycetes (*C. heterostrophus*, *A. brassicicola*, *S. nodorum*, and *A. pullulans*) and Onygenales (*C. immitis*, *H. capsulatum*, and *U. reesii*), lack of the NPS2 lineage in Eurotiales (*Aspergillus* sp.), as well as a lack of any ferrichrome synthetase in all hemiascomycete yeasts, zygomycetes, or chytrids surveyed is likely significant.

The NPS1/SidC lineage predates the divergence of ascomycetes and basidiomycetes as its members are present in both of these groups. In contrast, the

duplication into the two main NPS2 and NPS1/SidC lineages may have occurred in the ancestor of ascomycetes as the former lineage is only found within ascomycetes. The additional duplications within the NPS1/SidC lineage may have occurred also prior to the divergence of ascomycetes and basidiomycetes, as there are two distinct ferrichrome synthetase encoding genes from the NPS1/SidC lineage in both the basidiomycete *U. maydis* (Umfer3 and Umsid2) and the ascomycete *B. cinerea* (BC1G10928 and BC1G15494). This scenario would postulate an unlikely loss of one or the other of these genes in the majority of species examined. The other possibility is independent duplication of the NPS1/SidC type gene in certain species e.g., *U. maydis* and *B. cinerea*. However, in both ML and Bayesian phylogenetic analyses, the ascomycete proteins *B. cinerea* BC1G15494 and *F. graminearum* FG11026 grouped with, or outside of, basidiomycete proteins, suggesting an ancestral duplication of this lineage (Figure 2.3, see Appendix 2.3).

It is possible that the duplications within the NPS1/SidC lineage may be associated with production of different ferrichromes. *F. graminearum* NPS1 (FG11026), has recently been shown to produce malonichrome (two GLY, one ALA) (G Adam, BG Turgeon, unpublished) while certain other ascomycete members (e.g., *A. nidulans* SidC) of the NPS1/SidC lineage produce ferricrocin (two GLY, one SER). The two ferrichrome synthetases in *U. maydis* also produce distinct products; Umfer3 produces ferrichrome A (two SER, one GLY) and Umsid2 produces ferrichrome (3 GLY).

2.5.2 Evolution of Domain Architecture

In some respects, the C domain alone or in combination with the T domain can be considered the minimal evolutionary unit for NRPSs, as T-C units clearly occur in

the absence of A domains. T-C units may also be considered the minimal functional units for NRPS synthesis as they can be charged by nonadjacent A domains [4, 6, 17, 48, 49]. T-C units lacking an associated A domain could be created either through independent duplication of T-C units or through loss of an associated A domain from a complete A-T-C module. If complete A-T-C module repeats arise by tandem duplication, the C domain phylogenies may provide a more complete picture of the evolutionary history of duplications at the locus. The relationships observed between C domains of modules 3 and 5 and among modules 2, 4, and 6 of the NPS2 lineage (Figure 2.5A) and the partially ordered duplication history predicted by *C. heterostrophus* NPS2 C domains (Figure 2.6Aii) imply a series of tandem duplication events involving single or double complete A-T-C units as a possible hypothesis for the evolution of a hexamodular ferrichrome synthetase NRPS (Figure 2.6Aiv, Figure 2.8). These events would occur as follows: Step 1) duplication of module 1 to form a bimodular gene, Step 2) duplication of the bimodular gene (modules 1 and 2) to form a tetramodular gene (modules 1-4), and Step 3) duplication of modules 3 and 4 to form a hexamodular gene (modules 1-6) (Figure 2.8A, 2.6A).

These interpretations are based on algorithms for which it is assumed that there is no loss and no recombination, criteria that are clearly violated here for ferrichrome synthetases. We propose, however, that the C domains of *C. heterostrophus* NPS2 likely represent the full evolutionary history of ferrichrome synthetase modules. The chemical structure of ferrichromes (3 AA and 3 AHO) provides support for the notion of an ancestral gene with six complete modular units. Furthermore, our analyses (unpublished) and others [4] show little evidence for recombination within C domains. The tandem duplication hypothesis is based on these assumptions and is presented as one possible explanation for the diverse domain architectures. The phylogenetic relationships observed among A and C domains in both lineages are consistent with

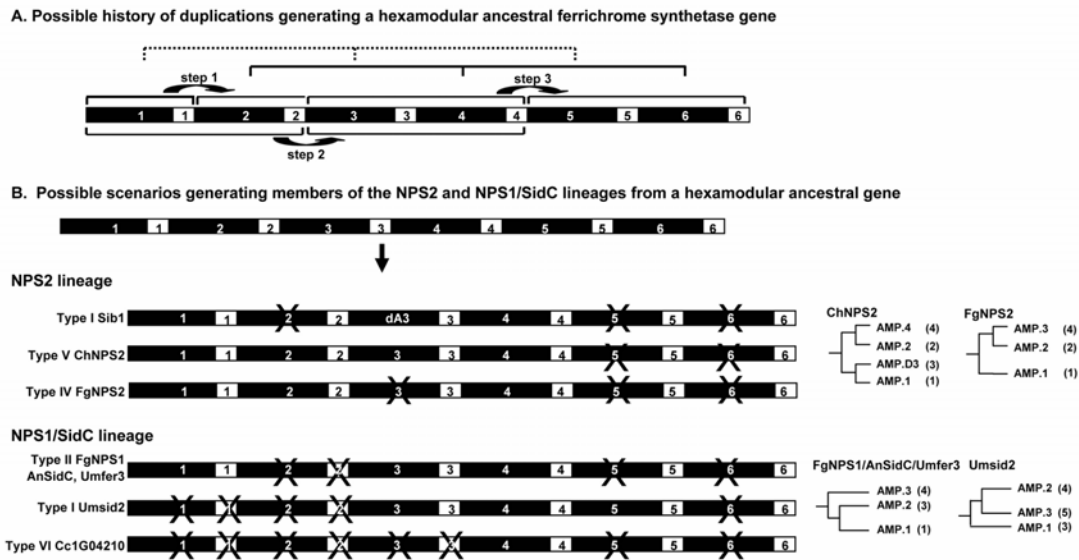


Figure 2.8. Models for evolution of a hexamodular ancestral ferrichrome synthetase gene and for generation of domain architectures of the extant types examined in this study. **A.** Possible origin of a hexamodular ancestral ferrichrome synthetase gene. We propose that a hexamodular gene arose by a series of duplication events. Step one: module 1 duplicates, forming module 1 and new module 2. Step two: modules 1 and 2 duplicate together, forming modules 1 and 2, and new modules 3 and 4. Step three: modules 3 and 4 duplicate together, forming modules 3 and 4, and new modules 5 and 6. This scenario predicts that modules 1, 3, and 5 (dotted lines) will show greater similarity to each other than to other modules. Similarly, modules 2, 4, and 6 (solid lines) will show greater similarity to each other than to modules 1, 3 and 5. **B.** Possible scenarios generating members of the NPS2 and NPS1/SidC lineages from a hexamodular ancestral gene. Trees to the right show relationships of extant AMP domains, based on Fig. 3. Numbers in parentheses indicate corresponding domain of hypothetical ancestral gene. Left side of figure indicates proposed losses of A (black boxes) or C (white boxes) domains, resulting in the extant gene.

this proposed tandem duplication history if one postulates the loss of module 5 and 6 A domains from both lineages and the additional loss of the complete module 2 (A-T-C) from the SidC lineage (Figures 2.6A iii-v and 2.6B iii-v with losses shown in red, Figures 2.4A, 2.4B). If these duplications occurred before the divergence of the majority of species examined, as supported by the reconciliation analysis, this scenario predicts that domains of modules 1, 3, and 5 (Figure 2.8A, top, dotted lines, Figures 2.6Av and 2.6Bv) will show greater similarity to each other than to other modules, as

will modules 2, 4, and 6 (Figures 2.8A, top, solid lines, Figures 2.6Av and 2.6Bv).

In general, these predictions are supported when the relationships of A or C domains from each lineage are examined. In particular, the relationships between modules 3 and 5 and between 4 and 6, which would have resulted from the final duplication are more strongly supported (Figures 2.4A, B, see Appendix 2.4A-D). The results are not consistent with recent independent duplication of T-C units giving rise to the final T-C repeat in most ferrichrome synthetases (Figure 2.2) as this latter mechanism would predict a closer relationship among C domains of modules 4, 5, and 6 which is not supported by C trees from either lineage. Instead, our analyses support the hypothesis of a hexamodular ancestor with six complete A-T-C modules, proposed previously by Schwecke [6], followed by loss of either complete A-T-C modules or individual A domains as the best hypothesis for the generation of the diverse domain architectures of the six ferrichrome synthetase domain structural types (Figure 2.8). In the NPS2 lineage, for example, both *C. heterostrophus* (representative of Type V) and *F. graminearum* (representative of Type IV) have 6 C domains, although they have only 4 and 3 A domains, respectively. Analyses of C domains of these proteins clearly indicate that the second C domain of Types V and IV are related (Figure 2.4, see Appendix 2.4A-D). The same is true for the third C domains. The difference in protein architecture in this region is presence/absence of an A domain between C2 and C3 (i.e., the *F. graminearum* gene appears to be missing the third A domain found in the *C. heterostrophus* protein). Similarly, the second C domain in *sib1* from *S. pombe* groups with the second C domain in *C. heterostrophus* NPS2 but lacks the corresponding A domain (Figure 2.4), suggesting loss of this domain in the *S. pombe* gene. Our data thus suggest that differential loss of A domains in different members of this lineage has resulted in the three distinct domain architectures. A recent study of the microcystin synthase gene cluster has shown recombination breakpoints within

NRPS A domains suggestive of recurrent A domain replacement [4]. Our analyses suggest that homologous recombination could also lead to complete loss of A domains.

For the NPS1/SidC lineage, *F. graminearum* NPS1, *A. nidulans* SidC and *U. maydis* fer3 all have 5 C domains and 3 A domains. A and C domain analyses of this lineage clearly indicate that there is a one to one relationship for all A and all C domains (Figure 2.4, see Appendix 2.4C-D). Examination of Umsid2, however, indicates that it has 3 A domains, but only 4 C domains; the module 1 A domain is related to module 2 A domains of the other members of this group, while both module 2 and 3 A domains are related to the C-terminal modules of other proteins in this lineage. Similarly the C domains from Umsid2 modules 2, 3, 4 are related to the C domains of modules 3, 4, 5 of the rest of the NPS1/SidC lineage. Umsid2 lacks the complete N-terminal A-T-C module of other NPS1/SidC members and retains the A domain corresponding to the module 4 C domain that our scenario postulates has been lost in other members of this lineage.

These data thus support the hypothesis [6] that the extant genes may have evolved from a hexamodular (A-T-C) ancestor and that repeated and independent losses of A domains or complete A-T-C modules may have given rise to the diverse domain architecture types observed in extant species.

2.5.3 Domain Architecture and Mechanism of Biosynthesis

How do ferrichrome synthetases differing in domain architecture, biosynthesize nearly identical chemical products? Several authors have suggested that T-C repeats can be used iteratively [17, 48, 49]. For example, Schwecke et al [6] have proposed a mechanism by which the functions of the missing *S. pombe* sib1 A domain

(which should accompany the second T-C) and the degenerate second A domain (Figure 2.2) are assumed by the first A domain, which charges both the second and third C domains in *cis*, thus attaching the three glycines required for the ferrichrome product. Similarly, some of the NPS2 lineage Type IV synthetases are predicted to make ferricrocin which contains two glycines and one serine. We speculate that the first A domain of this protein is used iteratively to attach two glycines by charging the T-C repeat after the second complete module. *U. maydis* *sid2* has only a single A domain predicted to code for glycine yet ferrichrome contains three glycines. Therefore, the first A domain must also be used iteratively. Similarly, the last A domain of Types II-V may also charge the final two T-C units at the C terminal ends of these proteins to assemble the three AHO groups that form the core iron binding group, common to all ferrichrome synthetases [6]. Interestingly, the *U. maydis* *sid2* protein, which has only a single terminal T-C, contains two complete A-T-C modules predicted to charge AHO. This protein thus must utilize an alternate mechanism to produce the three required AHO units and perhaps represents an intermediate step between a hexamodular ancestral gene with three complete A-T-C modules coding for AHO and a completely iterative system with a single A-T-C module coding for AHO followed by a T-C repeat that is used iteratively. Thus, loss of A domains in these NRPSs is compensated, likely, by iterative charging of T-C units. Type VI *C. cinerea* CC1G04120 is unusual in that it has only a single A domain and a T-C repeat. It is possible that this gene is incomplete due to assembly errors, or may function together with another NRPS to form the complete ferrichrome product. Alternatively, it may produce a product such as desdiserylglycerylferrirhodin (DDF) which consists of three AHO residues only.

The mechanisms controlling iterative use of NRPS domains are, to our knowledge, unknown. Here we observe that proteins with distinct domain

architectures produce nearly identical chemical products. Iterative synthesis provides yet another flexible mechanism for NRPS biosynthesis.

2.5.4 Substrate Specificity

Structural modeling results suggest that general features of the binding pocket such as size, hydrophobicity, and charge may be more important in determining substrate recognition than residues at fixed positions within the cavity. In homology based modeling of substrate specificity, small errors in the alignment between the experimental and the model sequence can lead to significant errors in the modeled structure. For this reason, we used an alignment of several experimental structures to optimize our alignments. We found that the A domains included in this study were remarkably conserved structurally and we were able to identify several conserved residue-patterns and structural features which aligned well in all the structures and served as markers to anchor our alignment of the experimental sequences, particularly near the residues that are supposed to form the wall of the binding site (the code). With careful attention to the alignment, we found that residues associated with the 10 or 13 AA ‘codes’ predicted to be important in substrate choice vary considerably and do not show a consistent pattern for A domains predicted to code the same substrate (Table 2.2, see Appendix 2.5). Thus, we found that the string of amino acids at the proposed ‘code’ positions was unable to predict substrates for any fungal A domain examined in this study. The 10AA code was originally deduced by extracting residues at positions predicted to interact with the Phe substrate in the 1_AMU_A domain from a multiple sequence alignment and is based on the assumption that, because A domains of NRPSs and other adenylyating enzymes show high structural similarity, the positions in the 1_AMU_A structure should be important for other substrates [9].

Recent studies, however, have shown that additional residues may be important for interacting with other substrates such as AHO [5, 6].

Our results from structural modeling and evolutionary analyses of functional residues point to key fragments within the binding pocket which surround and connect the α -helix and β -strand structural elements of the pocket, as general regions important for specificity. Our analyses also identified residue positions in addition to the 10 AA code positions within these fragments (229, 230, 240, 243, 280, 320, 322, and 326) which line the substrate pocket and are either positioned such that their side chains may interact with a substrate or are involved in shaping the size of the binding pocket (Table 2.2). Our study confirms [9] that D and K residues at positions 235 and 517 respectively (Table 2.2), adjacent to the N-terminal amino and C-terminal carboxyl groups, are conserved across all the sequences examined, and that they serve the general function of holding the amino and carboxyl groups of an amino acid substrate in the binding pocket and are not involved in recognition of a specific amino acid substrate.

We speculate that the residue positions showing a significant signal for functional divergence which fall outside of the binding pocket region on the surface of the protein (246-257 and 305-314) could have a role in either protein-protein interactions or interactions between the two subunits of the NRPS protein. One subunit contains both the substrate and AMP binding pockets while the other subunit covers the opening to the binding sites (Figure 2.7A). In the crystal structure of the related adenylating enzyme, acetyl CoA synthetase (1PG3_A), this second subunit may adopt two configurations in order to accomplish the two half-reactions of this enzyme: 1) adenylation of the substrate and 2) subsequent transfer to coenzyme A. Each configuration exposes a different set of residues to the active site [50, 51]. A similar mechanism may operate in NRPSs. Residues 305-314 on the surface of the first

subunit are not in a position to interact directly with the binding pocket, but could be involved in mediating interactions between the two subunits.

Thus, our results suggest that a rigid ‘code’ of specific amino acids at particular residue positions may not be the most reliable approach to predicting specificity of fungal NRPS A domains. Instead, the general chemical, physical, and structural features of the binding pocket may be more important. We conclude that methods of substrate prediction which evaluate chemical features of amino acids within these key regions may be better able to predict substrate specificity. Our findings await manipulation of key residues predicted to affect the chemical properties of the binding pocket, followed by examination of how this affects substrate choice.

2.6 Conclusions

Our results demonstrate two distinct lineages of ferrichrome synthetases in fungi and suggest that these genes are restricted to fission yeast, filamentous ascomycetes, and basidiomycetes. Phylogenetic analyses of domain architectures supports the hypothesis that the distinct domain architectures observed derive from a hexamodular ancestral gene through loss of individual A domains or complete A-T-C modules and support a series of tandem duplication events of single or double A-T-C modules as the mechanism generating this hexamodular ancestor.

Analyses of substrate specificity show that the proposed 10AA code was unable to infer substrate specificity for these fungal A domains. Structural modeling and evolutionary analyses of functional residues suggest that additional positions may play a role in substrate specificity. Our results confirm that two positions of the code which are conserved across all sequences examined, D235 and K517, likely do not play a role in amino acid substrate choice but instead serve the important function

of anchoring the substrate in the binding pocket through interaction with the amino and carboxyl groups respectively.

Authors' contributions

KEB selected and performed most of the phylogenetic analyses, DRR performed the protein structural modeling, BGT directed the research. KEB and BGT wrote the manuscript, with input from DRR.

Acknowledgements

BGT acknowledges the support of the Division of Molecular and Cellular Biosciences, National Science Foundation, the USDA Cooperative State Research Education and Extension Service, National Research Initiative and the BARD foundation. KEB thanks Jeff Doyle and Scott Kroken for discussion and comments on the manuscript and Conrad Schoch, Henk DeBakker, Dave Schneider and the Cornell Computational Biology Service Unit for computer resources and/or assistance in running Bayesian analyses.

APPENDIX 2.1

Appendix 2.1: Protein accession numbers used in this study

Species	GenBank Acc#	Genome ID ^a	Protein Name ^b	Ref.
<i>Alternaria brassicicola</i>	ABU42595	AB44259	NPS2	[18]
<i>Aspergillus fumigatus</i>	EAL92059	Afu3g15270	NRPS7	[23]
	EAL86616	Afu3g03350	NRPS3	[23]
	EAL91050	Afu3g17200	NRPS2 (SidC)	[23],
				[17]
<i>Aspergillus nidulans</i>	XP_753088	AN0607.3	SidC	[17]
<i>Aspergillus niger</i>	XP_001390952.1	Aspni1_207636	hypothetical protein	
<i>Aspergillus oryzae</i>	BAE59066	AO9002300528	Sid2/NRPS36	[23]
<i>Aspergillus terreus</i>	XP_001212122.1	ATEG02944.1	NRPS83	[23]
	XP_001217069.1	ATEG08448.1	NRPS82	[23]
	XP_001214251.1	ATEG05073.1	NRPS71/ SidC	[23]
<i>Aureobasidium pullulans</i>	AAD00581		peptide synthetase	[6]
<i>Botrytis cinereus</i>	XP_001550755.1	BC1G10928.1	hypothetical protein	
	XP_001557929.1	BC1G03511.1	hypothetical protein	
	XP_001546022.1	BC1G15494.1 ^d	hypothetical protein	
<i>Chaetomium globosum</i>	XP_001228767	CHGG02251.1	hypothetical protein	
	XP_001226019.1	CHGG10752.1	hypothetical protein	
	XP_001230007.1	CHGG03491.1	hypothetical protein	
	XP_001227470.1	CHGG09543.1	hypothetical protein	
<i>Coccidioides immitis (RS)</i>	XP_001247170.1	CIMG00941.1	hypothetical protein	
<i>Cochliobolus heterostrophus</i>	AAX09983		NPS1	[3]
	AAX09984		NPS2	[3]
	AAX09985		NPS3	[3]
	AAX09986		NPS4	[3]
	AAX09987		NPS5	[3]
	AAX09988		NPS6	[3]
	AAX09989		NPS7	[3]
	AAX09990		NPS8	[3]
	AAX09994		NPS9	[3]
	AAX09992		NPS10	[3]
	AAX09993		NPS11	[3]
	AAX09994		NPS12	[3]
	AY884198		NPS13	[3]
<i>Coprinus cinerea</i>	EAU88504.1	CC1G04210.1	hypothetical protein	
<i>Fusarium graminearum</i>	XP_391202.1	FG11026.2	NPS1	[18]
	XP_385548.1	FG05372.2	NPS2	[52]
<i>Histoplasma capsulatum</i>	XP_001544796.1	HCAG01843.1	hypothetical protein	
	XP_001538006.1	HCAG07428.1 ^{c,d}	hypothetical protein	
	XP_001538007.1	HCAG07429.1 ^{c,d}		
<i>Magnaporthe grisea</i>	XP_001407762.1	MGG12175.3	SSM1	[53]
<i>Neurospora crassa</i>	XP_960302	NCU07119.2	putative intracellular siderophore NPS	[54]
			fso1	[6]
<i>Omphalotus olearius</i>	AAX49356		fso1	[6]
<i>Schizosaccharomyces pombe</i>	CAB72227		sib1,SPAC23G3.02c	[6]

Appendix 2.1 Continued

<i>Sclerotinia sclerotiorum</i>	XP_001593263.1	SS1G06185.1	hypothetical protein	
	XP_001595604.1	SS1G03693.1	hypothetical protein	
<i>Stagonospora nodorum</i>		SNU02134.1	hypothetical protein	
<i>Trichoderma reesii</i>		69946 (JGI)	hypothetical protein	
<i>Uncinocarpus reesii</i>		UREG00890.1 ^c	hypothetical protein	
		UREG00891.1 ^c		
<i>Ustilago maydis</i>	XP_757581.1	UM01434.1	fer3	[55]
	AAB93493	UM05165.1	sid2	[48]
<i>Erwinia carotovora</i>	YP_049592		nonribosomal peptide	
<i>subsp. atroseptica</i>			synthetase	

^a Source as indicated in Materials and Methods

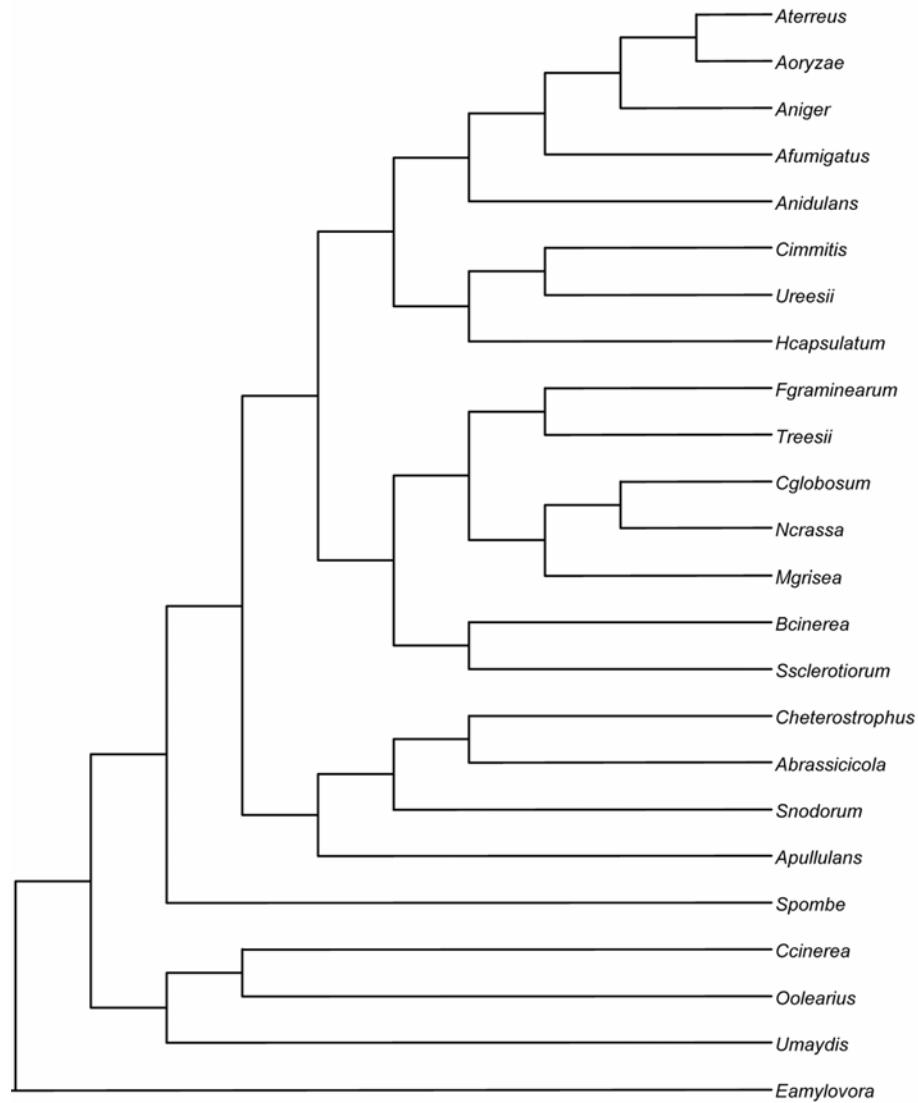
^b Common name

^c Two genes reannotated as a single gene

^d Incomplete gene

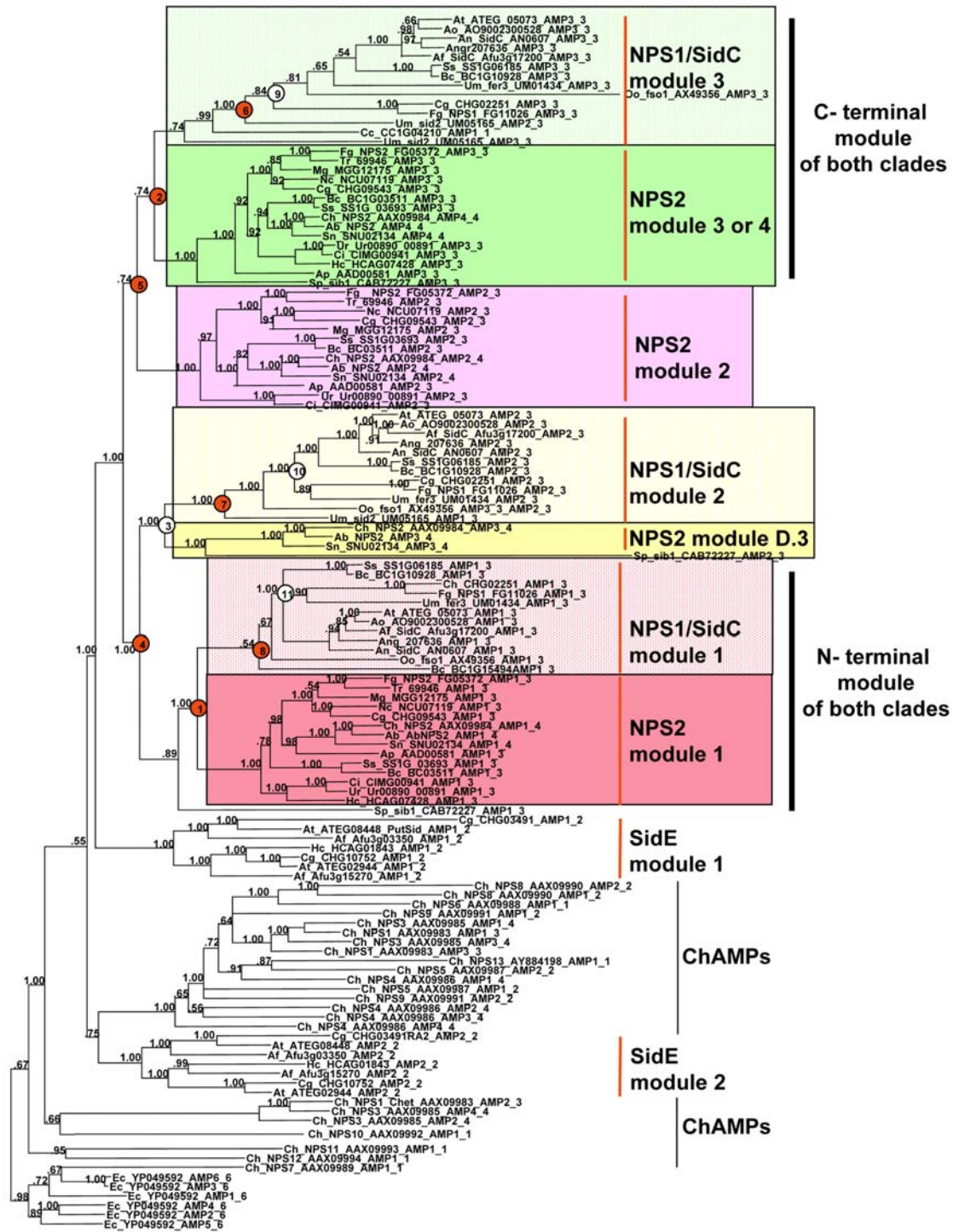
APPENDIX 2.2

Appendix 2.2. Species tree. Tree used for reconciliation analyses was adapted from four recent phylogenetic studies (See Materials and Methods). Dothideomycete taxa were placed as sister to other filamentous ascomycetes in the subphylum Pezizomycotina (see Materials and Methods).



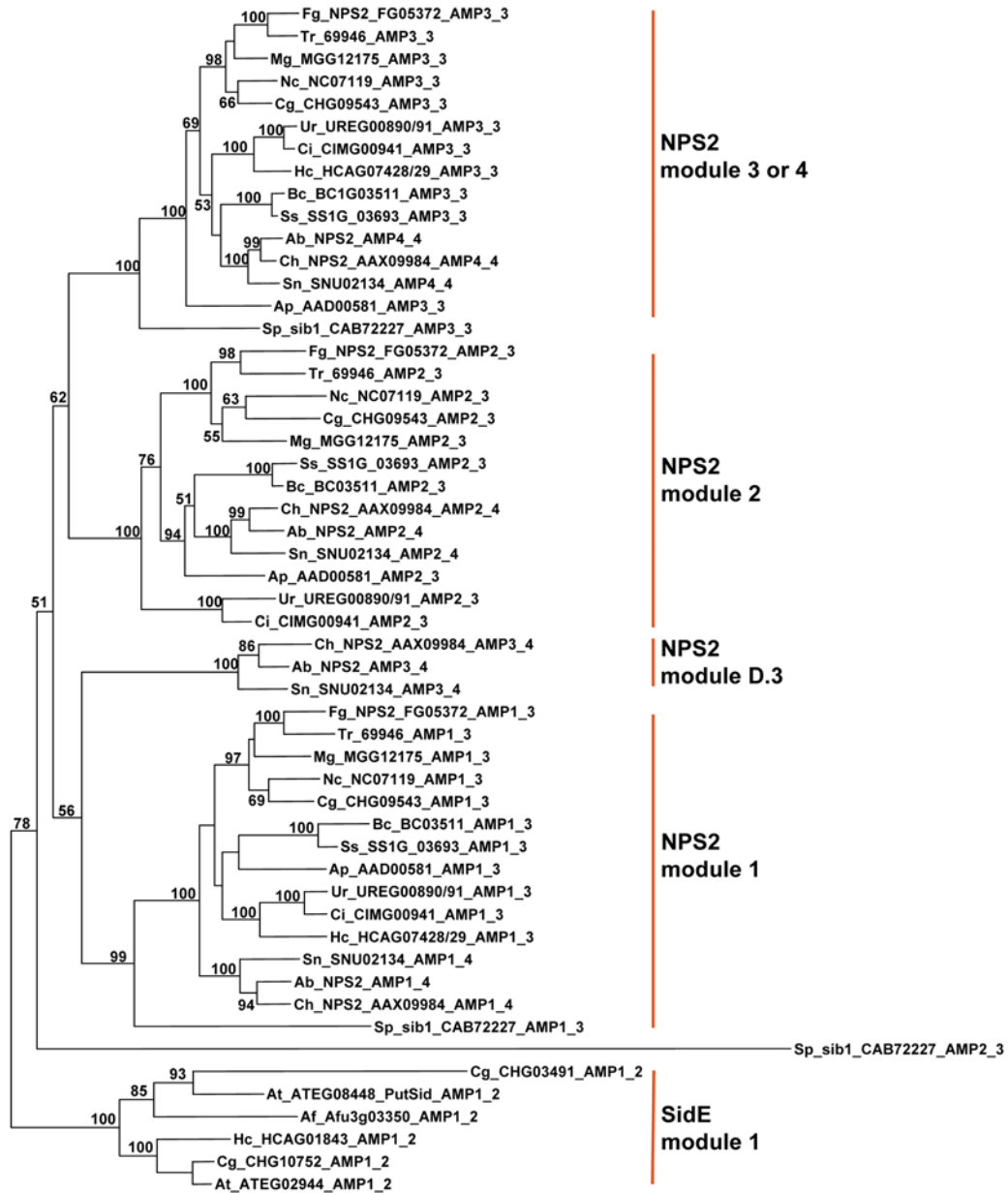
APPENDIX 2.3

Appendix 2.3. Bayesian analyses of all A domains examined in this study. As with the ML analysis (Figure 2.3), N-terminal A domains of both lineages group together and C-terminal domains of both lineages group together (thick vertical bars). NPS2, module 2 groups with the C-terminal modules, while NPS1/SidC module 2 and Dothideomycete NPS2 module D.3 group with the N-terminal modules. See Figure 2.3 for numbered node descriptions, species and Accession numbers, and nomenclature used. Bayesian posterior probabilities are indicated above branches. Note that in the Bayesian tree, the A domains of SidE module 1 group as in the ML tree (Figure 2.3).

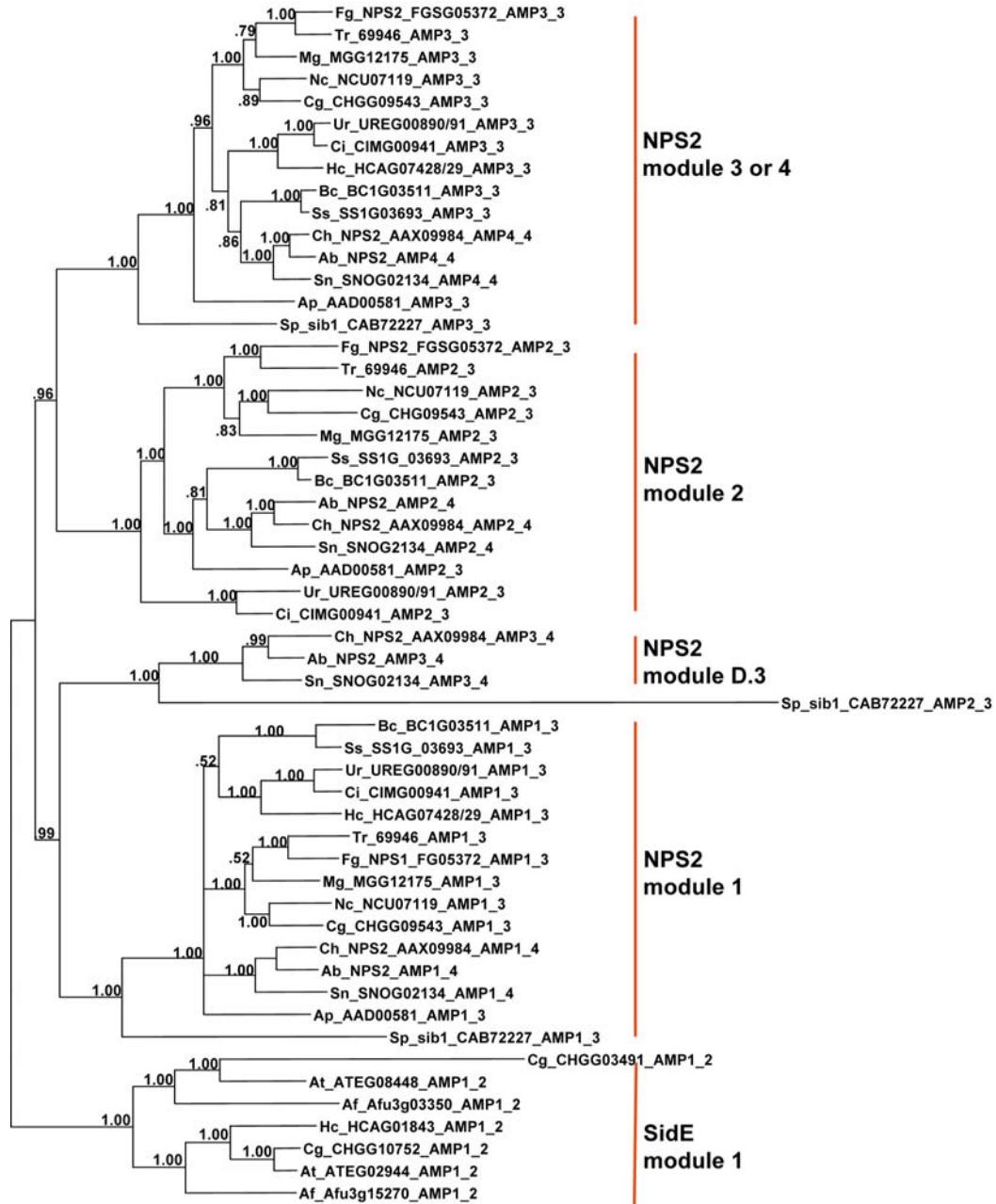


APPENDIX 2.4

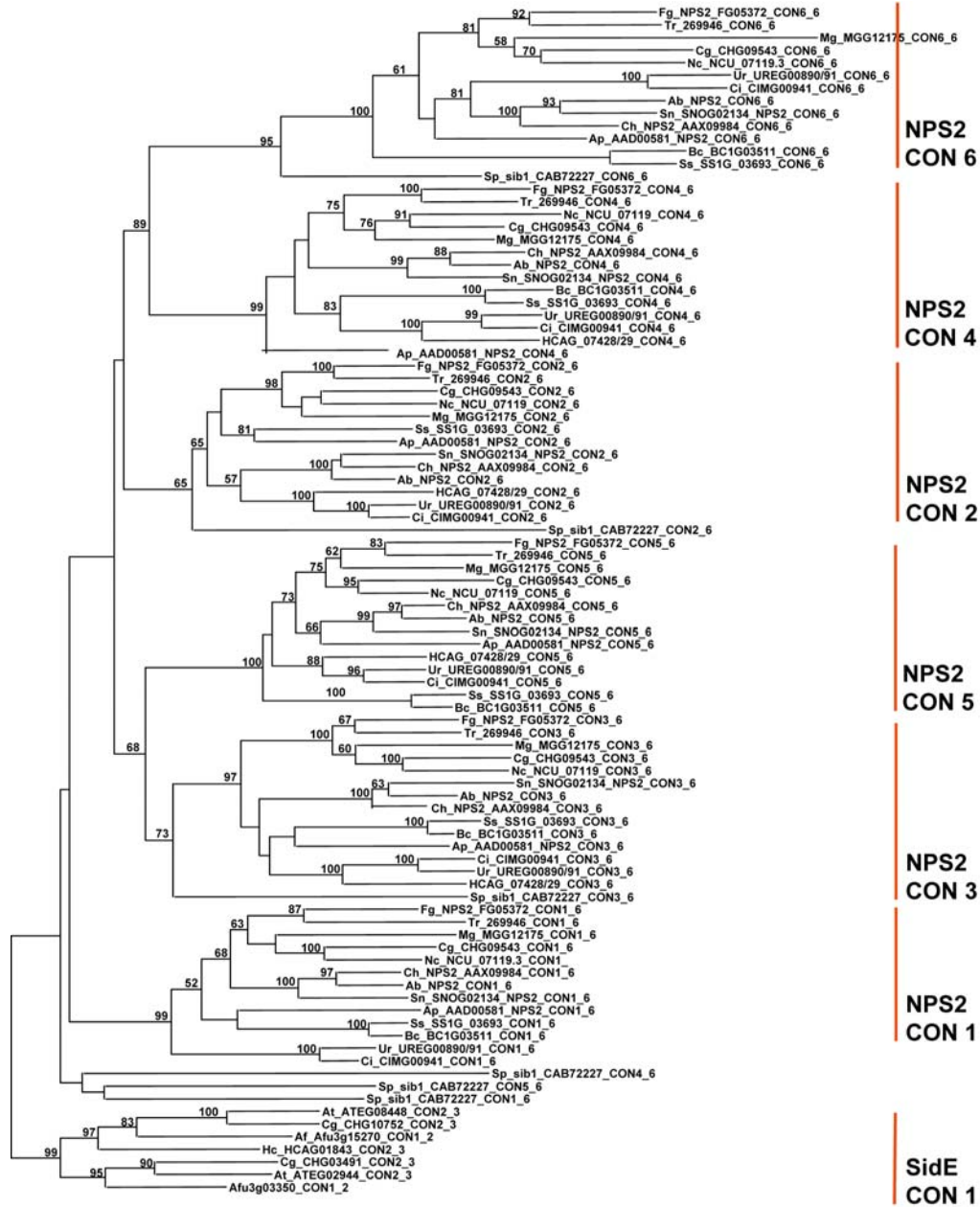
Appendix 2.4. Individual NPS2 and NPS1/SidC A domain lineage analyses. **A-B** Maximum likelihood (i) and Bayesian (ii) analyses of A and C domains, respectively, of ferrichrome synthetases in the NPS2 lineage. **C-D.** Maximum likelihood (i) and Bayesian (ii) analysis of A and C domains, respectively, of ferrichrome synthetases in the NPS1/SidC lineage. **A.** AMP domains. In both trees, A domains of module 2 group with those of C-terminal module 3 or 4 while A domains of Dothideomycete module D3 groups with those of N-terminal module 1. Bootstrap and posterior probability support respectively for these relationships are shown above branches. A domains of sib1 modules 1 and 3, group with other N and C-terminal module A domains, respectively. The degenerate A domain of sib1 module 2, varies in placement. Only module 1 and module 3 of *H. capsulatum* are included as the module 2 and 6 A domains are missing due to poor sequence quality. **B.** CON domains. In (i) and (ii) trees, C domains of modules 6 and 4 and those of module 5 or 3 group together. Bootstrap and posterior probability support for these relationships are shown above branches. The C domain of module 2 groups with modules 6 and 4 in the ML tree, but is unresolved in the Bayesian tree. Note, as indicated in the text, some SidE proteins have a N-terminal C domain. Here, for all SidEs, we used the C domain from the first complete (A-T-C) module. C domains of sib1 modules 3 and 6, group with the corresponding C domains of other NPS2 members, however all other sib1 C domains vary in placement. Only four C domains (C1, C3-5) of the *H. capsulatum* gene are shown as C2 and C6 are missing due to poor sequence quality. **C.** AMP domains. In both trees, A domains of module 1 and 2 group together while those of module 3 group separately. Bootstrap and posterior probability support respectively are shown above branches. *U. maydis* has two ferrichrome synthetases, fer3 and sid2. fer3 domains clearly group with the corresponding domains of the majority of the members of this lineage. *U. maydis* sid2 module 1 C domain, consistently groups with the module 2 C domains of the majority of the members of this lineage, while C domains of both sid2 modules 2 and 3 group with other module 3 C domains. In both trees, it is clear that *U. maydis* sid2 domains group separately from the fer3 domains, supporting the hypothesis of a duplication within this lineage. The A domains of FG11026 and CHG02251 clearly group separately from other ascomycete genes within this lineage supporting additional duplication within this lineage. **D.** CON domains. In (i) and (ii) trees, C domains of modules 6 and 4 and those of module 5 or 3 group together. Bootstrap and posterior probability support for these relationships respectively are shown above branches. The C domain of module 2, varies in placement while the C domain of module 1 also appears sister to the SidE outgroup.



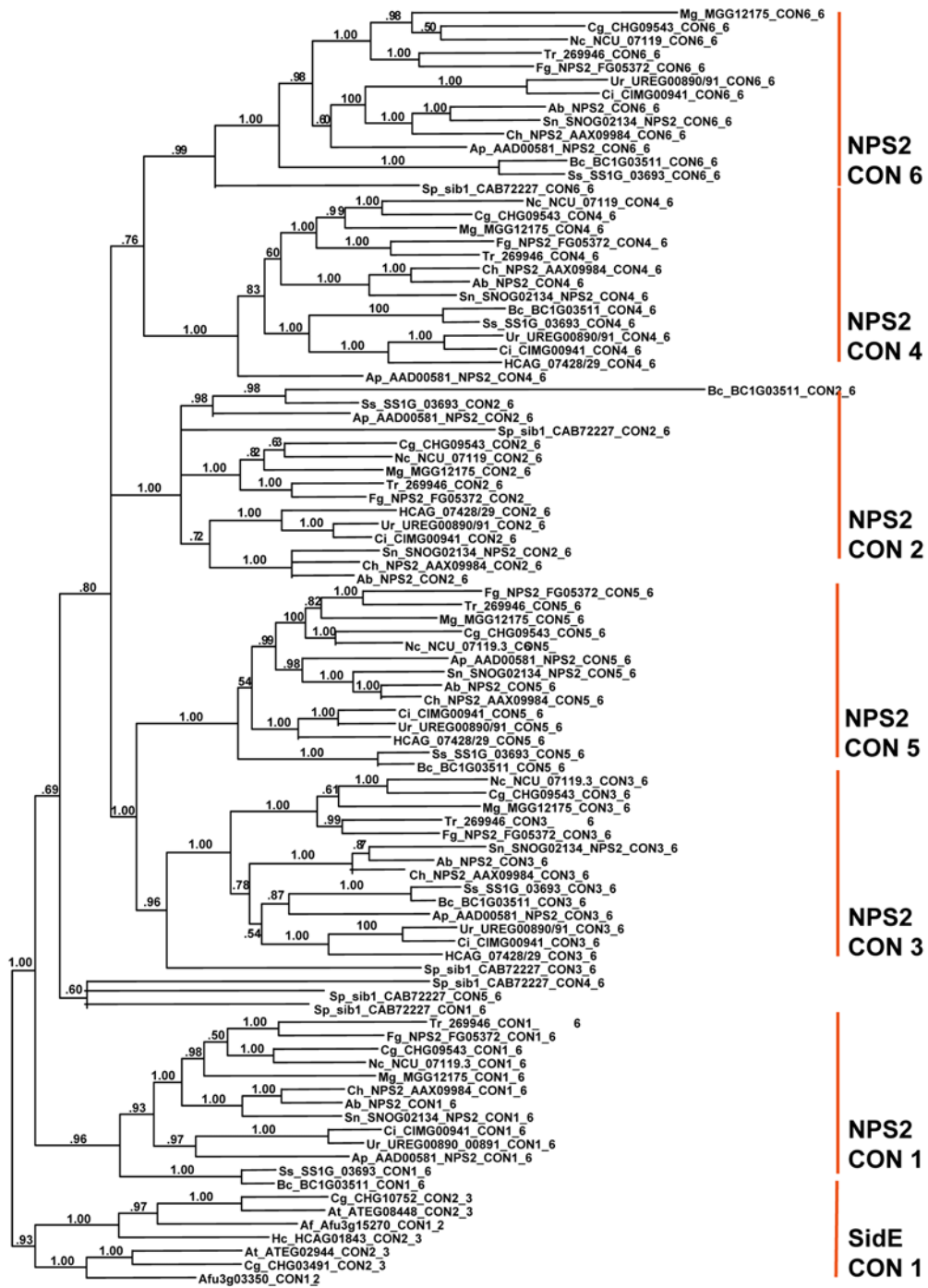
2.4Ai. NPS2 lineage AMP domains. Maximum Likelihood



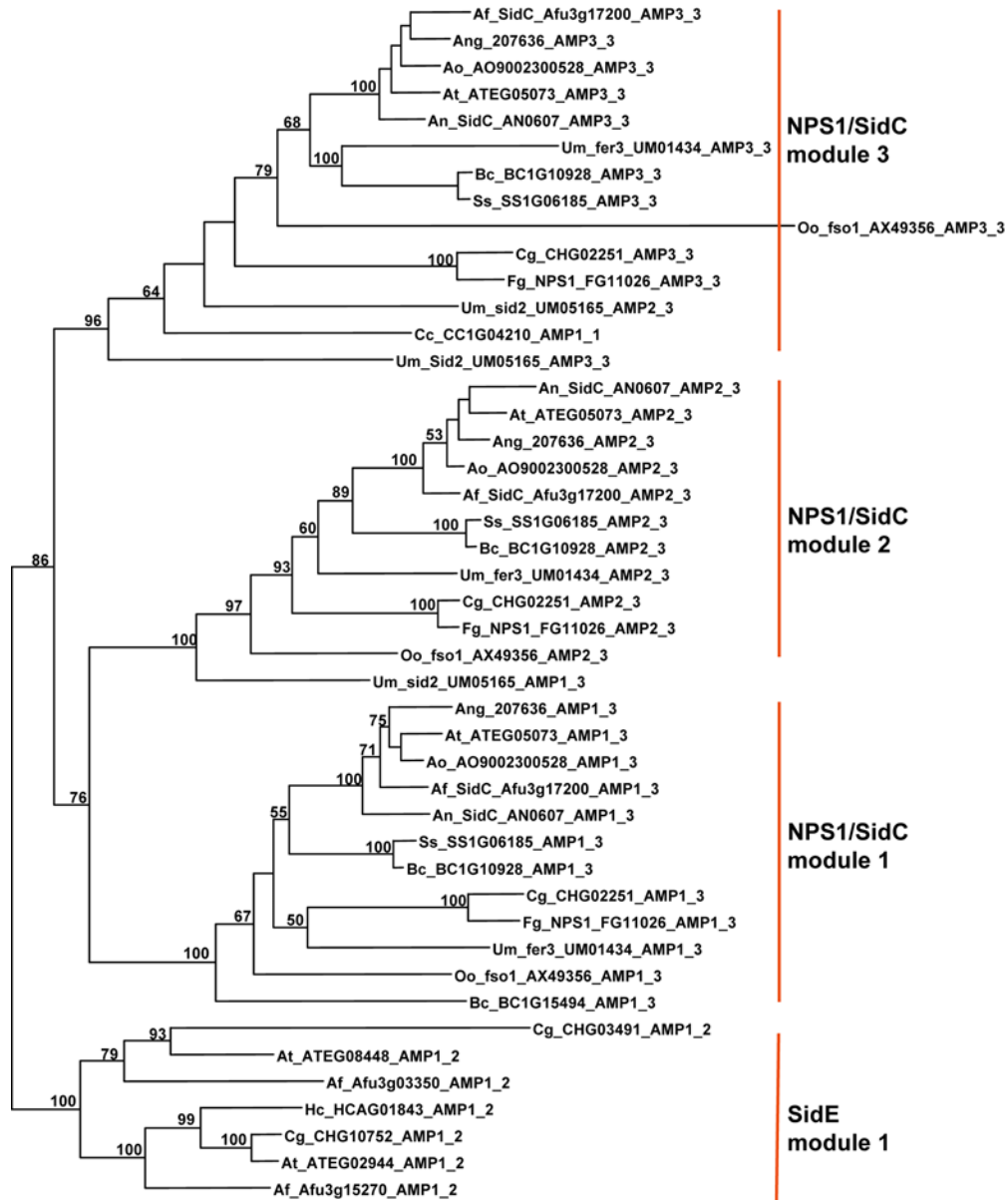
2.4Aii. NPS2 lineage AMP domains. Bayesian



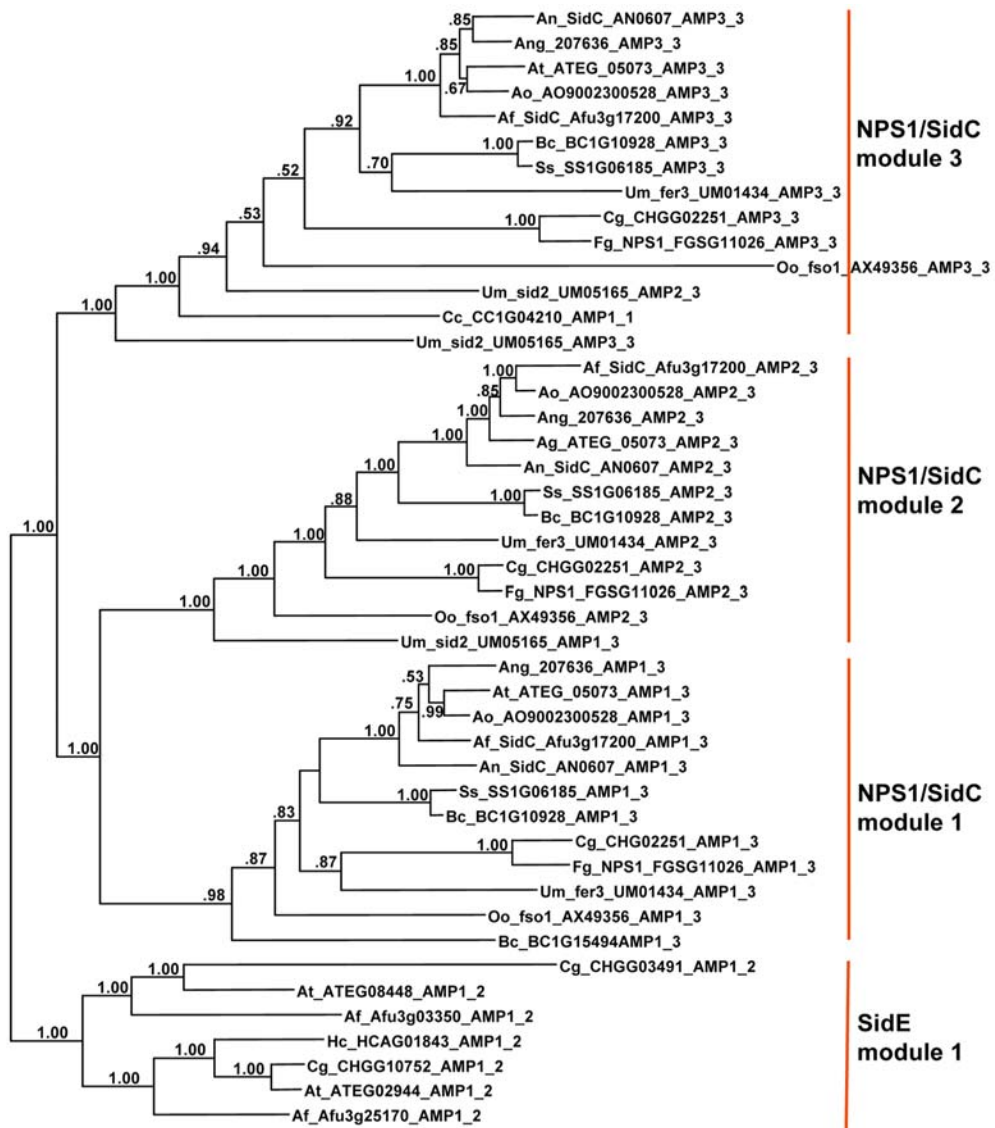
2.4Bi. NPS2 lineage CON domains. Maximum Likelihood



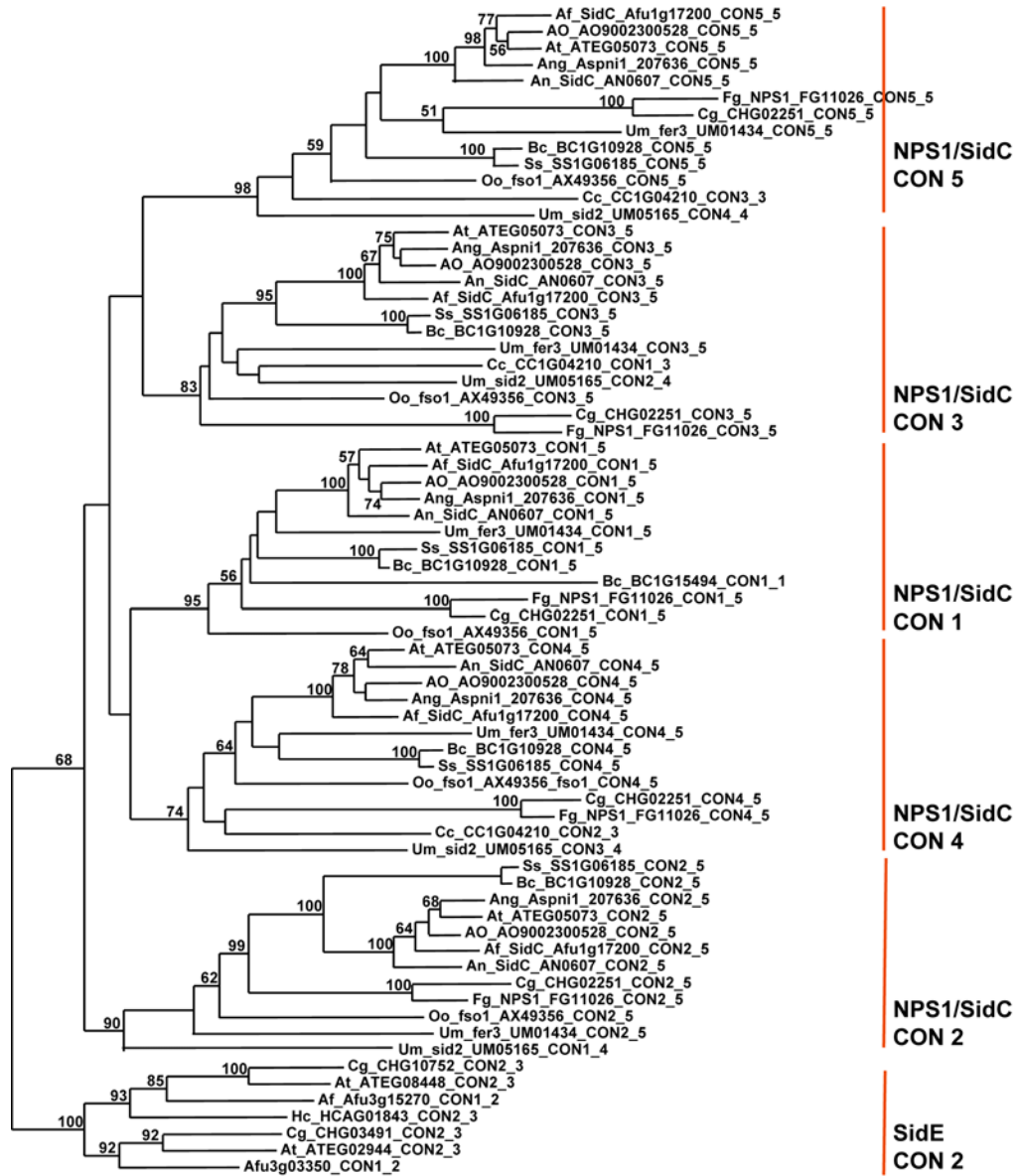
2.4Bii. NPS2 lineage CON domains. Bayesian



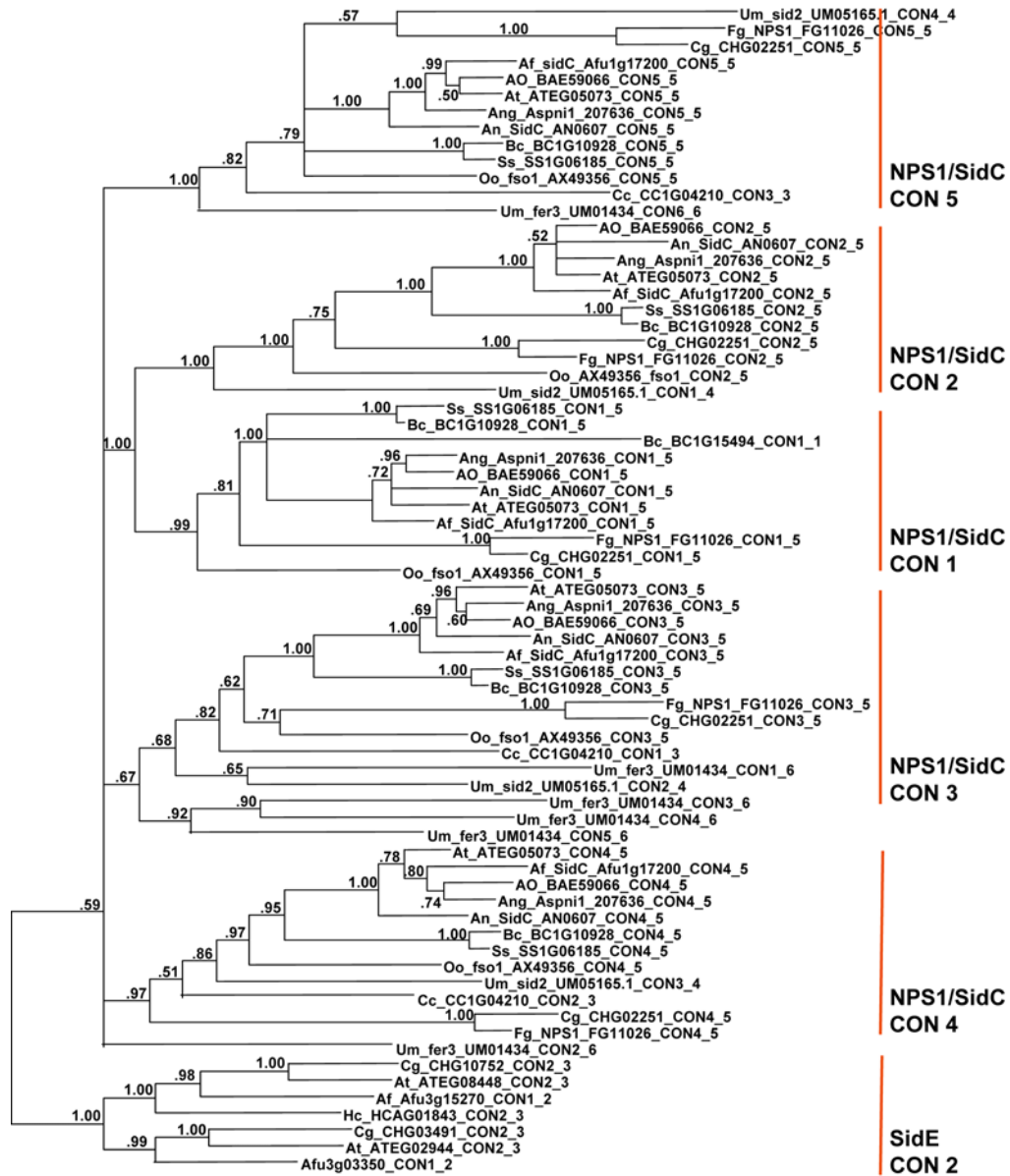
2.4Ci. NPS1/SidC lineage AMP domains. Maximum Likelihood



2.4Cii. NPS1/SidC lineage AMP domains. Bayesian



2.4Di. NPS1/SidC lineage CON domains. Maximum Likelihood



2.4Dii. NPS1/SidC lineage CON domains. Bayesian

Appendix 2.5. Amino acids corresponding to the 10AA code positions of selected bacterial and fungal NRPS AMP domains.

AminoAcid Substrate	Species	Accession #	NRPS/AMP domain ^a	10AA code position										Reference
				235	236	239	278	299	301	322	330	331	517	
GLY														
fungal:	<i>Hypocrea virens</i>	Q8NJJ1	TEX1_AMP2_18	D	I	G	M	V	V	G	V	L	K	[1, 2]
	<i>Tolypocladium inflatum</i>	Q09164	TOLIN_AMP7_11	D	I	Q	M	F	V	A	M	Q	K	[3, 4]
	<i>Schizosaccharomyces pombe</i>	SPAC23G3.02c	Sib1_AMP1_3	D	V	F	D	I	I	A	I	H	K	This study
		*												
	<i>Ustilago maydis</i>	UM05165.1**	sid2_AMP1_3	D	L	M	L	I	G	L	L	I	K	This study
	<i>Cochliobolus heterostrophus</i>	AAX09984*	NPS2_AMP4_4	D	M	Y	D	Y	I	S	F	C	K	This study
bacterial:	<i>Bacillus subtilis</i>	P45745	DHBF_AMP1_2	D	I	S	Q	L	G	L	I	W	K	[5]
	<i>Myxococcus xanthus</i>	Q50858	SafAMx1_AMP1_2	D	I	L	Q	L	G	L	V	W	K	[5]
	<i>Bacillus cereus</i>	Q81DQ0	GlycineAMPLigase_AMP1_1	D	I	L	Q	L	G	L	I	W	K	[5]
	<i>Bacillus anthracis</i>	Q81QP7	DHBF_AMP1_2	D	I	L	Q	L	G	L	I	W	K	[5]
	<i>Streptomyces chrysomallus</i>	Q9L8H4	ActinoIII_AMP2_3	D	I	L	Q	L	G	L	I	W	K	[5]
	<i>Nostoc sp.</i>	Q9RAH2	NosC_AMP2_3	D	I	L	Q	L	G	L	I	W	K	[5]
			Stachelhaus [2] consensus	None										
ALA														
fungal:	<i>Claviceps purpurea</i>	O94205	LPS1_AMP1_3	D	L	F	F	C	G	G	P	L	K	[2, 6-8]
	<i>Hypocrea virens</i>	Q8NJJ1	Tex1_AMP3_18	D	V	G	F	V	A	G	V	L	K	[9]
	<i>Hypocrea virens</i>	Q8NJJ1	Tex1_AMP8_18	D	I	F	V	V	A	G	V	I	K	[9]
	<i>Cochliobolus carbonum</i>	Q01886	HTS1_AMP2_4	D	A	G	G	C	A	M	V	A	K	[2, 10]
	<i>Tolypocladium inflatum</i>	Q09164	SimA_AMP11_11 (CssA)	D	V	F	I	Y	A	A	I	L	K	[2-4].
	<i>Cochliobolus carbonum</i>	Q01886	HTS1_Ccarb_AMP3_4	D	L	L	F	F	I	S	V	L	K	[2, 10]
	<i>Tolypocladium inflatum</i>	Q09164	SimA_AMP1_11 (CssA)	D	L	W	F	Y	I	A	V	V	K	[2-4]
bacterial:	<i>Streptococcus agalactiae</i>	P59591	DLTA_AMP1_1	D	L	M	T	F	D	A	V	A	K	[5]
	<i>Myxococcus xanthus</i>	Q50857	SafBMx1_AMP1_1	D	L	F	N	L	A	L	T	Y	K	[5][2]
	<i>Streptococcus pneumoniae</i>	P0A398	DLTA_AMP1_1	D	L	M	T	F	D	A	V	A	K	[5]
	<i>Lactobacillus rhamnosus</i>	P35854	DLTA_AMP1_1	D	L	M	V	F	C	T	V	A	K	[5]
	<i>Bacillus subtilis</i>	P39581	DLTA_AMP1_1	D	L	M	T	F	C	T	V	A	K	[5]
	<i>Staphylococcus aureus</i>	P68876	DLTA_AMP1_1	D	L	M	V	F	C	T	V	A	K	[5]
			Stachelhaus [2] consensus	D	L	L	F	G	I	A	V	L	K	[2]
ORN														
fungal:	<i>Claviceps purpurea</i>	O94205	LPS1_1_AMP2_3	D	L	V	G	M	A	A	V	G	K	[2, 8, 11]
	<i>Schizosaccharomyces pombe</i>	SPAC23G3.02c	Sib1_AMP3_3	D	V	L	D	I	G	F	I	G	K	This study
		*												
(AHO)	<i>Cochliobolus heterostrophus</i>	AAX09984*	NPS2_AMP4_4	D	V	L	D	I	G	G	I	G	K	This study
(AHO)	<i>Fusarium graminearum</i>	FG05372.1**	NPS2_AMP3_3	D	V	L	D	I	G	A	I	G	K	This study
(AHO)	<i>Aspergillus nidulans</i>	AN0607.3**	SidC_AMP3_3	D	P	L	S	T	G	A	I	G	K	This study
(AHO)	<i>Fusarium graminearum</i>	FG11026.1**	NPS1_AMP3_3	D	P	T	G	T	G	F	I	G	K	This study
(AHO)	<i>Ustilago maydis</i>	UM05165.1**	sid2_AMP3_3	D	V	I	D	M	G	A	I	G	K	This study

Appendix 2.5 Continued

bacterial:	<i>Brevibacillus parabrevis</i>	O30409	TycC_3_AMP5_6	D	V	G	E	I	G	S	I	D	K	[5]
	<i>Bacillus licheniformis</i>	O68007	BACB_AMP2_2	D	V	G	E	I	G	S	V	D	K	[5]
	<i>Mycobacterium smegmatis</i>	O87313	FxbB_AMP1_2	D	I	N	Y	W	G	G	I	G	K	[5]
	<i>Mycobacterium smegmatis</i>	O87314	FxbC_AMP1_4	D	M	E	N	L	G	L	I	N	K	[5][2]
	<i>Mycobacterium smegmatis</i>	O87314	FxbC_AMP3_4	D	M	E	N	L	G	L	I	N	K	[2, 5]
	<i>Bacillus subtilis</i>	O87606	FenC_AMP2_2	D	V	G	E	I	G	S	I	G	K	[2, 5]
	<i>Aneurinibacillus migulanus</i>	P0C063	GRSB_AMP3_4	D	V	G	E	I	G	S	I	D	K	[2, 5]
	<i>Bacillus subtilis</i>	P39845	PPS1_AMP2_2	D	V	G	E	I	G	S	I	D	K	[5]
			Stachelhaus [2] consensus	D	M	E	N	L	G	L	I	N	K	[2]
			Orn (1)											

^aNRPS A domains identified in the literature (referenced in final column) as coding for GLY, ALA, SER, or ORN were aligned with Toffee using the GrsA 1AMU structure as template. This alignment was inspected manually to insure consistency with a structural alignment of 1AMU and A domains from a number of other related A domains (1PG3_A, 1ULT_A, 1LC_I, 1T5D_X and 1MD9_A) and amino acids at positions corresponding to the 10AA code were extracted. The consensus “code” for each substrate determined by Stachelhaus et. al. [2] are shown.

Column 1. For ornithine, only a single representative is known from fungi. Domains identified in this study as activating AHO (N⁵-acetyl-N⁵-hydroxy-L- ornithine, N⁵-acyl-N⁵-hydroxy-L- ornithine), were included.

Column 3. All entries are uniprot (EMBL) accessions unless otherwise marked * = GenBank, ** = Broad Institute ID.

Column 4. NRPS name/A domain. For example, TEX1_AMP2_18 is the second A domain of a total of 18 in TEX1

REFERENCES

1. Van der Helm D, Winkelmann, G.: **Hydroxamates and polycarboxylates as iron transport agents (siderophores) in fungi**. In: *Metal Ions in Fungi*. Edited by Winkelmann, G. and Dennis, W. Vol. 11, series edn. New York, New York: Marcel Dekker; 1987: 39-98.
2. Haas H, Eisendle M, Turgeon BG: **Siderophores in fungal physiology and virulence**. *Ann Rev Phytopathol* 2008, **46**: 149-187.
3. Lee B, Kroken, S., Chou, D.Y.T., Robbertse, B., Yoder, O.C., and B.G. Turgeon: **Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress** *Eukaryotic Cell* 2005, **4**(3):545-555.
4. Fewer DP, Rouhiainen L, Jokela J, Wahlsten M, Laakso K, Wang H, Sivonen K: **Recurrent adenylation domain replacement in the microcystin synthetase gene cluster**. *BMC Evolutionary Biology* 2007, **7**.
5. Renshaw JC, Robson GD, Trinci APJ, Wiebe MG, Livens FR, Collison D, Taylor RJ: **Fungal siderophores: Structures, functions and applications**. In., Vol. 106; 2002: 1123-1142.
6. Schwecke T, Goettling K, Durek P, Duenas I, Kaeufer NF, Zock Emmenthal S, Staub E, Neuhof T, Dieckmann R, von Doehren H: **Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi**. *ChemBioChem* 2006, **7**:612-622.
7. Jalal MAF, Van der Helm, Dick: **Isolation and spectroscopic identification of fungal siderophores**. In: *CRC Handbook of Microbial Iron Chelates*. Edited by Winkelmann G. Boca Raton, FL: CRC Press; 1991: 235-269.
8. Lautru S, Challis GL: **Substrate recognition by nonribosomal peptide synthetase multi-enzymes**. *Microbiology (Reading)* 2004, **150**(Part 6):1629-1636.

9. Stachelhaus T, Mootz, Henning D. , and Marahiel, Monamed: **The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.** *Chemistry and Biology* 1999, **6**:493-505.
10. Challis GL, Ravel J, Townsend CA: **Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.** *Chemistry and Biology (London)* 2000, **7**(3):211-224.
11. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH: **Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution.** *BMC Evolutionary Biology* 2007, **7**:Article No.: 78.
12. Konz DaM, Mohamen: **How do peptide synthetases generate structural diversity?** *Chemistry and Biology* 1999, **6**:R39-R48.
13. Walton JD, Panaccione, Daniel G., and Hallen, Heather, E.: **Peptide synthesis without ribosomes.** In: *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine.* Edited by Tkacz J and Lange L. New York, New York: Kluwer Academic/Plenum Publishers; 2004.
14. Horowitz NG, Charlang, G., Horn, G. and Williams, N. : **Isolation and identification of the conidial germination factor of *Neurospora crassa*.** *J Bacteriology* 1976, **127**:135-140.
15. Matzanke BF, Bill, E., Trautwein, A., and Winklemann, G.: **Role of siderophores in iron storage in spores of *Neurospora crassa* and *Aspergillus ochraceus*.** *Journal of Bacteriology* 1987, **169**(12):5873-5876.
16. Matzanke BF: **Iron storage in fungi.** In: *Metal Ions in Fungi.* Edited by Winklemann G, and Dennis W., Vol. 11. New York, New York: Marcel Dekker Inc; 1994: 179-214.
17. Eisendle M, Oberegger, H., Zadra, I., and Haas, H.: **The siderophore system is essential for viability of *Aspergillus nidulans*: functional analysis of two genes encoding L-ornithine-N-5-monooxygenase (sidA) and a non-ribosomal peptide synthetase (SidC).** *Molecular Microbiology* 2003, **49**:359-375.

18. Oide S, Krasnoff SB, Gibson DM, Turgeon BG: **Intracellular siderophores are essential for ascomycete sexual development in heterothallic *Cochliobolus heterostrophus* and homothallic *Gibberella zeae*.** *Eukaryotic Cell* 2007, **6**(8):1339-1353.
19. Eisendle M, Schrettl M, Kragl C, Mueller D, Illmer P, Haas H: **The intracellular siderophore ferricrocin is involved in iron storage, oxidative-stress resistance, germination, and sexual development in *Aspergillus nidulans*.** *Eukaryotic Cell* 2006, **5**(10):1596-1603.
20. Oide S, Moeder W, Krasnoff S, Gibson D, Haas H, Yoshioka K, Turgeon BG: **NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes.** *Plant Cell* 2006, **18**(10):2836-2853.
21. Guindon S GO: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic Biology* 2003, **52**(5):696-704.
22. Welzel K, Eisfeld, Katrin, Antelo, Luis, Anke, Timm, Anke, Heidrun: **Characterization of the ferrichrome A biosynthetic gene cluster in the homobasidiomycete *Omphalotus olearius*.** *FEMS Microbiology Letters* 2005, **249**:157-163.
23. Cramer RA, Stajich, Jason E., Yamanaka, Yvonne, Dietrich, Fred S., Steinbach, William, and Perfect, John R.: **Phylogenomic analysis of non-ribosomal peptide synthetases in the genus *Aspergillus*.** *Gene* 2006, doi:10.1016/j.gene2006.07.008.
24. Eddy S: <http://hmmer.wustl.edu>.
25. Conti E, Stachelhaus, T., Marahiel, MA, and Brick, P.: **Structural basis for the activation of phenylalanine in the nonribosomal biosynthesis of gramicidin S** *EMBOJ* 1997, **16**:4174-4183.
26. O'Sullivan O, Suhre K, Abergel C, Higgins DG, Notredame C: **3DCoffee: Combining protein sequences and structures within multiple sequence alignments.** *Journal of Molecular Biology* 2004, **340**(2):385-395.

27. Abascal F, Zardoya, R., Posada, D.: **ProtTest: Selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**(9):2104-2105.
28. Ronquist RaJPH: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
29. Page RDM: **GeneTree: comparing gene and species phylogenies using reconciled trees.** *Bioinformatics* 1998, **14**(9):819-820.
30. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J *et al*: **Reconstructing the early evolution of Fungi using a six-gene phylogeny.** *Nature* 2006, **443**(7113):818-822.
31. Robbertse B, Reeves, John B., Schoch, Conrad L., and Spatafora, Joseph W.: **A phylogenomic analysis of the Ascomycota.** *Fungal Genetics and Biology* 2006, **43**:715-725.
32. Fitzpatrick DA, Logue, Mary E., Stajich, Jason E., and Butler, Geraldine: **A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis.** *BMC Evolutionary Biology* 2006, **6**:99.
33. Kuramae EE, Robert V, Echavarrri-Erasun C, Boekhout T: **Cophenetic correlation analysis as a strategy to select phylogenetically informative proteins: an example from the fungal kingdom.** *Bmc Evolutionary Biology* 2007, **7**.
34. Schoch CL, Shoemaker RA, Seifert KA, Hambleton S, Spatafora JW, Crous PW: **A multigene phylogeny of the Dothideomycetes using four nuclear loci.** *Mycologia* 2006, **98**(6):1041-1052.
35. Elemento O, Gascuel, Olivier, and Lefranc, Marie-Paule: **Reconstructing the duplication history of tandemly repeated genes.** *Molecular Biology and Evolution* 2002, **19**(3):278-288.
36. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Research* 1997, **25**(17):3389-3402.

37. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF: **Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements.** *Nucleic Acids Research* 2001, **29**(14):2994-3005.
38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The protein data bank.** *Nucleic Acids Research* 2000, **28**(1):235-242.
39. Shindyalov IN, Bourne PE: **Protein structure alignment by incremental combinatorial extension (CE) of the optimal path.** *Protein Engineering* 1998, **11**(9):739-747.
40. Sali A, Blundell TL: **Comparative protein modeling by satisfaction of spatial restraints.** *Journal of Molecular Biology* 1993, **234**(3):779-815.
41. Sanchez R, Sali A: **Evaluation of comparative protein structure modeling by MODELLER-3.** *Proteins-Structure Function and Genetics* 1997:50-58.
42. Sanchez R, Sali A: **Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(23):13597-13602.
43. Sanchez RaŠA: **Comparative protein structure modeling. Introduction and practical examples with MODELLER.** *Methods Mol Biol* 2000, **143**:97-129.
44. Kalinina OV, Mironov AA, Gelfand MS, Rakhmaninova AB: **Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families.** *Protein Science* 2004, **13**(2):443-456.
45. Gu XaVV, Kent: **DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family.** *Bioinformatics* 2002, **18**(3):500-501.
46. Gu X, Wang YF, Gu JY, Velden KV, Xu DP: **Predicting type-I (rate-shift) functional divergence of protein sequences and applications in functional genomics.** *Current Genomics* 2006, **7**(2):87-96.

47. Rausch C, Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H.: **Specificity predictions of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs).** *Nucleic Acids Research* 2005, **33**(18):5799-5808.
48. Yuan WM, Gentil, Guillaume D., Budde, Allen D., and Leong, Sally A.: **Characterization of the *Ustilago maydis* sid2 gene, Encoding a multidomain peptide synthetase in the ferrichrome biosynthetic gene cluster.** *Journal of Bacteriology* 2001, **183**(13):4040-4051.
49. Haas H: **Molecular genetics of fungal siderophore biosynthesis and uptake The role of siderophores in iron uptake and storage**, vol. 62; 2003: 316-330.
50. Jogl G, Tong L: **Crystal structure of yeast acetyl-Coenzyme A synthetase in complex with AMP.** *Biochemistry* 2004, **43**(6):1425-1431.
51. Linne U, Schafer A, Stubbs MT, Marahiel MA: **Aminoacyl-Coenzyme A synthesis catalyzed by adenylation domains.** *Febs Letters* 2007, **581**(5):905-910.
52. Tobiasen C, Aahman, Johan, Ravnholt, Kristin Slot, Bjerrum, Morten Jannick, Grell, Morten Nedergaard, and Geise, Henriette: **Nonribosomal peptide synthetase (NPS) genes in *Fusarium graminearum*, *F. culmorum* and *F. pseudograminearum* and identification of NPS2 as the producer of ferricrocin.** *Current Genetics* 2007, **51**:43-58.
53. Hof C, Eisfeld K, Welzel K, Antelo L, Foster AJ, Anke H: **Ferricrocin synthesis in *Magnaporthe grisea* and its role in pathogenicity in rice.** *Molecular Plant Pathology* 2007, **8**(2):163-172.
54. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S *et al*: **The genome sequence of the filamentous fungus *Neurospora crassa*.** *Nature* 2003, **422**(6934):859-868.
55. Eichhorn H, Lessing F, Winterberg B, Schirawski J, Kamper J, Muller P, Kahmann R: **A ferroxidation/permeation iron uptake system is required for virulence in *Ustilago maydis*.** *Plant Cell* 2006, **18**(11):3332-3345.

56. Wiest A, Grzegorski D, Xu BW, Goulard C, Rebuffat S, Ebbole DJ, Bodo B, Kenerley C: **Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase.** *Journal of Biological Chemistry* 2002, **277**(23):20862-20868.
57. Husi H, Schorgendorfer K, Stempfer G, Taylor P, Walkinshaw MD: **Prediction of substrate-specific pockets in cyclosporin synthetase.** *Febs Letters* 1997, **414**(3):532-536.
58. Lawen A, Traber R: **Substrate specificities of cyclosporin synthetase and peptolide Sdz 214-103 synthetase - Comparison of the substrate specificities of the related multifunctional polypeptides.** *Journal of Biological Chemistry* 1993, **268**(27):20452-20465.
59. Tudzynski P, Holter K, Correia T, Arntz C, Grammel N, Keller U: **Evidence for an ergot alkaloid gene cluster in *Claviceps purpurea*.** *Molecular and General Genetics* 1999, **261**(1):133-141.
60. Keller N, Tudzynski B: **Ergot alkaloids.** In: *The Mycota, Industrial Applications*. Edited by Osiewacz HD. Berlin: Springer; 2001: 157-181.
61. Walzel B, Riederer B, Keller U: **Mechanism of alkaloid cyclopeptide synthesis in the ergot fungus *Claviceps purpurea*.** *Chemistry & Biology* 1997, **4**(3):223-230.

CHAPTER 3

PHYLOGENOMICS REVEALS SUBFAMILIES OF FUNGAL NONRIBOSOMAL PEPTIDE SYNTHETASES AND THEIR EVOLUTIONARY RELATIONSHIPS

3.1 Abstract

Background: Nonribosomal peptide synthetases (NRPSs) are multimodular enzymes, found in fungi and bacteria, which biosynthesize peptides without the aid of ribosomes. Although their metabolite products have been the subject of intense investigation due to their life-saving roles as medicinals and injurious roles as mycotoxins and virulence factors, little is known of the phylogenetic relationships of the corresponding NRPSs or whether they can be ranked into subgroups of common function. We identified genes (*NPS*) encoding NRPS and NRPS-like proteins in 38 fungal genomes and undertook phylogenomic analyses in order to identify fungal NRPS subfamilies, assess taxonomic distribution, to evaluate levels of conservation across subfamilies, and to address mechanisms of evolution of multimodular NRPSs. We also characterized relationships of fungal NRPSs, a representative sampling of bacterial NRPSs, and related adenyating enzymes, including α -aminoadipate reductases (AARs) involved in lysine biosynthesis in fungi.

Results: Phylogenomic analysis identified nine major subfamilies of fungal NRPSs which fell into two main groups: one corresponds to *NPS* genes encoding primarily mono/bi-modular proteins which grouped with bacterial NRPSs and the other includes genes encoding primarily multimodular and exclusively fungal NRPSs. AARs shared a closer phylogenetic relationship to NRPSs than to other acyl-adenyating enzymes. Phylogenetic analyses and taxonomic distribution suggest that several mono/bi-

modular subfamilies arose either prior to, or early in, the evolution of fungi, while two multimodular groups appear restricted to and expanded in fungi. The older mono/bi-modular subfamilies show conserved domain architectures suggestive of functional conservation, while multimodular NRPSs, particularly those unique to euascomycetes, show a diversity of architectures and of genetic mechanisms generating this diversity.

Conclusions: This work is the first to characterize subfamilies of fungal NRPSs. Our analyses suggest that mono/bi-modular NRPSs have more ancient origins and more conserved domain architectures than multimodular NRPSs. It also demonstrates that the α -aminoacidate reductases involved in lysine biosynthesis in fungi are closely related to mono/bi-modular NRPSs. Several groups of mono/bi-modular NRPS metabolites are predicted to play more pivotal roles in cellular metabolism than products of multimodular NRPSs. In contrast, multimodular subfamilies of NRPSs are of more recent origin, are restricted to fungi, show less stable domain architectures, and biosynthesize metabolites which perform more niche-specific functions than mono/bi-modular NRPS products. The EAS NRPS subfamily, in particular, shows evidence for extensive gain and loss of domains suggestive of the contribution of domain duplication and loss in responding to niche-specific pressures.

3.2 Background

Nonribosomal peptide synthetases (NRPSs) are multimodular megasynthases which catalyze biosynthesis of small bioactive peptides (NRPs) via a thiotemplate mechanism independent of ribosomes [1-5]. NRPS encoding genes (*NPSs*) are plentiful in fungi and bacteria but are not known in plants or animals. The enzymes they encode biosynthesize a staggering diversity of chemical products because their

substrates can include both D and L forms of the 20 amino acids used in ribosomal protein synthesis, as well as non-proteinogenic amino acids such as ornithine, imino acids, and hydroxy acids such as α -amino adipic and α -butyric acids [1]. The natural functions of most NRPs for producing organisms are largely unknown, although recently it has become clearer that they play fundamental roles in fungal reproductive and pathogenic development, morphology, cell surface properties, stress management, and nutrient procurement [6-10] [11-15] in addition to better-known roles as toxins/mycotoxins involved in plant or animal pathogenesis or as life-saving pharmaceuticals such as antibiotics, immunosuppressants, and anticancer agents.

NRPSs use a set of core domains, known as a module, to accomplish peptide synthesis. A minimal module consists of three core domains: 1) an adenylation (A) domain which recognizes and activates the substrate via adenylation with ATP, 2) a thiolation (T) or peptidyl carrier protein (PCP) domain which binds the activated substrate to a 4'-phosphopantetheine (PP) cofactor via a thioester bond and transfers the substrate to 3) a condensation (C) domain which catalyzes peptide bond formation between adjacent substrates on the megasynthase complex [1]. Several specialized C-terminal domains involved in chain termination and release of the final peptide product have also been identified [16, 17]. In bacteria, chain release is most commonly effected by a thioesterase (TE) domain [18], which releases the peptide by either hydrolysis or internal cyclization [16, 17, 19]. In fungi, only a few NRPSs, such as the ACV synthetases, are known to release products via a TE domain and chain release is carried out by a variety of mechanisms, two of which predominate and occur less frequently in bacterial systems: 1) a terminal C domain, which catalyzes release by inter- or intra-molecular amide bond formation [16], and 2) a thioesterase NADP(H) dependent reductase (R) domain [20-23], which catalyzes reduction with NADPH to form an aldehyde. An additional mechanism, which has been reported

only in biosynthesis of fungal ergot alkaloids, involves nonenzymatic cyclization by formation of a diketopiperazine ring [16, 24].

NRPSs may contain additional modifying domains which alter the substrate during NRPS biosynthesis: 1) an epimerization (E) domain which catalyzes epimerization of an amino acid from the L to the D configuration [25], 2) an N-methylation (M) domain (methyltransferase) which catalyzes transfer of a methyl group from an S-adenosyl-methionine to an α -amino of the amino acid substrate, and 3) a specialized C domain termed a cyclization (Cyc) domain which catalyzes formation of oxazoline or thiazoline rings by internal cyclization of cysteine, serine, or threonine residues [26]. Additional tailoring enzymes which are not part of the NRPS may modify either the substrate or the final peptide product by glycosylation, hydroxylation, acylation, or halogenation [27, 28].

NRPSs may be monomodular, consisting of a single A-T-C module, or multimodular, consisting of repeated A-T-C modules. The suite of 14 NRPSs found in the genome of the Dothideomycete *Cochliobolus heterostrophus* is representative of the diversity of NPS genes in filamentous ascomycetes in that it contains a representative from most currently recognized groups of fungal NRPSs [10] [6], and, with the exception of duplicated copies of *ChNPS12*, the modular domain architectures of each encoded enzyme are distinct (Appendix 1). In addition to mono- and multi-modular NPSs, a hybrid gene (*ChNPS7;PKS24*) encoding an incomplete NRPS module (A-T) fused to a polyketide synthase (PKS) unit is present [10, 29]. Hybrid PKS;NRPS synthetases (e.g. ACE1, SYN2 in *Magnaporthe oryzae*, the reverse organization of *ChNPS7;PKS24*) are more common in filamentous fungi as well as in bacteria [30-34], although *C. heterostrophus* lacks a representative.

The evolutionary mechanisms giving rise to genes encoding enzymes with such diverse modular architectures are clearly complex. Likely mechanisms include:

1) tandem duplication and loss of individual modules or domains, 2) gene fusion/fission, and 3) recombination and/or gene conversion of individual modules or domains either within the same *NPS* or between different *NPS*s. It has been suggested that genes involved in secondary metabolite (small molecule) biosynthesis tend to be located in subtelomeric regions, a factor which may contribute to their rapid evolution by the aforementioned mechanisms [35, 36].

*NPS*s are generally recognized as a rapidly evolving gene class in fungi leading to few orthologs between species and highly discontinuous distributions [10, 37, 38]. However, as has been observed for members of other eukaryotic gene families (e.g., major histocompatibility complex [39], immune response [40], zinc-finger [41], reproductive [42], olfactory/chemosensory [43-47], MADS-box [48], and F-box gene families [49] among others), within each family, conservation and rates of gene duplication and loss are likely to vary among subgroups of genes encoding proteins of different function. In fact, some *C. heterostrophus* *NPS*s, *NPS2*, *NPS4*, *NPS6* and *NPS10* are conserved or moderately conserved across euascomycote fungi [8, 10, 50] and their NRP products are involved in basic cellular functions such as growth and development, reproduction, and pathogenesis [6-8]. The majority of *NPS*s, however, are highly discontinuously distributed across fungal taxa and even closely related species may share only a few homologs. Some, e.g., *Cochliobolus carbonum* *HTS1*, the gene encoding the NRPS for biosynthesis of HC-toxin [51], and *Alternaria alternata* apple pathotype *AMT*, the gene encoding the NRPS for biosynthesis of AM-toxin [52], appear unique even to one race or pathotype within a single species. These lineage-specific synthetases tend to have more specialized, niche-specific functions.

Higher rates of gene duplication and loss may reflect an adaptive response to selective pressure from pathogens, interactions with other organisms, or other environmental pressures. Recent work suggests that, in fungi, genes involved in

responses to stress are more likely to undergo duplication and loss than growth related genes [53]. Thus, we hypothesize that NRPSs with conserved functions involved in growth and development will show less variation in gene copy number and maintain a relatively conserved domain architecture in comparison with NRPSs with more niche-specific functions.

The multimodular structure of NRPSs and the complex mechanisms by which they evolve present challenges to phylogenetic analysis and consequently little work has been done to characterize phylogenetic relationships across this large class of megasynthases or to ask whether subclasses of common function can be identified, based on close relationships with NRPSs whose chemical products are known. In this study, we undertook phylogenomic analyses on a comprehensive dataset of fungal NRPS proteins to: 1) identify major subfamilies of NRPSs, 2) analyze patterns of distribution of these major subfamilies across fungal taxonomic groups, 3) understand relationships among selected bacterial NRPSs, fungal monomodular NRPS/NRPS-like proteins, fungal multimodular NRPSs, and related adenylyating enzymes, including α -aminoacidate reductases involved in lysine biosynthesis in fungi, 4) consider mechanisms of evolution of multimodular NRPSs, and 5) analyze patterns of NRPS gene and A domain duplication and loss across fungi.

3.3 Results and Discussion

3.3.1 Identification and Domain Structure of Candidate NRPSs

Candidate NRPSs extracted from each sequenced genome are listed in Appendix 3.2. Genus and species abbreviations for all organisms mentioned in this study are shown in Appendix 3.3. The proposed domain structure for each NRPS,

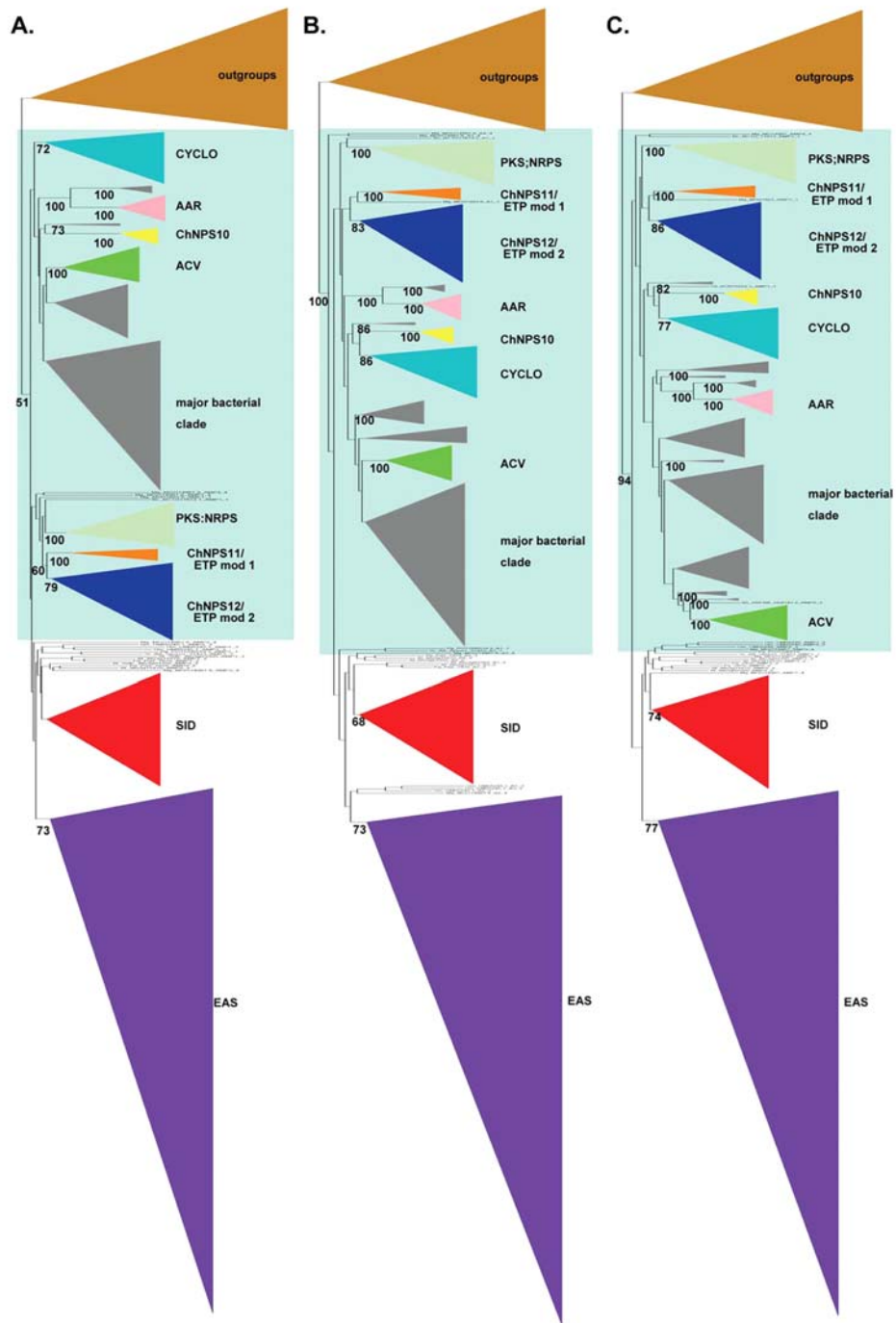
based on searches with our fungal-specific HMMER models Appendix 3.4 and the PFAM and Interpro databases, is shown in Appendix 3.2. The majority of multimodular NRPSs were composed of one or more standard NRPS modules (A-T-C) with or without modifying domains (E, M, etc), while most monomodular NRPSs lacked complete A-T-C modules and consisted of a single A domain or an A-T unit followed by a variety of C-terminal domains, several of which have not previously been identified as core NRPS domains (Appendix 3.2).

3.3.2 Phylogenomic Analysis and Subfamily Identification

All known NRPS/NRPS-like proteins formed a monophyletic group supported by greater than 90% bootstrap support in ML analyses and greater than 50% bootstrap support in the NJ analysis (Figure 3.1), separating them from most other known adenylating enzymes selected as potential outgroups, e.g., Acyl AMP ligases (AAL), CPS1 [54], Long Chain Fatty Acid ligases (LCFAL), Acetyl-CoA synthetases (ACoAS), and Ochratoxin synthetases (OCHRA)(Figure 3.1, Appendix 3.5). The α -aminoadipate reductases (AAR), homologs of *S. cerevisiae* Lys2 [23, 55-57], grouped within this well-supported clade of NRPS/NRPS-like proteins rather than with the other adenylating enzymes (Figures 3.1, 3.2, Appendix 3.6), suggesting that AARs are more closely related to NRPSs than to other adenylating enzymes.

The tree topologies resulting from phylogenetic analyses of individual A domains revealed two major groups of fungal NRPSs (Figure 3.1, Appendix 3.6). The first group (Figure 3.1, light blue rectangle) consists of primarily mono- or bi-modular fungal NRPSs which group with bacterial NRPS A domains.

Figure 3.1. Cartoons of tree topologies showing major NRPS subfamilies. All trees reflect phylogenetic analyses of the complete A domain dataset. **A.** NJ tree using a ML distance matrix created using the WAG plus gamma model. **B.** ML tree (PhyML) using the WAG plus gamma model. **C.** ML tree (RAxML) using the RTREVF plus gamma model. Bootstrap support greater than 50% is shown under branches. The light blue rectangle indicates primarily mono/bi-modular NRPS; the SID and EAS subclasses are primarily multimodular. Color coding for subfamilies: brown: adenylating enzyme outgroups; light green: fungal PKS;NRPS hybrid synthetases (PKS;NRPS); dark orange: ChNPS11/ETP module 1 synthetases (ChNPS11/ETP mod 1); dark blue: ChNPS12/ETP module 2 synthetases (ChNPS12/ETP mod 2); yellow: ChNPS10-like synthetases (ChNPS10); light blue: Cyclosporin synthetases (CYCLO); pink: α -aminoadipate reductases (AAR); dark green: ACV synthetases (ACV); red: siderophore synthetases (SID); purple: Euscomycete clade synthetases (EAS). The majority of bacterial sequences (dark gray) group together and contain some fungal A domains (ACV synthetases and the NPS;PKS hybrid (ChNPS7;PKS24) suspected of being horizontally transmitted from bacteria to fungi. The remaining bacterial A domains group with the mono/bi-modular AAR and ChNPS12/ETP mod 2 subfamilies.



Exceptions to the predominately mono/bi-modular fungal NRPS structures include the ACV synthetases and the clade containing A domains from the eleven modules of SimA (cyclosporin biosynthesis) [58] and from several related fungal NRPSs. The other large group contains exclusively fungal and primarily multimodular NRPSs and includes siderophore synthetases and a group we term the Euascomycete-only synthetases, as its members are restricted to euascomycetes. Both grouped together with greater than 97% bootstrap support in analyses of a reduced dataset which included selected representatives from each subfamily (Figure 3.2, red arrow, Appendix 3.7).

Phylogenetic analyses identified nine major subfamilies of fungal NRPSs. Subfamilies were defined as the most internal branch from the root node that formed a monophyletic group which was supported by greater than >70% bootstrap support, shared identical taxon composition across all three phylogenetic methods, and contained a representative fungal NRPS. These groups were named after a representative *C. heterostrophus* or other fungal NRPSs of well-known function in the group (Figures 3.1, 3.2, Appendix 3.6). Subfamilies include: 1) fungal PKS;NRPS hybrids, 2) ChNPS11/ETP toxin module 1 synthetases, ChNPS12-like /ETP module 2 toxin-like synthetases, 4) ChNPS10-like synthetases, 5) Cyclosporin synthetases (CYCLO), 6) α - amino adipate reductases (AAR), 7) ACV synthetases (ACV), 8) siderophore synthetases (SID), and 9) the Euascomycete-only synthetases (EAS). Deep phylogenetic relationships among mono/bi-modular subfamilies were unresolved and lacked bootstrap support (Figures 3.1, 3.2, Appendix 3.6A-C). A domains from a few ascomycete (BC1G11613.1, MGG 14967.5, MGG07803.5) and several urediniomycete (UM05245.1, Sr31423, and PGTG06519.1) proteins did not group with any of the major subfamilies and were not placed consistently in the trees when assessed by different phylogenetic methods.

Figure 3.2. ML phylogenetic tree (PhyML, WAG plus gamma) from the reduced A domain dataset. Branches corresponding to subfamilies are color coded as in Figure 3.1 and known products of NRPSs within each subfamily are shown to the right in parentheses. The representative *C. heterostrophus* NRPS A domains within each subfamily are indicated as red dots. Bootstrap values greater than 50% are shown above branches, where legibility makes this possible. This analysis shows stronger bootstrap support (97%) for grouping the exclusively fungal, multimodular subfamilies, SID and EAS subfamilies together (arrow). Double arrow indicates high bootstrap support (>85%) for grouping ChNPS11/ETP/ChNPS12 together.

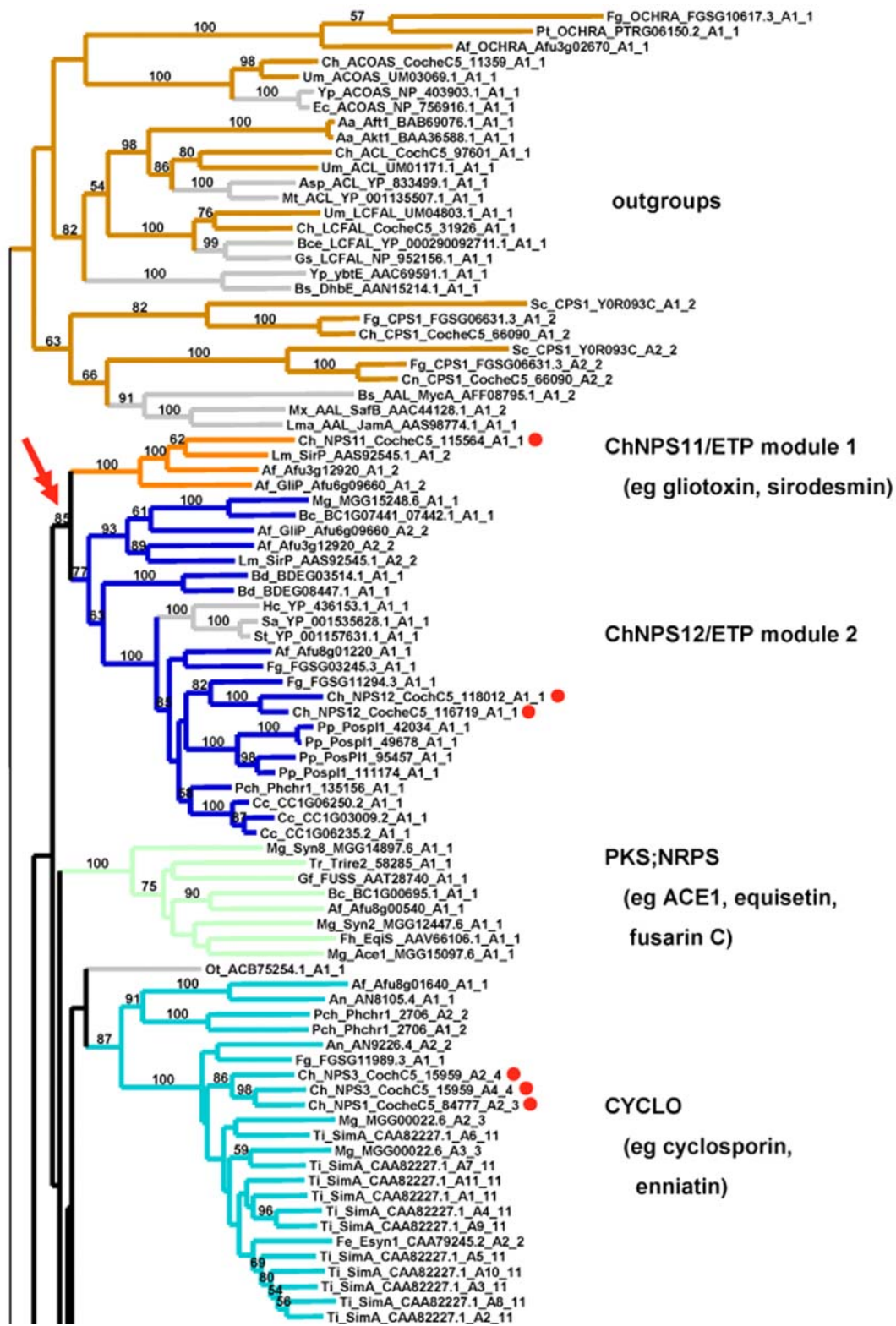


Figure 3.2 Continued

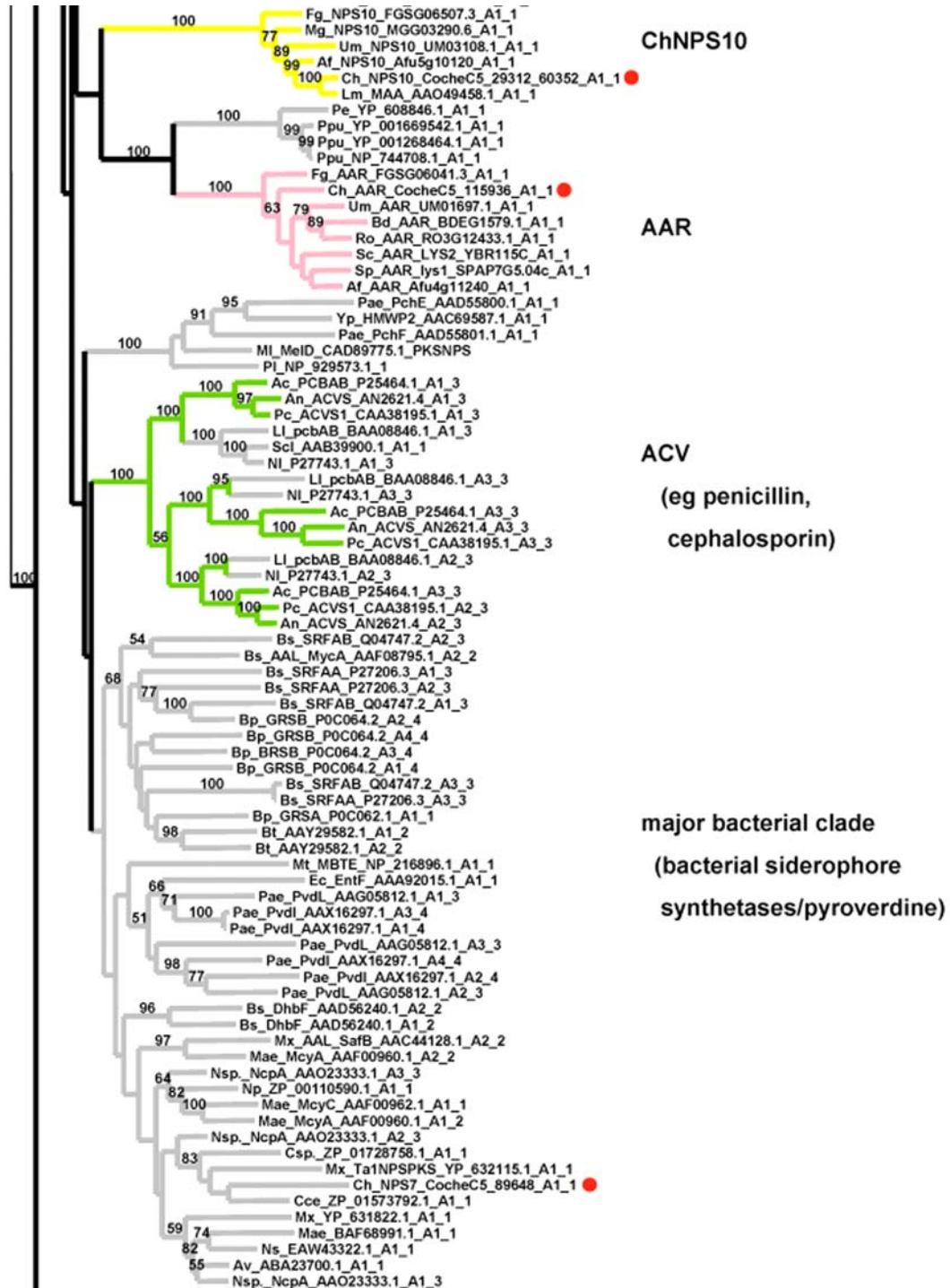


Figure 3.2 Continued



Homologs of bimodular *A. fumigatus* SidE, a putative siderophore synthetase [37], formed two clades corresponding to each module and consistently grouped with the SID subfamily but without bootstrap support in the larger phylogeny and with low bootstrap support (>50%) in the reduced phylogeny. We term this group SIDE but do not consider them as a major subfamily (Figure 3.2, 6A-C, Appendix 3.7).

3.3.3 Relationships Between Fungal and Bacterial NRPSs: Horizontal Transfer or Vertical Transmission and Massive Loss?

The majority of bacterial sequences (Appendix 3.8), identified as top hits in blast searches using a representative from each of the major fungal NRPS subfamilies to query the public databases, were eubacterial in origin and formed a monophyletic group (although lacking bootstrap support), which we term the major bacterial clade (Figures 3.1, 3.2, gray, Appendix 3.6). This clade contains two fungal representatives suspected of being horizontally transmitted from bacteria to fungi. One is the fungal ChNPS7;PKS24 hybrid NRPS;PKS synthetase which is nested within this clade; previous independent analyses of both the NRPS [10] and the PKS portion of this protein [29] found the same placement (Figure 3.2, Appendix 3.6). The other is the ACV synthetases, a group postulated to have been horizontally transferred from bacteria to fungi [59-64], which groups as sister to, or within, the major bacterial clade (Figures 3.1, 3.2, Appendix 3.6). Our analysis also shows that each of the three fungal ACV synthetase A domains groups with the corresponding bacterial A domain rather than forming separate clades of fungal and bacterial A domains. These results support previous claims of horizontal transfer based on observations of closer sequence similarity than expected between these fungal and bacterial genes [61-64] (Figure 3.2, Appendix 3.6).

In contrast, bacterial siderophore synthetases (eg. Pyoverdine (PvdD, PvdI, PvdJ, PvdL), yersiniabactin (ybtE), and Pyochelin (PchE, PchF)) group separately from fungal NRPSs (SID) that biosynthesize intracellular siderophores and fungal NRPSs (NPS6) in the EAS subfamily that biosynthesize extracellular siderophores (Figure 3.2, Appendix 3.6). This suggests that fungal and bacterial capacities to chelate iron via small molecule siderophores have evolved independently (Figure 3.2).

The remaining bacterial A domains included in this study that grouped with high bootstrap support with fungal A domains were associated with the ChNPS12-like/ETP module 2 and AAR subfamilies (Figure 3.2, Appendix 3.6). In the two cases of proposed horizontal transfer discussed above (e.g., ChNPS7 [10, 29], and ACV [59-64] synthetases), the fungal genes are nested within a large clade of bacterial sequences. The reverse phylogenetic situation is observed for bacterial genes grouping with the AAR and ChNPS12 subfamilies as, in these cases, bacterial NRPSs are nested within a large clade of fungal NRPSs (ChNPS12) or group as sister to fungal NRPSs (AARs). These placements suggest that either the fungal genes were transferred to bacteria or that the origin of these groups predates the divergence of eukaryotes and prokaryotes and the observed pattern reflects extensive loss or incomplete sampling from bacteria. Clearly, further sampling of bacterial sequences is needed to adequately address these hypotheses, but we favor the theory that these NRPS subfamilies may have originated prior to the divergence of prokaryotes and eukaryotes. We hypothesize that the lack of phylogenetic signal for resolving relationships among the fungal mono/bi-modular subfamilies may in part reflect an ancient and rapid radiation of these groups.

3.3.4 Distribution of NRPS Subfamilies Across Fungal Taxonomic Groups

The distribution of fungal NRPS subfamilies across the major fungal taxonomic groups supports previous findings that NRPSs are much more abundant in Eucoscomycetes than in Basidiomycetes and are scarce in Chytridiomycota, Zygomycota, Schizosaccharomycota, and Hemiascomycota [10] [65, 66]. The number and distribution of NRPSs in each subfamily are shown in Table 3.1. EAS and PKS;NRPS subfamilies were significantly overrepresented in Eucoscomycete taxa when evaluated by Fisher's exact tests, while ChNPS12-like synthetases were statistically overrepresented in Basidiomycete taxa (Table 3.1, asterisks). The Chytridiomycota, Zygomycota, Schizosaccharomycota, and Hemiascomycota contained only a few NRPSs. All Zygomycota and Hemiascomycota lacked genes encoding NRPS-type proteins other than a single AAR. The chytrid genome contained two additional NRPS-like proteins grouping with the ChNPS12/ETP module 2 subfamily, and the two Schizosaccharomycota taxa examined contained one additional NRPS for siderophore biosynthesis (Table 3.1). No subfamilies were statistically overrepresented in these groups.

Table 3.1. Numbers of NRPSs per subfamily across fungal taxonomic groups^a

Species	PKS; NRPS	NPS11/ETP mod 1	NPS12/ETP mod 2	NPS10	CYCLO	SID	ACV	AAR	EAS	Other ^b	Total
Ascomycota	*								*		
<i>A. fumigatus</i>	1	2 ^c	1	1	1	1	0	1	10	2	20
<i>A. nidulans</i>	1	0	1	1	2 ^d	1	1	1	11	0	19
<i>B. cinerea</i>	3 ^e	1	0	0	0	3	0	1	5	1	14
<i>C. immitis</i>	1	0	0	0	0	1	0	1	5	0	8
<i>C. heterostrophus</i>	0	1	2	1	2 ^d	1	0	1	6	1	15
<i>F. graminearum</i>	1	0	3	1	1	2	0	1	12	0	21
<i>M. oryzae</i>	6	1	1	1	1 ^d	1	0	1	4	2	18
<i>N. crassa</i>	0	0	0	0	0	1	0	1	2	0	4
<i>P. anserine</i>	4	0	1	0	0	1	0	1	4	1	12
<i>T. reesii</i>	2	2 ^c	1	1	0	1	0	1	5	0	13
Basidiomycota			*								
<i>C. cinerea</i>	0	0	3	0	0	1	0	1	0	0	5
<i>C. neoformans</i>	0	0	0	0	0	0	0	1	0	0	1
<i>L. bicolor</i>	0	0	0	0	0	0	0	1	0	0	1
<i>P. chrysosporium</i>	0	0	1	0	1	0	0	1	0	0	3
<i>P. stipitis</i>	0	0	0	0	0	0	0	1	0	0	1
<i>P. placenta</i>	0	0	8	0	0	0	0	2	0	0	10
<i>P. graminis</i>	0	0	0	0	0	0	0	1	0	1	2
<i>S. roseus</i>	0	0	0	0	0	0	0	1	0	1	2
<i>U. maydis</i>	0	0	0	1	0	2	0	1	0	1	5
Schizosaccharo- Mycota											
<i>S. japonicus</i>	0	0	0	0	0	1	0	1	0	0	2
<i>S. pombe</i>	0	0	0	0	0	1	0	1	0	0	2
Hemiascomycota											
all species	0	0	0	0	0	0	0	1	0	0	1

Table 3.1 Continued

Zygomycota											
<i>P. blakesleeanus</i>	0	0	0	0	0	0	0	1	0	0	1
<i>R. oryzae</i>	0	0	0	0	0	0	0	1	0	0	1
Chytridiomycota											
<i>B. dendrobatidis</i>	0	0	2	0	0	0	0	1	0	0	3
Microsporidia											
<i>E. cuniculi</i>	0	0	0	0	0	0	0	0	0	0	0

^a Based on inclusion in clades of phylogenetically defined subfamilies (Figure 3.1, Appendix 3.6).

^b Several proteins which grouped with NRPSs did not group with any of the 9 major subfamilies. These include homologs of SidE (Afu3g03350) in *A. fumigatus* and *P. anserina* species (Afu3g15270, Pa2_7870), several proteins in the urediniomycetes *U. maydis*, *S. roseus*, and *P. graminis* (UM05245.1, Sr31423, and PGTG06519), two proteins in *M. oryzae* (MGG 14967.5, MGG07803.5), and one in *B. cinerea* (BC1G11613.1).

^c *A. fumigatus* has two bi-modular NRPSs (Afu_6g09660, Afu_6g09660) as does *T. reesii* (Trire2_24586, Trire2_60458). The first modules of all four NRPSs group with the ChNPS11/ETP module 1 subfamily; the second modules group with the NPS12 subfamily. For tallying purposes, Afu_6g09660, Afu_6g0966, Trire2_24586, and Trire2_60458 were attributed to the ChNPS11/ETP module 1 subfamily.

^d ChNPS1 and ChNPS3 modules 1 and 3 group with the EAS subfamily, while ChNPS1 module 2 and ChNPS3 modules 2 and 4 group with the CYCLO subfamily. For tallying purposes, ChNPS1 and ChNPS3 were attributed to the CYCLO subfamily. Similarly, MGG00022.1 and AN9226 also have some A domains in the CYCLO subfamily and others in the EAS subfamily. For tallying purposes, these genes were included in the CYCLO subfamily.

^e *B. cinerea* contains 3 PKS;NRPS hybrids. For one of these BC1G15479.1, the A domain did not align well and was missing several core motifs. This protein was removed from the final phylogenetic analysis.

* PKS:NRPS and EAS subfamilies in Euascomycetes and the NPS12/ETP module 2 subfamily in Basidiomycetes are statistically over-represented

3.3.5 Lineage Specific Expansions and Contractions

When patterns of gene duplication and loss were analyzed for the total number of NRPSs/genome (combining all subfamilies) over the tree of fungi (Figure 3.3; Appendix 3.9), a highly significant expansion was found on the branch leading to Euascomycetes ($p = 7 \times 10^{-5}$). Significant expansions were also found within euascomycetes on the branches leading to the *Aspergillus* species ($p = .028$), to *F. graminearum* ($p = .011$) and to *M. oryzae* ($p = .032$). *N. crassa* showed a highly significant ($p = 5 \times 10^{-5}$) contraction in total number of NRPS genes (Figure 3.3), likely due to the efficiency of RIP and/or other genome defense mechanisms [67, 68].

Our data support previous findings [66], including our own [10], that unicellular fungi have few, if any, genes for secondary metabolism (Table 3.1, Figure 3.3). Ancestral reconstructions show that in hemiascomycete yeasts, this is due to loss of all NRPSs, except for a single AAR gene, that were present in basidiomycetes and inferred to be present in the ancestor of ascomycetes (Figure 3.3). However, both the fission yeast *S. pombe* and the unicellular basidiomycete yeast *Sporobolomyces roseus* contain one additional NRPS (a siderophore synthetase and an unknown, respectively) in addition to the single AAR gene, suggesting that a unicellular habit may not preclude the existence of secondary metabolite genes such as NRPSs. Patterns of expansion and contraction also do not seem to occur preferentially in fungal pathogens versus nonpathogens. While a number of pathogenic fungi (e.g., *F. graminearum*, *A. fumigatus*, and *M. oryzae*) do show evidence for expansions in numbers of NRPS, we also see expansion in the nonpathogen, *A. nidulans*.

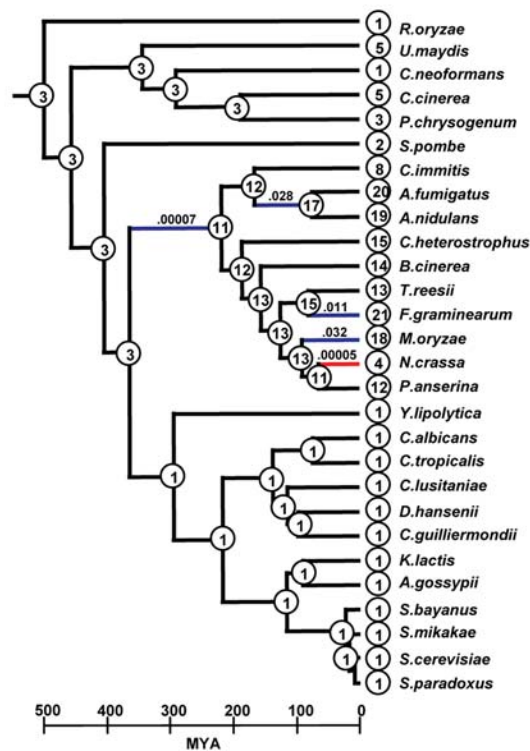


Figure 3.3. Lineage specific expansions and contractions in number of NRPS genes per genome. Inferred number of NRPS encoding genes at ancestral nodes mapped onto the ultrametric tree of fungi. Timescale in millions of years is shown along bottom. Branches with significant expansions (blue) or contractions (red) are shown with associated p-values above branches. The largest contraction in number of NRPSs occurs in *N. crassa* while the largest expansion occurs in the ancestor of euascomycetes. A highly significant expansion also occurs in *F. graminearum* and significant expansions occur in several other euascomycete taxa (e.g., *M. oryzae* and on the branch leading to the *Aspergillus* species).

3.3.6 Subfamily Distribution

AAR Subfamily: A single ortholog of *S. cerevisiae* Lys2, the AAR involved in reduction of α -amino adipic acid in the fungal lysine biosynthetic pathway [23, 69], was found in all fungi surveyed except the Microsporidian, *Encephalitozoon cuniculi*, an intracellular parasite which has lost the majority of genes involved in amino acid biosynthesis [70] and the basidiomycete *Postia placenta*, which appears to contain two (Table 3.1)(Appendix 3.2).

ChNPS11/ETP/ChNPS12: In a phylogeny of a reduced set of representative A domains from each subfamily (Figure 3.2), homologs of ChNPS11, ChNPS12, and the ETP toxin synthetases, GliP for Gliotoxin and SirP for Sirodesmin production, group together with strong bootstrap support (>80%), suggesting all share a common evolutionary origin. In the larger phylogeny of the complete dataset (Figure 3.1, Appendix 3.6), they formed two separate clades each supported by >70% bootstrap support, but lacked this level of support for the entire group. The first clade (ChNPS11/ETP module 1) includes the first module of the ETP toxin synthetases and monomodular ChNPS11. The second module of the ETP toxin synthetases, however, groups within a larger clade containing the two NRPSs from the chytrid genome, several eubacterial NRPSs, and a clade containing both euascomycete and basidiomycete homologs of ChNPS12 (ChNPS12/ETP module 2). While fungal NRPSs associated with ChNPS11 and ETP toxin synthetases are found only in Euascomycota, NRPSs from both eubacteria and from the most basal fungal group, Chytridiomycota, were nested within this larger clade with high bootstrap support (>80%) (Figures 3.1, 3.2).

ChNPS10, CYCLO, SID: Three subfamilies, monomodular ChNPS10, NRPSs grouping with SIMA (CYCLO), and NRPSs (SID) involved in intracellular (primarily) siderophore biosynthesis, contain representatives from both Basidiomycota and Euascomycota. While all euascomycetes and many basidiomycetes examined contain at least one representative from the SID subfamily (Table 3.1) [65], ChNPS10 and CYCLO are more discontinuously distributed and a representative is not found in all taxa (Table 3.1, Appendix 3.6).

ACV and PKS;NRPS subfamilies: PKS;NRPSs were restricted to and statistically overrepresented in euascomycetes. As has been noted previously [29, 71], all fungal PKS;NRPS hybrids fall into a single, well supported, monophyletic group,

which suggests a single origin (Table 3.1). However, not all ascomycetes have a representative of this group and the number of corresponding genes varies widely among taxa (Table 3.1). *C. heterostrophus*, for example, lacks a representative but *M. oryzae* has six. While ACV synthetases are found in both bacteria and fungi, within fungi, they appear restricted to Eurotiomycete and Hypocrealean taxa. This study did not identify any additional ACV synthetases in fungi apart from the known ones in *Penicillium chrysogenum*, *Aspergillus nidulans*, and *Cephalosporium acremonium* (Appendices 2, 6), supporting previous conclusions that their distribution is likely the product of one or more isolated horizontal transfer events [59, 61-64].

The Euascomycete (EAS): The EAS subfamily contains by far the greatest number of NRPSs and is both restricted to and statistically overrepresented in Euascomycetes (Table 3.1).

3.3.6 Hypothesized Origins Based on Taxonomic Distribution

Figure 3.4 shows the hypothesized origins of each subfamily based on taxonomic distribution of the oldest member of each group. By this criterion, the presence of bacterial sequences grouping within the ChNPS11/ETP module 1 and ChNPS12/ETP module 2 clade suggests that the origins of these groups may predate the divergence of eubacteria and eukaryotes (Figures 3.2, 3.4). The AAR subfamily must have arisen also either prior to or very early in the origin of the fungi as a representative is present in all fungi, including the most basal group, the Chytridiomycota (Table 3.1, Figures 3.2, 3.4). Since the SID, CYCLO, and ChNPS10 subfamilies all contain representatives from both Euascomycota and Basidiomycota, these groups must have evolved prior to the divergence of the Dikarya (Figure 3.4). The EAS, PKS;NRPS, and ACV synthetases contained only euascomycete

representatives. Both PKS;NRPS and EAS may thus have originated in the ancestor of euascomycetes (Figure 3.4).

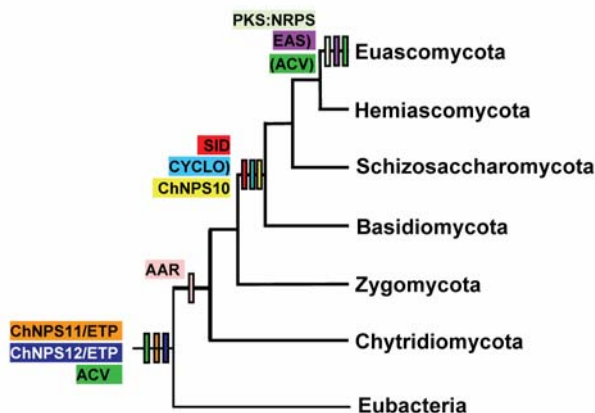


Figure 3.4. Hypothesized origins of major fungal NRPS subfamilies based on the oldest member of each subfamily. Subfamilies color coded as in Figure 3.1. AAR, and ChNPS11/ETP module 1 and ChNPS12/ETP module 2 likely originated prior to or early in the divergence of fungi. AAR genes are present in all fungi, while the ChNPS11/ETP/ChNPS12 clade contains representatives of the most ancestral fungal group, the Chytridiomycota, as well as bacterial sequences that nest with high bootstrap support within this clade. Although ACV genes are clearly present in eubacteria, they appear to have been horizontally transferred to euascomycete fungi, hence their dual placement. The CYCLO, ChNPS10, and SID subfamilies were found in Basidiomycota, Schizosaccharomycota, and Euascomycota and thus likely originated in an ancestor of the Dikarya. Fungal PKS;NRPS hybrids and EAS were found only in Euascomycetes.

As discussed above, the grouping of fungal ACV synthetase A domains with the corresponding A domains of bacterial ACV synthetases within a large clade of bacterial sequences provides evidence for horizontal transfer and suggests that this group originated within prokaryotes (Figure 3.4).

Thus, taxonomic distributions suggest a more ancient origin of several of the mono/bi-modular NRPS subfamilies (ChNPS11/ETP/ChNPS12, ACV, and AAR), possibly predating the divergence of eubacteria and fungi (Table 3.1, Figure 3.4). This hypothesis is tenable given that the strongly supported co-grouping of fungal and

bacterial outgroup adenylating enzymes (Figure 3.2, Appendix 3.6A-C) demonstrates that these enzyme classes originated prior to the divergence of bacteria and fungi. In contrast, the fungal-specific multimodular groups (SID and EAS), which group together with high bootstrap support in the reduced phylogeny (Table 3.1, Figure 3.2, Appendix 3.6A-C), appear to be of more recent origin and are restricted to and highly expanded in fungi.

3.3.7 Mono- and Bi- Modular NRPS Subfamilies

Unlike many of the multimodular NRPSs, most monomodular subfamilies lack a complete NRPS module (A-T-C) and consist of a single A domain or an A-T domain combination followed by a variety of C-terminal domains (Figure 3.5). Many of the mono/bi-modular groups show a conserved domain architecture across all members in a subfamily, suggesting their domain architectures may be functionally constrained. Available functional data suggest that the NRP products of several of these groups may play more central roles in cellular metabolism related to responses to oxidative stress and growth and development.

Whether monomodular NRPSs may act alone or in concert with non-NRPS proteins is currently unknown. However, in bacterial systems, both single A domains as well as A-T domain units, known as initiation modules, can interact with other NRPS proteins and accomplish biosynthesis by first activating and then transferring the activated substrate either to a C domain in the same NRPS or to a C domain in a different NRPS (nonlinear biosynthesis) [5].

AARs and Lysine Biosynthesis: AARs are conserved not only taxonomically but also in terms of domain structure. All have an identical structure consisting of an A-T unit followed by a thioester reductase (R) domain (IPR010080), a member of the

NAD(P)-binding Rossmann fold domain superfamily (SSF51735). There are two primary pathways for lysine biosynthesis, the diaminopimelic acid pathway (DAP), found predominantly in bacteria and plants, and the α -aminoadipate pathway (AAA), found primarily in fungi and a few bacteria [69]. As noted above, AARs catalyze reduction of α -aminoadipic acid in the AAA pathway [69]. The fact that AARs have a C-terminal R domain in common with several other NRPS subfamilies (PKS;NRPS, ChNPS10, EAS, discussed below) supports our conclusions based on phylogenetic relationships that AARs are more closely related to NRPSs than other adenylating enzymes (Figure 3.5).

Bacterial sequences grouping with fungal AAR are comprised of a single A domain followed by an acyl-transferase domain (PFAM01757) but lack the C-terminal R domain found in fungal AARs. We conclude that they are likely not involved in lysine biosynthesis in bacteria. Although there is evidence for the existence of lysine biosynthesis through the AAA pathway in some prokaryotes [72], current data suggests that these pathways do not include a step involving reduction of α -aminoadipic acid [72]. Thus, our data support previous conclusions that AARs are fungal-specific enzymes [73-75].

PKS;NRPS: Nearly all fungal PKS;NRPS hybrids have the same domain structure (KS-AT-M-KR-ACP-C-A-T-R) (Figure 3.5, Appendix 3.2). The terminal R domain has been reported previously in several PKS;NRPS hybrids [76-78].

ChNPS10: The ChNPS10 subfamily also has a conserved domain architecture across all genes in the subfamily, consisting of an A-T unit followed by two additional C-terminal domains. The first is a NAD(P) binding domain (IPR016040) also showing closest similarity to thioesterase reductase (R) domains and the second is a dehydrogenase domain with closest hits to ADH short chain dehydrogenases (IPR002198) (Figure 3.5).

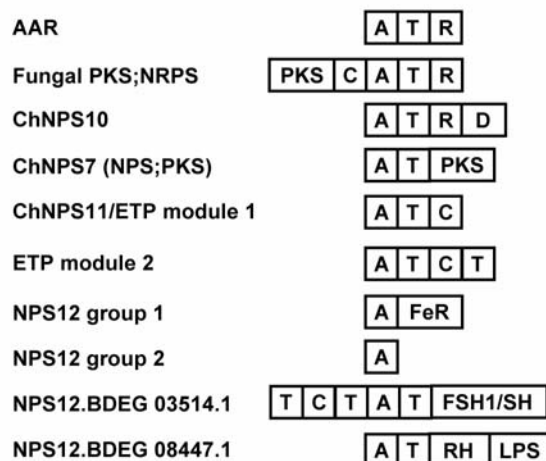


Figure 3.5. Conserved domain architectures for mono-bimodular NRPS subfamilies. The majority of mono-bimodular subfamilies have an A-T domain structure followed by various C-terminal domains. Only ChNPS11/ETP module 1 and ETP module 2 show complete A-T-C modules. The ChNPS12/ETP module 2 subfamily also contains representatives consisting of a single A domain. Domains: A = adenylation, T = thiolation, C = condensation, R = thioester reductase, D = ADH short chain dehydrogenase, PKS = polyketide synthase module, FeR = ferric reductase, FSH/SH = serine hydrolase, RH = polynucleotidyl transferase, Ribonuclease H, LPS = LPS-induced tumor necrosis alpha factor.

ChNPS11/ETP/ChNPS12: The large and highly diverse clade of ChNPS11/ETP/ChNPS12 homologs reveals the diversity of C-terminal domains that can follow A-T units and shows that, as for some bacterial NRPSs, fungal NRPS or NRPS-like proteins can consist of single A domains (Figures 3.5, 3.6).

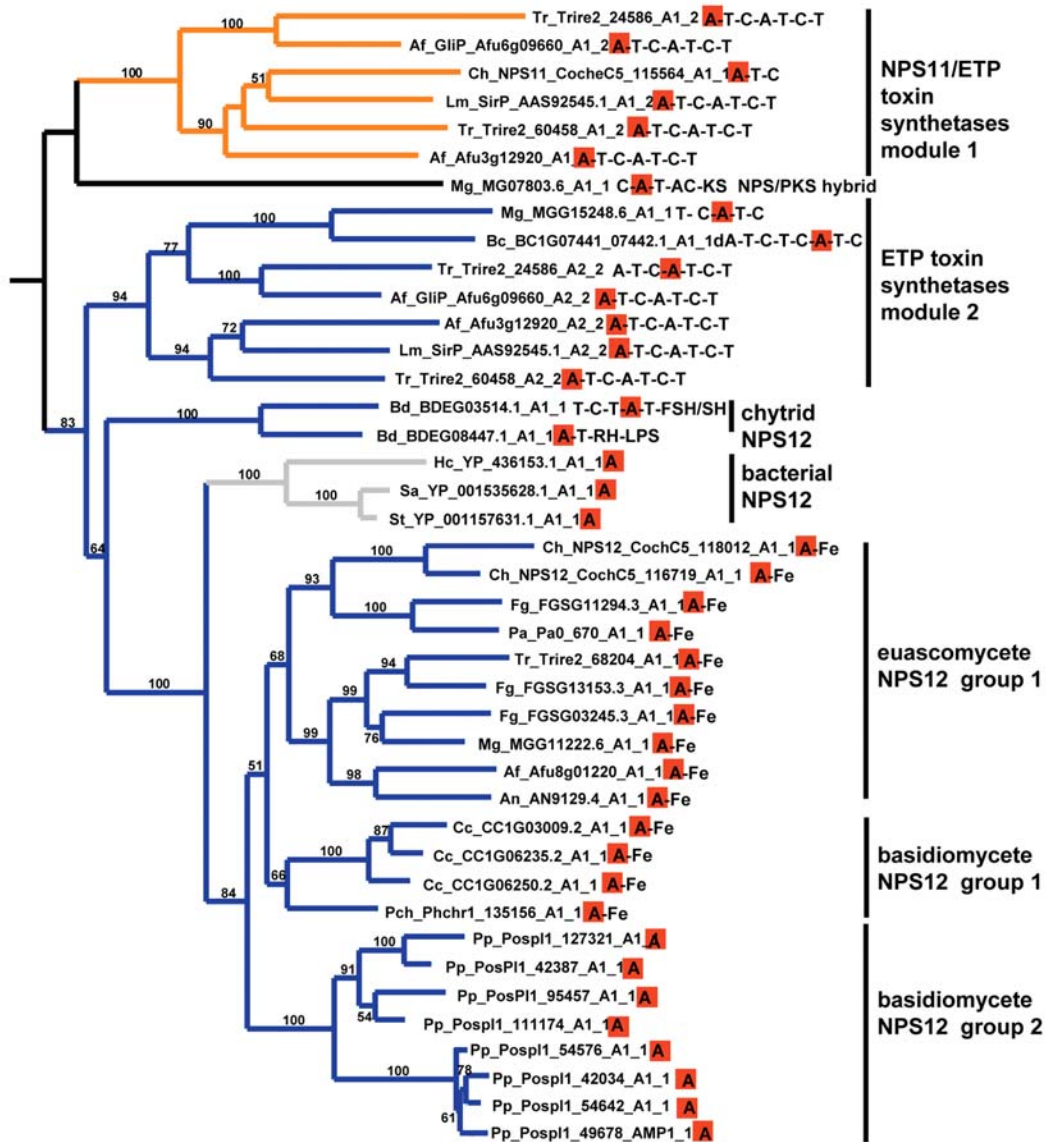
At the base of this group are monomodular ChNPS11 and module 1 of the bimodular ETP toxin synthetases, SirP and GliP, which contain complete A-T-C modules (Figure 3.6). Module 2 of SirP and GliP groups at the base of the ChNPS12/ETP module 2 clade. The second module of the ETP toxin synthetases contains a complete module followed by an additional T domain (A-T-C-T) (Figure 3.6). This group also contains several fungal proteins with an incomplete (MGG15248.6) or a degenerate (BC1G07441_07442.1) first module (Figures 3.2,

3.5). Nested within this clade is a group of bacterial NRPSs with a single A domain and two NRPS-like proteins from the chytrid *B. dentrobatidis* (Figure 3.6). One of the chytrid NRPSs (BDEG_03514.1) has a T-C-T-A-T domain architecture followed by a domain with similarity to FSH1 (IPR005645), a serine hydrolase domain. The other chytrid protein (BDEG_08447.1) has an A-T unit followed by two additional domains. The first shows closest similarity to polynucleotidyl transferase, Ribonuclease H fold (IPR012337), a domain associated with nucleic acid binding functions and found in a variety of proteins including HIV RNase H, transposases, and exonucleases [79, 80] (Figure 3.6).

The second domain shows closest similarity to the membrane-associated domain LPS-induced tumor necrosis factor alpha factor (LITAF, IPR006629, PF10601), which contains a characteristic cysteine rich zinc-binding motif found also in intracellular Zn²⁺ binding proteins and animal transcription factors. The zinc and DNA-binding domains found in the chytrid NRPSs are intriguing (Figure 3.5). Gliotoxin and Sirodesmin PL have been shown to inhibit viral reverse transcriptase [81] and general transcription [82], respectively. In the case of Sirodesmin PL, the addition of zinc and other IIB series metals (Hg and Cd) both decreases toxin production in *Leptosphaeria maculans* and also reverses the inhibition of transcription, suggesting interactions of Sirodesmin PL with either cellular zinc or zinc-containing metalloenzymes such as RNA polymerases [82, 83]. Whether these phenotypes relate to our identification of Zn-binding domains in the corresponding chytrid NRPS is unknown.

ChNPS12 (CocheC5_118012), and its paralog (CocheC5_116719) contain a single A domain followed by a domain showing closest similarity to a ferric reductase transmembrane domain (IPR013130).

Figure 3.6. Phylogeny of the ChNPS11/ETP/ChNPS12 subclade. Extracted from maximum likelihood (PhyML with WAG plus gamma substitution matrix) phylogeny of complete A domain dataset (Appendix 3.6B). Domain structure of each NRPS is shown to the right of species abbreviation and accession number. Orange highlighted A domains reflect corresponding A domain in the phylogeny. Orange branches = ChNPS11/ETP mod1 and blue = ChNPS12/ETP mod2 subfamilies. ChNPS11 is monomodular, while all other NRPSs in the ETP module 1 group are bimodular; all have complete A-T-C modules. The A domain from a *M. oryzae* NRPS;PKS (MG07803.6) also groups here. Members of the ChNPS12 subfamily show a diversity of C-terminal domains as described in text. The group includes two putative NRPSs from the chytrid, *B. dendrobatidis*, two proteins with either an incomplete (MGG15248.6) or a degenerate (BC1G07441_07442.1) first module and monomodular bacterial proteins consisting of single A domains. ChNPS12 homologs in Basidiomycete NPS12 group 2 consist of proteins with single A domains which appear to lack additional C-terminal domains and are highly expanded in the basidiomycete *Postia placenta*.



The closest homologs of ChNPS12 (Figure 3.7, ChNPS12 group 1) are present in both euascomycete and basidiomycete group 1 and have the same domain structure as the *C. heterostrophus* NPS12 proteins (Figure 3.6). Sister to all group 1 NPS12-like proteins is a group of proteins consisting of standalone A domains (Figure 3.6, NPS12 group 2). These were found only in the brown-rot fungus, *Postia placenta*, which carries eight closely related copies.

The monomodular bacterial NRPSs nested within the ChNPS12/ETP module 2 subfamily also consist of a standalone A domains. As noted earlier, for many bacterial NRPS systems (e.g., VibE, MxcE, and YbtE), single A domains may be involved in NRPS biosynthesis by activating and transferring the activated substrate to a different NRPS [5]. Only one example of this type of synthesis has been reported for fungi (e.g., *C. purpurea* ergot alkaloid biosynthesis) [5, 84], but our identification of these single fungal A domains grouping with other known NRPSs (e.g., ETP toxins) (Figures 3.2, 3.6, Appendix 3.6) suggests that this mechanism could be more common in fungi than previously appreciated.

The diversity of domain structures found within the ChNPS11/ETP/ChNPS12 group leads us to hypothesize that there may be several distinct functional groups within this clade.

3.3.9 Multimodular NRPS Subfamilies

The majority of multimodular NRPSs are found in the SID and EAS subfamilies. These subfamilies group together with high bootstrap support (>97%) in analyses of the reduced dataset (Figure 3.2). Analyses that included a larger number of bacterial sequences (KE Bushley and BG Turgeon, unpublished) support our phylogenetic and distribution data that the SID and EAS subfamilies are restricted to

fungi. As noted above, two subfamilies containing genes encoding multimodular NRPSs, the CYCLO and ACV synthetases, group with the primarily mono/bi-modular suite of NRPSs. (Table 3.1, Figure 3.2). SID synthetases show a relatively conserved domain architecture, are present in the majority of euascomycetes sampled, and are thought to have evolved by module duplication and selective loss of A domains or complete modules, as described in detail in Bushley et al. (2008) [65].

3.3.9.1 Diversity Within the EAS Subfamily

The EAS subfamily, in addition to containing the vast majority of fungal NRPSs, also shows the greatest diversity of both domain architecture and function (Figures 3.2, 3.7, Appendix 3.2). It includes proteins that are both structurally and functionally conserved (e.g. homologs of ChNPS6 which biosynthesize extracellular siderophores), as well as those that are highly lineage specific (e.g. HTS1 [51] and AMT [52] synthetases for host selective toxins, Tex1 [85] and other peptaibol synthetases, and ergot alkaloid synthetases). The highly diverse domain architectures and discontinuous distribution of corresponding A domains make the identification of orthologs across species extremely challenging.

ChNPS6/PerA: Perhaps the only group for which orthologs can be clearly identified are homologs of the most conserved NRPS in the EAS clade, ChNPS6, which biosynthesizes an extracellular iron scavenging siderophore that serves as a virulence factor for several fungi and is also involved in combating oxidative stress [6,10] (Figures 3.2, 3.6, 3.7). Although ChNPS6 appears to have undergone a gene duplication event, it is single copy in all species examined except *Trichoderma reesii* (Figure 3.7), which contains two paralogous copies. All ChNPS6 homologs have a highly conserved domain structure consisting of a single A-T-C module followed by a

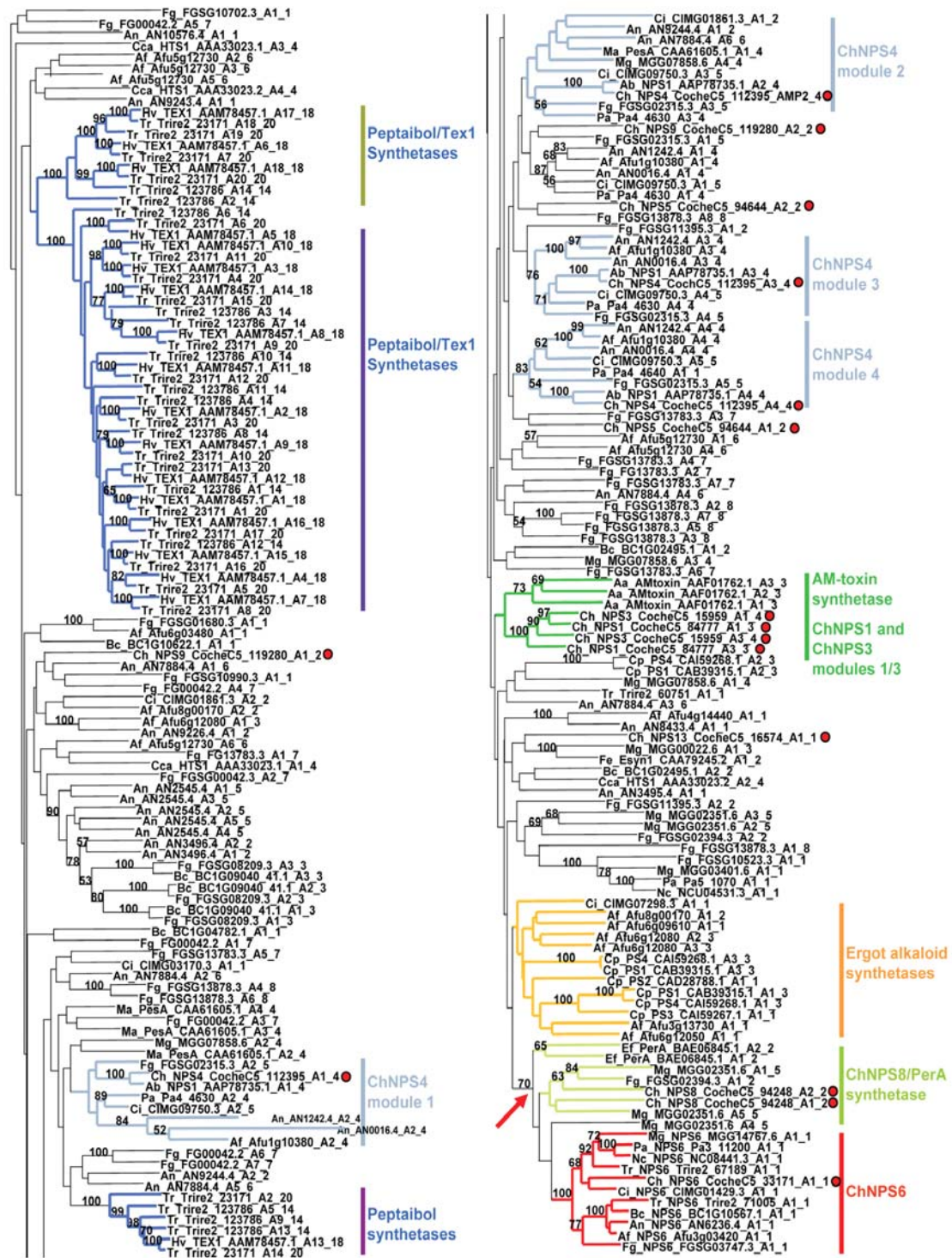
module with a degenerate A domain (dA-T-C) [10]. Sister to the ChNPS6 group is a clade containing both ChNPS8 and an *Epichloe festuca* NRPS, PerA; the latter NRPS mediates symbiotic interactions of *E. festuca* with its grass host by producing an NRP insect deterrent, peramine [86] (Figure 3.7, arrow).

Ergot Alkaloid Synthetases: NRPSs synthesizing ergot alkaloids consistently grouped sister to the ChNPS6 and ChNPS8/PerA clade but without bootstrap support (Figures 3.2, 3.7). These synthetases were found only in animal pathogens in the Eurotiales and grass endophytes such as *C. purpurea* (Figures 3.2, 3.7). Given that grass endophytes such as *C. purpurea* are thought to have an animal pathogenic ancestor [87] and that their ergot alkaloid NRP products have toxic effects on livestock and other animals [88-91], we hypothesize that NRPSs synthesizing ergot alkaloids originally evolved to function in animal pathogenesis.

Peptaibol Synthetases: Peptaibol synthetases, which were restricted to the Hypocrealean taxa examined in this study (*Trichoderma/Hypocrea*), also formed a well supported group. However, as discussed below, several modules of each peptaibol synthetase groups outside of the main clade (Table 3.1, Figures 3.2, 3.7)

Dothideomycete Host-Selective Toxin Synthetases: A domains of the *A. alternata* apple pathotype-specific AMT synthetase which produces the host-selective toxin, AM toxin, grouped consistently with modules 1 and 3 of ChNPS1 and ChNPS3 (discussed below). Modules of tetramodular *C. carbonum* HTS1, responsible for biosynthesis of another host selective toxin, the cyclic tetrapeptide, HC-toxin, grouped in disparate locations in the EAS clade such that clear homologs of HTS1 A domains were not recognizable in any of the species in our dataset (Figures 3.2, 3.7).

Figure 3.7. Phylogenetic analysis of the Euscomycete subclade. Tree extracted from the maximum likelihood (PhyML with WAG plus gamma substitution matrix) phylogeny of the complete A domain dataset (Appendix 3.6B). Branches defining subgroups of the EAS clade grouping with a *C. heterostrophus* NRPS A domains or with A domains from fungal NRPSs with known function are color coded: dark blue = peptaibol synthetases, light blue = ChNPS4 (clades grouping with each A domain of *C. heterostrophus* NPS4), green = AMT synthetases and Ch NPS1 and ChNPS3 modules 1 and 3, orange = ergot alkaloid synthetases, light green = ChNPS8/PerA synthetase, and red = homologs of ChNPS6 (extracellular siderophore synthetases). Of these groups, only the peptaibol synthetases, the clade containing NPS8/PerA/NPS6 synthetases (arrow), and ChNPS4 modules 3 and 4 have bootstrap support >70%. *C. heterostrophus* NRPS A domains are indicated (circles).



Each of the four A domains from each module of tetramodular ChNPS4 groups with strong support with the corresponding A domains of tetramodular AbNPS1 in the closely related Dothideomycete, *A. brassicae*. These A domains group within a larger clade containing *Metarhizium anisopliae* NRPS PesA although without bootstrap support (Figures 3.2, 3.7, Appendix 3.6). However, A domains from NRPSs found in other euascomycetes that group with each of the ChNPS4 modules contain from two to six modules. While some of these A domains are clearly related to those of ChNPS4, module duplication and loss obscures the history of this group.

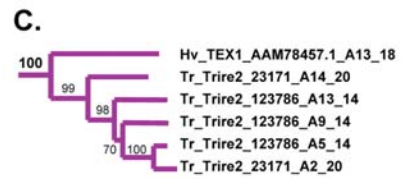
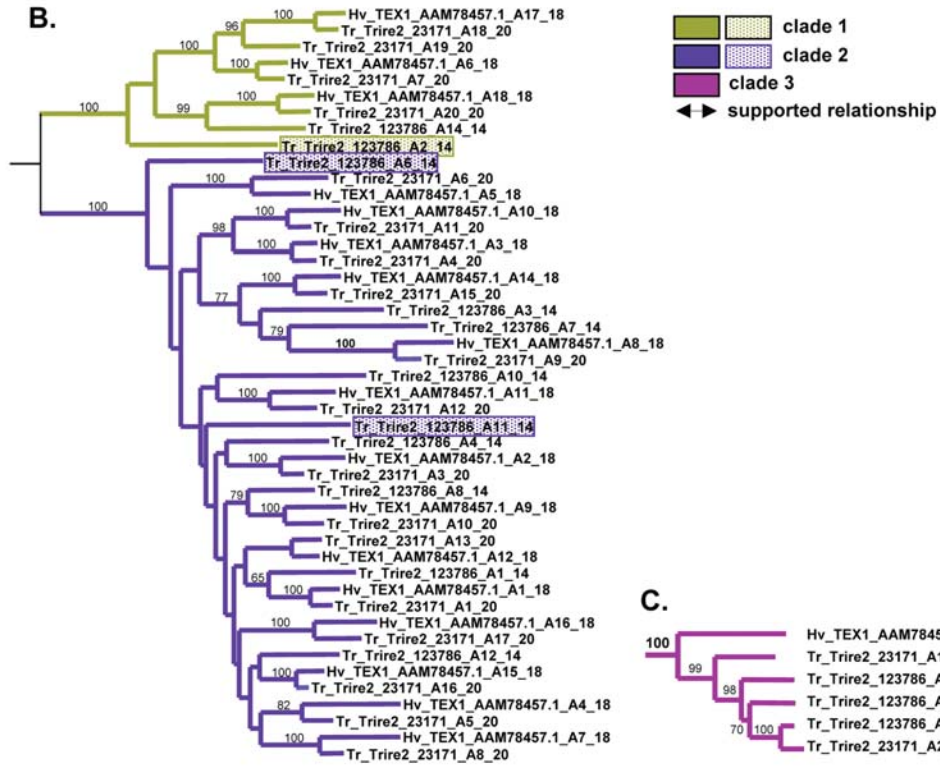
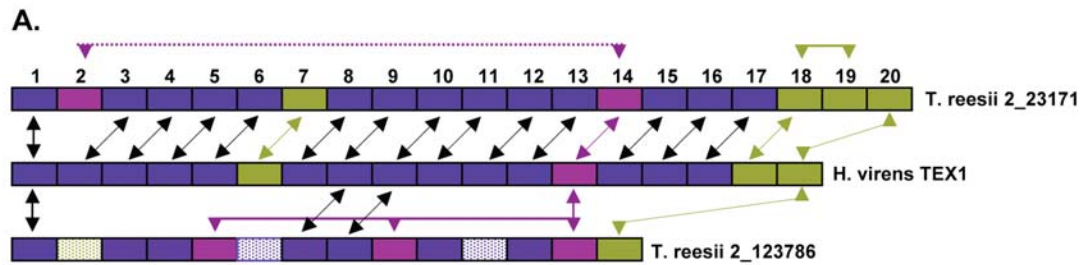
3.3.9.2 Evolutionary Mechanisms Giving Rise to Multimodular NRPSs

The greater diversity of domain architectures seen in multimodular NRPSs is likely due to the multiplicity of evolutionary mechanisms which may generate the corresponding multimodular genes. The EAS subfamily, in particular, contains NRPSs varying from monomodular proteins involved in ergot alkaloid biosynthesis (PS2 and PS4) and ChNPS6 (which has one complete and one degenerate A domain) to the eighteen module TEX1 synthetase responsible for peptaibol biosynthesis in *Trichoderma virens* (*Hypocrea virens*) [85] (Figures. 3.2, 3.7, Appendix 3.6). Several subgroups within the EAS illustrate some of the mechanisms by which the diverse domain architectures of multimodular NRPSs may arise.

Tandem Duplication: Cyclosporin synthetase (SimA) is a clear example of tandem duplication of modules of an NRPS in a single species (*Tolypocladium inflatum*). All eleven A domains from this protein group together as a single well supported monophyletic group (Figure 3.2) which also includes certain A domains from other fungal NRPSs, such as ChNPS1 module 2 and ChNPS3 modules 2 and 4.

Peptaibol synthetases illustrate a more complex process of tandem duplication of modules of an NRPS. Peptaibol synthetases are highly lineage specific and found only within the Hypocreales to date. Using *H. virens* TEX1 as a point of reference, we found that all modules of TEX1 group together in three separate, well-supported clades with modules of two peptaibol synthetases (Trire2_23171 and Trire2_123786) in the related species, *Trichoderma reesii* (Figures 3.2, 3.7). TEX1 module 13 falls outside of the other two TEX1 clades (Figures 3.2, 3.7, 3.8). The nearly one-to-one relationship between modules of TEX1 and modules of *T. reesii* Trire2_23171 suggests that tandem duplication of modules giving rise to these orthologous genes must have occurred prior to divergence of these two species (Figure 3.8). However, at least one additional internal duplication has occurred since divergence from an ancestral species (e.g., note the relationship between *T. reesii* Trire2_23171 modules 18 and 19) (Figure 3.8). The relationship of these two peptaibol synthetases with the *T. reesii* 14 module peptaibol synthetases, Trire2_123786 is less straightforward. However, we note that certain A domains from Trire2_123786 modules 2, 6, and 11 form widowed branches at the base of clades which contain A domains of at least two, and more often, all three peptaibol synthetases (Figure 3.8, stippled boxes). We hypothesize that these may be ancestral domains. Previous studies suggest that like *T. reesii*, *T. virens* also harbors additional NRPSs involved in peptaibol biosynthesis [92]

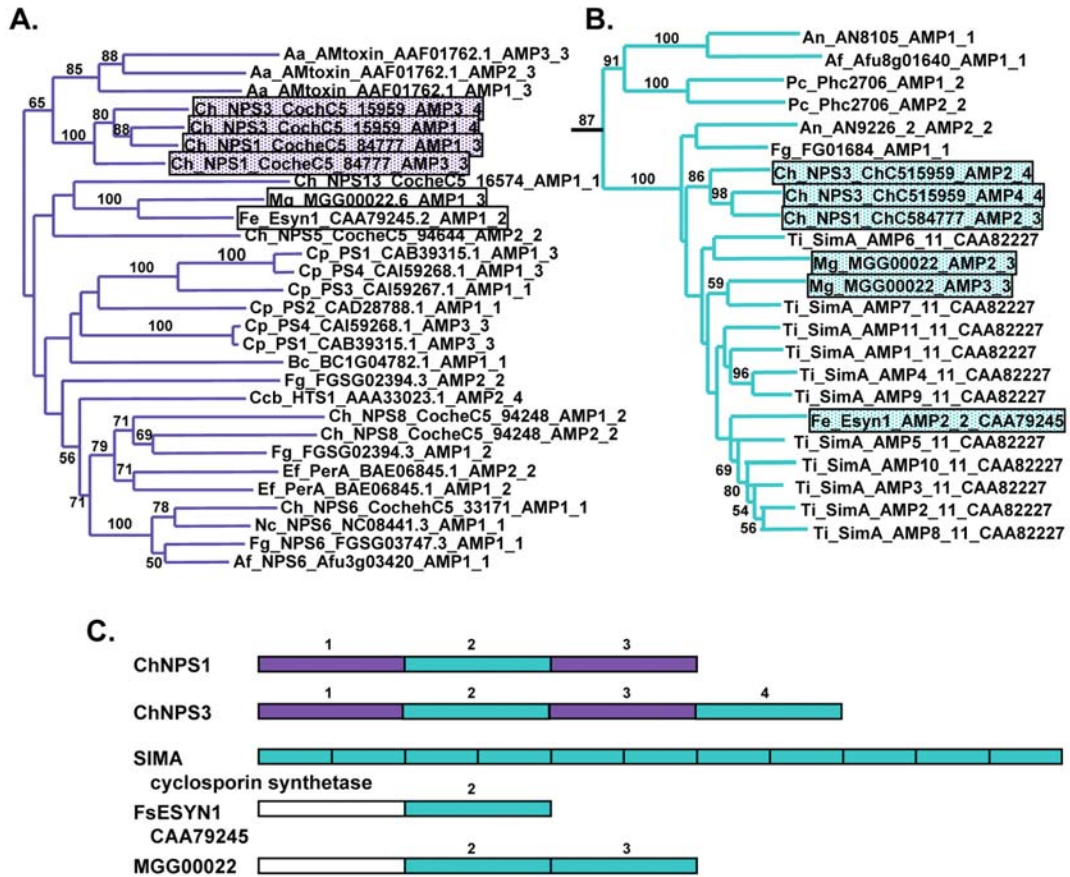
Figure 3.8. Modular organization of Peptaibol synthetases and proposed evolution by tandem duplication. A domains from peptaibol synthetases form three distinct, well-supported clades in the EAS subfamily (Figure 3.7). **A.** Modular structure of the *H. virens* TEX 1 peptaibol synthetase and two peptaibol synthetases in the related species, *T. reesii* (T.reesii 2_23171 and T.reesii 2_123786. Color coding corresponds to clades identified in phylogenetic analyses (B and C, and Figure 3.7). Arrows indicate bootstrap support for module relationships (B, C, and Figure 3.7). While T.reesii 2_23171 is clearly a homolog of TEX1, domain duplication of modules 18 to 19 or vice versa and addition of module 2 have occurred since the common ancestor of these species. **B.** Two of the peptaibol synthetases clades (light green and dark blue, Figure 3.7), group together as a monophyletic group but without bootstrap support. A domains shown in stippled boxes indicate modules from T.reesii 2_123786 which do not have a clear counterpart in the other peptaibols synthetases and may represent ancestral domains. **C.** The third clade (purple, Figure 3.7) groups in a distinct position within the EAS subtree.



Recombination: Two NRPSs found in *C. heterostrophus*, ChNPS1 and ChNPS3, demonstrate the potential role of recombination and modular rearrangement in the generation of multimodular NRPSs. Modules 1 and 3 of both ChNPS1 and ChNPS3 group within the EAS subfamily with AMT synthetase, a lineage specific NRPS found only in a single strain of related *A. alternata* [52] (Figures 3.2, 3.7, Appendix 3.6A-C). Module 2 of ChNPS1 and modules 2 and 4 of ChNPS3, however, group with the CYCLO synthetases among the mono/bi-modular NRPS subfamilies (Figure 3.2, Appendix 3.6). The phylogenetically unlinked locations of ChNPS1 and ChNPS3 modules in the larger phylogeny suggests that a recombination event must have given rise to the extant genes in *C. heterostrophus* (Figure 3.9). A domains of several other euascomycete NRPSs, for example, bimodular *Fusarium equiseti* Enniatin synthetase (FeESYN1) and trimodular *M. oryzae*, MGG00022, also show recombinant structures.

Module 1 A domains of both proteins group in the EAS clade with the *C. heterostrophus* pseudogene ChNPS13, but without bootstrap support (Figure 3.2), at positions distinct from modules 1 and 3 of ChNPS1 and ChNPS3. The C-terminal A domain of ESYN1 and the A domains of the final two modules of MGG00022 group in the CYCLO clade (Figures 3.2, 3.9), like module 2 of ChNPS1 and modules 2 and 4 of ChNPS3. Thus, homologs of modules of ChNPS1 and ChNPS3 appear in different combinations in other fungi and demonstrate that recombination plays an important role in the evolution of multimodular NRPSs.

Figure 3.9. Phylogenetic groupings and modular organization of ChNPS1 and ChNPS3 showing recombinant structure of these NRPSs. **A.** Modules 1 and 3 of both ChNPS1 and ChNPS3 group with AM toxin synthetase, a trimodular NRPS that biosynthesizes AM-toxin, an *Alternaria alternata* host-selective toxin. **B.** Module 2 of ChNPS1 and modules 2 and 4 of ChNPS3 group with A domains of cyclosporin synthetases (CYCLO) in a disparate position in the larger phylogeny compared to modules 1 and 3 which group in the EAS subfamily (Figure 3.2). **C.** Recombinant domain organization of ChNPS1 and ChNPS3. Blue boxes correspond to modules 2 and 4, purple boxes to modules 1 and 3. Note that single modules homologous to these domains are found in other euascomycete NRPSs. For example, Enniatin synthetases (Esyn1) and MGG00022.6 are also recombinant like ChNPS1 and ChNPS3 with one or more modules grouping with the cyclosporin subfamily (blue boxes) and others also within the EAS subfamily but in a distinct position from the ChNPS1 and ChNPS3 modules (clear boxes). Cyclosporin synthetases itself appears to have arisen by tandem duplication of modules within *T. inflatum*.



3.3.10 Stability of NRPS Gene Copy Number and Domain Architectures Across Subfamilies

Many multigene families experience gene duplication and loss and evolve by a birth-death process [93-96]. Variation in gene copy number resulting from gene duplication and loss is thought to be influenced by both functional and dosage requirements as well as random processes such as genomic drift [43, 44, 97, 98]. Recent studies suggest that functionally conserved genes, such as those involved in growth and development or other basic cellular processes, tend to experience both less variation in copy number [53] and more stable domain organizations [49] than genes involved in environmental and stress responses [53, 99].

For multimodular genes such as NRPSs, duplication and loss or birth-death evolution [93-95] can occur at two hierarchical levels: 1) at the level of the whole gene, and 2) at the level of domains within a gene (intragenic). In the latter case, genes encoding NRPSs whose products are involved in more conserved functions, such as the AARs, would be expected to have more stable domain architectures than those encoding proteins with niche-specific functions. The latter may experience less functional constraint allowing for flexible gain and loss of domains leading to diversity of domain structures. Because NRPS A domains are involved in substrate selection [100, 101], their loss or gain could result in a rapid change in the chemical product of an NRPS.

The range of variation in copy number of NRPS-encoding genes and in number of A domains/NRPS for each subfamily is shown for Euascomycete taxa only in Figure 3.10. Variation in gene copy number is the highest for the EAS subfamily but both the PKS;NRPS and ChNPS12 subfamilies also show substantial variation (Figure 3.10A).

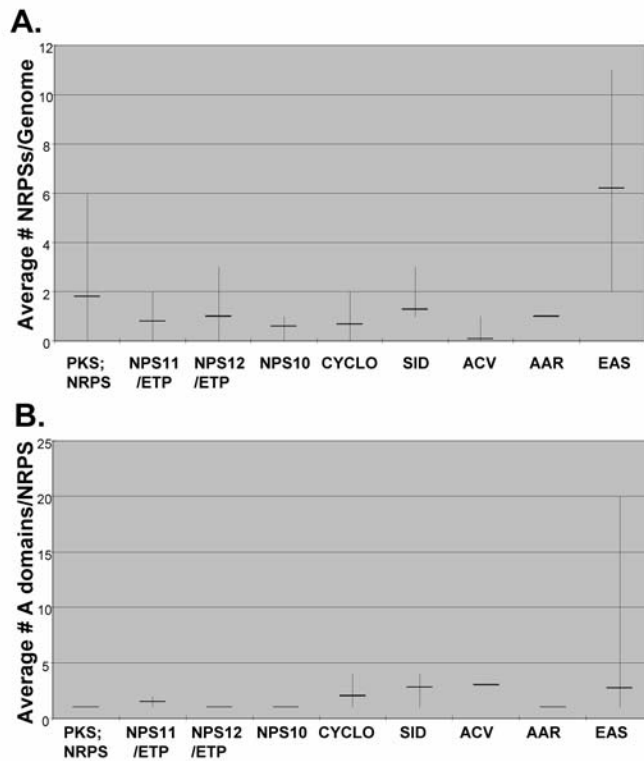


Figure 3.10. Number and range of NRPSs and A domains for each subfamily. **A.** Average and range (lowest to highest) number of NRPS-encoding genes in each subfamily per euascomycete genome shows that the EAS subfamily has both the highest average number of genes and the highest variation in copy number among species. PKS;NRPSs and ChNPS12 subfamilies also have substantial variation in numbers of NRPS-encoding genes among species. **B.** Average and range (lowest to highest) of the number of A domains/NRPS in euascomycete genomes for each subfamily shows that the EAS subfamily also has by far the greatest variation in number of A domains/NRPS followed by the CYCLO, and SID subfamilies.

The EAS subfamily also shows by far the greatest variation in number of A domains/NRPS, followed by CYCLO and SID subfamilies, suggestive of less stable domain architectures and higher rates of intragenic domain duplication for these three groups. All of the remaining mono/bi-modular subfamilies show remarkably conserved domain architectures (Figure 3.5, 3.10B), supporting available functional data which suggests these groups may have more central conserved roles in metabolism.

When we compared gene and domain duplication and loss in different

subfamilies across eukaryotes, no particular subfamily showed significant evidence for nonrandom expansion or contraction of number of genes. When patterns of the total number of A domains per subfamily were analyzed, the EAS subfamily was the only group which showed highly significant ($P < .00001$) deviation from a random birth-death process (data not shown). These results support other observations that gain and loss of domains is an important evolutionary force within the EAS subfamily and may represent an adaptive response to niche-specific environmental pressures.

3.3.11 Chain Termination Mechanisms

Our survey revealed that fungal NRPSs have a variety of C-terminal domains involved in chain termination. The most common for multimodular NRPSs is a C domain while for monomodular NRPSs it is an R domain (Appendix 3.2). R domains have previously been identified and shown to play a role in peptide release in fungal AARs [23, 56], a number of fungal PKS;NRPSs [76-78], and in a minority of bacterial NRPSs including SafA and MxcG [20, 21] and the PKS;NRPS hybrid, myxalamid [22]. Some multimodular NRPSs, however, also have a terminal R domain suggesting this may be a common release mechanism for fungal NRPSs. (Appendix 3.2). Two different release mechanisms have been identified for R type domains in fungal NRPSs, indicating the possibility of R domains subtypes. In fungal AAR's, the R domain reduces the enzyme bound α - amino adipic acid [23]. The C-terminal R domain in the fungal PKS;NRPSs for Equisetin biosynthesis (EqiS), however, catalyzes a Dieckmann condensation reaction, thus performing a function similar to bacterial TE domains [76]. Some mono- and some multi-modular NRPSs terminate in T domains (Appendix 3.2) although these have not been implicated previously in

chain release.

As noted previously, bacterial NRPSs generally have a TE domain at the C-terminal end for peptide release but TE domains have been found only in a few fungal NRPSs, notably the ACV synthetases [16]. We identified several other fungal NRPSs (AN2621.4, FGSG_11989.3, and Phchr1_2706), grouping with modules of Cyclosporin synthetase, which also contain a C-terminal TE domain (Appendix 3.2). However, our data suggest that TE domains are indeed rare in fungal NRPSs providing further support for the claims of horizontal transfer from bacteria to fungi of genes encoding ACV synthetases, and possibly these other fungal genes with TE domains.

3.4 Conclusions

Phylogenomic analysis identified nine major subfamilies of fungal NRPSs which fall into two main groups: 1) a group of primarily mono/bi-modular proteins (ChNPS10, AAR, ChNPS12, ChNPS11/ETP, PKS:NRPS, and CYCLO subfamilies) that group with bacterial NRPSs, and 2) a group of primarily multimodular proteins (EAS, SID) which appear both restricted to and highly expanded within fungi. Analyses demonstrate that α -aminoadipate reductases are more closely related to NRPSs than to other adenylation enzymes and provide further support for previous claims of horizontal transfer of certain NRPSs from bacteria to fungi. In addition, phylogenomic relationships among subfamilies, taxonomic distributions, structural conservation of domain architecture, and known functional data suggest that several of the mono/bi-modular groups are both older in origin and play more central roles in cellular metabolism. The highly expanded group of fungal multimodular genes, particularly the EAS subfamily, have less conserved domain architectures due to

domain/module duplication and loss, and tend to perform more niche-specific functions, typically considered the realm of “secondary” metabolites.

3.5 Materials and Methods

3.5.1 Identification of Putative NRPSs in Fungal Genomes

A set of fungal NRPSs with known chemical products was extracted from the NCBI database (Appendix 3.10), aligned using MUSCLE [102] with the 13 NRPSs identified previously in the Dothideomycete, *C. heterostrophus* C4 strain [10], and used to construct an initial HMMER model of fungal NRPS A domains using HMMER 2.0 (<http://HMMER.wustl.edu/>) (Appendix 3.11). This model was tested for specificity and ability to identify NRPSs proteins in fungal genomes for which NRPSs have been well characterized (e.g., *C. heterostrophus* and *Gibberella zeae/Fusarium graminearum*) and was found to correctly identify all known NRPSs in the genomes of these species as top hits. Protein datasets of a taxonomically representative sample of fungal genomes (Appendix 3.12) were downloaded and searched using both a local and global version of the fungal NRPS HMMER model. Proteins that were hit by our A domain model with an e-value less than 1 were considered possible NRPSs. A similar search strategy was employed on the nucleotide genome sequences using GENEWISE [103] and the same HMMER model to identify candidates that might have been missed or misannotated by automated gene calling programs. This approach did not identify any additional genes but did identify missed domains and also revealed a number of split gene annotations in the automated protein calls which we have reannotated. These included BC1G09040_09041.1, BC1G07441_07442.1, and FGSG11659.3 and FGSG11630.3 which we conclude

represents a single gene corresponding to the MIPS and version 2 broad annotation (FG_00042.1), (Appendix 3.2).

For each fungal genome, A domains from all candidate NRPSs were aligned, using MUSCLE [102], with A domains from the 12 NRPSs previously identified from *C. heterostrophus* [10] (Appendix 3.1) and with A domains from related adenyating enzymes in the AMP-binding family (PFAM PF00501) [e.g., acyl CoA ligases (ACoAL), acetyl CoA synthetases (ACoAS), acyl AMP ligases (AAL), homologs of *C. heterostrophus* CPS1 (CPS1) [54], long chain fatty acid ligases (LCFAL), and homologs of Ochratoxin synthetase (OCHRA) [104] (Appendix 3.5). An initial phylogenetic analysis was conducted using the WAG+G model in PhyML to define a set of candidate NRPS proteins for each genome. Proteins from each genome grouping within a monophyletic group containing A domains of the known *C. heterostrophus* NRPS proteins and separated from the outgroup proteins with consistently high bootstrap support (>90), were retained in the dataset as candidate NRPSs or NRPS-like proteins. We chose to use individual A domains, rather than to include only proteins containing a complete A-T-C module as has been used in previous studies [105] because the latter would miss several putative NRPS or NRPS-like proteins (e.g. *C. heterostrophus* NPS10 and NPS12 [10]) that lack a complete A-T-C module. In addition, freestanding A domains in bacterial NRPSs have been shown to catalyze NRPS biosynthesis by activating and transferring substrates *in trans* to separate NRPSs [5] and the evolutionary relationship between monomodular NRPS-like proteins and multimodular NRPSs was also of interest.

3.5.2 Annotation of Domain Architectures

All candidate proteins were annotated with our initial fungal NRPS A model

and the PFAM models for C (PF00668) and T (PF00550) domains. Using the domains identified in the dataset from this search, a refined set of fungal specific NRPS HMMER models was built for the A (FungalNPSAMP.hmm), C (FungalNPSCON.hmm), and T (FungalNPSTHIOL.hmm) domains (Appendix 3.4). These models more accurately identified C and T domains in NRPSs with known/manually curated annotations than the generic PFAM models and were thus used to annotate A-T-C domain structures of all candidate fungal NRPSs. In addition, all candidate proteins were used as queries against the PFAM and INTERPRO domain databases to identify additional non-canonical NRPS domains present in these proteins. A complete domain architecture was compiled for each protein by merging these two approaches (Appendix 3.2).

3.5.3 Phylogenomic Analyses

Representatives of both fungal and bacterial adenyating enzymes used as outgroups (Appendix 3.5) in identification of putative NRPSs were also used as outgroups in phylogenomic analyses. While all AARs grouped as putative NRPSs, to reduce the size of the dataset, only a taxonomically representative sample of the fungal AARs were included in the full phylogenetic analyses. Fungal A domains from NRPSs with known function and/or chemical products present in GenBank were also included (Appendix 3.10). To select a diverse group of bacterial proteins, a representative A domain of each subfamily of fungal NRPSs was used to query the nr protein database at NCBI and the top 5 bacterial protein hits for each, as well as a number of bacterial proteins with known chemical products, were selected (Appendix 3.8). The complete set of A domains were extracted from these 58 bacterial proteins for a total of 99 A domains.

All candidate NRPS and outgroup A domains were aligned with MUSCLE [102]. Portions of ambiguous alignment were first adjusted manually and then masked to remove columns in the alignment with > 30% gaps prior to phylogenetic analysis (Appendix 3.13). A few candidate A domains were partial (BC1G15479, FG11319, AN8504, and Pa3740) and were removed from the final analysis because they did not align well with other NRPSs. ProtTest [106] was used to identify an appropriate protein substitution matrix as it has been shown that spurious choice of a matrix can lead to inaccurate phylogenies [107]. The RtREV+G+F model had the best likelihood score for all criteria (AIC and BIC) except for AIC-1 with sample size corrected for the number of sites in the alignment, which identified WAG+G as the best model. Three methods were used for phylogeny construction: 1) Maximum likelihood (ML) using RaxML [108] with the RtREV+G+F substitution model, 2) ML using PhyML with the WAG+G model [109], and 3) Neighbor joining (NJ) using NEIGHBOR in PHYLIP [110] and a distance matrix created in TREEPUZZLE [111] with the WAG+G substitution model. We used a Gamma distribution with four rate categories to model rate variation in all analyses. Bootstrapping was performed to assess the robustness of the phylogeny. Bootstrap datasets of 500 replicates for ML analysis and 200 replicates for the NJ analyses were created using SEQBOOT in PHYLIP and analyzed by the respective methods.

Because bootstrap support has been observed to decline in larger datasets [112-114], we also performed analyses on a subset of the data containing representatives from each of the major subfamilies identified. This dataset was aligned separately with MUSCLE and also masked with slightly less stringent conditions to remove columns containing greater than 50% gaps (Appendix 3.14). Phylogenetic analyses were performed on this dataset using the same methods described above.

3.5.4 Subfamily Identification and Modeling

Fungal NRPS subfamilies were characterized as monophyletic groups defined by the most internal branch from the root above a bootstrap cutoff level (we chose 70%) [115, 116] that also shared identical taxon composition across all three phylogenetic methods and had fungal NRPS representation (Appendix 3.6). The SID group was a single exception in that in the full phylogenies (Figure 3.1, Appendix 3.6) maximum likelihood methods supported this clade with 68% and 74% bootstrap support while NJ did not provide support above 50% (Figure 3.1, Appendix 3.6). This clade is, however, supported by >80% bootstrap support in all phylogenetic methods in analysis of the reduced dataset (Figure 3.2, Appendix 3.7).

3.5.5 Distribution of NRPS Subfamilies Across Fungal Taxonomic Groups

To address patterns of distribution of NRPSs across fungal taxonomic groups, we tallied NRPS counts in Chytridiomycota, Zygomycota, Basidiomycota, Schizosaccharomycota, Hemiascomycota, and Euascomycota. Fisher's exact tests were used to test for associations between taxonomic groups and the proportion of genes in each NRPS subfamily.

3.5.6 Lineage Specific Expansions and Variation in Birth-Death Rates

We calculated and graphed the average and range of the number of genes encoding NRPSs in each subfamily per euascomycete genome and the number of A domains per NRPS for each subfamily to assess broad patterns of variation in numbers of genes and numbers of A domains/gene across subfamilies (Figure 3.10)

We used the method of Hahn et al. [117, 118], which applies a stochastic birth and death process along a phylogeny to test for statistically significant lineage specific expansions and contractions of 1) number of NRPS genes and 2) numbers of NRPS A domains/subfamily. For these analyses, we created an ultrametric species tree with the PL method in r8s [119] using the phylogeny of the concatenated protein dataset of Fitzpatrick et al. [120] (Appendix 3.9).

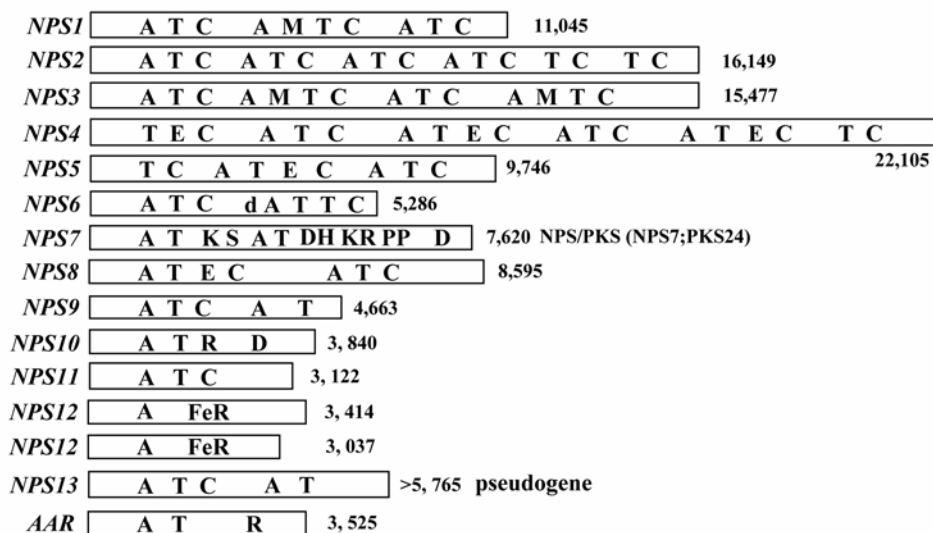
We performed two separate analyses using CAFÉ to look at patterns of gene and A domain expansions. The first analysis looked at patterns of the total number of NRPSs (e.g. all subfamilies combined) to look for broad patterns of expansions and contractions across the full tree of fungi (excluding *B. dendrobatidis*). The second analysis analyzed duplications and losses in each subfamily separately and was restricted to the euascomycete taxa because the birth-death model assumes that at least one gene of each subfamily is present in the common ancestor of all taxa. The ACV synthetase subfamily was excluded because parsimony inferred that this family had zero genes at the root. For all analyses, we used 1000 re-samplings and significant deviations from a random birth-death model were determined by viterbi p-values below .05.

Acknowledgements

BGT acknowledges, with gratitude, the US Department of Energy Joint Genome Institute (JGI) for their fungal genome program, in particular, for their support in generating the sequence of race O, strain C5, of *Cochliobolus heterostrophus* (http://genome.jgi-psf.org/CocheC5_1/CocheC5_1.home.html). KEB would like to thank J. Doyle, S. Kroken, and A. Siepel for discussion of phylogenetic analyses, J. Stajich and M. Hahn for discussions of ultrametric tree construction and birth-death analyses, and the Cornell Computational Biology Service Unit facility and staff, A.

Siepel, K. Nixon, and the CIPRES portal (http://www.phylo.org/sub_sections/portal/) for computer resources for phylogenetic and other computational analyses. BGT acknowledges the support of the Division of Molecular and Cellular Biosciences, National Science Foundation, the USDA Cooperative State Research Education and Extension Service, National Research Initiative and the BARD foundation.

APPENDIX 3.1



Appendix 3.1. Diagram of *Cochliobolus heterostrophus* NRPSs and their domain structure. Included are 12 NRPSs, one NRPS;PKS hybrid (NPS7/PKS24), one AAR, and one pseudogene (NPS13). Annotation of domains shows that, with the exception of the duplicated copy of ChNPS12, each has an unique domain architecture. Domain abbreviations: Adenylation (A), Thiolation (T), Condensation (C), Epimerization (E), Methylation (M), Dehydrogenase (D), Thioester reductase (R), Beta-ketosynthase (KS), Acyl transferase (AT), Dehydratase (DH), Ketoreductase (KR), and Ferric transmembrane reductase (FeR). Length of each gene in (bp) is shown to the right.

Appendix 3.2. Accession numbers, genomic locations, and domain architectures of NRPSs identified in fungal genomes

Species	Sequencing Center/Version #	Sequence Center ID	Subfamily/Group	# AMP	Known Genes	Chromosomal Location	Domain Annotation ^A
<i>Ashbya gossypii</i>		ADL346W	AAR	1		ChmrIV: 96109:100266	A-T-R
<i>Aspergillus fumigatus</i>	CADRE/TIGR annotation	Afu1g10380	EAS	4	<i>Pes1</i>	Chromosome 1: 2675699-2694887 -	A-T-E-C-A-T-C-A-T-C-A- A-(DNALigA1) ^B -T-E-C- T-C-T
		Afu1g17200	SID	3	<i>SidC</i>	Chromosome 1: 4688800-4703141 +	A-T-C-A-T-C-A-T-C-T- C-T-C
		Afu3g15270	SIDE	2	<i>SidE</i>	Chromosome 3: 4010522-4017637 +	A-T-C-A-T-C
		Afu3g03350	SIDE	2	<i>SidE</i>	Chromosome 3: 891335-898767 +	A-T-C-A-T-C
		Afu3g03420	EAS	1	<i>NPS6</i>	Chromosome 3: 908168-914474 +	A-T-C-dA-T-C
		Afu3g12920	ETP	2		Chromosome 3: 3429981-3437235 +	A-T-C-A-T-C-T
		Afu3g13730	EAS	1		Chromosome 3: 3619321-3623193 +	A-T-C
		Afu4g14440	EAS	1		Chromosome 4: 3815691-3817758 -	A
		Afu4g11240	AAR	1		Chromosome 4: 2934763-2939218 +	A-T-R
		Afu5g12730	EAS	6		Chromosome 5: 3314537-3340084 +	A-T-C-A-T-E-C-A-T-C- T-E-C-A-T-C-A-T-C-A- T-E-C
		Afu5g10120	NPS10	1	<i>NPS10</i>	Chromosome 5: 2603086-2606910 +	A-T-R-D
		Afu6g09660	ETP	2	<i>GliP</i>	Chromosome 6: 2352620-2359124 -	A-T-C-A-T-C-T
		Afu6g09610	EAS	1		Chromosome 6: 2339538-2343356 -	A-T-C
		Afu6g12050	EAS	1		Chromosome 6: 3013593-3017507 +	A-T-C
		Afu6g12080	EAS	3		Chromosome 6: 3023316-3035305 -	A-T-C-A-T-E-C-A-T-C
		Afu6g03480	EAS	1		Chromosome 6: 748395-753511 +	A-M-T-R-(PLP) ^B
Afu8g00170	EAS	2		Chromosome 8: 20854-27489 -	A-T-C-A-T-C		

Appendix 3.2 Continued

		Afu8g01220	NPS12	1		Chromosome 8: 286174-287750 +	A
		Afu8g00540	PKS;NPS	1		Chromosome 8: 117018-129323 +	KS-AT-M-KR-AC-C-A- T-R
		Afu8g01640	CYCLO	1		Chromosome 8: 430403-433447 -	A-T-R
<i>Aspergillus nidulans</i>	BROAD/Version 4	AN7884.4	EAS	6		Contig 43: 76779- 97630 -	A-T-C-A-T-E-C-A-T-C- A-T-C-A-T-C-A-T-R
		AN2545.4	EAS	5		Contig 17: 152220- 171385 +	T-E-C-A-T-E-C-A-T-C- A-T-C-A-T-C-A-T-C
		AN1242.4	EAS	4		Contig 2: 53533- 71726 + -	A-T-E-C-dA-C-A-T-C- A-(DNALigA1) ^B -T-E-C- T-C-T
		AN0016.4	EAS	4		Contig 45: 15391- 26703 -	A-T-E-C-dA-A-T-C-A- T-E-C-T-C-T
		AN2621.4	ACV	3	<i>ACV</i>	Contig 7: 721978- 736405 -	A-T-C-A-T-C-A-T-C-TE
		AN0607.4	SID	3	<i>SIDC</i>	Contig 59: 79653- 86681 +	A-T-C-A-T-C-A-T-C-T- C-T-C
		AN3496.4	EAS	2		Contig 172: 18331- 25266 -	T-C-A-T-C-A-T
		AN9244.4	EAS	2		Contig 107: 689-7055 +	A-T-C-T-C-(ESP) ^B -A
		AN6236.4	EAS	1	<i>NPS6</i>	Contig 59: 74210- 78819 -	A-T-C-dA-T-C
		AN3495.4	EAS	1		Contig 79: 149178- 154353 +	C-A-T-R
		AN10576.4	EAS	1		Contig 172: 14038- 17366 -	A-T-E-C
		AN9243.4	EAS	1		Contig 153: 165693- 167353 +	C-A
		AN8433.4	EAS	1		Contig 93: 164592- 168404 +	A
		AN5318.4	NPS10	1	<i>NPS10</i>	Contig 153: 92194- 104129 +	A-T-R-D
		AN8412.4	PKS;NPS	1		Contig 139: 270380- 273613 -	KS-AT-M-KR-AC-C-A- T-R
		AN8105.4	CYCLO	1		Contig 153: 399039- 401872 -	A-T-R
		AN8504.4 ^C	Incomplete			Contig 170: 179001- 186848 -	dA-T-C

Appendix 3.2 Continued

		AN9226.4	CYCLO/EAS	2		Contig 169: 212269-215635 +	A-T-C-A-M-T-C
		AN9129.4	NPS12	1		Contig 98: 7078-11515 +	A-FeR
		AN5610.4	AAR	1		Contig 98: 7078-11515 +	A-T-R
<i>Batrachochytrium dendrobatidis</i>	BROAD/Version 1	BDEG01579.1	AAR	1		Supercontig 1: 4224701-4229355 +	A-T-R
		BDEG03514.1	NPS12	1		Supercontig 4: 855028-860002 +	T-C-A-T-(FSH1) ^B
		BDEG08447.1	NPS12	1		Supercontig 16: 295121-301255 -	A-T-(RnaH) ^C -(LPS) ^B
<i>Botrytis cinerea</i>	BROAD/Version 1	BC1G10622.1	EAS	1		Supercontig 73: 102617-106612 -	A-T-C-R
		BC1G02495.1	EAS	2		Supercontig 8: 178697-187012 +	C-A-T-C-A-T-R
		BC1G10567.1	EAS	1	<i>NPS6</i>	Supercontig 72: 181976-187309 +	A-T-C-T-C
		BC1G03511.1	SID	3		Supercontig 13: 230111-245780 +	A-T-C-A-T-C-T-C-A-T-C-T-C-T-C
		BC1G10928.1	SID	3		Supercontig 75: 123029-134885 -	A-T-C-A-T-C-A-T-C-T-C
		BC1G04782.1	EAS	1		Supercontig 20: 111997-117883 -	M-T-C-A-(DNALigA1) ^B -T-R
		BC1G00695.1	PKS;NPS	1		Supercontig 2: 410037-422151 -	KS-AT-M-KR-AC-C-A-T-R
		BC1G15494.1	SID	1		Supercontig 180: 49908-52691 +	A-T-C
		BC1G09040_09041.1	EAS	3		Supercontig 52: 176500-186100 -	T-C-A-T-C-A-T-C-A-T-C-T
		BC1G15479.1 ^C	PKS;NPS	1		Supercontig 180: 3511-14382 -	KS-AT-M-KR-AC-C-A-T-R
		BC1G15703.1	PKS;NPS	1		Supercontig 196: 8963-16869 +	KS-AT-M-KR-AC-C-A-T-R
		BC1G07441_7442.1	ETP	1		Supercontig 42: 127734-130214 -	dA-T-C-T-C-A-T-C
		BC1G11613.1	OTHER	1		Supercontig 91: 14,000-16,000	A-T-(Hx) ^B

Appendix 3.2 Continued

		BC1G13197.1	AAR	1		Supercontig 116: 67595-71022 +	A-T-R
<i>Candida albicans</i>	BROAD/Version 1	CAWG_01102.1	AAR	1		Supercontig 1: 2601397-2605611 -	A-T-R
<i>Candida glabrata</i>	Genolevures/Version 1	CAGL0K07788g	AAR	1		Cagl0K:774352- 778476 -	A-T-R
<i>Candida guilliermondii</i>	BROAD/Version 1	PGUG_04759.1	AAR	1		Supercontig 6: 261996-266216 -	A-T-R
<i>Candida lusitanae</i>	BROAD/Version	CLUG_04446.1	AAR	1		Supercontig 5: 712467-716633 +	A-T-R
<i>Candida tropicalis</i>		CTR_04682.1	AAR	1		Supercontig 6: 899575-902463 +	A-T-R
<i>Coccidioides immitis</i>	BROAD/Version 3	CIMG09750.3	EAS	5		C. immitis RS: Chromosome 5: 2299495-2324157 -	A-T-E-C-A-T-C-A-T-T- C-C-A-C-A-T-C-T-C-T
		CIMG01429.3	EAS	1		C. immitis RS: Chromosome 1: 3743390-3749263 -	A-T-C-T-C
		CIMG03170.3	EAS	1		C. immitis RS: Chromosome 2: 630912-633926 -	C-A-T
		CIMG01861.3	EAS	2		C. immitis RS: Chromosome 1: 4899391-4906863 -	A-T-C-A-T-C
		CIMG07298.3	EAS	1		C. immitis RS: Chromosome 3: 4369404-4375473 -	A-T-C-T-C
		CIMG00941.3	SID	3		C. immitis RS: Chromosome 1: 2456614-2472250 -	A-T-C-A-T-C-T-C-A-T- C-T-C-T-C
		CIMG06629.3	PKS;NPS	1		C. immitis RS: Chromosome 3: 2485975-2498131 -	KS-AT-M-KR-AC-C-A- T-R
		CIMG01491.3	AAR	1		C. immitis RS: Chromosome 1: 3907134-3911528 -	A-T-R
<i>Cochliobolus heterostrophus</i>	JGI/Version 1	CocheC5_1_29312	NPS10	1	<i>NPS10</i>	CocheC5_1/scaffold_ 6:1384455-1390709	A-T-R-D

Appendix 3.2 Continued

		CocheC5_1_115564	ChNPS11/ETPm1	1	<i>NPS11</i>	CocheC5_1/scaffold_1:1126440-1130558	A-T-C
		CocheC5_1_118012	NPS12	1	<i>NPS12</i>	CocheC5_1/scaffold_11:763264-766881	A-FeR
		CocheC5_1_116719	NPS12	1	<i>NPS12</i>	CocheC5_1/scaffold_5:262551-266118	A-FeR
		CocheC5_1_15959	2CYCLO/2EAS	4	<i>NPS3</i>	CocheC5_1/scaffold_1:554870-569203	A-T-C-A-M-T-C-A-T-C-A-M-T-C
		CocheC5_1_84777	1CYCLO/2EAS	3	<i>NPS1</i>	CocheC5_1/scaffold_6:788062-801107	A-T-C-A-M-T-C-A-T-C
		CocheC5_1_115936	AAR	1	<i>AAR1</i>	CocheC5_1/scaffold_2:839177-843985	A-T-R
		CocheC5_1_77609	SID	4	<i>NPS2</i>	CocheC5_1/scaffold_33:136682-152804	A-T-C-A-T-C-A-T-C-A-T-C-T-C-T-C
		CocheC5_1_16574	EAS	1	<i>NPS13</i>	CocheC5_1/scaffold_1:2669103-2670314	A-T
		CocheC5_1_3317	EAS	1	<i>NPS6</i>	CocheC5_1/scaffold_25:568366-575312	A-T-C-dA-T-T-C
		CocheC5_1_94644	EAS	2	<i>NPS5</i>	CocheC5_1/scaffold_25:25248-35993	T-C-A-T-E-C-A-T-C
		CocheC5_1_94248	EAS	2	<i>NPS8</i>	CocheC5_1/scaffold_23:520703-531586	A-T-E-C-A-T-C
		CocheC5_1_119280	EAS	2	<i>NPS9</i>	CocheC5_1/scaffold_23:2629-8556	A-T-C-A-T
		CocheC5_1_112395	EAS	4	<i>NPS4</i>	CocheC5_1/scaffold_22:508445-531549	T-E-C-A-T-C-A-T-E-C-A-T-C-A-T-E-C-T-C
		CocheC5_1_89648	MBC	1	<i>NPS7</i>	CocheC5_1/scaffold_13:211976-223348	A-T-KS-AT-DH-KR-T-D
<i>Coprinus cinereus</i>	BROAD/Version 2	CC1G_03009.2	NPS12	1		Contig 177: 344918-348700 -	A-FeR
		CC1G_04210.2	SID	1		Contig 105: 74273-82054 -	A-T-C-T-C-T-C
		CC1G_06235.2	NPS12	1		Contig 194: 186935-190731 -	A-FeR
		CC1G_06250.2	NPS12	1		Contig 194: 233847-237602 -	A-FeR
		CC1G_15694.2	AAR	4		Contig 11: 933268-937996 -	A-T-R
<i>Cryptococcus neoformans</i>	BROAD/Version 1	CNAG_03588.1	AAR	1		Chromosome 8: 1345985-1350502 -	A-T-R

Appendix 3.2 Continued

<i>Debaromyces hansenii</i>	Genolevures/Version 1	DEHA2D07964g	AAR	1		Deha2D - 684912-653108	A-T-R
<i>Encephalitozoon cuniculi</i>	NCBI/Unannotated	None					
<i>Fusarium graminearum</i>	BROAD/Version 3	FGSG_11659.3 ^D	EAS	7	<i>NPS8</i>	F. graminearum: Supercontig 1: 144805-158252 -	A-T-C-A-T-C-A-T-C-A-T- T-C-A-T-C-A-T-C-A-T- C
		FGSG_11660.3 ^D			<i>NPS8</i>	F. graminearum: Supercontig 1: 162481-165335 -	
		FGSG_13783.3	EAS	6	<i>NPS18</i>	F. graminearum: Supercontig 7: 2213860-2242925 +	A-T-C-A-T-E-C-A-T-E- C-A-T-E-C-A-T-E-C-A- T-C
		FGSG_02315.3	EAS	5	<i>NPS4</i>	F. graminearum: Supercontig 1: 7439120-7462127 -	A-T-E-C-A-T-C-A-T-E- C-A-T-C-A-T-E-C-T-C
		FGSG_02394.3	EAS	2	<i>NPS15</i>	F. graminearum: Supercontig 1: 7670431-7677408 -	A-T-C-A-T-R
		FGSG_08209.3	EAS	3	<i>NPS7</i>	F. graminearum: Supercontig 5: 2569253-2582878 -	T-C-A-T-C-A-T-C-A-T- C
		FGSG_05372.3	SID	3	<i>NPS2</i>	F. graminearum: Supercontig 3: 2043338-2057970 +	A-T-C-A-T-C-T-C-A-T- C-T-C-T-C
		FGSG_11026.3	SID	3	<i>NPS1</i>	F. graminearum: Supercontig 8: 574013-588387 +	A-T-C-A-T-C-A-T-C-T- C-T-C
		FGSG_11395.3	EAS	2	<i>NPS14</i>	F. graminearum: Supercontig 9: 312651-319899 +	A-T-C-A-T-C
		FGSG_03747.3	EAS	1	<i>NPS6</i>	F. graminearum: Supercontig 2: 2809478-2815749 -	A-T-C-T-C
		FGSG_01680.3	EAS	1	<i>NPS16</i>	F. graminearum: Supercontig 1: 5534719-5539722 -	A-M-T-R-(PLP) ^B
		FGSG_13878.3	EAS	8	<i>NPS5</i>	F. graminearum: Supercontig 8: 693139-727216 +	A-C-A-T-E-C-A-T-E-C- A-T-E-C-A-T-E-C-A-T- E-C-A-T-E-C-A-T-R

Appendix 3.2 Continued

		FGSG_10990.3	EAS	1	<i>NPS9</i>	F. graminearum: Supercontig 8: 686748-689261 -	A-T
		FGSG_10523.3	EAS	1	<i>NPS3</i>	F. graminearum: Supercontig 7: 2137744-2145297 -	T-E-C-A-T-C-T
		FGSG_10702.3	EAS	1	<i>NPS17</i>	F. graminearum: Supercontig 7: 2680651-2682687 -	A
		FGSG_11294.3	NPS12	1	<i>NPS12</i>	F. graminearum: Supercontig 9: 577120-580295 +	A-FeR
		FGSG_06507.3	NPS10	1	<i>NPS10</i>	F. graminearum: Supercontig 4: 265804-269866 -	A-T-R-D
		FGSG_03245.3	NPS12	1	<i>NPS11</i>	F. graminearum: Supercontig 2: 4162913-4166150 -	A-FeR
		FGSG_11989.3	CYCLO	1	<i>NPS19</i>	F. graminearum: Supercontig 1: 5556950-5560682 +	A-M-T-TE
		FGSG_13153.3	NPS12	1	<i>NPS13</i>	F. graminearum: Supercontig 4: 3627883-3631142 -	A-FeR
		FGSG_06041.3	AAR	1		F. graminearum: Supercontig 3: 4113978-4117801 +	A-T-R
		FGSG_07798.3	PKS;NPS	1		F. graminearum: Supercontig 4: 4503929-4515787 -	KS-AT-M-KR-AC-C-A-T-R
		FGSG_11319.3 ^C	Incomplete			F. graminearum: Supercontig 9: 520033-520572 -	A
<i>Kluveromyces lactis</i> var. <i>lactis</i>	Genolevures/Version 1	KLLA0B09218g	AAR	1		KllaOB 805915-810072	A-T-R
<i>Laccaria bicolor</i>	JGI/Version 1	Lacbi_150981	AAR	1		scaffold_9:88330-92826	A-T-R

Appendix 3.2 Continued

<i>Magnaporthe oryzae</i>	BROAD/Version 6	MGG_07858.6	EAS	4		Supercontig 183: 547109-561691	A-T-C-A-T-C-A-T-C-A-T-C
		MGG_02351.6	EAS	5		Supercontig 186: 3232658-3248975 -	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C
		MGG_00022.6	2CYCLO/IEAS	3		Supercontig 194: 4090279-4102482 +	A-T-C-A-M-T-C-A-T-C
		MGG_09589.6	PKS;NPS	1		Supercontig 197: 596671-608484 -	KS-AT-M-KR-AC-C-A-T
		MGG_03290.6	NPS10	1	<i>NPS10</i>	Supercontig 190: 371071-374913 +	A-T-R-D
		MGG_07803.6	ChNPS11/ETPm1	1		Supercontig 183: 323819-328757 +	C-A-T-KS
		MGG_15248.6	ChNPS11/ETPm1	1		Supercontig 183: 125103-133264 +	T-C-A-T-C
		MGG_03401.6	EAS	1		Supercontig 190: 3214-11859 -	T-E-C-A-T-C
		MGG_14943.6	PKS;NPS	1		Supercontig 187: 767479-771399 -	KS-AT-M-KR-AC-C-A-T-R
		MGG_14897.6	PKS;NPS	1	<i>SYN8</i>	Supercontig 187: 2269449-2280593 -	KS-AT-M-KR-AC-C-A-T-R
		MGG_03810.6	PKS;NPS	1		Supercontig 187: 735169-748461 -	KS-AT-M-KR-AC-C-A-T-R
		MGG_12175.6	SID	3	<i>NPS2</i>	Supercontig 187: 2165442-2180544 -	A-T-C-A-T-C-T-C-A-T-C-T-C-T-C
		MGG_14967.6	OTHER	4		Supercontig 187: 2952752-3007190 +	A-T-C-T-C-A-T-C-T-C-A-T-E-C-A-T-C
		MGG_04949.6 ^C	Incomplete			Supercontig 21: 78-3628 -	A
		MGG_12447.6	PKS;NPS	1	<i>SYN2</i>	Supercontig 195: 2333033-2345385 -	KS-AT-M-KR-AC-C-A-T-R
		MGG_15097.6	PKS;NPS	1	<i>ACE1</i>	Supercontig 195: 2390311-2400256 +	KS-AT-M-KR-AC-C-A-T-R
		MGG_11222.6	NPS12	1	<i>NPS12</i>	Supercontig 196: 1938606-1939504 -	A-FeR
		MGG_14767.6	EAS	1	<i>NPS6</i>	Supercontig 196: 2997845-3004145+	A-T-C-dA-T-C
		MGG_02611.6	AAR	1		Supercontig 193: 2144939-2148616 +	A-T-R
		<i>Neurospora crassa</i>	BROAD/Version 3	NCU07119.3	SID	3	
NCU08441.3	EAS			1	<i>NPS6</i>	Contig 44: 204858-211015 +	A-T-C-dA-T-C

Appendix 3.2 Continued

		NCU04531.3	EAS	1		Contig 21: 527341-536094 -	T-T-C-C-A-T-C
		NCU03010.3	AAR	1		Contig 7: 160134-163731 -	A-T-R
<i>Phanaerochaete chrysosporium</i>	JGI/Version 1	Phchr1_2706	CYCLO	2		scaffold_11:866634-867772	A-M-T-C-A-T-TE
		Phchr1_135156	NPS12	1		scaffold_20:228019-231930	A-FeR
		Phchr1_161268	AAR	1		scaffold_2:1745018-1748359	A-T-R
<i>Phycomyces blakesleeanus</i>	JGI/Version 1	Phyb1_34455	AAR	1		Phyb1/scaffold_53:827-13373	A-T-R
<i>Pichia stipitis</i>	JGI/Version 2	Picst3_68020	AAR	1		Picst3/chr_6.1:353493-357715	A-T-R
<i>Podospora anserine</i>	Genoscope/Version 1	Pa0_240	PKS;NPS	1		SC_C_chrm6_seq:84078..97458	KS-AT-M-KR-AC-C-A-T-R
		Pa1_5210	PKS;NPS	1		SC_D_chrm1_seq:335192..336221	KS-AT-M-KR-AC-C-A-T-R
		Pa2_7870	SIDE	2		SC_B_chrm2_seq:427923..4279120	C-A-T-C-A-T-C
		Pa3_11200	EAS	1	<i>NPS6</i>	SC_C_chrm3_seq:250296..255914	A-T-C-T-C
		Pa4_4440	SID	3		SC_D_chrm4_seq:93610..108867	A-T-C-A-T-C-T-C-A-T-C-T-C-T-C
		Pa4_4630	EAS	4		SC_D_chrm4_seq:167899..184030	A-T-E-C-A-T-C-A-T-E-C-A-(DNALigA1) ^B -T
		Pa4_4640	EAS	1		SC_D_chrm4_seq:184183..192897	A-T-E-C-A-T-C-A-T-E-C-A-T-C-A-T-E-C-T-C-T
		Pa5_1070	EAS	1		SC_A_chrm5_seq:416202..424777	T-T-(PI4S) ^B -C-C-A-T-C
		Pa5_6830	PKS;NPS	1		SC_E_chrm5_seq:175989..188303	KS-AT-M-KR-AC-C-A-T-R
		Pa5_3740 ^C	Incomplete	1		SC_B_chrm5_seq:206605..208782	C-dA
		Pa6_10100	PKS;NPS	1		SC_D_chrm6_seq:622720..628038-	C-A-T-T-R

Appendix 3.2 Continued

		Pa0_670	NPS12	1	SC_A_chrm6.seq: 27621..30710	A-T-C
		Pa1_5110	AAR	1	SC_B_chrm1.seq:14 78700..1482367	A-T-R
<i>Postia placenta</i>	JGI/Version 1	Posp11_111174	NPS12	1	Posp11/scaffold_133: 154647-156802	A
		Posp11_95457	NPS12	1	Posp11/scaffold_133: 61362-63334	A
		Posp11_42387	NPS12	1	Posp11/scaffold_140: 198705-199322	A
		Posp11_127321	NPS12	1	Posp11/scaffold_133: 159978-162012	A
		Posp11_49678	NPS12	1	Posp11/scaffold_133: 50945-52834	A
		Posp11_54576	NPS12	1	Posp11/scaffold_34:3 68246-370135	A
		Posp11_42034	NPS12	1	Posp11/scaffold_43:2 96234-296941	A
		Posp11_54642	NPS12	1	Posp11/scaffold_34:3 61746-362573	A
		Posp11_109769	AAR	1	Posp11/scaffold_11:1 61253-165967	A-T-R
		Posp11_115736	AAR	1	Posp11/scaffold_12:3 40539-344819	A-T-R
<i>Puccinia graminis</i>	BROAD/Version 2	PGTG_06519.2	OTHER	1	Supercontig 15: 236832-241982 -	A-T-C
		PGTG_07683.2	AAR	1	Supercontig 20: 1032995-1037380 +	A-T-R
<i>Rhizopus oryzae</i>	BROAD/Version 1	RO3G_12433.1	AAR	1	Supercontig 10: 1228709-1233004 +	A-T-R
<i>Saccharomyces cerevisiae</i>	SGD,BROAD/Version 1	YBR115C/SCRG_02851.1	AAR	1	Chr2: 473920- 469742 -	A-T-R
<i>Saccharomyces bayanus</i>	BROAD/Unannotated BROAD/Unannotated	BROAD/Unannotated	AAR	1	contig_7_37512- 38720	A-T-R
<i>Saccharomyces mikatae</i>			AAR	1	contig_91_4043- 5791	A-T-R

Appendix 3.2 Continued

<i>Saccharomyces paradoxus</i>			AAR	1		contig_203_6588-4838	A-T-R
<i>Schizosaccharomyces japonicus</i>	BROAD/Version 2	SJAG_04031.2	SID	3	<i>sib1</i>	S. japonicus yFS275: Supercontig 5: 577316-591914 -	A-T-C-T-C-T-C-A-T-C-T-C-T-C
		SJAG_00869.2	AAR	1		Supercontig 1: 1769595-1773839 +	A-T-R
<i>Schizosaccharomyces pombe</i>	BROAD/Version 2	SPAP7G5.04c	AAR	1	<i>LYS2</i>	Chromosome 1: 3739162-3743421 -	A-T-R
		SPAC23G3.02c	SID	3	<i>sib1</i>	Chromosome 1: 854523-869527 -	A-T-C-T-C-T-A-T-T-C-T-C
<i>Sporobolomyces roseus</i>	JGI/Version 1	Sporo1_21452	AAR	1		Sporo1/scaffold_3:1748229-1753072	A-T-R
		Sporo1_31423	Other	1		Sporo1/scaffold_1:2360477-2373012	A-T-C-T
<i>Trichoderma reesii</i>	JGI/Version 2	Trire2_123786	EAS	14		Trire2/scaffold_26:276834-327620	KS-AT-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-R
		Trire2_23171	EAS	20	<i>TEX1</i> homolog	Trire2/scaffold_24:123560-193077	
		Trire2_58285	PKS;NPS	1		Trire2/scaffold_5:25618-37773	KS-AT-M-KR-AC-C-A-T-R
		Trire2_59315	PKS;NPS	1		Trire2/scaffold_6:34746-46569	KS-AT-M-KR-AC-C-A-T-R
		Trire2_60751	EAS	1		Trire2/scaffold_8:524121-526840	A-T-C
		Trire2_67189	EAS	1	<i>NPS6</i>	Trire2/scaffold_20:536612-542053	A-T-C-T-C
		Trire2_68204	NPS12	1		Trire2/scaffold_24:269353-272919	A-FeR
		Trire2_71005	EAS	1	<i>NPS6</i>	Trire2/scaffold_1:3561799-3563725	A-T-C-T-C
		Trire2_81014	NPS10	1	<i>NPS10</i>	Trire2/scaffold_22:47424-51332	A-T-R-D
		Trire2_24586	ChNPS11/ETPm1	2		Trire2/scaffold_1:2715208-2721935	A-T-C-A-T-C-T
		Trire2_60458	ChNPS11/ETPm1	2		Trire2/scaffold_7:1346092-1352757	A-T-C-A-T-C-T

Appendix 3.2 Continued

		Trire2_69946	SID	3		Trire2/scaffold_31:39 879-54649	A-T-C-A-T-C-T-C-A-T- C-T-C-T-C
		Trire2_4117	AAR	1		Trire2/scaffold_10:72 1901-725674	A-T-R
<i>Ustilago maydis</i>	BROAD/Version 1	UM05165.1	SID	3	<i>sid2</i>	Contig 188: 245412- 257254 +	A-T-C-A-T-C-A-T-C-T- C
		UM01434.1	SID	3	<i>fer3</i>	Contig 49: 92548- 107141 +	A-T-C-A-T-C-A-T-C-T- C-T-C
		UM05245.1	OTHER	3		Contig 191: 1-10972 -	A-T-C-A-T-C-A-T
		UM03108.1	NPS10	1	<i>NPS10</i>	Contig 105: 12433- 16395 +	A-T-R-D
		UM01697.1	AAR	1		Contig 66: 37971- 42527 +	A-T-R
<i>Yarrowia lipolytica</i>	Genolevures/Version 1	YALI0E06457g	AAR	1		Yali0E: 734132..738 373 +	A-T-R

^A DOMAIN CODES:

	Interpro #	PFAM	
A	IPR000873	PF00501	AMP-dependent synthetase and ligase
T	IPR006162	PF00550	Phosphopantetheine attachment site
C	IPR001242	PF00668	Condensation
E	IPR001509	PF01370	Epimerization
M	IPR013217	PF08242	Methyltransferase type 11 and type 12
R	IPR010080		Thioester reductase
D	IPR002198	PF00106	Short-chain dehydrogenase/reductase
KS	IPR014030	PF00109	Beta-ketosynthase
KR	IPR013968	PF08659	Keto-reductase
AT	IPR014043	PF00698	Acyl transferase
AC	IPR009081		Acyl carrier protein-like
TE	IPR001031	PF00975	Thioesterase
FeR	IPR013130	PF01794	Ferric reductase transmembrane domain

3BHS	IPR002225	PF01073	3-Beta hydroxysteroid dehydrogenase/isomerase
N4	IPR013120	PF07993	NAD_binding_4-male sterility factor
HX	IPR001451	PF00132	Bacterial transferase hexapeptide repeat
FSH1	IPR006660	PF03960	FSH1 - Serine Hydolase
PLP	IPR018319	PF03841	Pyridoxal phosphate-dependent transferase
LPS	IPR006629		LPS-induced tumor necrosis factor alpha factor
RnaH	IPR012337		Polynucleotidyl transferase, Ribonuclease H fold
ESP	IPR001638		Extracellular solute-binding protein, family 3
FabD	IPR016035		FabD/lysophospholipase-like
PI4S	IPR000215		Protease inhibitor I4, serpin
DNAIigA1	IPR016059		DNA Ligase A1

^B Domains in parentheses indicate domains which are noncanonical or unusual NRPS domains with hits greater than e-10

^C These NRPSs were removed from the final phylogenetic analyses as only a partial A domain was identified that did not align well with other sequences.

^D Our annotation of genomic DNA suggested that these genes (FGSG_11659.3 and FGSG_11660.3) should be merged to form a single gene with 7 A-T-C modules that corresponds to the FG00042.1 in the version 1 BROAD annotation of *F. graminearum* and are referred to as FG00042.1 in all trees, figures, and tables.

APPENDIX 3.3

Appendix 3.3. Species Abbreviations

Fungi

<i>Acremonium chrysogenum</i>	Ac
<i>Alternaria alternata</i>	Aa
<i>Alternaria brassicae</i>	Ab
<i>Ashbya gossypii</i>	Ag
<i>Aspergillus fumigatus</i>	Af
<i>Aspergillus nidulans</i>	An
<i>Batrachochytrium dendrobatidis</i>	Bd
<i>Botrytis cinerea</i>	Bc
<i>Candida albicans</i>	Ca
<i>Candida glabrata</i>	Cgl
<i>Candida guilliermondii</i>	Cgu
<i>Candida lusitanae</i>	Cl
<i>Candida tropicalis</i>	Ct
<i>Claviceps purpurea</i>	Cp
<i>Coccidioides immitis</i>	Ci
<i>Cochliobolus carbonum</i>	Cca
<i>Cochliobolus heterostrophus</i>	Ch
<i>Coprinus cinereus</i>	Cc
<i>Cryptococcus neoformans</i>	Cn
<i>Debaryomyces hansenii</i>	Dh
<i>Encephalitozoon cuniculi</i>	Ecu
<i>Epichloë festuca</i>	Ef
<i>Fusarium equiseti</i>	Fe
<i>Fusarium graminearum</i>	Fg
<i>Fusarium heterosporum</i>	Fh
<i>Gibberella fujikuroi</i>	Gf
<i>Hypocrea virens</i>	Hv
<i>Khuyveromyces lactis</i>	Kl
<i>Laccaria bicolor</i>	Lb
<i>Leptosphaeria maculans</i>	Lm
<i>Magnaporthe oryzae</i>	Mg
<i>Metarhizium anisopliae</i>	Ma
<i>Neurospora crassa</i>	Nc
<i>Penicillium chrysogenum</i>	Pc
<i>Phanaerochaete chrysosporium</i>	Pch
<i>Phycomyces blakesleeianus</i>	Pb
<i>Pichia stipitis</i>	Ps
<i>Podospira anserina</i>	Pa
<i>Postia placenta</i>	Pp
<i>Puccinia graminis</i>	Pg
<i>Pyrenophora tritici-repentis</i>	Pt
<i>Rhizopus oryzae</i>	Ro
<i>Saccharomyces bayanus</i>	Sb
<i>Saccharomyces cerevisiae</i>	Sc
<i>Saccharomyces mikatae</i>	Sm
<i>Saccharomyces paradoxus</i>	Spa
<i>Schizosaccharomyces japonicus</i>	Sj
<i>Schizosaccharomyces pombe</i>	Sp
<i>Sporobolomyces roseus</i>	Sr
<i>Trichoderma reesii</i>	Tr

<i>Tolypocladium inflatum</i>	Ti
<i>Ustilago maydis</i>	Um
<i>Yarrowia lipolytica</i>	Yl

Bacteria

<i>Anabaena variabilis</i>	Av
<i>Arthrobacter sp.</i>	Asp.
<i>Bacillus amyloliquefaciens</i>	Ba
<i>Bacillus subtilis</i>	Bs
<i>Brevibacillus brevis</i>	Bb
<i>Brevibacillus parabrevis</i>	Bp
<i>Brevibacillus texasporus</i>	Bt
<i>Burkholderia cenocepacia</i>	Bce
<i>Chlorobium ferrooxidans</i>	Cf
<i>Clostridium cellulolyticum</i>	Cce
<i>Crocospaera watsonii</i>	Cw
<i>Cyanothece sp.</i>	Csp.
<i>Dinoroseobacter shibae</i>	Ds
<i>Escherichia coli</i>	Ec
<i>Geobacter sulfurreducens</i>	Gs
<i>Hahella chejuensis</i>	Hc
<i>Herpetosiphon aurantiacus</i>	Ha
<i>Heliobacterium modesticaldum</i>	Hm
<i>Lynngbya majuscula</i>	Lma
<i>Lysobacter lactamgenus</i>	Ll
<i>Melittangium lichenicola</i>	Ml
<i>Microcystis aeruginosa</i>	Mae
<i>Micromonospora sp.</i>	Msp.
<i>Mycobacterium tuberculosis</i>	Mt
<i>Myxococcus xanthus</i>	Mx
<i>Nocardia lactamdurans</i>	Nl
<i>Nodularia spumigena</i>	Ns
<i>Nostoc punctiforme</i>	Np
<i>Nostoc sp.</i>	Nsp.
<i>Opitutus terrae</i>	Ot
<i>Photorhabdus luminescens</i>	Pl
<i>Pseudomonas aeruginosa</i>	Pae
<i>Pseudomonas entomophila</i>	Pe
<i>Pseudomonas fluorescens</i>	Pf
<i>Pseudomonas putida</i>	Ppu
<i>Pseudomonas syringae</i>	Psy
<i>Rhodococcus jostii</i>	Rj
<i>Roseobacter denitrificans</i>	Rd
<i>Salinispora arenicola</i>	Sa
<i>Salinispora tropica</i>	St
<i>Shewanella oneidensis</i>	So
<i>Stigmatella aurantiaca</i>	Sa
<i>Streptomyces clavuligerus</i>	Scl
<i>Streptomyces coelicolor</i>	Sco
<i>Yersinia pestis</i>	Yp

APPENDIX 3.4

Appendix 3.4. Profile HMMs for fungal-specific NRPS A, (3.4A) T (3.4B), and C (3.4C) domains. Zipped text file (file name extension .hmm to be used with the program package HMMER (<http://hmmer.janelia.org>)). Available upon request and included on CD in hard copy of thesis.

APPENDIX 3.5

Appendix 3.5. Fungal and Bacterial AMP-Binding Protein Outgroups

Species	NCBI Accession	Genome Sequencing Center ID	Protein
α-amino-adipate reductases			
<i>Aspergillus fumigatus</i>	XP_751705.1	Afu4g11240	
<i>Rhizopus oryzae</i>	XP_001879618.1	RO3G12433.1	
<i>Batrachochytrium dendrobatidis</i>	XP_001879618.1	BDEG_1579.1	
<i>Cochliobolus heterostrophus</i>		CocheC5_115936	
<i>Debaryomyces hansenii</i>	XP_001385417.1	DEHA0D08734g	
<i>Fusarium graminearum</i>	XP_386217.1	FGSG06041.3	
<i>Schizosaccharomyces pombe</i>	CAB88271.1	SPAP7G5.04c	Lys1
<i>Saccharomyces cerevisiae</i>	NP_009673.1	YBR115C/ SCRG_02851.1	Lys2
<i>Neurospora crassa</i>	XP_965396.1	NCU03010.3	
<i>Phycomyces blakesleeanus</i>	XP_001879618.1	Phyb11_34455	
<i>Ustilago maydis</i>	XP_757844.1	UM01697.1	
4-Coumarate/Acyl-CoA Ligases			
<i>Cochliobolus heterostrophus</i>		CocheC5_97601	
<i>Fusarium graminearum</i>	XP_383765.1	FGSG03589.3	
<i>Rhizopus oryzae</i>		RO3G05716.3	
<i>Ustilago maydis</i>	XP_757318.1	UM01171.1	
<i>Alternaria alternata</i>	BAB6907.1		Aft
<i>Alternaria alternata</i>	BAA36588.1		Akt1
<i>Mycobacterium tuberculosis</i>	YP_001135507.1	Mflv_4250	
<i>Arthrobacter sp.</i>	YP_833499.1	Arth_4024	
<i>Streptomyces coelicolor</i>	NP_628552.1	SCO4383	
<i>Dinoroseobacter shibae</i>	YP_001531603.1	Dshi_0253	
<i>Roseobacter denitrificans</i>	YP_682165.1	RD1_1868	
<i>Arthrobacter sp.</i>	YP_833499.1	Arth_4024	
<i>Streptomyces coelicolor</i>	NP_624638.1	SC5G9.20	
<i>Rhodococcus jostii</i>	YP_705267.1	RHA1_ro05328	
Acetyl CoA Synthetases			
<i>Aspergillus fumigatus</i>	XP_751720.1	Afu4g11080	
<i>Batrachochytrium dendrobatidis</i>		BDEG00471.1	
<i>Cochliobolus heterostrophus</i>		CocheC5_11359	
<i>Saccharomyces cerevisiae</i>	EDV09449.1	SCRG_05132.2	
<i>Schizosaccharomyces pombe</i>	NP_588291.1	SPCC417.14c	
<i>Ustilago maydis</i>	XP_759216.1	UM_03069.1	
<i>Escherichia coli</i>	NP_756916.1	c5064	
<i>Yersinia pestis</i>	NP_403903.1	YPO0253	
<i>Pseudomonas syringae</i>	NP_791649.1	PSPTO_1825	
<i>Shewanella oneidensis</i>	NP_718327.1	SO_2743	
Acyl AMP Ligases (AALs)			
<i>Aspergillus fumigatus</i>	XP_752870.1	Afu1g15010	Cps1
<i>Cochliobolus heterostrophus</i>	AAG53991.2	CocheC5_66090	Cps1
<i>Schizosaccharomyces pombe</i>	NP_593217.1	SPAC56F8.02	Cps1
<i>Fusarium graminearum</i>	AAP12366.1	FGSG_06631.3	Cps1
<i>Saccharomyces cerevisiae</i>	EDV10692.1	Y0R093C/SCRG_01491.2	Cps1
<i>Myxococcus xanthus</i>	AAC44128.1	U24657.1	SafB
<i>Mycobacterium tuberculosis</i>	YP_001284310.1	MRA_2967	FadD28
<i>Lyngbya majuscula</i>	AAS98774.1		JamA

Appendix 3.5 Continued

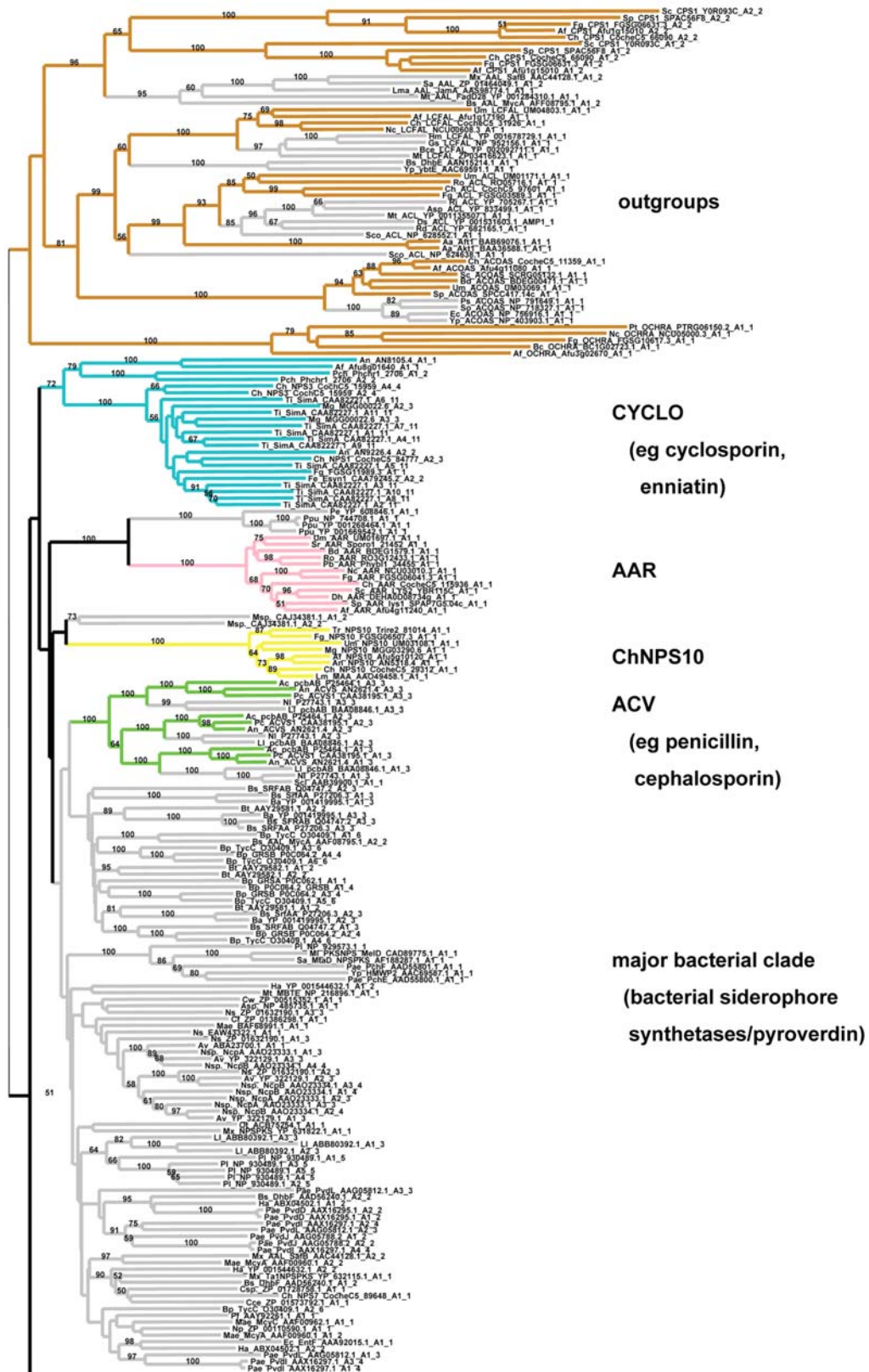
<i>Bacillus subtilis</i>	AAF08795.1		MycA
<i>Stigmatella aurantiaca</i>	ZP_01464049.1	STIAU_1156	
Long Chain Fatty Acid Acyl CoA Ligases (LCFAL)			
<i>Cochliobolus heterostrophus</i>		CocheC5_31926	
<i>Ustilago maydis</i>	XP_760950.1	UM04803.1	
<i>Aspergillus fumigatus</i>	XP_753087.1	Afu1g17190	
<i>Neurospora crassa</i>	XP_965748.1	NCU00608.3	
<i>Mycobacterium tuberculosis</i>	NP_217021.1	Rv2505c	
<i>Geobacter sulfurreducens</i>	NP_952156.1	GSU1103	
<i>Burkholderia cenocepacia</i>	YP_002092711.1	BCPG_01457.1	
<i>Heliobacterium modesticaldum</i>	YP_001678729.1	HM1_0093	
Ochratoxin (OCHRA)			
<i>Aspergillus fumigatus</i>	XP_748589.1	Afu3g02670	
<i>Pyrenophora tritici-repentis</i>	XP_001936483.1	PTRG_06150.2	
<i>Neurospora crassa</i>	XP_955820.1	NCU05000.3	
<i>Fusarium graminearum</i>	XP_390793.1	FGSG_10617.3	
<i>Botrytis cinerea</i>	XP_001558652.1	BC1G_02723.1	

Blank = none or not known

APPENDIX 3.6

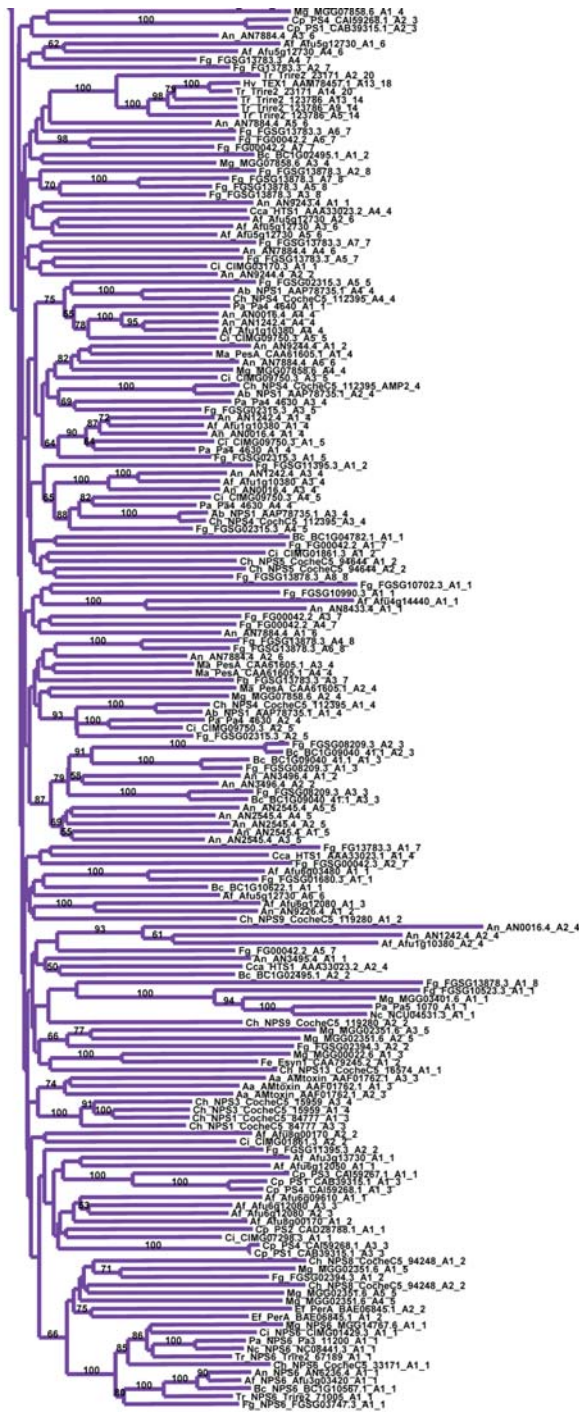
Appendix 3.6. Phylogenies resulting from analyses of the full A domain dataset.
A. NJ tree using a ML distance matrix created using the WAG plus gamma model,
B. ML tree (PhyML) using the WAG plus gamma model, and
C. ML tree (RAxML) using the RTREVF plus gamma model. Bootstrap support greater than 50% is shown under branches, where possible. Branches of monophyletic group defining subfamilies are color coded: brown: adenyating enzyme outgroups; light green: fungal PKS;NRPS hybrid synthetases (PKS:NRPS); dark orange: ChNPS11/ETP module 1 synthetases (ChNPS11/ETP mod1); dark blue: ChNPS12/ETP module 2 synthetases (ChNPS12/ETP mod2); yellow: ChNPS10-like synthetases (ChNPS10); light blue: Cyclosporin synthetases (CYCLO); pink: α -aminoadipate reductases (AAR); dark green: ACV synthetases (ACV); red: siderophore synthetases (SID); purple: Euascomycete clade synthetases (EAS). The majority of bacterial sequences (dark gray) group together and contain some fungal A domains (ACV synthetases and the NPS;PKS hybrid (ChNPS7;PKS24). The remaining bacterial A domains group with the mono/bi-modular AAR and ChNPS12/ETP mod 2 subfamilies.

Appendix 3.6A

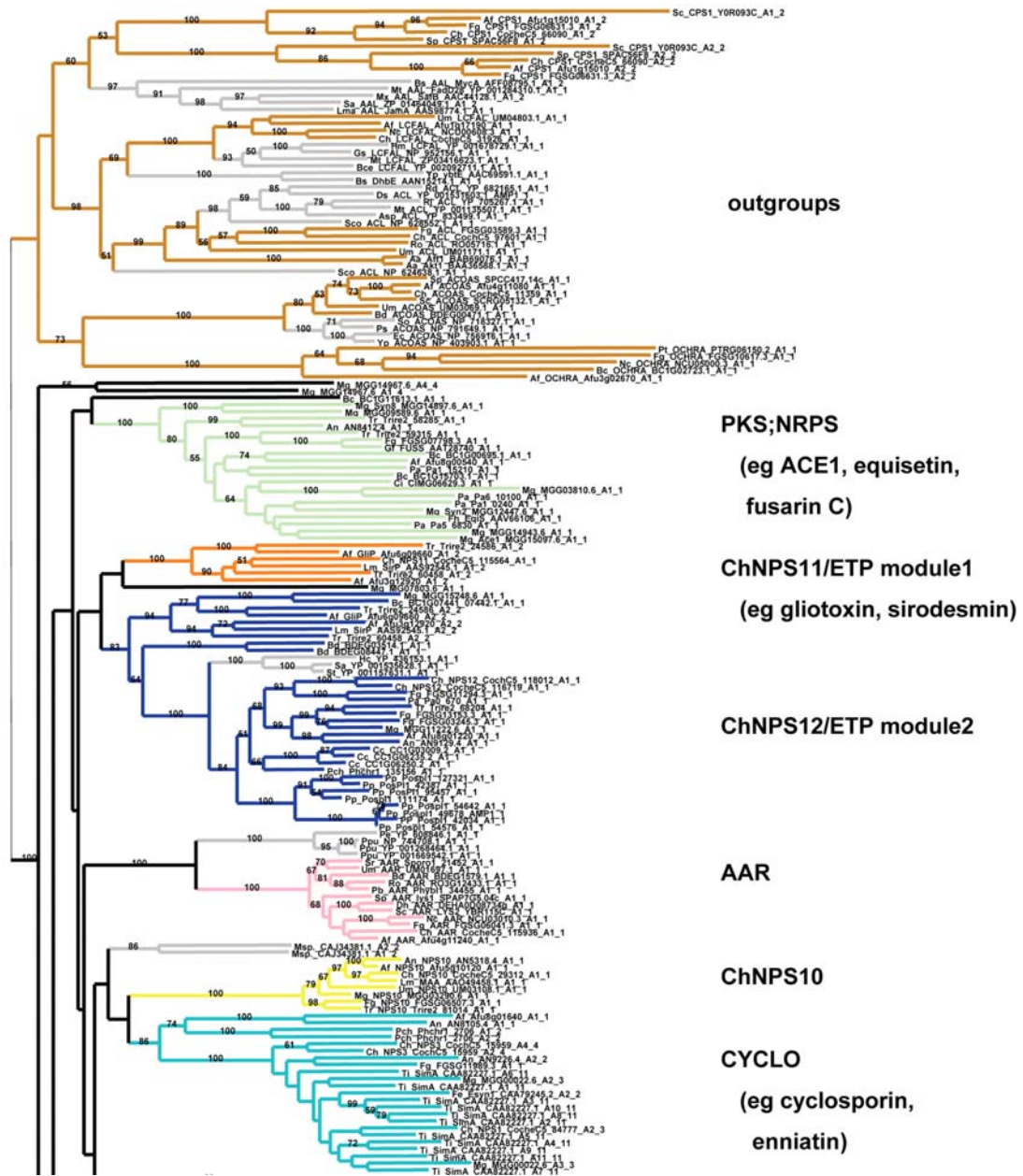


Appendix 3.6A Continued

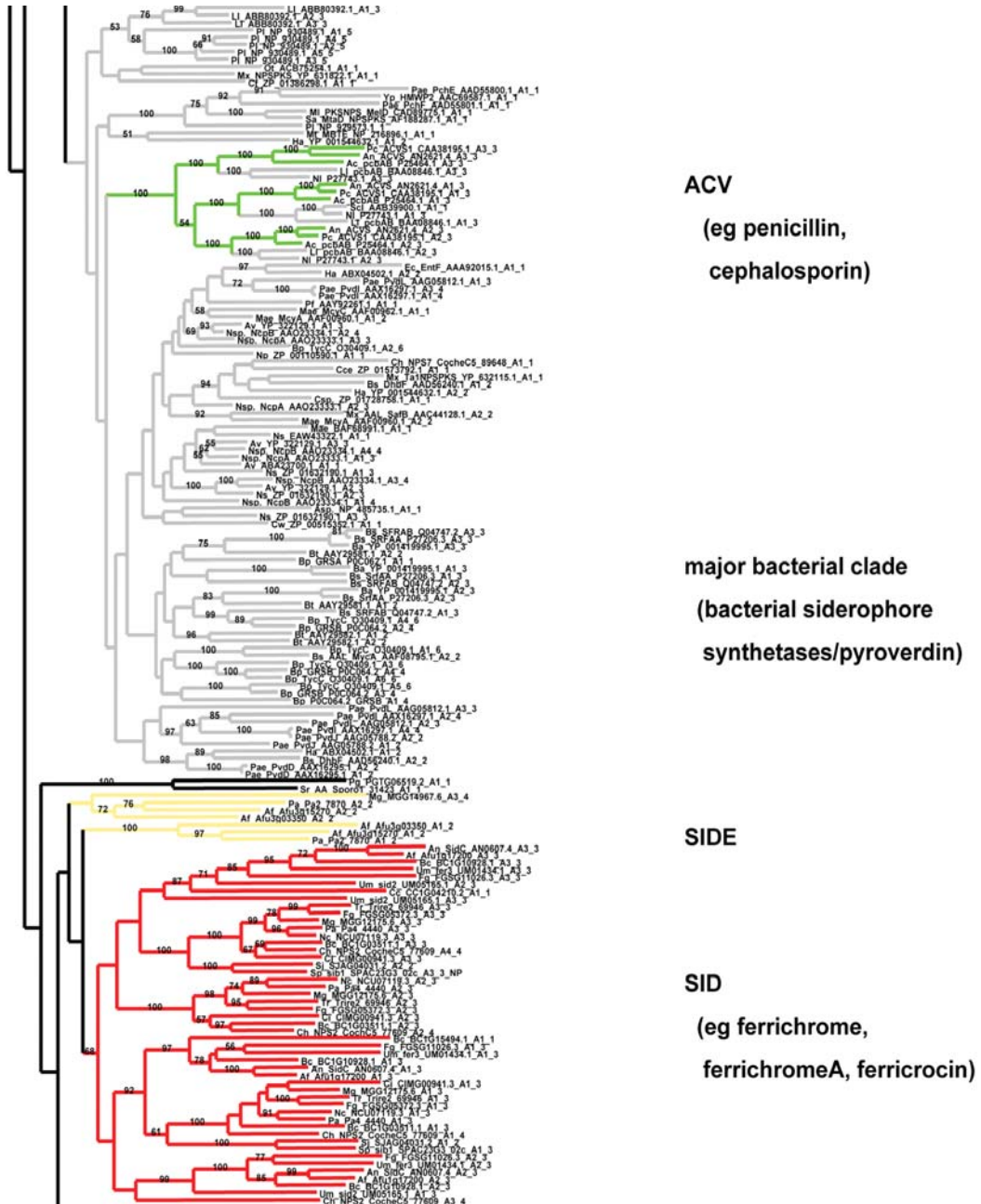
EAS Continued



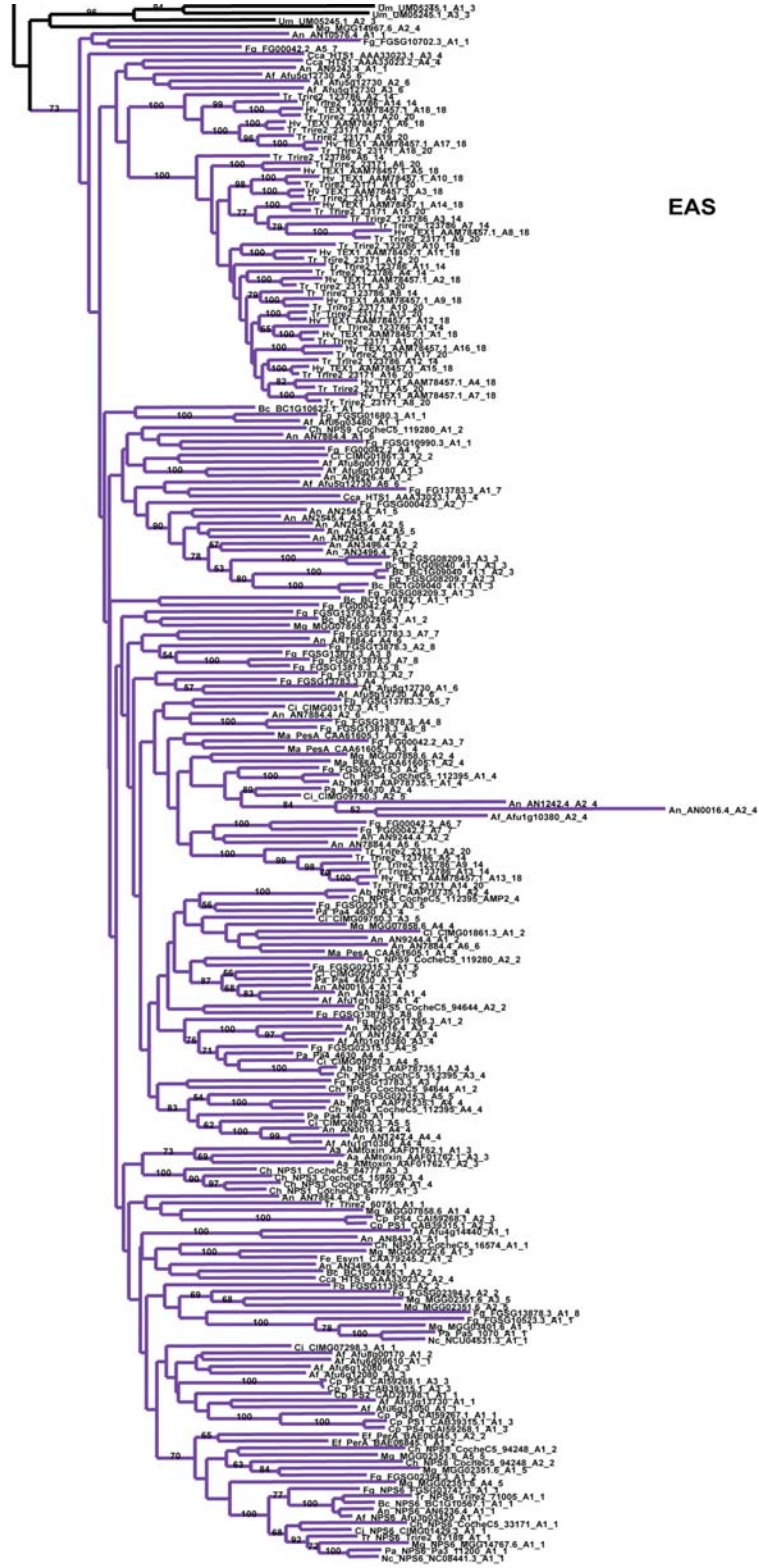
Appendix 3.6B



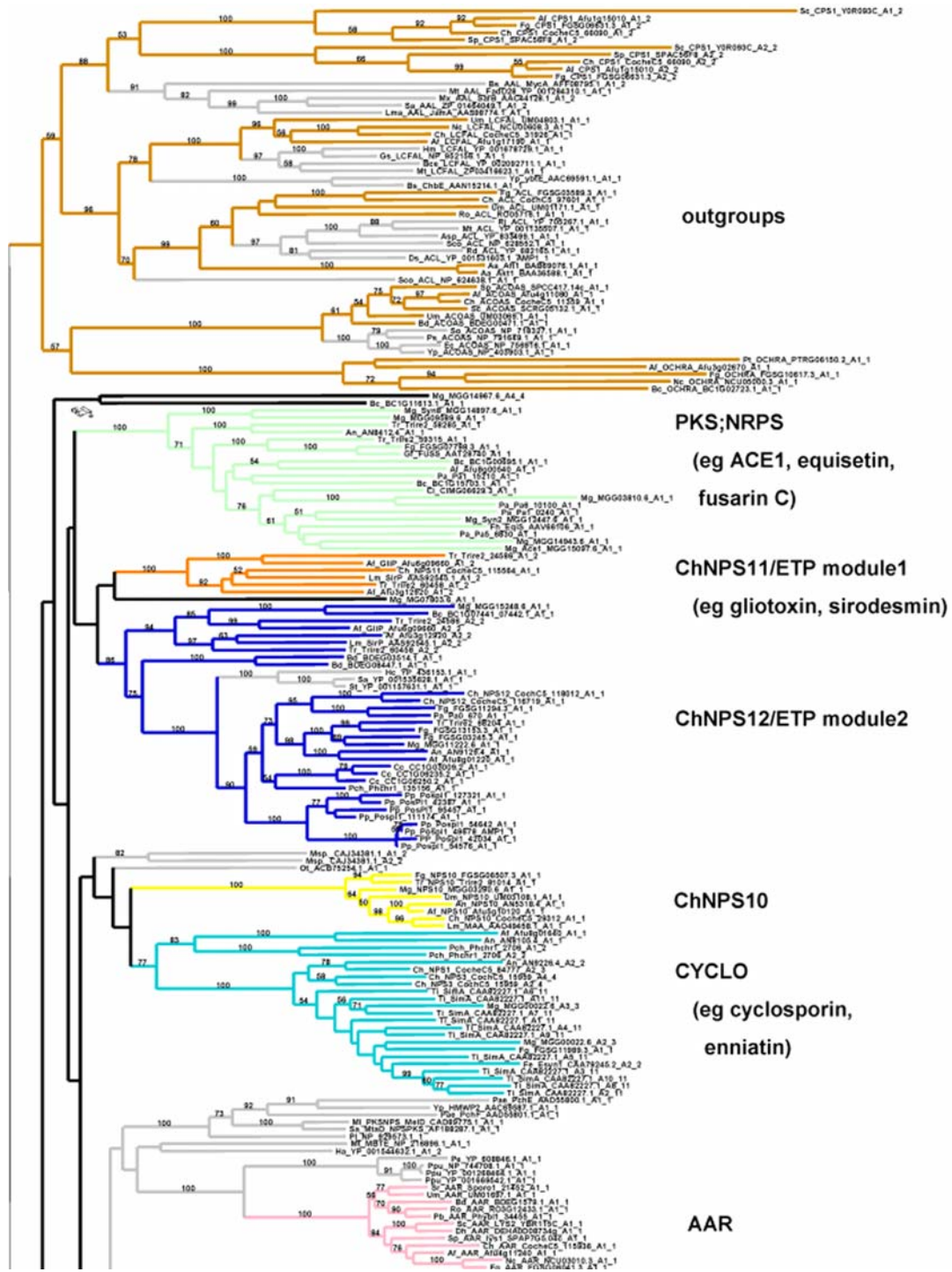
Appendix 3.6B Continued



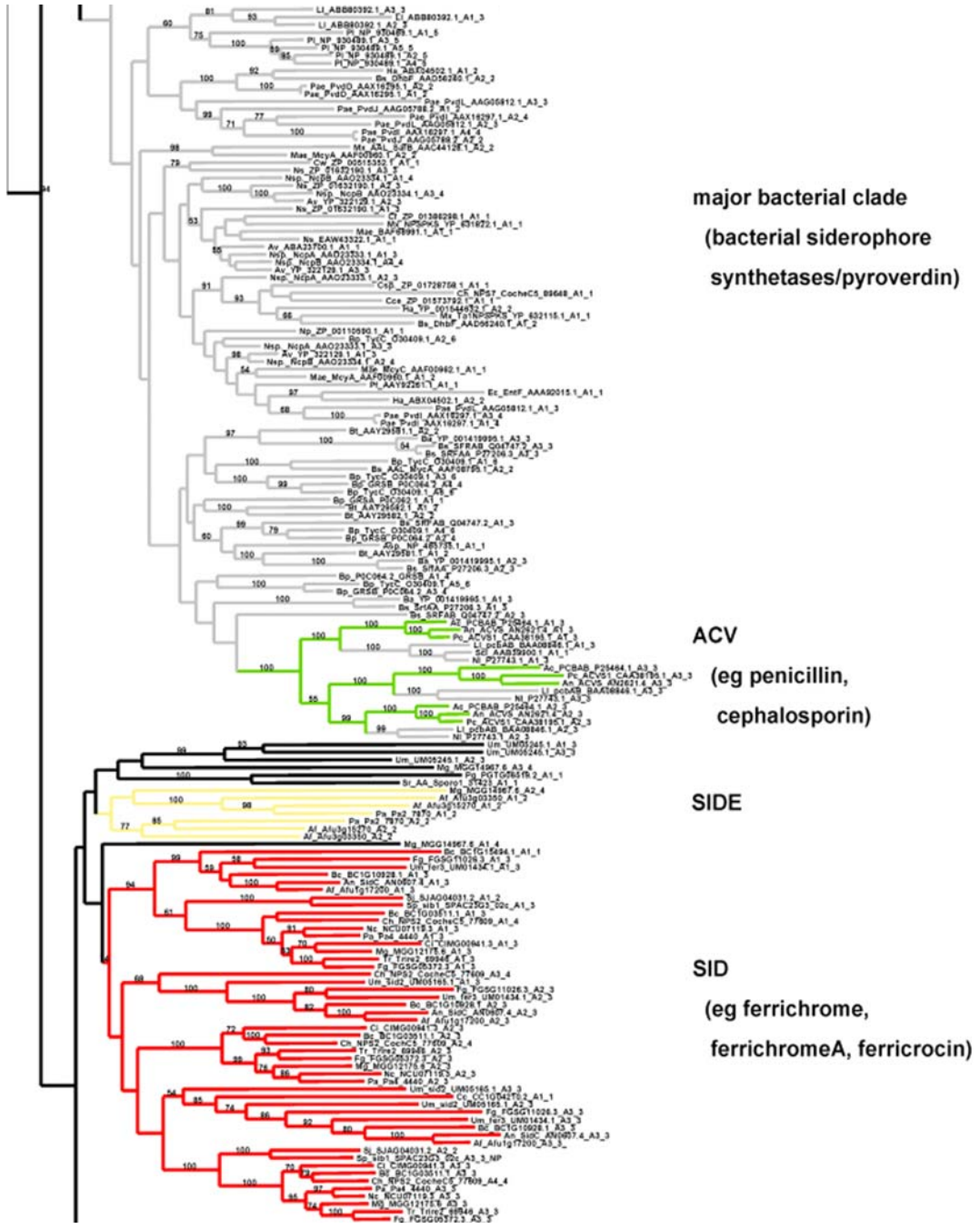
Appendix 3.6B Continued



Appendix 3.6C



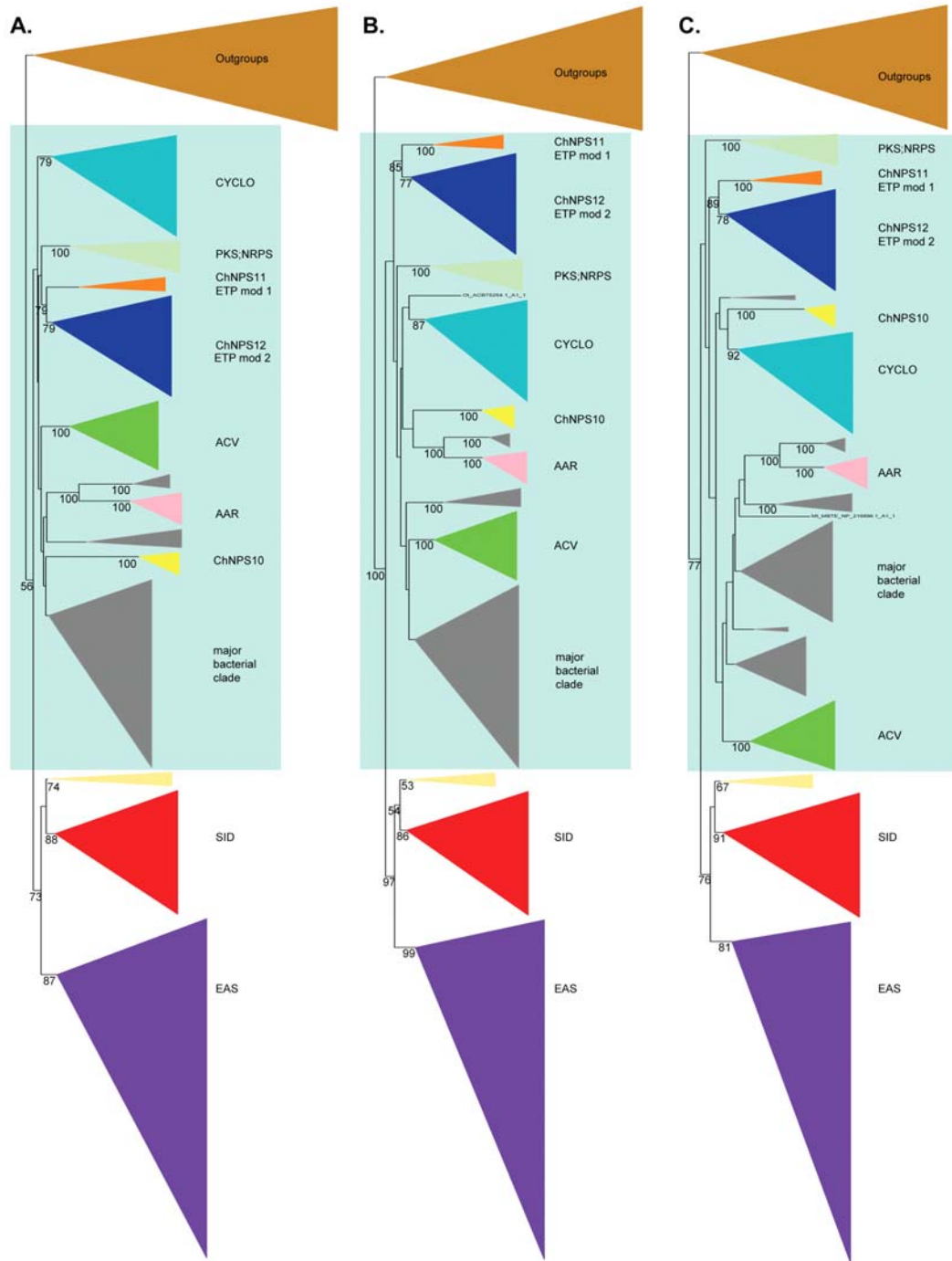
Appendix 3.6C Continued



Appendix 3.6C Continued



Appendix 3.7. Tree topologies from phylogenetic analyses of reduced dataset including representative A domains from each of the major fungal NRPS subfamilies. **A).** NJ tree using a ML distance matrix created using the WAG plus gamma model, **B).** ML tree (PhyML) using the WAG plus gamma model, and **C).** ML tree (RAxML) using the RTREVF plus gamma model. Bootstrap support greater than 50% is shown under branches. Color coding as in Appendix 3.6 and Figure 3.1. Topologies for the reduced dataset show stronger bootstrap support (>70%) for grouping the multimodular and exclusively fungal SID and EAS clades together, than do the trees resulting from analysis of the full A domain dataset.



APPENDIX 3.8

Appendix 3.8. Bacterial proteins used as outgroups

Species	NCBI Accession	Gene	Peptide Product	Reference
<i>Anabaena variabilis</i> (<i>Av</i>)	YP_322129.1			
<i>Anabaena variabilis</i> (<i>Av</i>)	ABA23700.1			
<i>Bacillus amyloliquefaciens</i> (<i>Ba</i>)	YP_001419995.1	SrfAA	Surfactin A	[121]
<i>Bacillus subtilis</i> (<i>Bs</i>)	AAD56240.1	DhbF	Bacillibactin	[122]
	P27206.3	SrfAA	Surfactin A	[123]
	AAN15214.1	DhbE	Bacillibactin	[124]
	Q04747.2	SrfAB	Surfactin B	[123]
<i>Brevibacillus brevis</i> (<i>Bb</i>)	P0C064.2	GrsB	Gramicidin B	[125]
	P0C062.1	GrsA	Gramicidin A	[126]
<i>Brevibacillus parabrevis</i> (<i>Bp</i>)	O30409.1	TycC	Tyrocidine C	[127]
<i>Brevibacillus texasporus</i> (<i>Bt</i>)	AA29581.1	BtD	BT Peptide	[128]
	AA29582.1	BtE	BT Peptide	[128]
<i>Chlorobium ferrooxidans</i> (<i>Cf</i>)	ZP_01386298.1			
<i>Clostridium cellulolyticum</i> (<i>Cce</i>)	ZP_01573792.1			
<i>Crocospaera watsonii</i> (<i>Cw</i>)	ZP_00515352.1			
<i>Cyanothece</i> sp. <i>CCY0110</i> (<i>Csp.</i>)	ZP_01728758.1			
<i>Escherichia coli</i> (<i>Eco</i>)	AAA92015.1	EntF	Enterobactin	[129]
<i>Hahella chejuensis</i> (<i>Hc</i>)	YP_436153.1			
<i>Herpetosiphon aurantiacus</i> (<i>Ha</i>)	YP_001544632.1			
	ABX04502.1			
<i>Lysobacter lactamgenus</i> (<i>Ll</i>)	BAA08846.1	pcbAB	Cephalosporin	[130]
	ABB80392.1	cpbI	(NRPS;PKS)	
<i>Melittangium lichenicola</i> (<i>Ml</i>)	CAD89775.1	MelD	Melithiazol	[131]
			(PKS;NRPS)	
<i>Microcystis aeruginosa</i> (<i>Mae</i>)	AAF00960.1	McyA	Microcystin	[132]
	AAF00962.1	McyC	Microcystin	[132]
	BAF68991.1	psm3B	(NRPS;PKS)	[133]
<i>Micromonospora</i> sp. (<i>Msp.</i>)	CAJ34381.1	tioY	Thiocoraline	
<i>Mycobacterium tuberculosis</i> (<i>Mt</i>)	NP_216896.1	MBTE	MBTE	[134]
			siderophore	
<i>Myxococcus xanthus</i> (<i>Mx</i>)	YP_631822.1		NRPS;PKS	[135]
	YP_632115.1	TaI	NRPS;PKS	[135]
<i>Nocardia lactamdurans</i> (<i>Nl</i>)	P27743.1	pcbAB	Cephamycin	[136]
<i>Nodularia spumigena</i> (<i>Ns</i>)	EAW43322.1			
	ZP_01632190.1			
	ZP_01632190.1			
<i>Nostoc punctiforme</i> (<i>Np</i>)	ZP_00110590.1			
<i>Nostoc</i> sp. (<i>Nsp.</i>)	AAO23333.1	NcpA	4-Methylproline	[137]
	AAO23334.1	NcpB	4-Methylproline	[137]
<i>Opiritatus terrae</i> (<i>Ot</i>)	ACB75254.1			
<i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> (<i>Pl</i>)	NP_929573.1		NRPS;PKS	[138]
	NP_930489.1			[138]
<i>Pseudomonas aeruginosa</i> (<i>Pae</i>)	AAD55800.1	PchE	Pyochelin	[139]
	AAD55801.1	PchF	Pyochelin	[139]
	AAX16295.1	PvdD	Pyoverdine	[140]
	AAX16297.1	PvdI	Pyoverdine	[140]

Appendix 3.8 Continued

	AAG05788.2	PvdJ	Pyoverdine	[141]
	AAG05812.1	PvdL	Pyoverdine	[141]
<i>Pseudomonas entomophila (Pe)</i>	YP_608846.1			[142]
<i>Pseudomonas fluorescens (Pf)</i>	AAAY92261.1			[143]
<i>Pseudomonas putida F1 (Ppu)</i>	YP_001268464.1			[144]
	YP_001669542.1			[144]
	NP_744708.1			[144]
<i>Salinispora arenicola CNS-205 (Sar)</i>	YP_001535628.1			
	YP_001157631.1			
<i>Salinispora tropica (St)</i>	YP_001157631.1			
<i>Stigmatella aurantiaca (Sau)</i>	AF188287.1	MtaA- MtaG	Myxothiazol	[145]
<i>Streptomyces clavuligerus (Scl)</i>	AAB39900.1	pcbAB	Penicillin	[146]
<i>Yersinia pestis (Yp)</i>	AAC69591.1	ybtE	Yersiniabactin	[147]
	AAC69587.1	HMWP2	Yersiniabactin	[147, 148]

Blank = unknown or not published

APPENDIX 3.9

Appendix 3.9. Ultrametric species tree used for CAFÉ analyses. Tree was created with the PL method in r8s [119] using the phylogeny of the concatenated protein dataset of Fitzpatrick et al. [120]. We used 5 calibration points (Dikarya = 452 MYA, Basidiomycetes = 340 MYA, Ascomycetes 400 MYA, Pezizomycetes = 215 MYA, and Sordariomycetes = 122 MYA) estimated previously by Taylor and Berbee [149] when fixing the 400 MYO fungal fossil *Paleopyrenomyces devonicus* at the origin of the ascomycetes. The root taxon *R. oryzae* was constrained to be less than the origin of the Fungi (495 MYA) estimated in this study [149]. Assigning the *Paleopyrenomyces* at the origins of ascomycota as opposed to the other suggested dates for this fossil (at the origins of Pyrenomyces and Sordariomycetes respectively) gives time estimates for the origins of Glomeromycota best coinciding with the radiation of land plants [149].

Ultrametric Tree:

(Roryz:480,((Umayd:340,((Ccin:185,Pchry:185):102,Cneo:287):53)BA:112,(Spomb:400,(((Afum:72,Anid:72):91,Cimm:163):52,(Chet:183,(Bcin:153,((Trees:78,fgram:78):44,(Mgris:88,(Ncras:62,Pans:62):26):34)SO:31):30):32)EA:145,(Ylip:290,(((calb:71,ctrop:71):63,((dhans:94,Cguill:94):17,Clus:111):23):79,((klact:87,Agoss:87):26,(Sbay:20,(Smik:14,(Scer:10,Spar:10):4):6):93):101):77):70):40)AS:52)DK:28);

APPENDIX 3.10

Appendix 3.10. Known fungal NRPSs used in constructing initial HMM model

Species	NCBI Protein Accession #	Sequencing Center ID ^a	NRPS Name/Product
<i>Alternaria alternata</i>	AAF01762.1		AMT/AM-toxin
<i>Alternaria brassicae</i>	AAP78735.1		NPS1
<i>Acremonium chrysogenum</i>	P25464.1		PCBAB/Cephalosporin
<i>Aspergillus fumigatus</i>	EAL88817.1	Afu6g09660	GliP/Gliotoxin
	EAL91592.1	Afu5g10120	NPS10
	EAL86624.1	Afu3g03420	NPS6/TAFC
<i>Aspergillus nidulans</i>	XP_660225.1	AN2621.4	ACVS/Penicillin
<i>Cochliobolus carbonum</i>	AAA33023.1		HTS1/HC-toxin
<i>Cochliobolus heterostrophus</i> ^c	AAX09983.1		NPS1
	AAX09984.1		NPS2/ferricrocin
	AAX09985.1		NPS3
	AAX09986.1		NPS4
	AAX09987.1		NPS5
	AAX09988.1		NPS6/coprogen
	AAX09989.1		NPS7
	AAX09990.1		NPS8
	AAX09991.1		NPS9
	AAX09992.1		NPS10
	AAX09993.1		NPS11
	AAX09994.1		NPS12
<i>Claviceps purpurea</i>	CAB39315.1		PS1/D-lysergic acid
<i>Epichloë festuca</i>	BAE06845.1		PerA/Peramine
<i>Fusarium equiseti</i>	CAA79245.2		Esyn1/Enniatin
<i>Fusarium graminearum</i>	XP_383923.1	FG03747.1	NPS6/coprogen
	XP_386683.1	FGSG_06507.3	NPS10
	XP_383923.1	FG03747.1	NPS6/coprogen
<i>Fusarium heterosporum</i>	AAV66106.1		EqiS/Equisetin
<i>Gibberella fujikuroi</i>	AAT28740.1		FUSS/Fusarin C
<i>Hypocrea virens</i>	AAM78457.1		TEX1/peptaibol
<i>Leptosphaeria maculans</i>	AAO49458.1		MAA
	AAS92545.1		SirP/sirodesmin PL
<i>Metarhizium anisopliae</i>	CAA61605.1		PesA
<i>Magnaporthe oryzae</i>	CAG28798.1	MGG_15097.6	Ace1
	XP_360747.1	MGG_03290.6	NPS10
	XP_364124.2	MGG_14767.6	NPS6/coprogen
	CAG28798.1	MGG_12447.6	Syn2
	CAH59193.1	MGG_12447.6	Syn8
	XP_963411.2	NCU_08441.3	NPS6/coprogen
<i>Neurospora crassa</i>	XP_963411.2	NCU_08441.3	NPS6/coprogen
<i>Penicillium chrysogenum</i>	CAA38195.1		ACVS1/Penicillin
	CAD28788.1		PS2/ergotamine
	CAI59267.1		PS3
	CAI59268.1		PS4/ergocryptine
<i>Schizosaccharomyces pombe</i>	CAB88271.1		Lys1/AAR
<i>Tolypocladium inflatum</i>	CAA82227.1		SimA/Cyclosporin
<i>Ustilago maydis</i>	XP_759255.1	UM03108.1	NPS10
	AAB93493.1	UM05165.1	sid2/ferrichrome
	XP_757581.1	UM01434.1	fer3/ferrichrome A

^a Blank = not applicable, ^b Blank = unpublished, ^c From *C. heterostrophus* strain C4

APPENDIX 3.11

Appendix 3.11. HMMER AMP domain models used as the initial model for NRPS identification. Zipped text files (file name extension .hmm to be used with the program package HMMER (<http://hmmer.janelia.org>). Available upon request and included on CD in hard copy of thesis.

APPENDIX 3.12

Appendix 3.12. Fungal Protein Datasets used in phylogenomic analyses

Classification/Species	Lifestyle	URL	Ref. ^a
Chytridiomycota			
<i>Batrachochytrium dendrobatidis</i> (JEL423)	animal pathogen	http://www.broad.mit.edu/annotation/genome/batrachochytrium_dendrobatidis/	
Zygomycota			
<i>Rhizopus oryzae</i> (RA99-880)	saprobe	http://www.broad.mit.edu/annotation/genome/rhizopus_oryzae/MultiHome.html	
<i>Phycomyces blakesleeanus</i> (NRRL1555)	saprobe	http://genome.jgi-psf.org/Phybl1/Phybl1.home.html	
Microsporidia			
<i>Encephalitozoon cuniculi</i> (GB-M1)	animal pathogen	http://www.genoscope.cns.fr/spip/Encephalitozoon-cuniculi-whole.html	[70]
Schizosaccharomycota			
<i>Schizosaccharomyces pombe</i> (972h)	saprobe	http://www.broad.mit.edu/annotation/genome/schizosaccharomyces_group/MultiHome.html	[150]
<i>Schizosaccharomyces japonicus</i> (yFS275)	saprobe	http://www.broad.mit.edu/annotation/genome/schizosaccharomyces_group/MultiHome.html	
Hemiascomycota			
<i>Ashbya gossypii</i> (ATCC 10895)	plant pathogen	Ashbya Genome Database: http://agd.vital-it.ch/index.html	[151]
<i>Candida albicans</i> (WO1)	animal pathogen	http://www.broad.mit.edu/annotation/genome/candida_group/MultiHome.html	[152]
<i>Candida glabrata</i> (CBS138)	animal pathogen	http://www.genolevures.org/cagl.html#	[153]
<i>Candida guilliermondii</i> (ATCC6260)	animal pathogen	http://www.broad.mit.edu/annotation/genome/candida_group/MultiHome.html	
<i>Candida lusitanae</i> (ATCC42720)	animal pathogen	http://www.broad.mit.edu/annotation/genome/candida_group/MultiHome.html	
<i>Candida tropicalis</i> (CBS94)	animal pathogen	http://www.broad.mit.edu/annotation/genome/candida_group/MultiHome.html	[154]
<i>Saccharomyces cerevisiae</i> (S288C)	saprobe	http://www.yeastgenome.org/	[155]
<i>Saccharomyces paradoxicus</i> (NRRLY-17217)	saprobe	Broad Institute, GenBank Accession AABZ00000000	[156]
<i>Saccharomyces bayanus</i> (MCYC623)	saprobe	Broad Institute, GenBank Accession AACAA00000000	[156]
<i>Saccharomyces mikatae</i> (IFO1815)	saprobe	Broad Institute, GenBank Accession AABZ00000000	[156]
<i>Debaryomyces hansenii</i> (CBS767)	saprobe	http://www.genolevures.org/deha.html	[153]
<i>Kluyveromyces lactis</i> var. <i>lactis</i> (CLIB210)	saprobe	http://www.genolevures.org/klla.html#	[153]
<i>Yarrowia lipolytica</i> (CLIB99)	saprobe	http://www.genolevures.org/yali.html#	[153]
Euascomycota			
<i>Aspergillus nidulans</i> (FGSC A4)	saprobe	http://www.broad.mit.edu/annotation/genome/aspergillus_group/MultiHome.html	[157]
<i>Aspergillus fumigatus</i> (Af293)	animal pathogen	CADRE: http://www.cadre-genomes.org.uk/aspergillus_links.html	[158]
<i>Magnaporthe oryzae</i> (70-15)	plant pathogen	http://www.broad.mit.edu/annotation/genome/magnaporthe_grisea/MultiHome.html	[159]
<i>Fusarium graminearum</i> (PH-1)	plant pathogen	http://www.broad.mit.edu/annotation/genome/fusarium_group/MultiHome.html	[160]
<i>Botrytis cinerea</i> (B05.10)	plant pathogen	http://www.broad.mit.edu/annotation/genome/botrytis_cinerea/	

Appendix 3.12 Continued

<i>Coccidioides immitis</i> (RS)	animal pathogen	http://www.broad.mit.edu/annotation/genome/coccidioides_group/MultiHome.html	
<i>Cochliobolus heterostrophus</i> (C5)	plant pathogen	http://genome.jgipsf.org/CocheC5_1/CocheC5_1_home.html	
<i>Neurospora crassa</i> (OR74A)	saprobe	http://www.broad.mit.edu/annotation/genome/neurospora/	[68]
<i>Podospora anserina</i> (DSM 980)	saprobe	http://podospora.igmors.u-psud.fr/	[161]
<i>Trichoderma reesii</i> (QM6a)	Mycoparasite	http://genome.jgi-psf.org/Trire2/Trire2.home.html	[162]
Basidiomycota:			
<i>Coprinopsis cinerea</i> (Okayama 7#130)	saprobe	http://www.broad.mit.edu/annotation/genome/coprinus_cinereus/MultiHome.html	
<i>Picia stipitis</i> (NRRL Y-11545)	saprobe	http://genome.jgi-psf.org/Picst3/Picst3.home.html	[163]
<i>Cryptococcus neoformans</i> var. <i>grubii</i> (serotype A – H99)	animal pathogen	http://www.broad.mit.edu/annotation/genome/cryptococcus_neoformans/MultiHome.html	[164]
<i>Puccinia graminis</i> (CRL 75-36-700-3)	plant pathogen	http://www.broad.mit.edu/annotation/genome/puccinia_graminis/	
<i>Postia placenta</i> (Mad-698-R)	saprobe	http://genome.jgi-psf.org/Pospl1/Pospl1.home.html	
<i>Phanaerochaete chrysosporium</i> (RP78)	saprobe	http://genome.jgi-psf.org/Phchr1/Phchr1.home.html	[165]
<i>Laccaria bicolor</i> (S238N-H82)	saprobe	http://genome.jgi-psf.org/Lacbi1/Lacbi1.home.html	[166]
<i>Sporobolomyces roseus</i>	saprobe	http://genome.jgi-psf.org/Sporo1/Sporo1.home.html	
<i>Ustilago maydis</i> (521)	plant pathogen	http://www.broad.mit.edu/annotation/genome/ustilago_maydis/	[167]

^a Blank = unpublished

APPENDIX 3.13

Appendix 3.13. MUSCLE alignment of 558 fungal and bacterial AMP domains used in phylogenetic analyses of the complete dataset. Zipped text file containing alignment in fasta format for visualization in sequence alignment editor such as ClustalX [122]. Available upon request and included on CD in hard copy of thesis.

APPENDIX 3.14

Appendix 3.14. MUSCLE alignment of the reduced dataset of fungal and bacterial AMP domains containing selected representatives of each major fungal subfamily and bacterial clades. Zipped text file containing alignment in fasta format for visualization in sequence alignment editor such as ClustalX [122]. Available upon request and included on CD in hard copy of thesis.

REFERENCES

1. Finking R, Marahiel MA: **Biosynthesis of nonribosomal peptides.** *Annual Review of Microbiology* 2004, **58**:453-488.
2. Sieber SA, Marahiel MA: **Learning from nature's drug factories: Nonribosomal synthesis of macrocyclic peptides.** *Journal of Bacteriology* 2003, **185**(24):7036-7043.
3. Grunewald J, Marahiel MA: **Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides.** *Microbiology and Molecular Biology Reviews* 2006, **70**(1):121-+.
4. Stein T, Vater J, Kruft V, Otto A, WittmannLiebold B, Franke P, Panico M, McDowell R, Morris HR: **The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates.** *Journal of Biological Chemistry* 1996, **271**(26):15428-15435.
5. Mootz HD, Schwarzer, Dirk, and Mohamed A. Marahiel: **Ways of assembling complex natural products on modular nonribosomal peptide synthetases.** *ChemBioChem* 2002, **3**:490-504.
6. Oide S, Moeder W, Krasnoff S, Gibson D, Haas H, Yoshioka K, Turgeon BG: **NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes.** *Plant Cell* 2006, **18**(10):2836-2853.
7. Oide S, Krasnoff SB, Gibson DM, Turgeon BG: **Intracellular siderophores are essential for ascomycete sexual development in heterothallic *Cochliobolus heterostrophus* and homothallic *Gibberella zeae*.** *Eukaryotic Cell* 2007, **6**(8):1339-1353.
8. Oide S: **Functional characterization of nonribosomal peptide synthetases in the filamentous ascomycete phytopathogen *Cochliobolus heterostrophus*.** *PhD.* Ithaca, NY: Cornell University; 2007.

9. Kim KH, Cho Y, La Rota M, Cramer RA, Lawrence CB: **Functional analysis of the *Alternaria brassicicola* nonribosomal peptide synthetase gene AbNPS2 reveals a role in conidial cell wall construction.** *Molecular Plant Pathology* 2007, **8**(1):23-39.
10. Lee BN, Kroken S, Chou DYT, Robbertse B, Yoder OC, Turgeon BG: **Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress.** *Eukaryotic Cell* 2005, **4**(3):545-555.
11. Hahn J, Dubnau D: **Growth stage signal transduction and the requirements for SRFA induction in development of competence.** *Journal of Bacteriology* 1991, **173**(22):7275-7282.
12. Schaeffer P: **Sporulation and the production of antibiotics, exoenzymes, and exotoxins.** *Bacteriol Rev* 1969, **33**(1):48-71.
13. Horinouchi S, Beppu T: **Autoregulatory factors of secondary metabolism and morphogenesis in actinomycetes.** *Crit Rev Biotechnol* 1990, **10**(3):191-204.
14. Marahiel MA, Stachelhaus T, Mootz HD: **Modular peptide synthetases involved in nonribosomal peptide synthesis.** *Chemical Reviews* 1997, **97**(7):2651-2673.
15. Challis GL, Ravel J, Townsend CA: **Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.** *Chemistry and Biology (London)* 2000, **7**(3):211-224.
16. Keating TA, Ehmann DE, Kohli RM, Marshall CG, Trauger JW, Walsh CT: **Chain termination steps in nonribosomal peptide synthetase assembly lines: Directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis.** *Chembiochem* 2001, **2**(2):99-107.
17. Keating TA, Walsh CT: **Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis.** *Current Opinion in Chemical Biology* 1999, **3**(5):598-606.

18. Schneider A, Marahiel MA: **Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis***. *Archives of Microbiology* 1998, **169**(5):404-410.
19. Kohli RM, Trauger JW, Schwarzer D, Marahiel MA, Walsh CT: **Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases**. *Biochemistry* 2001, **40**(24):7099-7108.
20. Pospiech A, Bietenhader J, Schupp T: **Two multifunctional peptide synthetases and an O-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumour agent saframycin Mx1 from *Mycococcus xanthus***. *Microbiology-UK* 1996, **142**:741-746.
21. Silakowski B, Kunze B, Nordsiek G, Blocker H, Hofle G, Muller R: **The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15**. *European Journal of Biochemistry* 2000, **267**(21):6476-6485.
22. Silakowski B, Nordsiek G, Kunze B, Blocker H, Muller R: **Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15**. *Chemistry & Biology* 2001, **8**(1):59-69.
23. Ehmman DE, Gehring AM, Walsh CT: **Lysine biosynthesis in *Saccharomyces cerevisiae*: Mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5**. *Biochemistry* 1999, **38**(19):6171-6177.
24. Walzel B, Riederer B, Keller U: **Mechanism of alkaloid cyclopeptide synthesis in the ergot fungus *Claviceps purpurea***. *Chemistry & Biology* 1997, **4**(3):223-230.
25. Pfeifer E, Pavelavrancic M, Vondohren H, Kleinkauf H: **Characterization of Tyrocidine synthetase-1 (Ty1) - Requirement of posttranslational modification for peptide biosynthesis**. *Biochemistry* 1995, **34**(22):7450-7459.

26. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH: **Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution.** *BMC Evolutionary Biology* 2007, **7**.
27. Walsh CT, Chen HW, Keating TA, Hubbard BK, Losey HC, Luo LS, Marshall CG, Miller DA, Patel HM: **Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines.** *Current Opinion in Chemical Biology* 2001, **5(5):525-534**.
28. Samel SA, Marahiel MA, Essen LO: **How to tailor non-ribosomal peptide products - new clues about the structures and mechanisms of modifying enzymes.** *Molecular Biosystems* 2008, **4(5):387-393**.
29. Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG: **Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100(26):15670-15675**.
30. Collemare J, Pianfetti M, Houille AE, Morin D, Camborde L, Gagey MJ, Barbisan C, Fudal I, Lebrun MH, Boehnert HU: ***Magnaporthe grisea* avirulence gene *ACE1* belongs to an infection-specific gene cluster involved in secondary metabolism.** *New Phytologist* 2008, **179(1):196-208**.
31. Maiya S, Grundmann A, Li X, Li SM, Turner G: **Identification of a hybrid PKS/NRPS required for pseurotin A biosynthesis in the human pathogen *Aspergillus fumigatus*.** *Chembiochem* 2007, **8(14):1736-1743**.
32. Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C: **Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*.** *Nature Chemical Biology* 2007, **3(4):213-217**.
33. Brendel N, Partida-Martinez LP, Scherlach K, Hertweck C: **A cryptic PKS-NRPS gene locus in the plant commensal *Pseudomonas fluorescens* Pf-5 codes for the biosynthesis of an antimitotic rhizoxin complex.** *Organic & Biomolecular Chemistry* 2007, **5(14):2211-2213**.
34. Rees DO, Bushby N, Cox RJ, Harding JR, Simpson TJ, Willis CL: **Synthesis of [1,2-C-13(2), N-15]-L-homoserine and its incorporation by the PKS-NRPS system of *Fusarium moniliforme* into the mycotoxin Fusarin C.** *Chembiochem* 2007, **8(1):46-50**.

35. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A *et al*: **Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus***. *PLOS Genetics* 2008, **4**(4).
36. Chang PK, Horn BW, Dörner JW: **Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic *Aspergillus flavus* isolates**. *Fungal Genetics and Biology* 2005, **42**(11):914-923.
37. Cramer RA, Stajich, J.E., Yvonne Yamanaka, Dietrich, F.S., Steinbach, William, J.S., and Perfect, J.R.: **Phylogenomic analysis of non-ribosomal peptide synthetases in the genus *Aspergillus***. *Gene* 2006, **383**(15):24-32.
38. Johnson R, Voisey C, Johnson L, Pratt J, Fleetwood D, Khan A, Bryan G: **Distribution of NRPS gene families within the *Neotyphodium/Epichloe* complex**. *Fungal Genetics and Biology* 2007, **44**(11):1180-1190.
39. Nei M, Hughes, A.L.: **Balanced polymorphism and evolution by the birth-and-death process in the MHC loci**. In: *11th Histocompatibility Workshop and Conference: 1992*: Oxford Univ. Press; 1992.
40. de Bono B, Madera M, Chothia C: **V-H gene segments in the mouse and human genomes**. *Journal of Molecular Biology* 2004, **342**(1):131-143.
41. Hamilton AT, Huntley S, Tran-Gyamfi M, Baggott DM, Gordon L, Stubbs L: **Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes**. *Genome Research* 2006, **16**(5):584-594.
42. Tian X, Pascal G, Fouchecourt S, Pontarotti P, Monget P: **Gene birth, death, and divergence: The different scenarios of reproduction-related gene evolution**. *Biology of Reproduction* 2009, **80**(4):616-621.
43. Nozawa M, Nei M: **Evolutionary dynamics of olfactory receptor genes in *Drosophila* species**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(17):7122-7127.

44. Nozawa M, Kawahara Y, Nei M: **Genomic drift and copy number variation of sensory receptor genes in humans.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(51):20421-20426.
45. Niimura Y: **Evolutionary dynamics of olfactory receptor genes in mammals.** *Genes & Genetic Systems* 2007, **82**(6):503-503.
46. Niimura Y, Nei M: **Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates.** *Journal of Human Genetics* 2006, **51**(6):505-517.
47. Niimura Y, Nei M: **Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice.** *Gene* 2005, **346**:13-21.
48. Nam J, Kim J, Lee S, An GH, Ma H, Nei MS: **Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(7):1910-1915.
49. Xu GX, Ma H, Nei M, Kong HZ: **Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(3):835-840.
50. Turgeon BG, Oide S, Bushley K: **Creating and screening *Cochliobolus heterostrophus* non-ribosomal peptide synthetase mutants.** *Mycological Research* 2008, **112**:200-206.
51. Scottcraig JS, Panaccione DG, Pocard JA, Walton JD: **The cyclic peptide synthetase catalyzing HC-Toxin production in the filamentous fungus *Cochliobolus-carbonum* is encoded by a 15.7-kilobase open reading frame.** *Journal of Biological Chemistry* 1992, **267**(36):26044-26049.
52. Johnson RD, Johnson L, Itoh Y, Kodama M, Otani H, Kahmoto K: **Cloning and characterization of a cyclic peptide synthetase gene from *Alternaria alternata* apple pathotype whose product is involved in AM-toxin synthesis and pathogenicity.** *Molecular Plant-Microbe Interactions* 2000, **13**(7):742-753.

53. Wapinski I, Pfeffer A, Friedman N, Regev A: **Natural history and evolutionary principles of gene duplication in fungi.** *Nature* 2007, **449**(7158):54-U36.
54. Lu SW, Kroken S, Lee BN, Robbertse B, Churchill ACL, Yoder OC, Turgeon BG: **A novel class of gene controlling virulence in plant pathogenic ascomycete fungi.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(10):5980-5985.
55. Suvarna K, Seah L, Bhattacharjee V, Bhattacharjee JK: **Molecular analysis of the LYS2 gene of *Candida albicans*: homology to peptide antibiotic synthetases and the regulation of the alpha-aminoadipate reductase.** *Current Genetics* 1998, **33**(4):268-275.
56. Eibel H, Philippsen P: **Identification of the cloned *S. cerevisiae* Lys2 gene by an integrative transformation approach.** *Molecular & General Genetics* 1983, **191**(1):66-73.
57. Sinha AK, Bhattach.Jk: **Lysine biosynthesis in *Saccharomyces* - Conversion of alpha-aminoadipate into alpha-aminoadipic delta-semialdehyde.** *Biochemical Journal* 1971, **125**(3):743-&.
58. Weber G, Schorgendorfer K, Schneiderscherzer E, Leitner E: **The peptide synthetase catalyzing Cyclosporin production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame.** *Current Genetics* 1994, **26**(2):120-125.
59. Aharonowitz Y, Cohen G, Martin JF: **Penicillin and Cephalosporin biosynthetic genes - Structure, organization, regulation, and evolution.** *Annual Review of Microbiology* 1992, **46**:461-495.
60. Brakhage AA, Al-Abdallah Q, Tuncher A, Sprote P: **Evolution of beta-lactam biosynthesis genes and recruitment of trans-acting factors.** *Phytochemistry* 2005, **66**(11):1200-1210.
61. Liras P, Martin JF: **Gene clusters for beta-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate?** *International Microbiology* 2006, **9**(1):9-19.

62. Buades C, Moya A: **Phylogenetic analysis of the Isopenicillin-N-synthetase horizontal gene transfer**. *Journal of Molecular Evolution* 1996, **42**(5):537-542.
63. Landan G, Cohen G, Aharonowitz Y, Shuali Y, Graur D, Shiffman D: **Evolution of Isopenicillin-N synthase genes may have involved horizontal gene-transfer**. *Molecular Biology and Evolution* 1990, **7**(5):399-406.
64. Penalva MA, Moya A, Dopazo J, Ramon D: **Sequences of Isopenicillin-N synthetase genes suggest horizontal gene transfer from prokaryotes to eukaryotes**. *Proceedings of the Royal Society of London Series B-Biological Sciences* 1990, **241**(1302):164-169.
65. Bushley KE, Ripoll DR, Turgeon BG: **Module evolution and substrate specificity of fungal nonribosomal peptide synthetases involved in siderophore biosynthesis**. *BMC Evolutionary Biology* 2008, **8**.
66. von Dohren H: **Biochemistry and general genetics of nonribosomal peptide synthetases in fungi**. In: *Molecular Biotechnology of Fungal Beta-Lactam Antibiotics and Related Peptide Synthetases*. vol. 88; 2004: 217-264.
67. Selker EU: **Genome defense and DNA methylation in *Neurospora***. *Cold Spring Harbor Symposia on Quantitative Biology* 2004, **69**:119-124.
68. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S *et al*: **The genome sequence of the filamentous fungus *Neurospora crassa***. *Nature* 2003, **422**(6934):859-868.
69. Velasco AM, Leguina JI, Lazcano A: **Molecular evolution of the lysine biosynthetic pathways**. *Journal of Molecular Evolution* 2002, **55**(4):445-459.
70. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretailade E, Brottier P, Wincker P *et al*: **Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi***. *Nature* 2001, **414**(6862):450-453.
71. Bohnert HU, Fudal I, Dioh W, Tharreau D, Nottoghem JL, Lebrun MH: **A putative polyketide synthase peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice**. *Plant Cell* 2004, **16**(9):2499-2513.

72. Nishida H, Nishiyama M, Kobashi N, Kosuge T, Hoshino T, and Yamane H: **A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: A key to the evolution of amino acid biosynthesis.** *Genome Research* 1999, **9**:1175-1183.
73. Xu HY, Andi B, Qian JH, West AH, Cook PF: **The alpha-amino adipate pathway for lysine biosynthesis in fungi.** *Cell Biochemistry and Biophysics* 2006, **46**(1):43-64.
74. Zabriskie TM, Jackson MD: **Lysine biosynthesis and metabolism in fungi.** *Natural Product Reports* 2000, **17**(1):85-97.
75. Kwang-Deuk A, Nishida H, Yoshiharu M, Yokota A: **Amino adipate reductase gene: A new fungal-specific gene for comparative evolutionary analyses.** *BMC Evolutionary Biology* 2002, **2**(6).
76. Sims JW, Schmidt EW: **Thioesterase-like role for fungal PKS-NRPS hybrid reductive domains.** *Journal of the American Chemical Society* 2008, **130**(33):11149-11155.
77. Sims JW, Fillmore JP, Warner DD, Schmidt EW: **Equisetin biosynthesis in *Fusarium heterosporum*.** *Chemical Communications* 2005(2):186-188.
78. Song ZS, Cox RJ, Lazarus CM, Simpson TJ: **Fusarin C biosynthesis in *Fusarium moniliforme* and *Fusarium venenatum*.** *Chembiochem* 2004, **5**(9):1196-1203.
79. Chattopadhyay D, Finzel BC, Munson SH, Evans DB, Sharma SK, Strakalaitis NA, Brunner DP, Eckenrode FM, Dauter Z, Betzel C *et al*: **Crystallographic analyses of an active HIV-1 ribonuclease H domain show structural features that distinguish it from the inactive form.** *Acta Crystallographica Section D-Biological Crystallography* 1993, **49**:423-427.
80. Rice P, Craigie R, Davies DR: **Retroviral integrases and their cousins.** *Current Opinion in Structural Biology* 1996, **6**(1):76-83.
81. Declercq E, Billiau A, Ottenheijm HCJ, Herscheid JDM: **Anti-reverse transcriptase activity of Gliotoxin analogs.** *Biochemical Pharmacology* 1978, **27**(5):635-639.

82. Rouxel T, Chupeau Y, Fritz R, Kollmann A, Bousquet JF: **Biological effects of Sirodesmin-PL, a phytotoxin produced by *Leptosphaeria maculans*.** *Plant Science* 1988, **57**(1):45-53.
83. Rouxel T, Kollmann A, Bousquet JF: **Zinc suppresses Sirodesmin PL toxicity and protects *Brassica napus* plants against the blackleg disease caused by *Leptosphaeria maculans*.** *Plant Science* 1990, **68**(1):77-86.
84. Hoffmeister D, Keller NP: **Natural products of filamentous fungi: enzymes, genes, and their regulation.** *Natural Product Reports* 2007, **24**(2):393-416.
85. Wiest A, Grzegorski D, Xu BW, Goulard C, Rebuffat S, Ebbole DJ, Bodo B, Kenerley C: **Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase.** *Journal of Biological Chemistry* 2002, **277**(23):20862-20868.
86. Tanaka A, Tapper BA, Popay A, Parker EJ, Scott B: **A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory.** *Molecular Microbiology* 2005, **57**(4):1036-1050.
87. Spatafora JW, Sung GH, Sung JM, Hywel-Jones NL, White JF: **Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes.** *Molecular Ecology* 2007, **16**(8):1701-1711.
88. Clay K, Cheplick GP: **Effect of ergot alkaloids from fungal endophyte-infected grasses on fall armyworm (*Spodoptera frugiperda*).** *Journal of Chemical Ecology* 1989, **15**(1):169-182.
89. Fiserova A, Pospisil M.: Role of ergot alkaloids in the **immune system**. In: *The Genus *Claviceps**. Edited by Kren V. CL. Amsterdam: Harwood; 1999: 451-467.
90. Panaccione DG, Cipoletti JR, Sedlock AB, Blemings KP, Schardl CL, Machado C, Seidel GE: **Effects of ergot alkaloids on food preference and satiety in rabbits, as assessed with gene-knockout endophytes in perennial ryegrass (*Lolium perenne*).** *Journal of Agricultural and Food Chemistry* 2006, **54**(13):4582-4587.

91. Cross D: **Ergot Alkaloid Toxicity**. In: *Clavicipitalean Fungi: Evolutionary Biology, Chemistry, Biocontrol, and Cultural Impacts*. Edited by James F. White Jr. CWB, Nigel L. Hywel Jones, Joseph W. Spatafora. New York, New York: Marcel Dekker, Inc.; 2003.
92. Wei XY, Yang FQ, Straney DC: **Multiple non-ribosomal peptide synthetase genes determine peptaibol synthesis in *Trichoderma virens***. *Canadian Journal of Microbiology* 2005, **51**(5):423-429.
93. Hughes AL, Nei M: **Evolution of the major histocompatibility complex-independent origin of nonclassical class-I genes in different groups of mammals**. *Molecular Biology and Evolution* 1989, **6**(6):559-579.
94. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(15):7799-7806.
95. Ota T, Nei M: **Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V-H gene family**. *Molecular Biology and Evolution* 1994, **11**(3):469-482.
96. Nei M, Rooney AP: **Concerted and birth-and-death evolution of multigene families**. *Annual Review of Genetics* 2005, **39**:121-152.
97. Nei M: **The new mutation theory of phenotypic evolution**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(30):12235-12242.
98. Nei M, Niimura Y, Nozawa M: **The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity**. *Nature Reviews Genetics* 2008, **9**(12):951-963.
99. Korbel JO, Kim PM, Chen X, Urban AE, Weissman S, Snyder M, Gerstein MB: **The current excitement about copy-number variation: how it relates to gene duplications and protein families**. *Current Opinion in Structural Biology* 2008, **18**(3):366-374.

100. Lautru S, Challis GL: **Substrate recognition by nonribosomal peptide synthetase multi-enzymes.** *Microbiology (Reading)* 2004, **150**(Part 6):1629-1636.
101. Stachelhaus T, Mootz HD, Marahiel MA: **The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.** *Chemistry and Biology (London)* 1999, **6**(8):493-505.
102. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research* 2004, **32**(5):1792-1797.
103. Birney E, Clamp M, Durbin R: **Genewise and genomewise.** *Genome Research* 2004, **14**:988-995.
104. Karolewicz A, Geisen R: **Cloning a part of the Ochratoxin A biosynthetic gene cluster of *Penicillium nordicum* and characterization of the ochratoxin polyketide synthase gene.** *Systematic and Applied Microbiology* 2005, **28**(7):588-595.
105. Rausch C, Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H.: **Specificity predictions of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs).** *Nucleic Acids Research* 2005, **33**(18):5799-5808.
106. Abascal F, Zardoya, R., Posada, D.: **ProtTest: Selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**(9):2104-2105.
107. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO: **Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified.** *Bmc Evolutionary Biology* 2006, **6**.
108. Stamatakis A, Hoover P, Rougemont J: **A rapid bootstrap algorithm for the RAxML web-servers.** *Systematic Biology* 2008, **75**(5):758-771.
109. Guindon S GO: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic Biology* 2003, **52**(5):696-704.

110. Felsenstein J: **PHYLIP (Phylogeny Inference Package) version 3.6 Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. 2005.**
111. Schmidt HA, Strimmer K, Vingron M, Haeseler Av: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
112. Bremner B, Jansen R, Oxelman B, Backlund M, Lantz H, KJ K: **More characters or more taxa for a robust phylogeny: case study from the coffee family (Rubiaceae).** *Systematic Biology* 1999, **48**:413-435.
113. Mitchell A, Mitter C, Regier JC: **More taxa or more characters revisited: Combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta : Lepidoptera).** *Systematic Biology* 2000, **49**(2):202-224.
114. Sanderson MJ, Wojciechowski MF: **Improved bootstrap confidence limits in large-scale phylogenies, with an example from *Neo-Astragalus* (leguminosae).** *Systematic Biology* 2000, **49**(4):671-685.
115. Auwerx J, Baulieu E, Beato M, Becker-Andre M, Burbach PH, Camerino G, Chambon P, Cooney A, Dejean A, Dreyer C *et al*: **A unified nomenclature system for the nuclear receptor superfamily.** *Cell* 1999, **97**(2):161-163.
116. Hodge T, Cope M: **A myosin family tree.** *Journal of Cell Science* 2000, **113**(19):3353-+.
117. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N: **Estimating the tempo and mode of gene family evolution from comparative genomic data.** *Genome Research* 2005, **15**(8):1153-1160.
118. De Bie T, Cristianini N, Demuth JP, Hahn MW: **CAFE: a computational tool for the study of gene family evolution.** *Bioinformatics* 2006, **22**(10):1269-1271.
119. Sanderson MJ: **r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock.** *Bioinformatics* 2003, **19**(2):301-302.

120. Fitzpatrick DA, Logue, Mary E., Stajich, Jason E., and Butler, Geraldine: **A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis.** *BMC Evolutionary Biology* 2006, **6**:99.
121. Chen XH, Koumoutsi A, Scholz R, Eisenreich A, Schneider K, Heinemeyer I, Morgenstern B, Voss B, Hess WR, Reva O *et al*: **Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42.** *Nature Biotechnology* 2007, **25**(9):1007-1014.
122. May JJ, Wendrich TM, Marahiel MA: **The Dhb operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester Bacillibactin.** *Journal of Biological Chemistry* 2001, **276**(10):7209-7217.
123. Fuma S, Fujishima Y, Corbell N, Dsouza C, Nakano MM, Zuber P, Yamane K: **Nucleotide-sequence of 5' portion of SrfA that contains the region required for competence establishment in *Bacillus subtilis*.** *Nucleic Acids Research* 1993, **21**(1):93-97.
124. May JJ, Kessler N, Marahiel MA, Stubbs MT: **Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(19):12120-12125.
125. Saito F, Hori K, Kanda M, Kurotsu T, Saito Y: **Entire nucleotide-sequence for *Bacillus brevis* Nagano Grs2 gene encoding Gramicidin-S synthetase-2 - a multifunctional peptide synthetase.** *Journal of Biochemistry* 1994, **116**(2):357-367.
126. Hori K, Yamamoto Y, Minetoki T, Kurotsu T, Kanda M, Miura S, Okamura K, Furuyama J, Saito Y: **Molecular cloning and nucleotide-sequence of the Gramicidin-S synthetase 1 gene.** *Journal of Biochemistry* 1989, **106**(4):639-645.
127. Mootz HD, Marahiel MA: **The Tyrocidine biosynthesis operon of *Bacillus brevis*: Complete nucleotide sequence and biochemical characterization of functional internal adenylation domains.** *Journal of Bacteriology* 1997, **179**(21):6843-6850.

128. Wu XF, Ballard J, Jiang YW: **Structure and biosynthesis of the BT peptide antibiotic from *Brevibacillus texasporus***. *Applied and Environmental Microbiology* 2005, **71**(12):8519-8530.
129. Rusnak F, Sakaitani M, Drucekhammer D, Reichert J, Walsh CT: **Biosynthesis of the *Escherichia coli* siderophore Enterobactin - Sequence of the EntF gene, expression and purification of EntF, and analysis of covalent phosphopantetheine**. *Biochemistry* 1991, **30**(11):2916-2927.
130. Kimura H, Miyashita H, Sumino Y: **Organization and expression in *Pseudomonas putida* of the gene cluster involved in Cephalosporin biosynthesis from *Lysobacter lactamgenus* YK90**. *Applied Microbiology and Biotechnology* 1996, **45**(4):490-501.
131. Weinig S, Hecht HJ, Mahmud T, Muller R: **Melithiazol biosynthesis: Further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases**. *Chemistry & Biology* 2003, **10**(10):939-952.
132. Tillett D, Dittmann E, Erhard M, von Dohren H, Borner T, Neilan BA: **Structural organization of Microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system**. *Chemistry & Biology* 2000, **7**(10):753-764.
133. Nishizawa A, Bin Arshad A, Nishizawa T, Asayama M, Fujii K, Nakano T, Harada K, Shirai M: **Cloning and characterization of a new hetero-gene cluster of nonribosomal peptide synthetase and polyketide synthase from the cyanobacterium *Microcystis aeruginosa* K-139**. *Journal of General and Applied Microbiology* 2007, **53**(1):17-27.
134. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE *et al*: **Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence (vol 393, pg 537, 1998)**. *Nature* 1998, **396**(6707):190-198.
135. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Ronning CM, Barbazuk WB, Blanchard M, Field C *et al*: **Evolution of sensory complexity recorded in a myxobacterial genome**. *Proc Natl Acad Sci USA* 2006, **103**(41):15200-15205

136. Coque JJR, Martin JF, Calzada JG, Liras P: **The Cephamycin biosynthetic genes PcbAB, encoding a large multidomain peptide synthetase, and PcbC of *Nocardia lactamdurans* are clustered together in an organization different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*.** *Molecular Microbiology* 1991, **5**(5):1125-1133.
137. Luesch H, Hoffmann D, Hevel JM, Becker JE, Golakoti T, Moore RE: **Biosynthesis of 4-Methylproline in cyanobacteria: Cloning of NosE and NosF genes and biochemical characterization of the encoded dehydrogenase and reductase activities.** *Journal of Organic Chemistry* 2003, **68**(1):83-91.
138. Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S, Bocs S, Boursaux-Eude C, Chandler M, Charles JF *et al*: **The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*.** *Nature Biotechnology* 2003, **21**(11):1307-1313.
139. Quadri LEN, Keating TA, Patel HM, Walsh CT: **Assembly of the *Pseudomonas aeruginosa* nonribosomal peptide siderophore Pyochelin: In vitro reconstitution of aryl-4,2-bisthiazoline synthetase activity from PchD, PchE, and PchF.** *Biochemistry* 1999, **38**(45):14941-14954.
140. Smith EE, Sims EH, Spencer DH, Kaul R, Olson MV: **Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*.** *Journal of Bacteriology* 2005, **187**(6):2138-2147.
141. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FSL, Hufnagle WO, Kowalik DJ, Lagrou M *et al*: **Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.** *Nature* 2000, **406**(6799):959-964.
142. Vodovar N, Vallenet D, Cruveiller S, Rouy Z, Barbe V, Acosta C, Cattolico L, Jubin C, Lajus A, Segurens B *et al*: **Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*.** *Nat Biotechnol* 2006, **24**(6):673-679.
143. Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA, Mavrodi DV, DeBoy RT, Seshadri R, Ren QH, Madupu R *et al*: **Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5.** *Nature Biotechnology* 2005, **23**(7):873-878.

144. Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, dos Santos V, Fouts DE, Gill SR, Pop M, Holmes M *et al*: **Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440**. *Environmental Microbiology* 2002, **4**(12):799-808.
145. Silakowski B, Schairer HU, Ehret H, Kunze B, Weinig S, Nordsiek G, Brandt P, Blocker H, Hofle G, Beyer S *et al*: **New lessons of combinatorial biosynthesis from myxobacteria - The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1**. *Journal of Biological Chemistry* 1999, **274**(52):37391-37399.
146. Yu H, Serpe E, Romero J, Coque JJ, Maeda K, Oelgeschlager M, Hintermann G, Liras P, Martin JF, Demain AL *et al*: **Possible involvement of the lysine epsilon-aminotransferase gene (Lat) in the expression of the genes encoding ACV synthetase (PcbAB) and Isopenicillin-N synthase (PcbC) in *Streptomyces clavuligerus***. *Microbiology-Uk* 1994, **140**:3367-3377.
147. Gehring AM, DeMoll E, Fetherston JD, Mori I, Mayhew GF, Blattner FR, Walsh CT, Perry RD: **Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis***. *Chemistry & Biology* 1998, **5**(10):573-586.
148. Gehring AM, Mori I, Perry RD, Walsh CT: **The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of Yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis* (vol 37, pg 11637, 1998)**. *Biochemistry* 1998, **37**(48):17104-17104.
149. Taylor JW, Berbee ML: **Dating divergences in the fungal tree of life: review and new analyses**. *Mycologia* 2006, **98**(6):838-849.
150. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S *et al*: **The genome sequence of *Schizosaccharomyces pombe* (vol 415, pg 871, 2002)**. *Nature* 2003, **421**(6918):94-94.
151. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi SD *et al*: **The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome**. *Science* 2004, **304**(5668):304-307.

152. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT *et al*: **The diploid genome sequence of *Candida albicans***. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(19):7329-7334.
153. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, de Montigny J, Marck C, Neuveglise C, Talla E *et al*: **Genome evolution in yeasts**. *Nature* 2004, **430**(6995):35-44.
154. Souciet JL, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, de Montigny J, Dujon B *et al*: **Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies**. *Febs Letters* 2000, **487**(1):3-12.
155. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M *et al*: **Life with 6000 genes**. *Science* 1996, **274**(5287):546-547.
156. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements**. *Nature* 2003, **423**(6937):241-254.
157. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J *et al*: **Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae***. *Nature* 2005, **438**(7071):1105-1115.
158. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C *et al*: **Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*** (vol **438**, pg **1151**, **2005**). *Nature* 2006, **439**(7075):502-502.
159. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Jin-Rong X, Huaqin P *et al*: **The genome sequence of the rice blast fungus *Magnaporthe grisea***. *Nature* 2005, **434**(7036):980(987).
160. Cuomo CA, Gueldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M *et al*: **The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization**. *Science* 2007, **317**(5843):1400-1402.

161. Paoletti M, Saupe SJ: **The genome sequence of *Podospora anserina*, a classic model fungus.** *Genome Biology* 2008, **9**(5).
162. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D *et al*: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*) (vol 26, pg 553, 2008).** *Nature Biotechnology* 2008, **26**(10):1193-1193.
163. Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H *et al*: **Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*.** *Nature Biotechnology* 2007, **25**(3):319-326.
164. Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA *et al*: **The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*.** *Science* 2005, **307**(5713):1321-1324.
165. Martinez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F *et al*: **Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78 (vol 22, pg 695, 2004).** *Nature Biotechnology* 2004, **22**(7):899-899.
166. Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al*: **The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis.** *Nature* 2008, **452**(7183):88-U87.
167. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O *et al*: **Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*.** *Nature* 2006, **444**(7115):97-101.
168. Page RDM: **TreeView: An application to display phylogenetic trees on personal computers.** *Computer Applications in the Biosciences* 1996, **12**(4):357-358.

169. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ: **Multiple sequence alignment with Clustal X**. *Trends in Biochemical Sciences* 1998, **23**(10):403-405.