



Phylogenomic analysis of non-ribosomal peptide synthetases in the genus *Aspergillus*

Robert A. Cramer Jr.^{a,*}, Jason S. Stajich^a, Yvonne Yamanaka^a, Fred E. Dietrich^a, William J. Steinbach^{a,b,1}, John R. Perfect^{a,c,1}

^a Duke University Medical Center, Department of Molecular Genetics and Microbiology, Durham, NC 27710, USA

^b Duke University Medical Center, Department of Pediatrics, Durham, NC 27710, USA

^c Duke University Medical Center, Department of Medicine, Durham, NC 27710, USA

Received 22 March 2006; received in revised form 29 June 2006; accepted 10 July 2006

Received by I.B. Rogozin

Abstract

Fungi from the genus *Aspergillus* are important saprophytes and opportunistic human fungal pathogens that contribute in these and other diverse ways to human well-being. Part of their impact on human well-being stems from the production of small molecular weight secondary metabolites, which may contribute to the ability of these fungi to cause invasive fungal infections and allergic diseases. In this study, we identified one group of enzymes responsible for secondary metabolite production in five *Aspergillus* species, the non-ribosomal peptide synthetases (NRPS). Hidden Markov models were used to search the genome databases of *A. fumigatus*, *A. flavus*, *A. terreus*, *A. nidulans*, and *A. oryzae* for domains conserved in NRPS proteins. A genealogy of adenylation domains was utilized to identify orthologous and unique NRPS among the *Aspergillus* species examined, as well as gain an understanding of the potential evolution of *Aspergillus* NRPS. mRNA abundance of the 14 NRPS identified in the *A. fumigatus* genome was analyzed using real-time reverse transcriptase PCR in different environmental conditions to gain a preliminary understanding of the possible functions of the NRPSs' peptide products. Our results suggest that *Aspergillus* species contain conserved and unique NRPS genes with a complex evolutionary history. This result suggests that the genus *Aspergillus* produces a substantial diversity of non-ribosomally synthesized peptides. Further analysis of these genes and their peptide products may identify important roles for secondary metabolites produced by NRPS in *Aspergillus* physiology, ecology, and fungal pathogenicity.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Fungal secondary metabolites; Invasive aspergillosis; Adenylation domain

1. Introduction

Fungi play significant roles in the lives of human beings providing humankind with many beneficial products, including

life-saving antibiotics, food products such as cheese and wine, and themselves as food. Many of these beneficial products are important pharmaceuticals including the immunosuppressant cyclosporine and the antibiotic penicillin. These natural products are most often classified as secondary metabolites (Keller et al., 2005). Secondary metabolites are distinguished from primary metabolites since they are not typically required for *in vitro* growth. Since many fungal secondary metabolites have antimicrobial activity, it has been hypothesized that they serve as a line of defense against other microorganisms in the environment.

One class of fungal secondary metabolites is the non-ribosomal peptides (NRPs), which are derived from proteinogenic and non-proteinogenic amino acids. NRPs are synthesized from large multidomain enzymes, the non-ribosomal

Abbreviations: NRPS, non-ribosomal peptide synthetase; NRPs, non-ribosomal peptides; A domain, adenylation domain; T domain, peptidyl carrier domain; C domain, condensation domain; TE, thioesterase domain; HMM, hidden Markov model; NJ, neighbor-joining; DMEM, Dulbecco's modified Eagle's medium; NRT, no reverse transcriptase; PKS, polyketide synthetase; BLAST, basic local alignment search tool; Pfam, Protein family.

* Corresponding author: Duke University Medical Center Box 3353, Durham, NC 27710, United States. Tel.: +1 919 681 2613; fax: +1 919 684 8902.

E-mail address: rrcramer@duke.edu (R.A. Cramer).

¹ WJS and JRP contributed equally to this manuscript.

peptide synthetases (NRPS) (Marahiel, 1992; Weber and Marahiel, 2001). NRPS contain modules that incorporate amino acids into the final peptide product. Each module contains a minimum of three domains, the adenylation domain (A domain) which is responsible for substrate specificity and activation, the peptidyl carrier domain (T domain) that covalently binds the substrate to the enzyme via a thioester linkage, and the condensation domain (C domain) that catalyzes peptide bond formation between an aminoacyl bound and peptidyl bound intermediate. Often with bacterial NRPS, the final peptide product is released from the NRPS template by a thioesterase domain (TE) found at the carboxy terminus of the NRPS. However, to date, very few TE domains have been identified in fungal NRPS.

Fungal NRPs have been found to have a wide range of biological activities. Some, like HC-toxin from *Cochliobolus carbonum*, are well known virulence factors in fungal–plant interactions that cause substantial crop damage, yield losses, and food supply contamination (Walton, 1987; Bennett, 1989; Panaccione et al., 1992; Osbourn, 2001). A few, like gliotoxin from *Aspergillus fumigatus*, are known to have deleterious effects on the immune system of humans (Gardiner et al., 2004; Szekeres et al., 2005). Yet, the role of NRPs and other fungal secondary metabolites in fungal–human interactions has not been extensively explored.

Recent evidence suggests that secondary metabolites produced by *A. fumigatus* may be important in the development of aspergillosis (Bok et al., 2005; Cramer et al., 2006). However, further studies are needed to address the hypothesis that NRPs and other *Aspergillus* secondary metabolites are part of the virulence composite for this opportunistic human fungal pathogen. In this study, we utilized a phylogenomics and functional genomics approach to identify NRPS genes in five *Aspergillus* species with whole genome sequences, *A. fumigatus*, *A. nidulans*, *A. flavus*, *A. oryzae*, and *A. terreus*. We focus on NRPS from *A. fumigatus* which causes the majority of invasive aspergillosis cases (Latge, 1999; Perfect et al., 2001).

2. Materials and methods

2.1. Non-ribosomal peptide synthetase identification

A multi-faceted bioinformatics approach was used to identify NRPS genes in the genomes of *A. fumigatus*, *A. nidulans*, *A. flavus*, *A. oryzae* and *A. terreus*. Profile hidden Markov Models (HMMs) of the 3 main NRPS domains (A, T, and C) were used to search protein sequence databases of the available genomes with HMMER (Sean Eddy, <http://hmmerr.wustl.edu/>) (Eddy, 1996). Proteins with at least 1 A, T, and C domain were annotated as NRPS. Putative NRPS amino acid sequences obtained from the HMMER search were used as queries in BLASTP searches with default parameters of the *Aspergillus* species genomes to obtain all possible NRPS-like sequences (*E* value cutoff 1e–10). Finally, using the putative NRPS A domain sequences from these analyses (a total of 190), hmmbuild and hmmcalibrate were used to create an HMM profile for the *Aspergillus* A domain amino acid sequences.

This profile was then used to search the *Aspergillus* proteomes with hmmsearch using an expected value cutoff of 1e–05 (Table 1).

The protein sequence databases used for each *Aspergillus* species are available at: *A. fumigatus* (<http://www.cadre.man.ac.uk/>), *A. nidulans* (<http://www.broad.mit.edu/annotation/fungi/aspergillus/>), *A. flavus* (<http://www.aspergillusflavus.org/genomics/>), *A. oryzae* (<http://www.bio.nite.go.jp/ngac/e/rib40-e.html>), and *A. terreus* (<http://fungal.genome.duke.edu/>). The unannotated *A. terreus* genome sequence was obtained from the Fungal Genome Initiative at the Broad Institute (http://www.broad.mit.edu/annotation/fungi/aspergillus_terreus/). Automated protein coding gene annotation of *A. terreus* was performed using a combination of *ab initio* (SNAP, Genezilla, AUGUSTUS), and protein alignment based gene prediction methods (Exonerate), and combined into a single gene prediction set using the tool GLEAN (Mackey et al. personal communication) (Korf, 2004; Majoros et al., 2004; Stanke and Waack, 2003; Slater and Birney, 2005).

Domain and module organization of predicted NRPS amino acid sequences was annotated using HMMER results from a search using NRPS amino acid sequences as queries against a locally installed Pfam database (*E* value cutoff 1e–03) (Finn et al., 2006). In addition, NRPS amino acid sequences were used as queries against the database of non-ribosomal peptide synthetases (<http://203.90.127.50/~zeeshan/webpages/nrpsall.html>) (Ansari et al., 2004). Discrepancies between the two prediction algorithms were resolved by manual inspection of the amino acid sequences of the disputed domain. When modules appeared to be missing specific domains, further manual annotation was performed to identify domains potentially missed by the prediction algorithms.

2.2. Phylogenomic and sequence analysis of *Aspergillus* NRPS proteins

A multiple alignment of *Aspergillus* A domains plus additional A domains from known fungal NRPS proteins acquired from GenBank was created using CLUSTALX and the BLOSUM62 scoring matrix (Thompson et al., 1997). A domains were defined with HMMER and their amino acid sequences extracted with a Perl script utilizing BioPerl (Stajich et al., 2002). The multiple alignment was edited manually to include the 6 known A domain motifs found in fungal NRPS A domains, this alignment was used to create unrooted neighbor-joining (NJ) and parsimonious trees using PFAAT and PHYLIP (<http://evolution.genetics.washington.edu/phytip.html>) packages respectively (Johnson et al., 2003). Due to the large size of the alignment, a portion of the alignment containing adenylation domains from putative NRPS predicted to make ETP toxins is presented (Fig. 3). Columns in the multiple alignment with >50% gaps were excluded from the NJ tree. Bootstrapping with 100 replications was performed with the unrooted NJ tree.

The NJ tree was virtually identical to the tree created with parsimony analysis. Due to the large size of the genealogy, the most well resolved portion of the tree is presented (Fig. 4). The full tree can be viewed online as Supplemental Fig. 1.

Table 1
Putative and known non-ribosomal peptide synthetases identified in this study

NRPS name in this study	Other names ^a	Database ID ^b	Predicted product/function
<i>Aspergillus fumigatus</i>			
NRPS1		Afu1g10380	
NRPS2	<i>sidC</i>	Afu1g17200	Ferrichrome siderophore
NRPS3	<i>sidE</i>	Afu3g03350	Putative unknown siderophore ^c
NRPS4	<i>nps6</i> , <i>sidD</i>	Afu3g03420	Oxidative stress/siderophore ^c
NRPS5		Afu3g12920	Putative unknown ETP toxin ^c
NRPS6		Afu3g13730	
NRPS7		Afu3g15270	Putative unknown siderophore
NRPS8		Afu5g12730	
NRPS9		Afu6g09610	
NRPS10	<i>gliP</i>	Afu6g09660	Gliotoxin biosynthesis
NRPS11		Afu6g12050	
NRPS12		Afu6g12080	
NRPS13	<i>ftmPS</i>	Afu8g00170	Fumitremorgin B or derivatives
NRPS14		Afu8g00540	PKS/NRPS hybrid ^c
<i>Aspergillus nidulans</i>			
NRPS15		AN0016.2	
NRPS16	<i>sidC</i>	AN0607.2	Ferricrocin siderophore
NRPS17		AN1242.2	
NRPS18		AN2545.2	
NRPS19	<i>acvA</i>	AN2621.2	Penicillin biosynthesis
NRPS20		AN3495.2	
NRPS21		AN3496.2	
NRPS22		AN4641.2	
NRPS23		AN6236.2	Oxidative stress/siderophore ^c
NRPS24		AN7884.2	
NRPS25		AN8412.2	PKS/NRPS hybrid ^c
NRPS26		AN8504.2	
NRPS27		AN9226.2	Putative <i>N</i> -methylated product ^c
NRPS28		AN9244.2	
<i>Aspergillus oryzae</i>			
NRPS29		AO090001000009	
NRPS30		AO090001000043	
NRPS31		AO090001000262	
NRPS32		AO090001000277	PKS/NRPS hybrid ^c
NRPS33		AO090005000993	
NRPS34		AO090011000043	
NRPS35		AO090020000380	Putative <i>N</i> -methylated product ^c
NRPS36	<i>sid2</i>	AO090023000528	Ferrichrome siderophore
NRPS37		AO090026000378	
NRPS38		AO090038000390	
NRPS39		AO090038000543	Penicillin biosynthesis ^c
NRPS40		AO090102000338	
NRPS41		AO090102000465	
NRPS42		AO090103000167	Oxidative stress/siderophore ^c
NRPS43		AO090103000223	
NRPS44		AO090103000224	PKS/NRPS hybrid ^c
NRPS45		AO090103000355	
NRPS46		AO090120000024	
<i>Aspergillus flavus</i>			
NRPS47		39.m05322	
NRPS48		40.m03083	
NRPS49		77.m03839	
NRPS50		77.m04289	

Table 1 (continued)

NRPS name in this study	Other names ^a	Database ID ^b	Predicted product/function
<i>Aspergillus flavus</i>			
NRPS51		77.m04350	PKS/NRPS hybrid ^c
NRPS52		80.m03224	Penicillin biosynthesis ^c
NRPS53		80.m03392	
NRPS54		80.m03659	PKS/NRPS hybrid ^c
NRPS55		80.m03675	
NRPS56		80.m03907	
NRPS57		92.m03107	
NRPS58		93.m03019	
NRPS59		93.m03587	
NRPS60		93.m03617	
NRPS61		92.m03802	
NRPS62		103.m02081	
NRPS63		103.m02216	
NRPS64		103.m02217	
NRPS65		103.m02275	Putative <i>N</i> -methylated product ^c
NRPS66		103.m02282	Oxidative stress/siderophore ^c
<i>Aspergillus terreus</i>			
NRPS67		ater_glean07417	
NRPS68		ater_glean03455	
NRPS69		ater_glean10081	
NRPS70		ater_glean02592	
NRPS71		ater_glean04710	Ferrichrome siderophore ^c
NRPS72		ater_glean10041	
NRPS73		ater_glean10362	
NRPS74		ater_glean06954	Putative unknown ETP toxin ^c
NRPS75		ater_glean11197	
NRPS76		ater_glean07040	PKS/NRPS hybrid ^c
NRPS77		ater_glean07366	PKS/NRPS hybrid ^c
NRPS78		ater_glean05849	
NRPS79		ater_glean09472	Oxidative stress/siderophore ^c
NRPS80		ater_glean09076	
NRPS81		ater_glean10104	
NRPS82		ater_glean10674	Putative unknown siderophore ^c
NRPS83		ater_glean06968	Putative unknown siderophore ^c
NRPS84		ater_glean10403	
NRPS85		ater_glean05764	
NRPS86		ater_glean07057	
NRPS87		ater_glean10082	
NRPS88		ater_glean10800	
<i>Other NRPS^d</i>			
<i>Tolypocladium inflatum</i>	<i>simA</i>	CAA82227	Cyclosporine biosynthesis
<i>Fusarium equiseti</i>	<i>esyn1</i>	CAA79245	Enniatin biosynthesis
<i>Leptosphaeria maculans</i>	<i>sirP</i>	AAS92545	Sirodesmin biosynthesis
<i>Penicillium chrysogenum</i>	<i>pcbaB</i>	P19787	Penicillin biosynthesis
<i>Ustilago maydis</i>	<i>sid2</i>	AAB93493	Ferrichrome siderophore
<i>Claviceps purpurea</i>	<i>ps2</i>	CAD28788	D-lysergyl peptide synthetase
<i>Claviceps purpurea</i>	<i>ps1</i>	CAB39315	D-lysergyl peptide synthetase
<i>Cochliobolus heterostrophus</i>	<i>nps6</i>	AAX09988	Oxidative stress/unknown siderophore

^a Other gene names published for respective NRPS.

^b Accession numbers from genome databases where sequences were identified.

^c Product/function predicted based on phylogenomic analysis in this study.

^d GenBank accession numbers listed.

2.3. mRNA abundance of *A. fumigatus* NRPS genes

NRPS gene transcription levels were measured in *A. fumigatus* strain AF293 grown under five different environmental conditions. In the first condition, ungerminated conidia were harvested from glucose minimal media (1% (w/v) glucose, 5% (v/v) 20× nitrate salts, 0.05% (v/v) trace elements) plates grown for four days at 37 °C. In the second condition, 5:1 solutions of spores:J774A.1 cells (American Type Culture Collection TIB-67) (1.25×10^6 spores/well, 2.5×10^5 macrophages/well) were incubated at 37 °C for eight hours under a 5% CO₂ atmosphere in Dulbecco's modified Eagle's media (DMEM) supplemented with 10% heat killed fetal calf serum (Gibco, Invitrogen Corporation, Carlsbad, CA). In the last three conditions spores were grown in liquid media and harvested as mycelia. Flasks with 50 mL of Czapek's Dox broth (Difco Laboratories, Sparks, MD), Sabouraud's dextrose broth (Difco Laboratories, Sparks, MD), or RPMI-1640 (Sigma Aldrich, St. Louis, MO) (three replicate flasks per broth) were inoculated with spores to give a final concentration of 1×10^7 spores/mL broth. After growing for 48 h in a shaking incubator at 37 °C, the mycelia in each flask were harvested. The fungal material collected from each of the five growth conditions was lyophilized overnight.

Total RNA was extracted from the lyophilized samples using the RNeasy total RNA isolation kit (Ambion, Austin, Tex.). For the liquid media conditions (Czapek's Sabouraud's, RPMI) separate RNA extractions were performed on the mycelia collected from three replicate flasks. Equal amounts (1 µg) of RNA extracted from each of the three replicate flasks were pooled. Possible genomic DNA contamination was removed by treating the RNA samples with DNaseI using the TURBO DNA-free kit (Ambion, Austin, Tex.). First-strand cDNA was created from the five conditions using 500 ng of DNaseI treated total RNA with the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA).

Real-time RT-PCR was conducted with 20 µl reaction volumes with the iQ SYBR green supermix (Biorad Laboratories, Hercules, CA), 2 µl of a 1:6 dilution of first-strand cDNA, and 0.4 µl of each 10 µM primer stock. Real-time PCR primers for the 14 predicted *A. fumigatus* NRPS genes and *A. fumigatus* actin gene were designed with Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Supplemental Table 1). Since oligo dT primer was used to prime the cDNA synthesis, and due to the large sequence length of NRPS genes, primers were all designed within 500 base pairs of the NRPS sequence's poly A+ tail.

No reverse transcriptase (NRT) controls were used to confirm elimination of contaminating genomic DNA. Real-time PCR was performed using an iQ Cyclor Real-Time PCR detection system (Biorad Laboratories, Hercules, CA). PCRs for each gene were done in triplicate and data were analyzed with the gene expression software included with the iQ Cyclor system. Melt curve analysis was performed after the PCR reaction was complete to confirm the absence of non-specific amplification products. The average cycle threshold of actin ($C_{t-actin}$) under each condition was used to normalize the average cycle threshold for each NRPS gene (C_{t-NRPS}) and calculate the relative abundance of the NRPS

transcripts. The relative abundance of each NRPS gene compared to actin was calculated as: Relative Abundance = $(2^{-\Delta C_t}) \times 100\%$, where $\Delta C_t = C_{t-actin} - C_{t-NRPS}$ (Lee et al., 2005).

3. Results

3.1. Non-ribosomal peptide synthetase identification

To identify the number of NRPS in available *Aspergillus* species genomes, bioinformatic analyses were utilized. HMMER and BLAST analyses identified multiple NRPS sequences in each *Aspergillus* genome, indicating that non-ribosomally produced peptides are abundant in the genus *Aspergillus*. All 5 *Aspergillus* species examined contained multiple NRPS genes, suggesting that *Aspergillus* species are capable of synthesizing a diverse range of non-ribosomally synthesized peptides (Table 2). *A. terreus* contained the highest number of NRPS with 22 followed by *A. flavus* with 20. *A. fumigatus*, *A. nidulans*, and *A. oryzae* contained 14, 14, and 18 NRPS, respectively as recently reported (Niernan et al., 2005). Only 3 NRPS among the 5 species were found to contain a thioesterase domain (TE), which is involved in cleaving the peptide product from the NRPS template. *A. flavus*, *A. nidulans*, and *A. oryzae* each contained 1 NRPS with a TE domain. These 3 NRPS are predicted to encode an NRPS involved in penicillin biosynthesis. All 5 species examined contained at least 1 polyketide synthetase (PKS)–NRPS hybrid protein (Silakowski et al., 2001).

3.2. *A. fumigatus* non-ribosomal peptide synthetases

The genome sequence of *A. fumigatus* contained three monomodular and ten multimodule NRPS in addition to one PKS/NRPS hybrid. Interestingly, 9 of the 14 NRPS were found on either chromosome three ($n=5$) or six ($n=4$). None of the *A. fumigatus* NRPS had predicted thioesterase domains, which function to release the synthesized peptide from the NRPS template. However, 10 of the NRPS ended with a unique C domain that may be involved in releasing the peptide from the NRPS template. This result was also recently found with NRPS from the plant pathogen *Cochliobolus heterostrophus* (Lee et al., 2005). All *A. fumigatus* NRPS had a traditional initiation module consisting of an A and T domain, and NRPS1 also

Table 2
Non-ribosomal peptide synthetases found in 5 species of *Aspergillus*

Species	Total NRPS	Monomodular	Multimodular	PKS/NRPS hybrid	Total A domains
<i>Aspergillus fumigatus</i>	14	3	10	1	31
<i>Aspergillus terreus</i>	22	5	15	2	43
<i>Aspergillus nidulans</i>	14	1	12	1	36
<i>Aspergillus flavus</i>	20	5	13	2	40
<i>Aspergillus oryzae</i>	18	4	12	2	40

NRPS1	ATE CA CAT CATE CT CT	6269
NRPS2	AT CAT CAT C TCTC	4763
NRPS3	AT CAT	2210
NRPS4	AT CAC	2083
NRPS5	AT CAT C	2306
NRPS6	ATC	1290
NRPS7	AT CAT C	2353
NRPS8	AT CATE CATCTE CAT CAT CATE C	8515
NRPS9	ATC	1135
NRPS10	AT CAT CT	2135
NRPS11	ATC	1160
NRPS12	AT CATE CAT C	3955
NRPS13	AT CAT C	2211
NRPS14	KS AT DH KR CATE N	3991

Fig. 1. Domain and module organization of *Aspergillus fumigatus* non-ribosomal peptide synthetases. Numbers on right signify number of amino acids in the NRPS protein. A = adenylation domain, T = thiolation domain, C = condensation domain, E = epimerization domain, KS = ketoacyl synthase, AT = acyl transferase, DH = dehydratase, KR = ketoreductase, N = NAD binding domain.

contained an epimerization domain immediately downstream of the initiation module (Fig. 1). NRPS8, NRPS12, and NRPS14 also had epimerization domains. NRPS8 was the largest protein containing 8515 amino acids (~25 kb gene) that encoded 1 initiation module and 5 elongation modules. This NRPS would be predicted to synthesize a peptide with 6 amino acid residues. However, the non-linear nature of modules containing more

domains than the linear arrangement of a typical elongation module (domain arrangement = CAT), may suggest repetitive or non-traditional use of these modules. The smallest NRPS was NRPS9 which contained 1135 amino acids that encoded a monomodular NRPS. Ten out of the 14 NRPS sequences contained introns with NRPS1 having the most with 6. NRPS6, NRPS8, and NRPS13 did not contain introns.

In the *A. fumigatus* genome, BLAST searches identified 6 additional sequences with similarity to NRPS. Further analysis of these sequences revealed that all 6 lacked the required C domain to be considered true NRPS. Two of these sequences, however, were found near genes considered to be involved in secondary metabolism, and thus may be involved in the synthesis of some unknown secondary metabolite. Sequence Afu4g1140 was found near a polyketide synthetase, and Afu5g10120 was located near multiple cytochrome P450 monooxygenases. The remaining 4 sequences were not found near genes typically thought to be involved in secondary metabolism.

3.3. Real-time PCR analysis of *A. fumigatus* NRPS mRNA abundance

To help gain insight into the possible functions of *A. fumigatus* NRPS genes, we profiled NRPS mRNA abundance from five unique environments. In all five environments, abundance of actin was greater than any of the NRPS (Fig. 2). NRPS4 had the highest mRNA abundance among the NRPS genes examined when *A. fumigatus* conidia were incubated with the macrophage cell line J774A.1. NRPS4, NRPS8, and

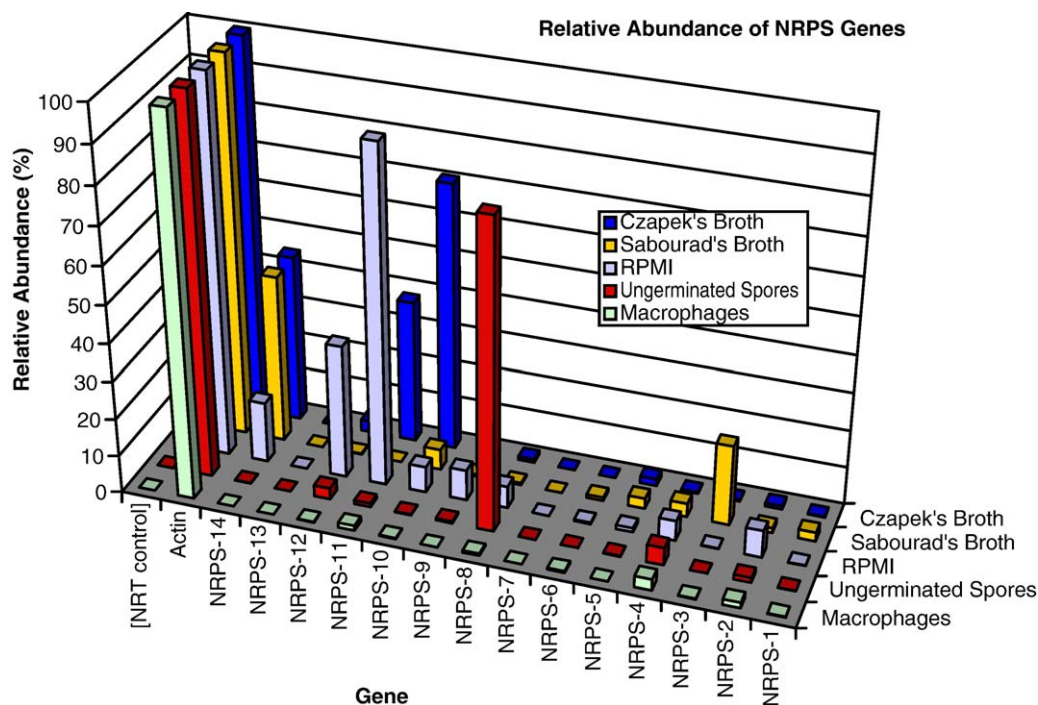


Fig. 2. mRNA abundance of *Aspergillus fumigatus* non-ribosomal peptide synthetases. mRNA abundance was examined in 5 environmental conditions: 48 h at 37 °C liquid cultures: Sabourad's broth (yellow), Czapek's Dox broth (blue), and RPMI-1640 (gray/light blue), 4 day old ungerminated spores from glucose minimal media (red), and incubation with J774A.1 macrophage cells for 8 h (light green). Relative abundance was calculated relative to changes in actin for each condition ($2^{-\Delta\Delta C_t}$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

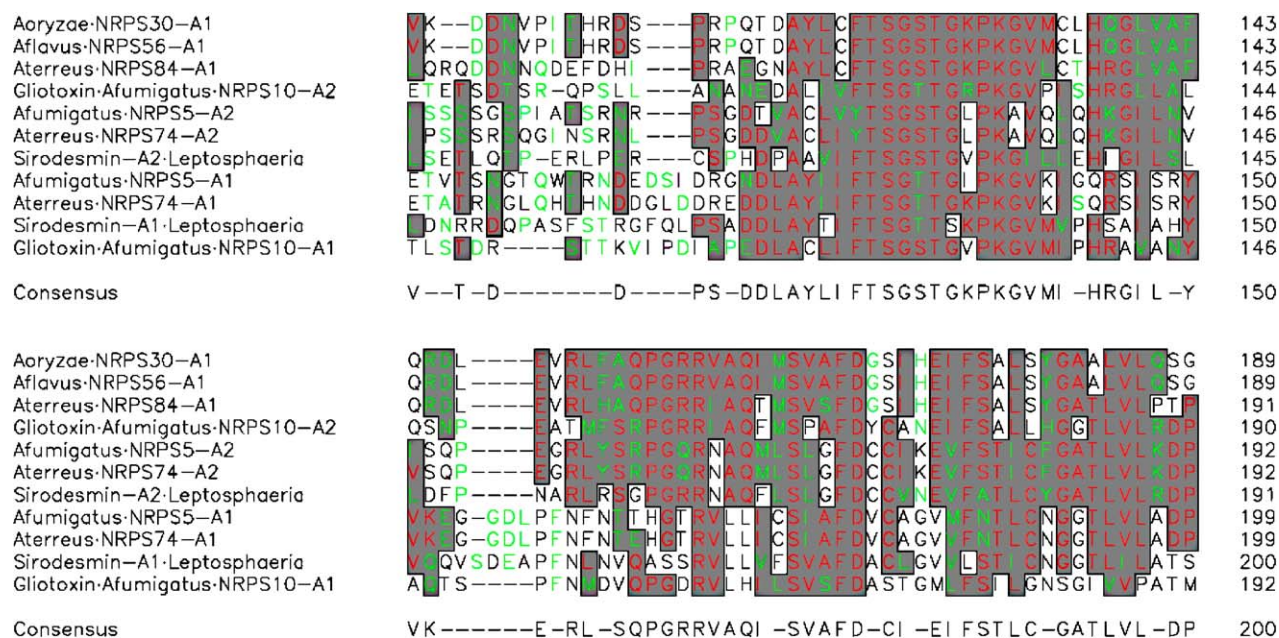


Fig. 3. Partial multiple alignment of adenylation domains created using ClustalX. Shown is the portion of the alignment containing the putative ETP toxins and positions 100–150.

NRPS12 were present in ungerminated conidia. In particular, NRPS8 was highly abundant in ungerminated conidia compared to its abundance in other conditions. NRPS11, NRPS12, and NRPS14 had the highest levels of abundance in RPMI-1640 medium. For NRPS11, RPMI-1640 medium was the condition in which its mRNA abundance was highest. NRPS14, the PKS/NRPS hybrid was highly abundant in Sabouraud's and Czapek's broth. In Czapek's broth, NRPS10, responsible for gliotoxin production (Cramer et al., 2006), was the most abundant. NRPS7 was the only NRPS that did not have significant abundance detected in any of the environments examined (relative abundances < 0.002).

3.4. Phylogenomic analysis of adenylation domains from *Aspergillus* NRPS

To identify potential NRPS orthologs in the genus *Aspergillus*, and examine potential mechanisms of NRPS evolution in this genus, a genealogy based on the most conserved domain in fungal NRPS, the A domain, was created. A total of 190 A domains from 88 *Aspergillus* NRPS were combined with 26 A domains from fungal NRPS with functionally characterized peptide products to create the genealogy (Fig. 4 and Supplemental Fig. 1). The unrooted NJ tree depicts a complex evolutionary history for A domains and their associated NRPS genes. A general trend found throughout the tree was that A domains from single NRPS genes are typically found in the same sub-clade. This allowed phylogenomic analyses to predict the putative function and the final peptide product made from these NRPS.

For example, the two A domains from the *A. fumigatus* NRPS responsible for gliotoxin production, NRPS10, were found in the same sub-clade, and all domains from the *A. nidulans* ferricrocin siderophore NRPS (NRPS16, or *sidC*), were found in a single sub-clade (Fig. 4). Based on these results,

sub-clades were identified that contained NRPS with characterized functions or products. When all A domains from a single NRPS were grouped within one of these sub-clades, the corresponding NRPS genes in other *Aspergillus* species were hypothesized to make similar products. However, only 6 of these functionally characterized sub-clades were identified in the tree.

These six sub-clades included a clade containing A domains from PKS/NRPS hybrids, a clade containing A domains known to incorporate amino acids *N*-methylated after adenylation and thiolation characterized by the presence of A domains from the cyclosporine NRPS from *Tolypocladium inflatum*, a clade containing A domains from NRPS known to be involved in the synthesis of epipolythiodioxopiperazine (ETP) toxins characterized by the presence of gliotoxin and sirodesmin producing NRPS, a clade containing A domains from NRPS involved in the synthesis of β -lactam antibiotics characterized by the ACV synthetase from *Penicillium chrysogenum*, a clade containing A domains from NRPS known to synthesize siderophores characterized by the *Ustilago maydis* *sid2* NRPS (Fig. 3), and a clade characterized by the presence of an NRPS involved in oxidative stress resistance, *nps6*, found in *C. heterostrophus* (Supplemental Fig. 1). Only the *nps6* oxidative stress clade and a clade of NRPS making an unknown product contained A domains and the corresponding NRPS from all five *Aspergillus* species. Based on these results, we conclude that very few NRPS are common to all investigated *Aspergillus* species.

Examples of individual A domains from a single NRPS more closely related to themselves rather than A domains from other NRPS are also found in the tree. For example, most of the A domains from *A. fumigatus* NRPS1 and NRPS8 were found grouped together in the tree (Supplemental Fig. 1). However, A domain 2 from *A. fumigatus* NRPS1 and A domain 6 from *A. fumigatus* NRPS8 were found in different parts of the tree and more closely related to A domains from other NRPS. Due to the

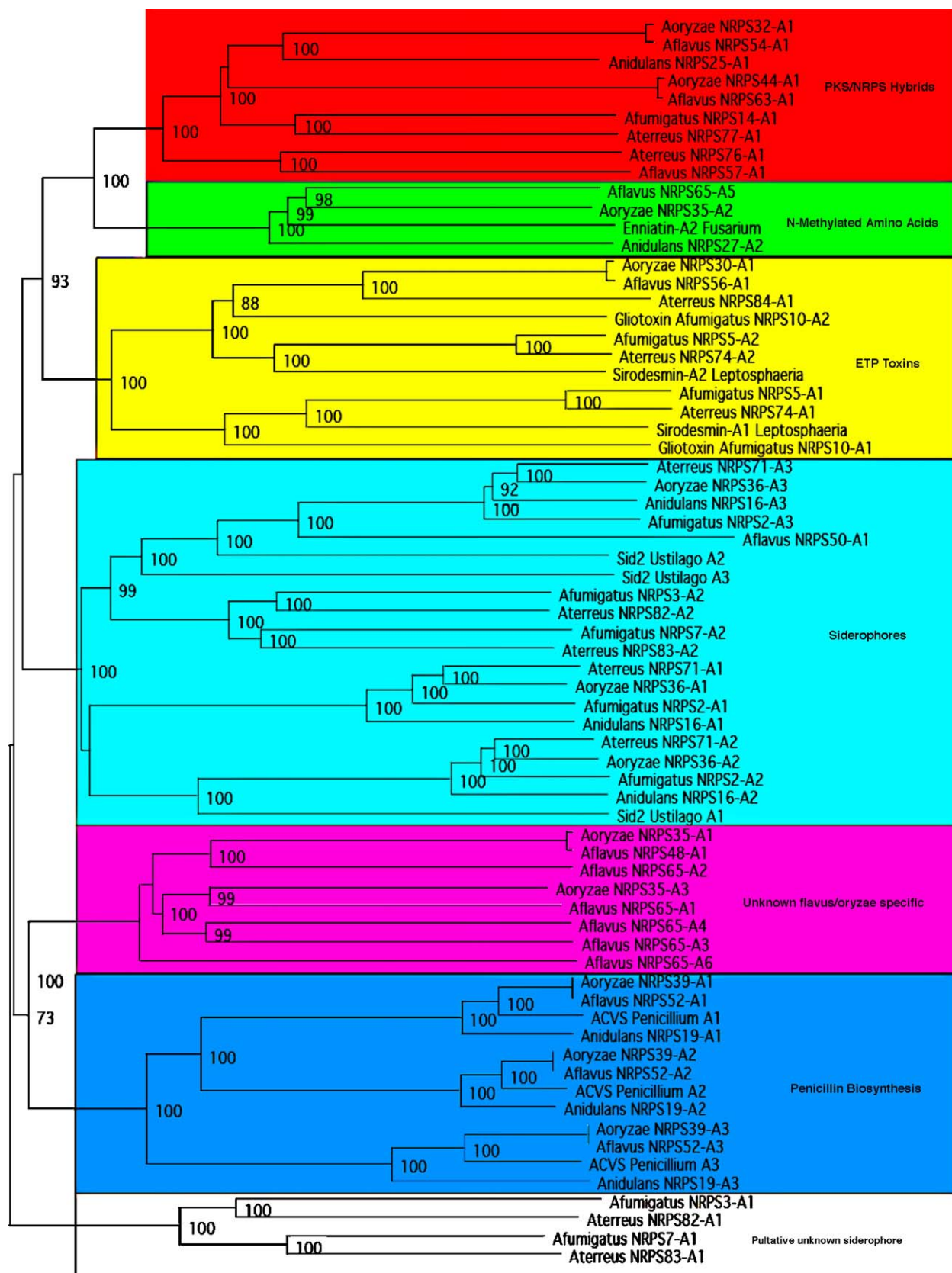


Fig. 4. Partial genealogy of *Aspergillus* adenylation (A) domains from non-ribosomal peptide synthetases (NRPS). Unrooted neighbor-joining tree created from CLUSTALW multiple alignment of 216 A domains. Colored blocks mark sub-clades with NRPS predicted to synthesize similar final peptide products. Branch length is indicative of number of amino acid changes. Numbers at nodes indicate bootstrap support, reported when $\geq 70\%$ for each clade performed with 100 replications. The tree portion of the tree presented is the most unambiguous. The entire tree can be viewed as Supplemental Fig. 1.

large sequence variation among A domains from the identified NRPS amino acid sequences, many parts of the tree were ambiguous. This result further strengthens the hypothesis that *Aspergillus* species produce a diverse array of small molecular weight peptides, many unique to individual *Aspergillus* species.

4. Discussion

Aspergillus species are important members of the ecological community. They are important plant and animal pathogens, and are also responsible for producing biologically active secondary metabolites that benefit humankind. The results of this study revealed that the genus *Aspergillus* has the potential to synthesize many unique non-ribosomal small molecular weight peptides. Synthesis of these peptides occurs via NRPS that have a complex evolutionary history in the genus *Aspergillus* as indicated by our A domain genealogy (Fig. 4 and Supplemental Fig. 1). Multiple evolutionary mechanisms appear to be responsible for the diversity of NRPS and NRPS products seen in the genus *Aspergillus*. This result is similar to previous findings with fungal PKS genes (Kroken et al., 2003). These evolutionary mechanisms may include a combination of A domain duplication, domain and module rearrangements, gene loss, and possible horizontal gene transfer.

Despite the suggested large diversity of NRPS genes in the A domain genealogy, we did find conserved orthologous NRPS present in *Aspergillus* species. The best example of this occurrence was found in a sub-clade of A domains from NRPS likely orthologous to the ferricrocin NRPS *sidC* of *A. nidulans* (*A. nidulans* NRPS16 in this study). Four of the 5 *Aspergillus* species examined contained A domains related to A domains from *sidC* (Fig. 4). The lone exception was *Aspergillus flavus*, which apparently does not synthesize a ferricrocin-like siderophore. We concluded that these A domains are from orthologous NRPS genes in other *Aspergillus* species. This hypothesis was further supported by the identical module arrangement (AT CAT ATC TCTCT) of the putative orthologous NRPS, and in the case of *A. fumigatus* NRPS2, previous gene expression studies showing expression of this NRPS (named *AfsidC*) under low iron conditions (Reiber et al., 2005). All four orthologous NRPS C termini ended with the unique domain arrangement of TCTC. This identical domain arrangement is found at the C terminus of the *U. maydis* *sid2* siderophore synthesizing NRPS, and functions to close the nascent tripeptide ring (Yuan et al., 2001). The occurrence of C and T domains at the carboxy terminus also is indicative of an iterative NRPS. C and T domains were often found at the C terminus of the majority of *Aspergillus* NRPS examined in this study.

Within this clade of ferrichrome siderophore synthesizing NRPS, a sub-clade containing the A domains from the second module of two *A. fumigatus* and two *A. terreus* NRPS was present. A domains from the first module were also found grouped together above this larger siderophore clade, suggesting that NRPS3 from *A. fumigatus* and NRPS82 from *A. terreus*, and *A. fumigatus* NRPS7 and *A. terreus* NRPS83 are two sets of orthologous pairs. In addition, the presence of the second module A domains within the siderophore clade may suggest these NRPS

produce siderophore like metabolites. This is significant since siderophores are well known virulence factors in bacterial pathogenesis and have recently been shown to be important in the development of invasive aspergillosis (Hissen et al., 2004, 2005). The potential ability to synthesize additional siderophores could contribute to the ability of *A. fumigatus* and *A. terreus* to survive in the iron poor environment of patient serum and increase the morbidity associated with invasive aspergillosis.

Only two sub-clades were found that contained orthologous NRPS from all five species of *Aspergillus*. The first clade was characterized by the presence of an NRPS known to be required for fungal plant pathogenesis and involved in oxidative stress tolerance, *nps6* from *C. heterostrophus* (Supplemental Fig. 1). The second clade did not contain a NRPS with a characterized product, and the final peptide product produced by these probable orthologs is unknown. To date, orthologs of *nps6* have been found in every ascomycete examined (Lee et al., 2005). The *A. fumigatus* ortholog is a strong candidate to be involved in invasive aspergillosis. It is well established that phagocytes use oxidative stress to kill conidia and hyphae of *A. fumigatus* (Levitz et al., 1986; Waldorf, 1989; Nessa et al., 1997; Marr et al., 2001). In our study, the *nps6* ortholog in *A. fumigatus*, NRPS4, was the NRPS expressed most in co-culture with the macrophage like cell line J774A.1 (Fig. 2). Gene replacement studies are currently underway to directly address the role of *A. fumigatus* NRPS4 in invasive aspergillosis.

The observation that all A domains from a single NRPS gene occurred in the same clade strongly suggests that domain duplication has played an important role in NRPS evolution. This finding further strengthens the hypothesis that A domain duplication and subsequent mutations that affect the specificity of substrate selection have helped give rise to the diversity of non-ribosomally synthesized peptides in the genus *Aspergillus*. In addition, there were examples of A domains from a single NRPS more similar to each other rather than to A domains from other NRPS, which may indicate that this NRPS was acquired after the *Aspergillus* species diverged. One example is NRPS8 from *A. fumigatus* (Supplemental Fig. 1).

It is tempting to hypothesize that species specific NRPS, such as NRPS8 and the gliotoxin biosynthesis NRPS10 from *A. fumigatus*, were acquired via horizontal gene transfer. Horizontal gene transfer has been hypothesized to account for the large diversity of secondary metabolites seen in filamentous fungi (Walton, 2000). However, gene duplication and gene loss are alternative hypotheses that can also explain the results in the genealogy. As more fungal genomes are sequenced, and more orthologous NRPS genes identified, this hypothesis may become more testable. Still, NRPS8 from *A. fumigatus* is a unique NRPS that warrants further study. NRPS8 transcripts were abundant in ungerminated spores suggesting that the small peptide produced is important in the early stages of fungal development (Fig. 2). *A. fumigatus* NRPS8 was the largest NRPS identified in this study, predicted to encode a small peptide with six amino acid residues. However, the presence of three epimerization domains and multiple C domains is inconsistent with traditional NRPS module organization, and may suggest that a more complex final product is synthesized.

Further studies to identify this peptide product may uncover why this NRPS is unique to *A. fumigatus* and not the other *Aspergillus* species examined in this study.

In summary, this study reports the characterization of the non-ribosomal biosynthetic capabilities of five *Aspergillus* species. Identifying and understanding the biological activities of these important small peptides is critical to gain a better understanding of the full impact of the aspergilli on humankind. Our phylogenomic analysis of A domains from *Aspergillus* NRPS suggests a complex evolutionary history for these multi-modular enzymes including gene duplications, modular rearrangements, and potential gene loss and/or horizontal gene transfer. The differential gene expression profiles seen with *A. fumigatus* NRPS and forthcoming analysis of NRPS mutants may in the near future help link specific metabolites with their corresponding NRPS. These links will be crucial for a fuller understanding of the diversification, maintenance, and evolution of NRPS genes within the genus *Aspergillus*, and possibly shed more light on their role in nature.

Acknowledgements

Robert A. Cramer Jr. is funded by the NIH/NIAID Molecular Mycology and Pathogenesis Training Program Contract No. 5 T32 AI052080. This project was funded in part by NIH/NIAID K08 award to William J. Steinbach, MD, Contract No. 1 K08 AI061149-01. Thanks to Dr. Andrew Alspaugh, Duke University Medical Center, for useful comments and technical assistance with the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.gene.2006.07.008](https://doi.org/10.1016/j.gene.2006.07.008).

References

- Ansari, M.Z., Yadav, G., Gokhale, R.S., Mohanty, D., 2004. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* 32, W405–W413.
- Bennett, J.W., 1989. Mycotoxin research: 1989. *Mycopathologia* 107, 65–66.
- Bok, J.W., et al., 2005. LaeA, a regulator of morphogenetic fungal virulence factors. *Eukaryot. Cell* 4, 1574–1582.
- Cramer Jr., R.A., et al., 2006. Disruption of a nonribosomal peptide synthetase in *Aspergillus fumigatus* eliminates gliotoxin production. *Eukaryot. Cell* 5, 972–980.
- Eddy, S.R., 1996. Hidden Markov models. *Curr. Opin. Struct. Biol.* 6, 361–365.
- Finn, R.D., et al., 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* 34, D247–D251.
- Gardiner, D.M., Cozijnsen, A.J., Wilson, L.M., Pedras, M.S., Howlett, B.J., 2004. The siderophore biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Mol. Microbiol.* 53, 1307–1318.
- Hissen, A.H., Chow, J.M., Pinto, L.J., Moore, M.M., 2004. Survival of *Aspergillus fumigatus* in serum involves removal of iron from transferrin: the role of siderophores. *Infect. Immun.* 72, 1402–1408.
- Hissen, A.H., Wan, A.N., Warwas, M.L., Pinto, L.J., Moore, M.M., 2005. The *Aspergillus fumigatus* siderophore biosynthetic gene sidA, encoding L-ornithine N5-oxygenase, is required for virulence. *Infect. Immun.* 73, 5493–5503.
- Johnson, J.M., Mason, K., Moallemi, C., Xi, H., Somaroo, S., Huang, E.S., 2003. Protein family annotation in a multiple alignment viewer. *Bioinformatics* 19, 544–545.
- Keller, N.P., Turner, G., Bennett, J.W., 2005. Fungal secondary metabolism — from biochemistry to genomics. *Nat. Rev., Microbiol.* 3, 937–947.
- Korf, I., 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5, 59.
- Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C., Turgeon, B.G., 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15670–15675.
- Latge, J.P., 1999. *Aspergillus fumigatus* and aspergillosis. *Clin. Microbiol. Rev.* 12, 310–350.
- Lee, B.N., Kroken, S., Chou, D.Y., Robbertse, B., Yoder, O.C., Turgeon, B.G., 2005. Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. *Eukaryot. Cell* 4, 545–555.
- Levitz, S.M., Selsted, M.E., Ganz, T., Lehrer, R.I., Diamond, R.D., 1986. In vitro killing of spores and hyphae of *Aspergillus fumigatus* and *Rhizopus oryzae* by rabbit neutrophil cationic peptides and bronchoalveolar macrophages. *J. Infect. Dis.* 154, 483–489.
- Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879.
- Marahiel, M.A., 1992. Multidomain enzymes involved in peptide synthesis. *FEBS Lett.* 307, 40–43.
- Marr, K.A., Koudadoust, M., Black, M., Balajee, S.A., 2001. Early events in macrophage killing of *Aspergillus fumigatus* conidia: new flow cytometric viability assay. *Clin. Diagn. Lab. Immunol.* 8, 1240–1247.
- Nessa, K., Palmberg, L., Johard, U., Malmberg, P., Jarstrand, C., Camner, P., 1997. Reaction of human alveolar macrophages to exposure to *Aspergillus fumigatus* and inert particles. *Environ. Res.* 75, 141–148.
- Nierman, W.C., et al., 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438, 1151–1156.
- Osborn, A.E., 2001. Tox-boxes, fungal secondary metabolites, and plant disease. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14187–14188.
- Panaccione, D.G., Scott-Craig, J.S., Pocard, J.A., Walton, J.D., 1992. A cyclic peptide synthetase gene required for pathogenicity of the fungus *Cochliobolus carbonum* on maize. *Proc. Natl. Acad. Sci. U. S. A.* 89, 6590–6594.
- Perfect, J.R., et al., 2001. The impact of culture isolation of *Aspergillus* species: a hospital-based survey of aspergillosis. *Clin. Infect. Dis.* 33, 1824–1833.
- Reiber, K., et al., 2005. The expression of selected non-ribosomal peptide synthetases in *Aspergillus fumigatus* is controlled by the availability of free iron. *FEMS Microbiol. Lett.* 248, 83–91.
- Silakowski, B., Kunze, B., Muller, R., 2001. Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. *Gene* 275, 233–240.
- Slater, G.S., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31.
- Stajich, J.E., et al., 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 12, 1611–1618.
- Stanke, M., Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 (Suppl 2), II215–II225.
- Szekeres, A., et al., 2005. Peptaibols and related peptaibiotics of *Trichoderma*. A review. *Acta Microbiol. Immunol. Hung.* 52, 137–168.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Waldorf, A.R., 1989. Pulmonary defense mechanisms against opportunistic fungal pathogens. *Immunol. Ser.* 47, 243–271.
- Walton, J.D., 1987. Two enzymes involved in biosynthesis of the host-selective phytotoxin HC-toxin. *Proc. Natl. Acad. Sci. U. S. A.* 84, 8444–8447.
- Walton, J.D., 2000. Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet. Biol.* 30, 167–171.
- Weber, T., Marahiel, M.A., 2001. Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure (Camb)* 9, R3–R9.
- Yuan, W.M., Gentil, G.D., Budde, A.D., Leong, S.A., 2001. Characterization of the *Ustilago maydis* sid2 gene, encoding a multidomain peptide synthetase in the ferriochrome biosynthetic gene cluster. *J. Bacteriol.* 183, 4040–4051.