

## Supplementary Materials for:

Acrophiarin (antibiotic S31794/F-1) from *Penicillium arenicola* shares biosynthetic features with both *Aspergillus*- and *Leotiomyce*-type echinocandins

Nan Lan<sup>1</sup>, Bruno Perlatti<sup>1</sup>, Daniel J. Kvitsek<sup>2</sup>, Philipp Wiemann<sup>2</sup>, Colin J.B. Harvey<sup>2</sup>, Jens Frisvad<sup>3</sup>, Zhiqiang An<sup>1</sup>, Gerald F. Bills<sup>1^</sup>

<sup>1</sup>Texas Therapeutics Institute, The Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, Texas 77054, USA

<sup>2</sup>Hexagon Bio, Menlo Park, California 94025, USA

<sup>3</sup>Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, DK-2800, Lyngby, Denmark

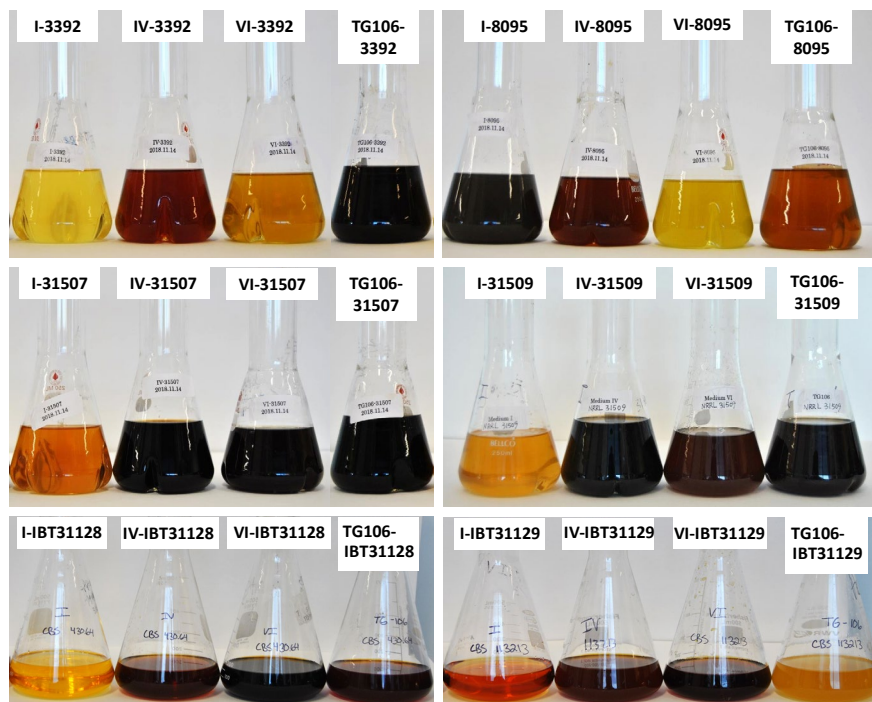


Fig. S1. Methanol extractions showing variation in pigmentation from four acrophiarin-producing strains of *P. arenicola*.

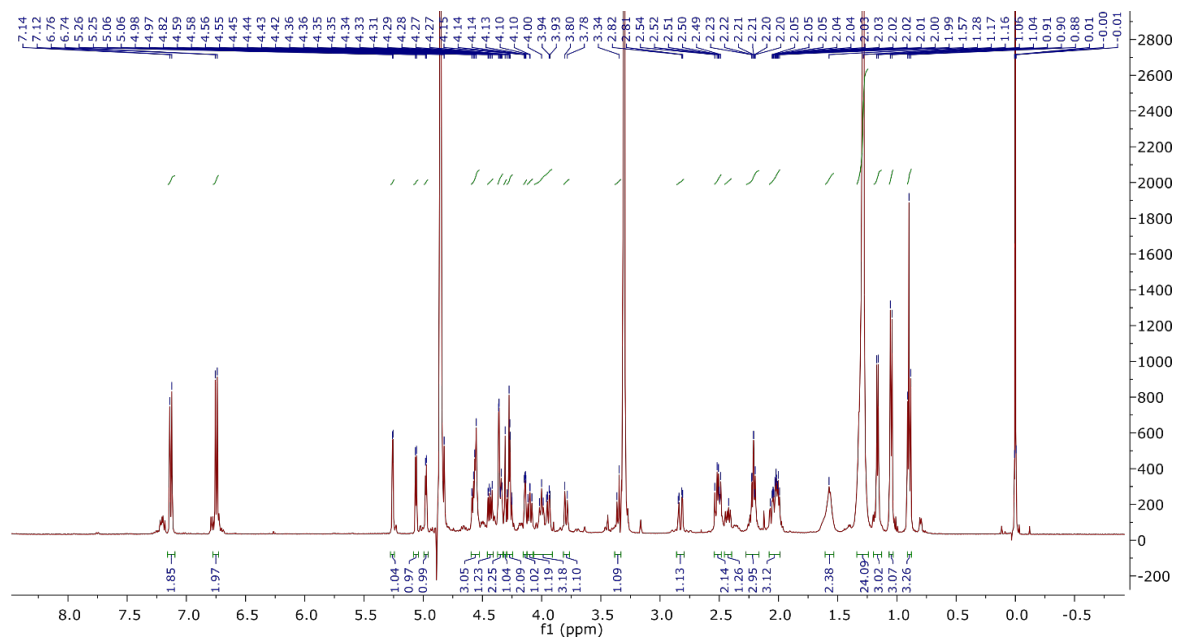


Fig. S2. <sup>1</sup>H NMR of acrophiarin obtained from *P. arenicola* NRRL 8095.

Table S1. The list of *P. arenicola* (NRRL 8095) acrophiarin BCG and surrounding genes.

(No PKS gene were detected).

Contig 7	
Gene	Products
g2.t1	hypothetical protein
g3.t1	hypothetical protein
g4.t1	Phosphipase C
g5.t1	putative P450
g6.t1	FAD dependent oxidoreductase
g7.t1	transporter
g8.t1	putative 2-hydroxyacid dehydrogenase
g9.t1	aspartate racemase protein
g10.t1	class II aldolase and adducin domain-containing protein
g11.t1	hypothetical protein
g12.t1	fungal specific transcription factor
g13.t1	hypothetical protein
g14.t1	hypothetical protein
<b>g15.t1 to g27.t1</b>	<b>acrophiarin gene cluster</b>
g28.t1	hypothetical protein
g29.t1	C6 finger domain-containing protein
g30.t1	alpha/beta-hydrolase
g31.t1	putative glyoxalase family protein
g32.t1	fumarylacetoacetate hydrolase protein
g33.t1	related to MFS drug efflux pump
g34.t1	Clavaminatase synthase-like protein
g35.t1	MFS general substrate transporter
g36.t1	Zn(II)2Cys6 transcription factor
g37.t1	hypothetical protein
g38.t1	glycosyltransferase family 71 protein
g39.t1	carbohydrate-binding protein

BLAST Results

[Questions/comments](#)

Blast 2 sequences

Job title: GLPKS4

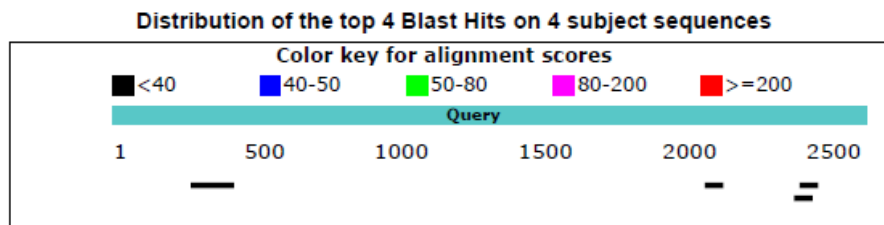
RID [B24SZ2KG114](#) (Expires on 04-14 06:37 am)

Query ID |cl|Query\_92081  
 Description GLPKS4  
 Molecule type amino acid  
 Query Length 2531

Subject ID 91 subjects  
 Description [See details](#)  
 Molecule type amino acid  
 Subject Length 52726  
 Program BLASTP 2.9.0+

Analyze your query with [SmartBLAST](#)

[Graphic Summary](#)



[Descriptions](#)

Sequences producing significant alignments:

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
contig7:g7.t2	26.2	26.2	2%	1.4	34.48%	Query_92092
contig7:g7.t1	26.2	26.2	2%	1.5	34.48%	Query_92091
contig7:g21.t1	25.8	25.8	2%	2.3	28.79%	Query_92109
contig7:g48.t1	23.9	23.9	5%	7.2	24.00%	Query_92143

## Alignments

contig7:g7.t2

Sequence ID: Query\_92092 Length: 541 Number of Matches: 1

Range 1: 44 to 97

Score	Expect	Method	Identities	Positives	Gaps	Frame
26.2 bits(56)	1.4()	Compositional matrix adjust.	20/58(34%)	29/58(50%)	4/58(6%)	

Features:

Query	2344	LNWAVAKSAVKPSLPGQNQLLIGVRSKSLSDPSNRVSKRDARMGAYLNTGSSTSAN	2401
		+ W S+V PSLPG Q+ G+ S+ + + SN VS +GA + G S N	
Sbjct	44	ITWDQGSSSVPSLPG-FQMKFGITSAANPTQISNFVSL---VYIGAGIGGGLSFFIN	97

contig7:g7.t1

Sequence ID: Query\_92091 Length: 542 Number of Matches: 1

Range 1: 44 to 97

Score	Expect	Method	Identities	Positives	Gaps	Frame
26.2 bits(56)	1.5()	Compositional matrix adjust.	20/58(34%)	29/58(50%)	4/58(6%)	

Features:

Query	2344	LNWAVAKSAVKPSLPGQNQLLIGVRSKSLSDPSNRVSKRDARMGAYLNTGSSTSAN	2401
		+ W S+V PSLPG Q+ G+ S+ + + SN VS +GA + G S N	
Sbjct	44	ITWDQGSSSVPSLPG-FQMKFGITSAANPTQISNFVSL---VYIGAGIGGGLSFFIN	97

contig7:g21.t1

Sequence ID: Query\_92109 Length: 7573 Number of Matches: 1

Range 1: 6426 to 6491

Score	Expect	Method	Identities	Positives	Gaps	Frame
25.8 bits(55)	2.3()	Compositional matrix adjust.	19/66(29%)	30/66(45%)	10/66(15%)	

Features:

Query	2024	EYGSMEIGKRDF-----IGKAQLNMDLFESNRSFFGVDLAKFDAARCQLLLVQM	2073
		E GS++ +G+RD IG+ + + S V+L K A Q+ LV	
Sbjct	6426	EDGSLIHLGRRDTQIKIRGQRVEIGEIEYQITQHLPEASTVAVELLKQGAQQQVSLVAA	6485
Query	2074	MEFYEK 2079	
		+EF E+	
Sbjct	6486	IEFIEE 6491	

contig7:g48.t1

Sequence ID: Query\_92143 Length: 507 Number of Matches: 1

Range 1: 155 to 299

Score	Expect	Method	Identities	Positives	Gaps	Frame
23.9 bits(50)	7.2()	Compositional matrix adjust.	36/150(24%)	51/150(34%)	18/150(12%)	

Features:

Query	267	RDGNIRSVIRATASNCDGKTPGITLPSSSESHEAMMRRAYKEA---CLDPTQTAFV----	319
		R D I ++ ATA N G G + S A + CL T A+	
Sbjct	155	RKKDMIFAIFGATAPN--GAVIGAVFAALFSQLAWWPWFVWMAIICLLTTALAYFVIPR	212
Query	320	--EAHGTGTKIGDPLEATAIARVFSSEKGVYIGSVK----PNVGHSEGASGVTSIMKAVL	373
		G I +E I F G+ + +V P VG V I+ A+	
Sbjct	213	IHHERKHGLTISGWIEELDIVGAF LGIVGLVLVNVAVNQAPIVGWQMPYVYVLLIVGALF	272
Query	374	ALENRTIPPNINFSTPNPQIPFEASNMKVA	403
		+ + P P IPF A N+ V+	
Sbjct	273	LVAFFWELQV---APKPLIPFHAFNLDVS	299

ST is a registered trademark of the National Library of Medicine

Fig. S3. NCBI blast results with GLPKS4 protein sequence as query and all the predicted proteins within contig 7. No significant matches found.

Table S2. Genomes used for phylogenetic reference reconstructions.

Species	strain	NCBI Taxonomy	Genome Center
<i>Glarea lozoyensis</i>	ATCC 20868	Leotiomycetes: Helotiales	PRJNA72229
<i>Pezizula radicola</i>	NRRL 12192	Leotiomycetes; Helotiales	PRJNA412280
<i>Aspergillus mulundensis</i>	DSM 5745	Eurotiomycetes; Aspergillaceae	PRJNA389331
<i>Aspergillus japonicus</i>	CBS 114.51	Eurotiomycetes; Aspergillaceae	PRJNA235073
<i>Aspergillus aculeatus</i>	ATCC 16872	Eurotiomycetes; Aspergillaceae	PRJNA82765
<i>Aspergillus versicolor</i>	CBS 583.65	Eurotiomycetes; Aspergillaceae	PRJNA207687
<i>Aspergillus sydowii</i>	CBS 593.65	Eurotiomycetes; Aspergillaceae	PRJNA207689
<i>Aspergillus nidulans</i>	FGSC A4	Eurotiomycetes: Aspergillaceae	PRJNA13961
<i>Aspergillus flavus</i>	NRRL 3357	Eurotiomycetes: Aspergillaceae	PRJNA38227
<i>Aspergillus niger</i>	ATCC 1015	Eurotiomycetes: Aspergillaceae	PRJNA15785
<i>Aspergillus oryzae</i>	RIB40	Eurotiomycetes: Aspergillaceae	PRJNA28175
<i>Aspergillus terreus</i>	NIH 2624	Eurotiomycetes: Aspergillaceae	PRJNA17637
<i>Coleophoma crateriformis</i>	BP5796	Leotiomycetes; Helotiales	PRJNA388491
<i>Coleophoma cylindrospora</i>	BP6252	Leotiomycetes; Helotiales	PRJNA389322
<i>Penicillium chrysogenum</i>	Wisconsin 54-1255	Eurotiomycetes: Aspergillaceae	PRJNA39879
<i>Talaromyces marneffeii</i>	ATCC 18224	Eurotiomycetes: Trichocomaceae	PRJNA32665
<i>Talaromyces stipitatus</i>	ATCC 10500	Eurotiomycetes: Trichocomaceae	PRJNA38857
<i>Arthrobotrys oligospora</i>	ATCC 24927	Orbiliomycetes: Orbiliaceae	PRJNA41495
<i>Chaetomium globosum</i>	CBS 148.51	Sordariomycetes: Chaetomiaceae	PRJNA12795
<i>Cordyceps militaris</i>	CM01	Sordariomycetes: Cordycipitaceae	PRJNA225510
<i>Fusarium oxysporum f. sp. lycopersici</i>	4287	Sordariomycetes: Nectriaceae	PRJNA18813
<i>Fusarium verticillioides</i>	7600	Sordariomycetes: Nectriaceae	PRJNA15553
<i>Neurospora crassa</i>	OR74A	Sordariomycetes: Sordariaceae	PRJNA13841
<i>Saccharomyces cerevisiae</i>	S288c	Saccharomycetes: Saccharomycetaceae	PRJNA245761

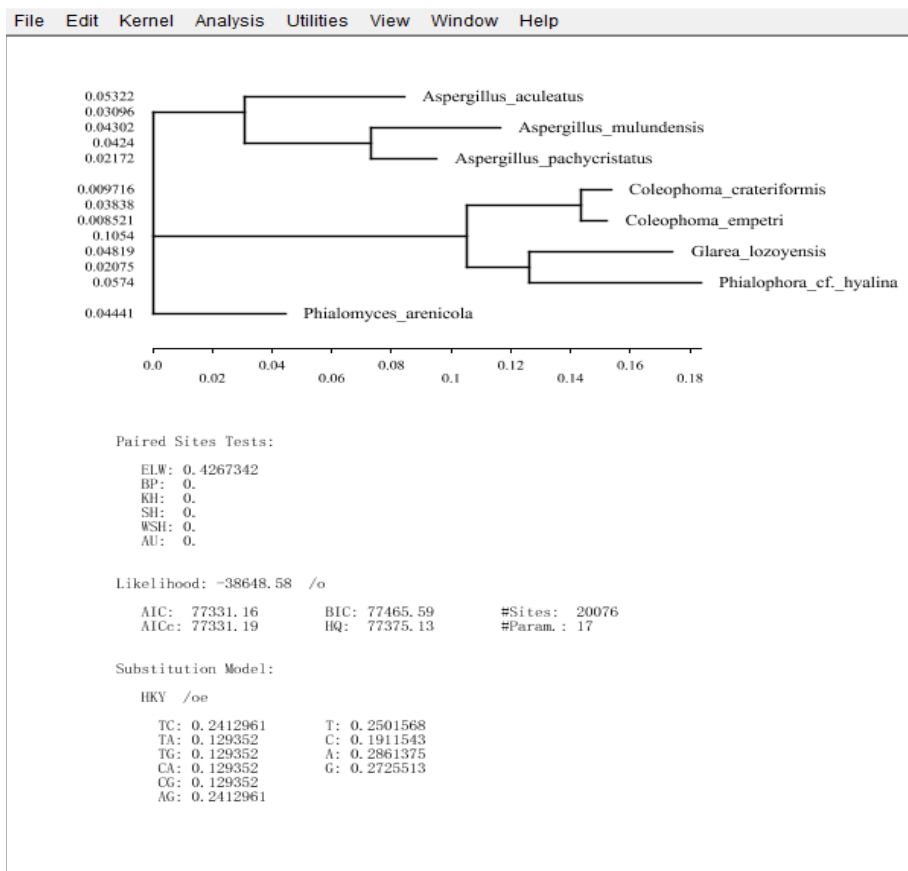


Fig. S4. Hypothesis testing of trees from a ten-set echinocandin pathway genes versus the current classification of the *Ascomycota* implement in TREEFINDER and subject to the Shimodaira-Hasegawa (SH) test and weighted Shimodaira-Hasegawa (WSH) test (Shimodaira H, 1999 Mol Biol Evol).

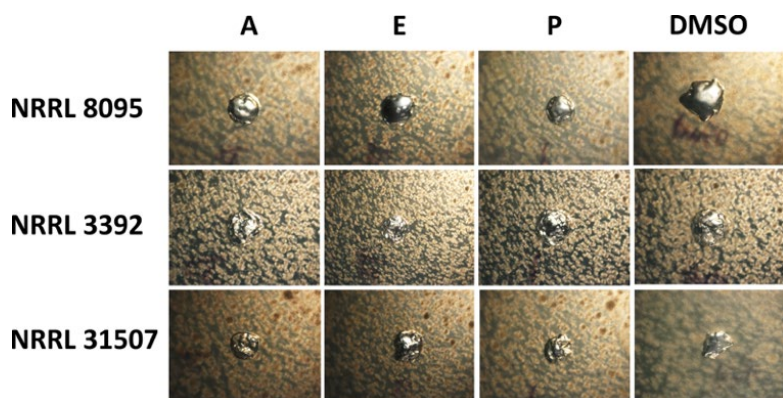


Fig. S5. Magnified view of inhibition zone of *P. arenicola* under stereomicroscope (NRRL 8095, NRRL 31507 and NRRL 3392) against echinocandins. A. Acrophiarin (250 ug/ml). E. Echinocandin B (250 ug/ml). P. Pneumocandin B<sub>0</sub> (250 ug/ml). Assay wells are 4 mm in diam.