BMC
Genomics

RESEARCH ARTICLE

Genomics-driven discovery of the pneumocandin biosynthetic gene cluster in the fungus *Glarea lozoyensis*

Li Chen[1,2†], Qun Yue[1†], Xinyu Zhang[1], Meichun Xiang[1], Chengshu Wang[3], Shaojie Li[1], Yongsheng Che[4], Francisco Javier Ortiz-López[5], Gerald F Bills[5,6], Xingzhong Liu[1*] and Zhiqiang An[6*]

Abstract

Background: The antifungal therapy caspofungin is a semi-synthetic derivative of pneumocandin B_0, a lipohexapeptide produced by the fungus *Glarea lozoyensis*, and was the first member of the echinocandin class approved for human therapy. The nonribosomal peptide synthetase (NRPS)-polyketide synthases (PKS) gene cluster responsible for pneumocandin biosynthesis from *G. lozoyensis* has not been elucidated to date. In this study, we report the elucidation of the pneumocandin biosynthetic gene cluster by whole genome sequencing of the *G. lozoyensis* wild-type strain ATCC 20868.

Results: The pneumocandin biosynthetic gene cluster contains a NRPS (GLNRPS4) and a PKS (GLPKS4) arranged in tandem, two cytochrome P450 monooxygenases, seven other modifying enzymes, and genes for L-homotyrosine biosynthesis, a component of the peptide core. Thus, the pneumocandin biosynthetic gene cluster is significantly more autonomous and organized than that of the recently characterized echinocandin B gene cluster. Disruption mutants of GLNRPS4 and GLPKS4 no longer produced the pneumocandins (A_0 and B_0), and the Δglnrps4 and Δglpks4 mutants lost antifungal activity against the human pathogenic fungus *Candida albicans*. In addition to pneumocandins, the *G. lozoyensis* genome encodes a rich repertoire of natural product-encoding genes including 24 PKSs, six NRPSs, five PKS-NRPS hybrids, two dimethylallyl tryptophan synthases, and 14 terpene synthases.

Conclusions: Characterization of the gene cluster provides a blueprint for engineering new pneumocandin derivatives with improved pharmacological properties. Whole genome estimation of the secondary metabolite-encoding genes from *G. lozoyensis* provides yet another example of the huge potential for drug discovery from natural products from the fungal kingdom.

Background

Fungi frequently cause deadly infections in immunocompromised patients resulting from HIV infection, cancer chemotherapy, and organ transplantation [1]. Until the introduction of caspofungin (CANCIDAS™) in 2001, antifungal therapy was limited to the use of polyenes (amphotericin B), azoles, and flucytosine which have high failure rates during management of fungal infection, while experiencing increasing clinical resistance [1]. The echinocandins are a class of antifungal lipopeptides targeting fungi via noncompetitive inhibition of the β-1,3-D-glucan synthase enzyme complex, leading to glucan polymer depletion in the fungal cell wall and resulting in osmotic instability and fungal cell lysis [1]. Human side effects to these chemicals are minimal because the target is absent in mammalian cells, and low dosing is used due to the drug's potent efficacy [1,2]. Thus far, three echinocandin-based agents have been approved for clinical use [1]. Caspofungin, a semi-synthetic derivative of pneumocandin B_0 (Figure 1a) which is a lipohexapeptide produced by the filamentous fungus *Glarea lozoyensis* (Figure 1b), was the first member of this class approved for human therapy; its registration was followed by micafungin

* Correspondence: liuxz@im.ac.cn; zhiqiang.an@uth.tmc.edu
†Equal contributors
[1]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
[6]Texas Therapeutics Institute, the Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA
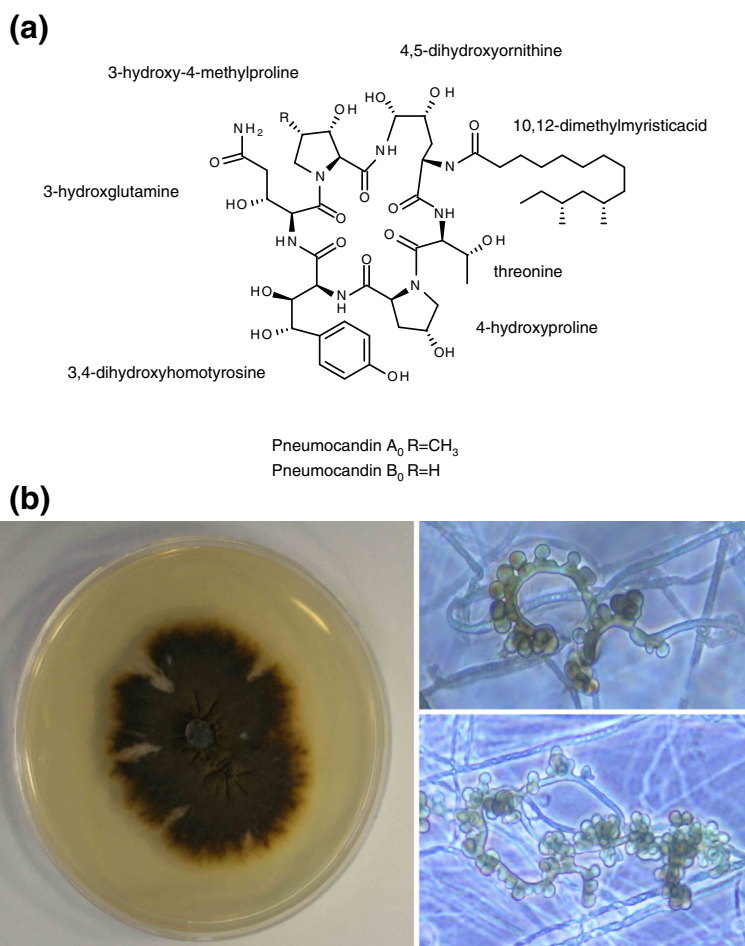Full list of author information is available at the end of the article

Figure 1 Pneumocandin structures and morphology of *Glarea lozoyensis*. (**a**) Chemical structures of pneumocandins. (**b**) Colony of *G. lozoyensis* on malt yeast agar (left panel); conidiophores and conidia of *G. lozoyensis* (right panels).

(MYCAMINE™) derived from FR901370 (WF11899A), a sulfonated hexapeptide produced by the fungus *Coleophoma empetri* [3], and lastly anidulafungin (ERAXIS™) derived from echinocandin B produced by the fungus *Aspergillus rugulosus* [4]. The three fungal metabolites share a common chemical structure of cyclic lipohexapeptide with *N*-acylated to either 10,12-dimethylmyristoyl (pneumocandins) or palmitoyl (FR901370) or linoleoyl (echinocandin B); their hexapeptide cores differ from each other by modifications on 4-hydroxyproline or dihydroxyhomotyrosine (FR901370 and pneumocandins possess 3-hydroxyglutamine, while echinocandin B has threonine substituted in the same position) [5-8]. Because of their high efficacy, they have become the first-line therapy for the treatment of invasive fungal infections [1].

Several cases of *in vivo* caspofungin resistance have been reported for *Candida* and *Aspergillus* species caused by mutations that reduce the drug sensitivity of the glucan synthase by several thousand-fold [9-12]. A compensatory cell wall remodeling mechanism elevating the chitin content has been found to be associated with caspofungin resistance in *C. albicans* [13-15]. Generation of pneumocandin derivatives with more desirable pharmacological properties via medicinal chemistry approaches has proven difficult [16,17]. Elucidation of the biosynthetic pathway to pneumocandins is the first step in applying pathway manipulation and biocombinatorial chemistry approaches to engineer new derivatives with broader spectra of activity and improved physiochemical characteristics to meet the challenges of broader efficacy and clinical resistance.

Based on the structure of pneumocandin, participation of a nonribosomal peptide synthetase (NRPS) and a polyketide synthase (PKS) are predicted for biosynthesis of the cyclic hexapeptide and the 10,12-dimethylmyristoyl side chain [18,19], respectively. Previous attempts to clone the NRPS and PKS gene cluster responsible for pneumocandin biosynthesis from *G. lozoyensis* have been unsuccessful [20,21]. Whole genome sequencing has proven to be an efficient approach in the identification of gene clusters of fungal secondary metabolites, such as

PKSs and NRPSs [22]. A recent genomic sequencing project of a pneumocandin B_0-overproducing mutant (ATCC 74030) derived from the wild-type (WT) strain of *G. lozoyensis* was inconclusive in identifying the pneumocandin biosynthetic cluster due to insufficient genome coverage [23]. In this study, we report the elucidation of the pneumocandin biosynthetic gene cluster by genome sequencing of the *G. lozoyensis* WT strain ATCC 20868. We also compare gene cluster organization with that of the recently published echinocandin B biosynthetic cluster [8,24]. In addition, analysis of the *G. lozoyensis* genome revealed a rich repertoire of secondary metabolite-encoding genes that once again illustrates the huge potential for drug discovery from natural products from the fungal kingdom.

Results

The genome characteristics of *G. lozoyensis*

Sequencing of the *G. lozoyensis* WT strain ATCC 20868 with an 80× genome coverage revealed a high resolution 39.6-megabase (Mb) genome with 0.5% repeat content. Reads were assembled into 22 scaffolds (>2 kb; N_{50}, 2.45 Mb) incorporating more than 99% of the total genomic base pairs (Figure 2a). The average gene density was 330 genes per Mb (Table 1). The 13,103 putative coding genes were assigned to different functional categories (Figure 2b). Consistent with previous studies by our group [25,26], a combined phylogenomic and phylogenetic analysis confirmed that *G. lozoyensis* belonged the same major phylogenetic lineage as the plant pathogenic fungi, *Sclerotinia sclerotiorum* and *Botrytis cinerea* [27], and the wood endophytic fungus, *Ascocoryne sarcoides*, the Helotiales [28] (Figure 3). A total of 4931 predicted proteins were assigned by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The top four categories in the KEGG functional classification were "Carbon Metabolism, Energy Metabolism, Amino Acid Metabolism, and Infectious Diseases" (Figure 4).

Strains of *Glarea lozoyensis* have been isolated from water, plant litter or soil samples [25,26]. However, the fungus has never been observed in nature, therefore its ecological role and trophic relationships remain unknown. It has been speculated that the fungus may be a plant or plant litter saprobe for the following reasons [25,26]. The fungus belongs to the same phylogenetic lineage as *Cyathicula* or *Crocicreas*, an inconspicuous group of fungi that are weak parasites, endophytes of living plants or saprobes of senescent plants and plant litter. In the laboratory, the fungus readily colonized and sporulated on sterilized hardwood [25]. Its asexual sporulation (Figure 1b) resembled that of a heterogeneous group of asexually reproducing fungi known as aero-aquatic fungi that often colonize plant debris in periodically inundated habitats [29,30]. Several recent studies have demonstrated a strong relationship between the suite of carbohydrate active enzymes (CAZymes,

http://www.cazy.org) in fungal genomes and their saprobic, parasitic or necrotrophic life strategies [31-33]. Such investigations have focused on those CAZymes involved in polysaccharide degradation and have contributed to a thorough understanding of the ecological role of a fungus. To infer whether *G. lozoyensis* might be a biotroph, saprotroph or necrotroph, we analyzed its complement of CAZy gene families and genes. The putative CAZymes in *G. lozoyensis* were identified using the CAZy annotation pipeline (http://mothra.ornl.gov/cgi-bin/cat.cgi) [34,35] and were compared to a selection of ascomycete and basidiomycete fungi (Figure 5a and 5b). At least 345 CAZymes in the five principal category families were identified in the genome (Figure 5a). This value is similar to the number of CAZymes found in known plant cell wall degrading ascomycetes, including the wood-inhabiting endophyte *A. sarcoides*, but significantly higher than the yeast *Saccharomyces cerevisiae*, and the plant biotrophic symbionts *Laccaria bicolor*, *Epichloë festucae*, and *Tuber melanosporum* (Figure 5a). A total of 180 glycoside hydrolases (GH) in 70 families were found in the *G. lozoyensis* genome, which is slightly less than average compared to other filamentous plant associated ascomycetes [36]. Likewise, the number of 67 glycosyl transferases (GT) in 35 families was also comparable to other plant inhabiting ascomycetes (Figure 5b). Average numbers of polysaccharide lyases (PL, 5), carbohydrate esterases (CE, 22) were found. However, a relatively abundant number of carbohydrate binding modules (CMB, 71) were identified. Therefore, its complement of genes associated with carbohydrate degradation and metabolism were consistent with those of other plant-associated ascomycetes.

G. lozoyensis genome revealed a rich repertoire of secondary metabolite-encoding genes

To identify the pathways involved in the synthesis of secondary metabolites in *G. lozoyensis*, we searched the genome for genes encoding key enzymes such as NRPS, PKS, terpene synthase (TS), and dimethylallyl tryptophan synthase (DMATS), which are essential for the biosynthesis of peptides, polyketides, terpenes, and alkaloids, respectively. The following secondary metabolite-encoding genes were dispersed among 49 gene clusters: six NRPSs, 24 PKSs, five polyketide synthase-nonribosomal peptide synthase hybrids (PKS-NRPS hybrids), 14 TSs, two DMATSs, 13 NRPS-like, one PKS-like, and one chalcone/stilbene synthase gene (Table 1). In addition to genes encoding the core enzyme(s), the majority of the 49 secondary metabolism gene clusters in *G. lozoyensis* contained genes encoding other biosynthesis enzymes, transcription regulators, and transporters. For example, about half of the gene clusters contained a gene encoding a Zn_2/Cys_6 or a C_2H_2 and C_2HC zinc transcriptional factor that could control the expression of genes within of its own cluster. Also, about
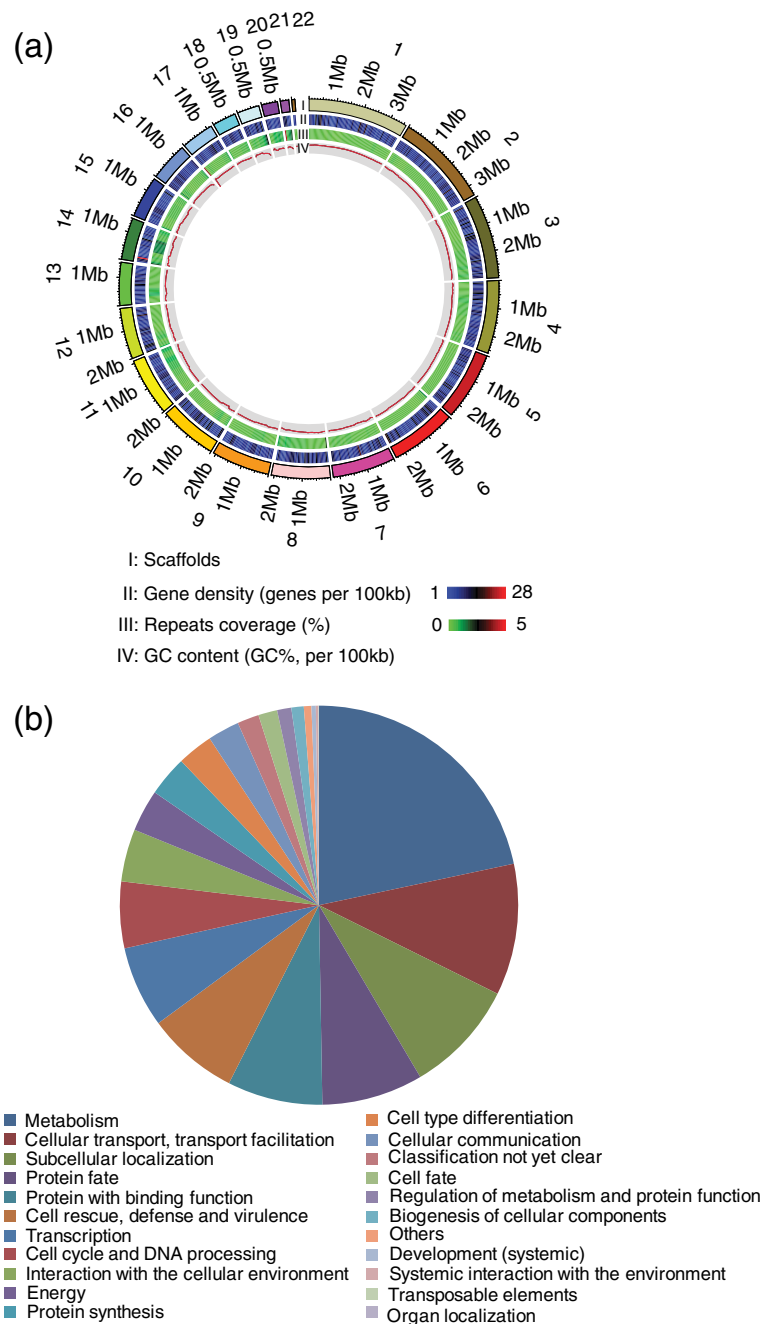
Figure 2 Genome features of *Glarea lozoyensis*. (**a**) General genome features of *G. lozoyensis*. I, 22 scaffolds (> 2 kb); II, gene density represented as number of genes per 100 kb; III, percentage of coverage of repetitive sequences; IV, GC content was estimated by the percent G + C in 100 kb. (**b**) Functional classificaton of proteins in the *G. lozoyensis* genome based on InterProScan analysis.

60% of the secondary metabolism clusters contained a gene encoding an ABC or a MFS transporter(s) that could export the metabolites produced by the enzymes encoded by the gene cluster (Additional file 1: Figure S1).

Biosynthetic capabilities of *G. lozoyensis*

An unexpected feature of the *G. lozoyensis* genome was its remarkable diversity of polyketide biosynthetic pathways and having at least 29 recognizable core PKS genes (Figure 6). Domain structure analysis revealed eight non-reducing PKSs, one partially-reducing PKS, four PKS-NRPS hybrids encoding partially reducing polyketides [37] and 16 PKSs encoding for highly reducing polyketides, including GLPKS4 and one PKS-NRPS hybrid (GLPKS3-NRPS) (Figure 6). A phylogenetic tree based on amino acid sequences of the ketosynthase

Table 1 General features of the *G. lozoyensis* genome

Features	
Assembly size (Mb)	39.6
Scaffold N_{50} (kb)	2453
Coverage (fold)	80
G + C content (%)	45.8
GC exonic (%)	49.06
GC intronic (%)	41.98
Repeat rate (%)	0.5
Protein-coding genes	13103
Gene density (per Mb)	330.38
Exons per gene	2.98
tRNAs	131
rRNAs	22
NRPSs	6
PKSs	24
PKS-NRPS hybrids	5
DMATSs	2
Terpene synthases	14
NRPS-like	13
PKS-like	1
Chalcone or stilbene synthase gene	1
Secondary metabolite gene clusters	49

Mb mega bases, *PKS* polyketide synthase, *NRPS* nonribosomal peptide synthetase, *PKS-NRPS hybrid* polyketide synthase-nonribosomal peptide synthetase hybrid, *DMATS* dimethylallyl tryptophan synthase, *tRNAs* transfer RNA, *rRNA* ribosomal RNA.

domains (KS) was constructed for the 24 PKSs and five PKS-NRPS hybrids in *G. lozoyensis* and 71 functionally characterized fungal PKSs encoding the products with known chemical structures (Figure 6, Additional file 2: Table S3). All four fungal-type PKS-NRPS hybrids (GLPKS26-NRPS, GLPKS27-NRPS, GLPKS28-NRPS, and GLPKS29-NRPS) were grouped with similar PKS-NRPS hybrids, such as those involved in the biosynthesis of the tetramic acids and HIV-1 integrase inhibitor equisetin (EqiS). Interestingly the four PKS-NRPS hybrids were also clustered with the HMG-CoA reductase inhibitor lovastatin (LDKS = LovB) which is proposed to be a truncated PKS-NRPS hybrid [38,39]. GLPKS8 and GLPKS9 were predicted to be non-reducing PKSs related to the PKSs responsible for biosynthesis of the metabolites mycophenolic acid and citrinin. GLPKS13 and three other *G. lozoyensis* PKSs (GLPKS10, GLPKS18, and GLPKS24) were grouped with the PKSs of lovastatin side chain (LNKS = LovF) [40] and the tetraketide acyl side chain of zaragozic acid A [41]. GLPKS19 and GLPKS11 shared significant homology with the T-toxin encoding gene CHPKS1 of *Cochliobolus heterostrophus* [42,43]. Six more *G. lozoyensis* PKSs (GLPKS4, GLPKS25,

GLPKS12, GLPKS7, GLPKS14, and GLPKS21) clustered with the hepato- and nephro-toxic fumonisin B_1 produced by *Gibberella fujikuroi* [44] and the solanapyrone Sol1 PKS of *Alternaria solani* [45]. The previously characterized GLPKS2, encoding for the biosynthesis of 6-methylsalicylic acid [20], grouped tightly with two other fungal 6-methylsalicylic acid PKSs, ATATX from *A. terreus* and MSAS from *Penicillium patulum* [46,47]. GLPKS1 has been previously identified as the *G. lozoyensis* melanin biosynthesis gene [21], and it clustered with other fungal di- and tetra-hydroxynaphthalene melanin biosynthesis genes, e.g. *Hypoxylon pulicicidum* (formerly *Nodulisporium* sp.) (NSPKS1) [48] and *Colletotrichum lagenarium* (CLPKS1) [49]. The ketosynthase sequence of *G. lozoyensis* GLPKS20 exhibited sequence similarities to genes involved in the biosynthesis of viridicatumtoxin [50]. Adjacent to the large groups of melanin and conidial pigment genes were the mycotoxin sterigmatocystin PKS (ANST) from *A. nidulans* [51] and the GLPKS5 from *G. lozoyensis*. Distantly related to the pigment PKSs was the *A. nidulans* orsellinic acid PKS protein OrsA [52,53], and GLPKS23 shared the same domain structure with OrsA. We speculated that orsellinic acid or related compounds may be produced by *G. lozoyensis*, and analysis of fermentations of *G. lozoyensis* confirmed that it produced isolecanoric acid (an orsellinic acid dimer) and pseudogyrophoric acid (a new orsellinic acid trimer) in certain culture media (Additional file 1: Figure S2). Therefore, we propose that GLPKS23 is responsible for orsellinic acid biosynthesis in *G. lozoyensis*. Cluster analysis revealed that a highly reducing PKS (GLPKS17) was proximal to a non-reducing PKS (GLPKS16) in the same cluster (Additional file 1: Figure S1). This tandem PKS structure was similar to that of the PKSs responsible for the biosynthesis of resorcylic acid lactones, e.g. radicicol and hypothemycin, and in fact, GLPKS16 appeared to be an ortholog of Hpm3 and RDC1 (Figure 6) [54,55].

NRPSs include modules that incorporate amino acids into the final peptide product. Each module minimally contains three domains, the adenylation domain (A domain), the thiolation domain (T domain), and the condensation domain (C domain). In addition to its abundant and diverse PKS pathways, the *G. lozoyensis* genome harbored six NRPS genes. Three NRPSs (GLNRPS2, GLNRPS3, GLNRPS5), contained a single module, encoding products with a single amino acid, the other three NRPSs (GLNRPS1, GLNRPS4, GLNRPS6), were multi-modular, encoding products with more than one amino acids (Figure 7). Gene cluster analysis revealed that GLNRPS1 (with two modules), GLNRPS2 and one NRPS-like genes located in the same cluster flanked by three clavaminate synthases (oxygenases) and MFS general substrate transporter genes (Additional file 1: Figure S1). These data indicated that a hydroxylation tetrapeptide product may
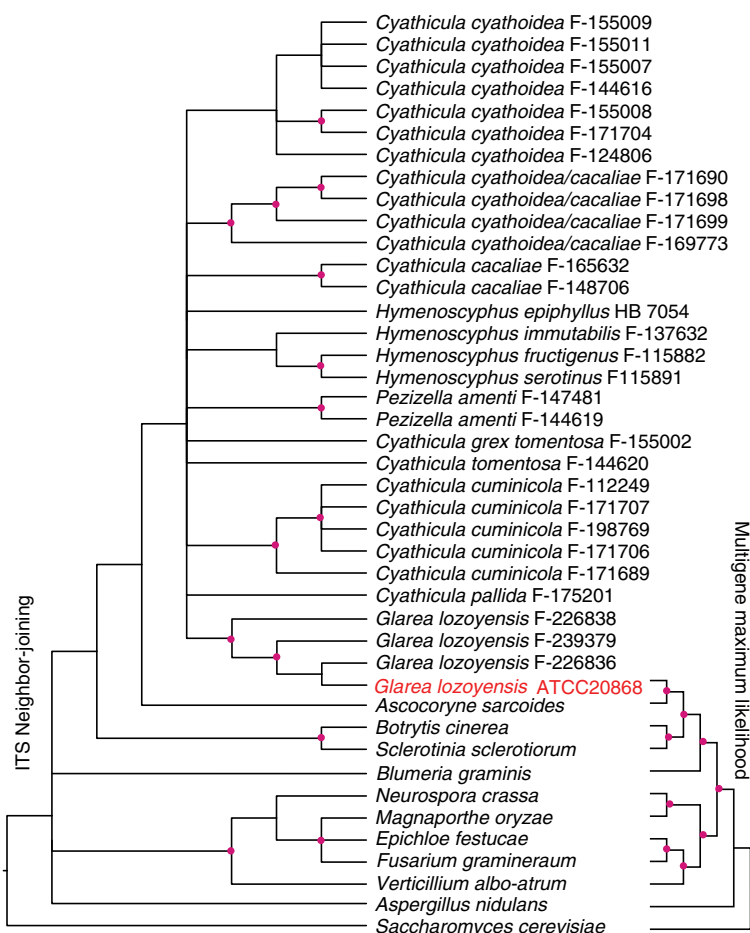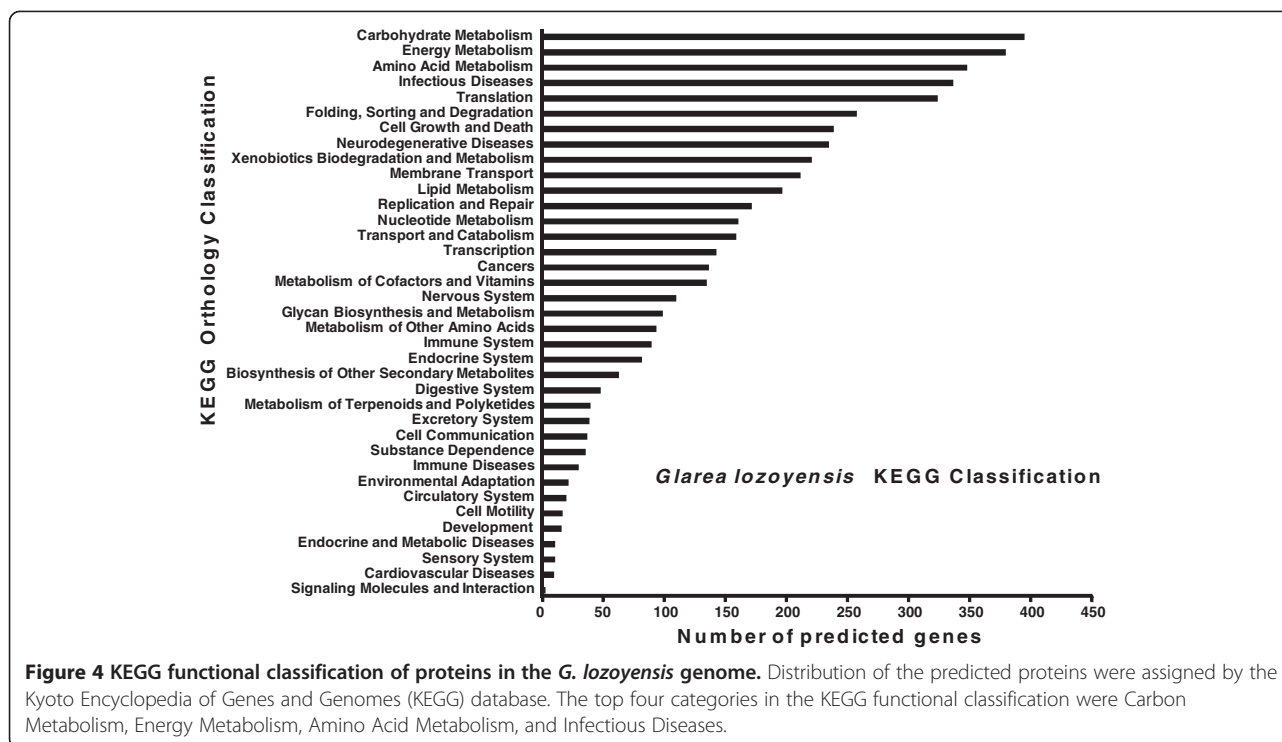
Figure 3 Phylogenetic analysis of *G. lozoyensis* using ITS sequences (brackets on the left) and genome protein sequences (brackets on the right). Left tree: The topology was estimated using neighbor-joining method based on the ITS sequence data from Peláez et al., 2011 [26] and the selected fungi on the right-side of the graphic. Right tree: A maximum likelihood phylogenomic tree showing evolutionary relationship of *G. lozoyensis* with selected ascomycete fungal species. The tree was constructed from the concatenated amino acid sequences of 878 common orthologous genes (Additional file 2: Table S2). The phylogenetic position of *G. lozoyensis* wild-type strain ATCC 20868 is marked in red. Branch nodes with greater than 60% support from 1000 bootstrapped pseudoreplicates are indicted with red dots in both trees. Both trees were rooted with S. cerevisiae.

be formed and excreted. GLNRPS4, with six modules that encode a hexapeptide product and located in a cluster bordered by various modifying enzymes, was proposed to be responsible for pneumocandin biosynthesis (Figure 7 and Figure 8a). Domain analysis revealed that GLNRPS6 had five modules, and module 1, module 3, module 5 contain one epimerization (E) domain respectively. The *glnrps6* was located in a cluster flanked by one MFS general substrate transporter gene, and thus suggested that a pentapeptide with three D-amino acids may be formed and excreted. Thirteen additional NRPS-like genes clusters were identified in *G. lozoyensis*, and some of them were located in clusters flanked by cytochrome P450, methyltranferase and transporter genes, thus indicating some hydroxylation and methylation products may be formed and excreted (Additional file 1: Figure S1).

To detect the classes of terpene synthases (TSs) in *G. lozoyensis*, the homologous sequences were analyzed by using BLAST at NCBI (http://www.ncbi.nlm.nih.gov/) (Additional file 2: Table S1). The richness of TSs, compared to related genome-sequenced fungi [27], revealed a great potential for *G. lozoyensis* to produce terpenoids. Three TS genes (GLAREA03340, GLAREA04931, GLAREA10578) encoded geranylgeranyl pyrophosphate synthase and geranylgeranyl transferase, and indicated these genes may be responsible for diterpene and carotenoid biosynthesis [56]. Two genes (GLAREA04679, GLAREA02940) encoding farnesyl pyrophosphate synthetase and farnesyl transferase indicated that sesquiterpenes may be formed [56]. Among these TS genes, only three (GLAREA11903, GLAREA03340, GLAREA08044) were located in gene clusters (Additional file 1: Figure S1). Two DMATS genes

Figure 4 KEGG functional classification of proteins in the *G. lozoyensis* genome. Distribution of the predicted proteins were assigned by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The top four categories in the KEGG functional classification were Carbon Metabolism, Energy Metabolism, Amino Acid Metabolism, and Infectious Diseases.

were found in the *G. lozoyensis* genome, and one gene (GLAREA04251) was located in a cluster downstream of another core PKS gene GLPKS9, signifying that a polyketide linked with dimethylallyl tryptophan may be the cluster's end product.

Identification of GLNRPS4 involving in pneumocandin biosynthesis in *G. lozoyensis*

The lipohexapeptide pneumocandin consists of two key components: a six-amino acid cyclic peptide and a 10,12-dimethylmyristoyl polyketide side chain [57]. Even though no such products are currently known from functionally characterized PKS-PKS hydrids [38], it is reasonable to consider that pneumocandins might be encoded by one of the PKS-NRPS hybrid proteins. However, in echinocandin B, the lipid side chain was thought to be derived from the cytoplasmic fatty acid pool [24]. Furthermore, domain analysis precluded the five PKS-NRPS hybrid proteins from pneumocandin biosynthesis because the hybrids contained only one A-T-C module, which could only incorporate one amino acid residue in the polyketide chain (Figure 7). Domain analysis of the six NRPS proteins showed that locus GLAREA10035 contained a NRPS with six A-T-C modules (designated as *glnrps4* and boxed) (Figure 7). Therefore, locus GLAREA10035 was the only plausible candidate. GLNRPS4, inferred to be responsible for the biosynthesis of the cyclic-hexapeptide core of the pneumocandins, comprised 7,192 amino acids and was

encoded by a gene with two introns (Additional file 1: Figure S3). GLNRPS4 encompassed 20 domains grouped into six modules each corresponding to one of the six amino acid incorporated monomers (Figure 7). The first module of GLNRPS4 had a unique T-C-A-T-C domain structure that differed from the other five modules which contained A-T-C domain structures. Two bioinformatics programs were used for substrate prediction, and both predicted that the third module encoded for proline [58,59]. However, neither program consistently predicted substrate specificities for the other five modules.

Analysis of the PKS-NRPS gene cluster for pneumocandin biosynthesis

Gene analysis of 50 kb of DNA flanking GLNRPS4 revealed a typical gene cluster for fungal secondary metabolite biosynthesis (Figure 8a). Immediately upstream of GLNRPS4 was the *glpks4* gene which encodes a PKS of 2,531 amino acids with eight introns (Additional file 1: Figure S3). Moreover, the PKS encoded by *glpks4* contained a methyltransferase domain that would be required for the biosynthesis of methyl group-containing fungal polyketides; the pneumocandin polyketide side chain contains two methyl groups (Figure 1a, Additional file 1: Figure S3) [5,57]. In addition to GLNRPS4 and GLPKS4, two other genes in this cluster stood out, GLAREA10021 encoding an acyltransferase and GLAREA10043 encoding an acyl-CoA ligase (Figure 8a). Labeling experiments at Merck revealed that GLPKS4 assembled a myristate from an acetyl starter,
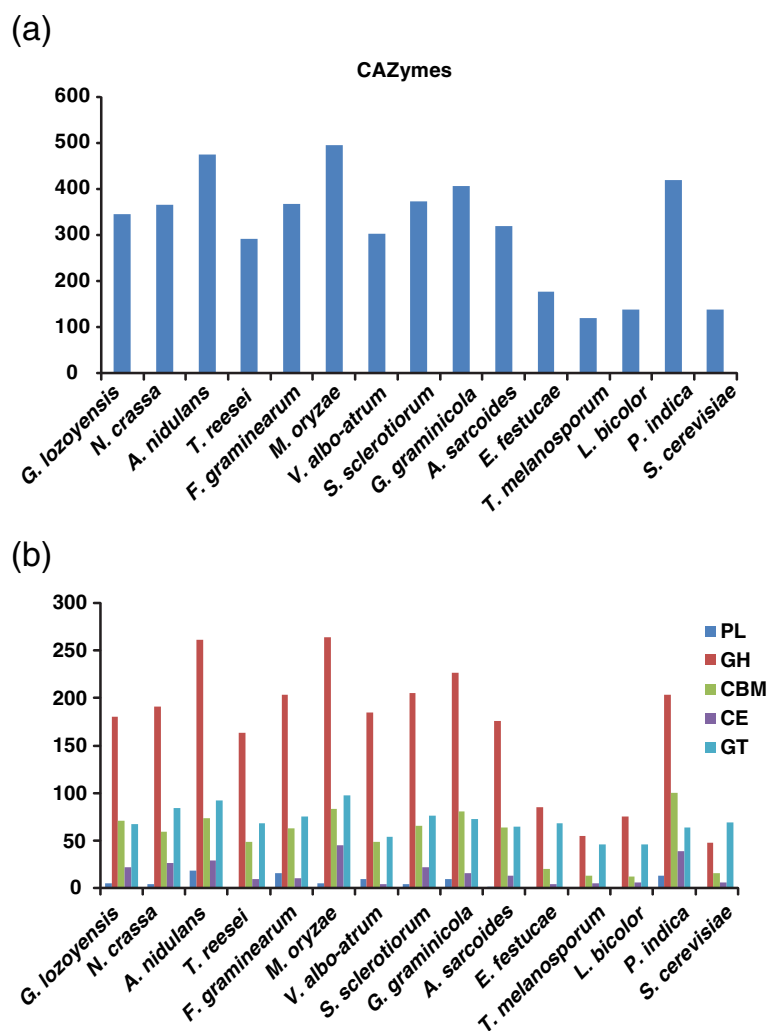
Figure 5 CAZymes (carbohydrate-active enzymes) analysis in the *G. lozoyensis* genome and other fungi. (a) Total number of CAZymes in different fungi (*Glarea lozoyensis, Neurospora crassa, Aspergillus nidulans, Trichoderma reesei, Fusarium graminearum, Aspergillus oryzae, Verticillium albo-atrum, Sclerotinia sclerotiorum, Glomerella graminicola, Ascocoryne sarcoides, Epichloë festucae, Tuber melanosporum, Laccaria bicolor, Piriformospora indica*, and *Saccharomyces cerevisiae*). (**b**) Number of different family of CAZymes in different fungi. PL: polysaccharide lyase, GH: glycoside hydrolase, CBM: carbohydrate-binding module, CE: carbohydrate esterases, and GT: glycosyltransferase. See Additional file 2: Table S2 for a detailed tabular summary.

whereas methionines provided two methyl groups to form the 10,12-dimethylmyristoyl side chain [18]. Although functional characterizations will be necessary to define how each gene contributes to the biosynthetic mechanism, based on the above analyses and those of the echinocandin B and emericellamide pathways [24,60], a hypothetical model of the pneumocandin biosynthetic pathway can be formulated from the four genes, GLNRPS4, GLPKS4, acyltransferase (GLAREA10021), and acyl-CoA ligase (GLAREA10043). The model predicts that 10,12-dimethylmyristoyl side chain is released from GLPKS4 as a carboxylic acid that is converted to a CoA thioester by the acyl-CoA ligase (GLAREA10043), and then loaded onto the acyltransferase (GLAREA10021).

The polyketide intermediate could then be shuttled to the first thiolation (T) domain of GLNRPS4, followed by its acylation to 4,5-dihydroxyorinithine to trigger elongation of the cyclic hexapeptide. Like other fungal NRPS and PKS gene clusters, the *glpks4* and *glnrps4* are positioned within a cluster that contains genes encoding for one or more cytochrome P450s, clavaminate synthase-like proteins (oxygenases), zinc finger transcription factors, and an ABC transporter (Figure 8a). It has been demonstrated that proline 3-hydroxylase and proline 4-hydroxylase, which are members of the 2-oxoglutarate-dependent dioxygenase class, can convert proline to 3-hydroxyproline and 4-hydroxyproline [61]. Two of the four oxygenases (GLAREA10033, GLAREA10041, GLAREA10042, and

Figure 6 (See legend on next page.)

(See figure on previous page.)
Figure 6 Domain prediction and phylogenetic analysis of polyketide synthases (PKSs) and polyketide synthases-nonribosomal peptide synthetase hybrids (PKS-NRPS hybrids) in *G. lozoyensis* and other characterized fungal PKSs. PKS and PKS-NRPS domains from *G. lozoyensis* were annotated by SMURF, anti-SMASH and SWISS-MODEL tools. SAT, starter unit acyltransferase domain; KS, ketosynthase domain; AT, acyltransferase domain; PT, product template domain; DH, dehydratase domain; ER, enoylreductase domain; KR, β-ketoacylreductase domain; MT, methyltransferase domain; ACP, acyl carrier protein; TE, thioesterase domain; A, adenylation domain; T, thiolation domain; C, condensation domain; R, reductive domain. Genealogy of PKSs and PKS-NRPSs was inferred by neighbor-joining analysis of the aligned amino acid sequences of the KS domains. Classification of PKSs and PKS-NRPSs sharing a common domain organization are highlighted by gray shading. Branch length indicates number of inferred amino acid changes. Red dots indicate branch nodes with >60% support. PKSs from *G. lozoyensis* are marked in red. See in Additional file 2: Table S3 for details of gene designations and their corresponding metabolites and references.

GLAREA10044) in the gene cluster were presumed to be involved in proline conversion. Two cytochrome P450 monooxygenases (GLAREA10030 and GLAREA10031) were classified in the CYP 512A family by the P450 database (http://www.cyped.uni-stuttgart.de/) which might be responsible for the hydroxylation of the amino acids. These oxygenases were also presumably involved in an oxidative mechanism for the conversion of leucine to methyl proline [19]. The putative zinc finger transcription regulator (GLAREA10050) belongs to the C_2H_2 and C_2HC zinc finger superfamily which are DNA-binding proteins and transcription factors [62]. Some members of this family are pathway-specific transcription regulators of secondary metabolite biosynthesis, e.g., Rua1 that activates the ustilagic acid biosynthesis gene cluster in *Ustilago maydis* [63]. Therefore, the zinc finger protein GLAREA10050 most likely regulates transcription of the *glpks4* and *glnrps4* genes. ABC transporters are ubiquitous membrane proteins with the ability to pump a variety of substrate specificities

of endogenous and exogenous toxic compounds [64,65]. The ABC transporter (GLAREA10036) in the cluster possibly secretes antifungal pneumocandins, thus avoiding of intracellular accumulation and ameliorating the toxicity to the producing cells.

Finally, a putative biosynthetic pathway for L-homotyrosine, the non-proteinogenic amino acid in the pneumocandin peptide core's fourth position, sits downstream of GLNPRS4 (Figure 8a). This set of five contiguous genes showed significant identity to the L-homotyrosine pathway of *E. rugulosa* [24] (Figure 8b), although the direction of transcription was inverted in two of the five genes, and consisted of GLAREA10037, an aconitase (62% identity to *hyt*D), GLAREA 10038, an isopropyl malate dehydrogenase (71% identity to *hty*C), GLAREA10039, a 2-isopropyl malate synthase (63.7% identity to *hyt*A), and GLAREA10040, an aminotransferase (64% identity to *hyt*B), and GLAREA10041, a non-heme dioxygenase (63.8% identity to *hyt*E). However, unlike the L-homotyrosine
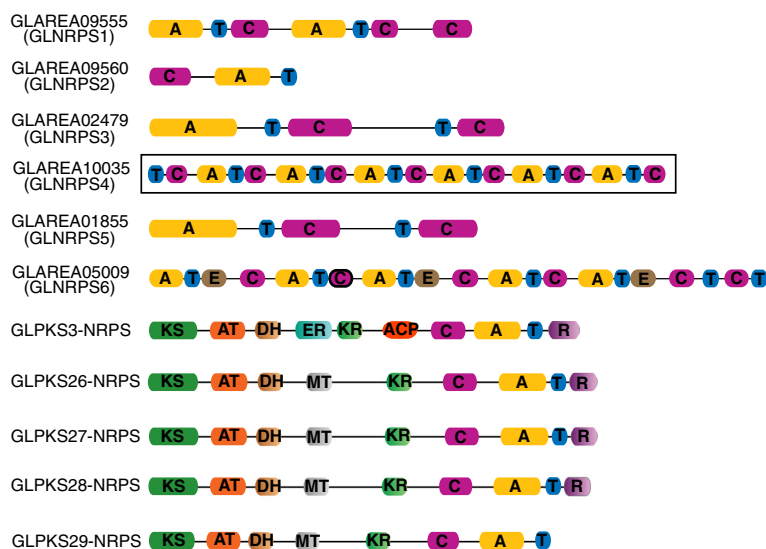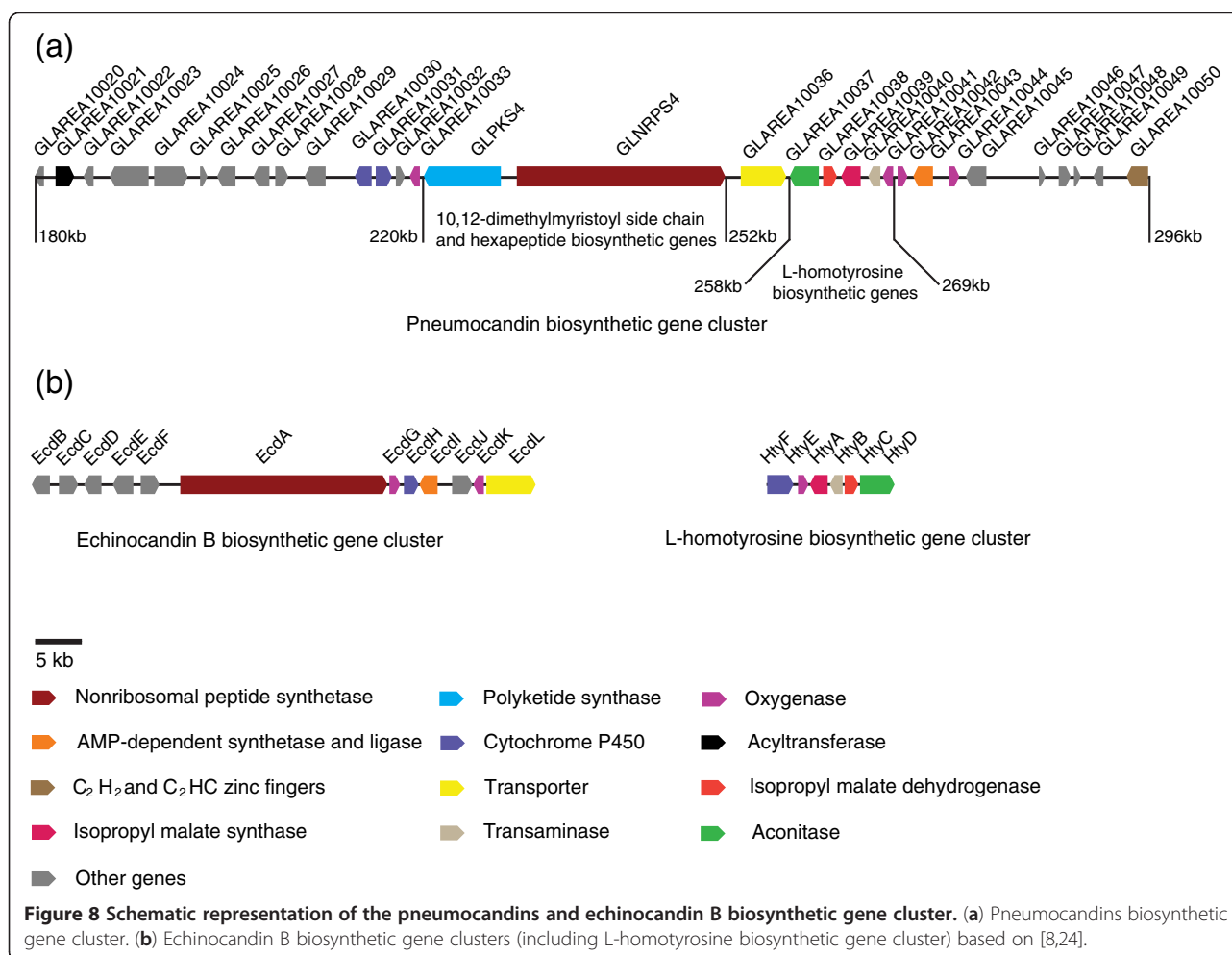

Figure 7 Schematic representations of the functional domains in nonribosomal peptide synthetase (NRPS) and polyketide synthase-nonribosomal peptide synthetase hybrid (PKS-NRPS hybrid) proteins in *G. lozoyensis*. A, adenylation domain; T, thiolation domain; C, condensation domain; E, epimerization domain; KS, ketosynthase domain; AT, acyltransferase domain; DH, dehydratase domain; ER, enoylreductase domain; KR, β-ketoacylreductase domain; MT, methyltransferase domain; ACP, acyl carrier protein; R, reductive domain. The pneumocandin-encoding NRPS (GLAREA10035 GLNPRS4) with six A-T-C modules is outlined.

Figure 8 Schematic representation of the pneumocandins and echinocandin B biosynthetic gene cluster. (**a**) Pneumocandins biosynthetic gene cluster. (**b**) Echinocandin B biosynthetic gene clusters (including L-homotyrosine biosynthetic gene cluster) based on [8,24].

pathway of *E. rugulosa*, the cytochrome P450 oxygenase gene corresponding to *hyt*F, was absent (Figure 8b).
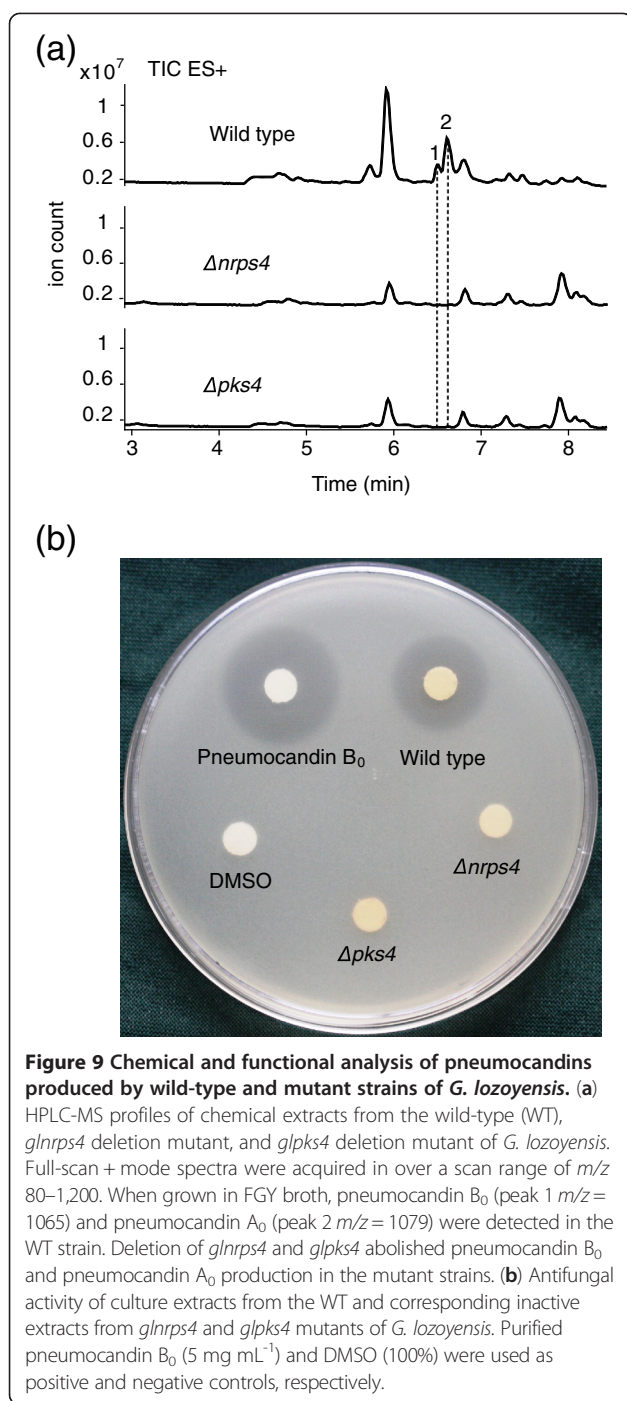
Functional analysis of *glpks4* and *glnrps4* in pneumocandin biosynthesis

To verify whether the gene cluster was responsible for pneumocandin biosynthesis, *glnrps4* and *glpks4* were knocked out by homologous replacement with an *Agrobacterium*-mediated transformation protocol developed previously for *G. lozoyensis*, and the deletions were verified by PCR analysis (Additional file 1: Figure S4b and S4c). Twelve and ten positive transformants were recovered for the GLNRPS4 and GLPKS4 knockouts, respectively. After growing the fungi in FGY medium and comparative analysis of the extracts by HPLC-MS using purified pneumocandin B_0 as a standard, the two major pneumocandins (A_0 and B_0) were produced by the *G. lozoyensis* WT strain as expected, but the pneumocandins were absent in the *glnrps4* and *glpks4* knockout mutants (Figure 9a). Consistent with earlier observations [5], the WT strain produced pneumocandin A_0 in larger quantities

than pneumocandin B_0 (Figure 9a, Additional file 1: Figure S4d). Antifungal assays showed that crude extracts from the WT strain caused zones of inhibition against the yeast *C. albicans*, whereas the crude extracts from mutants $\Delta glnrps4$ and $\Delta glpks4$ were inactive (Figure 9b). These results demonstrated that both *glnrps4* and *glpks4* were essential for biosynthesis of the pneumocandin core structure as predicted.

Discussion

Sequenced genomes are yielding substantial evidence for a richness of secondary metabolite pathways among the major kinds of fungi, well beyond that imagined to date, and the number of sequenced genomes is growing exponentially [66,67]. With the advance of next-generation sequencing technology, genome sequencing is evolving as an essential tool to decipher novel genes and gene clusters involved in biosynthesis of different metabolites in fungi [22,68]. For example, the biosynthetic pathway of the insecticidal cyclodepsipeptide destruxins was recently elucidated in the insect fungal pathogen *Metarhizium*

Figure 9 Chemical and functional analysis of pneumocandins produced by wild-type and mutant strains of *G. lozoyensis*. (**a**) HPLC-MS profiles of chemical extracts from the wild-type (WT), *glnrps4* deletion mutant, and *glpks4* deletion mutant of *G. lozoyensis*. Full-scan + mode spectra were acquired in over a scan range of *m/z* 80–1,200. When grown in FGY broth, pneumocandin B_0 (peak 1 *m/z* = 1065) and pneumocandin A_0 (peak 2 *m/z* = 1079) were detected in the WT strain. Deletion of *glnrps4* and *glpks4* abolished pneumocandin B_0 and pneumocandin A_0 production in the mutant strains. (**b**) Antifungal activity of culture extracts from the WT and corresponding inactive extracts from *glnrps4* and *glpks4* mutants of *G. lozoyensis*. Purified pneumocandin B_0 (5 mg mL^{-1}) and DMSO (100%) were used as positive and negative controls, respectively.

significantly higher than that of *A. sarcoides*, also of the Helotiaceae [28], and in the same order of magnitude as that of *B. cinerea*, *S. sclerotiorum*, and other sequenced Leotiomycetes [27,73]. Most of the ketosynthase domains of the 24 PKSs and five PKS-NRPS hybrids could be clustered with PKSs that were responsible for the biosynthesis of bioactive polyketides and polyketide-nonribosomal peptide hybrids (Figure 6). However, biosynthetic functions for only two of the 49 secondary metabolite-encoding genes in *G. lozoyensis* were previously validated (GLPKS1 for melanin and GLPKS2 for 6-methylsalicylic acid) [20,21]. Many secondary metabolites are fusions of nonribosomal peptides and polyketides, in which a PKSs and NRPSs interface and contribute to the same pathway end product [38,60,74]. Because the NRPS portion in each of the five PKS-NRPS hybrids in *G. lozoyensis* genome contains only one A-T-C module, one amino acid is predicted to be added to the polyketide produced by the PKS portion of the cluster, similar to ApdA in *A. nidulans* and ATEG00325 in *A. terreus*, which are involved in the biosynthesis of aspyridones and flavipucine, respectively [71,75].

Comparing the rich genetic potential for secondary metabolites in the *G. lozoyensis* genome, only pneumocandins were previously identified from the fungus. In an attempt to find additional chemistries, we identified isolecanoric acid and pseudogyrophoric acid as two new fermentation products of *G. lozoyensis* (Additional file 1: Figure S2). Therefore, majority of the secondary metabolites in *G. lozoyensis* remain to be characterized. Despite the advances in the field of microbial secondary metabolite biosynthesis, how the basic biology, ecology, and trophic strategies of microorganisms relate to their secondary metabolite production remains poorly understood. Application of efficient strategies to mine the metabolite-encoding gene clusters in *G. lozoyensis* and other poorly known fungi, while identifying their corresponding metabolites, presents a challenge and opportunity for natural products discovery.

GLNRPS4 and GLPKS4 are centrally located in the pneumocandin biosynthetic gene cluster, and how they cooperate with other genes in the cluster is still speculative. Even though they are independently transcribed and translated, their transcription is likely to be synchronized or co-regulated. The first module in GLNRPS4 has a unique T-C-A-T-C structure, and the first T domain in the T-C-A-T-C module is suggest to accept thiolated intermediates as found in emericellamide biosynthesis [60] or adenylated substrates similar to yersiniabactin biosynthesis [76]. Thus, the first T domain in the T-C-A-T-C module of GLNRPS4 could be responsible for accepting the incoming 10,12-dimethylmyristoyl side chain intermediate, whereas the second T domain would accept the 4,5-dihydroxyornithine adenylated by the module's A domain. Threonine, 4-hydroxyproline,

robertsii by genomic sequencing [69]. Genomic mining of several *Aspergillus* spp. has led to the elucidation of biosynthetic pathways of multiple bioactive compounds, including terrequinone A [70], emericellamide [60], aspyridones [71], pyripyropene A [72] and echinocandin B [24].

Genomic sequence analysis showed that *G. lozoyensis* has the potential to produce a diverse array of natural products. The genome was predicted to encode 49 gene clusters that contribute to its secondary metabolome,

4,5-dihydroxyhomotyrosine, 3-hydroxyglutamine and 3-hydroxyproline/3-hydroxy-4-methylproline would be sequentially added to the growing chain consistent with the *in silico* prediction that the A3 in GLNRPS4 is specific to proline [8,58,59]. Like many other NRPSs [60,77,78], the carboxyl terminal of GLNRPS4 lacks a thioesterase (TE) domain, suggesting that a dedicated TE is not required for pneumocandin cyclization. The last C domain of GLNRPS4 is proposed to be responsible for cyclization by condensation to form the peptide bond between 4,5-dihydroxyornithine and 3-hydroxyproline/3-hydroxy-4-methylproline. This proposal is consistent with the fact that the C domain has a HAEYD motif similar to the active site signature in the terminal C domain of cyclosporine synthetase (HSLYD) which is responsible for cyclization of cyclosporine in *Tolypocladium inflatum* and siderophore synthase SidC (HSLYD) involved in cyclization of the siderophore ferricrocin in *A. nidulans* [79,80]. The proposed biosynthetic sequence also parallels that proposed for echinocandin B [8,24]. Five of the six amino acids in the cyclic hexapeptide were hydroxylated, and hydroxylations of the two proline residues in pneumocandin B_0 were catalyzed by a proline-3-hydroxylase and a proline-4-hydoxylase [61]. The enzyme responsible for hydroxylation of 4-methylproline derived from leucine in pneumocandin A_0 may also be a proline 3-hydroxylase as 4-methylproline is an analogue of L-proline [19].

Other genes downstream of the GLNRPS4 that are likely involved the biosynthesis are the putative acyl-CoA ligase GLAREA10043 which shares 43% identity with EasD which converts polyketide carboxylic acid to a CoA thioester in emericellamide biosynthesis in *A. nidulans* [60]. The putative acyltransferase GLAREA10021 in the cluster shares more than 65% identity with the cholesterol acyltransferases from *Cordyceps militaris* [81]. Existence of these two genes suggests that the polyketide intermediate was first synthesized by GLPKS4, and then shuttled to the first T domain of GLNRPS4 mediated by the two enzymes, in a fashion similar to the emericellamide biosynthetic pathway [60]. Surprisingly and unlike the echinocandin B pathway [24], the putative pathway for the homotyrosine residue of the pneumocandin peptide core also sits downstream, and presumably L-homotyrosine biosynthesis is synchronized with the rest of the pathway.

The pneumocandin and echinocandin B pathways have some striking commonalities, yet obviously differ in their organization. The most obvious similarity is the high degree of identity between *ecdA* and *glnrps4* (60.8% identity over 22.7 kb, 55.2% identity over 7218 aa), and both have the same orientation in transcription and functional modules (TCATCATCATCATCATCATCT). Likewise, the genes of the L-homotyrosine pathway are highly similar, although their physical proximities to the core NRPS differ. Both pathways also contain a number

of oxygenases that, in the case of echinocandin B, tailor the multiple hydroxyl or diol groups of the amino acid core, but once again their physical location and order are significantly rearranged. However, the inclusion a PKS for side chain biosynthesis and its proximity for immediate loading onto the first thiolation domain, along with close proximity of the L-homotyrosine gene cluster and a possible zinc finger regulatory protein would likely confer greater metabolic autonomy to the pneumocandin pathway. The remarkable similarity between the echinocandin and pneumocandin pathways and especially the high degree of sequence homology between the amp-binding domains of GLNRPS4 and EcdA raises questions about pathway acquisition through horizontal gene transfer among fungi [82,83]. However, with only two echinocandin type pathways characterized thus far, speculation on why fungi from evolutionary lineages, Eurotiomycete (*E. rugulosa*) versus Leotiomycete (*G. lozoyensis*) that diverged 100s of millions of years ago, would share or converge on such similar molecular scaffolds is still premature. Elucidation of additional echinocandin type pathways in the Eurotiomycete, e.g., aculeacin and mulundocandin, and in the Leotiomycetes, e.g. FR901379 (WF11899A) and cryptocandin would yield evidence to determine a possible echinocandin progenitor and the probable directionality in gene recruitment or losses during the evolution of the echinocandin-pneumocandin gene clusters, as well as the significance of these potent cell wall-modifying metabolites to the fungi that produce them.

Elucidation of the pneumocandin biosynthetic pathway in *G. lozoyensis* paves the way for designing experimental procedures to enhance the production titer of the pneumocandins or engineering analogues with improved oral availability or broader spectrum of antifungal activities. Deletion of other PKS and NRPS genes could potentially reduce metabolic competition for substrates to GLPKS4 and GLNRPS4 and therefore increase the titers of pneumocandin B_0, in a manner similar to the disruption of GLPKS1 melanin gene in *G. lozoyensis* which doubled pneumocandin production titer [21]. Elimination, inactivation, addition or modification of the specificity of domains to GLPKS4 and GLNRPS4 could result in new pneumocandin derivatives via biocombinatorial chemistry approaches for the discovery and development of improved antifungal therapy.

Conclusion

The *Glarea lozoyensis* genome was sequenced, completely assembled and thoroughly annotated. The menu of secondary metabolites encoding genes was predicted from the genome, thus providing a greater understanding the complexity of primary and secondary metabolism in fungi from the yet poorly studied Leotiomycetes. The

biosynthetic gene cluster responsible for pneumocandin was predicted *in silico* and identified by core gene *glpks4* and *glnrps4* knockouts and bioassay experiments. The data from this study will form the basis for a more detailed functional analysis of pneumocandin biosynthetic pathways and enable the identification of other antifungal lipohexapeptide pathways in other fungi, of which both will be essential for increasing pneumocandin production and for generating new pneumocandin and echinocandin derivatives via biocombinatorial chemistry approaches.

Methods

Fungal and bacterial strains, vectors, and other reagents

The original pneumocandin producing strain of *G. lozoyensis* ATCC 20868 was obtained from American Type Culture Collection (ATCC) and was used as the wild-type recipient in *Agrobacterium*-mediated transformation experiments. The *Escherichia coli* strain DH5α was used in plasmid manipulations. *Agrobacterium tumefaciens* AGL-1 was described by Lazo *et al.* [84]. Plasmid pAg1-H3 was described by Zhang *et al.* [21], pEASY-T3 vector was from TransGen Biotech (Beijing, China), and pMD18-T vector was from Takara Biotech (Dalian, China). The pneumocandin B_0 standard was from Molcan Corporation (Ontario, Canada). LYCP-5 medium, FGY medium and conditions for *G. lozoyensis* fermentation were described by Connors *et al.* [85]. M-100 and IMAS mediums were described by Wang *et al.* [69]. Potato dextrose agar (PDA) and Sabouraud dextrose agar (SDA) were from Becton Dickinson (Franklin Lakes, New Jersey, USA). *E. coli* and *A. tumefaciens* AGL-1 were cultured as described by Zhang *et al.* [21]. Restriction endonucleases and DNA modifying enzymes were from New England Biolabs (Beverly, Massachusetts, USA).

DNA isolation and sequencing

Genomic DNA of *G. lozoyensis* ATCC 20868 was extracted as previously described by Zhang *et al.* [21]. Genomic DNA libraries with 500–800 bp inserts were constructed and sequenced with a Roche 454 GS FLX at the Chinese National Human Genome Center in Shanghai. A library with 3 kb inserts was constructed and sequenced with Illumina Genome Analyzer using the protocols as described for genomic sequencing of *Cordyceps militaris* [81]. The genome sequences were assembled using Newbler software (ver. 2.3) and SSPACE (http://www.baseclear.com/dna-sequencing/data-analysis/bioinformatics-tools/).

G. lozoyensis genome annotation, orthology and phylogenomic analyses

The *G. lozoyensis* genome was annotated with Augustus (http://bioinf.uni-greifswald.de/augustus) by referencing annotated genome of *Botrytis cinerea*. GeneID and GeneMark-ES were additionally used for open reading frames prediction in *G. lozoyensis* [86,87]. Repetitive sequences in the genome were identified by BLAST against the RepeatMasker library [88] and by *de novo* repetitive sequence search using RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html). Transfer RNAs (tRNAs) were identified with tRNAscan-SE [89]. Ribosomal RNAs (rRNAs) were predicted by a BLAST search with known rRNA modules from other fungal genomes. Whole genome protein families were classified by InterproScan analysis (http://www.ebi.ac.uk/interpro/) and BLAST against Kyoto Encyclopedia of Genes and Genomes (KEGG) database using KEGG Automatic Annotation Server (KAAS: http://www.genome.jp/kegg/kaas/). Carbohydrate-active enzymes from *G. lozoyensis* and reference fungi (Additional file 2: Table S2) were classified by local Blastp searching against a library of catalytic and carbohydrate-binding module enzymes [90]. PKS, NRPS, DMATS and related gene clusters were predicted by programs SMURF and anti-SMASH [68] and by manual annotation.

A total of 878 common orthologous genes were identified using the InParanoid pipeline in the selected fungal genomes (Additional file 2: Table S2) [91], and aligned with Clustal W (ver. 2.0). A maximum likelihood phylogenomic tree was created using the concatenated amino acid sequences in PAUP* 4.0 (beta 10 Win) with heuristic searches [92]. Characters were treated as unordered and gaps were regarded as missing data. Bootstrap support for internal branches was estimated by analysis of 1,000 pseudo replicates. Reference fungi used to construct the phylogenetic tree were described elsewhere (Additional file 2: Table S2 and ref. [26]). All internal transcribed spacer (ITS) sequences were aligned with Clustal W (ver. 2.0), and a neighbor-joining phylogenetic tree was generated with the program PAUP* 4.0 (beta 10 Win) using 1,000 bootstrap replicates and a Jukes-Cantor substitution model with pairwise deletion for gaps or missing data [92].

Phylogenetic analysis of PKS and PKS-NRPS genes

KS domains from fungi with PKS genes proven to be responsible for metabolites biosynthesis (Additional file 2: Table S3), PKSs and PKS-NRPS hybrids in *G. lozoyensis* (Additional file 2: Table S4) were identified by the program anti-SMASH [93] or visually in multiple alignments. All KS domains from PKS were aligned with Clustal X (ver. 2.0), and analyzed phylogenetically with MEGA 5.0 using a Jones-Taylor-Thornton substitution model, a pair-wise deletion for gaps or missing data, and a 1,000 bootstrap replications test [94]. The tree was rooted with the KS domain of the rat fatty acid synthase (Figure 6).

Gene knockout of *glnrps4* and *glpks4*

To verify function of the predicted pneumocandin gene cluster, *glnrps4* and *glpks4* deletion were conducted using method reported by Zhang *et al.* [21]. To verify function

of the predicted pneumocandin gene cluster, gene knock-out constructs for *glnrps4* and *glpks4* were created. Briefly, the flanking regions of the target genes were amplified using different primer pairs (Additional file 2: Table S5) and ligated into the binary vector of pAg1-H3 containing the hygromycin resistance gene to form pAg1-H3-nrps4 and pAg1-H3-pks4. The constructs were introduced into *G. lozoyensis* by *Agrobacterium*-mediated transformation using method reported by Zhang *et al.* [21] with slight modification. Conidia for transformation were harvested into sterile 0.05% Tween-20 followed with 2 times of wash with distilled water and then suspended into 0.5-1.0 mL sterile water (10^6 spores mL^{-1}). One hundred microliters of *G. lozoyensis* and 100 μL of *A. tumefaciens* (OD$_{660 nm}$ = 0.6-0.8) were mixed and spread on the IMAS agar plate and co-incubated at 28°C for 2 d. The co-culture of *A. tumefaciens* and *G. lozoyensis* was covered with M-100 supplemented with 300 μg mL^{-1} cefotaxime and 100 μg mL^{-1} hygromycin B, and incubated at 25°C for 2–3 weeks before isolating hygB resistant colonies. The transformants were purified by single conidium isolation and the gene knockout transformants were verified by PCR using multiple primers (Additional file 2: Table S5).

HPLC-MS analysis of pneumocandins

Fermentation and pneumocandin extraction protocols were described by Petersen *et al.* [95]. HPLC separation was performed on an Agilent Zorbax Extend-C18 1.8 μm 2.1 × 50 mm column using an Agilent 1200 Series system (Agilent, USA). The total flow rate was 0.3 mL min^{-1}; mobile phase A was with 0.1% formic acid and mobile phase B was acetonitrile. The total elution program was 25 min. Gradient elution began with 30% B for 0.5 min, changed to 70% B over 3.5 min, changed to 100% B over 8 min, maintained at 100% B for 5 min, to 30% B over 0.5 min, and re-stabilized for 7.5 min prior the next injection. The column temperature was maintained at 40°C. The injection volume was 10 μL.

Mass spectra were acquired with an Agilent Accurate-Mass Quadrupole–Time-of-Flight mass spectrometry (Q-TOF/MS) 6520 system in the positive ionization mode. For Q-TOF/MS conditions, fragmentor and capillary voltages were kept at 130 and 3,500 V, respectively. Nitrogen was supplied as the nebulizing and drying gas. Temperature of the drying gas was set at 30°C. The flow rate of the drying gas and the pressure of the nebulizer were 10 L min^{-1} and 25 psi, respectively. Full-scan spectra were acquired over a scan range of *m/z* 80–1,200 at 1.03 spectra s^{-1}.

Candida albicans zone of inhibition (ZOI) assays

Antifungal activity of the WT, *glnrps4* and *glpks4* gene deletion mutants of *G. lozoyensis* was measured by a zone of inhibition assay against the human fungal pathogen *Candida albicans* SC 5314. Ten-mL liquid culture of the wild-type or *glnrps4* and *glpks4* gene deletion mutants of *G. lozoyensis* were lyophilized in a vacuum freeze dryer, and 10 mL methanol were added and thoroughly mixed. After 1 h of orbital shaking, the mixtures were first centrifuged at low speed, the supernatant was transferred to glass tubes, and then DMSO (2 mL) was added to solubilize any metabolites precipitated during evaporation. The samples were concentrated to 2 mL under a warm N$_2$ stream during orbital shaking. The final samples were 5× whole broth equivalents including 100% DMSO relative to original culture volume.

Candida albicans SC 5314 cells grown on SDA plates were inoculated into 10 mL of Sabouraud dextrose broth and incubated overnight at 30°C. The *C. albicans* suspension was adjusted to an optical density of 0.4 at 660 nm and added to SDA in the proportion of 30 mL L^{-1}. Twenty-mL aliquots of the seeded agar media were poured into 9-cm Petri plates. Pneumocandin B$_0$ (5 mg mL^{-1}) and 100% DMSO were used as positive and negative controls. The extracts prepared from liquid culture of *G. lozoyensis* and the controls (10 μL) were applied to paper discs on the surface of the seeded assay plates. The plates were incubated at 30°C for approximately 20 h and ZOIs were measured and photographed.

Production, purification and identification of isolecanoric and pseudogyrophoric acids

Isolecanoric acid and the new compound pseudogyrophoric acid were isolated from the extract of *G. lozoyensis* ATCC 20868 grown in MV8 medium (V8 juice 200 mL, maltose 75 g, soy flour 1 g, L-proline 3 g, MES 16.2 g, distilled H$_2$O 800 mL) at 22°C on a rotary shaker at 220 rpm for 14 d. The isolation procedure, mass spectra are summarized in Additional file 1: Figure S2.

Additional files

Additional file 1: Figures that provide support information for the main text. Figure S1. provides the 49 secondary metabolite biosynthetic gene clusters in the *G. lozoyensis* genome. **Figure S2** lists materials and methods for purification and characterization of additional metabolites from *G. lozoyensis* ATCC 20868 grown on MV8 medium. **Figure S3** shows gene structure of GLNRPS4 and GLPKS4. **Figure S4** summarizes strategy for the construction of GLPKS4 and GLNRPS4 gene deletion mutants.

Additional file 2: Tables that provide support information for the main text. Table S1 shows the homologous analysis of terpene gene in *Glarea lozoyensis*. **Table S2** lists the fungal genomes used for phylogenomic analyses and CAZymes analysis. **Table S3** provides supporting data for Figure 6 and the list of 71 functionally characterized fungal PKSs and PKS-NRPSs hydrids, their gene designations, the principal pathway end product, and references. **Table S4** shows polyketide synthases (PKSs) in the *G. lozoyensis* genome used for phylogenetic tree construction. **Table S5** summarizes primers used for genes deletion and mutants verification.

Abbreviations
WT: Wild-type; NRPS: Non-ribosomal peptide synthetase; PKS: Polyketide synthase; Mb: Megabase; CAZymes: Carbohydrate active enzymes; GH: Glycoside hydrolases; GT: Glycosyl transferases; PL: Polysaccharide lyases; CE: Carbohydrate esterases; CBM: Carbohydrate binding modules; TS: Terpene synthase; DMATS: Dimethylallyl tryptophane synthase; PKS-NRPS hybrids: Polyketide synthase-nonribosomal peptide synthetase hybrids; KS: Ketosynthase domains; HR-PKS: Highly reduced PKS; NR-PKS: Non reduced PKS; A: Adenylation; T: Thiolation; C: Condensation; TE: Thioesterase; ITS: Internal transcribed spacer; ZOI: Zone of inhibition.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
ZA and XL designed research; LC, QY, F O-L and MX performed research; LC, XZ, CW, SL, YC and GB analyzed data; and LC, QY, GB, ZA, and XL wrote the paper. All authors read and approved the final manuscript.

Accession number
The Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ALVE00000000 *Glarea lozoyensis* ATCC 20868.

Author details
[1]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China. [2]University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China. [3]Key Laboratory of Insect Developmental and Evolutionary Biology, Institute of Plant physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, People's Republic of China. [4]Beijing Institute of Pharmacology & Toxicology, Beijing 100850, People's Republic of China. [5]Fundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía, Granada 18100, Spain. [6]Texas Therapeutics Institute, the Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

References
1. Denning DW: **Echinocandin antifungal drugs.** *Lancet* 2003, **362**:1142–1151.
2. Kurtz MB, Heath IB, Marrinan J, Dreikorn S, Onishi J, Douglas C: **Morphological effects of lipopeptides against *Aspergillus fumigatus* correlate with activities against (1,3)-beta-D-glucan synthase.** *Antimicrob Agents Chemother* 1994, **38**:1480–1489.
3. Cross SA, Scott LJ: **Micafungin: a review of its use in adults for the treatment of invasive and oesophageal candidiasis, and as prophylaxis against *Candida* infections.** *Drugs* 2008, **68**:2225–2255.
4. Raasch RH: **Anidulafungin: review of a new echinocandin antifungal agent.** *Exp Rev Anti-infec Ther* 2004, **2**:499–508.
5. Schwartz RE, Sesin DF, Joshua H, Wilson KE, Kempf AJ, Goklen KA, Kuehner D, Gailliot P, Gleason C, White R, *et al*: **Pneumocandins from *Zalerion arboricola*. I. Discovery and isolation.** *J Antibiot* 1992, **45**:1853–1866.
6. Nyfeler R, Keller-Schierlein W: **Metabolites of microorganisms. 143. Echinocandin B, a novel polypeptide-antibiotic from *Aspergillus nidulans* var. *echinulatus*: isolation and structural components.** *Helv Chim Acta* 1974, **57**:2459–2477.
7. Iwamoto T, Fujie A, Sakamoto K, Tsurumi Y, Shigematsu N, Yamashita M, Hashimoto S, Okuhara M, Kohsaka M: **WF11899A, B and C, novel antifungal lipopeptides. I. Taxonomy, fermentation, isolation and physico-chemical properties.** *J Antibiot* 1994, **47**:1084–1091.
8. Jiang W, Cacho RA, Chiou G, Garg NK, Tang Y, Walsh CT: **EcdGHK are three tailoring iron oxygenases for amino acid building blocks of the echinocandin scaffold.** *J Am Chem Soc* 2013, **135**:4457–4466.
9. Ramos A, Cuervas-Mons V, Noblejas A, Baños I, Duran P, Marcos R, Sánchez-Turrión V, Jiménez M, Arellano B, Corbacho C: **Breakthrough rhinocerebral mucormycosis in a liver transplant patient receiving caspofungin.** *Transplant Proc* 2009, **41**:1972–1975.
10. Madureira A, Bergeron A, Lacroix C, Robin M, Rocha V, de Latour RP, Ferry C, Devergie A, Lapalu J, Gluckman E: **Breakthrough invasive aspergillosis in allogeneic haematopoietic stem cell transplant recipients treated with caspofungin.** *Int J Antimicrob Agents* 2007, **30**:551–554.
11. Perlin DS: **Resistance to echinocandin-class antifungal drugs.** *Drug Resist Updat* 2007, **10**:121–130.
12. Thompson GR, Wiederhold NP, Vallor AC, Villareal NC, Lewis JS II, Patterson TF: **Development of caspofungin resistance following prolonged therapy for invasive candidiasis secondary to *Candida glabrata* infection.** *Antimicrob Agents Chemother* 2008, **52**:3783–3785.
13. Lee KK, MacCallum DM, Jacobsen MD, Walker LA, Odds FC, Gow NAR, Munro CA: **Elevated cell wall chitin in *Candida albicans* confers echinocandin resistance *in vivo*.** *Antimicrob Agents Chemother* 2012, **56**:208–217.
14. Plaine A, Walker L, Da Costa G, Mora-Montes HM, McKinnon A, Gow NAR, Gaillardin C, Munro CA, Richard ML: **Functional analysis of *Candida albicans* GPI-anchored proteins: roles in cell wall integrity and caspofungin sensitivity.** *Fungal Gen Biol* 2008, **45**:1404–1414.
15. Walker LA, Munro CA, De Bruijn I, Lenardon MD, McKinnon A, Gow NAR: **Stimulation of chitin synthesis rescues *Candida albicans* from echinocandins.** *PloS Path* 2008, **4**:e1000040.
16. Balkovec JM, Black RM, Hammond ML, Heck JV, Zambias RA, Abruzzo G, Bartizal K, Kropp H, Trainor C, Schwartz RE, *et al*: **Synthesis, stability, and biological evaluation of water-soluble prodrugs of a new echinocandin lipopeptide - Discovery of a potential clinical agent for the treatment of systemic candidiasis and *Pneumocystis carinii* pneumonia (Pcp).** *J Med Chem* 1992, **35**:194–198.
17. Yao J, Liu H, Zhou T, Chen H, Miao Z, Sheng C, Zhang W: **Total synthesis and structure–activity relationships of new echinocandin-like antifungal cyclolipohexapeptides.** *Eur J Med Chem* 2012, **50**:196–208.
18. Adefarati A, Giacobbe R, Hensens O, Tkacz J: **Biosynthesis of L-671,329, an echinocandin-type antibiotic produced by *Zalerion arboricola*: origins of some of the unusual amino acids and the dimethylmyristic acid side chain.** *J Am Chem Soc* 1991, **113**:3542–3545.
19. Adefarati AA, Hensens OD, Jones E, Tkacz J: **Pneumocandins from *Zalerion arboricola*. V. Glutamic acid-and leucine-derived amino acids in pneumocandin A$_0$ (L-671,329) and distinct origins of the substituted proline residues in pneumocandins A$_0$ and B$_0$.** *J Antibiot* 1992, **45**:1953–1957.
20. Lu P, Zhang A, Dennis LM, Dahl-Roshak AM, Xia YQ, Arison B, An Z, Tkacz JS: **A gene (pks2) encoding a putative 6-methylsalicylic acid synthase from *Glarea lozoyensis*.** *Mol Gen Genom* 2005, **273**:207–216.
21. Zhang A, Lu P, Dahl-Roshak AM, Paress PS, Kennedy S, Tkacz JS, An Z: **Efficient disruption of a polyketide synthase gene (pks1) required for melanin synthesis through Agrobacterium-mediated transformation of *Glarea lozoyensis*.** *Mol Gen Genom* 2003, **268**:645–655.
22. Brakhage AA: **Regulation of fungal secondary metabolism.** *Nat Rev Microbiol* 2013, **11**:21–32.
23. Youssar L, Gruning BA, Erxleben A, Gunther S, Huttel W: **Genome sequence of the fungus *Glarea lozoyensis*: the first genome sequence of a species from the Helotiaceae family.** *Euk Cell* 2012, **11**:250–250.
24. Cacho RA, Jiang W, Chooi YH, Walsh CT, Tang Y: **Identification and characterization of the echinocandin B biosynthetic gene cluster from *Emericella rugulosa* NRRL 11440.** *J Am Chem Soc* 2012, **134**:16781–16790.
25. Bills GF, Platas G, Peláez F, Masurekar P: **Reclassification of a pneumocandin-producing anamorph, *Glarea lozoyensis* gen. et sp. nov., previously identified as *Zalerion arboricola*.** *Mycol Res* 1999, **103**:179–192.
26. Peláez F, Collado J, Platas G, Overy DP, Martín J, Vicente F, Gónzlez del Val A, Basilio A, De la Cruz M, Tormo JR, *et al*: **Phylogeny and intercontinental distribution of the pneumocandin-producing anamorphic fungus *Glarea lozoyensis*.** *Mycology* 2011, **2**:1–17.
27. Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, Coutinho PM, de Vries RP, Dyer PS, Fillinger S: **Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*.** *PloS Gen* 2011, **7**:e1002230.

28. Gianoulis TA, Griffin MA, Spakowicz DJ, Dunican BF, Sboner A, Sismour AM, Kodira C, Egholm M, Church GM, Gerstein MB: **Genomic analysis of the hydrocarbon-producing, cellulolytic, endophytic fungus *Ascocoryne sarcoides*.** *PloS Gen* 2012, **8**:e1002558.

29. Fisher PJ, Davey RA, Webster J: **Degradation of lignin by aquatic and aero-aquatic hyphomycetes.** *Trans Brit Mycol Soc* 1983, **80**:166–168.

30. Fisher PJ, Sharma PD, Webster J: **Cellulolytic ability of aero-aquatic hyphomycetes.** *Trans Brit Mycol Soc* 1977, **69**:495–496.

31. Battaglia E, Benoit I, van den Brink J, Wiebenga A, Coutinho PM, Henrissat B, Vries RP: **Carbohydrate-active enzymes from the zygomycete fungus *Rhizopus oryzae*: a highly specialized approach to carbohydrate degradation depicted at genome level.** *BMC Genom* 2011, **12**:38.

32. van den Brink J, de Vries RP: **Fungal enzyme sets for plant polysaccharide degradation.** *Appl Microbiol Biotechnol* 2011, **91**:1477–1492.

33. Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, de Vries RP, Record E, Levasseur A, Baker SE, *et al*: **Genome sequence of the model mushroom *Schizophyllum commune*.** *Nat Biotechnol* 2010, **28**:957–U910.

34. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics.** *Nucleic Acids Res* 2009, **37**:D233–D238.

35. Yin YB, Mao XZ, Yang JC, Chen X, Mao FL, Xu Y: **dbCAN: a web resource for automated carbohydrate-active enzyme annotation.** *Nucleic Acids Res* 2012, **40**:W445–W451.

36. 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).** *Nat Biotechnol* 2008, **26**:553–560.

37. Boettger D, Hertweck C: **Molecular diversity sculpted by fungal PKS-NRPS hybrids.** *Chem Bio Chem* 2013, **14**:28–42.

38. Boettger D, Bergmann H, Kuehn B, Shelest E, Hertweck C: **Evolutionary imprint of catalytic domains in fungal PKS-NRPS hybrids.** *Chem Bio Chem* 2012, **13**:2363–2373.

39. Sims JW, Fillmore JP, Warner DD, Schmidt EW: **Equisetin biosynthesis in *Fusarium heterosporum*.** *Chem Commun* 2005, **14**:186–188.

40. Hendrickson L, Davis CR, Roach C, Nguyen DK, Aldrich T, McAda PC, Reeves CD: **Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene.** *Chem Biol* 1999, **6**:429–439.

41. Cox RJ, Glod F, Hurley D, Lazarus CM, Nicholson TP, Rudd BAM, Simpson TJ, Wilkinson B, Zhang Y: **Rapid cloning and expression of a fungal polyketide synthase gene involved in squalestatin biosynthesis.** *Chem Commun* 2004, **10**:2260–2261.

42. Yang G, Rose MS, Turgeon BG, Yoder O: **A polyketide synthase is required for fungal virulence and production of the polyketide T-toxin.** *Plant Cell Online* 1996, **8**:2139–2150.

43. Baker SE, Kroken S, Inderbitzin P, Asvarak T, Li BY, Shi L, Yoder OC, Turgeon BG: **Two polyketide synthase-encoding genes are required for biosynthesis of the polyketide virulence factor, T-toxin, by *Cochliobolus heterostrophus*.** *Mol Plant-Microbe Interac* 2006, **19**:139–149.

44. Proctor RH, Desjardins AE, Plattner RD, Hohn TM: **A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuro*i mating population A.** *Fungal Gen Biol* 1999, **27**:100–112.

45. Kasahara K, Miyamoto T, Fujimoto T, Oguri H, Tokiwano T, Oikawa H, Ebizuka Y, Fujii I: **Solanapyrone synthase, a possible diels-alderase and iterative type I polyketide synthase encoded in a biosynthetic gene cluster from *Alternaria solani*.** *Chem Bio Chem* 2010, **11**:1245–1252.

46. Beck J, Ripka S, Siegner A, Schiltz E, Schweizer E: **The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum* - Its gene structure relative to that of other polyketide synthases.** *Eur J Biochem* 1990, **192**:487–498.

47. Fujii I, Ono Y, Tada H, Gomi K, Ebizuka Y, Sankawa U: **Cloning of the polyketide synthase gene atX from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression.** *Mol Gen Genet* 1996, **253**:1–10.

48. Fulton T, Ibrahim N, Losada M, Grzegorski D, Tkacz J: **A melanin polyketide synthase (PKS) gene from *Nodulisporium* sp. that shows homology to the pks1 gene of *Colletotrichum lagenarium*.** *Mol Gen Genet* 1999, **262**:714–720.

49. Takano Y, Kubo Y, Shimizu K, Mise K, Okuno T, Furusawa I: **Structural analysis of PKS1, a polyketide synthase gene involved in melanin biosynthesis in *Colletotrichum lagenarium*.** *Mol Gen Genet* 1995, **249**:162–167.

50. Chooi YH, Cacho R, Tang Y: **Identification of the viridicatumtoxin and griseofulvin gene clusters from *Penicillium aethiopicum*.** *Chem Biol* 2010, **17**:483–494.

51. Yu JH, Leonard TJ: **Sterigmatocystin biosynthesis in *Aspergillus nidulans* requires a novel type I polyketide synthase.** *J Bacteriol* 1995, **177**:4792–4800.

52. Schroeckh V, Scherlach K, Nutzmann HW, Shelest E, Schmidt-Heck W, Schuemann J, Martin K, Hertweck C, Brakhage AA: **Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*.** *Proc Nat Acad Sci USA* 2009, **106**:14558–14563.

53. Sanchez JF, Chiang YM, Szewczyk E, Davidson AD, Ahuja M, Oakley CE, Bok JW, Keller N, Oakley BR, Wang CCC: **Molecular genetic analysis of the orsellinic acid/F9775 gene cluster of *Aspergillus nidulans*.** *Mol Biosyst* 2010, **6**:587–593.

54. Reeves CD, Hu Z, Reid R, Kealey JT: **Genes for the biosynthesis of the fungal polyketides hypothemycin from *Hypomyces subiculosus* and radicicol from *Pochonia chlamydosporia*.** *Appl Environ Microbiol* 2008, **74**:5121–5129.

55. Zhou H, Qiao KJ, Gao ZZ, Meehan MJ, Li JWH, Zhao XL, Dorrestein PC, Vederas JC, Tang Y: **Enzymatic synthesis of resorcylic acid lactones by cooperation of fungal iterative polyketide synthases involved in hypothemycin biosynthesis.** *J Am Chem Soc* 2010, **132**:4530–4531.

56. Keller NP, Turner G, Bennett JW: **Fungal secondary metabolism - From biochemistry to genomics.** *Nat Rev Microbiol* 2005, **3**:937–947.

57. Masurekar PS, Fountoulakis JM, Hallada TC, Sosa MS, Kaplan L: **Pneumocandins from *Zalerion arboricola*. II. Modification of product spectrum by mutation and medium manipulation.** *J Antibiot* 1992, **45**:1867–1874.

58. Prieto C, Garcia-Estrada C, Lorenzana D, Martin JF: **NRPSsp: non-ribosomal peptide synthase substrate predictor.** *Bioinformatics* 2012, **28**:426–427.

59. Rottig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O: **NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity.** *Nucleic Acids Res* 2011, **39**:W362–W367.

60. Chiang YM, Szewczyk E, Nayak T, Davidson AD, Sanchez JF, Lo HC, Ho WY, Simityan H, Kuo E, Praseuth A: **Molecular genetic mining of the *Aspergillus* secondary metabolome: Discovery of the emericellamide biosynthetic pathway.** *Chem Biol* 2008, **15**:527–532.

61. Petersen L, Olewinski R, Salmon P, Connors N: **Novel proline hydroxylase activities in the pneumocandin-producing fungus *Glarea lozoyensis* responsible for the formation of trans 3-and trans 4-hydroxyproline.** *Appl Microbiol Biotechnol* 2003, **62**:263–267.

62. MacPherson S, Larochelle M, Turcotte B: **A fungal family of transcriptional regulators: The zinc cluster proteins.** *Microbiol Mol Biol Rev* 2006, **70**:583–604.

63. Teichmann B, Liu L, Schink KO, Bölker M: **The C_2H_2 zinc finger transcription factor Rua1 activates the ustilagic acid biosynthesis gene cluster in *Ustilago maydis*.** *Appl Environ Microbiol* 2010, **76**:2633–2640.

64. Kovalchuk A, Driessen A: **Phylogenetic analysis of fungal ABC transporters.** *BMC Genom* 2010, **11**:177.

65. Linton KJ: **Structure and function of ABC transporters.** *Physiology* 2007, **22**:122–130.

66. Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, Kuske C, Magnuson JK, Martin F, Spatafora JW, *et al*: **Fueling the future with fungal genomics.** *Mycology* 2011, **2**:192–209.

67. An Z, Wang C, Liu X, Bennett JW: **China's fungal genomics initiative: a whitepaper.** *Mycology* 2010, **1**:1–8.

68. Fedorova ND, Moktali V, Medema MH: **Bioinformatics approaches and software for detection of secondary metabolic gene clusters.** *Meth Mol Biol* 2012, **944**:23–45.

69. Wang B, Kang QJ, Lu YZ, Bai LQ, Wang CS: **Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species.** *Proc Nat Acad Sci USA* 2012, **109**:1287–1292.

70. Bok JW, Hoffmeister D, Maggio-Hall LA, Murillo R, Glasner JD, Keller NP: **Genomic mining for *Aspergillus* natural products.** *Chem Biol* 2006, **13**:31–37.

71. Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C: **Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*.** *Nat Chem Biol* 2007, **3**:213–217.

72. Itoh T, Tokunaga K, Matsuda Y, Fujii I, Abe I, Ebizuka Y, Kushiro T: **Reconstitution of a fungal meroterpenoid biosynthesis reveals the

involvement of a novel family of terpene cyclases. *Nat Chem* 2010, **2**:858–864.

73. Gao Q, Jin K, Ying SH, Zhang YJ, Xiao GH, Shang YF, Duan ZB, Hu XA, Xie XQ, Zhou G, *et al*: **Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi** *Metarhizium anisopliae* **and** *M. acridum*. *PloS Gen* 2011, **7**:e1001264.

74. Fischbach MA, Walsh CT: **Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms.** *Chem Rev* 2006, **106**:3468–3496.

75. Gressler M, Zaehle C, Scherlach K, Hertweck C, Brock M: **Multifactorial induction of an orphan PKS-NRPS gene cluster in** *Aspergillus terreus*. *Chem Biol* 2011, **8**:198–209.

76. Miller DA, Luo LS, Hillson N, Keating TA, Walsh CT: **Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of** *Yersinia pestis*. *Chem Biol* 2002, **9**:333–344.

77. Bushley KE, Turgeon BG: **Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships.** *BMC Evol Biol* 2010, **10**:26.

78. Marahiel MA, Essen LO: **Nonribosomal Peptide Synthetases. Mechanistic and Structural Aspects of Essential Domains.** *Methods Enzymol* 2009, **458**:337–351.

79. 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.** *Curr Genet* 1994, **26**:120–125.

80. 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).** *Mol Microbiol* 2003, **49**:359–375.

81. Zheng P, Xia Y, Xiao G, Xiong C, Hu X, Zhang S, Zheng H, Huang Y, Zhou Y, Wang S: **Genome sequence of the insect pathogenic fungus** *Cordyceps militaris*, **a valued traditional Chinese medicine.** *Genom Biol* 2011, **12**:R116.

82. 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 Evol Biol* 2007, **7**:174.

83. Fitzpatrick DA: **Horizontal gene transfer in fungi.** *FEMS Microbiol Lett* 2012, **329**:1–8.

84. Lazo GR, Stein PA, Ludwig RA: **A DNA transformation–competent** *Arabidopsis* **genomic library in** *Agrobacterium*. *Nat Biotechnol* 1991, **9**:963–967.

85. Connors N, Petersen L, Hughes R, Saini K, Olewinski R, Salmon P: **Residual fructose and osmolality affect the levels of pneumocandins B$_0$ and co-produced by** *Glarea lozoyensis*. *Appl Microbiol Biotechnol* 2000, **54**:814–818.

86. Parra G, Blanco E, Guigó R: **Geneid in Drosophila.** *Genome Res* 2000, **10**:511–515.

87. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M: **Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training.** *Genome Res* 2008, **18**:1979–1990.

88. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase Update, a database of eukaryotic repetitive elements.** *Cytogenet Genome Res* 2005, **110**:462–467.

89. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25**:Q955–Q964.

90. Schomburg D, Schomburg I: **Enzyme databases.** *Meth Mol Biol* 2010, **609**:113–128.

91. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O, Sonnhammer ELL: **InParanoid 7: new algorithms and tools for eukaryotic orthology analysis.** *Nucleic Acids Res* 2010, **38**:D196–D203.

92. Swofford D: *PAUP* version 4.0. Phylogenetic Analysis Using Parsimony (and Other Methods).* Sunderland, Massachusetts, USA: Sinauer Associates; 2002.

93. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R: **AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.** *Nucleic Acids Res* 2011, **39**:W339–W346.

94. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**:2731–2739.

95. Petersen LA, Hughes DL, Hughes R, DiMichele L, Salmon P, Connors N: **Effects of amino acid and trace element supplementation on pneumocandin production by** *Glarea lozoyensis*: **impact on titer, analogue levels, and the identification of new analogues of pneumocandin B-0.** *J Indust Microbiol Biotechnol* 2001, **26**:216–221.