OPEN ACCESS Freely available online

PLoS GENETICS

Genome Sequencing and Comparative Transcriptomics of the Model Entomopathogenic Fungi *Metarhizium anisopliae* and *M. acridum*

Qiang Gao[1☯], Kai Jin[2☯], Sheng-Hua Ying[3☯], Yongjun Zhang[4☯], Guohua Xiao[1☯], Yanfang Shang[1], Zhibing Duan[1], Xiao Hu[1], Xue-Qin Xie[3], Gang Zhou[3], Guoxiong Peng[2], Zhibing Luo[4], Wei Huang[1], Bing Wang[1], Weiguo Fang[5], Sibao Wang[5], Yi Zhong[6], Li-Jun Ma[7], Raymond J. St. Leger[5], Guo-Ping Zhao[6], Yan Pei[4], Ming-Guang Feng[3]*, Yuxian Xia[2]*, Chengshu Wang[1]*

1 Key Laboratory of Insect Developmental and Evolutionary Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China, **2** College of Bioengineering, Chongqing University, Chongqing, China, **3** College of Life Sciences, Zhejiang University, Hangzhou, China, **4** Biotechnology Research Center, Southwest University, Chongqing, China, **5** Department of Entomology, University of Maryland, College Park, Maryland, United States of America, **6** Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China, **7** The Broad Institute, Cambridge, Massachusetts, United States of America

Abstract

Metarhizium spp. are being used as environmentally friendly alternatives to chemical insecticides, as model systems for studying insect-fungus interactions, and as a resource of genes for biotechnology. We present a comparative analysis of the genome sequences of the broad-spectrum insect pathogen *Metarhizium anisopliae* and the acridid-specific *M. acridum*. Whole-genome analyses indicate that the genome structures of these two species are highly syntenic and suggest that the genus *Metarhizium* evolved from plant endophytes or pathogens. Both *M. anisopliae* and *M. acridum* have a strikingly larger proportion of genes encoding secreted proteins than other fungi, while ~30% of these have no functionally characterized homologs, suggesting hitherto unsuspected interactions between fungal pathogens and insects. The analysis of transposase genes provided evidence of repeat-induced point mutations occurring in *M. acridum* but not in *M. anisopliae*. With the help of pathogen-host interaction gene database, ~16% of *Metarhizium* genes were identified that are similar to experimentally verified genes involved in pathogenicity in other fungi, particularly plant pathogens. However, relative to *M. acridum*, *M. anisopliae* has evolved with many expanded gene families of proteases, chitinases, cytochrome P450s, polyketide synthases, and nonribosomal peptide synthetases for cuticle-degradation, detoxification, and toxin biosynthesis that may facilitate its ability to adapt to heterogenous environments. Transcriptional analysis of both fungi during early infection processes provided further insights into the genes and pathways involved in infectivity and specificity. Of particular note, *M. acridum* transcribed distinct G-protein coupled receptors on cuticles from locusts (the natural hosts) and cockroaches, whereas *M. anisopliae* transcribed the same receptor on both hosts. This study will facilitate the identification of virulence genes and the development of improved biocontrol strains with customized properties.

Citation: Gao Q, Jin K, Ying S-H, Zhang Y, Xiao G, et al. (2011) Genome Sequencing and Comparative Transcriptomics of the Model Entomopathogenic Fungi *Metarhizium anisopliae* and *M. acridum*. PLoS Genet 7(1): e1001264. doi:10.1371/journal.pgen.1001264

Editor: Mark Achtman, Environmental Research Institute, University College Cork, Ireland

Received July 26, 2010; **Accepted** December 1, 2010; **Published** January 6, 2011

Funding: C Wang, Y Pei, and M-G Feng were supported by the Ministry of Science and Technology of China (grant no. 2009CB118904). C Wang was also supported by the Knowledge Innovation Program from Chinese Academy of Sciences (grant no. KSCX2-YW-G-037) and the National Hi-Tech Research and Development Program of China (grant no. 2009AA10Z112). Y Xia was supported by the Natural Science Foundation of China (grant no. 30771446). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mgfeng@zju.edu.cn (M-GF); yuxianxia@cqu.edu.cn (YX); cswang@sibs.ac.cn (CW)

☯ These authors contributed equally to this work.

Introduction

Most fungi with sequenced genomes are plants pathogens or saprophytes. However, there are also thousands of entomopathogenic fungal species that play a crucial role in controlling insect populations. The genus *Metarhizium* includes the best studied entomopathogenic fungi at the molecular and biochemical level. They have a world-wide distribution from the arctic to the tropics and colonize an impressive array of environments including forests, savannahs, swamps, coastal zones and deserts [1]. *Metarhizium* species are amongst the most abundant fungi isolated from soils with titers reaching 10^6 conidia per gram in grasslands [2]. The genus contains *M. anisopliae*, which has a broad host range, as well as specialists, such as the locust-specific pathogen *M. acridum*. These two species in particular have emerged as excellent model organisms to explore a broad array of questions in ecology and evolution, host preference and host switching, and to investigate the mechanisms of speciation. In addition, both *M. anisopliae* and *M. acridum* have been at the forefront of efforts to develop biocontrol alternatives to chemical insecticides in agricultural and disease-vector control programs, and many commercial products are on the market or under development [2–4].

Author Summary

Aside from playing a crucial role in natural ecosystems, entomopathogenic fungi are being developed as environmentally friendly alternatives for the control of insect pests. We conducted the first genomic study of two of the best characterized entomopathogens, *Metarhizium anisopliae* and *M. acridum*. *M. anisopliae* is a ubiquitous pathogen of >200 insect species and a plant growth promoting colonizer of rhizospheres. *M. acridum* is a specific pathogen of locusts. Important findings of this study included: 1) Both *M. anisopliae* and *M. acridum* have a very large number of genes encoding secreted proteins, and many of these play roles in fungus-insect interactions. 2) *M. anisopliae* has more genes than *M. acridum*, which may be associated with adaptation to multiple insect hosts. 3) Unlike *M. acridum*, the *M. anisopliae* genome contains many more transposase genes and shows no evidence of repeat-induced point mutations. The lack of repeat-induced mutations may have allowed the lineage-specific gene duplications that have contributed to its adaptability. 4) High-throughput transcriptomics identified the strategies by which these fungi overcome their insect hosts and achieve specificity. These genome sequences will provide the basis for a comprehensive understanding of fungal–plant–insect interactions and will contribute to our understanding of fungal evolution and ecology.

Our knowledge of the ecological impact of *M. anisopliae* and its potential as a biocontrol agent has recently been enhanced by the discovery that it colonizes plant roots where it may simultaneously act as a biofertilizer and biopesticide to boost plant growth [5]. Consistent with its broad lifestyle options, *M. anisopliae* exhibits an extremely versatile metabolism, enabling growth under various environmental conditions, with sparse nutrients and in the presence of compounds lethal to other fungi [6]. As the asexual stages (anamorphs) of medicinally valued *Cordyceps* spp. [7], *Metarhizium* spp. are prolific producers of enzymes and diverse secondary metabolites with activities against insects, fungi, bacteria, viruses and cancer cells [6,8,9]. In addition, the enzymes from *Metarhizium* spp. are frequently exploited as industrial catalysts [10,11]. *M. anisopliae* has also been used in studies on the immune systems of invertebrate model hosts to provide insights into emerging human pathogens [12], and it is a developing model for studies on aging [13,14].

In contrast to the versatile *M. anisopliae*, the specialist *M. acridum* is specific for certain locusts and grasshoppers [15]. However, like *M. anisopliae*, it is a producer of diverse cell types (e.g., conidia, hyphae, appressoria, unicellular blastospores, and multi-cellular hyphal bodies) that facilitate the infection of target insects via adhesion and penetration of the host cuticle, proliferation within tissues and the haemolymph, and eventual eruption through the host cadaver (Figure 1). *M. acridum* is mass produced and used on a large scale for locust control [16], whereas few other biological control agents have been such a commercial success because of poor efficacy compared to chemicals [17].

Although recent advances have identified the functions of several pathogenicity genes [18–22] and technical developments improved the virulence of *M. anisopliae* [23,24], the need to understand these fungi and expand their biotechnological potential requires sequenced genomes of *M. anisopliae* and *M. acridum*. Sequencing two related species that have evolved very different lifestyles will increase their utility as models, and provide insights into the evolution of pathogenicity. Such sequences will also allow for more rapid identification of genes encoding biologically active molecules and genes responsible for interactions between fungi, plants and insects. These findings could be further translated into the development of improved strains with customized properties that could potentially function as comprehensive plant symbionts to improve plant establishment and sustainable agriculture, particularly on marginal lands.

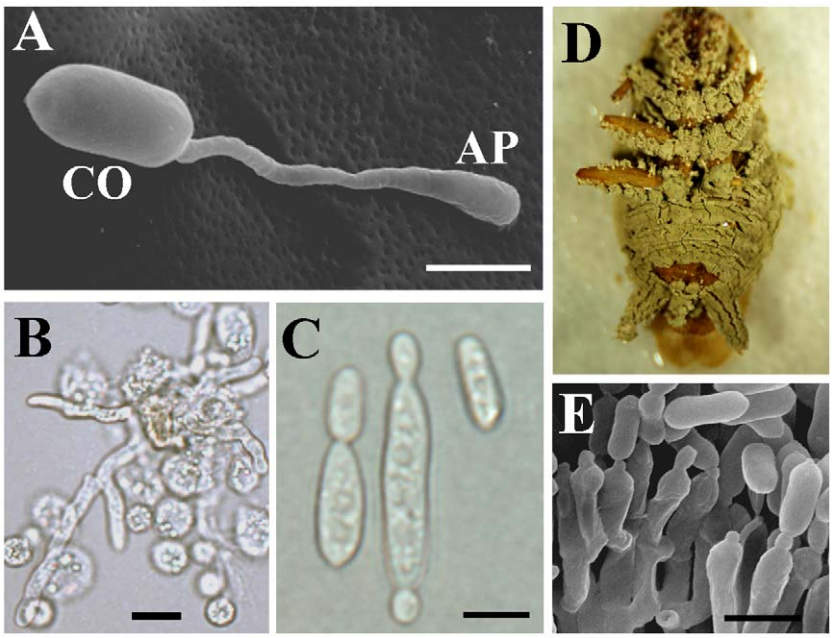

Figure 1. Major stages in the infection cycle of *Metarhizium*. (A) A germinating conidium producing an appressorium. (B) Mycelia attacked by hemocytes after cuticular penetration. (C) Budding yeast-type cells (blastospores) produced by the fungus to facilitate dispersal in insect hemocoel. (D) Cadaver showing emerging hyphae producing conidia (E). CO, conidium; AP, appressorium. Bar, 5 μm.
doi:10.1371/journal.pgen.1001264.g001

Results

Genome sequencing and general features

The genomes were each shotgun sequenced to ~100× coverage. The *M. anisopliae* genome (strain ARSEF 23) was assembled into 176 scaffolds (>1 kb; N50, 2.0 Mb) containing 1,271 contigs with a total size of 39.0 Mb (loci tagged as MAA). The *M. acridum* genome (strain CQMa 102) was assembled into 241 scaffolds (>1 kb; N50, 329.5 kb) containing 1,609 contigs with a genome size of 38.0 Mb (loci tagged as MAC) (Table 1). These assemblies closely correspond to the genome sizes of other Ascomycetes (Table S1). By mapping >6,000 unique expressed sequenced tagged sequences to the scaffolds, each genome was estimated to be >98% complete. *M. anisopliae* and *M. acridum* were predicted to have 10,582 and 9,849 protein coding genes, respectively, which is similar to the coding capacity of other Ascomycetes (Table S1). We examined homology relationships between *M. anisopliae* and *M. acridum*, and a set of eight other ascomycete genomes (Figure 2A). The results indicated that ~90% of the genes in both *Metarhizium* genomes have homologs ($E \leq 1\times10^{-5}$) in other Ascomycetes. In addition, *M. anisopliae* has 398 (3.8%) genes with matches restricted to *M. acridum* (*Metarhizium*-restricted genes) and 263 (2.5%) orphan sequences. *M. acridum* has 219 (2.2%) orphan sequences (Figure 2A). Further analysis of the *M. anisopliae* orphans showed that 21.3% had matches in bacteria, 3.4% in animals and 3.8% in viruses. Similarly, 13.3%, 5.5% and 2.7% of the *M. acridum* orphans had matches in bacteria, animals and viruses, respectively, consistent with possible horizontal gene transfer events.

The proportion of genes encoding secreted proteins is remarkably large, being 17.6% (1,865 proteins) in *M. anisopliae* and 15.1% (1,490 proteins) in *M. acridum* as compared to 7–10% in plant pathogens [25] and ~5% in *N. crassa* [26] or *A. nidulans* [27]. As expected, many of the secreted proteins are in families which could have roles in colonization of insect tissues, such as proteases (Table S2). However, 32.2% of *M. anisopliae* and 28.7% of *M. acridum* secreted proteins had no conserved domains or functionally characterized homologs. Of these, ~22% were *Metarhizium*-restricted genes and ~4% were orphan genes in either genome.

Table 1. Main features of the *M. anisopliae* and *M. acridum* genomes.

Features	*M. anisopliae*	*M. acridum*
Size (Mb)	39.04	38.05
Coverage (fold)	100×	107×
G+C content (%)	51.49	49.91
Repeat rate (%)	0.98	1.52
Protein-coding genes	10582	9849
Gene density (genes per Mbp)	271.1	258.8
Exons per gene	2.8	2.7
Secreted proteins	1865	1490
GPI proteins[1]	158	142
Unique proteins	615	434
tRNA	141	122
Pseudogenes	363	440

[1]Glycosylphosphatidylinositol-anchored cell wall proteins.

doi:10.1371/journal.pgen.1001264.t001

Syntenic and phylogenetic relationships

Pairwise comparison indicated that the two *Metarhizium* genome structures have large areas of synteny (Figure 2B, Figure S1A). The lineage specific regions of *M. anisopliae* and *M. acridum* contain high densities of transposases, species-specific genes, genes encoding proteins with unknown functions and pseudogenes (Figure S1B). Similar lineage-specific regions were found in *Fusarium* spp. [28]. Ninety nine percent of the *M. anisopliae* genome comprises non-repetitive sequences, and the orthologs shared with the *M. acridum* genome display an average 89.8% amino acid identity. The two *Metarhizium* species are therefore more closely related than the three *Aspergillus* species *A. nidulans*, *A. fumigatus* and *A. oryzae* which share only 68% average sequence identity [29]. A phylogenomic analysis revealed that *M. anisopliae* and *M. acridum* lineages diverged about 33–43 million years (MY) ago and are most closely related to the mutualistic plant endophyte *Epichloe festucae* (divergence time 88–114 MY) and to the wheat head blight fungus *Fusarium graminearum* (divergence time 144–187 MY) (Figure 2C).

Transposases and repeat-induced point mutation

The specialist *M. acridum* harbors more repetitive elements than *M. anisopliae* but the latter has many more transposases (Table S2). Most of these are DNA transposases (97/148 in MAA; 12/20 in MAC), with subclasses *hAT* (45/97) and *Helitron* (26/97) being particularly abundant in *M. anisopliae*. The *Copia* (17) and LINE (26) retrotransposons are also abundant in the genome of *M. anisopliae*, while *M. acridum* has only three LINE elements and does not contain *Copia* (Figure 3A). Transcriptome analysis (see below) showed that most (>65%) of the transposase genes were transcribed by the *Metarhizium* hyphae during the infection process (Table S3).

The number of putative transposases in the *M. acridum* genome is lower by at least a factor of five than in most Ascomycetes, including *M. anisopliae* (Table S2). This could be explained by repeat induced point mutations (RIP) introducing CpG to TpA transitions in duplicated sequences during the sexual cycle [30]. This mutational bias is observed in *M. acridum* (RIP index, 2.17) but not in *M. anisopliae* (RIP index, 1.09) (Figure 3B). Consistent with *Neurospora crassa* which has efficient RIP [31], the genome of *M. acridum* contained twice as many duplicated pseudogenes (254 versus129) as did that of *M. anisopliae*. The *M. anisopliae* genome contains more processed and fragmented pseudogenes caused by mobile elements (234 versus 186), consistent with transposons making a greater contribution to genetic instability in *M. anisopliae* (Table S4). The production of stable biocontrol agents for commercialization might therefore benefit from disabling transposable elements.

Virulence associated genes

An InterproScan analysis identified 2,710 protein families (containing 7,178 proteins) in *M. anisopliae* and 2,658 families (containing 6,615 proteins) in *M. acridum*. A stochastic birth and death model [32] showed that relative to *M. acridum*, 42 families including transporters, transcription factors, cytochrome P450s, proteases and lipases were expanded and three families (protein kinase, aminotransferase and transpeptidase) were contracted in *M. anisopliae* (Table S5). This resulted in *M. anisopliae* having more genes in most functional categories except for those involved in signal transduction (Figure 4, Table 2).

To find potential virulence-associated genes, a whole genome blast analysis was conducted against the pathogen-host interaction (PHI) gene database, a collection of experimentally verified pathogenicity, virulence and effector genes from fungi, oomycetes and bacteria [33]. We identified 1,828 putative PHI genes in *M.*

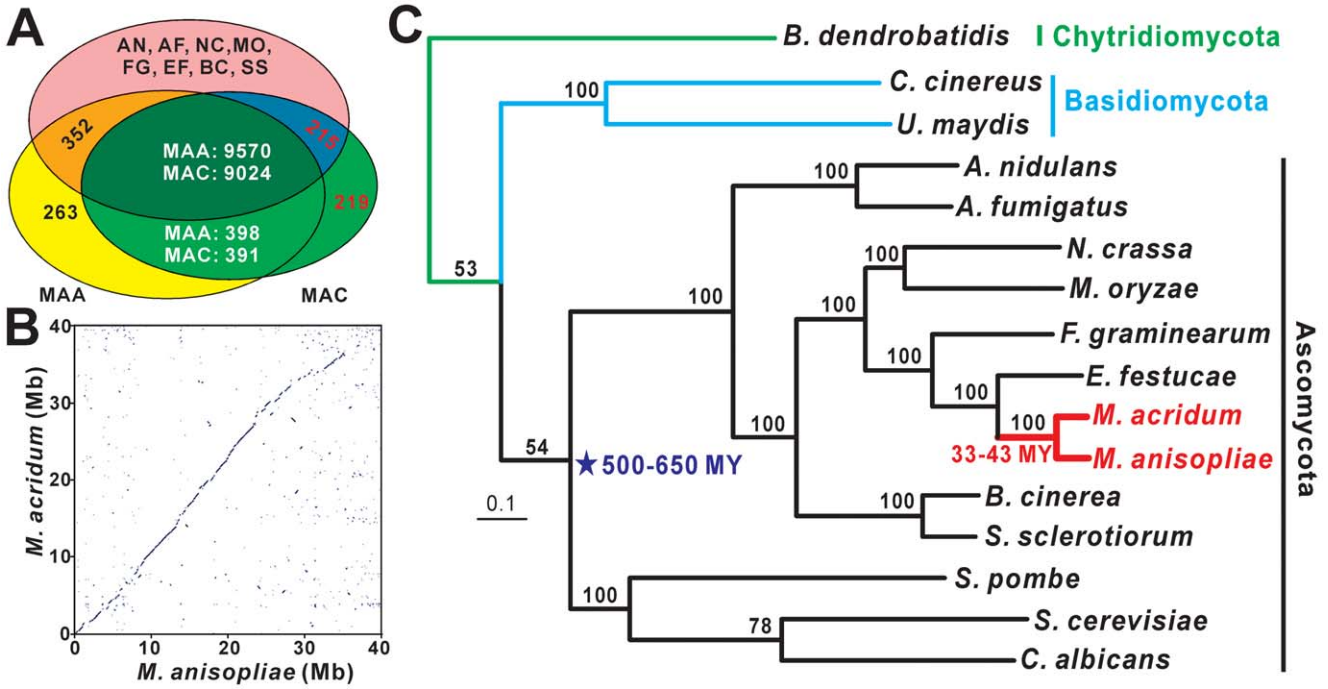

Figure 2. Homology, syntenic, and phylogenomic relationships of *M. anisopliae* and *M. acridum*. (A) Predicted proteins in *M. anisopliae* (MAA) and *M. acridum* (MAC) were compared with the genome encoding proteins of *Aspergillus nidulans* (AN), *A. fumigatus* (AF), *Neurospora crassa* (NC), *Magnaporthe oryzae* (MO), *Fusarium graminearum* (FG), *Epichloë festucae* (EF), *Botrytis cinerea* (BC) and *Sclerotinia sclerotiorum* (SS). The diagram was constructed with a cut off *E*-value $<1\times10^{-5}$. (B) Dot blot of *M. anisopliae* and *M. acridum* using ordered scaffold data. (C) Phylogenetic tree constructed using the Dayhoff amino acid substitution model showing the evolutionary relationships of 16 fungal species. MY = million years.
doi:10.1371/journal.pgen.1001264.g002

anisopliae (17.3% of its genes, belonging to 383 protein families) and 1,629 putative PHI genes in *M. acridum* (16.5%, 371 families), of which 1,331 genes were orthologous. Although there are no entries from entomopathogenic fungi in the PHI-base, we proceeded on the assumption that the proof of pathogenicity/virulence of a gene in one fungus also suggests a pathogenicity/virulence function in other fungi [34]. In accordance with this assumption, the search of the PHI database yielded several already characterized *M. anisopliae* pathogenicity determinants, including subtilisins (see below) and hydrophobins (small cell wall proteins) that have pleiotropic functions in *M. anisopliae* including attachment of spores to hydrophobic surfaces [35]. The class 2 (MAA_01182 and MAC_09507) and class 1 (MAA_10298 and MAC_04376) hydrophobins had significant similarity with PHI sequences from plant pathogenic fungi. The previously characterized adhesin, MAD1 (MAA_03775) required for specific binding to insect host surfaces [20], resembled EAP1 (PHI acc: 517) from the human pathogen *Candida albicans*. However, the adhesin MAD2 (MAA_03807) required for binding to plant surfaces [20], had no significantly similar sequence in the PHI database. Orthologs to both MAD1 (MAC_00987) and MAD2 (MAC_00953) were found in the *M. acridum* genome.

Using the PHI-base content with a focus on ascomycetes, Sexton and Howlett found many parallels in the infection mechanisms used by plant and animal pathogens [36]. To determine how many plant pathogen PHI genes are also found in *Metarhizium*, we screened the *F. graminearum* and *M. oryzae* genomes against the PHI-base and identified 2,053 genes (in 398 families) and 1,713 genes (in 427 families), respectively, representing about 16% of gene contents in these two fungi (Table S6). Approximately, 70% of these genes are orthologous to PHI sequences in *M. anisopliae* and *M. acridum*. Fewer *Metarhizium* orthologs were found in animal pathogenic fungi such as *C. albicans*, which could be explained by *Metarhizium* being more closely related to plant pathogens (Figure 2C) as well as the animal pathogens lacking appressoria (infection structures) during host penetration [4].

Gene families involved in degrading insect cuticles

Insect pathogens such as *Metarhizium* spp. need to penetrate the protein-chitin rich insect cuticle and solubilize host tissues for nutrition. Therefore, they would be expected to secrete large numbers of degradative enzymes. Indeed, *M. anisopliae* has more genes encoding secreted proteases (132) than other sequenced fungi (Table S2). The trypsin family has the highest relative expansion among the proteases with 32 genes in *M. anisopliae*, almost twice as many as *M. acridum* and 6 to 10 times as many as the other taxa evaluated (Figure 5A, Table S2). A chymotrypsin (MAA_07484) that might have been imported from bacteria through horizontal gene transfer [37] and two trypsins that were recently duplicated in *M. anisopliae* (MAA_05135 and MAA_05136) are missing from the *M. acridum* genome (Table S7). Subtilisins (55 in MAA and 43 in MAC; 7 to 31 in other fungi) (Figure 5B, Table S8) and aspartyl proteases (33 in MAA and 25 in MAC; 9 to 21 in other fungi) (Table S9) are also expanded in *M. anisopliae* due to lineage-specific duplications (Figure S1C). Most of the *Metarhizium* subtilisins (48 in MAA and 37 in MAC) and aspartyl proteases (27 in MAA and 23 in MAC) had significant matches in the PHI-base. Subtilisins assist in the infection processes of *M. anisopliae* by degrading host cuticles, providing nutrition and disabling antimicrobial peptides [38]. The importance of *Metarhizium* aspartyl proteinases has not been demon-

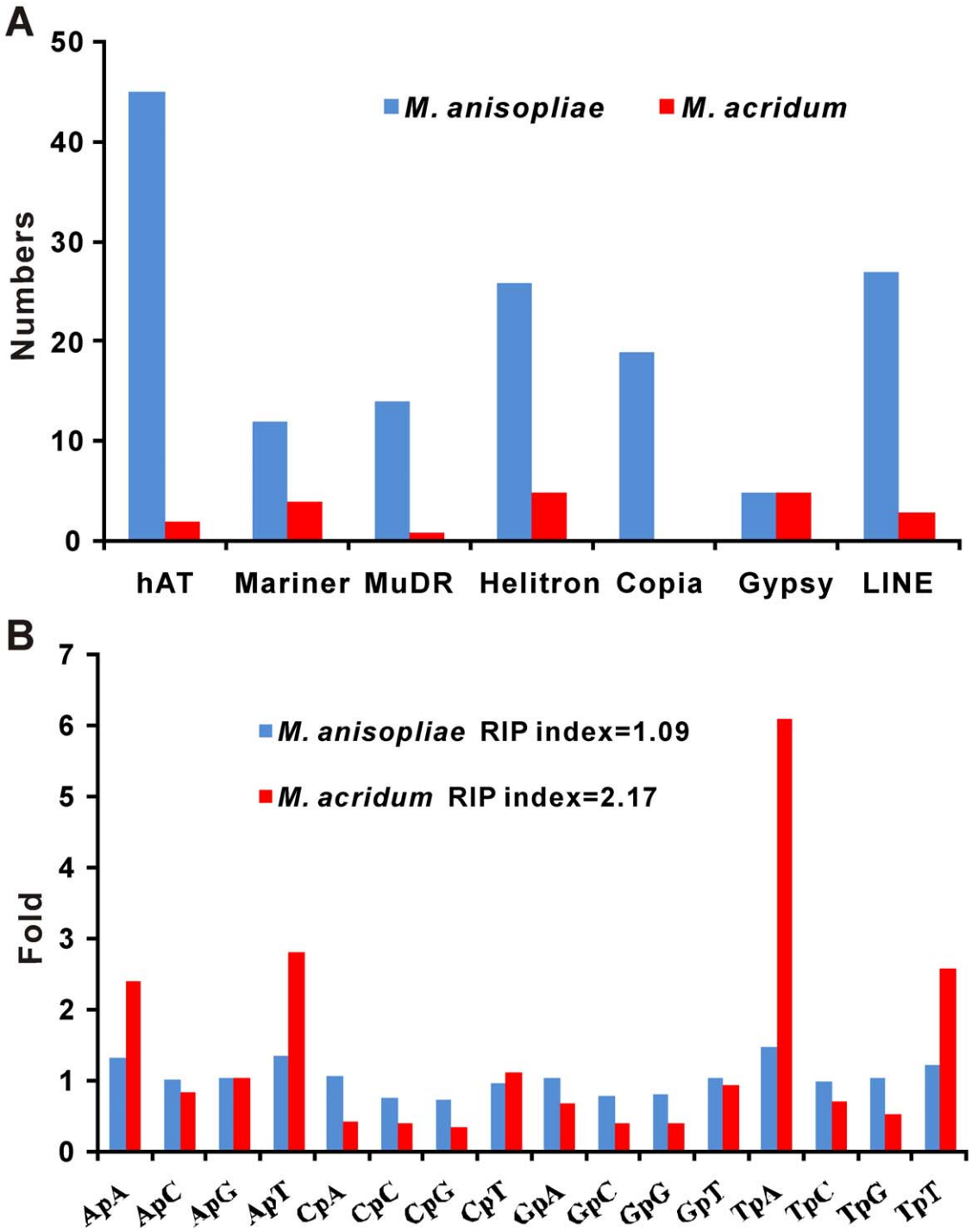

Figure 3. Families of transposase genes and estimation of RIP. Families of transposase genes (A) and estimation of RIP (B) in *M. anisopliae* and *M. acridum*.
doi:10.1371/journal.pgen.1001264.g003

strated but they resemble the aspartyl proteases that assist the human pathogen *C. albicans* by degrading cell surface molecules [39].

Many plant pathogens need glycoside hydrolases, pectate lyases and cutinases to degrade the plant cuticle (waxy layer) and cell wall. The number of glycoside hydrolases (GH) possessed by *M. anisopliae* (156) and *M. acridum* (140) is close to the average for plant pathogenic fungi (150) (Table S10). However, only ~20% of the *Metarhizium* GH genes (36 in MAA and 29 in MAC) were similar to PHI catalogued sequences as compared to 44% (70 genes) in *F. graminearum* and 29% (57 genes) in *M. oryzae* (Table S6). The plant pathogens in particular have additional GH3 cellulases while *Metarhizium* spp. lack the GH11 family of xylanases. GH3 and GH11 family genes are catalogued in the PHI-base. Overall, fewer genes were associated with plant utilization in *Metarhizium* than in plant pathogens. This included fewer putative cutinases (2 in

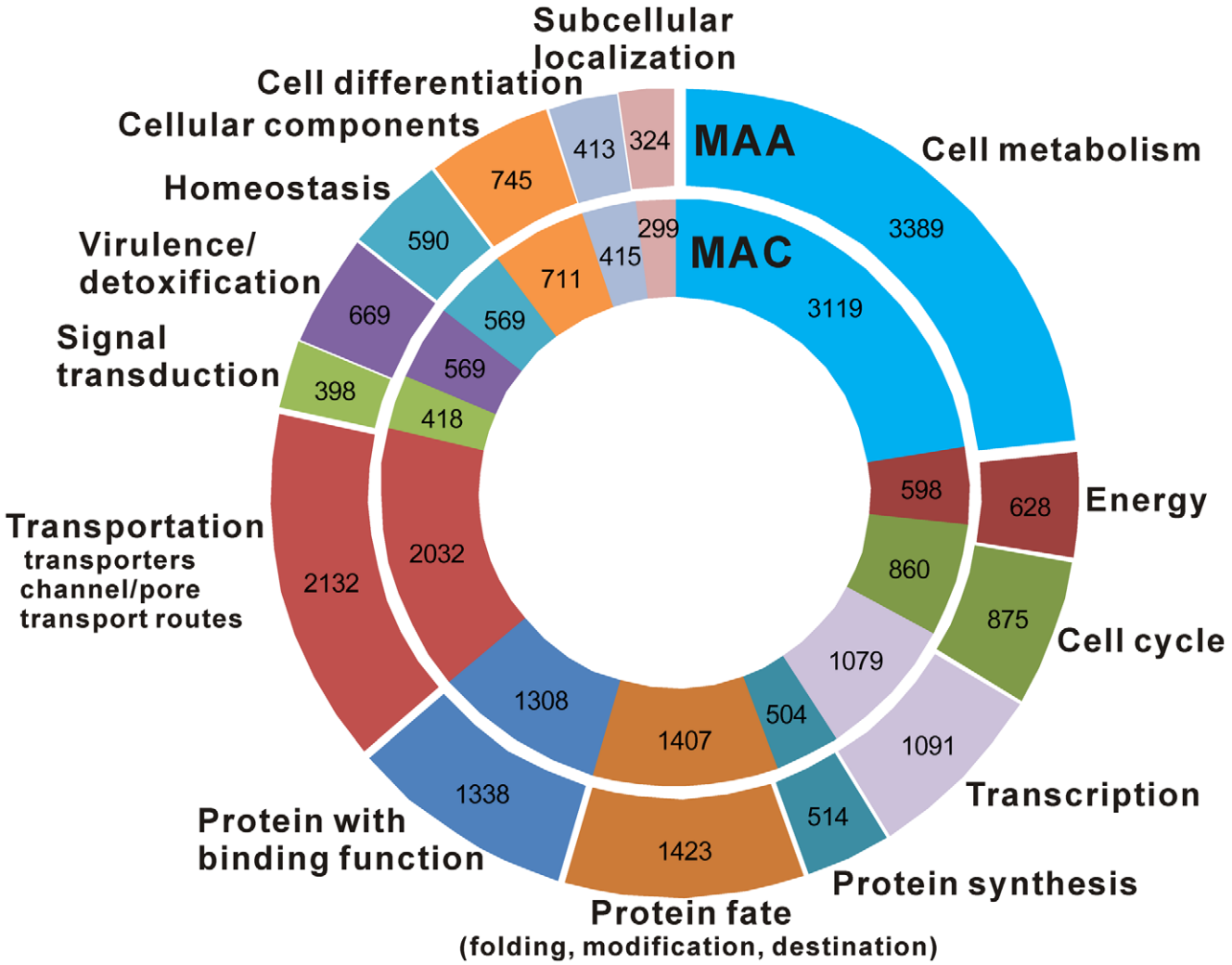

Figure 4. Functional classification and comparison of *M. anisopliae* and *M. acridum* proteins. Each circle represents the relative fraction of genes represented in each of the categories for each genome. The gene numbers are also shown.
doi:10.1371/journal.pgen.1001264.g004

Metarhizium spp. 8 to 18 in plant pathogens) and pectate lyases (7 in *Metarhizium* spp.; 9 to 25 in plant pathogens). However, the GH16 family of xyloglucan/xyloglucosyl transferases involved in decomposition of plant cell walls is well represented in the *Metarhizium* genomes (18 in MAA and 16 in MAC; 6 to 16 in plant pathogens) (Table S10). More predictably, GH18 chitinases involved in the digestion of insect cuticle chitin [40], are over represented in *Metarhizium* (30 in MAA and 21 in MAC; 5 to 14 in plant pathogens) (Figure 5C, Table S6). The few chitinases included in the PHI database are involved in fungal developmental processes, as chitin is not a substrate found in animal and plant hosts.

The number of genes for secreted lipases (12 in MAA, 5 in MAC) is well above the average found in other fungi, and 9 *M. anisopliae* and 5 *M. acridum* lipases showed significant similarity to genes in the PHI-base, as compared to 3 lipases each in *F. graminearum* and *M. oryzae* (Table S6). The role of individual *Metarhizium* lipases in pathogenicity has not been demonstrated, although a lipase activity inhibitor blocks infection processes in *M. anisopliae* [41]. Lipases MAA_03127 and MAC_09232 showed best-hit relationships with an extracellular lipase FGL1 (PHI acc: 432) that is a virulence factor in *F. graminearum* [42].

Gene families for transportation and detoxification of compounds

Metarhizium genomes encode a large number of transporters (484 in MAA and 441 in MAC) (Table S11). Most transporters belong to the major facilitator superfamily (MFS) (269 in MAA; 236 in MAC) but the ATP-binding cassette (ABC) is also well represented (56 in MAA; 51 in MAC) (Table S11). Most of the ABC transporters (52/56 in MAA and 46/51 in MAC) and many of the MFS transporters (124/269 in MAA and 96/236 in MAC) were similar to genes catalogued in the PHI database (Table S6). The ABC transporters are usually implicated in defending the pathogen from host-produced secondary metabolites, whereas MFS proteins are typically involved in the transport of a wide range of substrates and may function as nutrient sensors [43]. Interestingly, both *Metarhizium* species have more amino acid and peptide transporters than do other fungi (60 in MAA and 57 in MAC; 29 to 38 in other fungi), consistent with their being able to access a range of protein degradation products from insect sources. Homologs of these genes are absent from the PHI database. The only *Metarhizium* transporter with an experimentally determined function is the sucrose and galactoside transporter MRT

Table 2. Comparison of selected protein families in *M. anisopliae* and *M. acridum* and their percentage of expression on cockroach (CO) and locust (LO) cuticles during pre-penetration growth.

Protein families	*M. anisopliae*			*M. acridum*		
	No.	% CO_exp	% LO_exp	No.	% CO_exp	% LO_Exp
Subtilisins	**55**	54.5	61.8	**43**	53.5	65.1
Trypsins	**32**	26.7	33.3	**17**	76.5	64.5
Aspartic proteases	**34**	61.8	75.3	**26**	76.9	80.8
Chitinases	**30**	43.3	46.7	**21**	61.9	76.2
Lipases	**47**	80.9	89.4	**41**	87.8	92.7
Cytochrome P450	**123**	52.8	54.5	**100**	70.0	75.0
Dehydrogenases	**271**	82.7	84.1	**236**	85.2	86.0
Acyl-CoA N-acyltransferases	**40**	90.0	92.5	**36**	86.1	86.1
Monooxygenases	**30**	70.0	70.0	**23**	91.3	91.3
Transporters	**484**	82.9	84.8	**441**	86.4	83.8
SM backbone genes[1]	**61**	39.3	44.3	**42**	57.1	69.0
GPCR-like protein	**78**	70.5	75.6	**62**	95.2	91.9
Protein kinases	**161**	83.2	85.1	**192**	86.5	89.6
Transcription factors	**510**	89.6	91.0	**439**	94.4	96.3
Average	128	64.9	68.1	112	79.2	82.2

No., number of paralogous genes.
% CO_Exp and LO_Exp indicate percentage of paralogous genes expressed by *M. anisopliae* or *M. acridum* infecting locust and cockroach cuticles, respectively.
[1]Secondary metabolite backbone biosynthetic genes.
doi:10.1371/journal.pgen.1001264.t002

(belonging to the MFS superfamily), which is required by *M. anisopliae* for rhizosphere competence but not for virulence [44]. There are 6 MRT homologs in *M. anisopliae* and 5 in *M. acridum* but 12 in *F. graminearum* and 26 in *M. oryzae*, suggesting these genes could be generally important for establishing plant-fungus relationships.

Additional evidence about lifestyle could be found in the relatively large number of genes involved in detoxification in both *Metarhizium* genomes (Table 2, Table S2) as these potentially contribute to interactions with insect hosts (Table S6). However, families of dehydrogenases, acyl-CoA N-acetyltransferases, monooxygenases and cytochrome P450s (CYP) were preferentially expanded in *M. anisopliae* relative to *M. acridum* (Table 2, Table S5). One third of the dehydrogenases (92/271 in MAA and 80/236 in MAC) were putative PHI genes (Table S6). *M. anisopliae* was particularly enriched in zinc-containing alcohol dehydrogenases (17 in MAA; 7 in MAC) required for the biosynthesis of mannitol, a crucial factor for stress tolerance and virulence in the animal pathogen *Cryptococcus neoformans* [45]. The monooxygenases in particular might be involved in rapid elimination of insect polyphenolics by ortho-hydroxylation of phenols to catechols [46].

The genome of *M. anisopliae* encodes 123 highly divergent CYP genes verses 100 CYPs in *M. acridum* (Figure 5D, Table S12). Ninety of the *M. anisopliae* CYPs and 69 of the *M. acridum* CYPs are similar to sequences in the PHI-base (Table S6). CYP genes are involved in oxygenation steps during alkane assimilation and the biosynthesis of secondary metabolites as well as with detoxification [47]. *M. anisopliae* efficiently metabolizes the alkanes that are a major component of the surface layer of the insect cuticle (epicuticle) [48]. Although the CYP52 subfamily is particularly important for alkane oxidation [49], *M. anisopliae* has only a single CYP52 (MAA_06634) compared to four in *M. acridum* (Table S12). However, MAA_06634 and its ortholog in *M. acridum* (MAC_09267) were highly expressed (see below) by *M. anisopliae* and *M. acridum* when infecting either cockroach or locust cuticles (Figure S5A). The other CYP genes up-regulated on cuticles were mostly involved in detoxification. *M. anisopliae* and *M. acridum* are predicted to contain four and two CYP504s, respectively. CYP504s are used by fungi to degrade phenylacetate [50], an antimicrobial compound found in plants and insects [51]. The subfamily CYP53 is also represented in the PHI database as it is responsible for detoxification of benzoate and its derivatives [52]. *M. anisopliae* and *M. acridum* have two and one CYP53 genes, respectively.

The subfamily CYP5081 involved in biosynthesis of helvolic acid, an antibiotic toxin [53], consists of four closely localized CYP loci (PHI genes) in *M. anisopliae* (MAA_06585, MAA_06586, MAA_06589 and MAA_06593) that are absent in *M. acridum*. All four CYP5081 genes were expressed by *M. anisopliae* infecting cuticles (Figure 5D). Both *M. anisopliae* and *M. acridum* have three CYP genes putatively encoding lipid dioxygenases (CYP6001: MAA_04954 and MAC_00208; CYP6002: MAC_05834; CYP6003: MAA_03481 and MAC_00918; CYP6004: MAA_0003) and two lipoxygenases (MAA_06278 and MAA_01260; MAC_01254 and MAC_9416). Oxylipins, the end products of these genes, allow *Aspergillus nidulans* to colonize plant seeds [54], and seeds are also a habitat for *M. anisopliae* [55], implying that a similar strategy is employed by *Metarhizium* to establish plant-fungus relationships.

Core genes for biosynthesis of secondary metabolites

M. anisopliae is a prolific producer of secondary metabolites including insecticidal destruxins [56], but with the exception of the serinocyclins [57] and NG-391 [58], the genes involved in their biosynthesis are unknown. However, diagnostic genes for secondary metabolite production include those encoding polyketides and non-ribosomal peptides (the most prominent classes of fungal secondary metabolites), as well as those responsible for modifications of the core moiety (a peptide or polyketide) such as genes encoding dehydrogenases, methyltransferases, acetyl transferases, prenyl-transferases, oxireductases and CYPs [36]. Consistent with expressed sequence tag studies [59], *M. anisopliae* appears to possess a much greater potential for the production of secondary metabolites than *M. acridum* or most other fungi (Tables S2 and S13). The *M. anisopliae* genome encodes 14 putative non-ribosomal peptide synthases (NRPS), 24 polyketide synthases (PKS) and 5 NRPS-PKS hybrid genes, which is more than *M. acridum* (13 NRPS genes, 13 PKS genes and 1 NRPS-PKS hybrid) and the average in other Ascomycetes (7 NRPS, 12 PKS genes and 1 NRPS-PKS) (Table S13). NRPSs and PKSs are strongly associated with pathogenicity in many plant pathogenic fungi and are well represented in the PHI database. As in other fungi, *Metarhizium* NRPS and PKS genes were often clustered together with genes that modify their products. One cluster suggests that *Metarhizium* might produce prenylated alkaloids (Figure S2). *M. anisopliae* possesses putative NRPS-like antibiotic synthetases (MAA_08272) consistent with defending the cadaver against microbial competitors. It also possesses a putative bassianolide synthetase (MAA_07513), a virulence factor of the insect pathogen *Beauveria bassiana* [60]. The NRPS-like proteins MAA_07148 and MAC_06316 are most similar to ACE1, a PKS/NRPS hybrid that confers avirulence to *M. grisea* during rice infection [61]. *M. anisopliae* NRPS MAA_00969 is similar (43% identity) to HTS1, the key enzyme responsible for the biosynthesis of the host-selective HC-toxin that confers the specificity of *Cochliobolus carbonum* to maize [62]. Sixteen out of 24 PKS and 5/14 NRPS genes in *M. anisopliae* are species specific

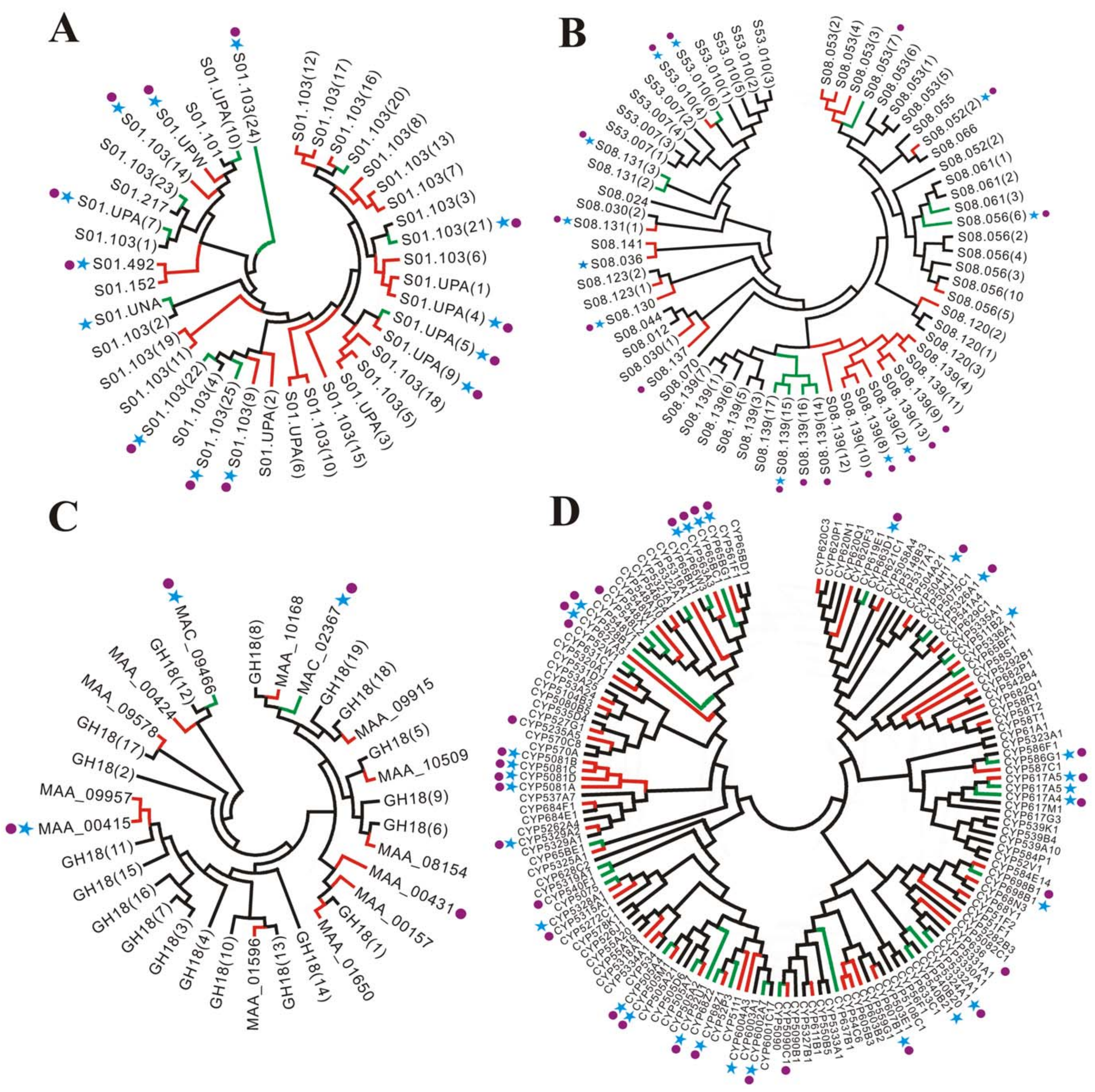

Figure 5. Unrooted phylogenetic trees showing differences in gene expansion in *M. anisopliae* and *M. acridum*. (A) Trypsins; (B) Subtilisins; (C) GH18 chitinases; (D) Cytochrome P450s. Black branches identify orthologous loci in *M. anisopliae* and *M. acridum*. Red and green branches identify genes that are only present in *M. anisopliae* or *M. acridum*, respectively. Lineage specific genes expressed by these species on cockroach cuticle are marked with a blue asterisk. Lineage specific genes expressed on locust cuticle are marked with a purple circle. Protease family classification refers to Table S7 for trypsins and Table S8 for subtilisins.
doi:10.1371/journal.pgen.1001264.g005

versus 4/13 PKS and 5/13 NRPS in *M. acridum*, suggesting lineage specific expansion of these families in both *Metarhizium* species. However, it is reassuring for present and future commercialization of these fungi that we found no orthologs of genes for the biosynthesis of the human mycotoxins gliotoxin and aflatoxin.

Signal transduction

To recognize and adapt to invertebrate environments such as the insect cuticle, hemolymph and cadaver, *Metarhizium* spp. need to rapidly respond to changes in nutrient availability, osmolarity and the host immune system [18,63]. In *Magnaporthe*, the *Pth11*-like G-protein coupled receptor (GPCR) is a PHI gene (PHI-base acc: 404) because it mediates cell responses to inductive cues [64]. *M. anisopliae* and *M. acridum* have 54 and 40 putative PTH11-like GPCRs, respectively compared to an average of 32 in other fungi (Table S2, Table S14). The *Metarhizium* sequences could be grouped into six subfamilies (Figure S3). G protein alpha subunits have been extensively studied in fungi and many are required for pathogenicity because they transduce extracellular signals leading to infection-specific development [65]. Distinct roles for three G

protein alpha subunit genes have been revealed in *M. grisea*, *A. nidulans* and *N. crassa*. A fourth G-alpha protein has been identified in the plant pathogens *Stagonospora nodorum* (SNOG_06158) [66], *Ustilago maydis* (UM05385) [67], and the saprophyte *A. oryzae* (BAE63877) [68]. Each of the *Metarhizium* genomes also contain four G-alpha genes. The genes MAA_03488 and MAC_04984 show best hits (>30% similarity) with SNOG_06158, UM05383 and BAE63877, suggesting they may be orthologous. SNOG_06158 is the most highly up-regulated *S. nodorum* G-alpha gene *in planta* [66]. Likewise, MAA_03488 and MAC_04984 are the most highly expressed G-alpha genes during infection of either cockroach or locust cuticles (see below, Table S20).

The chief mechanism used by bacteria for sensing their environment is based on two conserved proteins: a sensor histidine kinase (HK) and an effector response regulator (RR) that functions as a molecular switch controlling diverse activities. In fungi, two component pathways mediate environmental stress responses and hyphal development [69]. *M. anisopliae* and *M. acridum* have 10 and 9 histidine kinases, respectively compared to 3 to 20 in other fungi (Table S2).

To regulate cell function, *M. acridum* has 192 protein kinases as compared to 161 in *M. anisopliae* which is still above the average (143) found in other fungi (Tables S5 and S15). Much of the *M. acridum* expansion involves cyclin dependent and cell division control kinases, suggesting that *M. acridum* has a particularly complex signal transduction cascade controlling cell division. As signal transduction is a critical part of fungal development and infection processes, and accordingly most of the kinases had orthologs in the PHI database (124/161 in MAA and 137/192 in MAC). The high frequency of pseudogenes among kinases (*M. acridum*, 1:6; *M. anisopliae*, 1:8), compared to transporters (*M. anisopliae*, 1:82; *M. acridum*, 1:33) and other gene families suggests that protein kinases have a particularly high rate of turnover (Table S16). Differentially lost genes tend to function in accessory roles so these kinases might have had redundant functions in signal transduction that changed rapidly under strong selective constraints.

Following signaling transduction, physiological responses are regulated by activation of different transcription factors. *M. anisopliae* has 510 putative transcription factors compared to 439 in *M. acridum*, the difference being largely due to *M. anisopliae* having more C2H2 zinc finger and Zn2/Cys6 transcription factors (Tables S5 and S17). These families are also expanded in some *Aspergilli*, where the characterized examples are involved in regulating diverse aspects of primary and secondary metabolism, including protein and polysaccharide degradation [70]. The cAMP response element-binding (CREB) protein, a basic leucine zipper transcription factor (bZIP), is a major downstream transcription factor for cAMP/PKA pathways in mammals [71]. CREB has not been characterized in fungi; however, our transcriptome data shows that a putative bZIP transcription factor (MAA_02048 or MAC_02758) is highly expressed by each *Metarhizium* species coincident with up-regulation of protein kinase A (see below). The physiological role(s) of MAA_02048 are currently under investigation.

Comparative transcriptome analysis

Insect bioassays confirmed that *M. acridum* killed locusts but not cockroaches, while *M. anisopliae* killed both insects (Figure S4). In order to identify the putative signal transduction and metabolic pathways involved in formation of infection structures, we used RNA-Seq to compare transcriptional responses of *M. anisopliae* and *M. acridum* to infection of the optically clear hind wings of adult locusts and cockroaches, respectively. A time period of 24 hours was chosen to focus on the crucial processes involved in prepenetration growth e.g., adhesion to epicuticle, germination and production of appressoria [72].

After sequencing >2.5 million tags for each treatment, it was calculated that >82% of predicted *M. anisopliae* genes and >88% of predicted *M. acridum* genes were expressed during pre-penetration growth. This included more than 80% of the *M. anisopliae* and *M. acridum* genes with sequences similar to those in the PHI database (Table S18). Germination and growth by *M. anisopliae* and *M. acridum* on either insect triggered high level expression of genes associated with translation (e.g., ribosomal proteins) and post-translational modifications (e.g., heat shock proteins) (Figure S5, Table S19). However, otherwise, the two fungi differed greatly from each other in their transcriptional responses to each cuticle, and to a lesser extent the two cuticles elicited different responses from each fungus (Figure 6, Figure S6). The orthologs of many differentially expressed genes are involved in appressorial formation and function in plant pathogens (Table S19), including Cas1 from *Glomerella cingulata* and Mas1 from *M. oryzae* [73]. Three of these genes were among the five most highly expressed *M. acridum* genes on locust cuticle. Their expression levels were ~2-fold lower on cockroach cuticle, similar to a previously characterized cuticle binding adhesin, Mad1 [20]. This is also consistent with a previous study which showed that *M. acridum* up-regulated (~3-fold) a single Mas1-like gene (MAC_03649) in the extracts of locust cuticular lipids but this gene was down-regulated in extracts from beetles (~4-fold) or cicadas (~2-fold) [72].

Formation of appressoria would be expected to involve significant modifications of the germ tube cell wall. Between 6 to 10% of the genes highly expressed by *M. anisopliae* and *M. acridum* on host cuticles encoded cell wall proteins. However, cell wall remodeling may be a greater feature of post penetration development because a microarray study showed that ~20% of insect hemolymph-induced genes were involved in cell wall formation [74]. Evidently, different subsets of genes are required before and after penetration of the cuticle. Suppression-subtractive hybridization identified 200 genes expressed by *M. acridum* in the hemolymph of locusts [75], and only eight of these genes involved in translation were among the 100 genes that were most highly expressed by pre-penetration germlings.

About 60% of the transcripts expressed by *M. anisopliae* in liquid cultures containing insect cuticles encoded secreted products, including many proteases [76], as compared to ~20% of the transcripts in pre-penetration germlings (Table S19), indicating that growth in culture does not mimic the environment experienced on the insect surfaces. Despite the lineage-specific expansion of protease gene families in *Metarhizium* spp., only a few proteases were abundantly expressed by either species on insect epicuticles. Two trypsins were highly expressed by *M. anisopliae* on cuticle surfaces, but similar to most subtilisins, they were not expressed by *M. acridum*. Early expression of proteases triggered by nitrogen starvation may allow *M. anisopliae* to sample the cuticle, resulting in further induction of proteases that could digest the proteinaceous procuticular layer [76]. Consistent with this hypothesis, both *Metarhizium* species expressed several genes involved in recognition of nitrogen starvation signals, including MAA_03429 and MAC_02501, which resemble the STM1-like GPCR responsible for triggering adaptation to nitrogen starvation in fission yeast *Schizosaccharomyces pombe* [77] (Table S20).

The profile of dehydrogenases produced on insect cuticles was used to highlight metabolic pathways that participate in pre-penetration growth. The expression profile of dehydrogenases produced on locust and cockroach cuticles was highly correlated ($r = 0.96$) in *M. anisopliae*, but much less so in *M. acridum* ($r = 0.69$). The most abundant dehydrogenase transcripts expressed by *M.*

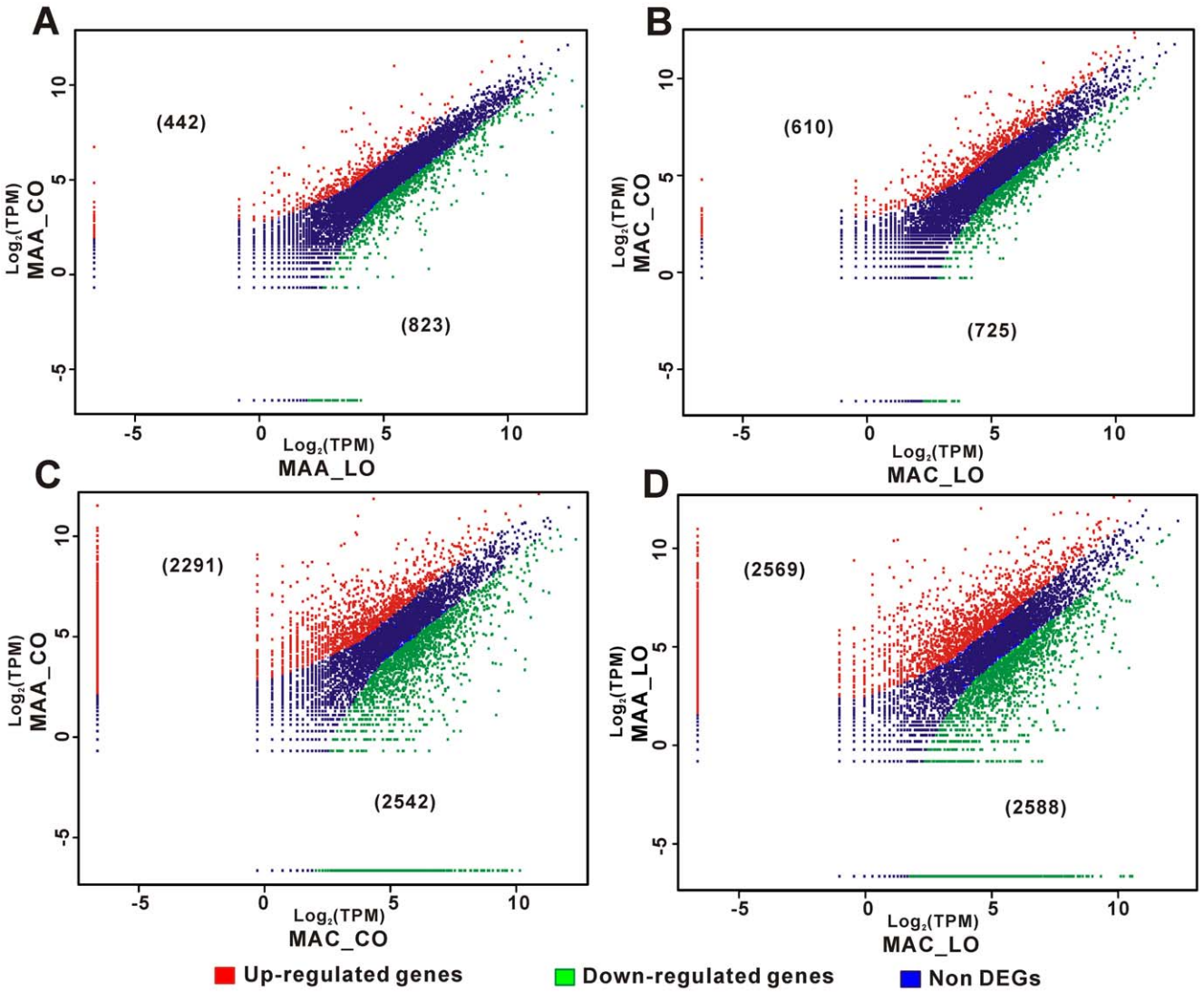

Figure 6. Differential gene expression by *M. anisopliae* (MAA) and *M. acridum* (MAC) on locust (LO) and cockroach (CO) hind wings. Genes differentially expressed by *M. anisopliae* (A) and *M. acridum* (B) infecting cockroach versus locust cuticles. Genes differentially expressed by *M. anisopliae* versus *M. acridum* on cockroach (C) and locust (D) cuticles. The figures in parentheses are the number of genes significantly up- or down-regulated by each fungus.
doi:10.1371/journal.pgen.1001264.g006

anisopliae on both cuticles included enzymes involved in glycolysis, the citric acid cycle and the oxidative branch of the pentose phosphate pathway. Genes involved in metabolizing intracellular lipids, proteins and amino acids were also highly expressed, showing that lipids are an important nutrient reserve, and that there is a high turnover of proteins during the formation of appressoria as previously suggested for *M. oryzae* [78].

Similar to previous observations [21], *M. acridum* germlings only produce appressoria on locust cuticle, and these visually resemble the appressoria produced by *M. anisopliae* on both insect cuticles (Figure 7). Consistent with early host recognition events being key to establishing specificity, *M. acridum* but not *M. anisopliae* transcribed different *Pth11*-like GPCR genes on locust and cockroach cuticles (Figure S3A). The up-regulation of G-protein alpha subunit, phosphatidyl inositol-specific phospholipase C, protein kinase C, Ca/calmodulin-dependent kinase and extracellular signal-regulated protein kinases indicate that the mitogen-activated protein kinase pathway was strongly activated by *M. anisopliae* during infection of both insects and by *M. acridum* infecting locust cuticle. Unexpectedly, *M. acridum* expressed adenylate cyclase and protein kinase A at higher levels on cockroach cuticle even though appressoria formation was not induced (Figure 7, Table S20). Most of the up-regulated signal-tranduction genes were similar to known PHI genes that regulate infection processes in other fungi (Table S6). Overall, our results suggest that *M. anisopliae* and *M. acridum* are able to differentiate diverse host-related stimuli on locust and cockroach cuticles using distinct or shared signaling pathways involving PTH11-like GPCRs, calcium-dependant pathways and MAP kinases that are probably under subtle and sophisticated cross-pathway controls.

Discussion

Recent improvements in next generation sequencing technology and bioinformatics now allows the *de novo* assembly of high quality eukaryotic genomes [79,80]. We used such an approach to provide

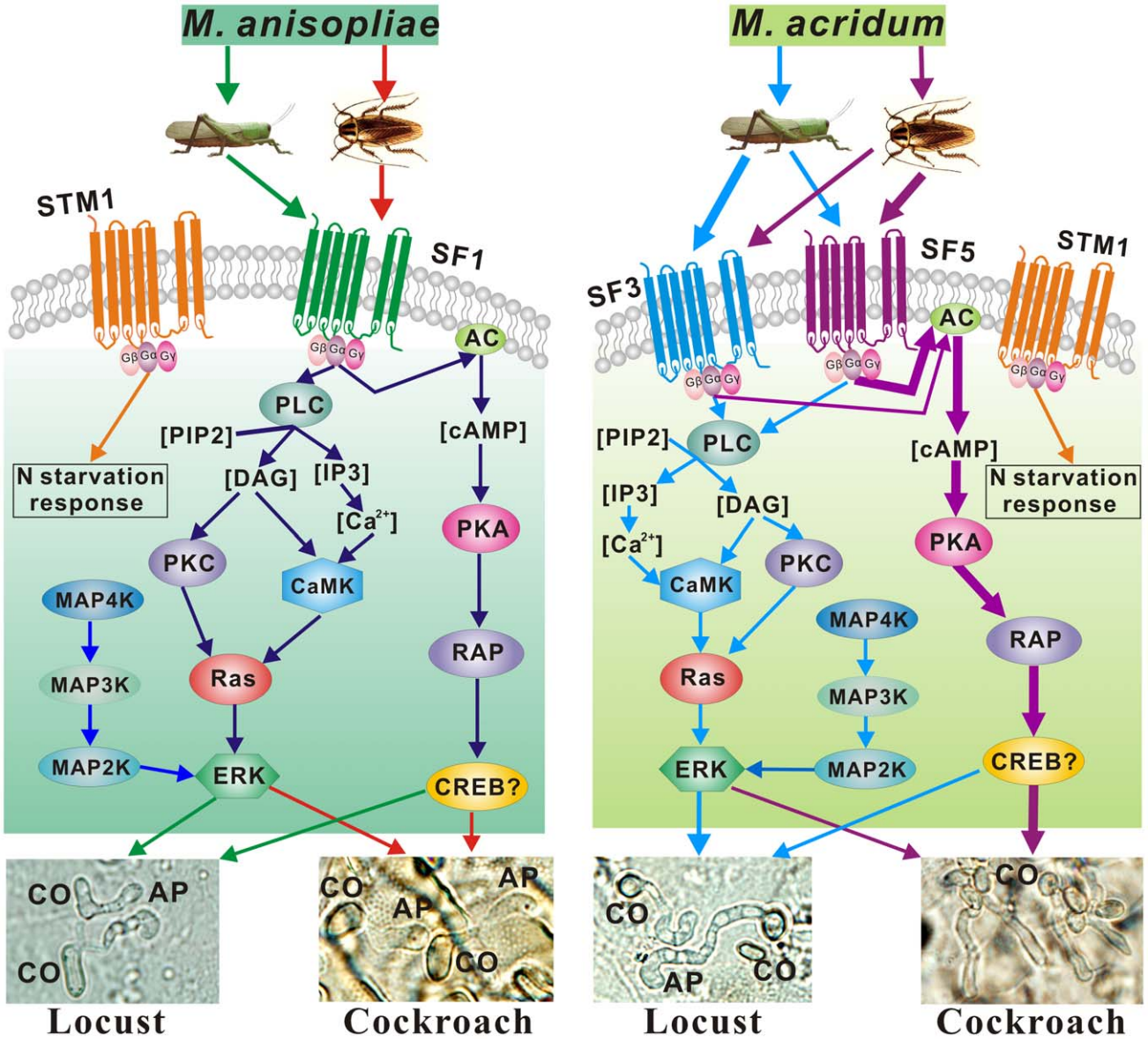

Figure 7. Differentially regulated signaling pathways employed by *M. anisopliae* and *M. acridum* infecting cockroach and locust cuticles. Both the MAP kinase and cAMP dependent protein kinase A (PKA) pathways were activated by *M. anisopliae* and *M. acridum* infecting cockroach and locust cuticles. PLC, phosphatidyl inositol-specific phospholipase C; PIP2, phosphatidylinositol 4,5-bisphosphate; IP3, inositol 1,4,5-triphosphate; DAG, diacylglycerol; PKC, protein kinase C; CaMK, calcium/calmodulin regulated kinase; ERK, extracellular signal-regulated protein kinase; CREB, a basic leucine zipper transcription factor that is a potential cAMP response element-binding protein; CO, conidium; AP, appressorium. Thicker arrows indicate pathways that are more highly expressed by *M. acridum* on either locust or cockroach cuticles.
doi:10.1371/journal.pgen.1001264.g007

the first draft sequences of insect fungal pathogens *M. anisopliae* and *M. acridum*. *Metarhizium* species are the best studied insect pathogenic fungi and thus serve as an excellent starting point for gaining a broad perspective of issues in insect pathology. Sequencing two related species that evolved very different life styles provides a powerful method to derive lists of candidate genes controlling pathogenicity, host specificity and alternative saprophytic life styles. By using the experimentally verified pathogen-host interaction (PHI) gene reference database [33], we found that >16% of the genes encoded by each genome have significant similarities with genes involved in pathogenicity in other fungi, oomycetes or bacteria. Our study also highlighted secreted proteins which are markedly more numerous in *Metarhizium* spp. than in plant pathogens and non-pathogens, pointing to a greater complexity and subtlety in the interactions between insect pathogens and their environments. High resolution RNA-Seq transcriptomic analyses found that the two *Metarhizium* spp. have highly complicated finely-tuned molecular mechanisms for regulating cell differentiations in response to different insect hosts. These were the first large scale transcriptome studies done with insect pathogenic fungi grown under simulated insect parasitism rather than in liquid cultures.

Whole genome analyses indicated that *Metarhizium* spp. are closer to plant endophytes and plant pathogens than they are to animal pathogens like *A. fumigatus* and *C. albicans*. The finding suggested that *Metarhizum* may have evolved from fungi adapted to

grow on plants even though they now infect insects. This inference is supported by the consistent existence of genes for plant degrading enzymes within *Metarhizium* genomes (Table S2). In contrast, fungal pathogens of humans that are seldom recovered from soil, such as *Coccidioides*, exhibit few of these enzymes or none [81]. Even necrophytes such as *Trichoderma reesei* lack many families of plant cell wall degrading enzymes [82], and the existence of such families in *Metarhizium* spp. implies that these species are able to utilize living plant tissues. Potentially, these enzymes could also facilitate colonization of root surfaces but this must remain speculative because the genetic basis for rhizosphere competence is largely obscure in fungi [5,83]. Our identification of the full repertoire of *Metarhizium* genes will help to identify genes responsible for life on the plant root.

M. acridum contains fewer transposase genes than *M. anisopliae* which might be due to differences in repeat-induced point mutation (RIP). Both *M. anisopliae* and *M. acridum* have orthologs (MAA_03836 and MAC_00922) of the *N. crassa* RIP defective gene ($E \leq 10^{-80}$), the only gene known to be required for RIP [84]. The retention of this gene suggests that *M. anisopliae* might have undergone RIP at some stage in its evolution, even though its genome currently shows no bias towards C:G to T:A mutations. Creating new genes through duplication is almost impossible when RIP is very efficient [31], so the apparent loss of RIP in *M. anisopliae* may have been a compromise for the massive expansion of some gene families, though at the cost to *M. anisopliae* of increased transposition. *M. acridum* has a strong RIP bias, but RIP is only functional when meiosis is frequent. *Cordyceps taii* has been described as the sexual type (teleomorph) of *Metarhizium taii* [7,85] but the sexual stages of *M. acridum* (and *M. anisopliae*) are unknown. However, both *Metarhizium* species have a complement of apparently functional genes whose orthologs in *N. crassa* and *A. nidulans* are known to be required for meiosis and sexual development (Table S21). These include putative α-mating type genes and genes with similarity to a high mobility group (HMG) mating type gene, suggesting that they may have the potential to be either self (homothallic) or non-self (heterothallic) fertile under favorable conditions [26]. More studies are required to understand the importance of the RIP mechanisms in the evolution of *Metarhizium* genomes and to determine the frequency of meiosis. Discovering whether *M. anisopliae* and *M. acridum* undergo sexual reproduction also has important implications for understanding the evolution of new strains of these pathogens.

Alternatively to an undiscovered sexual stage, the conservation of sex genes in an asexual species could be due to a recent loss of sexuality, pleiotropy or parasexual recombination following heterokaryon formation [86]. The well known parasexual cycle that occurs in some fungi including *Metarhizium* provides another mechanism for hybridization [87]. As with sexual hybridization there are numerous barriers between vegetative fusions of different fungal species with the major one being vegetative incompatability, which results from heterokaryon incompatability proteins that block exchange of DNA [88]. *M. acridum* has fewer (25 genes) heterokaryon incompatibility proteins than *M. anisopliae* (35 genes), which suggests that *M. acridum* may be less reproductively isolated than *M. anisopliae*. However, it is likely that *M. acridum* with its more specialized lifestyle and narrow environmental range encounters fewer genetically distinct individuals than the more opportunistic *M. anisopliae* (Table S2).

The evolutionary transition of *Metarhizium* spp. to insect pathogenicity must have involved adaptations to insect-based nutrition, as indicated by the large number of proteases, lipases and chitinases that can digest insect cuticles and the host body (Table S6). Except for the lipid outer epicuticle, most of the barriers and nutritional resources in insects are proteinaceous, and *Metarhizium* has a full set of proteases including many different subtilisins, trypsins, chymotrypsins, metalloproteases, aspartyl proteases, cysteine proteases and exo-acting peptidases. The chymotrypsins are *M. anisopliae* specific, and may have been acquired by a horizontal gene transfer event [37]. Otherwise, the ~2–10-fold expanded repertoire of various families of secreted proteases has been the result of preservation by natural selection of duplicated genes. These may have facilitated the adaptation to heterogenous environments. Thus, the abundance of aspartyl proteases and carboxypeptidases (active at low *p*H) and subtilisins and trypsins (active at high *p*H), reflects the ability of *M. anisopliae* to grow in media with a wide range of *p*H values [89]. The ability to produce large quantities of secreted proteases will obviously assist in the rapid degradation of insect host barriers, but the diversity of different proteases might also have been selected because insects frequently exploit anti-fungal protease inhibitors [38].

With the exception of the trypsins, most of the proteases with orthologs in the PHI-base (Table S6) are reported to have a major function in degrading host barriers. Fungal trypsins are regarded as markers of pathogenicity as they are almost exclusively found in pathogens of plants, animals or fungi [90]. *M. anisopliae* has more trypsins than any other sequenced fungus, including *M. acridum*, which indicates a recent evolution of this gene family by gene duplication in *M. anisopliae* (Figure 1C). We also infer that differential gene loss has occurred due to the existence of six trypsin pseudogenes in *M. anisopliae* (Table S16). At least two active trypsins are expressed during insect infection [91], but the role of these trypsins in disease has not been demonstrated. The only sequences similar ($E < 1 \times 10^{-10}$) to *Metarhizium* trypsins in the PHI database are from the oomycete plant pathogen *Phytophthora sojae* (PHI acc.: 652 and 653). Plants produce diverse glucanases to degrade pathogen cell walls, and the *P. sojae* trypsins quench this by degrading the glucanases [92]. It is feasible that a similar strategy could occur in insect-fungus interactions since the β-glucan recognition proteins, β-1,3-glucanases and β-1,4-glucanases involved in insect immune responses are similar to the anti-fungal glucanases produced by plants [93].

To date ~20 *Metarhizium* genes that contribute to infectious capacity have been described [4]. These have provided important new insights into the novel mechanisms by which pathogens evade host immunity by masking cell wall components with a collagen [18], differentially attach to insects or plants with different adhesins [20] and regulate intracellular lipid stores with a perilipin [21]. Some of these genes, like the collagen MCL1 (MAA_01665), seem to be specifically associated with pathogenicity in *M. anisopliae*, showing that analysis of orphan (species-specific) genes will be crucial for a full understanding of pathogenicity. Other genes and gene families are generally associated with pathogenicity and can be predicted with the help of the PHI database. The >370 families of genes categorized as containing PHI genes in *Metarhizium* therefore represent good leads for dissecting the molecular genetics of pathogenicity. Many families like the crotonases involved in fatty acid metabolism [94], the *PacC* transcription factor that mediates the environmental *p*H signal, and the suppressers of defense responses such as catalases and superoxide dismutases have been well documented as virulence factors in diverse pathogens of plants and animals [36]. It would be surprising if they were not involved in *Metarhizium* infection processes.

Other sequences identified from comparisons with the PHI database may be less generic in their impact on pathogenesis. As well as secreted proteins, the interaction between a pathogen and

its host is to a large extent orchestrated by the proteins that are localized to the cell wall or cell membrane, and these categories are well represented in the PHI database. Plant and animal pathogens frequently have a subset of extracellular membrane proteins containing an eight-cysteine domain referred to as CFEM. In plant pathogens, CFEM-containing proteins function as cell-surface receptors or signal transducers, or as adhesion molecules in host-pathogen interactions [64]. Deletion of CFEM-containing proteins produces a cascade of pleiotropic effects in *C. albicans*, most effecting cell-surface-related properties including the ability to form biofilms [95]. The genomic sequences reveal that *Metarhizium* species also have CFEM-containing proteins (MAA_03310, MAA_04981 and MAA_07591 in *M. anisopliae*; MAC_09015 and MAC_09359 in *M. acridum*), and functional analysis is underway to investigate the role they play in *M. anisopliae* development and pathogenesis. There are many other putative PHI protein families that need to be verified as virulence or pathogenicity determinants in *Metarhizium*. For example, CheY-like domain proteins are response regulators in some bacterial two-component signaling systems [96], but their roles in fungi remain to be determined. *Metarhizium* spp. have an average number of histidine kinases compared to other filamentous fungi, and yet *M. anisopliae*, *M. acridum*, *F. graminearum* and *M. oryzae* have 4, 3, 2 and 0 CheY-like proteins, respectively (Table S6), indicating that *M. anisopliae* is comparatively well supplied with putative effector proteins that promote responses to stimuli. Much more unexpectedly, *M. anisopliae* has 6 (*M. acridum* has 1) homologs of heat-labile enterotoxins that play important roles in bacterial pathogenesis [97]. The HMG proteins involved in fungal sexuality are also required for fungal pathogenicity [98]. Both *Metarhizium* species have four HMG proteins that are predicted to be PHI genes (Table S6).

M. anisopliae produces large numbers of proteins and secondary metabolites that might be dedicated to host interaction and countering insect defenses [6]. The identity and molecular functions of most secondary metabolite encoding genes remain to be determined in *Metarhizium*, and it will be intriguing to investigate which of their products are required for pathogenicity and or host specificity. However, with respect to the number of PKS and NRPS genes, *M. anisopliae* appears to possess a greater potential for the production of secondary metabolites than *M. acridum* and other sequenced Ascomycetes. *M. anisopliae* kills hosts quickly via toxins and grows saprophytically in the cadaver. In contrast, *M. acridum* causes a systemic infection of host tissues before the host dies which suggests limited production of toxins, or none [99]. The presence of NRPS MAA_00969 in *M. anisopliae* is remarkable as almost all similar genes encoding host selective toxins were found in the Dothideomycetes [100]. It is unlikely that MAA_00969 and HTST1 (encoding the HC-toxin) evolved independently, and one possibility is that MAA_00969 was acquired by an interspecific horizontal gene transfer event. There is no evidence to date that *M. anisopliae* has a relationship with any plant species that would require a specific toxin, and there are no reports of host-specific toxins in fungal pathogenesis of animals or insects.

Specialization in host range in various *Metarhizium* lineages is associated with a reduction in the range of molecules the fungi can utilize for nutrition or are able to detoxify [101]. Consistent with this is the deficit of dehydrogenases (DHGs) in *M. acridum* relative to *M. anisopliae* or saprophytic fungi (Table 2). *M. anisopliae* also has more cytochrome P450s (CYPs), which are used by fungi to detoxify diverse substrates [102]. Thus, the additional CYPs and DHGs encoded by *M. anisopliae* may enable it to detoxify the toxic repertoires of multiple insect hosts, as compared to *M. acridum* that infects only locusts. CYPS and DHGs also contribute to production of different secondary metabolites by oxidation (CYPs) and dehydroxylation (DHGs) of the backbone structures produced by the PKSs and NRPSs [103]. *M. anisopliae*'s PKS and NRPS clusters contain 18 CYPs and 21 DHGs, while *M. acridum*'s PKS and NRPS clusters contain 3 CYPs and 12 DHGs. The insecticidal destruxin A–F subclasses produced by *M. anisopliae* have the same backbone structure, but more than 30 different analogues [104]. These analogs presumably derive principally from the action of CYPs or DHGs, but the molecular mechanisms have not been determined.

Comparative global transcriptional studies of *M. anisopliae* and *M. acridum* provided a broad-based analysis of gene expression during early colonization processes, particularly in terms of the genes involved in host recognition, metabolic pathways and pathogen differentiation (Figure 7, Figure S6). About 20% of the genes most highly expressed by both *Metarhizium* species are putative PHI genes (Table S19). In spite of the abundance of protease genes in the *Metarhizium* genomes only a few proteases, mostly the trypins were expressed in the early stages of infection. As mentioned above, the trypsins could possibly serve as suppressors of host defenses. Studies in a range of plant pathogens suggest that early infection is characterized by the catabolism of internal lipid stores and that polymerized substrates are used after the readily available substrates are exhausted [65,66]. The transcriptome of *M. anisopliae* shows that it also uses internal lipid stores early in infection, which is consistent with a previous study [21]. Proteases and chitinases are secreted later at very high levels to digest the protein-chitin procuticles [23]. The occurrence of a stress condition during the early phase of the interaction with the insect host was indicated by the massive up-regulation of heat shock proteins (HSPs). MAA_04726 and MAC_01954 are similar ($E<1\times10^{-160}$) to an HSP90 from *C. albicans* that is a crucial virulence factor governing cell drug resistance and morphogenetic transition [105]. The other highly expressed *Metarhizium* HSPs (e.g., HSP30s and HSP70s) are considered to be part of a general defense response and did not resemble sequences in the PHI database.

In spite of differences in infection procedures, we were able to identify some concordance between up-regulated *Metarhizium* genes and metabolic networks up-regulated by *M. oryzae* [78] and the mycoparasite *Trichoderma harzianum* [106]. In particular, during early host colonization, they all up-regulated pathways associated with translation, post-translational modification, and amino acid and lipid metabolism. *Metarhizium* spp. also resemble *M. oryzae* and *T. harzianum* in that pathogenicity is associated with nitrogen deprivation and related stresses, indicating that at least some of the physiological conditions on insects, plants and fungal hosts might be similar. For example, the *S. pombe* STM1 gene links environmental nitrogen with cell differentiation [77]. The up-regulation of similar STM1-like receptors by the three pathogenic fungi could be a common mechanism for linking low levels of nitrogen on the host surfaces with differentiation of infection structures.

In spite of their different host ranges, developmental processes within *M. anisopliae* and *M. acridum* are very similar, e.g. formation of appressoria and blastospores. However, comparatively analyzing their host-invading transcriptomes suggested that recognition might be determined in part by regulatory controls that exclusively limit expression of genes for pathogenicity-related developmental processes to individual hosts. Functional characterization should elucidate whether the expansion in *M. anisopliae* of several families of signal receptors and response elements is indicative of functional redundancy and/or reflective of a need for fine-tuned sensing of

the host environments. The differentially regulated *Pth11* GPCR genes are clear early candidates for further functional analysis to confirm their role as regulators of pathogenicity, and to investigate how their function varies between strains with different host ranges. Such studies could define the core set of host-specific transcripts and identify targets for effectively altering host range.

In conclusion, we have identified significant differences in gene contents and transcriptional regulations between *M. acridum* and *M. anisopliae*, that have led to the latter having a wider biochemical repertoire available for infecting multiple hosts. The genomic sequences will facilitate identifying candidate genes for manipulation to increase the benefits of applying *Metarhizium* not just as an insecticide but also potentially as a biofertilizer. The range of exploitable fungal virulence genes is enormous as besides the putative PHI genes, other virulence factors such as the systems for evading host immunity are of particular interests.

Materials and Methods

Fungal strains

M. anisopliae strain ARSEF 23 has been studied in the laboratory for more than 40 years [107]. It is a generalist insect pathogen that successfully infects locusts, caterpillars, flies, crickets and beetles, amongst others, and is classified as a Group A strain (good germination in many media and production of appressoria against a hard hydrophobic surface in yeast extract medium) [101]. *M. acridum* CQMa 102 can only infect acridids and is being mass produced for large-scale locust control in China [16]. It is classified as a Group D strain (little or no germination in yeast extract or glucose media). A recent taxonomic revision assigns *M. anisopliae* ARSEF 23 to a new species, *viz.*, *M. robertsii* [108].

Genome sequencing and assembly

The genomes of *M. anisopliae* and *M. acridum* were sequenced with the next generation sequencing technology Illumina. DNA libraries with 200 bp, 2 kb and 6 kb inserts were constructed and sequenced with the Illumina Genome Analyzer sequencing technique at the Beijing Genomics Institute at Shenzhen with protocols described previously for the giant panda genome [80]. The genome sequences were assembled using SOAPdenovo [109]. For syntenic relationship analysis, the scaffolds of both genomes were oriented by MEGABLAST for dot plotting and a pair-wise comparison with an Argo Genome Browser [110].

Annotation

Annotations of the genomic sequences of *M. anisopliae* and *M. acridum* were performed with Augustus [111], specifically trained with >6000 unique sequenced *Metarhizium* ESTs, and the annotated information of *F. graminearum* was incorporated as a reference. An *ab initio* predictor, GeneMark [112] was additionally used for ORF prediction with both *Metarhizium* genomes. Thorough manual checks were conducted on parallel comparisons of the results from two prediction methods. All questionable ORFs were individually subjected to Blast searches against the NCBI curated refseq_protein database and the individual prediction with the best hit was selected for each ORF. Pseudogene identification was conducted with the pipeline of PseudoPipe [113]. Transfer RNAs (tRNAs) were predicted with tRNAscan-SE [114] and ARAGORN [115]. Secreted proteins were identified by SignaIP 3.0 analysis (http://www.cbs.dtu.dk/services/SignalP/).

Orthology and phylogenomic analysis

Ortholog conservation in fungi was characterized with Inparanoid 7.0 [116]. In total, 946 orthologous proteins were acquired and aligned with Clustal X 2.0 [117]. A maximum likelihood phylogenomic tree was created using the concatenated amino acid sequences with the program TREE-PUZZLE using the Dayhoff model [118]. The divergence time between species was estimated with the Langley-Fitch method with r8s [119] by calibrating against the reassessed origin of the Ascomycota at 500–650 million years ago [120].

Protein family classification and evolution analysis

Whole genome protein families were classified by InterproScan analysis (http://www.ebi.ac.uk/interpro/) in combination with the Treefam methodology that defines a protein family as a group of genes descended from a common ancestor [121]. To identify potential pathogenicity and virulence genes, whole genome blast searches were conducted against protein sequences in the pathogen-host interaction database (version 3.2, http://www.phi-base.org/) ($E<1\times10^{-5}$). The families of proteases were additionally classified by Blastp against the MEROPS peptidase database (http://merops.sanger.ac.uk/). Transporters were classified based on the Transport Classification Database (http://www.tcdb.org/tcdb/). The cytochrome P450s were named according to Dr. Nelson's P450 database (http://drnelson.utmem.edu/CytochromeP450.html). G-protein coupled receptors, protein kinases, transcription factors and GH families were classified by Blastp against GPCRDB (http://www.gpcr.org/7tm/), KinBase (http://kinase.com/), Fungal Transcription Factor Database (http://ftfd.snu.ac.kr/) and CAZy database (http://www.cazy.org/), respectively. All *Metarhizium* genes with significant hits (E value $\leq 10^{-5}$) to GPCRDB sequences and that contained 7 transmemebrane helices (analyzed with http://www.cbs.dtu.dk/services/TMHMM/) were included as putative GPCRs. To analyze fungal secondary metabolite pathways, the genome annotation data from both species were coordinated and analyzed with the program SMURF (http://www.jcvi.org/smurf/index.php). The evolution of protein family size variation (expansion or contraction) was analyzed using CAFE [32].

Repeat and repeat-induced point mutation (RIP) analysis

Genome repetitive elements were analyzed by Blast against the RepeatMasker library (open 3.2.8) (http://www.repeatmasker.org/) and with the Tandem Repeat Finder [122]. RIP index was determined with the software RIPCAL by reference against the non-repetitive control families [30]. The transposons/retrotransposons encoding transposases/retrotransposases were classified by Blastp analysis against the Repbase (http://girinst.org/).

Transcriptome analysis

The hind wings from locusts (*Locusta migratoria*) and cockroaches (*Periplaneta americana*) were collected and surface sterilized in 37% H_2O_2 (5 min), washed in sterile water twice and dipped in conidial suspensions (2×10^7 spores per ml) of *M. anisopliae* ARSEF 23 or *M. acridum* CQMa 102 for 20 seconds. The inoculated wings were placed on 1% water agar and incubated at 25°C for 24 hrs. The wings with fungal cells were homogenized in liquid nitrogen and the total RNA was extracted with a Qiagen RNeasy mini kit plus on-column treatment with DNase I. Messenger RNA was purified from 6 μg total RNA. After reverse transcription into double strand cDNA for tag preparation according to the massively parallel signature sequencing protocol [123], it was sequenced with an Illumina technique. We omitted tags from further analysis if only one copy was detected or it could be mapped to different transcripts [124]. Other tags were mapped to the genome or annotated genes by allowing if they possessed no more than one nucleotide mismatch. The abundance of each tag was converted to

transcripts per million for quantitative comparison between samples. We used the test of false discovery rate (FDR≤0.001) to estimate the level of differential gene expression by each species under different induction conditions [125].

Insect bioassays

Metarhizium anisoliae and *M. acridum* were tested for their ability to kill adult locusts *Locusta migratoria* and cockroaches *Periplaneta americana*. For these experiments, conidia from each species were applied topically by immersion of cold-immobilized insects into aqueous suspensions of 5×10^8 spores per ml. Each treatment was replicated three times with 15 insects per replicate and the experiments were repeated twice. Mortality was recorded every 12 hours and the lethal time values for 50% mortality (LT_{50}) were estimated [18].

Accession numbers

The Whole Genome Shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession ANDI00000000 for *Metarhizium acridum* and ANDJ00000000 for *Metarhizium anisopliae*, respectively. The RNA-seq data have been deposited at NCBI GEO repository with accession numbers GSM612996, GSM612997, GSM612998 and GSM612999 for the samples of *M. anisopliae* infection of locust, *M. anisopliae* infection of cockroach, *M. acridum* infection of locust and *M. acridum* infection of cockroach, respectively.

Supporting Information

Figure S1 Comparative structural mapping of selected regions from *Metarhizium anisopliae* (MAA) and *M. acridum* (MAC) genomes. (A) An example of highly syntenic relationships between selected regions from MAA and MAC; (B) Examples of lineage specific regions from MAA and MAC containing high densities of species-specific genes, proteins with unknown functions and transposases; (C) Selected syntenic regions showing occurrence of subtilisin duplications in MAA with no counterpoints in MAC.
Found at: doi:10.1371/journal.pgen.1001264.s001 (1.69 MB TIF)

Figure S2 Schematic comparison of a conserved gene cluster in *Metarhizium anisopliae* and *M. acridum* putatively involved in synthesis of secondary metabolites, including ergot alkaloids. The core genes dimethylallyl tryptophan synthase (DMAT, an aromatic prenyltransferase that carries out the first committed step of ergot alkaloid biosynthesis) and nonribosomal peptide synthetases (NRPS) were predicted by SMURF (http://www.jcvi.org/smurf/index.php). Ergot alkaloid proteins A and B are involved in forming the D ring of the ergot alkaloid structural framework. Dynamin GTPases are involved in endocytosis and secretion. Ergot alkaloids are toxins and important pharmaceuticals that are produced biotechnologically on an industrial scale.
Found at: doi:10.1371/journal.pgen.1001264.s002 (0.55 MB TIF)

Figure S3 Characterization of predicted Pth11-like G-protein coupled receptors (GPCR) encoded in the *Metarhizium anisopliae* (MAA) and *M. acridum* (MAC) genomes. (A) Neighbor-joining phylogeny of GPCRs from MAA and MAC showing that they are clustered into six subfamilies. The arrows indicate that the GPCR gene most highly transcribed by MAA on either locust or cockroach cuticles is MAA_06268, whereas the GPCR genes most highly expressed by MAC on locust and cockroach cuticles are MAC_00494 and MAC_03366, respectively; (B) The number of GPCR genes in each subfamily in *M. anisopliae* and *M. acridum*.
Found at: doi:10.1371/journal.pgen.1001264.s003 (1.56 MB TIF)

Figure S4 Bioassays of *Metarhizium anisopliae* and *M. acridum* against adult locusts (A) and cockroaches (B). The median lethal times (LT_{50}) vs. locusts (*Locusta migratoria*) were 8.60±0.15 and 6.54±0.37 days for *M. anisopliae* and *M. acridum*, respectively. For cockroaches (*Periplaneta americana*), the LT_{50} for *M. anisopliae* was 10.10±0.37 days. *M. acridum* was non-pathogenic to cockroaches.
Found at: doi:10.1371/journal.pgen.1001264.s004 (0.54 MB TIF)

Figure S5 Differential gene expression profiling for selected orthologous genes from *Metarhizium anisopliae* and *M. acridum* infecting locust and cockroach cuticles, respectively. (A) Cytochrome P450 genes; (B) Serine protease subtilisin genes; (C) Trypsin genes; (D) GH18 family of chitinase genes. The Heat Map figures were generated using the log2 ratio of corresponding MAA gene transcription per million (TPM) data against MAC TPM data, i.e. $Log2^{(MAA\_TPM/MAC\_TPM)}$. The figures show *M. anisopliae* genes that are up-regulated (red) and down-regulated (green) relative to *M. acridum*.
Found at: doi:10.1371/journal.pgen.1001264.s005 (1.37 MB TIF)

Figure S6 Functional classification of differentially expressed genes. Number of genes expressed by *M. anisopliae* (A) and *M. acridum* (B) infecting cockroach (blue bars) and locust (red bars) cuticles. The number of genes differentially expressed by *M. anisopliae* (blue bars) and *M. acridum* (red bars) on locust (C) and cockroach (D) cuticles.
Found at: doi:10.1371/journal.pgen.1001264.s006 (1.51 MB TIF)

Table S1 The number of protein-coding genes and genome size of *Metarhizium* species and other Ascomycetes.
Found at: doi:10.1371/journal.pgen.1001264.s007 (0.06 MB DOC)

Table S2 Sizes of selected gene families in *Metarhizium* spp. and other Ascomycetes.
Found at: doi:10.1371/journal.pgen.1001264.s008 (0.08 MB DOC)

Table S3 Expression of putative transposase coding genes by *Metarhizium anisopliae* and *M. acridum* during infection of cockroach (CO) and locust (LO) cuticles.
Found at: doi:10.1371/journal.pgen.1001264.s009 (0.06 MB DOC)

Table S4 The number of pseudogenes (by class) in the *Metarhizium anisopliae* and *M. acridum* genomes.
Found at: doi:10.1371/journal.pgen.1001264.s010 (0.05 MB DOC)

Table S5 Expanded (highlighted in green) and contracted (highlighted in yellow) protein families in *Metarhizium anisopliae* and *M. acridum*.
Found at: doi:10.1371/journal.pgen.1001264.s011 (0.10 MB DOC)

Table S6 Major protein families of *Metarhizium* spp. and plant pathogenic fungi implicated in the pathogen-host interactions.
Found at: doi:10.1371/journal.pgen.1001264.s012 (0.18 MB DOC)

Table S7 Trypsin proteases encoded in *Metarhizium anisopliae* and *M. acridum* genomes, arranged by MEROPS family.
Found at: doi:10.1371/journal.pgen.1001264.s013 (0.08 MB DOC)

Table S8 Subtilisin proteases encoded in *Metarhizium anisopliae* and *M. acridum* genomes, arranged by MEROPS family.
Found at: doi:10.1371/journal.pgen.1001264.s014 (0.10 MB DOC)

Table S9 Aspartyl proteases encoded in *Metarhizium anisopliae* and *M. acridum* genomes, arranged by MEROPS family.
Found at: doi:10.1371/journal.pgen.1001264.s015 (0.08 MB DOC)

Table S10 Carbohydrate-degrading enzymes in *Metarhizium* species and other fungi, arranged by GH family.
Found at: doi:10.1371/journal.pgen.1001264.s016 (0.10 MB DOC)

Table S11 Identification of genes coding for membrane transporters in *Metarhizium anisopliae* and *M. acridum*.
Found at: doi:10.1371/journal.pgen.1001264.s017 (0.08 MB DOC)

Table S12 Cytochrome P450 (CYP) genes encoded in *Metarhizium anisopliae* and *M. acridum* genomes, arranged by CYP family.
Found at: doi:10.1371/journal.pgen.1001264.s018 (0.12 MB DOC)

Table S13 The number of core genes involved in the biosynthesis of secondary metabolites in different fungi.
Found at: doi:10.1371/journal.pgen.1001264.s019 (0.06 MB DOC)

Table S14 The number of G-protein coupled receptors encoded in *Metarhizium anisopliae* and *M. acridum* genomes, arranged by class.
Found at: doi:10.1371/journal.pgen.1001264.s020 (0.06 MB DOC)

Table S15 The number of genes coding for protein kinases in *Metarhizium anisopliae* and *M. acridum*, arranged by class.
Found at: doi:10.1371/journal.pgen.1001264.s021 (0.06 MB DOC)

Table S16 Classification of conserved domains retained in the pseudogenes of *Metarhizium anisopliae* and *M. acridum*.
Found at: doi:10.1371/journal.pgen.1001264.s022 (0.06 MB DOC)

Table S17 The number of genes coding for transcription factors in *Metarhizium anisopliae* and *M. acridum*, arranged by family.
Found at: doi:10.1371/journal.pgen.1001264.s023 (0.08 MB DOC)

Table S18 General features of the high throughput transcriptome analysis.
Found at: doi:10.1371/journal.pgen.1001264.s024 (0.06 MB DOC)

Table S19 The 100 most highly expressed genes in *Metarhizium anisopliae* and *M. acridum* infecting locust and cockroach cuticles.
Found at: doi:10.1371/journal.pgen.1001264.s025 (0.21 MB DOC)

Table S20 Differential expression of genes involved in signal transduction.
Found at: doi:10.1371/journal.pgen.1001264.s026 (0.10 MB DOC)

Table S21 Sexuality-related genes in *Metarhizium anisopliae* and *M. acridum* and their orthologs in *Aspergillus nidulans* and *Neurospora crassa*.
Found at: doi:10.1371/journal.pgen.1001264.s027 (0.14 MB DOC)

Acknowledgments

The authors appreciate the help of Prof. David Nelson for cytochrome P450 classification and Dr. Jongsun Park for transcription factor analysis. We also acknowledge the sequencing and assembly help from Beijing Genomics Institute at Shenzhen, China.

Author Contributions

Conceived and designed the experiments: RJ St. Leger, C Wang. Performed the experiments: Q Gao, SH Ying, Y Shang, Z Duan, W Huang, B Wang. Analyzed the data: Q Gao, K Jin, SH Ying, Y Zhang, G Xiao, Y Shang, Z Duan, X Hu, XQ Xie, G Zhou, G Peng, Z Luo, W Huang, B Wang, W Fang, S Wang, Y Zhong, LJ Ma, GP Zhao, Y Pei, MG Feng, Y Xia, C Wang. Contributed reagents/materials/analysis tools: Y Pei, MG Feng, Y Xia. Wrote the paper: RJ St. Leger, C Wang.

References

1. Zimmerman G (2007) Review on safety of the entomopathogenic fungus *Metarhizium anisopliae*. Biocontrol Sci Technol 17: 879–920.
2. Lomer CJ, Bateman RP, Johnson DL, Langewald J, Thomas M (2001) Biological control of locusts and grasshoppers. Annu Rev Entomol 46: 667–702.
3. de Faria MR, Wraight SP (2007) Mycoinsecticides and Mycoacaricides: A comprehensive list with worldwide coverage and international classification of formulation types. Biol Control 43: 237–256.
4. St. Leger RJ, Wang CS (2010) Genetic engineering of fungal biocontrol agents to achieve greater efficacy against insect pests. Appl Microbiol Biotechnol 85: 901–907.
5. St. Leger RJ (2008) Studies on adaptations of *Metarhizium anisopliae* to life in the soil. J Invertebr Pathol 98: 271–276.
6. Roberts DW, St. Leger RJ (2004) *Metarhizium* spp., cosmopolitan insect-pathogenic fungi: mycological aspects. Adv Appl Microbiol 54: 1–70.
7. Liu ZY, Liang ZQ, Whalley AJ, Yao YJ, Liu AY (2001) *Cordyceps brittlebankisoides*, a new pathogen of grubs and its anamorph, *Metarhizium anisopliae* var. *majus*. J Invertebr Pathol 78: 178–182.
8. Isaka M, Kittakoop P, Kirtikara K, Hywel-Jones NL, Thebtaranonth Y (2005) Bioactive substances from insect pathogenic fungi. Acc Chem Res 38: 813–823.
9. Kim HG, Song H, Yoon DH, Song BW, Park SM, et al. (2010) *Cordyceps pruinosa* extracts induce apoptosis of HeLa cells by a caspase dependent pathway. J Ethnopharmacol 128: 342–351.
10. Silva WOB, Santi L, Berger M, Pinto AFM, Guimaraes JA, et al. (2009) Characterization of a spore surface lipase from the biocontrol agent *Metarhizium anisopliae*. Proc Biochem 44: 829–834.
11. Pereira JL, Noronha EF, Miller RN, Franco OL (2007) Novel insights in the use of hydrolytic enzymes secreted by fungi with biotechnological potential. Lett Appl Microbiol 44: 573–581.
12. Gottar M, Gobert V, Matskevich AA, Reichhart JM, Wang C, et al. (2006) Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. Cell 127: 1425–1437.
13. Wang C, Butt TM, St. Leger RJ (2005) Colony sectorization of *Metarhizium anisopliae* is a sign of ageing. Microbiology 151: 3223–3236.
14. Li L, Pischetsrieder M, St. Leger RJ, Wang CS (2008) Associated links among mtDNA glycation, oxidative stress and colony sectorization in *Metarhizium anisopliae*. Fungal Genet Biol 45: 1300–1306.
15. Driver F, Milner RJ, Trueman JWH (2000) A Taxonomic revision of *Metarhizium* based on sequence analysis of ribosomal DNA. Mycol Res 104: 135–151.
16. Peng GX, Wang ZK, YinYP, Zeng DY, Xia YX (2008) Field trials of *Metarhizium anisopliae* var. *acridum* (Ascomycota: Hypocreales) against oriental migratory locusts, *Locusta migratoria manilensis* (Meyen) in Northern China. Crop Prot 27: 1244–1250.
17. Hajek AE, McManus ML, Delalibera I (2007) A review of introductions of pathogens and nematodes for classical biological control of insects and mites. Biol Control 41: 1–13.
18. Wang C, St. Leger RJ (2006) A collagenous protective coat enables *Metarhizium anisopliae* to evade insect immune responses. Proc Natl Acad Sci U S A 103: 6647–6652.
19. Duan Z B, Shang YF, Gao Q, Zheng P, Wang CS (2009) A phosphoketolase Mpk1 of bacterial origin is adaptively required for full virulence in the insect-pathogenic fungus *Metarhizium anisopliae*. Environ Microbiol 11: 2351–2360.
20. Wang C, St. Leger RJ (2007) The MAD1 adhesin of *Metarhizium anisopliae* links adhesion with blastospore production and virulence to insects, and the MAD2 adhesin enables attachment to plants. Eukaryot Cell 6: 808–816.
21. Wang C, St. Leger RJ (2007) The *Metarhizium anisopliae* perilipin homolog MPL1 regulates lipid metabolism, appressorial turgor pressure, and virulence. J Biol Chem 282: 21110–21115.
22. Fang W, Fernandes EK, Roberts DW, Bidochka MJ, St. Leger RJ (2010) A laccase exclusively expressed by *Metarhizium anisopliae* during isotropic growth is involved in pigmentation, tolerance to abiotic stresses and virulence. Fungal Genet Biol 47: 602–607.

23. St. Leger RJ, Joshi L, Bidochka MJ, Roberts DW (1996) Construction of an improved mycoinsecticide overexpressing a toxic protease. Proc Natl Acad Sci U S A 93: 6349–6354.
24. Wang C, St. Leger RJ (2007) A scorpion neurotoxin increases the potency of a fungal insecticide. Nat Biotechnol 25: 1455–1456.
25. Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, et al. (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317: 1400–1402.
26. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422: 859–868.
27. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, et al. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. Oryzae*. Nature 438: 1105–1115.
28. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, et al. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464: 367–373.
29. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, et al. (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet 4: e1000046. doi:10.1371/journal.pgen.1000046.
30. Hane, JK, Oliver RP (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics 9: 478.
31. Galagan JE, Selker EU (2004) RIP: the evolutionary cost of genome defense. Trends Genet 20: 417–423.
32. de Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22: 1269–1271.
33. Winnenburg R, Urban M, Beacham A, Baldwin TK, Holland S, et al. (2008) PHI-base update: additions to the pathogen host interaction database. Nucleic Acids Res 36: D572–D576.
34. Baldwin TK, Winnenburg R, Urban M, Rawlings C, Koehler J (2006) The pathogen-host interactions database (PHI-base) provides insights into generic and novel themes of pathogenicity. Mol Plant Microbe Interact 19: 1451–1462.
35. St Leger RJ, Staples RC, Roberts DW (1992) Cloning and regulatory analysis of starvation-stress gene, ssgA, encoding a hydrophobin-like protein from the entomopathogenic fungus, *Metarhizium anisopliae*. Gene120: 119–1124.
36. Sexton AC, Howlett BJ (2006) Parallels in fungal pathogenesis on plant and animal hosts. Eukaryot Cell 5: 1941–1949.
37. Screen SE, St. Leger RJ (2000) Cloning, expression, and substrate specificity of a fungal chymotrypsin. Evidence for lateral gene transfer from an actinomycete bacterium. J Biol Chem 275: 6689–6694.
38. Bagga S, Hu G, Screen SE, St. Leger RJ (2004) Reconstructing the diversification of subtilisins in the pathogenic fungus *Metarhizium anisopliae*. Gene 324: 159–169.
39. Schaller M, Borelli C, Korting HC, Hube B (2005) Hydrolytic enzymes as virulence factors of *Candida albicans*. Mycoses, 48: 365–377.
40. Fan Y, Fang W, Guo S, Pei X, Zhang Y, et al. (2007) Increased insect virulence in *Beauveria bassian*a strains overexpressing an engineered chitinase. Appl Environ Microbiol 73: 295–302.
41. Beys da Silva WO, Santi L, Schrank A, Vainstein MH (2010) *Metarhizium anisopliae* lipolytic activity plays a pivotal role in *Rhipicephalus* (*Boophilus*) *microplus* infection. Fungal Biol 114: 10–15.
42. Voigt CA, Schäfer W, Salomon S (2005) A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. Plant J 42: 364–375.
43. Morschhauser J (2010) Regulation of multidrug resistance in pathogenic fungi. Fungal Genet Biol 47: 94–106.
44. Fang W, St. Leger RJ (2010) *Mrt*, a gene unique to fungi, encodes an oligosaccharide transporter and facilitates rhizosphere competency in *Metarhizium robertsii*. Plant Physiol 154: 1549–1557.
45. Suvarna K, Bartiss A, Wong B (2000) Mannitol-1-phosphate dehydrogenase from *Cryptococcus neoformans* is a zinc-containing long-chain alcohol/polyol dehydrogenase. Microbiology 146: 2705–2713.
46. Itoh S, Fukuzumi S (2007) Monooxygenase activity of type 3 copper proteins. Acc Chem Res 40: 592–600.
47. Nelson DR (1999) Cytochrome P450 and the individuality of species. Arch Biochem Biophys 369: 1–10.
48. Jarrold SL, Moore D, Potter U, Charnley AK (2007) The contribution of surface waxes to pre-penetration growth of an entomopathogenic fungus on host cuticle. Mycol Res 111: 240–249.
49. Pedrini N, Crespo R, Juárez MP (2007) Biochemistry of insect epicuticle degradation by entomopathogenic fungi. Comp Biochem Physiol C Toxicol Pharmacol 146: 124–137.
50. Ferrer-Sevillano F, Fernandez-Canon JM (2007) Novel phacB-encoded cytochrome P450 monooxygenase from *Aspergillus nidulans* with 3-hydroxyphenylacetate 6-hydroxylase and 3,4-dihydroxyphenylacetate 6-hydroxylase activities. Eukaryot Cell 6: 514–520.
51. Mendonca Ade L, da Silva CE, de Mesquita FL, Campos Rda S, Do Nascimento RR, et al. (2009) Antimicrobial activities of components of the glandular secretions of leaf cutting ants of the genus *Atta*. Antonie Van Leeuwenhoek 95: 295–303.
52. Podobnik B, Stojan J, Lah L, Krasevec N, Seliskar M, et al. (2008) CYP53A15 of *Cochliobolus lunatus*, a target for natural antifungal compounds. J Med Chem 51: 3480–3486.
53. Mitsuguchi H, Seshime Y, Fujii I, Shibuya M, Ebizuka Y, et al. (2009) Biosynthesis of steroidal antibiotic fusidanes: functional analysis of oxidosqualene cyclase and subsequent tailoring enzymes from *Aspergillus fumigatus*. J Am Chem Soc 131: 6402–6411.
54. Tsitsigiannis DI, Keller NP (2006) Oxylipins act as determinants of natural product biosynthesis and seed colonization in *Aspergillus nidulans*. Mol Microbiol 59: 882–892.
55. Hu G, St. Leger RJ (2002) Field studies using a recombinant mycoinsecticide (*Metarhizium anisopliae*) reveal that it is rhizosphere competent. Appl Environ Microbiol 68: 6383–6387.
56. Wang C, Skrobek A, Butt TM (2004) Investigations on the destruxin production of the entomopathogenic fungus *Metarhizium anisopliae*. J Invertebr Pathol 85: 168–174.
57. Moon YS, Donzelli BG, Krasnoff SB, McLane H, Griggs MH, et al. (2008) *Agrobacterium*-mediated disruption of a nonribosomal peptide synthetase gene in the invertebrate pathogen *Metarhizium anisopliae* reveals a peptide spore factor. Appl Environ Microbiol 74: 4366–4380.
58. Donzelli BG, Krasnoff SB, Churchill AC, Vandenberg JD, Gibson DM (2010) Identification of a hybrid PKS-NRPS required for the biosynthesis of NG-391 in *Metarhizium robertsii*. Curr Genet 56: 151–162.
59. Freimoser FM, Hu G, St. Leger RJ (2005) Variation in gene expression patterns as the insect pathogen *Metarhizium anisopliae* adapts to different host cuticles or nutrient deprivation *in vitro*. Microbiology 151: 361–371.
60. Xu Y, Orozco R, Kithsiri Wijeratne EM, Espinosa-Artiles P, Leslie Gunatilaka AA, et al. (2009) Biosynthesis of the cyclooligomer depsipeptide bassianolide, an insecticidal virulence factor of *Beauveria bassiana*. Fungal Genet Biol 46: 353–364.
61. Böhnert HU, Fudal I, Dioh W, Tharreau D, Notteghem JL, et al. (2004) A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. Plant Cell 16: 2499–2513.
62. Panaccione DG, Scott-Craig JS, Pocard JA, Walton JD (1992) A cyclic peptide synthetase gene required for pathogenicity of the fungus *Cochliobolus carbonum* on maize. Proc Natl Acad Sci USA 89: 6590–6594.
63. Wang CS, Duan ZB, St. Leger RJ (2008) The MOS1 osmosensor of *Metarhizium anisopliae* is required for adaptation to insect host hemolymph. Eukaryot Cell 7: 302–309.
64. Kulkarni RD, Thon MR, Pan H, Dean RA (2005) Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. Genome Biol 6: R24.
65. Solomon PS, Tan KC, Sanchez P, Cooper RM, Oliver RP (2004) The disruption of a Galpha subunit sheds new light on the pathogenicity of *Stagonospora nodorum* on wheat. Mol Plant Microbe Interact 17: 456–466.
66. Hane JK, Lowe RG, Solomon PS, Tan KC, Schoch CL (2007) Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. Plant Cell 19: 3347–3368.
67. Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T, et al. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature 444: 97–101.
68. Lafon A, Han KH, Seo JA, Yu JH, d'Enfert C (2006) G-protein and cAMP-mediated signaling in aspergilli: a genomic perspective. Fungal Genet Biol 43: 490–502.
69. Lin CH, Chung KR (2010) Specialized and shared functions of the histidine kinase- and HOG1 MAP kinase-mediated signaling pathways in *Alternaria alternata*, a filamentous fungal pathogen of citrus. Fungal Genet Biol 47: 818–827.
70. Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, et al. (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. Nat Biotechnol 25: 221–231.
71. Sands WA, Palmer TM (2008) Regulating gene transcription in response to cyclic AMP elevation. Cell Signal 20: 460–466.
72. Wang C, St. Leger RJ (2005) Developmental and transcriptional responses to host and nonhost cuticles by the specific locust pathogen *Metarhizium anisopliae* var. *acridum*. Eukaryot Cell 4: 937–947.
73. Irie T, Matsumura H, Terauchi R, Saitoh H (2003) Serial analysis of gene expression (SAGE) of *Magnaporthe grisea*: genes involved in appressorium formation. Mol Genet Genomics 270: 181–189.
74. Wang C, Hu G, St. Leger RJ (2005) Differential gene expression by *Metarhizium anisopliae* growing in root exudate and host (*Manduca sexta*) cuticle or hemolymph reveals mechanisms of physiological adaptation. Fungal Genet Biol 42: 704–718.
75. Zhang C, Xia Y (2009) Identification of genes differentially expressed *in vivo* by *Metarhizium anisopliae* in the hemolymph of *Locusta migratoria* using suppression-subtractive hybridization. Curr Genet 55: 399–407.
76. Freimoser FM, Hu G, St. Leger RJ (2005) Variation in gene expression patterns as the insect pathogen *Metarhizium anisopliae* adapts to different host cuticles or nutrient deprivation *in vitro*. Microbiology 151: 361–371.
77. Chung KS, Won M, Lee SB, Jang YJ, Hoe KL, et al. (2001) Isolation of a novel gene from *Schizosaccharomyces pombe*: stm1^{+} encoding a seven-transmembrane loop protein that may couple with the heterotrimeric Galpha 2 protein, Gpa2. J Biol Chem 276: 40190–40201.
78. Oh Y, Donofrio N, Pan H, Coughlan S, Brown DE, et al. (2008) Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus *Magnaporthe oryzae*. Genome Biol 9: R85.

79. Nowrousian M, Stajich JE, Chu M, Engh I, Espagne E, et al. (2010) *De novo* assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. PLoS Genet 6: e1000891. doi:10.1371/journal.pgen.1000891.
80. Li R, Fan W, Tian G, Zhu H, He L, et al. (2010) The sequence and *de novo* assembly of the giant panda genome. Nature 463: 311–317.
81. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, et al. (2009) Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. Genome Res 19: 1722–1731.
82. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol 26: 553–560.
83. Dearnaley JDW (2006) The fungal endophytes of *Erythrorchis cassythoides* - Is this orchid saprophytic or parasitic? Aust Mycol 25: 51–57.
84. Freitag M, Williams RL, Kothe GO, Selker EU (2002) A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. Proc Natl Acad Sci U S A 99: 8802–8807.
85. Sung GH, Hywel-Jones NL, Sung JM, Luangsa-Ard JJ, Shrestha B, et al. (2007) Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. Stud Mycol 57: 5–59.
86. Paoletti M, Rydholm C, Schwier EU, Anderson MJ, Szakacs G, et al. (2005) Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. Curr Biol 15: 1242–1248.
87. Messias CL, Azevedo JL (1980) Parasexuality in the deuteromycete *Metarhizium anisopliae*. Trans Br Mycol Soc 75: 473–477.
88. Glass NL, Kaneko I (2003) Fatal attraction: nonself recognition and heterokaryon incompatibility in filamentous fungi. Eukaryot Cell 2: 1–8.
89. St. Leger RJ, Joshi L, Roberts DW (1998) Ambient pH is a major determinant in the expression of cuticle-degrading enzymes and hydrophobin by *Metarhizium anisopliae*. Appl Environ Microbiol 64: 709–713.
90. Dubovenko AG, Dunaevsky YE, Belozersky MA, Oppert B, Lord JC (2010) Trypsin-like proteins of the fungi as possible markers of pathogenicity. Fungal Biol 114: 151–159.
91. St Leger RJ, Joshi L, Bidochka MJ, Rizzo NW, Roberts DW (1996) Biochemical characterization and ultrastructural localization of two extracellular trypsins produced by *Metarhizium anisopliae* in infected insect cuticles. Appl Environ Microbiol 62: 1257–1264.
92. Rose JK, Ham KS, Darvill AG, Albersheim P (2002) Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counter-defense mechanism by plant pathogens. Plant Cell 14: 1329–1345.
93. Tanaka H, Ishibashi J, Fujita K, Nakajima Y, Sagisaka A, et al. (2008) A genome-wide analysis of genes and gene families involved in innate immunity of *Bombyx mori*. Insect Biochem Mol Biol 38: 1087–1110.
94. Piekarska K, Mol E, van den Berg M, Hardy G, van den Burg J (2006) Peroxisomal fatty acid beta-oxidation is not essential for virulence of *Candida albicans*. Eukaryot Cell 5: 1847–1856.
95. Pérez A, Pedrós B, Murgui A, Casanova M, López-Ribot JL (2006) Biofilm formation by *Candida albicans* mutants for genes coding fungal proteins exhibiting the eight-cysteine-containing CFEM domain. FEMS Yeast Res 6: 1074–1084.
96. Wolanin PM, Webre DJ, Stock JB (2003) Mechanism of phosphatase activity in the chemotaxis response regulator CheY. Biochemistry 42: 14075–14082.
97. Liang S, Hajishengallis G (2010) Heat-labile enterotoxins as adjuvants or anti-inflammatory agents. Immunol Invest 39: 449–467.
98. Lu JP, Feng XX, Liu XH, Lu Q, Wang HK, et al. (2007) Mnh6, a nonhistone protein, is required for fungal development and pathogenicity of *Magnaporthe grisea*. Fungal Genet Biol 44: 819–829.
99. Samuels RI, Charnley AK, Reynolds SE (1988) The role of destruxins in the pathogenicity of 3 strains of *Metarhizium anisopliae* for the tobacco hornworm, *Manduca sexta*. Mycopathologia 104: 51–58.
100. Wolpert TJ, Dunkle LD, Ciuffetti LM (2002) Host-selective toxins and avirulence determinants: what's in a name? Annu Rev Phytopathol 40: 251–285.
101. St. Leger RJ, May B, Allee LL, Frank DC, Roberts DW (1992) Genetic differences in allozymes and in formation of infection structures among isolates of the entomopathogenic fungus *Metarhizium anisopliae*. J Invertebr Pathol 60: 89–101.
102. George HL, VanEtten HD (2001) Characterization of pisatin-inducible cytochrome p450s in fungal pathogens of pea that detoxify the pea phytoalexin pisatin. Fungal Genet Biol 33: 37–48.
103. Ehrlich KC, Chang PK, Scharfenstein LL, Jr., Cary JW, Crawford JM, et al. (2010) Absence of the aflatoxin biosynthesis gene, *norA*, allows accumulation of deoxyaflatoxin B1 in *Aspergillus flavus* cultures. FEMS Microbiol Lett 305: 65–70.
104. Molnar I, Gibson DM, Krasnoff SB (2010) Secondary metabolites from entomopathogenic Hypocrealean fungi. Nat Prod Rep 27: 1241–1275.
105. Shapiro RS, Uppuluri P, Zaas AK, Collins C, Senn H, et al. (2009) Hsp90 orchestrates temperature-dependent *Candida albicans* morphogenesis via Ras1-PKA signaling. Curr Biol 19: 621–629.
106. Lorito M, Woo SL, Harman GE, Monte E (2010) Translational research on trichoderma: from 'omics to the field. Annu Rev Phytopathol 48: 395–417.
107. Rangel DE, Butler MJ, Torabinejad J, Anderson AJ, Braga GU, et al. (2006) Mutants and isolates of *Metarhizium anisopliae* are diverse in their relationships between conidial pigmentation and stress tolerance. J Invertebr Pathol 93: 170–182.
108. Bischoff JF, Rehner SA, Humber RA (2009) A multilocus phylogeny of the *Metarhizium anisopliae* lineage. Mycologia 101: 512–530.
109. Li R, Yu C, Li Y, Lam TW, Yiu SM, et al. (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966–1967.
110. Engels R, Yu T, Burge C, Mesirov JP, DeCaprio D, et al. (2006) Combo: a whole genome comparative browser. Bioinformatics 22: 1782–1783.
111. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics 24: 637–644.
112. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Res 18: 1979–1990.
113. Zhang Z, Carriero N, Zheng D, Karro J, Harrison PM, et al. (2006) PseudoPipe: an automated pseudogene identification pipeline. Bioinformatics 22: 1437–1439.
114. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964.
115. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32: 11–16.
116. Ostlund G, Schmitt T, Forslund K, Köstler T, Messina DN, et al. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res 38: D196–203.
117. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
118. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504.
119. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia 98: 838–849.
120. Lücking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT (2009) Fungi evolved right on track. Mycologia 101: 810–822.
121. Li H, Coghlan A, Ruan J, Coin LJ, Hérichè JK, et al. (2006) TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res 34: D572–D580.
122. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573–580.
123. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18: 630–634.
124. Mourier T, Willerslev E (2010) Large-scale transcriptome data reveals transcriptional activity of fission yeast LTR retrotransposons. BMC Genomics 11: 167.
125. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29: 1165–1188.