

RESEARCH ARTICLE

Open Access



Genome sequence and comparative analysis of clavicipitaceous insect-pathogenic fungus *Aschersonia badia* with *Metarhizium* spp.

Yamini Agrawal, Tarun Narwani and Srikrishna Subramanian*

Abstract

Background: *Aschersonia badia* [(*Ab*) Teleomorph: *Hypocrella siamensis*] is an entomopathogenic fungus that specifically infects scale insects and whiteflies. We present the whole genome sequence of *Ab* and its comparison with two clavicipitaceous fungi *Metarhizium robertsii* (*MR*: generalist entomopathogen) and *M. acridum* (*MAC*: acridid-specific entomopathogen) that exhibit variable host preferences. Here, through comparative analysis of pathogen-host interacting genes, carbohydrate active enzymes, secondary metabolite biosynthesis genes, and sexuality genes, we explore the proteins with possible virulence functions in clavicipitaceous fungi. Comprehensive overview of GH18 family chitinases has been provided to decipher the role of chitinases in clavicipitaceous fungi that are either host specific or generalists.

Results: We report the 28.8 Mb draft genome of *Ab* and its comparative genome analysis with *MR* and *MAC*. The comparative analyses suggest expansion in pathogen-host interacting gene families and carbohydrate active enzyme families in *MR*, whilst their contraction in *Ab* and *MAC* genomes. The multi-modular NRPS gene (*dtxS1*) responsible for biosynthesis of the secondary metabolite destruxin in *MR* is not conserved in *Ab*, similar to the specialist pathogen *MAC*. An additional siderophore biosynthetic gene responsible for acquisition of iron was identified in *MR*. Further, the domain survey of chitinases suggest that the CBM50 (LysM) domains, which participate in chitin-binding functions, were not observed in *MAC*, but were present in *Ab* and *MR*. However, apparent differences in frequency of CBM50 domains associated with chitinases of *Ab* and *MR* was identified, where *MR* chitinases displayed a higher proportion of associated CBM50 domains than *Ab* chitinases.

Conclusions: This study suggests differences in distribution of *dtxS1* and chitinases in specialists (*Ab* and *MAC*) and generalists (*MR*) fungi. Our analysis also suggests the presence of a siderophore biosynthetic gene in the *MR* genome which perhaps aids in enhanced virulence potential and host range. The variation in association of CBMs, being higher in generalists (*MR*) and lower in specialists (*Ab* and *MAC*) fungi may further be responsible for the differences in host affiliation.

Keywords: Chitinase, Destruxin, Host-specificity, Pathogen-host interactions

* Correspondence: krishna@imtech.res.in
CSIR-Institute of Microbial Technology, Sector 39-A, Chandigarh 160036, India



Background

The fungal family Clavicipitaceae (Hypocreales, Ascomycota) includes many insect-pathogens such as *Metarhizium* spp. and these fungi have been exploited for their mycoinsecticidal abilities. Family Clavicipitaceae, as defined in modern taxonomy [Clavicipitaceae *sensu stricto* (*s.s.*)], includes three lineages specific to scale insects and whiteflies, referred to as *Hypocrella*, *Regiocrella* and *Torrubiella*; and one generalist lineage, referred to as *Metacordyceps* [1]. In this study, we have selected *Aschersonia badia* (*Ab*), an insect-pathogenic fungus belonging to the *Hypocrella* lineage for whole genome sequencing.

Aschersonia spp. are insect-pathogenic fungi that specifically infect whiteflies (Homoptera, Aleyrodidae) and scale insects (Homoptera, Coccidae). These are predominantly found in tropical and sub-tropical areas. *Aschersonia* spp. [*A. aleyrodis* (Teleomorph: *Hypocrella libera*)] were one of the first fungal entomopathogens utilized as a biocontrol agent [2] and was reported as the cause of epizootics amongst whitefly populations in greenhouses and guava and citrus groves during the 20th century in various parts of the world such as Azerbaijan, Bulgaria, China, Florida, Jamaica, Japan and Russia [3–6]. Moreover, *Aschersonia* spp. show adaptation to low relative humidity [7], perseverance on plant exteriors [8], and compatibility with insect parasitoids [9] in the management of whitefly pests. However, these fungi take a long time to grow in culture, and are not effective against all host stages and this has limited their successful exploitation against insect pests [10].

Metarhizium spp. (Clavicipitaceae *s.s.*, Metacordyceps) are well known entomopathogenic fungi and are best suited targets for biocontrol measures. *Metarhizium robertsii* (*MR*; previous name *M. anisopliae*) is a broad-spectrum insect pathogenic fungus that has been approved by the United States Environmental Protection Agency (USEPA) as an active ingredient for pest control [11, 12]. *M. acridum* (*MAC*) is a specialist pathogen against locust [13]. These diverse features of clavicipitaceous entomopathogens drive our interest towards comparison of *Ab* genome with the *MAC* and *MR* genomes that are adapted to different host ranges and lifestyles in order to get insights into the evolution of host affiliation of fungal entomopathogenicity.

Fungal entomopathogens are known to exhibit contact-based infection through the host cuticle unlike bacterial and viral entomopathogens which are required to be ingested by the target pest [14]. The primary structural component of host cuticle (arthropod exoskeleton) and cell walls of filamentous fungi is chitin which is the second most abundant natural biopolymer after cellulose. Enzymatic degradation is a crucial step in entry of the fungal spore into the insect body where chitinases play an important role in hydrolyzing the chitin-rich insect cuticle

[15]. *Metarhizium* spp. are known to be prolific producers of chitinases, and many of the *MR* chitinases have been suggested to be involved in pathogenicity [16, 17]. However, in-depth analysis of chitinases and their relation with pathogenic lifestyles of clavicipitaceous entomopathogens have not been explored. Therefore, in the present study, we explore the diversity of chitinases in *Ab*, *MAC* and *MR*.

Methods

Fungal strain and maintenance

Ab strain MTCC 10142 was retrieved from the Microbial Type Culture Collection (MTCC), CSIR-Institute of Microbial Technology, Chandigarh, India. As per MTCC records, the *Ab* strain was isolated from Homopteran larva. Fungal culture was grown on potato dextrose agar medium and incubated at 25 °C for 20–25 days. DNA isolation from the fresh mycelia was performed using ZR Fungal/Bacterial DNA kit (Zymo Research, Catalogue number D6005) as per the instructions provided in user manual.

Genome sequencing, assembly and annotation

2x100 paired-end shotgun sequencing (average insert size of 350 bp) of the *Ab* genome was performed using Illumina HiSeq 1000 technology at the Centre for Cellular and Molecular Platforms (C-CAMP) Bangalore, India. The reads were filtered using NGS QC Toolkit v2.2.3 [18] with length cutoff 70 % and base call quality greater than 20; reads were also trimmed for adapters using FastQC [19] and CLCbio wb6.0 genomics workbench (<http://www.clcbio.com>). The datasets were filtered for the presence of reads from bacterial DNA that may be present as contaminants. These were then further processed for genome assembly. The high quality reads were assembled into contigs using SPAdes 2.5.1 [20]. Before running SPAdes as an assembler, its 'read-error-correction' module was used to rectify any orientation issues and InDels in the paired reads. Then its Assembly protocol was run at K-mer range 51–69 with step of +2 and '-careful' option. The contigs obtained from these assemblies were scaffolded using SSPACE v2.0 [21]. Completeness measure was predicted using Core Eukaryotic Genes Mapping Approach (CEGMA) v2.5 with default parameters [22].

Genome annotation was performed using MAKER pipeline with Augustus as the gene predictor [23, 24]. *Fusarium graminearum* (*Fg*) was selected, as a model for gene calling, from the Augustus pre-trained dataset. EST and protein homology evidence of *Fg* was also provided to MAKER. The repeats were masked using RepeatMasker v4.0.3 (<http://www.repeatmasker.org/>). *MAC* CQMa 102 (PRJNA38715) and *MR* (PRJNA230500) genomes were reannotated using the same methodology for

consistency. Both reannotated and the original set of genes/proteins from *MAC* and *MR* genomes were used for the computational analyses. The predicted proteins were subjected to BLASTp against the non-redundant (nr) database at the National Centre for Biotechnology Information (NCBI) using an E-value cutoff of 10^{-5} . The draft genome of *Ab* (*Hypocrella siamensis*) was deposited at NCBI-GenBank under the accession number JMQE00000000.

Phylogenetic analysis

To validate the phylogenetic position of *Ab*, *MAC* and *MR*, four gene markers: mitochondrial ATP6 (*atp6*), the largest and second largest subunits of RNA polymerase II (*rpb1* and *rpbII*) and β -tubulin (*tub*) genes were retrieved from the *Ab*, *MAC* and *MR* genomes and were concatenated. These concatenated sequences were included in the four gene concatenated sequence dataset of hypocrealean fungi [1]. Maximum Likelihood (ML), Maximum Parsimony (MP) and Neighbor-Joining (NJ) phylogenetic reconstruction was performed using Randomized Axelerated Maximum Likelihood (RAxML) v8.0 [25], Phylogenetic Analysis Using Parsimony (PAUP) v4.0 [26] and Molecular Evolutionary Genetics Analysis (MEGA) v6.0 [27], respectively. Ambiguously aligned regions were excluded from the analysis using the trimAl tool [28]. Bootstrap analysis with 1000 replicates was performed.

Cluster of orthologous (COG) groups

The COG clusters were created by subjecting the predicted proteomes to BLASTp against the COG database [29] 2014 update (<http://www.ncbi.nlm.nih.gov/COG/>) with an E-value threshold of 10^{-5} . For each COG mapping, functional category assignment was given based on COG category letter associations. Reciprocal best BLAST hits (E-value threshold of 10^{-10}) obtained using PERL script of Proteinortho v2.0 [30] were used to characterize the orthologous proteins among all three genomes. Computational Analysis of gene Family Evolution (CAFE) v3.0 [31] was used to estimate the gene family expansions and contractions in *Ab*, *MAC* and *MR* for the CAZymes glycoside hydrolases (GHs, 55 families), glycosyltransferases (GTs, 37 families), carbohydrate esterases (CEs, 10 families), auxiliary activities (AAs, 10 families) and carbohydrate binding modules (CBMs, 17 families) and the trasposase gene families (DNA transposase gene families: 10, Retrotransposase gene families: 7). ML Newick tree prepared through MEGA was used as an input for CAFE. *P*-value cut-off of 0.01 was selected to calculate the significant changes in gene numbers.

Protein domain search

Secondary metabolite biosynthetic clusters were identified by using the predicted proteome as query on the

antiSMASH server [32]. Pathogen Host interacting (PHI) partners were identified by subjecting the predicted proteomes to BLASTp against the PHI database [33] v3.6 with an E-value threshold of 10^{-5} . The domain architecture of chitinases and destruxins were annotated using the Pfam v27.0 and SMART v7.0 databases [34–36]. Molecular weight was evaluated using the Protparam tool [37].

Phylogenetic analysis of non ribosomal peptide synthase (NRPS) biosynthetic gene types

NRPS genes identified from the three genomes *Ab*, *MAC* and *MR* were combined with the phylogenomic NRPS dataset reported previously [38]. Their phylogenetic analysis was performed as described above. Multiple sequence alignment file used for the phylogenetic analysis is provided as Additional file 1 (A, B).

Sequence similarity of destruxin gene was calculated using SIAS server (<http://imed.med.ucm.es/Tools/sias.html>). After running BLASTp [Query: *dtxS1* (*MR*), Database: *Ab*_Proteome], the alignments of interest (*dtxS1* and *Ab-Node74-Gene7905*) were uploaded in SIAS server and % identities and % similarities were obtained.

Carbohydrate active enzymes (CAZy) search

CAZY classes Glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs) and carbohydrate binding modules (CBMs) were searched from the CAZY database. CAZY families were assigned to the proteins by subjecting the predicted proteome of all the three genomes to dbCAN web server [39].

Transposable elements (TEs) and Repeat induced point mutation (RIP) analysis

TEs were classified by subjecting the genome sequences of *Ab*, *MAC* and *MR* to BLASTn against the Repbase (<http://girinst.org>) libraries of RepeatMasker (<http://www.repeatmasker.org>). RIP indices were estimated using the RIPCAL v1.0 [40].

Identification of MAT idiomorphs

Mating type genes control sexual development in fungi [41]. Putative mating type loci were identified by the presence of conserved domains and the sequence similarities to the corresponding MAT genes in other filamentous fungi. Conserved alpha domain, Proline–Proline–Phenylalanine domain, High–Mobility–Group (HMG) domain with a DNA binding site and MatA HMG box with a DNA binding site were searched to identify the mating type loci MAT1-1-1, MAT1-1-2, MAT1-1-3 and MAT1-2-1, respectively [42, 43]. Synteny for the mating type locus among all the three genomes was analyzed by locating the position of MAT locus and their flanking genes. The absence of any gene was reconfirmed by mapping

back the reads on these genes from related fungi such as *Metarhizium majus*, *Trichoderma reesei*, *Fusarium graminearum*, *Beauveria bassiana*, *Cordyceps militaris*, *Ophiocordyceps sinensis*, *Hirsutella thompsonii*, *Isaria farinose*, etc. This was done mainly to verify that they were not missing in the three genomes due to assembly errors.

Glycoside hydrolase family 18 (GH18) chitinases search

The chitinase sequences were searched in the predicted proteomes of *Ab*, *MAC* and *MR* genomes using ProfileScan (<http://www.csd.hku.hk/bruhek/gcgdoc/profilescan.html>) with the conserved active site signature motif [LIVMF]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E (Prosite No. PS01095) [44]. The identified chitinase sequences from the three genomes were aligned using PCMA [45] and profile hidden Markov model was built using HMMER v3.1b1 [46]. This HMMER profile was further used to scan the predicted proteome of the genomes to find other chitinase sequences that have mutations/SNPs at conserved consensus motifs. GH18 chitinase sequences of fungal origin were selected from a previously reported dataset [47]. All retrieved chitinases from the three genomes and GH18 family chitinase sequences of fungal origin were assembled and aligned using MAFFT v7.0 [48]. For subgroup A, B and C fungal chitinases, three separate alignment files were prepared as previously described [47, 49, 50] (Additional file 2).

Results

Phylogenetic relationships of *Ab* and *Metarhizium* species

The phylogenetic analysis grouped *Ab*, *MAC* and *MR* within the Clavicipitaceae *s.s.* clade (Additional file 3: Figure S1). *Ab* MTCC 10142 strain was clustered within the *Hypocrella* clade whereas *MAC* and *MR* were clustered within *Metacordyceps* clade (ML, MP and NJ analysis parameters presented 100 % support values) thus validating the phylogenetic identification of the organisms under study. The phylogenetic tree displayed the paraphyletic occurrence of family Clavicipitaceae into clades A, B and C, *viz.*, Clavicipitaceae *s.s.*, Ophiocordycipitaceae and Cordycipitaceae, respectively and the grouping of Clavicipitaceae *s.s.* clade into four lineages: three specific to scale insects and whiteflies, *Hypocrella*, *Torrubiella* and *Claviceps*; with one as a generalist, *Metacordyceps* (Additional file 3: Figure S1).

Genome features

The Illumina HiSeq 1000 shotgun sequencing of *Ab* genome resulted in 12,732,904 paired-end (PE) reads (insert size of 350 bp and length 101 bp). A total of 11,990,083 high-quality reads after filtering and trimming were assembled into 2,034 contigs (maximum contig length as 246 Kb and minimum as 600 bp) with a coverage of 90x.

Genome size as assembled with Illumina PE data was 28.8 Mb (N50: 27770). Assembly of the core genomic regions was predicted to be 94 % complete based on CEGMA analysis. Table 1 lists the genome sequencing, assembly and annotated features for *Ab*. The total number of genes obtained from *Ab*, *MAC* and *MR* genomes was 9292, 10853 and 12880, respectively. The core gene pool is represented by 5586 genes, common in all three genomes and the predicted unique gene content for *Ab*, *MAC* and *MR* was 3402, 3352 and 5247 genes, respectively (Additional file 3: Figure S2A). A comparative distribution of functional features of predicted proteins suggests that about 18 % function in cellular processes and signaling, 19 % in information storage and processing, and 36–39 % in metabolism (Additional file 3: Figure S2B).

Genes involved in biosynthesis of secondary metabolites

Secondary metabolites are quite diverse in fungi and vary substantially in their structure and biological activity. Some of the secondary metabolites secreted by fungi aid in pathogenicity mechanisms and toxicity reactions while others are useful for therapeutic purposes [51]. The most common gene types involved in their synthesis include NRPS: non-ribosomal peptide synthases, PKS: polyketide synthases, and TS: terpene synthases. All the predicted genes involved in synthesis of secondary metabolites were identified from the three genomes where *MR* shows higher abundance of these predicted genes than *Ab* and *MAC* genomes (Table 2). One of the appealing observations is the presence of siderophores in the *MR* genome (Table 2) that could perhaps be related to its enhanced pathogenicity and host range [52–54].

NRPS are modular enzymes involved in the biosynthesis of diverse low molecular weight bioactive metabolites that participate in various fundamental metabolic functions of fungi such as growth and development, virulence, niche specific functions, stress responses, etc.

Table 1 Genome sequencing, assembly and annotated features

Features	<i>Ab</i>
Genome Size	28.8 Mbp
PE reads	12,732,904
High quality reads	11,990,083
N50 contigs	27,770
Contigs	2,034
Max. contig length	246 Kbp
Min. contig length	600 bp
Coverage	90x
Total length	28,856,701
GC (%)	53.1
Total genes annotated	9,292

Table 2 Gene types involved in the biosynthesis of secondary metabolites in *Ab*, *MAC* and *MR* genomes

Gene type	<i>Ab</i>	<i>MAC</i>	<i>MR</i>
NRPS	9	13	15
NRPS-T1PKS	5	4	4
NRPS-Terpene	0	2	0
NRPS-T1PKS-Terpene	0	1	1
T1PKS	11	10	16
T3PKS	1	0	0
Unusual HgID/E-like PKS	2	0	1
Terpene	7	6	13
Terpene-T1PKS	1	0	2
Other	8	5	13
Siderophore	0	0	1
Total	44	41	66

[38, 55]. A set of three core domains (A-T-C) forms the basic module of NRPS where A is adenylation domain for recognizing and activating the substrate (initiation), T or PCP is thiolation domain or peptidyl carrier protein (PCP) for binding and transferring the activated substrate (elongation) and C is the condensation domain for forming peptide bonds between elongated amino acid (aa) chain (termination) [38, 56]. Additionally, accessory domains like epimerization (E) domain responsible for altering the aa configuration from L to D form, N-methylation domain (nMT) responsible for transfer of the methyl group to the substrate may also be present in the NRPS module [38, 57]. The number and order of these modules on each NRPS genes is known to govern the structural variability of the resulting product [38, 58, 59].

The domain architecture of all predicted NRPS genes in the three genomes were compared (Fig. 1). These genes were nomenclatured as *X_nrps-z*, where *X* is the

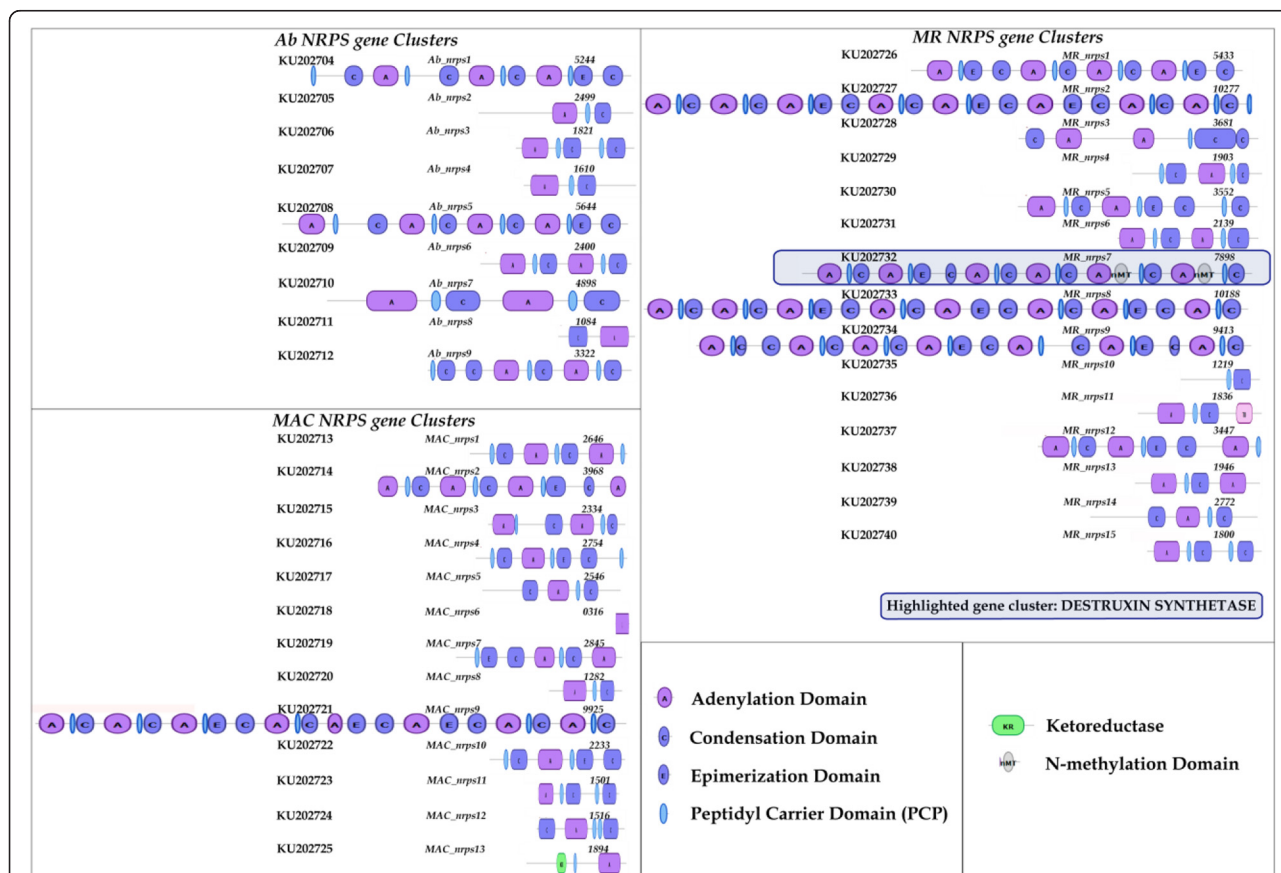


Fig. 1 Secondary metabolite biosynthetic NRPS genes and their domain distributions in *Ab*, *MAC* and *MR* genomes, as identified by antiSMASH server. A total of 9, 13 and 15 NRPS genes were identified in *Ab*, *MAC* and *MR* genomes. The core module of NRPS contains three domains (A-T-C), A: adenylation domain for initiation, T/PCP: thiolation domain/peptidyl carrier protein for elongation, and C: condensation domain for termination. The hexa-modular structure of NRPS gene (along with nMT in each of the last two A-T-C modules) encodes destruxin synthetase (*dxs1*), an insecticidal host-specific gene. This hexa-modular structure of NRPS gene and nMT are only seen in *MR* highlighted gene, suggesting presence of *dxs1* in *MR* genome while its absence in *Ab* and *MAC* genomes. GenBank accession number of each gene is shown on left side of the gene schematic. The numbers by the end of each gene schematic denote the gene length (amino acids)

organism label (*Ab*, *MAC* or *MR*) and *z* is the designated number. Conservation of basic A-T-C domains was seen in all predicted genes, with the exception of *Ab_nrps8*, *MAC_nrps6*, *MAC_nrps13* and *MR_nrps10*. *MR* genome contained many multimodular NRPS genes in comparison to the other two genomes suggesting their evolution to perform niche adaptation functions [38]. One such insecticidal host-specific multimodular gene is destruxin synthetase (*dtxS1*), which is reported selectively in the *MR* genome, and is absent in the *MAC* genome [58–60]. This *dtxS1* contains the hexa-modular structure of NRPS (along with nMT domain in each of the last two A-T-C modules) and is postulated to be one of the factors in evolution of fungal affiliation to diverse host niches [59, 60]. In the present analysis, none of the NRPS genes was observed to have six A-T-C modules in *Ab* and *MAC*. Further, no nMT in conjugation with A-T-C module was observed, suggesting the absence of this gene in the specialist pathogen *Ab*, as reported previously for *MAC* [58, 59]. However to further confirm this, *Ab* gene homologous to the *dtxS1* gene of *MR* was identified based on sequence similarity in the *Ab* genome. BLASTp indicated 81 % similarity (31 % identity) of *Ab*-Node74 gene-7905 (*Ab*-nrps-7905 (~5.2 kb) with *dtxS1* gene (~8 kb) of *MR*. No similar gene in *MAC* was identified. Therefore, domains of *MR dtxS1* gene and *Ab*-nrps-7905 were compared. It revealed presence of only three complete A-T-C modules in *Ab*-nrps-7905 that cannot be referred to as destruxin synthetase.

Phylogenetic relationships of predicted NRPS genes in *Ab*, *MAC* and *MR* were further assessed to identify their orthologous genes and to trace their duplication and divergence history (Fig. 2, S3). List of predicted NRPS gene cluster sequences identified in *Ab*, *MAC* and *MR* genomes is provided in Additional file 4. Fungal NRPS have been classified into two main groups: mono/bimodular NRPS of ancient origin, involving bacterial and fungal NRPS (subfamilies clustered: α-aminoadipate reductases (AAR), ChNPS10-like synthetases, ChNPS11/ETP toxin module 1 synthetases, ChNPS12/ETP toxin module 2 synthetases, Cyclosporin synthetases (CYCLO) and PKS-NRPS); and multimodular NRPS of recent origin, involving exclusively fungal NRPS (Subfamilies clustered: Euascomycete-only synthetases (EAS) and siderophore synthetases (SID)) [38]. In the present phylogenomic analysis, we observed the clustering of predicted genes from the three genomes mostly in the EAS subfamily that is specific for the ascomycete lineage (Fig. 2). One gene each from *Ab*, *MAC* and *MR* (*Ab_nrps7*, *MAC_nrps6* and *MR_nrps3*) grouped in the SID subfamily (Fig. 2), responsible for iron uptake [61]. Both EAS and SID subfamilies are observed to exhibit low conservation of their domain architecture owing to autonomous domain gain/loss mechanisms that likely

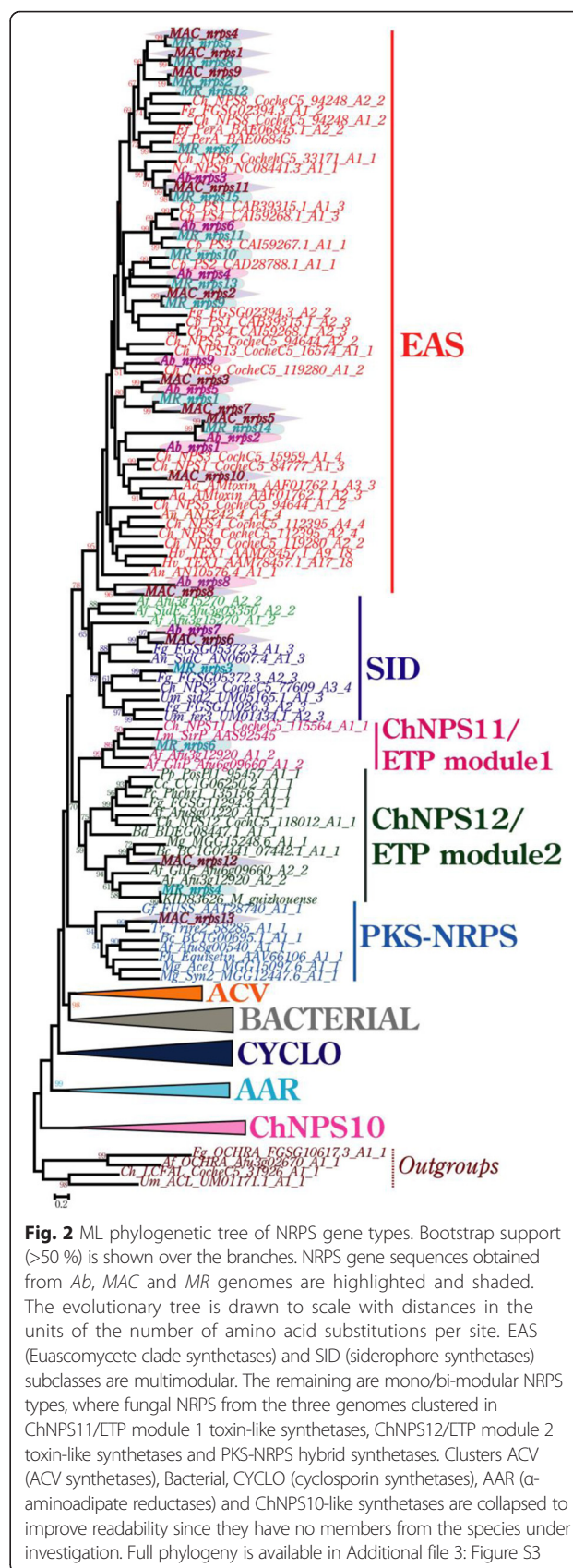


Fig. 2 ML phylogenetic tree of NRPS gene types. Bootstrap support (>50 %) is shown over the branches. NRPS gene sequences obtained from *Ab*, *MAC* and *MR* genomes are highlighted and shaded. The evolutionary tree is drawn to scale with distances in the units of the number of amino acid substitutions per site. EAS (Euascomycete clade synthetases) and SID (siderophore synthetases) subclasses are multimodular. The remaining are mono/bi-modular NRPS types, where fungal NRPS from the three genomes clustered in ChNPS11/ETP module 1 toxin-like synthetases, ChNPS12/ETP module 2 toxin-like synthetases and PKS-NRPS hybrid synthetases. Clusters ACV (ACV synthetases), Bacterial, CYCLO (cyclosporin synthetases), AAR (α-aminoadipate reductases) and ChNPS10-like synthetases are collapsed to improve readability since they have no members from the species under investigation. Full phylogeny is available in Additional file 3: Figure S3

aid in specialized niche adaptive roles [38]. *MR_nrps6* was observed to cluster with ChNPS11/ETP toxin module 1 synthetases whereas *MAC_nrps12* and *MR_nrps4* were clustered with ChNPS12/ETP toxin module 2 synthetases subfamilies (Fig. 2).

Pathogen-host interacting (PHI) genes

The identification of entomopathogenicity-related genes is important in order to understand the pathogenicity mechanisms of the fungal taxa that could be explored in development of mycoinsecticidal strategies. PHI database (<http://www.phi-base.org/>) catalogues 3944 genes and 6473 interactions (<http://www-phi4.phibase.org/releaseNote.htm>) to be involved in pathogenicity. In our analysis, we obtained 1901, 2093 and 2378 of PHI genes from *Ab*, *MAC* and *MR* genomes respectively. Figure 3

shows the important steps in pathogen-host interactions and the distribution of statistically significant (not evolved due to random chance) PHI genes (with CAFE *P*-value < 0.01) in the three genomes as evidenced from the CAFE analysis. PHI genes are known to be involved in host recognition, signaling, adhesion, appressoria development, formation of infection structures, host colonization, conidiation, spore germination, cuticle penetration, nutrition, etc. [62–66]. We identified genes with matches to those in PHIbase that have demonstrated roles in insect-pathogenicity in the three genomes. The detail of each PHI gene identified is provided in Additional file 5. Our analyses suggest a variability in the distribution of these genes in the three genomes, indicating a contraction of PHI families in *Ab* (CAFE average expansion is negative = -1.45) and *MAC*

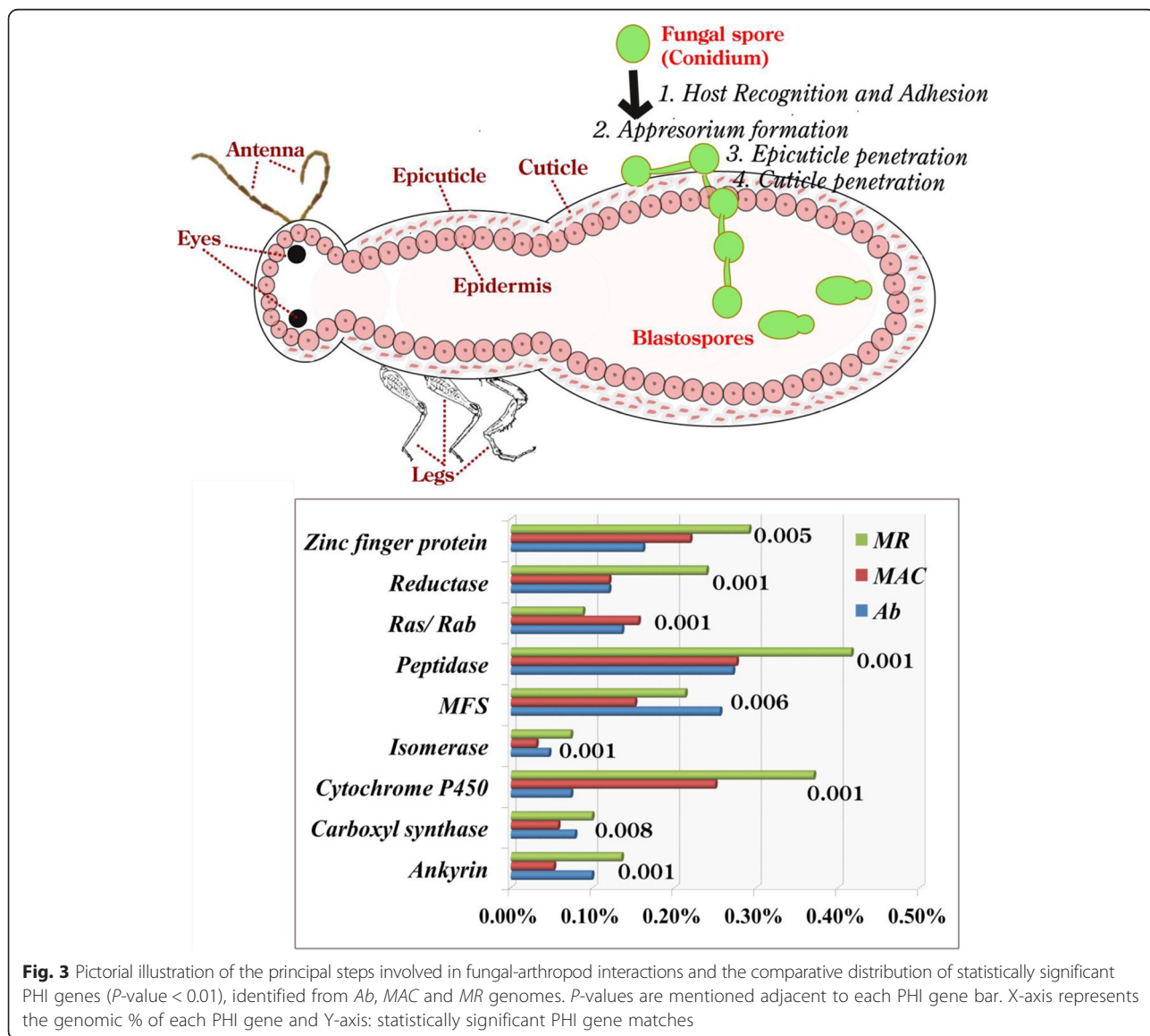


Fig. 3 Pictorial illustration of the principal steps involved in fungal-arthropod interactions and the comparative distribution of statistically significant PHI genes (*P*-value < 0.01), identified from *Ab*, *MAC* and *MR* genomes. *P*-values are mentioned adjacent to each PHI gene bar. X-axis represents the genomic % of each PHI gene and Y-axis: statistically significant PHI gene matches

(CAFE average expansion is negative = -0.88), whereas they are significantly expanded in *MR* (CAFE average expansion = 1.46) as suggested by previous reports where also larger gene families were observed in generalists genomes as compared to the specialists genomes due to dynamic loss and gain of genes [13, 60].

Carbohydrate active enzymes

A total of six classes of carbohydrate active enzymes as provided in CAZy database have been compared. The genes from each class: carbohydrate binding modules (CBMs), glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterase (CE), and auxiliary activities (AAs) are provided in Additional files 6, 7, 8, 9, 10 and 11, respectively.

The prevalence of associated functional modules, CBMs was found to be quite variable in the three fungal genomes (Additional file 6). CBM38 (inulin binding function) and CBM54 (xylan, glucan and chitin binding function) were observed to be unique to *Ab*. CBM1 (cellulose binding domain) family members were observed in *MAC* and *MR*. CBM18 and CBM50 (LysM) family modules known to be associated with chitinase catalytic domains and are implicated to bind chitin were found to be commonly present in all three genomes. However, their frequency was lowest in *MAC* genome. CAFE analysis suggests a contraction of CBM families in *Ab* and *MAC* genomes, whereas they are expanded in the *MR* genome.

GHs, GTs, PLs, CEs and AAs are the catalytic enzyme classes that direct the lysis, synthesis or alteration of carbohydrate moieties [67–70]. The three genomes show the predominance of GH16 (xyloglucan: xyloglucosyltransferase), GH18 (chitinase) and GH76 (α -1,6-mannanase) family enzymes (Additional file 7). GH9 (endoglucanase) and GH120 (β -xylosidase) were observed to be unique to the *Ab* genome. GH54 (α -L-arabinofuranosidase; β -xylosidase), GH88 (d-4,5-unsaturated β -glucuronidase), GH95 (fucosidase) and GH117 (α -1,3-L-neoagarooligosaccharide hydrolase) were exclusive to the *MR* genome. GT2 (cellulose/chitin synthase) and GT32 (α -1,6-mannosyltransferase/N-acetylglucosaminyltransferase) were commonly observed in the three genomes (Additional file 8). *Ab* displayed exclusive occurrence of PL10 (pectate lyase) family enzyme, whereas *MAC* and *MR* displayed PL7 (alginate lyase), PL8 (hyaluronate lyase) and PL20 (endo- β -1,4-glucuronan lyase) family enzymes (Additional file 9). CE1 (acetyl xylan esterase; carboxylesterase) and CE10 (arylesterase; carboxyl esterase; acetylcholinesterase; sterol esterase) family enzymes were predominantly present in all the three genomes (Additional file 10). AA3 (glucose-methanolcholine (GMC) oxidoreductases) and AA7 (oligosaccharide oxidase) were the most frequent families in the three

genomes (Additional file 11). CAFE analysis indicated the contraction of CAZY families (GHs, GTs, CEs and AAs) in *Ab* and *MAC* genomes, whereas their expansion in *MR* genome.

Repeat induced point mutation (RIP) and diversity of transposable elements (TE)

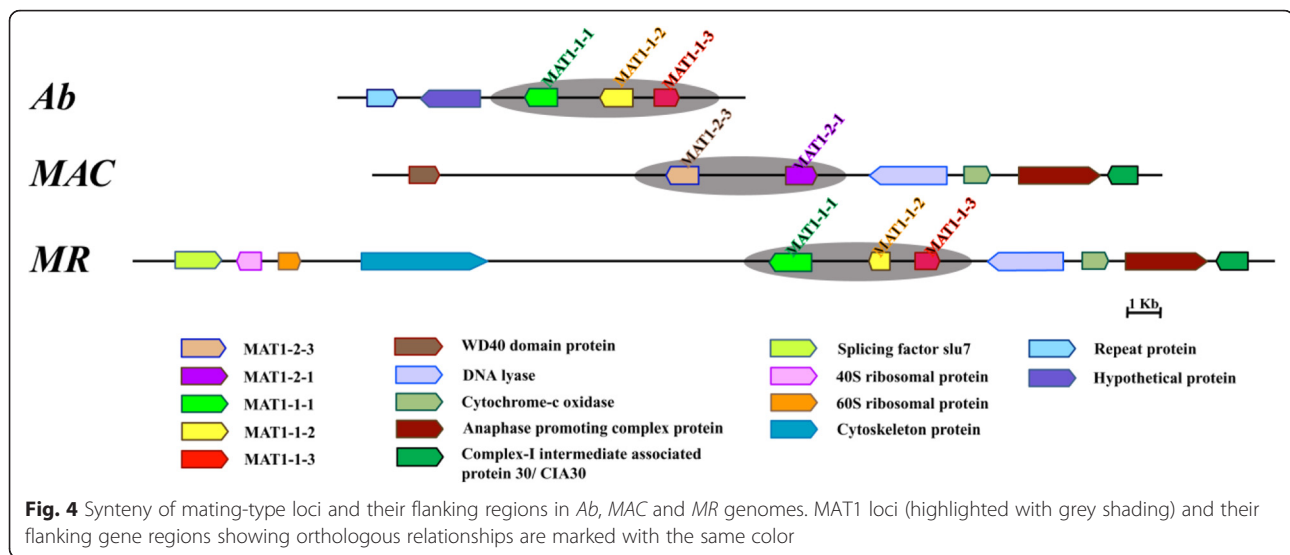
Repeat-induced point mutation (RIP) is a protection system in fungal genomes that operates on duplicated sequences and checks the development of TEs through hypermutation [71, 72]. A genome-wide RIP analysis was performed on the *Ab*, *MAC* and *MR* genomes. Summary of RIP signatures predicted in the three genomes is provided in Additional file 12. RIP frequency index (CpA + TpG)/(ApC + GpT) was calculated as 0.72 in *Ab*, 0.94 in *MAC* and 1.06 in *MR* (Additional file 13), where the threshold was 1.03 and the RIP index \leq 1.03 indicated RIP activity. Therefore, *Ab* and *MAC* genomes were suggested to present significant RIP activity consistent with earlier reports supporting occurrence of RIPs in specialists genome whereas their absence from non-specialists genomes [13, 60]. The frequency of putative transposase genes suggests their expansion in *MR* genome as compared to the *Ab* and *MAC* genomes (Additional file 14). This observed difference is suggested to be related to the effect of RIP activity [13, 60].

Mating type

The three genomes show a variety of common fungal mating genes (Additional file 15). The mating type genes MAT1-1-1, MAT1-1-2 and MAT1-1-3 were identified in *Ab* and *MR*. Interestingly, syntenic analysis revealed the presence of mating type genes MAT1-2-3 and MAT1-2-1 in place of MAT1-1-1, MAT1-1-2 and MAT1-1-3 in *MAC* and the conservation of genes flanking the MAT1-2-3 was initially reported in the genus *Fusarium* (Teleomorph: *Gibberella*) (*Nectriaceae*, *Hypocreales*) [73] that was recognized based on its expression profile, displaying upregulation during the sexual developmental process similar to other MAT loci. However, no known functional domain was identified in this MAT1-2-3 locus. Therefore, in *MAC* genome, this putative MAT1-2-3 locus was identified based on sequence similarity with *Fusarium* MAT1-2-3 locus. *Ab* is already known to be sexually fertile based on the detection of its teleomorph *Hypocrella siamensis* [74].

Phylogenetic classification of putative cuticle degrading chitinases

Fungal chitinases are known to belong exclusively to the GH18 family [75–77]. Based on amino acid sequence similarity of their GH18 domains [49, 77], these have been classified into three subgroups A, B and C. These



display differences in their enzymatic properties—exo for sg-A and sg-C, whereas endo for sg-B; the presence or absence of carbohydrate binding modules (CBMs), absent in sg-A, whereas present in sg-B at C-terminal and in sg-C at N-terminal—and the architecture of their substrate binding cleft [75–77]. In this study, we identified 16, 23 and 28 chitinase genes in *Ab*, *MAC* and *MR* genomes, respectively. Their properties are presented in Additional file 16. These chitinases were nomenclatured based on their increasing molecular weights and are represented as *X Chit-z*, where *X* is the organism label (*Ab*, *MAC* or *MR*) and *z* is the designated number. A few of the *MAC* and *MR* chitinases were named on the basis of sequence similarity between *MAC* and *MR* chitinases. The three datasets, *viz.*, subgroup A, B and C were analyzed to reconstruct their phylogenetic relationships.

Subgroup-A (sg-A) dataset

This dataset contained 85 chitinase sequences, including 20 sequences retrieved from *Ab*, *MAC* and *MR* genomes (Fig. 5a). *Ab* chitinases were only observed in clades A-V and A-II. However, *MAC* and *MR* chitinases were observed in all sg-A clades: A-V, A-IV, A-III and A-II. The majority of chitinase sequences from the three genomes grouped with orthologs from *Trichoderma reesei* (*Hypocrea jecorina*) [Ascomycota, Hypocreales, Hypocreaceae] and other plant pathogenic fungi.

Subgroup-B (sg-B) dataset

This dataset included 68 chitinase sequences, including 29 sequences retrieved from *Ab*, *MAC* and *MR* genomes (Fig. 5b). All chitinases were clustered with *H. jecorina* chitinases. Interestingly, *MAC* was found to contain more chitinases than *MR*. However, *Ab* contained the least number of sg-B chitinases. Further, *Ab Chit-1*,

MAC Chit-3, *MAC Chit-7* and *MR Chit-7* were observed to cluster with *H. jecorina Chit-15*, forming a separate clade in sg-B.

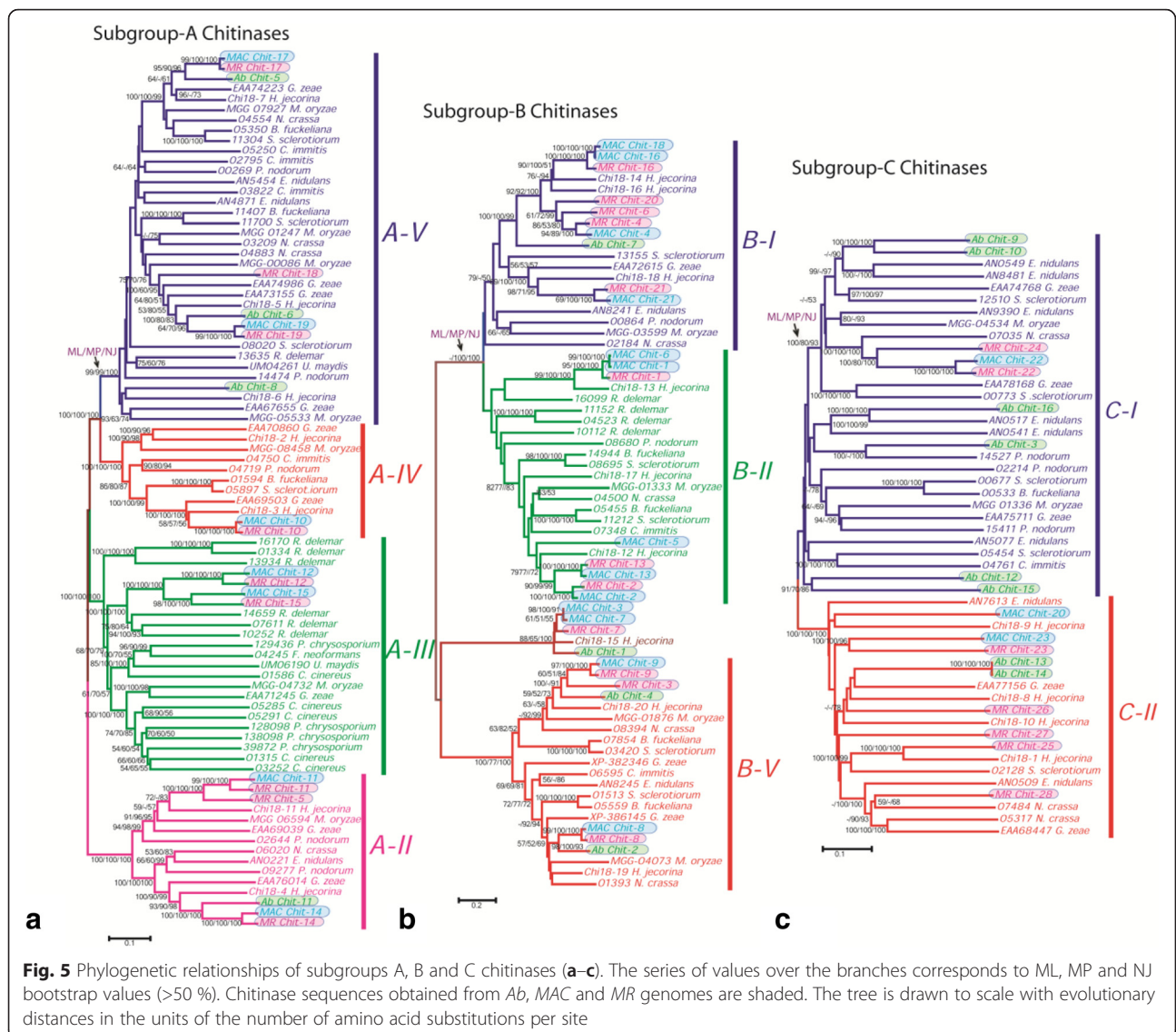
Subgroup-C (Sg-C) dataset

This dataset included 50 chitinase sequences, including 18 sequences retrieved from *Ab*, *MAC* and *MR* genomes (Fig. 5c). The number of sg-C chitinases was higher in the *Ab* genome as compared to the *MR* and *MAC* genomes.

Domain architecture of chitinases

The domain architecture for all the chitinases identified by sequence similarity from the three genomes is presented in Fig. 6. It reveals interesting details about domain distribution in chitinases present in the three genomes. Intriguingly, an additional domain glycerophosphoryldiester phosphodiesterase (GPDP) in sg-A classified chitinase *Ab Chit-11* was observed in *Ab* genome. This enzyme GPDP is known for glycerophospholipid metabolism and is indicated to be associated with extracellular events like cell wall organization [78]. Furthermore, the distribution of CBM50 (or LysM domain) known for glycan (chitin) binding functions and predicted as virulence factors [79] was quite variable in the three genomes. Only *Ab* and *MR* genomes displayed the presence of these domains whereas no CBM50 domain was identified in the *MAC* genome. However, higher numbers of CBM50 domains in comparison to *Ab* were observed in *MR* genome. Further, about 82 % (23/28) of chitinases have a signal peptide (SP) signature sequence in *MR* while only 50 % in the other two species contain the SP (Fig. 6).

The sg-B chitinases are known to possess CBMs (CBM1) as their C-terminal domains [76]. Our analysis



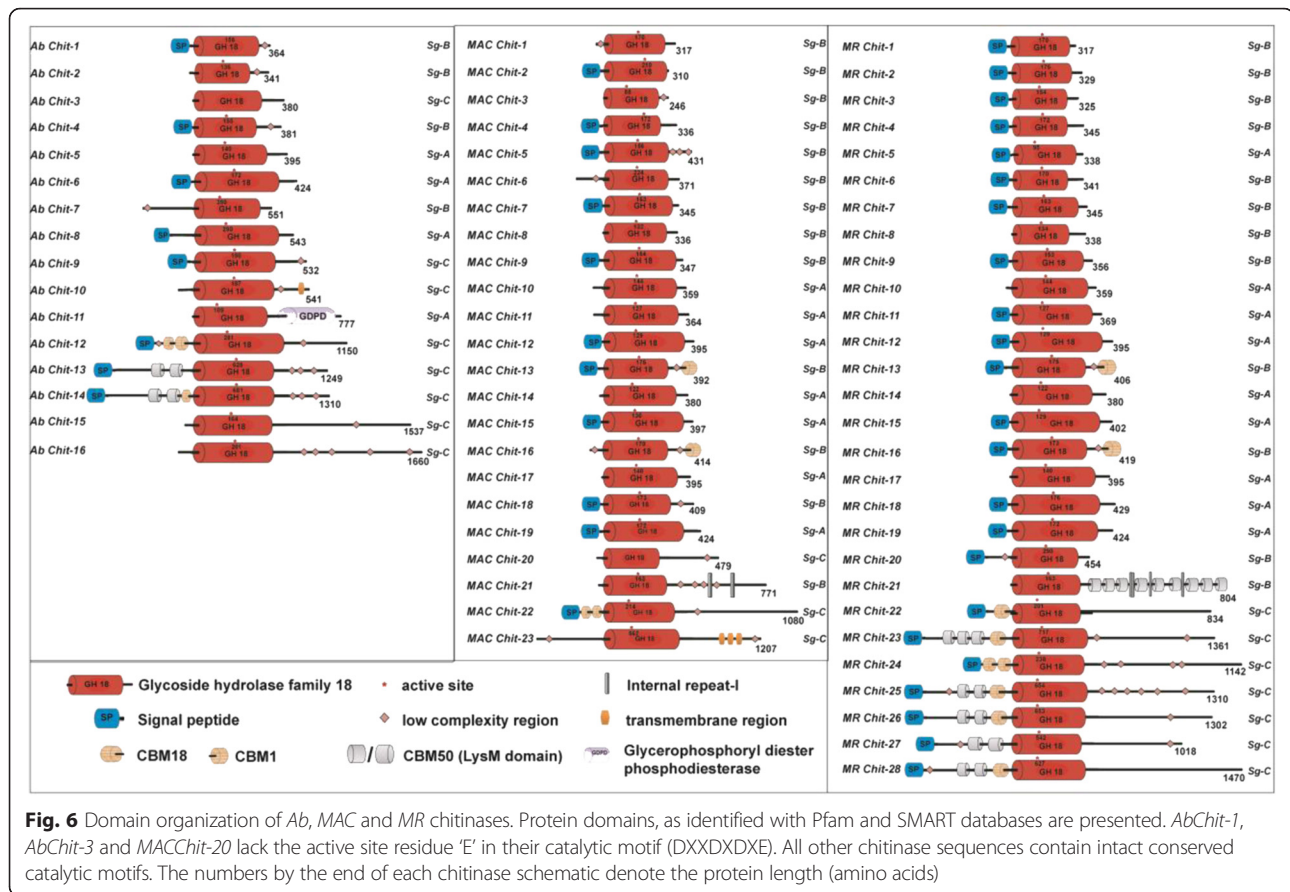
shows CBM1 is present in the *MAC* and *MR* genomes but absent in the *Ab* genome. Multiple (nine) copies of C-terminal CBM50 domains were observed in *MR* genome.

The sg-C chitinases too displayed some interesting differences. Most of sg-C chitinases contained the N-terminal CBM18 and CBM50 domains in *Ab* and *MAC* genomes. However, *MR* genomes showed regular characteristics of sg-C chitinases, i.e., high molecular weight and presence of multiple domains.

Discussion

It is often suggested that diversity in pathogen-host relationships contributes to species divergence but the knowledge of genetic mechanisms responsible for speciation and varied host interactions in fungal entomopathogens have been quite limited [80, 81]. The entomopathogenic

fungal lineages possess a number of potential pathogenicity associated genes in their genome to acclimatize to varied niches [60, 82–85]. The evolutionary dynamics of host affiliation observed in Clavicipitaceae have been aptly explained by the host-habitat hypothesis where the proximity in the habitat has been related to the acquisition of new hosts [86]. *Metacordyceps* subfamily in Clavicipitaceae lineage is of interest as it contains pathogenic species varying from specialists (with a narrow host range), transitional (with intermediate host range) to generalists (with broad host range) categories [59, 60]. Comparative analyses of their genomes has already been reported that suggests the association of generalists with loss of RIPs (genome defense) and in turn expansion of gene/protein families, whereas the retention of RIPs in specialists fungi [13, 60]. In the current study, comparative genomics of *Ab* belonging to the *Hypocrella* subfamily with *Metacordyceps*



members *MAC* and *MR* further aids in understanding the genomic basis for host niche adaptations in the Clavicipitaceae lineage. *Ab*, being a specialist fungus shows an evolutionary pattern similar to the specialist genome *MAC*, with a similar occurrence of RIPs and in the evolution of protein families (contraction). This study also supports that during their evolution generalist fungi have lost the active RIP mechanism which may have aided broader host affiliation by the retention of duplicated genes and in turn protein family expansions. Generalist' pathogens are suggested to have evolve higher numbers of multi-modular NRPS genes (such as *dtxS1*), siderophores, CBMs and chitinases as contributory factors in acquiring new virulence mechanisms to enhance their pathogenicity potential [13, 60].

Multimodular NRPS, *dtxS1* as a virulence factor

The hexamodular NRPS gene responsible for the production of destruxins, identified as *dtxS1* gene in *MR* genome was reported to be involved in the pathogenicity of generalist fungus *MR* [59]. It was postulated that acquisition of *dtxS1* gene in *Metarhizium* lineages was in coordination with the evolution of host specificity [59]. Our genome analysis suggests that *Ab* genome does not contain the NRPS *dtxS1* multi-domain gene as

reported for specialist pathogens [58–60]. The NRPS subfamilies are known to cluster into two main groups: mono/bimodular NRPS (of ancient origin, bacterial and fungal NRPS) and multimodular NRPS (recent origin, exclusively fungal NRPS) [38]. The multimodular NRPS subfamilies were identified to contain variable domain architecture in response to the niche adaptation roles, illustrating frequent incidences of domain gain and loss events, unlike mono/bimodular NRPS that are less variable in architecture [38, 55]. The destruxin biosynthesis NRPS gene belongs to the multimodular NRPS cluster, is likely to have originated via a gene/domain duplication event in *MR* to form NRPS *dtxS1* gene. Gain of multimodular NRPS gene (such as *dtxS1*) could therefore be suggested as responsible for the generalist behavior and enhanced virulence of *MR* [59, 60].

Siderophores as virulence factors

Iron is an essential micronutrient for life and the ability of pathogens to cope with limited iron within the host organism represents an evolutionary force for the development of virulence [54, 87]. Siderophores are high-affinity small molecules that function in iron uptake [88]. Pathogenic fungi have evolved the production of these siderophores for acquisition of iron from the host organism they infect,

to ensure their own survival [54, 87, 89]. Maintenance of iron homeostasis therefore suggests association of fungal pathogenicity and siderophore biosynthesis [52–54]. There are reports from pathogenic fungi such as *Alternaria alternata* [90], *Aspergillus fumigatus* [53, 54, 91], *Cryptococcus neoformans* [92], *Fusarium graminearum* [93], *Magnaporthe oryzae* [94], *Metarhizium robertsii* [95] that correlate iron acquisition and virulence. For example, mutagenesis studies on murine model of invasive aspergillosis caused by *Aspergillus fumigatus* displayed complete lack of virulence on elimination of the entire siderophore biosynthetic gene (Δ *sidA* mutant) demonstrating critical role of siderophores in pathogenicity and thereby in fungal-host interactions [91, 96, 97]. Likewise, another mutagenesis study of two NRPS enzymes *sidD* (intracellular siderophore) and *sidC* (extracellular siderophore) investigated the involvement of siderophores in pathogenesis in *MR* [95]. Based on these studies, it could therefore be proposed that fungal entomopathogens have evolved the genes involved in siderophore biosynthesis to raise their virulence potential.

Chitinases as virulence factors

Chitinases have been suggested as the determining factor of virulence in fungal entomopathogens with their involvement in degradation of arthropod cuticle [50, 98]. In the present study, we observe a higher proportion of chitinases in the genomes of the insect pathogens (*Ab*, *MAC* and *MR*) compared to other non-insect pathogens like *Aspergillus* spp. and *Rhizopus* spp. [13, 99]. Sg-B and sg-C chitinases are commonly known to be involved in virulent and aggressive functions towards insects in entomopathogenic fungi [100–102] likely due to the presence of additional domains like CBM1, CBM18 and CBM50 responsible for chitin binding functions [77]. Interestingly, only the *MR* genome is observed to contain chitinases (*MR Chit-9* classified under sg-B clade) with multiple CBM50 domains at the C-terminal. Similar observation could also be viewed in sg-C classified *MR* chitinases with the presence of multiple copies of N-terminal CBM18 and CBM50 domains. The presence of comparatively expanded numbers of accessory domains, CBMs in *MR* genome may relate to evolution of multiple host adaptations [103, 104]. Moreover, the complete lack of LysM effectors (CBM50 domain) in *MAC* genome may also be associated with its pathogenicity potential and in turn host-specific behavior [79, 105]. This could also be supported with overall distribution of CBM1, CBM18 and CBM50 domains in *Ab*, *MAC* and *MR* genomes.

Mating type

Mating type genes are the governing players of sexual functions where the presence or absence of two compatible mating type idiomorphs, MAT1-1 and MAT1-2 decides

homothallism (self-fertility) or heterothallism (self-sterility but cross-fertility) in filamentous ascomycetes [41, 106]. Clavicipitaceous fungi too show diversity in their mode of mating [86, 107, 108]. The present analysis based on sequence similarity searches identified putative mating type idiomorphs, MAT1-1 (MAT1-1-1, MAT1-1-2 and MAT1-1-3) in *Ab* suggesting its potential for sexual fertility, which could also be supported by its known teleomorphic form *Hypocrella siamensis* [74]. Mating type genes, MAT1-2-3 and MAT1-2-1 in place of MAT1-1-1, MAT1-1-2 and MAT1-1-3 were identified based on synteny analysis in *MAC* and the conservation of genes flanking the MAT1-1 idiomorph in *MAC* and *MR* (Fig. 4).

Conclusions

The present study is an attempt to understand the evolution of entomopathogenicity in Clavicipitaceae s.s. members in relation to their adaptations to host-range. In this study, we suggest the involvement of multimodular NRPS genes, siderophores, CBMs and chitinases in generalist clavicipitaceous fungal entomopathogen *MR* that could have evolved for diverse host affiliations, as compared to specialist entomopathogens *Ab* and *MAC*.

Availability of supporting data

Additional files 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16.

Additional files

Additional file 1: Multiple sequence alignment of NRPS genes. (ZIP 225 kb)

Additional file 2: Multiple sequence alignment of subgroup A,B and C fungal chitinases. (DOC 593 kb)

Additional file 3: Phylogeny and comparative genomics of clavicipitaceous fungi *Ab*, *MAC*, and *MR*. (PDF 1319 kb)

Additional file 4: List of accession numbers for predicted NRPS gene cluster sequences identified in *Ab*, *MAC* and *MR* genomes. (XLS 10 kb)

Additional file 5: Major protein families of *Ab*, *MAC* and *MR* genomes implicated in the pathogen-host interactions. (XLS 20 kb)

Additional file 6: Analysis of carbohydrate-binding module (CBM) containing proteins, arranged by CBM family. (XLS 14 kb)

Additional file 7: Carbohydrate-degrading enzymes, arranged by GH family. (XLS 67 kb)

Additional file 8: Carbohydrate-degrading enzymes, arranged by GT family. (XLS 23 kb)

Additional file 9: Carbohydrate-degrading enzymes, arranged by PL family. (XLS 7 kb)

Additional file 10: Carbohydrate-degrading enzymes, arranged by CE family. (XLS 15 kb)

Additional file 11: Carbohydrate-degrading enzymes, arranged by AA family. (XLS 15 kb)

Additional file 12: Summary of RIP signatures in *Ab*, *MAC* and *MR* genomes. (XLS 46 kb)

Additional file 13: RIP frequency index in *Ab*, *MAC* and *MR* genomes. (XLS 8 kb)

Additional file 14: Putative transposase genes encoded by *Ab*, *MAC* and *MR*. (XLS 10 kb)

Additional file 15: Sexuality related genes in *Ab*, *MAC* and *MR* genomes. (XLS 22 kb)

Additional file 16: Properties of chitinases retrieved from the genomes of *Ab*, *MAC* and *MR*. The active site position, number of amino acid, molecular weight (MW), presence of signal peptide and domains (CBM, LysM and others in addition to the catalytic domain) have been analyzed for the respective genes in the genome and listed. (XLS 21 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SKS: conceived the idea and designed the research. TN: genome assembly, annotation and data analysis. YA: data analysis, phylogeny and manuscript writing. All authors: manuscript review. All authors read and approved the final manuscript.

Acknowledgements

Dr. Belle Damodara Shenoy is acknowledged for critical inputs to an earlier draft of the manuscript. This work was financially supported by the project "Expansion and modernization of Microbial Type Culture Collection and Gene Bank (MTCC)" jointly supported by Council of Scientific and Industrial Research (CSIR) Grant No. BSC0402 and Department of Biotechnology (DBT) Govt. of India Grant No. BT/PR7368/INF/22/177/2012. *Ab* isolate for genome sequencing was obtained from MTCC, Chandigarh, India. NGS data was obtained from C-CAMP next-generation genomics facility, Bangalore, India. YA was supported by senior research fellowship from CSIR [31/050 (0315)/2010-EMR-I] and TN was supported from junior research fellowship from the CSIR XII five-year plan network project Genomics and informatics Solutions for Integrating Biology (GENESIS).

Received: 23 December 2015 Accepted: 5 May 2016

Published online: 17 May 2016

References

- Sung GH, Hywel-Jones NL, Sung JM, Luangsa-Ard JJ, Shrestha B, Spatafora JW. Phylogenetic classification of *Cordyceps* and the clavicipitaceae fungi. *Stud Mycol*. 2007;57:5–59. doi:10.3114/sim.2007.57.01.
- Berger EW. Natural enemies of scale insects and whiteflies in Florida. *Florida State Plant Board Quarterly Bulletin*. 1921;5:141–54.
- Borner C. Aleyrodidae (Aleurodina). In: Sorauer CP, editor. *Handbuch der Pflanzenkrankheiten, Tierische Schädlinge an Nutzpflanzen, Teil 2, Lieferung 4. Heteroptera und Homoptera*. Berlin: Paul Parey; 1956. p. 331–59.
- Evans HC, Hywel-Jones N. Aspects of the genera *Hypocrella* and *Aschersonia* as pathogens of coccids and whiteflies. *Vth International Colloquium on Invertebrate Pathology and microbial Control*. Adelaide: Society for Invertebrate Pathology; 1990. p. 111–5.
- McCoy CW, Samson RA, Boucias DG. Entomogenous fungi. In: Ignoffo CM, Mandava NB, editors. *CRC Handbook of Natural Pesticides*. Boca Raton: CRC Press; 1988.
- Fawcett HS. *Citrus diseases and their control*. 2nd ed. New York: McGraw-Hill; 1936.
- Fransen JJ. Control of greenhouse whitefly, *Trialeurodes vaporariorum*, by the fungus *Aschersonia aleyrodis*. *Bulletin SROP*. 1987;10:57–61.
- Meeke ETM, Van VS, Joosten NN, Franssen JJ, Lenteren JC. Persistence of the fungal whitefly pathogen, *Aschersonia aleyrodis*, on three different plant species. *Mycol Res*. 2000;104:1234–40.
- Fransen JJ, Van-Lenteren JC. Host selection and survival of the parasitoid *Encarsia formosa* on greenhouse whitefly, *Trialeurodes vaporariorum*, in the presence of hosts infected with the fungus *Aschersonia aleyrodis*. *Entomol Exp Appl*. 1993;69:239–49.
- Ramakers PMJ, Samson RA. *Aschersonia aleyrodis*, a fungal pathogen of whitefly. II. Application as a biological insecticide in glasshouses. *Zeitschrift für angewandte Entomologie*. 1984;97:1–8.
- Strasser H, Hutwimmer S, Burgstaller W. Metabolite Toxicology of Fungal Biocontrol Agents. In: Ehlers R-U, editor. *Regulation of Biological Control Agents*. Netherlands: Springer; 2011. p. 191–213.
- Genthner FJ, Chancy CA, Couch JA, Foss SS, Middaugh DP, George SE, et al. Toxicity and Pathogenicity Testing of the Insect Pest Control Fungus *Metarhizium anisopliae*. *Arch Environ Contam Toxicol*. 1998;35(2):317–24. doi:10.1007/s002449900382.
- Gao Q, Jin K, Ying SH, Zhang Y, Xiao G, Shang Y, et al. Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet*. 2011;7(1):e1001264. doi:10.1371/journal.pgen.1001264.
- Thomas MB, Read AF. Can fungal biopesticides control malaria? *Nat Rev Microbiol*. 2007;5(5):377–83. doi:10.1038/nrmicro1638.
- Charnley AK. Fungal pathogens of insects: cuticle degrading enzymes and toxins. *Adv Bot Res*. 2003;40:241–321.
- Niassy S, Subramanian S, Ekese S, Bargul JL, Villinger J, Maniania NK. Use of *Metarhizium anisopliae* chitinase genes for genotyping and virulence characterization. *Biomed Res Int*. 2013;2013:465213. doi:10.1155/2013/465213.
- Bhanu Prakash GV, Padmaja V, Jami SK, Kirti PB. Expression of chitinase genes of *Metarhizium anisopliae* isolates in lepidopteran pests and on synthetic media. *J Basic Microbiol*. 2012;52(6):628–35. doi:10.1002/jobm.201100274.
- Patel RK, Jain M. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS One*. 2012;7(2):e30619.
- Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 28 Apr 2016.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77. doi:10.1089/cmb.2012.0021.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2011;27(4):578–9. doi:10.1093/bioinformatics/btq683.
- Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 2007;23(9):1061–7. doi:10.1093/bioinformatics/btm071.
- Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. 2008;18(1):188–96. doi:10.1101/gr.6743907.
- Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 2005; 33(Web Server issue):W465–7. doi:10.1093/nar/gki458.
- Stamatakis A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*. 2014. doi:10.1093/bioinformatics/btu033.
- Swofford DL. PAUP*: Phylogenetic analysis using parsimony (* and other methods). Version 4.0. Sunderland: Sinauer Associates; 2003.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9. doi:10.1093/molbev/mst197.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25(15):1972–3. doi:10.1093/bioinformatics/btp348.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000;28(1):33–6.
- Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*. 2011;12: 124. doi:10.1186/1471-2105-12-124.
- Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol*. 2013;30(8):1987–97. doi:10.1093/molbev/mst100.
- Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*. 2011;39(Web Server issue):W339–46. doi:10.1093/nar/gkr466.
- Winnenburg R, Baldwin TK, Urban M, Rawlings C, Kohler J, Hammond-Kosack KE. PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res*. 2006;34(Database issue):D459–64. doi:10.1093/nar/gkj047.
- Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res*. 2012;40(Database issue):D302–5. doi:10.1093/nar/gkr931.

35. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res.* 2012;40(Database issue):D290–301. doi:10.1093/nar/gkr1065.
36. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A.* 1998;95(11):5857–64.
37. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: John MW, editor. *The Proteomics Protocols Handbook*. Humana Press; 2005. p. 571–607. doi:10.1385/1-59259-890-0:571.
38. Bushley KE, Turgeon BG. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol Biol.* 2010;10:26. doi:10.1186/1471-2148-10-26.
39. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2012;40(Web Server issue):W445–51. doi:10.1093/nar/gks479.
40. Hane JK, Oliver RP. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics.* 2008;9:478. doi:10.1186/1471-2105-9-478.
41. Lee SC, Ni M, Li W, Shertz C, Heitman J. The evolution of sex: a perspective from the fungal kingdom. *Microbiol Mol Biol Rev.* 2010;74(2):298–340. doi:10.1128/MMBR.00005-10.
42. Bushley KE, Li Y, Wang WJ, Wang XL, Jiao L, Spatafora JW, et al. Isolation of the MAT1-1 mating type idiomorph and evidence for selfing in the Chinese medicinal fungus *Ophiocordyceps sinensis*. *Fungal Biol.* 2013;117(9):599–610. doi:10.1016/j.funbio.2013.06.001.
43. Klux V, Nowrousian M, Ringelberg C, Loros JJ, Dunlap JC, Pöggeler S. Functional Characterization of MAT1-1-Specific Mating-Type Genes in the Homothallic Ascomycete *Sordaria macrospora* Provides New Insights into Essential and Nonessential Sexual Regulators. *Eukaryot Cell.* 2010;9(6):894–905. doi:10.1128/ec.00019-10.
44. Jablonowski D, Fichtner L, Martin VJ, Klassen R, Meinhardt F, Stark MJR, et al. *Saccharomyces cerevisiae* cell wall chitin, the *Kluyveromyces lactis* zymocin receptor. *Yeast.* 2001;18(14):1285–99. doi:10.1002/yea.776.
45. Pei J, Sadreyev R, Grishin NV. PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics.* 2003;19(3):427–8. doi:10.1093/bioinformatics/btg008.
46. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39 suppl 2:W29–37. doi:10.1093/nar/gkr367.
47. Karlsson M, Stenlid J. Evolution of family 18 glycoside hydrolases: diversity, domain structures and phylogenetic relationships. *J Mol Microbiol Biotechnol.* 2009;16(3-4):208–23. doi:10.1159/000151220.
48. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol.* 2013;30(4):772–80. doi:10.1093/molbev/mst010.
49. Seidl V, Huemer B, Seiboth B, Kubicek CP. A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS J.* 2005;272(22):5923–39. doi:10.1111/j.1742-4658.2005.04994.x.
50. Agrawal Y, Khatri I, Subramanian S, Shenoy BD. Genome sequence, comparative analysis and evolutionary insights into chitinases of entomopathogenic fungus *Hirsutiella thompsonii*. *Genome Biol Evol.* 2015;7(3):916–30. doi:10.1093/gbe/evv037.
51. Molnar I, Gibson DM, Krasnoff SB. Secondary metabolites from entomopathogenic Hypocrealean fungi. *Nat Prod Rep.* 2010;27(9):1241–75. doi:10.1039/c001459c.
52. Miethke M, Marahiel MA. Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev.* 2007;71(3):413–51. doi:10.1128/MMBR.00012-07.
53. Schrettel M, Haas H. Iron homeostasis—Achilles' heel of *Aspergillus fumigatus*? *Curr Opin Microbiol.* 2011;14(4):400–5. doi:10.1016/j.mib.2011.06.002.
54. Haas H. Fungal siderophore metabolism with a focus on *Aspergillus fumigatus*. *Nat Prod Rep.* 2014;31(10):1266–76. doi:10.1039/c4np00071d.
55. Amoutzias GD, Van de Peer Y, Mossialos D. Evolution and taxonomic distribution of nonribosomal peptide and polyketide synthases. *Future Microbiol.* 2008;3(3):361–70. doi:10.2217/17460913.3.3.361.
56. Finking R, Marahiel MA. Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol.* 2004;58:453–88. doi:10.1146/annurev.micro.58.030603.123615.
57. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol.* 2007;7:78. doi:10.1186/1471-2148-7-78.
58. Giuliano GDB, Krasnoff SB, Moon YS, Churchill AC, Gibson DM. Genetic basis of destruxin production in the entomopathogen *Metarhizium robertsii*. *Curr Genet.* 2012;58(2):105–16. doi:10.1007/s00294-012-0368-4.
59. Wang B, Kang Q, Lu Y, Bai L, Wang C. Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species. *Proc Natl Acad Sci U S A.* 2012;109(4):1287–92. doi:10.1073/pnas.1115983109.
60. Hu X, Xiao G, Zheng P, Shang Y, Su Y, Zhang X, et al. Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. *Proc Natl Acad Sci U S A.* 2014;111(47):16796–801. doi:10.1073/pnas.1412662111.
61. Oide S, Moeder W, Krasnoff S, Gibson D, Haas H, Yoshioka K, et al. NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes. *Plant Cell.* 2006;18(10):2836–53. doi:10.1105/tpc.106.045633.
62. St. Leger RJ, Staples RC, Roberts DW. Cloning and regulatory analysis of starvation-stress gene, *ssgA*, encoding a hydrophobin-like protein from the entomopathogenic fungus, *Metarhizium anisopliae*. *Gene.* 1992;120(1):119–24.
63. Wang C, St. Leger RJ. 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.* 2007;6(5):808–16. doi:10.1128/EC.00409-06.
64. Zhang Y, Zhang J, Jiang X, Wang G, Luo Z, Fan Y, et al. Requirement of a mitogen-activated protein kinase for appressorium formation and penetration of insect cuticle by the entomopathogenic fungus *Beauveria bassiana*. *Appl Environ Microbiol.* 2010;76(7):2262–70. doi:10.1128/AEM.02246-09.
65. Freimoser FM, Hu G, St Leger RJ. Variation in gene expression patterns as the insect pathogen *Metarhizium anisopliae* adapts to different host cuticles or nutrient deprivation in vitro. *Microbiology.* 2005;15(2):361–71. doi:10.1099/mic.0.27560-0.
66. Donatti AC, Furlaneto-Maia L, Fungaro MH, Furlaneto MC. Production and regulation of cuticle-degrading proteases from *Beauveria bassiana* in the presence of *Rhammatocerus schistocercoides* cuticle. *Curr Microbiol.* 2008;56(3):256–60. doi:10.1007/s00284-007-9071-y.
67. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical J.* 1991;280(2):309–16.
68. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* 2008;77:521–55. doi:10.1146/annurev.biochem.76.061005.092322.
69. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical J.* 2010;432(3):437–44. doi:10.1042/BJ20101185.
70. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZY database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels.* 2013;6(1):41. doi:10.1186/1754-6834-6-41.
71. Hood ME, Katawczik M, Giraud T. Repeat-Induced Point Mutation and the Population Structure of Transposable Elements in *Botrytis violaceum*. *Genetics.* 2005;170(3):1081–9. doi:10.1534/genetics.105.042564.
72. Horns F, Petit E, Yockteng R, Hood ME. Patterns of Repeat-Induced Point Mutation in Transposable Elements of Basidiomycete Fungi. *Genome Biol Evol.* 2012;4(3):240–7. doi:10.1093/gbe/evs005.
73. Martin SH, Wingfield BD, Wingfield MJ, Steenkamp ET. Structure and evolution of the *Fusarium* mating type locus: new insights from the *Gibberella fujikuroi* complex. *Fungal Genet Biol.* 2011;48(7):731–40. doi:10.1016/j.fgb.2011.03.005.
74. Mongkolsamrit S, Luangsa-ard JJ, Spatafora JW, Sung G-H, Hywel-Jones NL. A combined ITS rDNA and β -tubulin phylogeny of Thai species of *Hypocrella* with non-fragmenting ascospores. *Mycol Res.* 2009;113(6-7):684–99. http://dx.doi.org/10.1016/j.mycres.2009.02.004.
75. Adrangi S, Faramarzi MA. From bacteria to human: a journey into the world of chitinases. *Biotechnol Adv.* 2013;31(8):1786–95. doi:10.1016/j.biotechadv.2013.09.012.
76. Hartl L, Zach S, Seidl-Seiboth V. Fungal chitinases: diversity, mechanistic properties and biotechnological potential. *Appl Microbiol Biotechnol.* 2012;93(2):533–43. doi:10.1007/s00253-011-3723-3.
77. Seidl V. Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions. *Fungal Biol Rev.* 2008;22(1):36–42. http://dx.doi.org/10.1016/j.fbr.2008.03.002.
78. Hayashi S, Ishii T, Matsunaga T, Tominaga R, Kuromori T, Wada T, et al. The glycerophosphoryl diester phosphodiesterase-like proteins SHV3 and its homologs play important roles in cell wall organization. *Plant Cell Physiol.* 2008;49(10):1522–35. doi:10.1093/pccp/pcn120.
79. Kombrink A, Thomma BP. LysM effectors: secreted proteins supporting fungal life. *PLoS Pathog.* 2013;9(12):e1003769. doi:10.1371/journal.ppat.1003769.

80. Stukenbrock EH. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist*. 2013;199(4):895–907. doi:10.1111/nph.12374.
81. Giraud T, Refregier G, Le Gac M, de Vienne DM, Hood ME. Speciation in fungi. *Fungal Genet Biol*. 2008;45(6):791–802. doi:10.1016/j.fgb.2008.02.001.
82. Duan Z, Shang Y, Gao Q, Zheng P, Wang C. A phosphoketolase Mpk1 of bacterial origin is adaptively required for full virulence in the insect-pathogenic fungus *Metarhizium anisopliae*. *Environ Microbiol*. 2009;11(9):2351–60. doi:10.1111/j.1462-2920.2009.01961.x.
83. Gardiner DM, Kazan K, Manners JM. Cross-kingdom gene transfer facilitates the evolution of virulence in fungal pathogens. *Plant Sci*. 2013;210:151–8. doi:10.1016/j.plantsci.2013.06.002.
84. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BP, et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog*. 2011;7(7):e1002137. doi:10.1371/journal.ppat.1002137.
85. Zhao H, Xu C, Lu HL, Chen X, St. Leger RJ, Fang W. Host-to-pathogen gene transfer facilitated infection of insects by a pathogenic fungus. *PLoS Pathog*. 2014;10(4):e1004009. doi:10.1371/journal.ppat.1004009.
86. Kepler RM, Sung GH, Harada Y, Tanaka K, Tanaka E, Hosoya T, et al. Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and Ustilaginoidae (Clavicipitaceae). *Am J Bot*. 2012;99(3):552–61. doi:10.3732/ajb.1100124.
87. Johnson L. Iron and siderophores in fungal-host interactions. *Mycol Res*. 2008;112(2):170–83. <http://dx.doi.org/10.1016/j.mycres.2007.11.012>.
88. Budzikiewicz H. Siderophores from Bacteria and from Fungi. Iron Uptake and Homeostasis in Microorganisms. Norfolk: Caister Academic Press; 2010.
89. Scharf DH, Heinekamp T, Brakhage AA. Human and Plant Fungal Pathogens: The Role of Secondary Metabolites. *PLoS Pathog*. 2014;10(1):e1003859.
90. Chen L-H, Yang SL, Chung K-R. Resistance to oxidative stress via regulating siderophore-mediated iron acquisition by the citrus fungal pathogen *Alternaria alternata*. *Microbiology*. 2014;160(5):970–9. doi:10.1099/mic.0.076182-0.
91. Schrettl M, Bignell E, Kragl C, Sabiha Y, Loss O, Eisendle M, et al. Distinct Roles for Intra- and Extracellular Siderophores during *Aspergillus fumigatus* Infection. *PLoS Pathog*. 2007;3(9):e128.
92. Saikia S, Oliveira D, Hu G, Kronstad J. Role of ferric reductases in iron acquisition and virulence in the fungal pathogen *Cryptococcus neoformans*. *Infect Immun*. 2014;82(2):839–50. doi:10.1128/IAI.01357-13.
93. Greenshields DL, Liu G, Feng J, Selvaraj G, Wei Y. The siderophore biosynthetic gene SID1, but not the ferroxidase gene FET3, is required for full *Fusarium graminearum* virulence. *Mol Plant Pathol*. 2007;8(4):411–21. doi:10.1111/j.1364-3703.2007.00401.x.
94. Hof C, Eisfeld K, Welzel K, Antelo L, Foster AJ, Anke H. Ferricrocin synthesis in *Magnaporthe grisea* and its role in pathogenicity in rice. *Mol Plant Pathol*. 2007;8(2):163–72. doi:10.1111/j.1364-3703.2007.00380.x.
95. Giuliano GDB, Gibson DM, Krasnoff SB. Intracellular siderophore but not extracellular siderophore is required for full virulence in *Metarhizium robertsii*. *Fungal Genet Biol*. 2015;82:56–68. doi:10.1016/j.fgb.2015.06.008.
96. Hissen AHT, Wan ANC, Warwas ML, Pinto LJ, Moore MM. The *Aspergillus fumigatus* Siderophore Biosynthetic Gene *sidA*, Encoding L-Ornithine N(5)-Oxygenase, Is Required for Virulence. *Infect Immun*. 2005;73(9):5493–503. doi:10.1128/iai.73.9.5493-5503.2005.
97. Schrettl M, Bignell E, Kragl C, Joechl C, Rogers T, Arst HN, et al. Siderophore Biosynthesis But Not Reductive Iron Assimilation Is Essential for *Aspergillus fumigatus* Virulence. *J Exp Med*. 2004;200(9):1213–9. doi:10.1084/jem.20041242.
98. St. Leger RJ, Wang C. Genetic engineering of fungal biocontrol agents to achieve greater efficacy against insect pests. *Appl Microbiol Biotechnol*. 2010;85(4):901–7. doi:10.1007/s00253-009-2306-z.
99. Xiao G, Ying SH, Zheng P, Wang ZL, Zhang S, Xie XQ, et al. Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci Rep*. 2012;2:483. doi:10.1038/srep00483.
100. Baratto CM, Dutra V, Boldo JT, Leiria LB, Vainstein MH, Schrank A. Isolation, characterization, and transcriptional analysis of the chitinase *chi2* Gene (DQ011663) from the biocontrol fungus *Metarhizium anisopliae* var. *anisopliae*. *Curr Microbiol*. 2006;53(3):217–21. doi:10.1007/s00284-006-0078-6.
101. da Silva MV, Santi L, Staats CC, da Costa AM, Colodel EM, Driemeier D, et al. Cuticle-induced endo/exoacting chitinase CHIT30 from *Metarhizium anisopliae* is encoded by an ortholog of the *chi3* gene. *Res Microbiol*. 2005;156(3):382–92. doi:10.1016/j.resmic.2004.10.013.
102. de las Mercedes Dana M, Limon MC, Mejias R, Mach RL, Benitez T, Pintor-Toro JA, et al. Regulation of chitinase 33 (*chi33*) gene expression in *Trichoderma harzianum*. *Curr Genet*. 2001;38(6):335–42.
103. Guillen D, Sanchez S, Rodriguez-Sanoja R. Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol*. 2010;85(5):1241–9. doi:10.1007/s00253-009-2331-y.
104. Liu K, Zhang W, Lai Y, Xiang M, Wang X, Zhang X, et al. *Drechlerella stenobrocha* genome illustrates the mechanism of constricting rings and the origin of nematode predation in fungi. *BMC Genomics*. 2014;15:114. doi:10.1186/1471-2164-15-114.
105. Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack KE, et al. Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiol*. 2011;156(2):756–69. doi:10.1104/pp.111.176347.
106. Ni M, Feretzaki M, Sun S, Wang X, Heitman J. Sex in fungi. *Annu Rev Genet*. 2011;45:405–30. doi:10.1146/annurev-genet-110410-132536.
107. Pattemore JA, Hane JK, Williams AH, Wilson BA, Stodart BJ, Ash GJ. The genome sequence of the biocontrol fungus *Metarhizium anisopliae* and comparative genomics of *Metarhizium* species. *BMC Genomics*. 2014;15:660. doi:10.1186/1471-2164-15-660.
108. Chaverri P, Liu M, Hodge KT. A monograph of the entomopathogenic genera *Hypocrella*, *Moelleriella*, and *Samuelsia* gen. nov. (Ascomycota, Hypocreales, Clavicipitaceae), and their *aschersonia*-like anamorphs in the Neotropics. *Studies in Mycology*. 2008;60:1–66. doi:10.3114/sim.2008.60.01.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

