BMC Microbiology

Identifying candidate *Aspergillus* pathogenicity factors by annotation frequency

Kayla K. Pennerman[1*] , Guohua Yin[2], Anthony E. Glenn[1] and Joan W. Bennett[2]

Abstract

Background: Members of the genus *Aspergillus* display a variety of lifestyles, ranging from saprobic to pathogenic on plants and/or animals. Increased genome sequencing of economically important members of the genus permits effective use of "-omics" comparisons between closely related species and strains to identify candidate genes that may contribute to phenotypes of interest, especially relating to pathogenicity. Protein-coding genes were predicted from 216 genomes of 12 *Aspergillus* species, and the frequencies of various structural aspects (exon count and length, intron count and length, GC content, and codon usage) and functional annotations (InterPro, Gene Ontology, and Kyoto Encyclopedia of Genes and Genomes terms) were compared.

Results: Using principal component analyses, the three sets of functional annotations for each strain were clustered by species. The species clusters appeared to separate by pathogenicity on plants along the first dimensions, which accounted for over 20% of the variance. More annotations for genes encoding pectinases and secondary metabolite biosynthetic enzymes were assigned to phytopathogenic strains from species such as *Aspergillus flavus*. In contrast, *Aspergillus fumigatus* strains, which are pathogenic to animals but not plants, were assigned relatively more terms related to phosphate transferases, and carbohydrate and amino-sugar metabolism. Analyses of publicly available RNA-Seq data indicated that one *A. fumigatus* protein among 17 amino-sugar processing candidates, a hexokinase, was up-regulated during co-culturing with human immune system cells.

Conclusion: Genes encoding hexokinases and other proteins of interest may be subject to future manipulations to further refine understanding of *Aspergillus* pathogenicity factors.

Keywords: *Aspergillus*, Comparative gene annotation, Comparative protein annotation, Hexokinase, Pathogenicity factors

Background

Only a minority of fungi are pathogenic, mostly on plants, while the majority of fungal species are saprobes or mutualists [1]. Numerous studies have investigated the underlying genetics of fungal pathogenicity as well as the environmental, biological and pathogenic relevance of fungi to basic science and human affairs [2]. The ascomycetous genus

Aspergillus provides an intriguing model for studying differentiation among opportunistic pathogens (both plant and animal) and saprobic decomposers. All members of this genus live largely as saprobes. However, several species are able to cause rots on living plant tissues and/or invasive aspergillosis in immunocompromised mammals [3–5]. Systemic aspergillosis is a life-threatening disease posing considerable public health and economic concerns. In recent decades, there has been a rise in the numbers of cases of invasive aspergillosis, likely due to an increase in immunosuppressive chemotherapy treatments, as well as organ and

* Correspondence: kkpennerman@gmail.com
[1]United States Department of Agriculture, Toxicology and Mycotoxin Research Unit, Athens, GA 30605, USA
Full list of author information is available at the end of the article

Pennerman *et al. BMC Microbiology*   (2020) 20:342

Page 2 of 11

stem cell transplantations [6–8]. Invasive aspergillosis is usually caused by strains of *A. fumigatus*; less common *Aspergillus* causative agents are *A. fischeri*, *A. flavus*, *A. nidulans*, *A. niger* and *A. terreus* [9]. Human pathogenesis by *Aspergillus* species is complex and requires the normally saprotrophic *Aspergillus* spp. to adapt to the environment of the human lung [10]. Hospitalizations due to *Aspergillus* infections cost the United States an estimated 1.2 billion dollars annually [11]. One *Aspergillus* species, *A. sydowii*, is salt-tolerant pathogen of coral and humans that may become more destructive with continued global warming, along with other pathogenic species [12].

Opportunistic plant infections by *Aspergillus* species are also common after drought, insect damage or other environmental stresses. In particular, infection by *A. flavus* and *A. parasiticus* strains cause large economic losses in agriculture due to associated contamination with mycotoxins, notably aflatoxins. Aflatoxin contamination alone costs the United States 50 million to 1.7 billion dollars each year [13]. *A. flavus* is not only the primary causative agent of *Aspergillus* infections and aflatoxin contamination in crops, it is simultaneously the second most common cause of aspergillosis in human patients. In contrast, while *A. fumigatus* is the most common cause of human and veterinary aspergillosis, it is not known to cause disease in any plant host [14–16]. Essentially, *Aspergillus* pathogens can be grouped by those that may infect both plants and animals (phytopathogenic), and those that are not known to cause diseases in plants (non-phytopathogenic). This distinction is not sufficiently explored in published research to date.

Different infection strategies are required to cause disease in different hosts; thus, pathogenic species and strains must have genes that enable disease-causing infections and suppress resistance responses. The expression of relevant genes may be influenced by environmental conditions such as nutrient composition and response to host defenses [17]. Mammalian pathogens have to evade circulating immune system cells while the phytopathogens must be able to penetrate sturdy cell walls. Further, animal fungal pathogens preferentially disperse by hyphae or arthroconidia instead of conidia, are frequently dimorphic and often lack a known complete sexual cycle [18, 19]. Plant pathogens have hydrolytic enzymes to degrade cuticles and plant cell walls, and the ability to form appressoria [19]. Phytopathogenicity genes necessary for disease development include those that are involved in host recognition, signaling, secondary metabolite synthesis, cell wall integrity, appressorial formation, degradation of host cuticle and cell wall, uptake of nutrients and genes with unknown roles [20]. For both plant and animal pathogens, the ability to withstand abiotic stresses within the host environment, such as hypoxia, oxidative burst and mammalian body temperatures, are likely additional virulence factors [16, 19, 21–24].

In this study, multiple comparisons of genome annotations, largely represented by *A. flavus* and *A. fumigatus*, were used to address the discrepancy between the host ranges of phytopathogenic and non-phytopathogenic *Aspergillus* species. The recent substantial increase in genomic sequencing of filamentous fungi permits more focused comparisons among strains between and within species [25]. Both structural (high-throughput prediction of the composition and arrangement of physical motifs) and functional (high-throughput prediction of the biological uses of gene products) genomics are routinely performed on newly-sequenced genomes. Usually, genome size, GC content, number of predicted genes and predicted functions of those genes are reported in genome announcements. Biological interpretations of large data sets may uncover hitherto overlooked genes that define what a fungus can or cannot do. A consistent method of gene prediction across a set of genomes and genome annotations is a useful starting place to identify genes that contribute to phenotypes of interest. Such a method has the benefit of not requiring wet laboratory work, complementing transcriptomic or proteomic experiments that test predictions and quantify expression levels, and leading to new hypotheses for targeted gene manipulations.

The present study demonstrates the utility of this approach by looking for new pathogenicity genes within the *Aspergillus* genus. The identified gene candidates are corroborated by reports on the effects of *Aspergillus* deletion mutants on pathogenicity or virulence, and by published transcriptomic data. To the authors' knowledge, this the first report of a method to search for genes contributing to a phenotype of interest by gene annotation frequencies across a genus. The long-term objective of our research is to develop a pipeline for extracting new, mycologically-relevant information from the wealth of genomics data stored in public databases. These bioinformatics findings will guide hypotheses that can be tested later via methods such as gene deletion. Our immediate goals were to 1) conduct a comparative analysis of annotations for 216 *Aspergillus* genomes, 2) identify previously-unidentified pathogenicity factors and 3) use previously-reported information to determine if the candidate pathogenicity factors are known to affect virulence or change expression level during infection processes.

Results

Functional annotation relative frequencies clustered strains by species

A total of 216 *Aspergillus* genomes were initially included in the study (Additional file 1). The structural

aspects of predicted genes were similar regardless of host association (Additional files 2, 3, 4). The ranges are given in parentheses for the following attributes: gene counts (8833 to 14,749), average gene lengths (703.8 to 1815.7 bp), average exon counts (1.9 to 3.6), average exon lengths (279.2 to 540.8 bp), average intron counts (1.3 to 2.6) and average intron length (77.3 to 92.8 bp). GC content was lower in intronic sequences (32.3 to 40.8%) compared to exonic sequences (52.1 to 58.4%). According to ANOVA, the number of predicted genes and average gene GC content differed among the species (p-value $< 2^{-16}$). By t-test, there were no statistically significant differences between the species classified as phytopathogenic or not. Only *A. terreus* strains had significantly different exonic and gene GC content compared to all other species. Some codon bias in favor of AAG for lysine and GAG for glutamate instead of equivalent codons was observed (Additional file 3). The most frequent amino acids in the translated sequences were alanine, glycine, leucine and serine; the least frequent were cysteine and tryptophan (Additional file 4). A t-test indicated that non-phytopathogenic strains had significantly fewer (45.6 versus 79.5 average clusters, p-value = 8.8^{-14}) predicted secondary metabolite clusters (Additional file 5).

The first and second principal components of principal component analysis (PCA) results explained 73.3% of the variance among structural annotations and 22.3 to 38.7% of the variances of relative frequencies of functional annotations. Both structural and functional annotations yielded grouping of strains that largely clustered with other members of the same species (Fig. 1;

Additional file 6). There was an unsurprising overlap of the closely-related *A. flavus, A. oryzae, A. parasiticus* and *A. sojae* species. *A. niger* and *A. tubingensis* strains clustered near each other as well. With the exceptions of phytopathogenic *A. terreus* and non-phytopathogenic *A. sydowii*, most of the species were separated by pathogenicity along the first principal. Use of the third to tenth principal components did not improve graphed separation. Strains *A. sydowii* BOBA1 and *A. terreus* T3_Kankrej were removed from PCA figures of functional annotations as they did not cluster with any other strain and increased the spans of the first and second principal components by more than 100% (Fig. 1; Additional file 6). The inclusion of the two strains did not greatly affect the PCA results using the structural annotations and were included in Additional file 6a. Due to the extreme outlying natures of strains BOBA1 and T3_Krankrej, they were not used in further analyses.

Amino-sugar terms were assigned more frequently to non-phytopathogens

Differentially-assigned annotations (DAAs) were defined as annotations with significantly different relative frequencies between phytopathogenic and non-phytopathogenic strains within a set of InterPro (IPR), Gene Ontology (GO) or Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation terms. Most (80.8 to 91.5%) of the AUGUSTUS-predicted genes in each strain were assigned at least one functional annotation term (Additional file 5). Of the retrieved 211 genes from PHI-base, 209 were annotated with 325 IPR, 322 GO and 379 KEGG terms (Additional file 7).
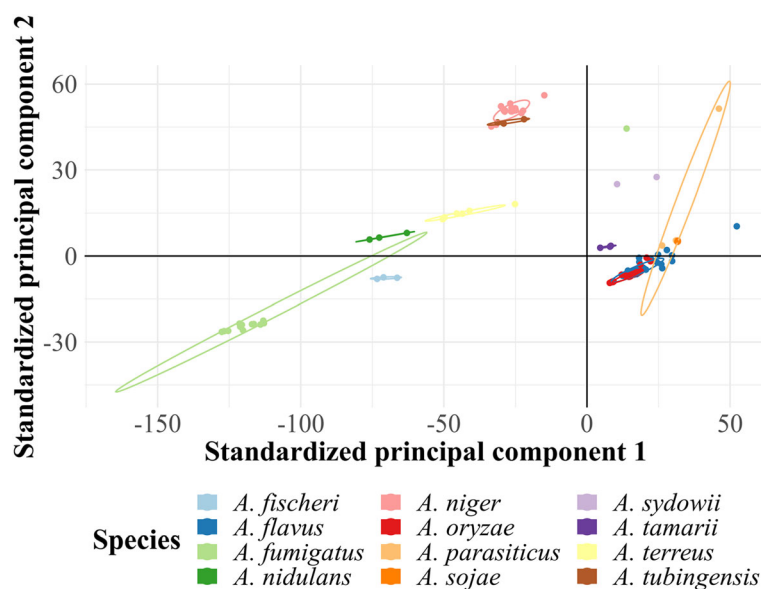

Fig. 1 IPR functional annotation cluster by *Aspergillus* species. Non-phytopathogenic strains trend to the left of the plot. Other PCA plots and the corresponding scree plots are shown in Additional file 6

Pennerman *et al. BMC Microbiology*   (2020) 20:342

Page 4 of 11

Between the phytopathogenic and non-phytopathogenic species, 316 IPR, 214 GO and 1603 KEGG terms were found to have significantly different relative frequencies of assignment. Among these, 58 IPR, 51 GO terms and 74 KEGG terms matched those applied to *Aspergillus* genes in the PHI-base database. Limiting the search to only functionally annotated PHI-base genes that non-lethally affect pathogenicity and/or virulence of *Aspergillus* when mutated yielded 32 IPR, 20 GO terms and 52 KEGG DAAs (Fig. 2, Additional files 8, 9). The matched IPR DAAs that were higher in phytopathogenic strains relative to number of predicted genes included annotations related to fatty acid synthesis (IPR026025), oxidoreduction (IPR036812, IPR023210), zinc permease (IPR004698), secondary metabolite synthesis (IPR001227, IPR001242, IPR010071, IPR014030, IPR016035, IPR020801, IPR020841, IPR032088, IPR042099) and pectin degradation (IPR011050). Matched IPR annotations assigned more frequently to the non-phytopathogenic species included phosphatases and kinases (IPR000719, IPR011009, IPR036457), and sugar hydrolases and transferases (IPR001830, IPR017853). Multiple amino acid and sugar metabolism IPR annotations that were not shared between the list of DAAs and PHI-base gene annotations were over-represented in *A. fumigatus* compared to *A. flavus*. These 21 terms predicted a chitinase, a galactose mutarotase, glycoside hydrolases, glucoamylases and a peptidoglycan deacetylase (Additional file 10, highlighted in blue).

DAA GO and KEGG terms related to peptidoglycan metabolic processes (GO:0009254), glycosaminoglycan metabolic processes (GO:0030203), carbohydrate metabolic processes (GO:0005975), peptidoglycan turnover (GO:0009254), amino sugar and nucleotide sugar metabolism (KO00520), and galactose metabolism (KO00052) were commonly enriched in the non-phytopathogenic strains *A. fischeri*, *A. flavus* and *A. sydowii*, but not in *A. nidulans*. Except for *A. terreus*, terms related to phenol-containing compound metabolic processes (GO: 0018958) and histidine metabolism (KO00340) were enriched in the phytopathogenic strains. Further, *A. flavus*, *A. oryzae*, *A. parasiticus*, *A. sojae* and *A. tamarii* had enrichment of a glycosaminoglycan degradation term (KO00531).

The major difference between the phytopathogens and non-phytopathogens was the number of translated amino acid sequences annotated to a broad category of amino and/or sugar (amino-sugar) metabolism and modification terms. This trend was also observed when comparing the enrichments between the full proteomes and sub-proteomes of *A. flavus* NRRL 3357 and *A. fumigatus* Af293. The strains had 226 and 131 proteins, respectively, back-matched from the DAAs. Sub-proteomic enrichment analyses identified many enriched GO and KEGG terms (Fig. 3; Additional file 11). *A. flavus* NRRL 3357 uniquely had enrichment of arylsulfatase
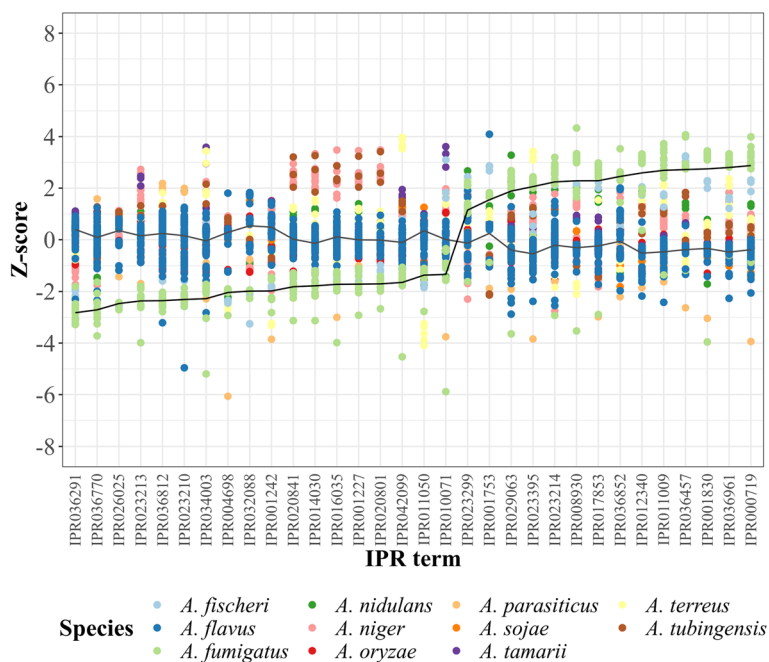
Fig. 2 Relative frequencies of IPR terms associated with pathogenicity and/or virulence. Only annotations with significantly different relative frequencies between phytopathogenic and non-phytopathogenic species, and with IPR annotations shared with PHI-base genes are shown. Z-scores were calculated using percentages of total annotations for a strain. Two lines track the mean Z-scores for *A. flavus* and *A. fumigatus*. Similar graphs for GO and KEGG annotations are shown in Additional file 8. Annotation terms and definitions are listed in the same order as the x-axes in Additional file 9

activity (GO:0004065), transferase activity (GO:0016603, GO:0016755), 121 polyketide and secondary metabolite synthesis terms (Additional file 10, highlighted in orange) and fatty acid synthases (K00665). The sub-proteomic annotation of *A. fumigatus* Af293 was enriched in functions related to amino-sugar metabolism (GO:0000270, GO:0009254, GO:0030203, K00844, K12407), carbohydrate metabolism (GO:0005975, GO:0005984, GO:0005991, GO:0005992, GO:0009311, GO:0009312, GO:0046351), phosphorylation (GO:0004672, GO:0006468, GO:0016310, GO:0016772, GO:0016773, K02216, K07198, K08794, K08811), nucleotide binding (GO:0000166, GO:0005524, GO:0017076, GO:GO:0030554, GO:0032553, GO:0032555, GO:0032559, GO:0035639, GO:0097367, GO:1901265, K11665), oxidation-reduction (K22727, K22728), steroid biosynthesis (K00512, K21445, K22726) and chromatin structure (K14440).

Annotation terms related to amino-sugar metabolism IPR037950, GO:0000270, GO:0009254, GO:0030203, K00844, K12407, K13748, K19223 and K21471 were assigned to 12 *A. flavus* NRRL 3357 and 17 *A. fumigatus* Af293 AUGUSTUS-predicted protein-coding genes. These predicted *A. flavus* and *A. fumigatus* genes were then matched to 9 and 17 NCBI-curated proteins of *A. flavus* NRRL 3357 (XP_002372155.1, putative hexokinase; XP_002372244.1, putative hexokinase; XP_002373069.1, integral membrane protein; XP_002373361.1, polysaccharide deactylase family protein; XP_002373849.1, hexokinase family protein; XP_002375551.1, hexokinase family

protein XprF; XP_002379109.1, putative oxidoreductase; XP_002382200.1, putative hexokinase Kxk; XP_002385204.1, UPF0075 domain protein) and *A. fumigatus* Af293 (XP_746328.1, UPF0075 family protein; XP_747575.1, polysaccharide deactylase family protein; XP_747679.1, putative hexokinase; XP_747854.1, putative glucokinase GlkA; XP_747946.1, putative LysM domain protein; XP_748231.1, hypothetical protein AFUA_5G0110; XP_749104.1, putative hexokinase; XP_749181.1, putative hexokinase; XP_749720.1, putative hexokinase Kxk; XP_751374.1, putative C6 transcription factor; XP_752897.1, polysaccharide deactylase family protein; XP_753281.1, putative NlpC/P60-like cell-wall peptidase; XP_753725.1, polysaccharide deactylase family protein; XP_755146.1, hexokinase family protein; XP_755905.1, putative hexokinase family protein XprF; XP_755969.1, putative glucokinase; XP_001481464.1, hypothetical protein AFUA_6G09315), respectively.

A hexokinase gene is upregulated in response to co-inoculation with human cells

The percentages of sequencing reads from NCBI SRA accession PRJNA560197 that mapped to the *A. fumigatus* genome were low, probably reflecting lower fungus to human ratios in the co-cultures compared to the experiments performed to yield the reads for the PRJEB1583 dataset (Additional file 12). For the former, sequencing reads mapped to 48.8 to 66.5% of

Fig. 3 Enrichment of GO and KEGG annotation terms. Enrichment of GO biological process terms for **a**. *A. flavus* and **b**. *A. fumigatus*, and **c**. KEGG pathways for both species on the KEGG Mapper map for *A. flavus* metabolic pathways. Enrichment was based on sub-proteomic annotations compared to annotations of the full predicted proteomes. Enriched GO terms are in yellow-shaded boxes. Enriched KEGG pathways are highlighted in the indicated colors. The original diagram of *A. flavus* KEGG metabolic pathways is shown in Additional file 11

A. fumigatus NCBI-annotated genes; the PRJEB1583 reads mapped to 93.5 to 98.1% of the genes. One gene, encoding the hexokinase XP_749720.1, was significantly (p-value = 0.0006) up-regulated when *A. fumigatus* was co-cultured with human dendritic cells for 4 h compared to incubation alone (Fig. 4). The gene also had increased expression when the fungus was cultured in a similar nutrient medium with macrophage-like cells for 1 h. However, at 2 h, the gene expression lowered back to the initial level. Non-significant upregulation was commonly observed for the gene encoding the glucokinase XP_747854 (p-value = 0.18). Genes encoding these proteins were not in the PHI-base.

Discussion

The aim of this study was to identify candidate protein-coding genes within the *Aspergillus* genus that may contribute to pathogenicity. A comparative protein annotation approach was employed in which differential frequencies of annotation assignments were used to find predicted annotations that may be over or underrepresented (DAAs) within a set of genomes. The structural annotations did not yield a meaningful distinction between the phytopathogenic and non-phytopathogenic *Aspergillus* strains, though the larger number of secondary metabolite gene clusters in the phytopathogenic species tracked with the fact that fungal plant pathogens have larger secretomes than non-phytopathogens [26]. In contrast, all three sets of IPR, GO and KEGG functional annotations indicated that genes functioning in amino-sugar metabolism were assigned relatively more frequently to genes of the non-phytopathogenic strains

A. fischeri, *A. fumigatus* and *A. sydowii*. Six of the *A. fumigatus* predicted genes retrieved from PHI-base were predicted to be involved in sugar metabolism, being annotated as glycosyl transferases, glycoside hydrolases, glucan synthases or mannosidases. Five of these genes cause loss of virulence or death when mutated: *FKS1* (PHI-base accession PHI2533), *AGS1* (PHI3902), *AGS2* (PHI3903), *AGS3* (PHI3904) and *GEL2* (PHI434) [27–29]. One gene, *tslA* (PHI7121; trehalose synthase), increases virulence after mutation [30]. The one carbohydrate metabolism-related *A. flavus* PHI-base gene (*PECA*; PHI88) was a pectinase [31].

The agreement among predicted overrepresented annotations led to the hypothesis that the identified genes play important roles in the pathogenicity of *A. fumigatus*, which might be indicated by differential expression during infection and disease progression in human hosts compared to growth in single-species cultures. Briefly, invading *A. fumigatus* is subject to phagocytosis by macrophages mediated by antigen-presenting dendritic cells. Analyses of publicly available RNA-Seq datasets indicated that one *A. fumigatus* Af293 hexokinase gene encoding the protein XP_749720.1 out of 17 observed candidates has time-dependent increased expression when the fungus is incubated with human immune cells. This gene is not known to have been previously studied as a pathogenicity factor. Its nucleotide sequence did not match (E-value ≤10) any human genes in the NCBI database, indicating that human RNA sequencing reads extracted from co-cultures should not substantially map to this hexokinase gene in the *A. fumigatus* genome. If the hexokinase is a pathogenicity factor and has a sufficiently different structure from human
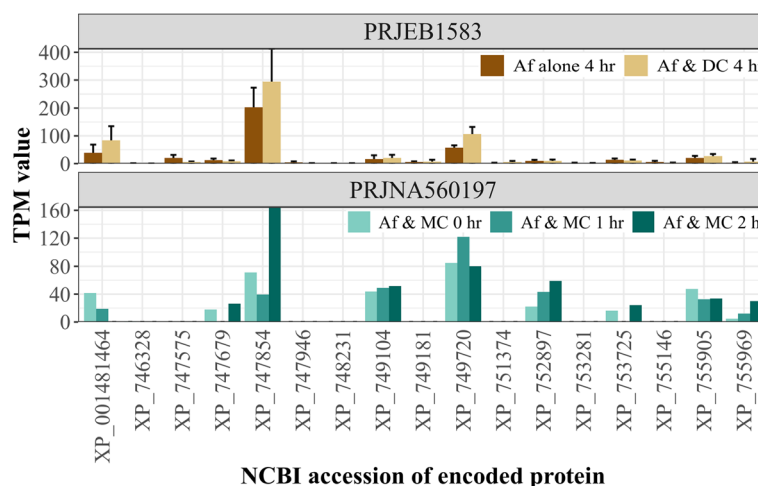

Fig. 4 Gene expression during co-culture with human immune cells. Expression of *A. fumigatus* (Af) amino-sugar processing genes during co-culture with dendritic cells (DC) for 4 h or macrophages (MC) for 0, 1 or 2 h. RNA transcripts were quantified as transcript per million reads (TPM) values for the 17 *A. fumigatus* Af293 genes of interest. Error bars represent one standard deviation above the mean. Facets are labeled by the NCBI SRA Project accession. The gene encoding XP_749720.1, a hexokinase, was significantly up-regulated during co-culture with dendritic cells

Pennerman *et al. BMC Microbiology* (2020) 20:342

Page 7 of 11

proteins, it may be useful as a target for inhibitory drugs to treat aspergillosis.

While it is not immediately clear what roles sugar-metabolizing genes may play in aspergillosis caused by *A. fumigatus* beyond energy production and storage, it can be hypothesized that the genes are involved in fungal pathogen signaling on the cell wall. Pathogenic fungi interact with host immune systems via chemical signatures called pattern-associated molecular patterns (PAMPs) and host pattern recognition receptors. Fungal PAMPs include complex carbohydrates in the cell walls which bind Toll-like receptors and C-type lectin receptors found on animal mononuclear phagocytes [32–36]. This binding initiates signaling cascades that induce the release of cytokines, phagocytosis and cell death. Most fungal PAMPs in mammalian hosts are glucose-containing macromolecules, including mannoproteins, chitin and β-glucans [37].

The higher number of genes functioning in amino-sugar processing could also be related to *N*-acetylglucosamine synthesis and/or breakdown. This amino-sugar has structural and functional roles in fungal cell walls, and in cell signaling for expression of virulence genes of *Candida albicans* and pseudohyphal morphogenesis of *Candida* and *Yarrowia* species [38]. Hexokinases, specifically, are involved in morphogenesis and virulence along with their nominal roles in life-supporting sugar and *N*-acetylglucosamine metabolism in *A. fumigatus* and *Candida albicans* [39–42]. There are several reported disease-associated glycosylated antigens, detoxifying catalases and host-adhering sialic acids from *A. fumigatus* [10, 43–49]. *A. fumigatus* conidial surfaces have 3 to 20 times more sialic acids than the less virulent or non-pathogenic *Aspergillus* species *A. auricomus*, *A. ornatus* and *A. wentii* [50]. Perhaps, the *A. fumigatus* candidate pathogenicity factors identified here could be involved in the synthesis of those macromolecules. Fungal carbohydrates and bacterial peptidoglycans additionally induce plant innate immunity [51–53]. Altogether, the overrepresentation of amino-sugar processing genes in *A. fumigatus* compared to *A. flavus* may suggest that *A. fumigatus* PAMPs are modified in unique, species-specific ways to be misrecognized or induce improper responses by immunocompromised animal hosts and/or are easily recognized by plant hosts. Further experimental study is required to properly test the above hypotheses and to understand the differences among the strains used here.

While the authors are unaware of strain idiosyncrasies that may result in one strain being better classified as plant pathogenic or not, the strains had noticeable differences in genome characterizations. The genus itself is not amenable to neat categorization between phytopathogenic and non-phytopathogenic species. Differences in

strain virulence and/or host preference may be implied by isolation source (Additional file 1). These differences may have been obscured here due to grouping by named species. The non-phytopathogenic species *A. fumigatus*, *A. nidulans* and *A. sydowii* can be endophytes [54–56]. Despite not being associated with plant diseases, *A. fumigatus* has genes encoding cellulases, hemicellulases and pectinases, but no gene sets uniquely shared with non-*Aspergillus* human fungal pathogens [57]. In this study, *A. terreus* was found to have a lower number of proteins annotated as pectinase-like proteins compared to other phytopathogenic species, which may reflect that *A. terreus* infects leaves where pectin levels are lower compared to pectin levels in fruits and seeds [58–60]. *A. nidulans* had fewer amino-sugar metabolism genes than the other non-phytopathogenic species. *A. nidulans*, compared to *A. fumigatus*, induces a weaker oxidative burst by human immune cells and is phagocytosed at a slower rate by rodent macrophages [61].

Conclusions

A direct correlation between frequencies of amino-sugar processing genes and virulence in animal hosts by *Aspergillus* strains could support the hypothesis that the amino-sugar genes of interest are involved in pathogen-host recognition. An application of the comparative protein annotation method used here to additional transcriptomic or proteomic data would help further identify and test pathogenicity candidates by their expression patterns.

Methods

Genomes, gene predictions and annotations

Genomic sequences were retrieved from the NCBI Genome database [62]. Only species with at least three different strain genomes publicly available by April 14, 2020 were included, totaling 217 genomes from 12 species of *Aspergillus*: three from *A. fischeri*, 64 from *A. flavus*, 14 from *A. fumigatus*, three from *A. nidulans*, 17 from *A. niger*, 92 from *A. oryzae*, three from *A. parasiticus*, five from *A. sojae*, three from *A. sydowii*, three from *A. tamarii*, seven from *A. terreus* and three from *A. tubingensis* (Additional file 1). Strain *A. terreus* ATCC 20542 was excluded due to an abnormally low genome size (138.52 kbp). *Aspergillus* species were classified according to reported ability to cause disease and persistent rot in live plants in environmental settings. Therefore, *A. flavus*, *A. niger*, *A. parasiticus*, *A. tamarii*, *A. terreus* and *A. tubingensis* were labeled as phytopathogenic and were compared to the non-phytopathogenic species *A. fischeri*, *A. fumigatus*, *A. nidulans* and *A. sydowii*. *A. oryzae* and *A. sojae* comprise domesticated strains of *A. flavus* and *A. parasiticus*, respectively, and may not reflect natural loss of phytopathogenicity [63–65]. These two species were excluded from statistical tests comparing phytopathogenic

to non-phytopathogenic species. Secondary metabolite gene clusters were predicted for all strains using antiSMASH at default settings [66].

Gene predictions were performed on both DNA strands using AUGUSTUS version 3.0 trained on *A. nidulans* for all *Aspergillus* strains [67]. The resulting genomic annotations for each strain were parsed to calculate gene counts, average gene length (from start codon to stop codon, inclusive), average exon frequency per gene, average exon length (inclusive of stop codon), average intron frequency per gene, average intron length, average gene GC content (from start codon to stop codon, inclusive), average exonic GC content (inclusive of stop codon), average intronic GC content and codon usage (exclusive of stop codons TAA and TAG). Translated sequences were functionally annotated using InterProScan version 5.39 and KofamScan version 1.2 [68, 69], and the frequencies of each assigned (E-value $\leq 1^{-50}$) IPR, GO and KEGG annotation term were counted. Frequencies of annotations for each strain were normalized as percentages of total predicted gene count for the strain. The 211 *Aspergillus* genes present in PHI-base version 4.9 were uploaded after being experimentally studied as pathogenicity or virulence factors by other researchers [70]. The PHI-base set of genes comprised 18 from *A. flavus*, 189 from *A. fumigatus* and 4 from *A. nidulans*. These genes were retrieved and subjected to annotation by InterProScan and KofamScan.

Identification of *A. flavus* and *A. fumigatus* proteins with enriched annotations

Two-way analysis of variance (ANOVA) was used to determine if species identity was a factor in multiple genomic structural aspects. DAAs were identified using PCA and independent *t*-tests. Statistical tests were performed with counts of assigned annotations normalized as percentages of total annotations for a strain. Figures were generated using R packages ape, DECIPHER, dplyr, factoextra, FactoMineR, ggdendro and ggplot2 [71–78]. For *t*-tests, relative differences of at least 0.01% with Benjamini-Hochberg corrected *p*-values $< 1^{-4}$ (false discovery rate = 0.1%) were considered significant. Excluding the non-hierarchical IPR terms, GOATOOLS and Fisher's exact test (α-level = 1^{-3} for GO annotations; 1^{-6} for KEGG annotations) were used to identify enriched GO and KEGG terms compared to the full list of proteomic annotations [79]. DAAs between phytopathogenic versus non-phytopathogenic species were back-matched to the predicted proteomes to produce subproteomes. The derived proteins of interest had functional annotation terms only present in the list of DAAs. In other words, proteins with annotation terms not in the DAA list were excluded from the sub-proteomes. Enrichment analyses with GOATOOLS and Fisher's exact test also were performed in a second alternative method, comparing the

annotated sub-proteomes to the annotated full proteomes. GOATOOLS and KEGG Mapper were used to generate figures with the enriched terms [78, 79].

Transcriptomic data analysis of *A. fumigatus* cultured with human immune cells

RNA-Seq reads were retrieved from the NCBI SRA database project accessions PRJEB1583 and PRJNA560197 [80, 81]. The datasets comprised RNA-Seq reads from *A. fumigatus* incubated alone or with human dendritic cells (PRJEB1583; read accessions ERR236917, ERR236920, ERR236932, ERR236939, ERR236940, ERR236942, ERR236948, ERR236949, ERR236951, ERR236953, ERR236959, ERR236962, ERR236963, ERR236972), or with macrophage-like cells (PRJNA560197; read accessions SRR9965307, SRR9965308, SRR9965309). The sequencing reads were processed as previously described with slight modifications [23]. Briefly, read quality was checked, then aligned to the *A. fumigatus* Af293 genome (NCBI GenBank assembly GCA_000002655.1) guided by the respective GFF3 file [82]. Utilizing BLAST, AUGUSTUS-predicted *A. fumigatus* Af293 proteins with enriched DAAs of interest were matched to accessions in the NCBI Protein and Nucleotide databases with a maximum E-value of 1^{-100} [83]. Presence of the corresponding genes was assessed in all experimental groups: *A. fumigatus* spores incubated alone or co-cultured with human dendritic cells (5:1 spores:human cells) in complete RPMI 1640 medium at 37 °C for 4 h (PRJEB1583, seven replicates), and *A. fumigatus* cultured with a human leukemia cell line differentiated into macrophage-like cells (2:1 spores:human cells) for 0, 1 or 2 h at 37 °C in a modified RPMI 1640 medium (PRJNA560197, one replicate). For the PRJEB1583 experiment, *t*-tests (α-level = 1^{-3}) were performed to compare gene expression quantified as transcripts per million reads.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-020-02031-y.

Additional file 1: Table S1. Strains included in the present study.

Additional file 2: Figure S1. Structural aspects of predicted *Aspergillus* genes. **a.** Number of genes; **b.** average exon count per gene; **c.** average intron count per gene; **d.** average gene length; **e.** average exon length per gene; **f.** average intron length per gene; **g.** average gene GC content; **h.** average exon GC content; **i.** average intron GC content.

Additional file 3: Figure S2. Amino acid usage among predicted *Aspergillus* genes.

Additional file 4: Figure S3. Codon usage among predicted *Aspergillus* genes.

Additional file 5: Table S2. Overview of functional annotation of *Aspergillus* strains.

Pennerman *et al. BMC Microbiology*   (2020) 20:342

Page 9 of 11

Additional file 6: Figure S4. Structural and functional annotation cluster by *Aspergillus* species. PCA and scree plots for **a., b.** structural aspects; **c.** IPR terms; **d., e.** GO terms; **f., g.** KEGG terms.

Additional file 7: Table S3. Functional annotation of pathogenicity/virulence *Aspergillus* genes.

Additional file 8: Figure S5. Relative frequencies of **a.** GO and **b.** KEGG terms associated with virulence in *Aspergillus*. Annotation terms and definitions are listed in the same order in Additional file 9.

Additional file 9: Table S4. Definitions of annotations shown in Fig. 2 and Additional file 8. Amino-sugar annotations mentioned in the text are highlighted in blue.

Additional file 10: Table S5. Differentially-assigned annotations not assigned to PHI-base genes. Overrepresented IPR amino-sugar terms are highlighted in blue. Enriched GO and KEGG polyketide and secondary metabolite synthesis terms are highlighted in orange.

Additional file 11: Figure S6. Original diagram of *A. flavus* KEGG metabolic pathways. Updates to the image are available via KEGG Mapper [79].

Additional file 12: Table S6. RNA-Seq metadata and read mapping results.

Abbreviations
ANOVA: Analysis of variance; DAA: Differentially-assigned annotation; GO: Gene Ontology; IPR: InterPro; KEGG: Kyoto Encyclopedia of Genes and Genomes; PCA: Principal component analysis

Acknowledgements
Not applicable.

Authors' contributions
KKP designed and performed the study, analyzed and interpreted the results, and wrote the manuscript. GY, AEG and JWB provided valuable insights for designing the study, interpreting results and revising the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Sources of genomic data and summarized results reported in the article are included in this published article and its additional files.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests. Use of a company or product name by the United States Department of Agriculture does not imply approval or recommendation of the product to the exclusion of others that may be suitable.

Author details
[1]United States Department of Agriculture, Toxicology and Mycotoxin Research Unit, Athens, GA 30605, USA. [2]Department of Plant Biology, Rutgers University, The State University of New Jersey, New Brunswick, NJ 08901, USA.

References
1. Dighton J, White JF. The fungal community. Its organization and role in the ecosystem. Boca Raton: CRC Press; 2017. https://doi.org/10.1201/9781315119496.
2. Heitman J, Howelett BJ, Crous PW, Stukenbrock EH, James TY, Gow NAR. The fungal kingdom. Washington DC: ASM Press; 2018. https://doi.org/10.1128/9781555819583.
3. Goldman GH, Osmani SA. The aspergilli: genomics, medical aspects, biotechnology, and research methods. Boca Raton: CRC Press; 2008. https://doi.org/10.1201/9781420008517.
4. Bennett JW. *Aspergillus*: a primer for the novice. Med Mycol. 2009;47:S5–S12. https://doi.org/10.1080/13693780802712515.
5. Gupta VK. New and future developments in microbial biotechnology and bioengineering: *Aspergillus* system properties and applications. Amsterdam: Elsevier; 2016. https://doi.org/10.1016/C2014-0-00304-8.
6. Latgé J-P. *Aspergillus fumigatus* and aspergillosis. Clin Microbiol Rev. 1999;12: 310–50.
7. Latgé J-P, Steinbach WJ. Aspergillus fumigatus and aspergillosis. Washington DC: ASM Press; 2008.
8. Vallabhaneni S, Benedict K, Derado G, Mody RK. Trends in hospitalizations related to invasive aspergillosis and mucormycosis in the United States, 2000–2013. Open Forum Infect Dis. 2017;4:ofw268.
9. Lass-Flörl C, Cuenca-Estrella M. Changes in the epidemiological landscape of invasive mould infections and disease. J Antimicrob Chemother. 2017;72:i5–i11. https://doi.org/10.1093/jac/dkx028.
10. Nesbitt JR, Steves EY, Schonhofer CR, Cait A, Manku SS, Yeung JHF, Bennet AJ, McNagny KM, Choy JC, Hughes MR, Moore MM. The *Aspergillus fumigatus* sialidase (Kdnase) contributes to cell wall integrity and virulence in amphotericin B-treated treated mice. Front Microbiol. 2017;8. https://doi.org/10.3389/fmicb.2017.02706.
11. Benedict K, Jackson BR, Chiller T, Beer KD. Estimation of direct healthcare costs of fungal diseases in the United States. Clin Infect Dis. 2019;68:1791–7. https://doi.org/10.1093/cid/ciy776.
12. Soler-Hurtado MM, Sandoval-Sierra JV, Machordom A, Diéguez-Uribeondo J. *Aspergillus sydowii* and other potential fungal pathogens in gorgonian octocorals of the Ecuadorian Pacific. PLoS One. 2016;11(11):e0165992. Published 2016 Nov 30. https://doi.org/10.1371/journal.pone.0165992.
13. Mitchell NJ, Bowers E, Hurburgh C, Wu F. Potential economic losses to the US corn industry from aflatoxin contamination. Food Addit Contam Part A Chem Anal Control Expo Risk Assess. 2016;33:540–50. https://doi.org/10.1080/19440049.2016.1138545.
14. Balajee SA, Kano R, Baddley JW, Moser SA, Marr KA, Alexander BD, Andews D, Kontoyiannis DP, Perrone G, Peterson S, Brandt ME, Pappas PG, Chiller T. Molecular identification of *Aspergillus* species collected for the transplant-associated infection surveillance network. J Clin Microbiol. 2009;47:3138–41. https://doi.org/10.1128/JCM.01070-09.
15. Steinbach WJ, Marr KA, Anaissie EJ. Azie N, Quan SP Meier-Kriesche, Apewokin S, horn DL. Clinical epidemiology of 960 patients with invasive aspergillosis from the PATH Alliance registry. J Inf Secur. 2012;65:453–64. https://doi.org/10.1016/j.jinf.2012.08.003.
16. Lamoth F. *Aspergillus fumigatus*-related species in clinical practice. Front Microbiol. 2016;7. https://doi.org/10.3389/fmicb.2016.00683.
17. Pfliegler WP, Pócsi I, Győri Z, Pusztahelyi T. The Aspergilli and their mycotoxins: metabolic interactions with plants and the soil biota. Front Microbiol. 2020;10:2921. https://doi.org/10.3389/fmicb.2019.02921.
18. Calderone RA, Cihlar RL. Fungal pathogenesis: principles and clinical applications. New York: Marcel Dekker; 2002.
19. Sexton AC, Howlett BJ. Parallels in fungal pathogenesis on plant and animal hosts. Eukaryot Cell. 2006;5:1941–9. https://doi.org/10.1128/EC.00277-06.
20. van de Wouw AP, Howlett BJ. Fungal pathogenicity genes in the age of 'omics'. Mol Plant Pathol. 2011;12:507–14. https://doi.org/10.1111/j.1364-3703.2010.00680.x.
21. Casadevall A. Cards of virulence and the global virulome for humans. Microbe. 2006;1:359–64. https://doi.org/10.1128/microbe.1.359.1.
22. Mead ME, Knowles SL, Raja HA, Beattie SR, Kowalski CH, Steenwyk JL, Silva LP, Chiaratto J, Ries LNA, Goldman GH, Cramer RA, Oberlies NH, Rokas A. Characterizing the pathogenic, genomic, and chemical traits of *Aspergillus fischeri*, a close relative of the major human fungal pathogen *Aspergillus*

fumigatus. mSphere. 2019;4:e00018–9. https://doi.org/10.1128/mSphere.00018-19.

23. Pennerman KK, Yin G, Bennett JW, Hua ST. *Aspergillus flavus* NRRL 35739, a poor biocontrol agent, may have increased relative expression of stress response genes. J Fungi (Basel). 2019;5:53. https://doi.org/10.3390/jof5020053.

24. Knowles SL, Mead ME, Sliva LP, Raja HA, Steenwyk JL, Goldman GH, Oberlies NH, Rokas A. Gliotoxin, a known virulence factor in the major human pathogen *Aspergillus fumigatus* is also biosynthesized by its nonpathogenic relative *Aspergillus fischeri*. mBioi. 2020;11:e03361–19. https://doi.org/10.1128/mBio.03361-19.

25. Nierman WC, May G, Kim HS, Anderson MJ, Chen D, Denning DW. What the *Aspergillus* genomes have told us. Med Mycol. 2005;43:S3–5. https://doi.org/10.1080/13693780400029049.

26. Krijger J, Thon MR, Deising HB, Wirsel SGR. Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. BMC Genomics. 2014;15. https://doi.org/10.1186/1471-2164-15-722.

27. Mouyna I, Morelle W, Vai M, Monod M, Léchenne B, Fontaine T, Beauvais A, Sarfati J, Prévost M-C, Henry C, Latgé J-P. Deletion of *GEL2* encoding for a β (1–3) glucanosyltransferase affects morphogenesis and virulence in *Aspergillus fumigatus*. Mol Microbiol. 2005;56:1675–88. https://doi.org/10.1111/j.1365-2958.2005.04654.x.

28. Hu W, Sillaots S, Lemieux S, Davison J, Kauffman S, Breton A, Linteau A, Xin C, Bowman J, Becker J, Jiang B, Roemer T. Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. PLoS Pathog. 2007;3:e24. https://doi.org/10.1371/journal.ppat.0030024.

29. Beauvais A, Bozza S, Kniemeyer O, Formosa C, Balloy V, Henry C, Roberson RW, Dague E, Chignard M, Brakhage AA, Romani L, Latgé J-P. Deletion of the α-(1,3)-glucan synthase genes induces a restructuring of the conidial cell wall responsible for the avirulence of *Aspergillus fumigatus*. PLoS Pathog. 2013;9:e1003716. https://doi.org/10.1371/journal.ppat.1003716.

30. Thammahong A, Caffrey-Card AK, Dhingra S, Obar JJ, Cramer RA. Aspergillus fumigatus trehalose-regulatory subunit homolog moonlights to mediate cell wall homeostasis through modulation of chitin synthase activity. mBio. 2017;8:e00056–17. https://doi.org/10.1128/mBio.00056-17.

31. Shieh MT, Brown RL, Whitehead MP, Cary JW, Cotty PJ, Cleveland TE, Dean RA. Molecular genetic evidence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of *Aspergillus flavus* in cotton bolls. Appl Environ Microbiol. 1997;63:3548–52. https://doi.org/10.1128/AEM.63.9.3548-3552.

32. Mambula SS, Sau K, Henneke P, Golenbock DT, Levitz SM. Toll-like receptor (TLR) signaling in response to *Aspergillus fumigatus*. J Biol Chem. 2002;277: 39320–6. https://doi.org/10.1074/jbc.M201683200.

33. Bellocchio S, Montagnoli C, Bozza S, Gaziano R, Rossi G, Mambula SS, Vecchi A, Mantovani A, Levitz SM, Romani L. The contribution of the toll-like/IL-1 receptor superfamily to innate and adaptive immunity to fungal pathogens *in vivo*. J Immunol. 2004;172:3059–69. https://doi.org/10.4049/jimmunol.172.5.3059.

34. Serrano-Gómez D, Domínguez-Soto A, Ancochea J, Jimenez-Heffernan JA, Leal JA, Corbí AL. Dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin mediates binding and internalization of *Aspergillus fumigatus* conidia by dendritic cells and macrophages. J Immunol. 2004;173: 5635–43. https://doi.org/10.4049/jimmunol.173.9.5635.

35. Serrano-Gómez D, Leal JA, Corbí AL. DC-SIGN mediates the binding of *Aspergillus fumigatus* and keratinophylic fungi by human dendritic cells. Immunobiol. 2005;210:175–83. https://doi.org/10.1016/j.imbio.2005.05.011.

36. Gersuk GM, Underhill DM, Zhu L, Marr KA. Dectin-1 and TLRs permit macrophages to distinguish between different *Aspergillus fumigatus* cellular states. J Immunol. 2006;176:3717–24. https://doi.org/10.4049/jimmunol.176.6.3717.

37. Sorrell TC, Chen SCA. Fungal-derived immune modulating molecules. In: Fallon PG, editor. Pathogen-derived immunomodulatory molecules. Advances in experimental medicine and biology, vol. 666. New York: Springer; 2009. https://doi.org/10.1007/978-1-4419-1601-3_9.

38. Konopka JB. *N*-acetylglucosamine functions in cell signaling. Scientifica. 2012. https://doi.org/10.6064/2012/489208.

39. Yamada-Okabe T, Sakamori Y, Mio T, Yamada-Okabe H. Identification and characterization of the genes for *N*-acetylglucosamine kinase and *N*-acetylglucosamine-phosphate deacetylase in the pathogenic fungus *Candida albicans*. Eur J Biochem. 2001;268:2498–505. https://doi.org/10.1046/j.1432-1327.2001.02135.x.

40. Fleck CB, Brock M. Aspergillus fumigatus catalytic glucokinase and hexokinase: expression analysis and importance for germination, growth, and conidiation. Eukaryot Cell. 2010;9:1120–35. https://doi.org/10.1128/EC.00362-09.

41. Rao KH, Ghosh S, Natarajan K, Datta A. *N*-acetylglucosamine kinase, HXK1 is involved in morphogenetic transition and metabolic gene expression in *Candida albicans*. PLoS One. 2013;8:e53638. https://doi.org/10.1371/journal.pone.0053638.

42. Laurian R, Dementhon K, Doumèche B, Soulard A, Noel T, Lemaire M, Cotton P. Hexokinase and glucokinases are essential for fitness and virulence in the pathogenic yeast *Candida albicans*. Front Microbiol. 2019;10: 327. https://doi.org/10.3389/fmicb.2091.00327.

43. Monod M, Paris S, Sanglard D, Jaton-Ogay K, Bille J, Latgé JP. Isolation and characterization of a secreted metalloprotease of *Aspergillus fumigatus*. Infect Immun. 1993;61:4099–104.

44. Beauvais A, Monod M, Debeaupuis JP, Diaquin M, Kobayashi H, Latgé J-P. Biochemical and antigenic characterization of a new dipeptidylpeptidase isolated from *Aspergillus fumigatus*. J Biol Chem. 1997;272:6238–44. https://doi.org/10.1074/jbc.272.10.6238.

45. Bouchara JP, Sanchez M, Chevailler A, Marot-Leblond A, Lissitzky JC, Tronchin G, Chabasse D. Sialic acid-dependent recognition of laminin and fibrinogen by *Aspergillus fumigatus* conidia. Infect Immun. 1997;65:2717–24.

46. Wasylnka JA, Moore MM. Adhesion of *Aspergillus* species to extracellular matrix proteins: evidence for involvement of negatively charged carbohydrates on the conidial surface. Infect Immun. 2000;68:3377–84. https://doi.org/10.1128/IAI.68.6.3377-3384.2000.

47. Yuen KY, Chan CM, Chan KM, Woo PCY, Che XY, Leung ASP, Cao L. Characterization of AFMP1: a novel target for serodiagnosis of aspergillosis. J Clin Microbiol. 2001;39:3830–7. https://doi.org/10.1128/JCM.39.11.3830-3837.2001.

48. Paris S, Wysong D, Debeaupuis JP, Shibuya K, Philippe B, Diamond RD, Latgé JP. Catalases of *Aspergillus fumigatus*. Infect Immun. 2003;71:3551–62. https://doi.org/10.1128/iai.71.6.3551-3562.2003.

49. Warwas ML, Watson JN, Bennet AJ, Moore MM. Structure and role of sialic acids on the surface of *Aspergillus fumigatus* conidiospores. Glycobiol. 2007; 17:401–10. https://doi.org/10.1093/glycob/cwl085.

50. Wasylnka JA, Simmer MI, Moore MM. Differences in sialic acid density in pathogenic and non-pathogenic *Aspergillus* species. Microbiol. 2001;147: 869–77. https://doi.org/10.1099/00221287-147-4-869.

51. Erbs G, Silipo A, Aslam S, de Castro C, Liparoti V, Flagiello A, Pucci P, Lanzetta R, Parrilli M, Molinaro A, Newman M-A, Cooper RM. Peptidoglycan and muropeptides from pathogens *Agrobacterium* and *Xanthomonas* elicit plant innate immunity: structure and activity. Chem Biol. 2008;15:438–48. https://doi.org/10.1016/j.chembiol.2008.03.017.

52. Liu B, Li J-F, Ao Y, Qu J, Li Z, Su J, Zhang Y, Liu J, Feng D, Qi K, He Y, Wang J, Wang H-B. Lysin motif–containing proteins LYP4 and LYP6 play dual roles in peptidoglycan and chitin perception in rice innate immunity. Plant Cell. 2012;24:3406–19. https://doi.org/10.1105/tpc.112.102475.

53. Silipo A, Erbs G, Shinya T, Dow JM, Parrilli M, Lanzetta R, Shibuya N, Newman M-A, Molinaro A. Glyco-conjugates as elicitors or suppressors of plant innate immunity. Glycobiol. 2010;20:406–19. https://doi.org/10.1093/glycob/cwp2.

54. Liu JY, Song YC, Zhang Z, Wang L, Guo ZJ, Zou WX, Tan RX. Aspergillus fumigatus CY018, an endophytic fungus in *Cynodon dactylon* as a versatile producer of new and bioactive metabolites. J Biotech. 2004;114:279–87. https://doi.org/10.1016/j.jbiotec.2004.07.008.

55. Song X-Q, Zhang X, Han Q-J, Li X-B, Li G, Li R-J, Jiao Y, Zhou J-C, Lou H-X. Xanthone derivatives from *Aspergillus sydowii*, an endophytic fungus from the liverwort *Scapania ciliata* S. lac and their immunosuppressive activities. Phytochem Lett. 2013;6:318–21. https://doi.org/10.1016/j.phytol.2013.03.012.

56. Xu G, Yang S, Meng L, Wang B-G. The plant hormone abscisic acid regulates the growth and metabolism of endophytic fungus *Aspergillus nidulans*. Sci Rep. 2018;8. https://doi.org/10.1038/s41598-018-24770-9.

57. Tekaia F, Latgé JP. *Aspergillus fumigatus*: saprophyte or pathogen? Curr Opin Microbiol. 2005;8:385–92. https://doi.org/10.1016/j.mib.2005.06.017.

58. Fischer RL, Bennett AB. Role of cell wall hydrolases in fruit ripening. Annu Rev Plant Physiol Plant Mol Biol. 1991;42:675–703. https://doi.org/10.1146/annurev.pp.42.060191.003331.

59. Harholt J, Suttangkakul A, Vibe SH. Biosynthesis of pectin. Plant Physiol. 2010;153:384–95. https://doi.org/10.1104/pp.110.156588.

60. Louis B, Roy P, Sayanika DW, Talukdar NC. *Aspergillus terreus* Thom a new pathogen that causes foliar blight of potato. Plant Path Quar. 2013;3:29–33. https://doi.org/10.5943/ppq/3/1/5.

61. Gresnigt MS, Becker KL, Leenders F, Alonso MF, Wang X, Meis JF, Bain JM, Erwig LP, van de Veerdonk FL. Differential kinetics of *Aspergillus nidulans* and *Aspergillus fumigatus* phagocytosis. J Innate Immun. 2018;10:145–60. https://doi.org/10.1159/000484562.

62. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nuc Acid Res. 2019;47:D94–9. https://doi.org/10.1093/nar/gky989.

63. Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, Klich MA, Tabb DL, McDonald WH, Rokas A. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. Curr Biol. 2012;22:1403–9. https://doi.org/10.1016/j.cub.2012.05.033.

64. Kim KM, Lim J, Lee JJ, Hurh B-S, Lee I. Characterization of *Aspergillus sojae* isolated from Meju, Korean traditional fermented soybean brick. J Microbiol Biotechnol. 2017;27:251–61. https://doi.org/10.4014/jmb.1610.10.

65. Watarai N, Yamamoto N, Sawada K, Yamada T. Evolution of *Aspergillus oryzae* before and after domestication inferred by large-scale comparative genomic analysis. DNA Res. 2019;26:465–72. https://doi.org/10.1093/dnares/dsz024.

66. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nuc Acid Res. 2019;47:W81–7. https://doi.org/10.1093/nar/gkz310.

67. Stanke M, Schoeffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinform. 2006;7:62.

68. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinform. 2014;30:1236–40. https://doi.org/10.1093/bioinformatics/btu031.

69. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinform. 2019:btz859. https://doi.org/10.1093/bioinformatics/btz859.

70. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, de Silva N, Martinez MC, Pedro H, Yates AD, Keywan Hassani-Pak K, Hammond-Kosack KE. PHI-base: the pathogen–host interactions database. Nuc Acid Res. 2019;48:D613–20. https://doi.org/10.1093/nar/gkz904.

71. Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. J Stat Softw. 2008;25:1–18. https://doi.org/10.18637/jss.v025.i01.

72. de Vries A, Ripley BD. ggdendro: Create dendrograms and tree diagrams using "ggplot2". R package version 0.1–20; 2016.

73. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.

74. Wright ES. Using DECIPHER v2.0 to analyze big biological sequence data in R. R J. 2016;8:352–9. https://doi.org/10.32614/RJ-2016-025.

75. Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinform. 2018;35:526–8. https://doi.org/10.1093/bioinformatics/bty633.

76. Kassambara A, Mundt F. factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.6; 2019.

77. Wickham H, François R, Henry L, Müller K. dplyr: A grammar of data manipulation. R package version 0.8.5; 2020.

78. Klopfenstein DV, Zhang L, Pedersen B, Ramírez F, Vesztrocy AW, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M, Dampier W, Dessimoz C, Flick P, Tang H. GOATOOLS: a Python library for gene ontology analyses. Sci Rep. 2018;8:10872. https://doi.org/10.1038/s41598-018-28948-z.

79. Kanehisa M, Sato Y. KEGG mapper for inferring cellular functions from protein sequences. Protein Sci. 2020;29:28–35. https://doi.org/10.1002/pro.3711.

80. Leinonen R, Sugawara H, Shumway M. International nucleotide sequence database collaboration. The sequence read archive. Nuc Acid Res. 2011;39:D19–21. https://doi.org/10.1093/nar/gkq1019.

81. Seo H, Kang S, Park YS, Yun CW. The role of zinc in gliotoxin biosynthesis of *Aspergillus fumigatus*. Int J Mol Sci. 2019;20:6192. https://doi.org/10.3390/ijms20246192.

82. Nierman WC, Pain A, Anderson MJ, et al. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature. 2006;438:1151–6. https://doi.org/10.1038/nature04332.

83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1999;215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.