The master regulator MAT1-1-1 of fungal mating binds to its targets via a conserved motif in the human pathogen *Aspergillus fumigatus*

Barbara Ramšak,[1] Jessica Markau,[1] Tobias Pazen,[2] Tim A. Dahlmann [ID],[1] Sven Krappmann,[2] and Ulrich Kück [ID] [1],*

[1]Allgemeine und Molekulare Botanik, Ruhr-Universität Bochum, 44780 Bochum, Germany
[2]Mikrobiologisches Institut – Klinische Mikrobiologie, Immunologie und Hygiene, Universitätsklinikum Erlangen und Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg, 91054 Erlangen, Germany

*Corresponding author: Allgemeine und Molekulare Botanik, Universitätsstraße 150, Ruhr-University Bochum, D-44801 Bochum, Germany. ulrich.kueck@rub.de

Abstract

Mating-type transcription factors are master regulators of sexually related signal transduction pathways in fungi; however, their recognition of specific DNA sequences from target genes is widely undetermined. Here, we identified and characterized the DNA-binding sequence of the MAT1-1-1 alpha-box domain transcription factor from the human pathogen *Aspergillus fumigatus*. In order to explore MAT1-1-1 DNA-binding targets, we used the previously reported MAT1-1-1 binding motif from *Penicillium chrysogenum*, in a bioinformatics approach. We identified 18 *A. fumigatus* genes carrying the MAT1.1 sequence in their upstream region, among them genes for the α-pheromone precursor (PpgA), G-protein-coupled pheromone receptor (PreA), and for TomA, an unidentified protein. To validate our prediction further, quantification of transcript levels showed a decrease in expression of *ppgA*, *tomA*, and others in a *MAT1-1* deletion strain. For a functional analysis of the binding sites, truncated variants of the *A. fumigatus MAT1-1-1* gene were introduced into *Escherichia coli* for heterologous expression. The yield of recombinant protein was further optimized for the AfMAT1-1-1$_{78-235}$ variant that harbors an extended alpha-box domain. AfMAT1-1-1$_{78-235}$ bound to a subset of the most strongly upregulated genes: *ppgA*, *preA*, and *tomA*. The DNA-binding specificity was confirmed by testing mutated binding sequences, as well as performing competition experiments with specific and non-specific sequences. Finally, equilibrium dissociation constants of 1.83 ± 0.1 and $1.45 \pm 0.26\,\mu M$ were determined for AfMAT1-1-1$_{78-235}$ and fusion protein GST-AfMAT1-1-1$_{78-235}$. Collectively, these findings provide further insights into AfMAT1-1-1-mediated gene expression and imply that alpha-box domain regulators from other members of Eurotiales control fungal development in a conserved manner.

Keywords: mating type; transcription factor; DNA-binding motif; *Escherichia coli* protein synthesis; human pathogen *Aspergillus fumigatus*

Introduction

Sexual reproduction is a complex and ubiquitous process that occurs in eukaryotic organisms (Goodenough and Heitman 2014; Speijer *et al.* 2015). Despite the huge diversity of sexual programs found in fungi, mating-type (*MAT*) loci are a common feature for all fungal sections (Fraser and Heitman 2004; Idnurm *et al.* 2008; Coelho *et al.* 2017). The loci are called idiomorphs to indicate that they do not represent alleles of a single gene, despite their identical genomic localization (Metzenberg and Glass 1990). In Euascomycetes, the number of *MAT* genes varies and they usually code for transcription factors (Debuchy and Turgeon 2006). While members of the Sordariomycetes have rather complex *MAT* loci carrying several genes, *MAT* loci in members of the Eurotiales are simply organized and often contain a single functional reading frame for either a high mobility group (HMG) or an alpha-box DNA-binding domain transcription factor (Pöggeler 2007; Klix *et al.* 2010; Debuchy *et al.* 2014; Dyer and Kück 2017).

Belonging to the group of Eurotiales, the filamentous ascomycete *Aspergillus fumigatus* is a human pathogenic mold, mostly associated with invasive aspergillosis in immunocompromised patients (Fraser 1993; Hohl and Feldmesser 2007; Latgé and Chamilos 2019). *Aspergillus fumigatus* was considered to be an obligate asexual organism for many years, until genomic analysis revealed the presence of genes essential for sexual reproduction (Pöggeler 2002; Varga 2003; Paoletti *et al.* 2005). *Aspergillus fumigatus* is considered to be a heterothallic (self-sterile) fungus, therefore the mating process starts with the recognition between two compatible strains carrying *MAT1-1* and *MAT1-2* idiomorphs (Paoletti *et al.* 2007; O'Gorman *et al.* 2009). Recognition is initiated by α-type (PpgA) or a-type pheromones binding to the cognate G-protein-coupled pheromone receptors (PreA and PreB). Signal transduction activates a highly conserved mitogen-activated protein kinase pathway which triggers the transcriptional activation of proteins required for the cellular response (Elion 2000; Pöggeler 2002; Dyer *et al.* 2003; Seo *et al.* 2004). The expression of an α-type

pheromone precursor PpgA in *A. fumigatus* is under the transcriptional control of MAT proteins (Szewczyk and Krappmann 2010; Yu *et al.* 2018). Recently, the *MAT1-2* idiomorph was found to encode an additional transcript, *MAT1-2-4*, whose deletion resulted in impaired fruiting body formation (Yu *et al.* 2017). The complete and functional sexual cycle with formation of fruiting bodies (cleistothecia) and sexual spores can be induced under laboratory conditions (O'Gorman *et al.* 2009; Szewczyk and Krappmann 2010; Sugui *et al.* 2011; Ashton and Dyer 2019). Thus, *A. fumigatus* proved to be a suitable model for investigating a reproduction system requiring two compatible partners for sexual mating in Euascomycetes.

MAT proteins from filamentous ascomycetes have been demonstrated to function beyond their previously understood role in sexual development (Wada *et al.* 2012; Böhm *et al.* 2013; Becker *et al.* 2015; Kim *et al.* 2015; Mahmoudjanlou *et al.* 2020). Notably, despite their different lifestyles, MAT1-1-1 orthologs from *A. fumigatus* and *Penicillium chrysogenum* have also been shown to regulate a similar number of transcriptional targets; however, among those target genes, most of them were obviously not directly involved in mating (Becker *et al.* 2015; Yu *et al.* 2018). For instance, it was shown that both transcription factors have an impact on secondary metabolite production (Böhm *et al.* 2013; Yu *et al.* 2018). Furthermore, population studies of *A. fumigatus* strains revealed an association between *MAT1-1* isolates and increased virulence of those isolates (Alvarez-Perez *et al.* 2009; Cheema and Christians 2011). However in another study, a pair of isogenic strains with opposite mating-type loci showed similar pathogenicity (Losada *et al.* 2015). The expression of genes for several secondary metabolites is mating-type dependent in *A. fumigatus* (Yu *et al.* 2018), thus sexual identity might have a distinct, albeit indirect effect on virulence. Epipolythiodioxopiperazine gliotoxin serves as a prominent example of a secondary metabolite that affects the virulence of *A. fumigatus* in a context-dependent manner (Kwon-Chung and Sugui 2009). In contrast, others, such as pseurotin or fumagillin, whose expression is specifically regulated by *MAT1*-encoded transcriptional regulators, have so far shown no contribution to pathogenicity. Further evidence for an overlap of fungal secondary metabolism and virulence arises from studies on the global regulator LaeA (Sugui *et al.* 2007).

Although MAT transcription factors have been functionally explored over the past two decades, relatively few studies have answered the question of how MAT transcription factors regulate expression of their downstream genomic targets, thus the protein–DNA interactions of MAT transcription factors remain poorly understood. Recently, a MAT1-1-driven transcriptome analysis in *P. chrysogenum* and *A. fumigatus* demonstrated that the MAT1-1-1 orthologs from these two fungi share a significant number of common target genes (Böhm *et al.* 2013; Becker *et al.* 2015; Yu *et al.* 2018). However, the specific DNA-binding sites of the *A. fumigatus* MAT1-1-1 transcription factor remain to be determined.

The alpha-box domains and HMG domains of MAT proteins from ascomycetes exhibit similar functional and structural residues, therefore they might bind to DNA in a similar fashion (Jackson *et al.* 2013). Since eukaryotic HMG domains were shown to be sufficient for DNA binding (Schlierf 2002; Jauch *et al.* 2012), we hypothesized that the region of AfMAT1-1-1 encoding an alpha-box DNA-binding domain would be essential for the site-specific interaction with DNA.

In this study, we performed a detailed analysis of the AfMAT1-1-1 binding to the highly conserved DNA-binding motif which is enriched in genes that were shown to be regulated by AfMAT1-1-1 (Yu *et al.* 2018) or are known to be involved in sexual reproduction in aspergilli (Dyer and O'Gorman 2012). We demonstrate that the extended alpha-box domain of AfMAT1-1-1 is sufficient for the sequence-specific DNA recognition. Furthermore, the regulatory binding sites of MAT1-1-1 orthologs from *A. fumigatus* and *P. chrysogenum* are highly conserved, suggesting common mechanisms for MAT driven gene regulation within members of the Eurotiales.

Materials and methods
Transcriptional profiling by RT-qPCR
Transcript levels of selected candidates were quantified in the MAT1-1 wild-type isolate D141 and its congenic mat1-1Δ deletion mutant (Szewczyk and Krappmann 2010) by real-time quantitative PCR (RT-qPCR). Total RNA serving as input was extracted from ground mycelia of three biological replicates using TRI Reagent (Sigma-Aldrich, St. Louis, Missouri) and peqGOLD Phase Trap reaction tubes (PEQLAB, Erlangen, Germany). Adequate quality was checked on a NanoDrop UV-spectrophotometer to further analyze only samples showing A260/A280 values greater than 1.8. Purification including DNAse digestion was then performed with filters of the innuPREP Plant RNA Kit (Analytic Jena, Jena, Germany). Reverse transcription was carried out with the High Capacity cDNA Reverse Transcription Kit from Applied Biosystems without RNAse inhibitor using 1 μg RNA as input for reverse transcription by the MultiScribeTM reverse transcriptase and following the recommended incubation protocol of 10 min at 25°C, 120 min at 37°C, 5 min at 85°C, and storage at 4°C. The resulting cDNA samples served as qPCR templates in a ViiA 7 real-time PCR system equipped with a 384-well block module (Applied Biosystems, Foster City, CA, USA). All reactions were performed in MicroAmp optical 384-well reaction plates covered with qPCR adhesive seal sheets (4titude) in a standard volume of 8 μl for each individual reaction containing 25 ng cDNA as template, primers at 800 nM concentration each, and 4 μl of the SYBRTM Select Master Mix. Pairs of oligonucleotides used for quantification of ppgA (AFUB_072280), tomA (AFUB_063230), cp9 (AFUB_064360), and AFUB_070880 transcript levels are presented in Supplementary Table S1. Reactions were performed in technical triplicates, and equilibrium dissociation curves were plotted to determine specificity of the PCR runs. Aliquots of purified total RNA served as "no reverse transcriptase" negative control to assess the amount of DNA contamination and did generally not result in a PCR product. For the analysis of RT-qPCR results, the QuantStudio software (Version 1.3; Applied Biosystems) was used. Quantitative analyses were performed according to the comparative threshold cycle method ($2^{-\Delta\Delta Ct}$ method; Livak and Schmittgen 2001), using transcript levels of the β tubulin-encoding gene tubA (AFUA_1G10910) as established internal reference.

Sequence comparison
Identification of orthologs for all MAT1-1-1 target genes reported for *P. chrysogenum* (Becker *et al.* 2015) in *A. fumigatus* was performed by a reciprocal BLASTP analysis of the corresponding amino acid sequences using the publicly available genome annotations of *P. chrysogenum* Wisconsin 54-1255 (van den Berg *et al.* 2008), *A. fumigatus* A1163 (Fedorova *et al.* 2008), and *A. fumigatus* Af293 (Nierman *et al.* 2005), as described previously (Altschul 1997; Traeger *et al.* 2013). The orthologs of *P. chrysogenum* MAT1-1-1 targets in *A. fumigatus* were compared to the recently

published list of 59 upregulated genes in MAT1-1-1 overexpression strain MAT1-1 OE in the *A. fumigatus* D141 background (Yu *et al.* 2018). Selected matching hits were used for sequence motif analyses.

In silico sequence motif analysis

Sequence motif analyses were performed by using the programs MEME (Bailey and Elkan 1995) and FIMO (Grant *et al.* 2011), both parts of the MEME Suite (Bailey *et al.* 2009). Therefore, the 1000 nt regions upstream of all *A. fumigatus* A1163 coding sequences were determined by custom made Perl scripts and were considered as putative promoter sequences. Identification of putative *A. fumigatus* MAT1-1-1 (AfMAT1-1-1) binding sites within the promoter sequences was performed using the previously reported *P. chrysogenum* MAT1-1-1 (PcMAT1-1-1) binding motif MAT1.1 (Becker *et al.* 2015) as the input for FIMO (ρ-value < 0.0001). For a *de novo* motif search, we used MEME with the zoops (zero or one occurrence per sequence) option, both in the classical and differential enrichment mode.

Cloning and protein expression

The full-length ORF of *A. fumigatus* MAT1-1-1 was amplified by PCR from a cDNA clone using gene-specific primers AfMAT1-1_f and AfMAT1-1_r (Supplementary Table S1), which harbored *Nco*I and *Eco*RI restriction sites, respectively. The PCR product was purified and cloned in-frame into the bacterial expression vector pGEX-4T-1-TEV, which had been previously digested with *Nco*I and *Eco*RI, resulting in the pGEX-AfMAT1-1-1 expression plasmid (Supplementary Figure S2). Truncated variants of the AfMAT1-1-1 (amino acid 1–197, 1–235, 1–254, 78–197, 78–235, 78–254, and 78–368) were amplified by PCR from a cDNA clone using DNA primers: AfMAT1-1_f, AfMAT1-1_r, AfMAT1-1_F4_r, AfMAT1-1_F6_r, AfMAT1-1_d_f, and AfMAT1-1_d_r (Supplementary Table S1). Respective PCR products were integrated into the expression vector pGEX-4T-1-TEV using the described cloning procedure, resulting in plasmids: pGEX-AfMAT1-1-1_F1, pGEX-AfMAT1-1-1_F2, AfMAT1-1-1_F3, AfMAT1-1-1_F4, AfMAT1-1-1_F5, AfMAT1-1-1_F6, and AfMAT1-1-1_F7 (Supplementary Table S2). The integrity of the protein constructs was verified by Sanger sequencing (Eurofins Genomics, Ebersberg, Germany) using primer pair pGex_f and pGex_r.

Protein expression analysis

The expression of the recombinant full-length AfMAT1-1-1 and its truncated plasmid constructs (Supplementary Table S2) were transformed into *Escherichia coli* BL21 (DE3) plysS competent cells (Invitrogen, Carlsbad, CA, USA). After transformation, optimal growth conditions were tested and are specified in Supplementary File S1.

High-yield purification of the recombinant AfMAT1-1-1$_{78–235}$, protein quantification, and western blotting

For high-yield purification of GST-AfMAT1-1-1$_{78–235}$, we followed in principle the protocol by Sommerkamp *et al.* (2019), with some major modifications, as outlined in Supplementary File S2. Protein quantification and western blotting were modified according to published methods (Becker *et al.* 2015) and are specified in Supplementary File S3.

Oligonucleotides used in electrophoretic mobility shift assays

In total, 1000 nt upstream regions of the genes *ppgA*, *preA*, and *tomA* were used to design 29-bp long oligonucleotide DNA probes for electrophoretic mobility shift assays (EMSAs). The annealing of oligonucleotides was performed by combining equimolar amounts of complementary DNA as described in Becker *et al.* (2015). After annealing, double-stranded (ds) oligonucleotides were labeled using [γ^{32}P] ATP (6000 Ci/mmol; Hartmann Analytic, Braunschweig, Germany) and T4 polynucleotide kinase (Promega, Madison, WI, USA). Oligonucleotides (Supplementary Table S1) were purchased by Eurofins Genomics (Ebersberg, Germany).

EMSAs and shift-western analysis

EMSAs and shift-western analyses were performed as described by Becker *et al.* (2015) with the following minor exceptions. ^{32}P-labeled dsDNA oligonucleotides (∼100 cps) were incubated with varying protein concentrations in the presence of 8 μl binding buffer [125 mM phosphate buffer, 375 mM KCl, 0.25% bovine serum albumin (w/v), 0.25% Tween 20 (v/v), 12.5% glycerol, 12.5 ng poly (dI/dC)] (Thermo Scientific, Waltham, MA, USA) in a final volume of 20 μl at room temperature for 20 min. The same absolute amounts of the samples (20 μl) were resolved on 15% polyacrylamide gels at 4°C. The GST-AfMAT1-1-1$_{78–235}$ binding efficiency was determined in mutational analysis by measuring the signal intensities using ImageJ (Schneider *et al.* 2012). The DNA-binding affinity of GST-AfMAT1-1-1$_{78–235}$ was determined using the relation: Fraction$_{(bound)}$/Fraction$_{(bound + unbound)}$. Protein complexes with the mutated DNA samples were normalized to the complexes formed between protein and non-mutated DNA samples. Quantification of EMSA gels to determine the DNA-binding affinity of GST-AfMAT1-1-1$_{78–235}$ and AfMAT1-1-1$_{78–235}$ was carried out using ImageJ (Schneider *et al.* 2012).

Data availability

Analyzed data are provided in the attached supplementary files. Plasmids and sequences are available upon request.

Supplementary material is available at figshare DOI: https://doi.org/10.25387/g3.13122746.

Results
Bioinformatic analyses identify putative AfMAT1-1-1 binding sites within the upstream regulatory regions of target genes

Based on the *A. fumigatus* Af293 genome annotation (Nierman *et al.* 2005), 1000 nt regions upstream of translation initiation sequences of each *A. fumigatus* gene were used to find individual motifs of the *P. chrysogenum* MAT1-1-1 (PcMAT1-1-1) binding motif (MAT1.1) using FIMO, part of the MEME suite 5.1.1 (Bailey *et al.* 2009). We identified MAT1.1 in 1731 (14.2%) of 9630 upstream sequences of *A. fumigatus* Af293. We then narrowed our search by selecting sex and development related genes from three different data sets: 59 differentially overexpressed genes found in a transcriptome study from an *A. fumigatus* MAT1-1 overexpression strain (Yu *et al.* 2018), 81 genes know to be involved in sexual reproduction in other aspergilli (Dyer and O'Gorman 2012), and 243 genes previously identified to be targets of the *P. chrysogenum* MAT1-1-1 transcription factor (Becker *et al.* 2015). Using the intersection of the 1731 putative target genes and genes that were found in at least two of the three afore mentioned data sets, we

Table 1 MAT1.1 motif occurrence in the promotor regions of genes regulated by AFMAT1-1-1 or other sexually associated genes from *Aspergilli* species

ID	Description[a]	# of motifs	Distance to TIS[b]	DNA strand	Sequence (5′→3′)	Motif ρ-value	ΔMAT-1-1 DE (qRT-PCR, $2^{-\Delta\Delta Ct}$)[d]	MAT-1-1 OE (RNA-seq, $\log_2$)
AFUB_072280	mating alpha-pheromone PpgA	1	−315	+	TTATTGAG	6.14E-05	2003.3	8.11
AFUB_063230	conserved hypothetical protein TomA	2	−380, −406	+ −	TTATTGAG CTATTGAT	3.96E-05	268.8	7.97
AFUB_064360	serine carboxypeptidase, putative Cp9	2	−238, −254	−	TTATTGAG	1.79E-05	255.1	5.67
AFUB_055410	a-pheromone receptor PreA	2	−244, −519	− −	TTATTGAG CTATTGAG	1.79E-05	nd	5.41
AFUB_070880	conserved hypothetical protein	2	−248, −457	+ −	CTATTGAT CTATTGAT	3.96E-05	286.4	5.13
AFUB_016640	conidial hydrophobin RodB	1	−392	−	TTATTGAG	6.14E-05	nd	4.35
AFUB_094560	DNA methyltransferase-like DmtA	1	−867	+	TTATTGAG	6.14E-05	nd	1.91
AFUB_032300	protein kinase, putative	1	−103	+	CTATTGAC	7.93E-05	nd	1.76
AFUB_021310	maleylacetoacetate isomerase MaiA	1	−393	+	CTATTGAT	1.79E-05	nd	1.74
AFUB_007030	conserved hypothetical protein	1	−609	+	CTATTGAT	3.96E-05	nd	1.69
AFUB_078190	conserved hypothetical protein	1	−870	−	CTATTGAT	3.96E-05	nd	1.64
AFUB_021320	transcription factor (zinc-finger) SilA	1	−687	+	CTATTGAG	1.79E-05	nd	1.61
AFUB_070440	oxidoreductase, putative	1	−42	−	CTATTGAG	1.79E-05	nd	1.60
AFUB_099080	ferric-chelate reductase, putative	2	−328, −436	−	CTATTGAC	7.93E-05	nd	1.59
AFUB_099650[c]	NO response (flavohaemoglobin) FhbA	1	−878	+	CAATTGAG	9.72E-05	nd	−1.0
AFUB_043130[c]	MAP kinase, kinase STE7	1	−649	+	CTATTGAT	3.96E-05	nd	0.24
AFUB_020110[c]	transcription factor (forkhead) FhpA	1	−732	+	CTATTGAT	3.96E-05	nd	−1.08
AFUB_059790[c]	phosducin-like protein PhnA	1	−454	−	CAATTGAG	9.72E-05	nd	−0.1

[a] As obtained from the AspGD website: http://www.aspergillusgenome.org/.
[b] TIS, translational initiation site.
[c] *Aspergillus nidulans* orthologs, associated with sexual reproduction in *A. nidulans* (Dyer and O'Gorman 2012).
[d] *n*-Fold transcript levels in *Aspergillus fumigatus* MAT1-1-1 wild-type isolate in comparison to its congenic *mat1-1-1Δ* deletion strain.

identified the 18 genes that were viewed as the most promising candidates for AfMAT1-1-1 binding (Table 1).

Among the genes identified, we found sexual development related genes, such as *ppgA* (AFUB_072280) and *preA* (AFUB_055410), encoding a sex pheromone and a pheromone receptor, and *cp9* (AFUB_064360), encoding a putative serine carboxypeptidase. The fourth gene, *tomA* (AFUB_063230) encodes a putative protein of unknown function, which has orthologs in different members of the Eurotiomycetes (Supplementary Figure S1).

Other potential targets of AfMAT1-1-1 required for functional sexual development include *DmtA*, which encodes DNA methyltransferase A involved in fruiting body development (Lee *et al.* 2008); *SilA* and *FhbA* as repressors of sex (Han *et al.* 2008; Baidya *et al.* 2011; Dyer and O'Gorman 2012); *PhnA* encoding a phosducin-like protein required for G-protein signaling (Seo *et al.* 2004); *Ste7*, a MAP kinase kinase that acts as an activator of pheromone transduction pathway (Bayram *et al.* 2012); and the transcription factor *FhpA* required for cleistothecia production (Dyer and O'Gorman 2012) (Table 1).

To validate the mating-type locus-controlled transcriptional expression of selected candidates, transcript levels of *ppgA*, *tomA*, *cp9*, and *AFUB_070880* were quantified in *MAT1-1* wild-type and *mat1-1Δ* deletion strains, as described in the *Materials and methods* section. RT-qPCR data confirmed the AfMAT1-1-1-dependent transcription of these genes (Figure 1A). Based on this, and since the upstream regions of *ppgA*, *tomA*, and *preA* showed at least one putative binding site match with the PcMAT1-1-1 binding motif (Figure 1B), we analyzed the interaction between these exemplary promoters and the AfMAT1-1-1 transcription factor in more detail.

Cloning and expression of variants of the AfMAT1-1-1 protein

To determine whether AfMAT1-1-1 binds to the discovered binding sites in the upstream regulatory regions of *ppgA*, *preA*, and *tomA*, we set out to generate soluble AfMAT1-1-1 recombinant protein for *in vitro* DNA-binding assays. The full-length *AfMAT1-1-1* cDNA was fused with a glutathione S-transferase (GST-tag) at the N-terminus (Supplementary Figure S2). The two protein moieties were linked by a 1.8 kDa polylinker carrying a tobacco etch virus (TEV) protease recognition site (Figure 2A). Although the GST-AfMAT1-1-1 protein with a predicted molecular weight of 68.6 kDa was expressed, we found that the recombinant protein was predominantly in the insoluble fraction, and western blot analysis indicated that its expression level was low (Figure 2B).

We used several approaches to improve expression rate and solubility of full-length AfMAT1-1-1. At first, we examined the *AfMAT1-1-1* sequence for the presence of codons that are rarely used in *E. coli*. Analysis of the native *AfMAT1-1-1* gene revealed eight rarely used codons in *E. coli* (Supplementary Figure S3). The corresponding triplets encode for amino acids arginine, proline, leucine and the "UGA" stop codon. Therefore, we expressed the full-length AfMAT1-1-1 and C-terminally shortened variant GST-AfMAT1-1-1$_{1-197}$ in *E. coli* strain Rosetta, which carries a plasmid encoding codons tRNAs for AGG, AGA, AUA, CUA, CCC, and CGA that are rarely used in *E. coli*. Heterologous expression of the gene variants in Rosetta cells did not lead to increased protein expression compared to *E. coli* BL21 pLysS cells (Supplementary Figure S4). A likely explanation for this observation is that other factors such as disordered regions present in the N- and C-terminus predominantly affect translational process. Based on obtained result, we did not decide for codon replacement strategy. Instead, we set out to construct several truncated variants of AfMAT1-1-1 (Figure 2A).

To avoid constructing non-functional truncated proteins, the alignment of amino acid sequences in the extended alpha-box domains was revised using JPred (Drozdetskiy *et al.* 2015) and IUPred2A (Mészáros *et al.* 2018) servers (Supplementary Figure S5). Secondary structure prediction of AfMAT1-1-1 suggested
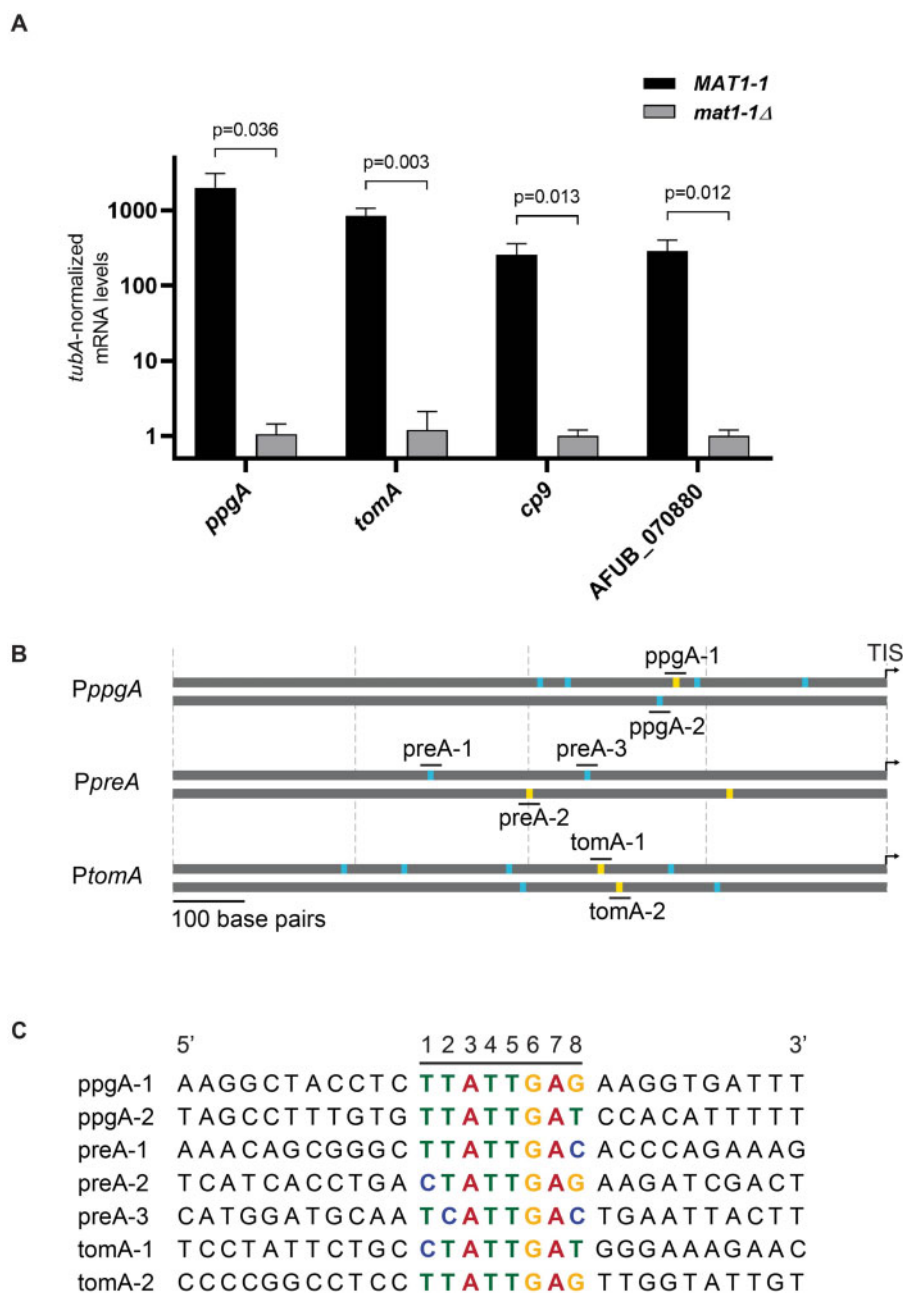
Figure 1 MAT1.1 binding motif from *Penicillium chrysogenum* is enriched in genes regulated by *Aspergillus fumigatus* MAT1-1-1. (A) Transcript levels, as monitored by qRT-PCR analyses ($n = 3$), of the presumably AfMAT-1-1-1 regulated genes *ppgA*, *tomA*, *cp9*, and *AFUB_070880* during prolonged cultivation in minimal medium identify their promoters as *bona fide* targets for the corresponding mating-type factor. (B) Positions of the MAT1.1 binding motif as identified by Becker *et al.* (2015), within the 1000 nt upstream regulatory regions of genes *ppgA*, *preA*, and *tomA*. The translation initiation site (TIS) indicates the translational start codon of the open reading frames. Nucleotide sequences used in EMSA experiments are marked. Putative AfMAT1-1-1 binding sites are indicated in different color, depending on the FIMO ρ-values for MAT1.1 (yellow: $\rho < 0.0001$; blue: $0.0001 \leq \rho < 0.001$). (C) The aligned sequences derived from the gene targets of AfMAT1-1-1 and were used in EMSAs. Nucleotides of the core consensus motif are indicated by colors. Numbering of the nucleotides corresponds to their position within the motif.

that the conserved alpha-box domain contains three helices. A fourth helix is predicted adjacent to the alpha-box (Supplementary Figure S5A). The alpha-box domain (amino acids 78–139) together with the downstream sequence extending to amino acid 184 form a so-called MATalpha_HMG-box (Pfam PF04769) domain, which was defined previously (Martin *et al.* 2010). The central region of AfMAT1-1-1 (amino acids 83–218), including the MATalpha_HMG-box domain, is predicted to be globular, whereas the N- and C-terminal parts of the protein appears to be naturally disordered (Supplementary Figure S5B).

Considering the secondary structure and globularity of AfMAT1-1-1, we designed a set of protein constructs with truncated N- and/or C-termini, which all contained the complete MATalpha_HMG-box domain (Figure 2A). We tested the variant constructs for optimal protein synthesis and binding activity toward tom1-2 and tomA-1 probes (Supplementary Figure S6). The DNA-binding activity of the variant $AfMAT1\text{-}1\text{-}1_{1-235}$, expressing the complete N-terminal part with DNA-binding domain was not significantly altered in comparison with variants AfMAT1-1-1_{78-235} and AfMAT1-1-1_{78-254} encoding only the region of
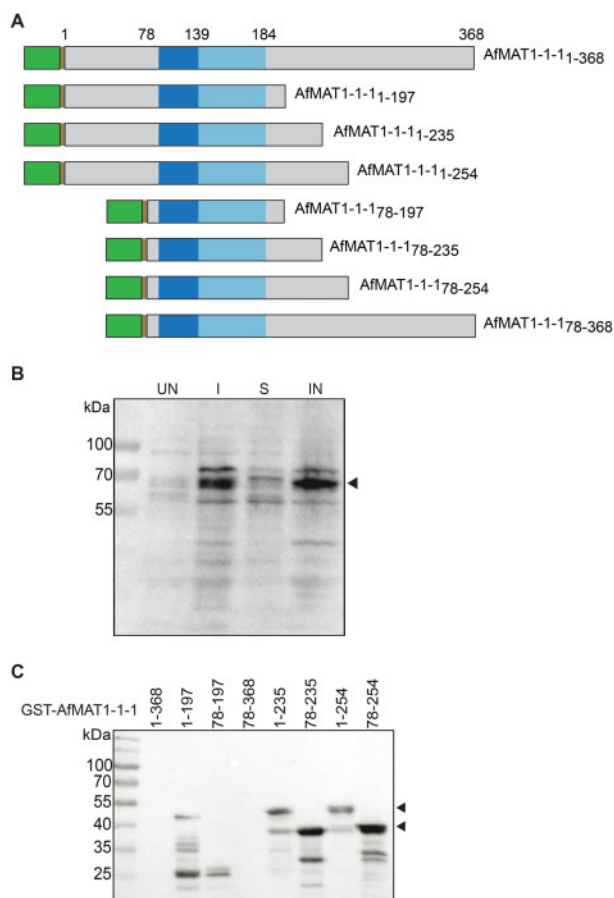
Figure 2 Overexpression of full-length and truncated recombinant AfMAT1-1-1 proteins. (A) Schematic representation of the full-length recombinant AfMAT1-1-1 and a set of AfMAT1-1-1 truncated protein constructs containing MATalpha_HMG (PF04769) domain (light blue) with a conserved N-terminal adjacent region, alpha-box domain (dark blue). Green box corresponds to the GST-tag with theoretical MW of 26 kDa and a brown box corresponds to polylinker containing recognition sequence for TEV protease. (B) Western blot analysis of full-length GST-AfMAT1-1-1 (theoretical MW = 66.8 kDa). The protein was mainly present in the insoluble fraction (IN). Lanes: uninduced (UN) and induced (I) fraction of *Escherichia coli* cells; soluble fraction (S) and insoluble (IN) protein fraction of the induced cell culture. (C) A small-scale expression experiment of truncated GST-AfMAT1-1-1 constructs. Soluble fractions of the proteins, which are schematically presented in (A), were analyzed by western blot assay using antibodies against GST. Black arrow heads indicate non-degraded proteins. For theoretical MWs, see Supplementary Table S3.

DNA-binding domain. Finally, we have chosen construct GST-AfMAT1-1-1$_{78–254}$ for a detailed biochemical analysis, since this is the smallest peptide and showed the best solubility and minimal protein degradation after purification from *E. coli* (Figure 2C).

AfMAT1-1-1$_{78–235}$ remains stable in solution after removal of the GST-tag

AfMAT1-1-1$_{78–254}$ contains the entire MATalpha_HMG-box domain with an additional 51 residues at the C-terminus (Figure 2A). Following high-yield purification (see Supplementary File S2), we noted that after SDS-PAGE analysis the final GST-AfMAT1-1-1$_{78–254}$ product appeared approximately 7 kDa smaller than the predicted 44.1 kDa (Figure 3, A and B). We subsequently removed the GST-tag, and after a second purification, GST-tag and GST-tagged TEV were separated from AfMAT1-1-1$_{78–254}$

(Figure 3C). After size-exclusion chromatography (Figure 3D), AfMAT1-1-1$_{78–254}$ eluted from a Superdex^{TM} 75 size-exclusion column as a rather symmetric peak corresponding to a molecular weight of about 28.4 kDa (Supplementary Figure S7). Since the theoretical molecular weight of AfMAT1-1-1$_{78–254}$ was calculated to be 18 kDa, the protein is likely to exist as a dimer in aqueous solution. SDS-PAGE analysis of eluted fractions indicated that protein migrates as a single band, confirming that the purified protein was nearly homogeneous (Figure 3D). Our yields of pure protein estimated from a Bradford protein assay were on average 30–35 mg per 4 l of bacterial culture.

In vitro binding specificity of GST-AfMAT1-1-1$_{78–235}$

We examined sequence-specific DNA binding of the GST-AfMAT1-1-1$_{78–254}$ using a radiolabeled 29-bp dsDNA probes in EMSA. The oligonucleotide sequences were based on the *ppgA*, *preA*, and *tomA* promoters and they all contained the predicted MAT1.1 motif (Figure 1C and Supplementary Table S1). GST-AfMAT1-1-1$_{78–235}$ bound to all the dsDNA oligonucleotides (Figure 4, A–C), including the dsDNA oligonucleotides ppg1-2 and tom1-2, originating from the *P. chrysogenum* promoters of *ppg1* and *tom1* genes (Figure 4, A and C). The ppg1-2 and tom1-2 were shown previously to be bound by PcMAT1-1-1 (Becker *et al.* 2015). In contrast, the dsDNA oligonucleotides derived from the *A. fumigatus* *ppgA* and *tomA* promoters were not bound by the PcMAT1-1-1 ortholog (Figure 4, A and C). Taken together, this analysis indicates that the protein region from 78 to 235 amino acids encoded by *AfMAT1-1-1*$_{78–235}$ is sufficient for forming DNA-protein complexes. Based on the dsDNA probes bound by GST-AfMAT1-1-1$_{78–235}$, we redefined the consensus binding motif as (T/C)(T/C)ATTGA(G/T/C) (Figure 4D). We conclude that the AfMAT1-1-1$_{78–235}$ peptide tolerates variation in the nucleotides flanking the core "ATTGA" of the binding motif.

To address the question of whether AfMAT1-1-1$_{78–235}$ acts in a sequence-specific manner, we introduced nucleotide substitutions within the AfMAT1-1-1$_{78–235}$ binding motif in dsDNA oligonucleotides ppgA-1 and tomA-1 (Figure 5A). Nucleotide A in position 3 of the consensus binding motif "TTATTGAG" (ppgA-1) and "CTATTGAT" (tomA-1) was substituted by a G, resulting dsDNA oligonucleotide probes ppgA_m1 and tomA-1_m1, and then further substitution of nucleotide G in position 6 to an A resulted in oligonucleotides ppgA-1_m2 and tomA-1_m2 (Supplementary Table S1). The binding efficiency of GST-AfMAT1-1-1$_{78–235}$ was quantified by EMSAs (Figure 5, A and B). The first single nucleotide substitution at position 3 in the motif significantly reduced binding of GST-AfMAT1-1-1$_{78–235}$. Binding decreased by 50% for tomA-1_m1 and 90% for ppgA-1_m1 compared to the wild-type sequences. Unexpectedly, the second nucleotide substitution in oligonucleotides ppgA-1_m2 and tomA-1_m2 did not further reduce binding of GST-AfMAT1-1-1$_{78–235}$. Thus a single nucleotide change within the AfMAT1-1-1 binding motif at position 3 significantly alters binding of GST-AfMAT1-1-1$_{78–235}$, indicating that binding of AfMAT1-1-1$_{78–235}$ to DNA is sequence specific (Figure 5B).

To confirm the observations above, we tested the DNA sequence specificity of the protein domain by performing a competition assay (Figure 5C). Non-radiolabeled ppgA-1 served as a specific competitor in EMSAs, preventing non-specific nucleotide-protein interactions, while non-radiolabeled dsDNA oligonucleotide ppgA-1_m2 was used to investigate the effect of mutations in the AfMAT1-1-1 binding site on protein binding. Incubating GST-AfMAT1-1-1$_{78–235}$ with increasing concentrations of specific competitor oligonucleotide ppgA-1 significantly decreased the signal intensity of the shifted bands (Figure 5C, left
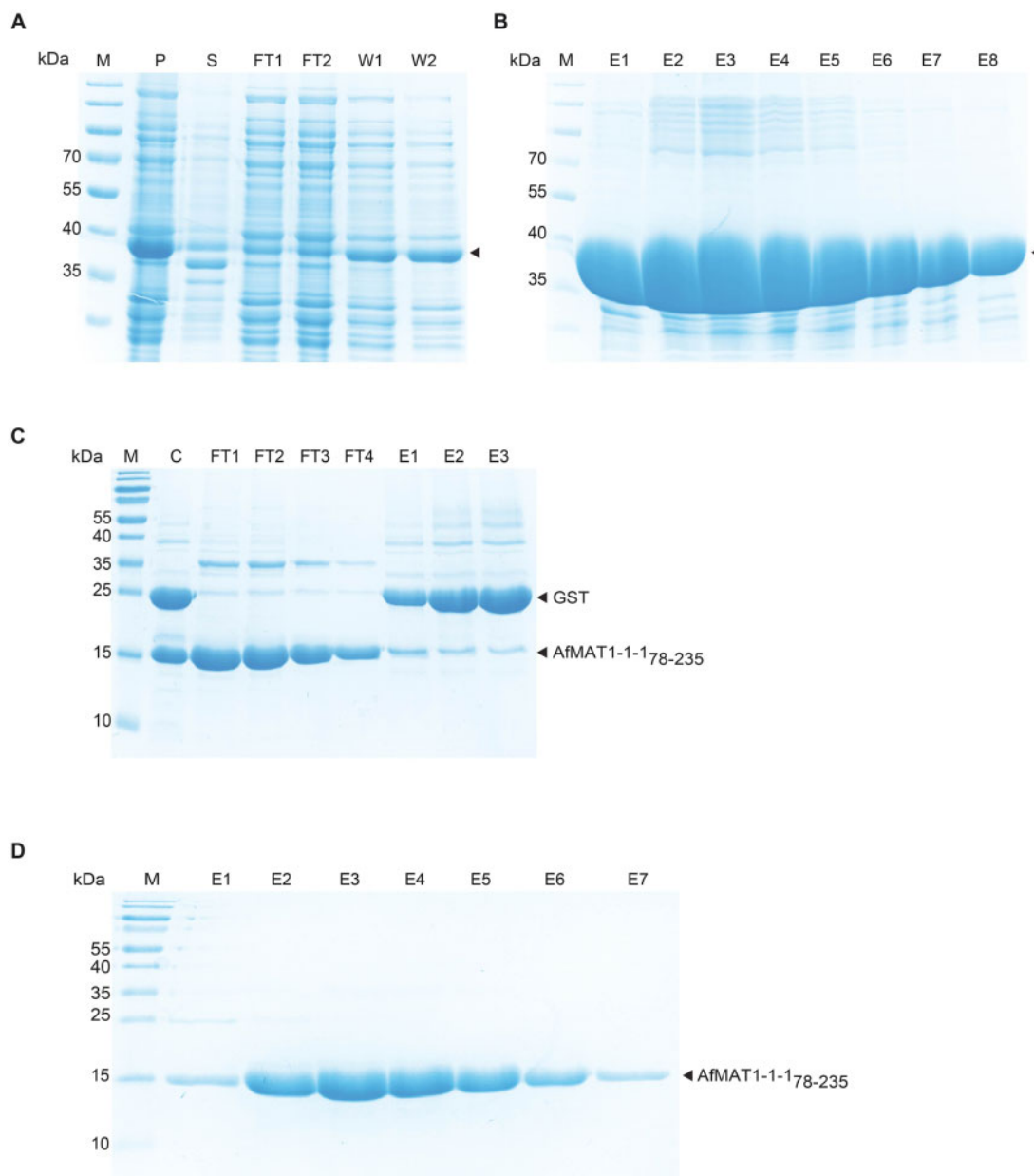
Figure 3 High-yield purification of the recombinant GST-tagged AfMAT1-1-1$_{78-235}$ from *Escherichia coli*. (A, B) GST-MAT1-1-1$_{78-235}$ was first purified using 20 ml GSTrap^{TM} (GE Healthcare) column, and the 2 ml fractions obtained throughout purification were analyzed by 10% SDS-PAGE. The lanes are labeled as follows: "M," molecular weight marker; "P," non-soluble fraction; "S," soluble fraction; "FT1–2," flow-through fractions containing unbound protein; "W1–2," wash fractions; "E1–8," eluted fractions (2 ml) after first affinity chromatography. Arrow heads indicate the GST-AfMAT1-1-1$_{78-235}$. (C) Fifteen percent SDS-PAGE analysis of the second glutathione affinity purification step. This time batch purification procedure was used to remove the cleaved GST-tag from AfMAT1-1-1$_{78-235}$ protein. Cleavage mixture was applied to the glutathione agarose beads and fractions were collected as follows: "M," molecular weight marker; "C," cleavage mixture; "FT1–4," flow-through fractions containing MAT1-1-1$_{78-235}$; "E1–3," fractions with GST-tag and GST-TEV protease. GST and AfMAT1-1-1$_{78-235}$ are indicated. (D) Protein fractions obtained after size-exclusion chromatography (SEC). Fractions "FT1–4" (see Figure 4D) were concentrated and applied to the size-exclusion Superdex^{TM} column. AfMAT1-1-1$_{78-235}$ was eluted from the column in 2 ml fractions labeled as "E1–E7." The protein concentrations obtained were 2.8 mg/ml ("E2"), 4.1 mg/ml ("E3"), 4 mg/ml ("E4"), and 2.7 mg/ml ("E5") as determined by Bradford protein assay. The arrow head indicates AfMAT1-1-1$_{78-235}$.

panel). However, increasing concentrations of the non-specific competitor ppgA-1_m2 only slightly decreased the signal intensity, thus confirming our results with the specific competitor. Western analysis using an anti-GST antibody further confirmed that oligonucleotide DNA binding is mediated by the GST-AfMAT1-1-1$_{78-235}$ protein (Figure 5C, right panel). In summary, our results show that binding of GST-AfMAT1-1-1$_{78-235}$ to ppgA-1 and tomA-1 target motifs is sequence specific. Additionally, the third A nucleotide of the consensus motif "(T/C)TATTGA(G/T)"

seems to be important for AfMAT1-1-1$_{78-235}$ sequence-specific DNA binding.

Next, we assessed the binding affinity of GST-AfMAT1-1-1$_{78-235}$ and GST-free AfMAT1-1-1$_{78-235}$ by determining the equilibrium dissociation constants (K_D). Increasing amounts of purified proteins were incubated with the fixed concentration of dsDNA oligonucleotide tomA-1 (Figure 6). Both proteins interacted with the target DNA, which decreased the amount of radiolabeled tomA-1 probe (Figure 6, A and B). The binding affinity of the proteins was
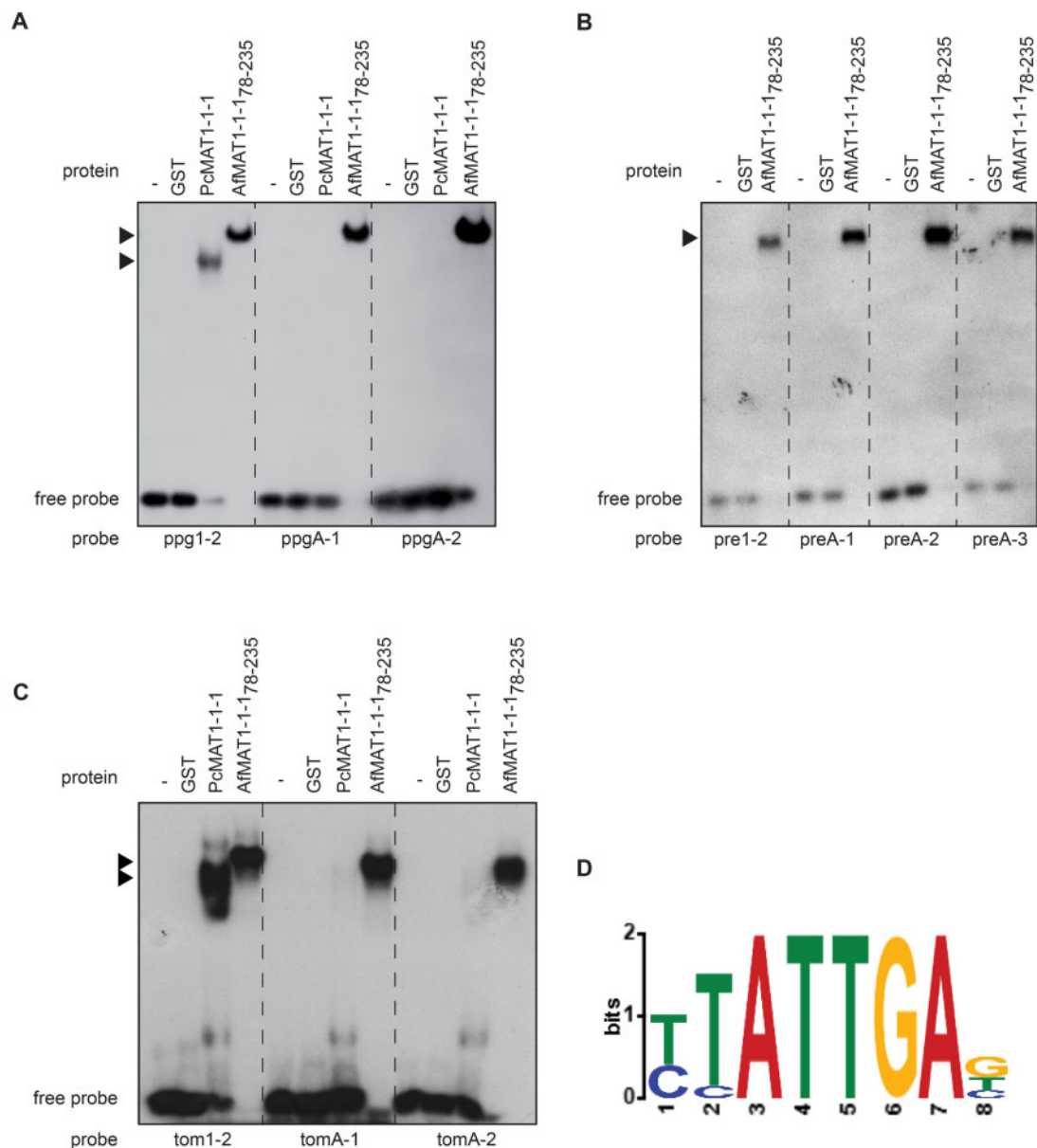
Figure 4 GST-AfMAT1-1-1$_{78-235}$ interacts with the identified DNA-binding motifs within the promoter regions of selected genes. (A) EMSA using GST-AfMAT1-1-1$_{78-235}$ and full-length PcMAT1-1-1 as an internal control. Radiolabeled ppg1-2 from *Penicillium chrysogenum*, and ppgA-1 and ppgA-2 from *Aspergillus fumigatus* were used. For both EMSAs, 4 µg (4.5 µM) of protein was mixed with the DNA. The formation of the protein–DNA complexes is indicated by arrow heads. (B) EMSA using purified GST-AfMAT1-1-1$_{78-235}$ and radiolabeled oligonucleotide probes, derived from the upstream region of the *pre1* and *preA*. Pre1-2 probe derived from *P. chrysogenum*, and preA-1, preA-2, and preA-3 from *A. fumigatus*. The lanes are labeled as follows: "–," dsDNA probe without protein; "GST," GST protein; "MAT1-1-1$_{78-235}$," GST-tagged AfMAT1-1-1$_{78-235}$. (C) EMSA using GST-AfMAT1-1-1$_{78-235}$ and full-length PcMAT1-1-1. Radiolabeled tom1-2 from *P. chrysogenum*; and tomA-1 and tomA-2 from *A. fumigatus* were used. For both EMSAs, 4 µg (4.5 µM) of protein was mixed with the labeled DNA. Arrow heads indicate formation of the protein–DNA complexes. (D) The position weight matrix (PWM) logo of the AfMAT1-1-1 DNA-binding motif generated from 10 experimentally verified transcription factors binding sites (see A–C) using the MEME suite (Bailey *et al.* 2009).

quantified using an established method (Heffler *et al.* 2012) and was further evaluated by nonlinear regression analysis. The K_D values of 1.45 ± 0.26 and $1.83 \pm 0.1\,\mu M$ were obtained for GST-AfMAT1-1-1$_{78-235}$ and AfMAT1-1-1$_{78-235}$. Since the K_D values of the proteins were very similar, DNA binding of AfMAT1-1-1$_{78-235}$ was not significantly altered by removal of the GST-tag.

Discussion

Here, we used a bioinformatics analysis, together with biochemical and DNA-binding studies to answer the question of whether

the AfMAT1-1-1 binding sites can be defined based on the known binding motif of the MAT1-1-1 protein from *P. chrysogenum*, a closely related member of Eurotiales. After considering the codon usage of the recombinant fungal gene and choosing different *E. coli* host strains, we found that a N-terminally tagged AfMAT1-1-1$_{78-235}$ construct, containing the MAT alpha-box domain, was best suited for our *in vitro* DNA-binding assays. The purified AfMAT1-1-1 construct interacts specifically with the promoter sequences from *ppgA*, *preA*, and *tomA*, which were shown to be strongly transcriptionally regulated in a *A. fumigatus* overexpression strain (Yu *et al.* 2018). From our result we conclude that the
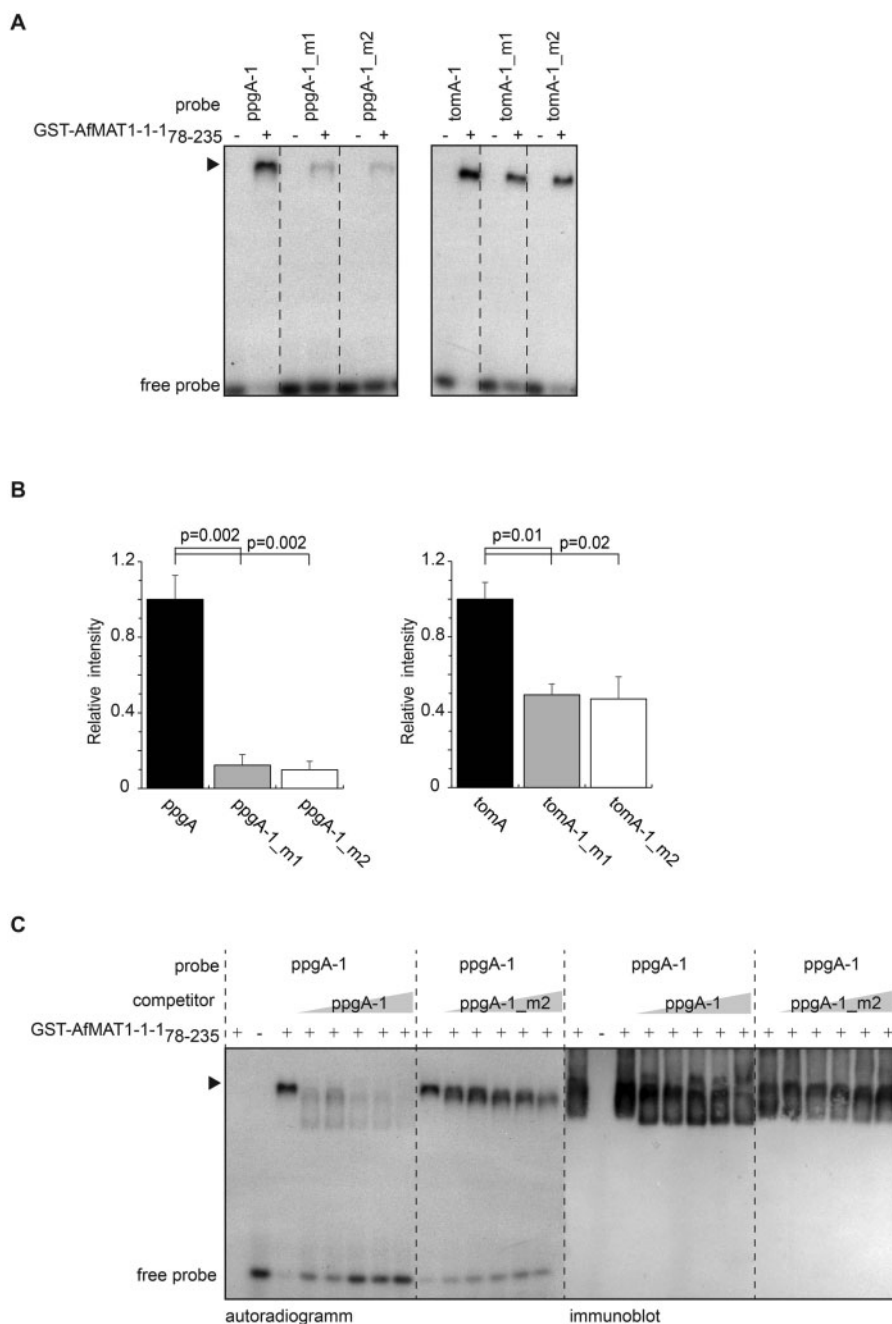
Figure 5 Single nucleotide substitution in the core binding motif decreases DNA binding of GST-AfMAT1-1-1$_{78-235}$ protein. (A) EMSA of the GST-AfMAT1-1-1$_{78-235}$ incubated with different radiolabeled dsDNA probes, as indicated. (B) Comparison of GST-AfMAT1-1-1$_{78-235}$ DNA-binding activity binding between the non-mutated and mutated DNA probes. Shown is the relative intensity determined from the autoradiograms as described in the *Materials and methods* section. Graphs show mean ± SD ($n = 3$). The ρ-values are indicated. Student's *t*-test. (C) Competition experiment, using radiolabeled ppgA-1 in the presence of unlabeled ppgA-1 (specific competitor) and unlabeled ppgA-1_m2 (non-specific competitor). Unlabeled probes with increasing concentration from 2.5 to 16.5 μM were incubated with the protein. Shift-Western experiment (GST-immunoblot) demonstrated the presence of GST-AfMAT1-1-1$_{78-235}$/DNA complexes. In total, 4 μg (=4.5 μM) of protein was used per each reaction. Arrow heads indicate formation of protein–DNA complexes.

N-terminal part of AfMAT1-1-1 has no direct impact on DNA-binding specificity. However, we cannot exclude, that the C-terminal part of AfMAT1-1-1 increases the DNA-binding specificity, since we were unable to express variant *AfMAT1-1-1$_{78-368}$*, which encodes the complete C-terminal part of the protein.

Sexual development in ascomycetes is controlled by mating-type locus encoded transcription factors that carry either a HMG or an alpha-box domain. While HMG domains are ubiquitous and their protein–DNA interactions are molecularly well defined

(Philley and Staben 1994; Štros *et al.* 2007; Ait Benkhali *et al.* 2013), alpha-box domain transcription factors have a rather limited distribution and are less well studied. The alpha-box domain orthologs from ascomycetes are very similar; however comparative studies of alpha-box domain proteins from the two protoascomycetes *Saccharomyces cerevisiae* and *Candida albicans* showed that the DNA-binding specificity of the *C. albicans* Matα1 changed after the divergence of *S. cerevisiae* and *C. albicans* (Carr 2004; Tuch *et al.* 2008; Baker *et al.* 2011). In *S. cerevisiae*, the transcription of
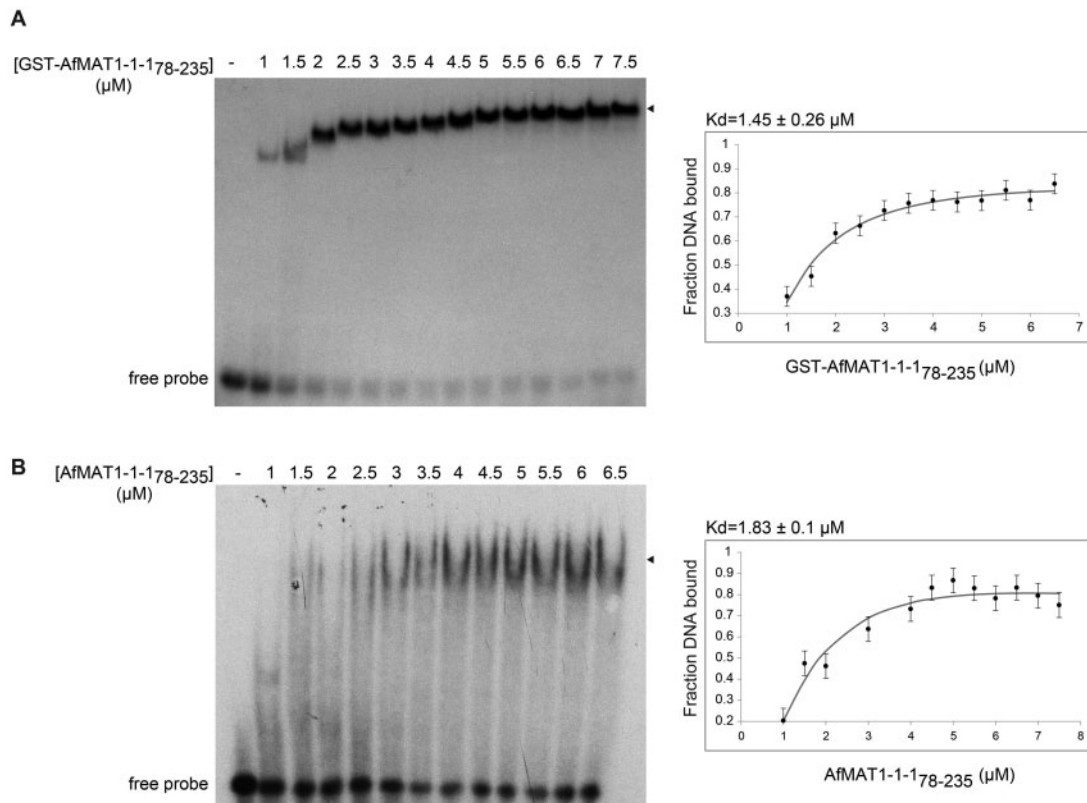
Figure 6 Determination of the equilibrium dissociation constants for GST-AfMAT1-1-1$_{78-235}$ and AfMAT1-1-1$_{78-235}$. (A, B left panels) EMSA titrations of the purified proteins with the tomA-1. Increasing concentrations of proteins GST-AfMAT1-1-1$_{78-235}$ (1–7.5 µM) and AfMAT1-1-1$_{78-235}$ (1–6.5 µM) were incubated with the radiolabeled tomA-1 dsDNA oligonucleotide. Protein/DNA complexes were separated through a native polyacrylamide gel, and the radiation signals from each band were obtained from the EMSA profiles. The intensity of the signals was quantified using ImageJ (Schneider *et al.* 2012). Arrow heads indicate the position of the protein–DNA complexes. (A, B right panels) Determination of the equilibrium dissociation constant (K_D) for the GST-AfMAT1-1-1$_{78-235}$ and AfMAT1-1-1$_{78-235}$. The fractions of bound and free DNA from A (for GST-AfMAT1-1-1$_{78-235}$) and B (for AfMAT1-1-1$_{78-235}$) were plotted as a function of the protein concentration and fitted with nonlinear regression analysis. The averaged K_D values from three assays with the standard error for each protein, GST-AfMAT1-1-1$_{78-235}$ and AfMAT1-1-1$_{78-235}$, are shown in the plot.

α-specific genes required for mating-type specification is activated by the alpha-box domain transcription factor MATα1. However, the transcriptional activation of the MATα1 target genes in *S. cerevisae* and *C. albicans* is only achieved when the MADS-box transcription factor Mcm1 forms a complex with MATα1 (Carr 2004). This probably explains a faster divergence of alpha1 DNA-binding specificity in these two yeasts compared to the two euascomycete compared in this study. In *P. chrysogenum*, the full-length alpha-box domain transcription factor MAT1-1-1 itself is sufficient for sequence-specific DNA binding (Becker *et al.* 2015). The fact that *A. fumigatus* and *P. chrysogenum* are phylogenetically closely related and that MAT1-1-1 orthologs from these two species share some common target genes explains why their binding motifs are at least partially conserved (Becker *et al.* 2015; Yu *et al.* 2018).

We succeeded in purifying a truncated version of AfMAT1-1-1, where the C-terminal 133 amino acids residues were deleted. This modification improved significantly the solubility of the construct compared to the full-length AfMAT1-1-1 protein. We speculate that the presence of a long intrinsically disordered region in the C-terminus (amino acids 268–294) of AfMAT1-1-1 could cause low solubility of the protein in aqueous solutions (Supplementary Figure S3B). This notion is consistent with studies where long intrinsically disordered regions were shown to be abundant in eukaryotic transcription factors and proved to be prone to aggregation (Ward *et al.* 2004; Uemura *et al.* 2018).

Using the most soluble protein construct, we show that the extended alpha-box domain of AfMAT1-1-1 interacts specifically with the promoter sequences from *ppgA*, *preA*, and *tomA* without the additional protein cofactor *in vitro*. Thus, confirming that the main components of the pheromone response pathway *ppgA* and *preA* are direct targets of the *A. fumigatus* MAT1-1-1 protein. This is consistent with data obtained for Matα1 transcription factor from *S. cerevisiae* (Galgoczy *et al.* 2004) or *P. chrysogenum* (Becker *et al.* 2015). Interestingly, we detected a conserved target of MAT1-1-1, *tomA*, which was previously discovered in two independent investigations (Becker *et al.* 2015; Yu *et al.* 2018). Here, *tomA* was one of the top targets of AfMAT1-1-1. In *P. chrysogenum*, the *tomA* ortholog (*Pc20g00090*) showed the highest statistically significant peak value in a PcMAT1-1-1 ChIP-seq analysis (Becker *et al.* 2015), while in *A. fumigatus*, *tomA* showed the highest differential expression in a comparative transcriptome analysis between a MAT1-1-1 overexpression strain and the wild type (Yu *et al.* 2018). TomA orthologs are generally present in members of the Eurotiomycetidae, a subclass of the ascomycetes, which besides *Aspergillus* and *Penicillium* species also includes the human pathogens *Histoplasma capsulatum*, *Coccidioides* spp, and *Blastomyces dermatitidis*. The conserved putative NLS "RKRRA" was identified at the N-terminus of TomA orthologs, suggesting TomA is a nuclear protein (Supplementary Figure S1). Other genomic targets, such as conidial hydrophobin *RodB*, DNA methyltransferase-like *DmtA*, and transcription factor *SilA*, also

appear to be direct targets of AfMAT1-1-1, which is in line with findings that MAT1-1-1 proteins have highly diverse functions in developmental processes. Future ChIP-seq experiments will be necessary to identify all target genes of MAT1-1-1 transcription factors *in vivo*.

MAT1-1-1 transcription factors from *A. fumigatus* and *P. chrysogenum* share reasonable similarity, with 36% sequence identity, whereas the alpha-boxes themselves have 58% amino acids in common (Supplementary Figure S5). Thus, the alpha-box domains are the most conserved region within MAT1-1-1 orthologs from diverse ascomycetes (Martin *et al.* 2010), yet any conservation of regions extending outside these domains is extremely limited. A previous study demonstrated that the transcription factor DNA-binding preference can be predicted on the basis of amino acid sequence similarity within other closely related members of the same transcription factor family (Alleyne *et al.* 2009). For example, cross-species conservation of regulatory regions of transcription factors was shown for human and mouse HMG domains from Sox proteins (Kamachi and Kondoh 2013). Moreover, preservation of *cis*-regulatory elements in orthologous promoter regions was observed in closely related ascomycetes (Gasch *et al.* 2004). The binding motifs of MAT1-1-1 proteins from *A. fumigatus* and *P. chrysogenum* are highly similar. The core "$A_3T_4T_5G_6A$" sequence is present in both; however, variations of nucleotides occur outside the central consensus motif. Furthermore, the same core sequence is also recognized by the MATα1 in *S. cerevisiae* (Baker *et al.* 2011). A single base pair substitution of the first A_3 to a G within the AfMAT1-1-1 core consensus motif in ppgA-1 and tomA-1 oligonucleotides led to significantly decreased DNA binding, indicating the importance of this nucleotide for the DNA-binding activity of the protein. Surprisingly, an additional nucleotide substitution (G_6 &cenveo_unknown_entity_wingdings3_F022; A) did not further alter the binding of AfMAT1-1-1$_{78-235}$. Similarly, a single nucleotide substitution of the third base within the core of MAT1-1-1 binding motif severely diminished formation of the protein–DNA complexes in *P. chrysogenum* (Becker *et al.* 2015). Therefore, we reasoned that nucleotides present in the core motif are important for the binding of MAT1-1-1 orthologs.

In summary, our study revealed that the DNA-binding motif of MAT1-1-1 from *A. fumigatus* resembles the motif reported for MAT1-1-1 from *P. chrysogenum*. Using a purified transcription factor fragment, we gained new insights into the molecular basis of DNA binding by fungal MAT alpha-box domains. The identification of putative AfMAT1-1-1 binding sites within the upstream regulatory regions of conserved target genes strengthens the idea that this process is highly conserved within members of Eurotiales (Eurotiomycetidae), despite their multiple lifestyle transitions. Additional *in vivo* studies will proof whether shortened versions of the MAT1-1-1 locus are sufficient to restore fertility in the sterile ΔMAT1-1-1 strain of *A. fumigatus*.

Acknowledgments

Literature cited

Ait Benkhali J, Coppin E, Brun S, Peraza-Reyes L, Martin T, *et al.* 2013. A network of HMG-box transcription factors regulates sexual cycle in the fungus *Podospora anserina*. PLoS Genet. 9:e1003642.

Alleyne TM, Pena-Castillo L, Badis G, Talukder S, Berger MF, *et al.* 2009. Predicting the binding preference of transcription factors to individual DNA k-mers. Bioinformatics. 25:1012–1018.

Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Alvarez-Perez S, Blanco J, Alba P, Garcia M. 2009. Mating type and invasiveness are significantly associated in *Aspergillus fumigatus*. Med Mycol. 48:1–6.

Ashton GD, Dyer PS. 2019. Culturing and mating of *Aspergillus fumigatus*. Curr Protoc Microbiol. 54:e87.

Baidya S, Cary JW, Grayburn WS, Calvo AM. 2011. Role of nitric oxide and flavohemoglobin homolog genes in *Aspergillus nidulans* sexual development and mycotoxin production. Appl Environ Microbiol. 77:5524–5528.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, *et al.* 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37:W202–W208.

Bailey TL, Elkan C. 1995. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. Mach Learn. 21: 51–80.

Baker CR, Tuch BB, Johnson AD. 2011. Extensive DNA-binding specificity divergence of a conserved transcription regulator. Proc Natl Acad Sci U S A. 108:7493–7498.

Bayram Ö, Bayram ÖS, Ahmed YL, Maruyama J, Valerius O, *et al.* 2012. The *Aspergillus nidulans* MAPK module AnSte11-Ste50-Ste7-Fus3 controls development and secondary metabolism. PLoS Genet. 8:e1002816.

Becker K, Beer C, Freitag M, Kück U. 2015. Genome-wide identification of target genes of a mating-type α-domain transcription factor reveals functions beyond sexual development. Mol Microbiol. 96:1002–1022.

Böhm J, Hoff B, O'Gorman CM, Wolfers S, Klix V, *et al.* 2013. Sexual reproduction and mating-type–mediated strain development in the penicillin-producing fungus *Penicillium chrysogenum*. Proc Natl Acad Sci U S A. 110:1476–1481.

Carr EA. 2004. α1-Induced DNA bending is required for transcriptional activation by the Mcm1–α1 complex. Nucleic Acids Res. 32: 2298–2305.

Cheema MS, Christians JK. 2011. Virulence in an insect model differs between mating types in *Aspergillus fumigatus*. Med Mycol. 49: 202–207.

Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: the Basidiomycota. Microbiol Spectr. 5 (3). doi: 10.1128/microbiolspec.FUNK-0046-2016

Debuchy R, Berteaux-Lecellier V, Silar P. 2014. Mating systems and sexual morphogenesis in Ascomycetes. In: KA Borkovich, DJ Ebbole, editors. Cellular and Molecular Biology of Filamentous Fungi. Washington, DC: ASM Press. p. 499–535.

Debuchy R, Turgeon BG. 2006. Mating-type structure, evolution, and function in Euascomycetes. In: U Kües, R Fischer, editors. Growth, Differentiation and Sexuality. Berlin/Heidelberg, Germany: Springer-Verlag. p. 293–323.

Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 43: W389–W394.

Dyer PS, Kück U. 2017. Sex and the imperfect fungi. Microbiol Spectr. 5 (3). doi: 10.1128/microbiolspec.FUNK-0043-2017.

Dyer PS, O'Gorman CM. 2012. Sexual development and cryptic sexuality in fungi: insights from *Aspergillus species*. FEMS Microbiol Rev. 36:165–192.

Dyer PS, Paoletti M, Archer DB. 2003. Genomics reveals sexual secrets of *Aspergillus*. Microbiology. 149:2301–2303.

Elion EA. 2000. Pheromone response, mating and cell biology. Curr Opin Microbiol. 3:573–581.

Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, *et al.* 2008. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet. 4:e1000046.

Fraser JA, Heitman J. 2004. Evolution of fungal sex chromosomes. Mol Microbiol. 51:299–306.

Fraser RS. 1993. Pulmonary aspergillosis: pathologic and pathogenetic features. Pathol Annu. 28:231–277.

Galgoczy DJ, Cassidy-Stone A, Llinas M, O'Rourke SM, Herskowitz I, *et al.* 2004. Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 101:18069–18074.

Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, *et al.* 2004. Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol. 2:e398.

Goodenough U, Heitman J. 2014. Origins of eukaryotic sexual reproduction. Cold Spring Harb Perspect Biol. 6:a016154.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics. 27:1017–1018.

Han S-Y, Ko J-A, Kim J-H, Han K-Y, Han K-H, *et al.* 2008. Isolation and functional analysis of the *silA* gene that controls sexual development in response to light in *Aspergillus nidulans*. Korean J Mycol. 36:189–195.

Heffler MA, Walters RD, Kugel JF. 2012. Using electrophoretic mobility shift assays to measure equilibrium dissociation constants: GAL4-p53 binding DNA as a model system. Biochem Mol Biol Educ. 40:383–387.

Hohl TM, Feldmesser M. 2007. *Aspergillus fumigatus*: principles of pathogenesis and host defense. Eukaryot Cell. 6:1953–1963.

Idnurm A, Walton FJ, Floyd A, Heitman J. 2008. Identification of the sex genes in an early diverged fungus. Nature. 451:193–196.

Jackson D, Lawson T, Villafane R, Gary L. 2013. Modeling the structure of yeast MATα1: an HMG-box motif with a C-terminal helical extension. Open J Biophys. 3:1–12.

Jauch R, Ng CKL, Narasimhan K, Kolatkar PR. 2012. The crystal structure of the Sox4 HMG domain–DNA complex suggests a mechanism for positional interdependence in DNA recognition. Biochem J. 443:39–47.

Kamachi Y, Kondoh H. 2013. Sox proteins: regulators of cell fate specification and differentiation. Development. 140:4129–4144.

Kim H-K, Jo S-M, Kim G-Y, Kim D-W, Kim Y-K, *et al.* 2015. A large-scale functional analysis of putative target genes of mating-type loci provides insight into the regulation of sexual development of the cereal pathogen *Fusarium graminearum*. PLoS Genet. 11:e1005486.

Klix V, Nowrousian M, Ringelberg C, Loros JJ, Dunlap JC, *et al.* 2010. 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. 9:894–905.

Kwon-Chung KJ, Sugui JA. 2009. What do we know about the role of gliotoxin in the pathobiology of *Aspergillus fumigatus*? Med Mycol. 47:S97–S103.

Latgé J-P, Chamilos G. 2019. *Aspergillus fumigatus* and aspergillosis in 2019. Clin Microbiol Rev. 33:e00140-18.

Lee DW, Freitag M, Selker EU, Aramayo R. 2008. A cytosine methyltransferase homologue is essential for sexual development in *Aspergillus nidulans*. PLoS One. 3:e2531.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2-\Delta\Delta CT$ method. Methods. 25:402–408.

Losada L, Sugui JA, Eckhaus MA, Chang YC, Mounaud S, *et al.* 2015. Genetic analysis using an isogenic mating pair of *Aspergillus fumigatus* identifies azole resistance genes and lack of MAT locus's role in virulence. PLoS Pathog. 11:e1004834.

Mahmoudjanlou Y, Dahlmann TA, Kück U. 2020. Molecular analysis of mating type loci from the mycophenolic acid producer *Penicillium brevicompactum*: phylogeny and MAT protein characterization suggest a cryptic sexual life cycle. Fungal Biol. 124:821–833.

Martin T, Lu S-W, van Tilbeurgh H, Ripoll DR, Dixelius C, *et al.* 2010. Tracing the origin of the fungal α1 domain places its ancestor in the HMG-Box superfamily: implication for fungal mating-type evolution. PLoS One. 5:e15199.

Mészáros B, Erdős G, Dosztányi Z. 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46:W329–W337.

Metzenberg RL, Glass NL. 1990. Mating type and mating strategies in *Neurospora*. BioEssays. 12:53–59.

Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, *et al.* 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature. 438:1151–1156.

O'Gorman CM, Fuller HT, Dyer PS. 2009. Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. Nature. 457:471–474.

Paoletti M, Rydholm C, Schwier EU, Anderson MJ, Szakacs G, *et al.* 2005. Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. Curr Biol. 15:1242–1248.

Paoletti M, Seymour FA, Alcocer MJC, Kaur N, Calvo AM, *et al.* 2007. Mating type and the genetic basis of self-fertility in the model fungus *Aspergillus nidulans*. Curr Biol. 17:1384–1389.

Philley ML, Staben C. 1994. Functional analyses of the *Neurospora crassa* MT a-1 mating type polypeptide. Genetics. 137:715–722.

Pöggeler S. 2002. Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. Curr Genet. 42:153–160.

Pöggeler S. 2007. MAT and its role in the homothallic ascomycete *Sordaria macrospora*. In: J Heitman, JW Kronstad, JW Taylor, LA Casselton, editors. Sex in Fungi. Washington, DC: ASM Press. p. 171–188.

Schlierf B. 2002. Cooperative binding of Sox10 to DNA: requirements and consequences. Nucleic Acids Res. 30:5509–5516.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 9:671–675.

Seo J-A, Han K-H, Yu J-H. 2004. The *gprA* and *gprB* genes encode putative G protein-coupled receptors required for self-fertilization in *Aspergillus nidulans*. Mol Microbiol. 53:1611–1623.

Sommerkamp JA, Frankenberg-Dinkel N, Hofmann E. 2019. Crystal structure of the first eukaryotic bilin reductase Gt PEBB reveals a flipped binding mode of dihydrobiliverdin. J Biol Chem. 294:13889–13901.

Speijer D, Lukeš J, Eliáš M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. Proc Natl Acad Sci U S A. 112:8827–8834.

Štros M, Launholt D, Grasser KD. 2007. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell Mol Life Sci. 64:2590–2606.

Sugui JA, Losada L, Wang W, Varga J, Ngamskulrungroj P, *et al.* 2011. Identification and characterization of an *Aspergillus fumigatus* "supermater" pair. MBio. 2:e00234-11.

Sugui JA, Pardo J, Chang YC, MüLlbacher A, Zarember KA, *et al.* 2007. Role of laeA in the regulation of *alb1*, *gliP*, conidial morphology, and virulence in *Aspergillus fumigatus*. Eukaryot Cell. 6:1552–1561.

Szewczyk E, Krappmann S. 2010. Conserved regulators of mating are essential for *Aspergillus fumigatus* cleistothecium formation. Eukaryot Cell. 9:774–783.

Traeger S, Altegoer F, Freitag M, Gabaldon T, Kempken F, *et al.* 2013. The genome and development-dependent transcriptomes of *Pyronema confluens*: a window into fungal evolution. PLoS Genet. 9: e1003820.

Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD. 2008. The evolution of combinatorial gene regulation in fungi. PLoS Biol. 6:e38.

Uemura E, Niwa T, Minami S, Takemoto K, Fukuchi S, *et al.* 2018. Large-scale aggregation analysis of eukaryotic proteins reveals an involvement of intrinsically disordered regions in protein folding. Sci Rep. 8:678.

van den Berg MA, Albang R, Albermann K, Badger JH, Daran J-M, *et al.* 2008. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. Nat Biotechnol. 26:1161–1168.

Varga J. 2003. Mating type gene homologues in *Aspergillus fumigatus*. Microbiology. 149:816–819.

Wada R, Maruyama J, Yamaguchi H, Yamamoto N, Wagu Y, *et al.* 2012. Presence and functionality of mating type genes in the supposedly asexual filamentous fungus *Aspergillus oryzae*. Appl Environ Microbiol. 78:2819–2829.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol. 337:635–645.

Yu Y, Amich J, Will C, Eagle CE, Dyer PS, *et al.* 2017. The novel *Aspergillus fumigatus* MAT1-2 - 4 mating-type gene is required for mating and cleistothecia formation. Fungal Genet Biol. 108:1–12.

Yu Y, Blachowicz A, Will C, Szewczyk E, Glenn S, *et al.* 2018. Mating-type factor-specific regulation of the fumagillin/pseurotin secondary metabolite supercluster in *Aspergillus fumigatus*. Mol Microbiol. 110:1045–1065.

Communicating editor: N. Louise Glass