Article (Discoveries)

Fungal Transcription factors: Markers of genetic innovation, network rewiring and conflict between genomics and transcriptomics.

Oier Etxebeste[1,≠].

[1]Laboratory of Biology, Department of Applied Chemistry, Faculty of Chemistry, University of the Basque Country (UPV/EHU), 20018 San Sebastian, Spain.

[≠] Corresponding author: Oier Etxebeste; E-mail: oier.echeveste@ehu.eus; Tel: (+34) 943 018517; Fax: (+34) 943 015270; Laboratory of Biology, Dept. of Applied Chemistry, Faculty of Chemistry, University of the Basque Country, Manuel de Lardizabal, 3, 20018, San Sebastian.

Running title: Genomic/transcriptomic analysis of *A. nidulans* transcription factors.

One-sentence summary: An *in silico* study of the predicted set of transcriptional regulators in the fungus *Aspergillus nidulans*, their conservation patterns and the correlation between gene annotations and transcriptomic data

Keywords: Transcription factor, transcriptional regulation, development, stress response, bioinformatics, genomics, transcriptomics, gene duplication, gene regulatory network, network rewiring.

Abstract

Gene Regulatory Networks (GRNs) are shaped by the democratic/hierarchical relationships among transcription factors (TFs) and associated proteins, together with the cis-regulatory sequences (CRSs) bound by these TFs at target promoters. GRNs control all cellular processes, including metabolism, stress-response, growth and development. Due to the ability to modify morphogenetic and developmental patterns, there is the consensus view that the reorganization of GRNs is a driving force of species evolution and differentiation. Duplication of genes coding for TFs, their divergent sequence evolution and gain/loss/modification of CRSs are events causing GRN rewiring. Fungi (mainly Saccharomycotina) have served as a reference kingdom for the study of GRN evolution. Here, I studied the genes predictably coding for TFs in the fungus *Aspergillus nidulans* (Pezizomycotina). Bioinformatics results support the view that duplication events of TF-coding genes potentially impact the species levels. Discrepancy between gene-annotation and transcriptomic data suggests that duplication of genes potentially coding for TFs are not always accompanied by the generation of functional forms and modification of regulatory potential. Overall, this study establishes a novel theoretical framework in synthetic biology, since overexpression of silent or primarily non-functional TF forms would provide additional tools for assessment of the mechanisms triggering GRN rewiring.

Introduction

Transcription factors (hereafter referred to as TFs) are proteins capable of binding DNA at cis-regulatory sequences (CRSs) of target genes. TF-CRS binding repress or promote transcription by RNA polymerase. The biological impact of most TFs cannot be completely assessed individually but collectively, as interacting units of a network or interconnected networks of TFs, associated proteins and CRSs which are generally known as Gene Regulatory Networks or GRNs (Park and Wang 2018). The architecture of GRNs ranges from simple, such as that controlling mating-type in yeasts (Nocedal and Johnson 2015) to complex, multistep hierarchical and democratic relationships formed by multiple TFs and associated proteins, such as the biofilm network of *Candida albicans* or the embryonic stem cell network of *Mus musculu*s (compared by (Sorrells and Johnson 2015)).

GRNs control all cellular processes, how cells interact with the environment and how they adapt to changes. Fundamentally, they determine growth and developmental patterns of all species. Considering that every single cellular, organ, tissue or organismal morphology is the phenotypic manifestation of the specific architectures of and connections among GRNs, introduction or removal of regulatory units within GRNs, or modifications in the connections among regulatory units, have the potential to render modified developmental phenotypes. Thus, it is generally accepted that rewiring of GRNs is an important driving force of species evolution and differentiation (Park and Wang 2018).

There are different mechanisms for modifying the structure of GRNs (Voordeckers et al. 2015). The emergence of new regulators (rather than the incorporation of individual target genes; “regulator-first” model; see references within (Sorrells and Johnson 2015)) is one of the major events, and occurs mainly through duplication of genes coding for TFs and divergent evolution of each paralog sequence. Amino acidic changes in TFs can precede gains of CRSs (Britton et al. 2020), leading to neofunctionalization (the development of new functions) or subfunctionalization (retention of a subset of the functions of the original paralog) (Sorrells and Johnson 2015; Voordeckers et al. 2015).

The kingdom fungi is composed of 2-4 million species (Hawksworth and Lücking 2017). The most primitive clades of the fungal tree of life are those of

Blastocladiomycota, Chytridiomycota and Zoopagomycota, while the three main clades correspond to Mucoromycota, Ascomycota and Basidiomycota (Spatafora et al. 2017). While species belonging to the class of Saccharomycetes reproduce mainly asexually by budding, most species of these three clades form filamentous multicellular colonies and more complex reproductive structures. Saccharomycetes have traditionally been used to study how GRNs are shaped and evolve (Nocedal and Johnson 2015; Sorrells and Johnson 2015; Britton et al. 2020). Recent comparative studies carried out in Basidiomycetes and Ascomycetes have shown that TFs controlling (asexual and/or sexual) developmental patterns emerged gradually in evolution (Etxebeste et al. 2019; Kiss et al. 2019; Krizsán et al. 2019). Furthermore, the formation of main clades such as that of Basidiomycota or that of the subphylum Pezizomycotina was accompanied by a boost in the number of TF-coding genes. This stepwise emergence of TFs has remarkable implications regarding, for example, the control of developmental patterns. In the order Eurotiales, the transcription factor BrlA plays a role as a bottleneck in the formation of asexual structures known as conidiophores (Adams et al. 1998). BrlA emerged much later than other upstream and downstream TFs of the same GRN (de Vries et al. 2017; Ojeda-López et al. 2018), suggesting that it caused rewiring events in pre-existent ancient networks (Etxebeste et al. 2019).

The starting point of this work is an update of the inventory of potential TF-coding genes elaborated by Wortmann and colleagues in the genome of the filamentous reference fungus *Aspergillus nidulans* (Wortman et al. 2009). Comparative searches show that there are TFs with a within-genus conservation pattern, suggesting that the stepwise emergence of TFs mentioned above impacts the species level. Multiple TFs, such as the C2H2-type BrlA, SteA and MsnA, probably share a common ancestor, showing that duplication events are recurrent. This seems to be a general trend in fungi, and is a predominant mechanism in the most represented family of DNA-binding domains (DBDs) of Ascomycetes: Zn2Cys6 clusters and fungal-specific TF domains. Nevertheless, discrepancies between gene-annotation and transcriptomic data, a high proportion of which are observed in genes coding for Zn2Cys6 TFs, suggests that the emergence of genes potentially coding for TFs is not always accompanied by the generation of functional forms and subsequent extension of the regulatory potential of the cell. By analyzing available RNA-seq data, examples of overlapping transcriptional responses to different stimuli are shown, such as nitrogen starvation and the emergence

of filamentous cells to the air environment. This implies the existence of interconnections among GRNs. Overall, our results suggest that filamentous fungal genomes constitute valuable resources to study how the set of regulatory proteins and the regulatory networks they form are organized and evolve. Furthermore, modification of the coding sequences of primarily non-functional or silent TF forms, and their functional assessment, would provide additional tools for studying mechanisms of GRN rewiring.

Results

The set of potential TFss in the ascomycete fungus *Aspergillus nidulans* is dominated by binuclear zinc clusters

Aspergillus nidulans is an ascomycete of the Eurotiomycetes class (subphylum Pezizomycotina). It has been traditionally used as a reference system for the study of filamentous fungal biology (Etxebeste and Espeso 2020), and thus, a large number of TFs have been characterized functionally and their genetic/molecular inter-relationships determined. TFs and GRNs controlling in fungi, i.e., stress-response, carbon and nitrogen metabolism, light sensing and the induction/progression of developmental cycles were characterized for the first time in this Aspergillus species. Taking advantage of this knowledge, recent works have focused on the study of how some of these GRNs were structured during the evolution of the kingdom fungi (de Vries et al. 2017; Ojeda-López et al. 2018; Etxebeste et al. 2019). An interesting feature of the TFomes of Ascomycota, in general, is the remarkable expansion of zinc cluster-type TFs, while in Basidiomycota and early diverging fungi, C2H2-type TFs are the most represented family (Shelest 2017). To explore new features of its TFome, the conservation and expression patterns of the genes encoding these TFs, and uncover new traits of the organization and evolution of GRNs in the kingdom fungi, firstly, I updated the list of TFs potentially encoded in the genome of *A. nidulans*. With that aim, the list published by Wortman and colleagues was used as a reference (Wortman et al. 2009) and updated (Table S1) based on several works available in the literature and the FungiDB database (Basenko et al. 2018). The information available about each predicted TF, the extension of their regulatory domains (IPD: Interpro Regulatory Domains) and their amino acid sequences were retrieved from the FungiDB database and the InterPro server (Mitchell et al. 2019). Table S1 lists the 923 IPDs elaborated in the predicted 520 TFs, together

with their full-length and IPD amino acid sequences. The tables in Figure 1A show the number of TFs and IPDs of each family. The updated distribution shows that *A. nidulans* follows the same trend as Ascomycota species (Shelest 2017). Zn2Cys6 family TFs are a 63.7% (331 of 520) of the predicted TFome and Zn2Cys6 IPDs are a 36.0% (332 of 923) of the total number of predicted DNA-binding domains. Both are followed by Fungal-specific regulatory TFs/IPDs, most of the times associated with zinc clusters, and C2H2-type TFs/IPDs (14.0%, 73 of 520, and 19.7%, 182 of 923, of C2H2-type TFs and IPDs, respectively) (Figure 1A). The values given for Zn2Cys6-type TFs and IPDs were obtained after the observation that the pre-existing annotation of specific genes (i.e., *An11962*, *An1705*, *An5464*, *An5510*, *An7076*, *An7584*, *An8103* or *An8501*) was probably incorrect as their 5´-UTR or 5´-coding regions included non-predicted or non-processed introns as well as regions coding for Zn2Cys6 IPDs (see next section for discrepancy between gene annotation and transcriptomic data). This suggests that the percentage of Zn2Cy6-type TFs or IPDs coded by the genome of *A. nidulans* is even higher than initially predicted.

Figure 1B corresponds to a draft tree ("rough tree" generated with MAFFT and edited with iTOL) differentiating each IPD family. Again, the remarkable expansion of zinc clusters in combination with Fungal specific TF domains (light and dark orange clades, respectively) can be clearly seen, followed by C2H2-type IPDs (dark grey clade). Of a total of 520 potential TFs, 118 (22.7%) had a standard name, meaning that they have been functionally characterized to some extent (Table S2). That suggests that 77.3% (402 out of 520) of the predicted TFome of *A. nidulans* remains to be functionally characterized. The FungiDB database predicted the presence of a signal peptide in 16 potential TFs (3.1%) while SignalP was much more restrictive (only 4; 0.8%) (Table S3), suggesting that the number of TFs targeted to the endoplasmic reticulum is probably low. The number of potential TFs with predicted transmembrane (TM) domains (Table S3) was 24 (4.6%) according to the FungiDB database or the TMHMM website (a less restrictive prediction by Phobius suggested that the number of TM domain-containing TFs was 143; 27.5%). This suggests that there may be a significant subset of TFs participating in signal transduction mechanisms between membraneous compartments and nuclei, as has been described, for example, for SrbA/An7661, an SREBP-type (sterol regulatory element-binding protein) TF of *A. nidulans* (Bat-Ochir et al. 2016). Nevertheless, the number of active TM domain-containing TFs will in all

probability be lower, according to the RNA-seq data that will be analyzed in more detail below.

Discrepancy between genomic annotation and transcriptomic data

Comparative analyses among species are, in general, done based on genome annotations. However, the reliability of eukaryotic genome annotations is considered low, due in part to the low percentage of eukaryotic genomes comprising protein-coding exons and that gene coding regions are interrupted by introns (Salzberg 2019). Thus, I wondered what percentage of the potential *A. nidulans* TFome is transcribed as predicted and if there was a subset of genes rendering TF forms different from the predicted full-length forms or even non-functional. With that aim, RNA-seq experiments published by us previously (Garzia et al. 2013; Oiartzabal-Arano et al. 2015; Picazo et al. 2020) and transcriptomic data available in the FungiDB database were used as a reference to compare genomic annotations and the most probable transcript sequences (Figure 2).

The information of RNA-seq analysis for each predicted TF-coding gene is included in Table S4 (column B: RNA-seq data). It was not possible to assess the annotation of approximately 27 genes (out of 520; 5.2 %; Figure 2A and Table S4; the expression levels of *An3391*, encoding a Zn2Cys6-type TF, are shown in Figure 2B as an example of this first group). The coding sequences of approximately 336 genes/transcripts are processed as predicted (64.6%; Figure 2A) in the culture conditions and genetic backgrounds analyzed (coverage of RNA-seq reads at the *An2290/steA locus*, coding for a regulator of sexual development (Vallim et al. 2002), is shown as an example in Figure 2C). In multiple cases, alternative splicing events were identified in transcripts but the corresponding genes were included in this second group because the position and extension of the introns were correctly annotated. Intron retention (IR) was the most common alternative splicing form, together with the identification of alternative start or end sites. The expression levels of *An1569*, encoding a Zn2Cys6 TF, are shown as an example in Figure 2D. Retention of the first intronic sequence leads to a premature stop codon that would generate a truncated peptide containing only the Zn2Cys6 IPD (amplification of the dotted orange square in the bottom panel of Figure 2D). Considering the low number of reads showing processing of that first intron, it can be

suggested that the concentration of the full-length form of the protein will be, in all probability, very low.

Genes/transcripts needing some type of re-annotation were included in a third group (approximately 157 out of 520; 30.2 %; Figure 2A). Of them, 124 were predicted to code for Zn2Cys6 and/or Fungal sp. IPDs, 17 for C2H2, 5 for bHLH, 3 for bZIP, 3 for homeodomain, 1 for a myb-like, 2 for GATA-type, 1 for Copper Fist, 1 for WOPR, 1 for Velvet and 1 for Forkhead domain IPDs. In multiple cases, the discrepancy between annotation and transcriptomic data was limited to the 5´- or 3´-end of the gene. This can be seen in the cases of *An11098*, encoding a Zn2Cys6-type TF in which an alternative start site was identified (Figure 2E), and *An2782*, also coding for a binuclear zinc cluster in which the third predicted intron was retained, modifying the reading frame and the stop signal (Figure 2F). Transcriptomic data also suggested that the coding sequences of a subset of genes within this third group clearly diverge from the predicted ones. Three examples are shown in Figure 2. The start of the *An1812/jlbA* transcript, which is upregulated under nitrogen starvation (Strittmatter et al. 2001), is as annotated but non-processing of the first intron leads to a premature stop (Figure 2G). The probable start codon of the coding sequence will be that for Met175, with the bZIP domain predictably extending from residues 194 to 257 (the predicted polypeptide in FungiDB and AspGD databases is 258 amino acids long). An9221 is predicted to code for a TF with a signal peptide (SP) and six transmembrane (TM) domains in the N-terminus of the protein (first 249 amino acids), and a Zn2Cys6 IPD between residues 267 and 303 (Figure 2H). The number of reads mapping to the first two exons and introns is very low but suggests that the two introns are retained, leading to a premature stop (bottom panel in Figure 2H). The probable start codon is that for Met246, just before the binuclear zinc cluster domain. Finally, the analysis of *An5252* is shown in Figure 2I. First of all, and surprisingly, the predicted protein sequence of An5252 is identical to that of An9240, suggesting that they are not paralogs but rather correspond to a genome assembly/annotation error. Despite the low number of reads mapping to this gene (Figure 2I), RNA-seq data suggest that only intron 2 is processed, while intron 1 is retained. Retention of intron 1 probably leads to a premature stop codon and a first coding region covering codons for M1 and G90 (plus 9 codons our of frame prior to a stop). A second coding region would start with the codon for M127, implying the loss of the first predicted C2H2 domain (dark grey square bold in red). The predicted second

C2H2 IPD would be included in this second, hypothetical coding region (M127-D168 plus 20 codons out of frame).

A subset of genes potentially coding for TFs, mainly within the binuclear zinc cluster family, remains silent

To analyze the expression patterns of the genes encoding the potential TFome of *A. nidulans*, we retrieved RNA-seq data from multiple transcriptomic comparisons carried out using *A. nidulans* as the reference system and publicly available in the FungiDB database or other repositories (see Materials and Methods). Table S5 includes FPKM (Fragments Per Kilobase Million) and TPM (Transcripts Per Kilobase Million) data for each TF-coding transcript, which are ordered according to the IPD family the corresponding protein belongs to (dual specificity TFs, those containing more than one type of IPD (Shelest 2017), were included in all the families they belong to and, thus, repeated). First, heatmaps were built for each comparison. Clustering was avoided to check if there could be any expression pattern associated with a specific IPD family. Figure S1 shows that it is not the case.

Since an important number of the transcripts in Table S5 showed low or null expression, all values below 5 in the case of FPKM (heatmaps 1-6 and 9-10) or TPM (7th and 8th heatmaps) were removed from the table to build Figure 3. The increase in the number of gray areas, principally in the sections corresponding to Zn2Cys6-Fungal sp. or aflatoxin regulatory IPD families, followed by some C2H2-type TFs, suggests that an important fraction (at least 53 out of 520; 10.2%) of the TFome of *A. nidulans* tends to remain silent under the tested experimental conditions/genetic backgrounds (others are silent in most but not all culture conditions/genetic backgrounds). That is expected in the case of TF-coding genes of secondary metabolite gene clusters (SM clusters). Nevertheless, among those 53 genes that remain silent in all conditions/backgrounds analyzed, only 13 are predicted or shown to be members of SM clusters (*An0533*, *An1029*, *An2025*, *An3250*, *An3391*, *An6440*, *An7076*, *An7913*, *An8103*, *An8111*, *An8509*, *An8916* and *An9236*; Table S5; Figure 3). As described above, annotation of others, such as *An5252*/*An9240*, is clearly incorrect (see Figure 2 and Table S4), suggesting that no functional TF forms are translated from some predicted genes in detectable levels.

Figure 3 and Figure S1 also suggest that specific culture conditions and genetic backgrounds trigger a change in the general expression pattern of the *A. nidulans* TFome (see black bars in Figures 3 and S1). Clustering of heatmaps corresponding to the effect of the addition of sugarcane bagasse as the main nutrient source (Figure 4A and 4D) (Brown et al. 2016), the effect of nitrogen starvation (Figure 4B and 4E) (Sibthorp et al. 2013) and a null *sltA* (salt sensitivity) background (Figure 4C and 4F) (Picazo et al. 2020) supports this possibility. Qualitatively, a majority of TF-coding genes are upregulated in the first two culture conditions but downregulated in the *sltA*Δ mutant.

Duplication of TF-coding genes impacts the species level

In a previous work, and using basically BLAST and HMM analyses, the conservation pattern of 33 regulators of fungal development was analyzed in 503 fungal proteomes representing all phyla and subphyla, and most known classes (Etxebeste et al. 2019). The different conservation patterns observed suggested stepwise emergence of transcription factors. Interestingly, the conservation of a subset of those transcriptional regulators was restricted to the class and order levels, and in specific cases almost to the species level, as apparently occurred with MoConX4 of *Magnaporthe oryzae*. A similar trend was observed by Kriszán and colleagues (Krizsán et al. 2019). Consequently, it was suggested that the emergence of some TFs may be recent, and may have led to the rewiring of the transcriptional networks controlling development and to further morphological diversification.

The FungiDB database includes an OrthoMCL identifier for each gene (Fischer et al. 2011). This preliminary analysis of orthologs and paralogs within the FungiDB database gave the opportunity to make a first screening of *A. nidulans* TFs with a restricted phylogenetic distribution (not shown), and made focus the search on bHLH-type TFs. Thirteen genes predictably coding or shown to code for TFs of this family were identified by the FungiDB database. Five of them had an standard name (*An1298/glcD*, *An7553/devR*, *An7661/srbA*, *An7734/anbH1* and *An8271/palcA*; (Caddick et al. 1986; Arst et al. 1990; Caruso et al. 2002; Tüncher et al. 2004; Bat-Ochir et al. 2016)), while the remaining eight had only a systematic name. By sequence similarity, An4394/UrdA was initially described to belong to this family of TFs (Oiartzabal-Arano et al. 2015; Pandit et al. 2018). However, the Interpro website did not identify a basic helix-loop-

helix domain within its primary structure. Thus, it was included in Table S1 but excluded from the following analysis.

First, and to analyze the evolution of *A. nidulans* bHLH-s, sequences of proteins predicted to contain this type of transcriptional regulatory domain were retrieved from the FungiDB database. The search was limited to *Aspergillus* species included in this database, plus *Penicillium rubens*, *Talaromyces stipitatus* and *T. marneffei*. The phylogenetic tree (Figure 5A) suggested that all but four *A. nidulans* bHLH sequences (An0396, An2811, An2392 and An5078) are conserved in the family Trichocomaceae.

In the case of *An5078*, RNA-seq data strongly suggest that the predicted two introns are not processed, modifying the start codon to that in position 92 (Figure 5B). BLAST searches and HMMER analyses gave several possible orthologs of An5078. Nevertheless, a reverse BLAST of each specific hit against the genome database of *A. nidulans* at the AspGD database (confirmatory reverse retrieval; (Etxebeste et al. 2019)) returned An5078 as the first hit in very few cases (Table S6), confirming a conservation pattern for An5078 restricted to less than ten *Aspergillus* species and suggesting that it is a recent duplication of An7170.

The phylogenetic tree in Figure 5A included An0396 in the An5078 clade. That could be a consequence of a hypothetic incorrect annotation of the gene. However, RNA-seq data showed that the predicted intron is processed and the predicted protein sequence is in all probability correct (Figure 5C). BLAST and HMMER searches, as well as confirmatory reverse retrievals, showed conservation of An0396 exclusively in *A. mulundensis*, with the rest of hits being orthologs of An5078, first, and An7170 in second instance. These results strongly suggest that An0396 corresponds to an even more recent duplication of An5078.

An2392 is predicted to be a TF of the An10297-containing (coding for an NRPS-like protein) secondary metabolite gene cluster. Despite the low expression levels, mapping of RNA-seq reads suggest that the predicted second intron is not processed, leading to a premature stop codon and the loss of the last two exons (Figure 5D). BLAST and HMMER searches, together with confirmatory reverse retrievals, strongly suggest that An2392 corresponds to a duplication of An6295 and that its conservation pattern is highly restricted, a common feature of genes belonging to secondary metabolite gene

clusters (Nützmann et al. 2018; Rokas et al. 2018). Finally, despite the low expression levels, RNA-seq data suggested that the annotation of *An2811* is incorrect and that the probable aminoacidic sequence is shorter than that predicted but includes a bHLH domain (Figure 5E). BLAST and HMMER searches gave 19 probable orthologs and in all cases the confirmatory reverse retrieval against the *A. nidulans* genome at the AspGD database returned An2811 as the first hit. The length of all these hits was in the range of 83 (*A. nidulans*) and 169 aa-s (*A. thermomutatus*). Thus, results suggest that the restricted phylogenetic distribution of An2811 shown in Figure 5A is an artifact and that it is conserved at least in the genus *Aspergillus*. Overall, the analysis of the bHLH family of TFs in *A. nidulans* suggests that this family is prone to duplication and that the most recent duplication events impacted the species level (see Discussion).

The Zn2Cys6 and C2H2 families of TFs as paradigms of gene duplication events extending the transcriptional regulatory potential of *A. nidulans*

The biggest family of TFs in *A. nidulans* in particular, and Ascomycetes in general (Shelest 2017), is that of zinc clusters. In addition, gene duplication and the generation of paralogs is one of the main events causing rewiring of GRNs and extension of the regulatory potential (see the introduction and discussion of this work). To update and extend the initial characterization of zinc clusters carried out by Todd and Andrianopoulos (Todd and Andrianopoulos 1997), the CD-Hit website (sequence identity cut-off of 0.3; see Materials and Methods) was used to search for paralogs within this family of TFs (Figure 6). CD-Hit identified clusters of up to six paralogous Zn2Cys6 TFs, totaling 139 members of this family that could have arisen by gene duplication. This number means that, at least, 42.0% of all predicted *A. nidulans* Zn2Cys6-type TFs would be a result of paralogy (potential clusters of 2, 3, 4, 5 and 6 paralogs are indicated in Figure 6A with brown, green, purple, red and blue lines, respectively; circles indicate TFs predictably involved or shown to be involved in the control of secondary metabolism while stars indicate those with a standard name). The biggest three clusters, composed of six paralogs each, were taken as a reference. Unexpectedly, the six members of one of them, An7896, An2036, An3385, An0533, An6788 and An1029 are all regulators of secondary metabolite gene clusters, which contrasts with the aforementioned general idea that these clusters were observed to be highly specific. A BLAST alignment using An3385 as query shows the high score, expect and coverage values (Figure 6B; the sequence alignment can be seen in Figure

S2). A similar pattern was observed in the remaining two clusters of six paralogs (Figure 6C and 6D; sequence alignments shown in Figures S3 and S4, respectively).

Overall, the above results correlate with the idea of gene duplication being a main event in the extension of the regulatory potential of organisms and suggest that duplication of genes coding for Zn2Cys6 TFs has played a key role in the expansion of this family of regulators in Ascomycota. Nevertheless, it remains to be investigated why has been this family of TFs, and not the others, the one expanded in this phylum (see Discussion).

As a comparison, the same procedure was followed with C2H2-type TFs, the second biggest family of TFs in *A. nidulans*. The CD-Hit suite predicted paralog clusters of 2 or 3 members (highlighted by brown and green lines in Figure 7A, respectively; again, circles indicate TFs predictably involved or shown to be involved in the control of secondary metabolism, while stars indicate those C2H2-type TFs with a standard name). A total of 18 (16 considering that An5252 and An9240 are probably the same protein; see above) out of 73 potential C2H2-type TFs were predicted to be a result of duplication, a clearly lower percentage (24.7%; 21.9% if An5252 and An9240 are not considered as paralogs) compared to that estimated in the case of Zn2Cys6 TFs (42.0%).

The case of the developmental regulator BrlA (Boylan et al. 1987) is paradigmatic among C2H2-type TFs of *A. nidulans*, since it is conserved exclusively in the order Eurotiales, and maybe in Chaetothyriales (Eurotiomycetes), but occupies a central position in the genetic pathways that control conidiation (de Vries et al. 2017; Ojeda-López et al. 2018; Etxebeste et al. 2019). Indeed, BrlA links the upstream developmental activation (signal transduction) pathways (UDA) and the central developmental pathway (CDP) that controls most of the morphological changes that lead to conidia production (Etxebeste and Espeso 2020). Interestingly, the promoter of *brlA* and its gene products developed the ability to recruit upstream TFs and transcriptionally control downstream TF-coding genes that are more widely conserved in fungi. Thus, it constitutes a model system for the study of how the emergence of new regulators can trigger GRN rewiring (Etxebeste et al. 2019).

BLAST and HMM analyses carried out to determine the conservation pattern of BrlA in the proteomes of 503 fungal species (Etxebeste et al. 2019) returned hits for BrlA in

species outside Eurotiomycetes. Nevertheless, according to the confirmatory reverse retrieval against the *A. nidulans* proteome within the AspGD database, these hits corresponded, in most of the cases, to orthologs of SteA (not shown), a TF widely conserved in fungi and necessary for sexual development (Figure 2C) (Vallim et al. 2002). Trying to obtain clues on the origin and the ancestor(s) of BrlA, new BLAST analyses were carried out at the *Aspergillus nidulans* (www.aspgd.org) or the *Saccharomyces cerevisiae* (www.yeastgenome.org) genome databases using BrlA as the query sequence. The sequences of the C2H2 IPDs of the best hits were retrieved, aligned and the corresponding phylogenetic tree generated (Figure 7B). BrlA was included in the same clade as the C2H2 IPDs of MsnA (and the Msn paralogs of *S. cerevisiae*, Msn2, Msn4 and Com2; necessary for viability under stress conditions; (Siggers et al. 2014)) and next to the clade including the C2H2 domain of SteA and the carbon catabolite repressor CreA (Dowzer and Kelly 1991). Indeed, the sequences aligned in Figure 7C conserved all but one of the residues of *S. cerevisiae* Msn paralogs predicted to contact DNA (green and red arrows, respectively) (Siggers et al. 2014). The only divergence was the substitution of a Gln in the second C2H2 IPD of Msn proteins and SteA by an Ala in BrlA. Thus, it is tempting to suggest that the C2H2-type transcriptional regulatory region of the asexual developmental regulator BrlA belongs to the same family of those of MsnA and SteA, regulators of stress response and sexual development, respectively. However, and as expected, the C2H2 domains of MsnA (residues T463-H515) could not replace functionally those of BrlA (residues F320-H375), since a recombinant strain expressing a BrlA::HA_{3x} chimera bearing this replacement (M1-Q319 of BrlA, T463-H515 of MsnA, and S376-E432 of BrlA) showed an aconidial phenotype (Figure 7D).

Overlap of GRNs: nitrogen starvation and the control of asexual development in *A. nidulans*

To understand the structure, function and evolution of complex GRNs, scientists hierarchically decompose these networks into manageable and intelligible subsystems (see references in (Verd et al. 2019)). This approach has been informative, going back again to the example of the control of asexual development in *A. nidulans*, with UDA and CDP pathways and the link established between them by BrlA (see the previous section of this work). Nevertheless, GRNs are, in general, connected with other networks. Indeed, TFs inducing sexual development, such as NsdD, act as repressors of

asexual development by directly binding to the promoter of *brlA* and inhibiting its expression (Lee et al. 2014; Lee et al. 2016). In addition, the expression of *brlA* is induced or repressed depending on several environmental stimuli, such as light/darkness, the presence of O_2/CO_2, salt or osmotic stress conditions or nutrient saturation/starvation (see references in (Etxebeste and Espeso 2020)). For example, emergence of hyphae to the air environment induces the expression of *brlA* through the UDA pathway but at the same time, nitrogen starvation also induces the expression of *brlA* and conidia production in submerged culture (Skromne et al. 1995; Etxebeste et al. 2009).

To explore the relationship between nitrogen starvation and the induction of asexual development in *A. nidulans*, I took advantage of RNA-seq data available in the FungiDB database and comparing the transcriptomic profiles in AMM or complete (ACM) culture media with those under nitrogen starvation (4 and 72 hours) (Sibthorp et al. 2013). First, genes with a fold change (FC) equal to or above three, and those with a FC ≤ 3, were filtered for the comparison of samples collected under nitrogen starvation (72h) and in AMM that used ammonium as the nitrogen source (Table S7). Figure 8A shows that most genes were upregulated under nitrogen starvation (72h) compared to ACM or AMM (supplemented with either nitrate or ammonium) media (black squares). The ratio of genes with a FC ≥ 3 to those with a FC ≤ 3 was 2.84 (2,219 compared to 781, respectively), suggesting that nitrogen starvation induced the expression of more than 20% of the genes estimated in *A. nidulans* (Figure 8B). Genes encoding TFs of the Zn2Cys6, fungal-specific (usually associated with zinc clusters) and C2H2 families were, respectively, the top 1, top 4 and top 7 in the list of Interpro domains upregulated (FC ≥ 3) under nitrogen starvation (Figure 8C). Indeed, only 18 TF-coding genes were downregulated (FC ≤ 3) compared to the 196 TF-coding genes that were upregulated (FC ≥ 3) under nitrogen starvation (ratio: 10.89) (Figure 8D; Table S8).

Interestingly, multiple genes encoding known regulators of asexual and sexual development showed a FC ≥ 3 under nitrogen starvation, such as UDAs *flbC*, *flbD*, *urdA*, *hbxA*, *hbxB* or *devR*, and CDPs *brlA*, *wetA*, *stuA* and *medA* (Figure 8E) (Boylan et al. 1987; Marshall and Timberlake 1991; Miller et al. 1992; Busby et al. 1996; Tüncher et al. 2004; Garzia et al. 2010; Kwon et al. 2010; Oiartzabal-Arano et al. 2015; Son et al. 2020). This observation reveals transcriptional links to the experimental observation

described years ago that nitrogen starvation induced conidia production (Skromne et al. 1995).

In addition, and although genes encoding key regulators of nitrogen metabolism such as *nirA*, *areA* and *areB* were, as expected, also included in the list of 196 TFs, and this could misdirect the search, it was hypothesized that new regulators of *A. nidulans* development could be identified. Thus, five uncharacterized TF-coding genes upregulated under nitrogen starvation (72h) were deleted (Figure 8F). Multiple criteria were defined for selection of these five genes. First, orthologs of genes that in other fungi such as *Neurospora crassa* are important for development were included (i.e. according to the FungiDB database *An10192* is the ortholog of *N. crassa tah-2*). Second, genes significantly deregulated after the induction of asexual development were included (*An1500* and *An4324*, respectively) (Garzia et al. 2013). Third, genes with one (*An10192*) or no paralogs (*An1500*, *An2001*, *An4324* and *An4527*) were prioritized. Finally, genes coding for TFs of the most numerous families were selected (*An10192* and *An2001* encode zinc clusters, *An1500* encodes a C2H2-type TF, *An4324* encodes a bZIP-type TF, and *An4527* encodes a cMyb-type TF). In comparison with the parental wild-type and the remaining null mutant strains generated, the Δ*An2001/nsiD* (nitrogen starvation-induced developmental regulator) mutant showed partial inhibition of radial growth and conidia production, mainly in ACM supplemented with nitrate as the main nitrogen source (Figure 8G and Figure S5A and S5C). After 144 hours of culture at 37 ºC, while the reference wild-type strain produced cleistothecia in ACM (red arrowheads in Figure 8G and Figure S5B and S5D), the Δ*nsiD* mutant was blocked in the development of sexual structures. Overall, results show that analyses of the transcriptomic responses of *A. nidulans* to nitrogen starvation can help in the identification of hitherto uncharacterized regulators of development. In the case of NsiD, our group plans to elucidate the functional relationship with known regulators of growth and sexual/asexual development as well as the main features of its regulatory activity. At the moment, it can be stated that it shows moderate/high expression levels under most of the conditions and in most of the genetic backgrounds tested through RNA-seq in *A. nidulans* (Table S5), that its three introns are processed as predicted (Table S4), and that it is conserved in the subphylum Pezizomycotina (not shown). In addition, and considering that nitrogen starvation or the use of different nitrogen sources can acidify (i.e. ammonium) or alkalinize (i.e. nitrate) the culture medium, it can be

hypothesized that specific regulators of nitrogen metabolism and/or stress response could act, indirectly or directly by binding to the promoter of *brlA*, as developmental activators.

Discussion

How GRNs evolve and what can fungi teach us about the process

There are three main properties that help us understand how GRNs evolve (see references in (Igler et al. 2018)). Robustness is related to the ability of a system to sustain its functionality in the face of perturbation (Kitano 2004; Noman et al. 2015), which, in the case of GRNs, could be understood as the extent a phenotypic trait persists after mutation (Igler et al. 2018). Tunability can be defined as the changes in gene expression levels, for example, due to gain, loss or mutation of CRSs, or mutation of the DBDs of TFs. Finally, evolvability, beyond the generic ability to evolve or the capacity of an evolving system to generate or facilitate adaptive change, is associated with the capacity of GRNs to modify or acquire novel regulatory connections (Igler et al. 2018; Verd et al. 2019). Gene duplication and divergent sequence evolution of paralogs are a central enablers for these three parameters and, thus, for the rewiring and reorganization of GRNs. As mentioned in the introduction of this work, duplication and sequence divergence will probably cause gains or losses of CRSs, and neofunctionalization or subfunctionalization would modify connections within or among gene networks (Sorrells and Johnson 2015; Voordeckers et al. 2015; Britton et al. 2020). Consequently, all of these events could directly impact robustness, tunability and evolvability of GRNs.

Duplication of TF-coding genes and their divergent sequence evolution is also a main mechanism for GRN evolution in fungi. The availability of hundreds of fungal genomes and proteomes, representing most of the clades of the fungal tree of life (Spatafora et al. 2017; Li et al. 2020), gives fungal researchers the possibility to elucidate when each family of transcriptional regulators and each TF emerged (de Mendoza et al. 2013; Todd et al. 2014; Etxebeste et al. 2019; Krizsán et al. 2019; de Mendoza and Sebé-Pedrós 2019). *A. nidulans* has been used as a reference fungus since the 1940s. Major transcriptional regulators of multiple GRNs such as those controlling sexual and asexual development, stress response, carbon and nitrogen metabolism, secondary metabolism,

and so on, have been functionally characterized in *A. nidulans*, making it a valuable system for the study of how GRNs are rewired and evolve after the emergence and duplication of new regulators.

Knowledge thus gained can, of course, be extended to other fungal systems. Recently, new experimental approaches have been developed, focused on the expression of genes encoding regulators of *A. nidulans* asexual reproduction with the aim of modifying developmental patterns of species such as *Monascus ruber* or *Histoplasma capsulatum* (Ojeda-López et al. 2018; Mead et al. 2020). Although the trials were unable to trigger major morphological changes, they paved the way for the establishment of simple and reliable approaches for the experimental study of GRN rewiring and its consequences on organism evolution.

Tracking of the emergence of fungal TFs have shown us that it occurs gradually, with specific clades exhibiting a burst in the number of new regulators (Etxebeste et al. 2019; Krizsán et al. 2019). Furthermore, results support the view that this process is currently active, with the existence of TFs with an extremely narrow or almost species-specific distribution. In plants, bHLHs form the second largest family of TFs, where they are key regulators of important metabolic, physiological and developmental processes (Pires and Dolan 2010a). This family of TFs had been shown to be remarkably expanded approximately coincident with the appearance of the first land plants, but has subsequently remained relatively conserved in this group of plants (Pires and Dolan 2010b; Pires and Dolan 2010a). In molluscs, Bao and colleagues proposed that bHLH genes showed high evolutionary stasis at the family level but considerable within-family diversification by tandem gene duplication (TGD) (Bao et al. 2017). The results obtained here from the analysis of the set of bHLH-type TFs of *A. nidulans* correlate with the idea of high within-family diversification; though not through a tandem gene duplication (see below for Zn2Cys6-type TFs). Thus, members of this family of TFs could be viewed as a paradigm of intra-genus gene duplication. *An2392* probably emerged as a duplication of *An6295*. *An5078* seems to be a duplication of *An7170* and, interestingly, *An0396* is probably an even more recent duplication of *An5078*.

But the largest family of TFs in *A. nidulans* in particular and Ascomycetes in general is that of Zn2Cys6-type TFs (Shelest 2017), occurring to a large extent due to multiple events of gene duplication. Shelest described that, among the most abundant TF-coding

gene families, not all of them expand, not all of those expanding do it at the same rate, and not all of them expand following comparable trends in all fungal lineages (Shelest 2017). Zinc clusters grew much faster in Ascomycetes, suggesting a particular role and evolutionary history of Zn2Cys6-type TFs (Shelest 2017). An interesting question is what makes this family so prone to successful gene duplication in Ascomycetes (the same question could be made for C2H2-type TFs in Basidiomycetes and early diverging fungi).

Gene duplication can occur by several mechanisms, including whole- or segmental-genome duplication, and single gene duplication, which includes tandem, proximal, retro-transposed, DNA-transposed and dispersed duplications (Zhang 2003; Lallemand et al. 2020). Whole-genome duplication events have been described in fungi (Corrochano et al. 2016) but not in *Aspergilli* (Khaldi and Wolfe 2008). Figure S6 and Table S9 show the genomic distribution of those genes encoding the Zn2Cys6-type TFs shown in Figure 6A and that could constitute clusters of paralogs according to the CD-Hit algorithm. It has been described that specific mechanisms of TGD might lead to increasing or decreasing copy numbers in gene families (Lallemand et al. 2020). However, based on their chromosomal distribution, TGD seems not to be the main mechanism of gene duplication of those probable paralog clusters of Zn2Cys6 TFs of *A. nidulans*. Segmental duplications (SDs) are defined as genomic segments longer than 1 Kb that are repeated within the genome with at least 90% sequence identity (Bailey et al. 2001). But, as mentioned previously, Khaldi and Wolfe described that they found no clues of SDs in *Aspergilli* (Khaldi and Wolfe 2008). Dispersed duplications (DDs) have been defined as the generation of two gene copies that are neither neighboring nor colinear (Qiao et al. 2019), a definition which fits with the genomic distribution of the members of the hypothetic Zn2Cys6 paralog clusters of *A. nidulans* (Figure 6A, Figure S6 and Table S9). Despite the fact that the mechanisms enabling DDs remain unclear and have been defined as unpredictable and occurring through random patterns (Qiao et al. 2019), dispersed duplication can be proposed as one of the main duplication events causing the expansion of Zn2Cys6-type TFs in *A. nidulans*.

Interestingly, there are Zn2Cys6 paralog clusters that could have been generated following different duplication mechanisms. One of the biggest Zn2Cys6 paralog clusters is composed of regulators of secondary metabolite gene clusters (An7896, An2036, An3385, An0533, An6788 and An1029; Figure 6B). A BLAST search against

the genome of *A. nidulans* using the sequence of An7903 (the polyketide synthase, PKS, of the *dba* cluster) as the query (Gerke and Braus 2014), suggested that the PKSs An2032, An3386, An0523 and An1034 (the expect, score and coverage values of An6791 were considerably lower) could also be paralogs (at least, these were the best hits). Similar correlations were observed in the case of the esterases/lipases An7899, An2031, An0529, An6789 and An1032, or the FAD-binding monoxygenases An7902, An2033, An3382, An0530 and An1033. Additional BLAST searches against the *A. nidulans* genome also linked An7893 with An0526, An11584 (located between An7897 and An7898) with An0527, or An7898 with An0528. It has been proposed that the presence of the *dba* cluster in *A. nidulans* is a consequence of a horizontal gene transfer (HGT) event from *Talaromyces stipitatus* (Nguyen et al. 2015; Oiartzabal-Arano et al. 2015), while the donor of the cluster defined by An2036 probably was *Coccidioides immitis* (Oiartzabal-Arano et al. 2015) or *Penicillium italicum* (Nguyen et al. 2015). The rest of the secondary metabolite gene clusters corresponding to the aforementioned Zn2Cys6 paralog group have not been linked with HGT events (Nguyen et al. 2015), and could correspond to internal duplication events (followed by divergent evolution). Again, location of these clusters in different chromosomes make the possibility of TGD events unlikely. Furthermore, the presence in the cluster defined by the PKS An0523 of a DDE1 transposon-related protein, An0532, predictably containing a Tc5 transposase DNA-binding domain (IPR006600), a DDE superfamily endonuclease domain (IPR004875) and a Homeodomain-like domain (IPR009057), make duplication via the action of transposable elements plausible. Overall, quantification of the importance of each duplication mechanism in the expansion of Zn2Cys6-type TFs in the *A. nidulans* genome is beyond the scope of this work and will require deeper bioinformatic analyses, but these observations support that DDs and transposon-mediated duplication mechanisms are more relevant than others.

The most likely possibility after duplication is pseudogenization, which stipulates the complete loss of one copy with the other copy persisting under purifying selection (Lallemand et al. 2020). The pseudogene would then be free to accumulate deleterious mutations, as could be the case of *An0737* or *An1887* (which are probably duplications of *An1518* and *An10300*, respectively; pseudogenes were removed from the initial list of TF-coding genes in Table S1). However, multiple duplications are fixed and maintained in genomes, and different models have been proposed to explain how

neofunctionalization, subfunctionalization or functional redundancy occur (Innan and Kondrashov 2010; Lallemand et al. 2020). Examples of neo-/subfunctionalization, or functional redundancy within Zn2Cys6 paralog clusters could be those of An7610/XlnR (xylan degradation), An0388/AraR (pentose catabolism) and An10550/GalR (galactose conversion), which are involved in the control of carbohydrate metabolism (Tamayo et al. 2008; Kowalczyk et al. 2015), and form a paralog cluster with An9458 and An2672. Also the cases of An1848/RosA and An5170/NosA, both involved in the control of sexual development (Vienken et al. 2005; Vienken and Fischer 2006). A third example could be the paralog cluster composed of An1425/FarB and An7050/FarA (fatty-acid utilization) (Hynes et al. 2006) together with An9117, An3863 and An1077. It has been proposed that the variability in the general structure of CRSs bound by Zn2Cys6 TF-s at target promoters could be a main reason supporting their fixation and expansion in the genomes of Ascomycetes (Shelest 2017). This variability would allow highly specific recognition but at the same time flexibility to produce a large number of different variants. Nevertheless, fixation of paralog genes in a genome should also include the generation of CRSs in their own promoter regions, which would be bound by upstream TFs in order to induce their expression. In this scenario, Figure 3 and Tables S6 indicate that a large proportion of genes encoding Zn2Cys6-type TFs show very low expression levels or remain silent. That includes paralog clusters in which some paralogs are expressed while others remain silent (Table S9). It remains to be determined if this observation reflects the requirement of the corresponding functions under highly specific environmental/culture conditions (as could be the case of regulators of secondary metabolite gene clusters) or, alternatively, that these genes are in a post-duplication stage 1) prior to gains of up- and/or downstream CRSs (fixation) or 2) prior to pseudogenization. In the future, the potential of silent TF-coding genes for GRN rewiring will have to be assessed by comparing the phenotypes of their null or overexpression mutants and by assessing which regulatory connections within GRNs are altered.

The second largest family is that of C2H2-type TFs, which includes several regulators of sexual and asexual development and stress response, such as SteA, MsnA, BrlA, CrzA, PacC, FlbC, SltA, NsdC, AslA or RocA. BrlA, SteA and MsnA are remarkable due to their function in the control of asexual development, sexual development and stress response, respectively, as well as the high conservation within the C2H2 domains

of the amino acids that are predicted to bind DNA (Siggers et al. 2014). Although the high sequence divergence of these three regulators outside the DBDs will in all probability introduce specificities for DNA binding, it could be hypothesized that the DNA target sequence of BrlA could be similar to the consensus target of yeast Msn-like TFs (5´-A/CGGGG-3´) (Siggers et al. 2014). As can be seen, similarities compared to BrlA Response Elements (BREs; 5´-CAAGGGG-3´), defined by (Chang and Timberlake 1993) based on the promoter of *rodA*, are remarkable, although it is unknown if this sequence is the consensus bound by BrlA in all of its target promoters. A plausible possibility is that BrlA, SteA and MsnA compete for the same DNA targets in specific promoters. In correlation with this hypothesis are, on the one hand, the increase in conidia production of the null *msnA/nrdA* strain and the aconidial phenotype of overexpression and multicopy strains (see references within (Dyer and O'Gorman 2012)). On the other hand, the *fluffy* phenotype described after the replacement of the C2H2 domains of BrlA with those of MsnA suggest that there are specificities in CRSs. Protein-DNA binding experiments with BrlA and MsnA constructs in wild-type and the corresponding null genetic backgrounds would render important information on 1) the evolutionary origin of BrlA, 2) its DNA-binding mechanism, and 3) how gene duplication can trigger rewiring of transcriptional networks and the modification of homeostasis/developmental responses. These strategies would contribute to the consolidation of filamentous fungi as reference systems to study GRN rewiring mechanisms.

Methods

Database of the potential TFs encoded by the *A. nidulans* genome

A preliminary list of *A. nidulan*s genes coding for TFs was obtained from (Wortman et al. 2009) and was updated with the addition of genes functionally characterized in the literature as coding for TFs and genes annotated as coding for TFs in the FungiDB database (Basenko et al. 2018). Amino acid sequences of each potential TF were retrieved from the FungiDB database. Genes annotated as pseudogenes and containing stop signals in their predicted coding sequence were not considered.

Bioinformatic analyses

Annotated gene and protein sequence analyses were carried out in FungiDB and AspGD databases. Presence of putative signal peptides was analyzed using Interpro (Mitchell et al. 2019) and SignalP (Almagro Armenteros et al. 2019). Presence of putative transmembrane (TM) domains within protein sequences was analyzed using Interpro, TopCons (Tsirigos et al. 2015), TMHMM (Krogh et al. 2001) and Phobius (Käll et al. 2007).

Protein BLASTs were carried out at the NCBI´s website (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). Protein sequences were aligned using Clustal Omega (Sievers et al. 2011) while clustering of IPDs was carried out using rough clustering at MAFFT, version 7 (Katoh et al. 2019). Phylogenetic trees were generated using MEGAX (Kumar et al. 2018). Phylogenetic (MEGAX) and rough trees (MAFFT) were edited using iTOL (Letunic and Bork 2019) and Archaeopteryx version 0.9921 (Han and Zmasek 2009). Outlier IPDs were removed from the rough tree. Homology searches were carried out in the HMMER web server, which bases biosequence analyses on Hidden Markov Models (Potter et al. 2018). Full fasta sequences of the hits identified by HMMER were retrieved and analyzed in the NCBI Blast site (align two or more sequences) to obtain query coverage values. CD-Hit (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=Server%20home) was used to identify and cluster possible paralogs among Zn2Cys6 and C2H2-type TFs (Li and Godzik 2006)

RNA-seq data analyses

Expression values of each gene annotated as coding for TFs were obtained as FPKM or TPM values from RNA-seq experiments available in the literature and corresponding to different genetic backgrounds and culture conditions. FPKM values corresponding to 1) the use of different nitrogen sources or nitrogen starvation, 2) null *veA* or 3) null *mtfA* genetic backgrounds, or 4) a null *clrB* genetic background were retrieved from the FungiDB database (Sibthorp *et al.*, 2013; Coradetti *et al.*, 2013; Lind *et al.*, 2015). TPM values corresponding to 5) samples comparing alkaline pH (8.0) or sodium stress (1M NaCl) to standard culture conditions (Aspergillus minimal medium or AMM), in both the wild-type and a null *sltA* background, were obtained from (Picazo et al. 2020). FPKM values comparing 6) samples collected before and after the induction of conidiation (vegetative growth, VG, and asexual development, AD), both in a null *flbB*

strain and its isogenic wild-type reference were obtained from (Garzia et al. 2013; Oiartzabal-Arano et al. 2015). Raw RNA-seq data deposited under accession numbers 7) E-MTAB-6996 (null *sclB* background) (Thieme et al. 2018), 8) PRJNA588808 (null *vosA* background) (Kim et al. 2020), 9) GSE72316 (null *osaA* background) (Alkhayyat et al. 2015), and 10) PRJNA294437 (use of waste steam-exploded sugarcane bagasse as a nutrient source) (Brown et al. 2016), were retrieved and processed at the Galaxy website (Afgan et al. 2018). *bam* files were generated from *fastq* files using RNA Star (Dobin et al. 2012), and Stringtie (Pertea et al. 2015) or CuffNorm (Trapnell et al. 2010) were used to generate FPKM-based or TPM-based expression tables. To check the annotation of genes potentially coding for TFs and the processing or non-processing of predicted introns, *bam* files corresponding to RNA-seq data published by us (Garzia et al. 2013; Oiartzabal-Arano et al. 2015; Picazo et al. 2020) were analyzed using Integrative Genome Viewer (IGV; (Robinson et al. 2011)), in combination with the set of transcriptomics data available in the FungiDB database. GO enrichment analysis were carried out using ShinnyGO (Ge et al. 2019). Heatmaps were generated using Heatmapper (Babicki et al. 2016) and Morpheus, a versatile matrix visualization and analysis software developed by the Broad Institute (https://software.broadinstitute.org/morpheus).

Generation of recombinant strains of *A. nidulans*

The Fusion-PCR procedure (Yang et al. 2004) was used to generate linear DNA constructs in which the region coding for the C2H2-type DBD of BrlA (codons for F320-H375) had been replaced by the region coding for the C2H2 DBD of MsnA (codons for T463-H515). The rest of the *brlA* coding sequence (*brlA* has no introns) was maintained as in the wild-type version (codons for M1-Q319 and S376-E432). First, a $brlA::ha_{3x}::pyrG::brlA_{3'\text{-}UTR}$ construct was generated. With that aim, 3 fragments were amplified and fused: 1) the promoter and the coding region of *brlA* (3.6 Kb; oligonucleotides BrlA-PP1 and BrlA-GSP2) (Table S10); 2) the sequence coding for the HA_{3x} tag plus $pyrG^{Afum}$ as selection marker (2.2 Kb; oligonucleotides BrlA-GFP1 and BrlA-GFP2); and 3) 1.5 Kb of the 3´-UTR region of *brlA* (1.6 Kb; oligonucleotides BrlA-GSP3 and BrlA-GSP4).

This construct was used as template in new PCR reactions to amplify and fuse: 1) the promoter of *brlA* and its coding sequence until the codon for Q319 (3.4 Kb;

oligonucleotides BrlA-PP1 and BrlA-ZnFDw; Table S10); 2) the region coding for the C2H2 DBD of MsnA (0.2 Kb; genomic DNA of a wild-type strain as template; oligonucleotides $MsnA_{ZnF}$forBrlAUp and $MsnA_{ZnF}$forBrlADw); and 3) the end of the coding sequence of *brlA* (codons for S376-E432) plus $ha_{3x}::pyrG^{Afum}$ and the 3´-UTR region of *brlA* (3.8 Kb; oligonucleotides BrlA-ZnFUp and BrlA-GSP4).

Fusion-PCR cassettes, $brlA::ha_{3x}::pyrG^{Afum}::brlA_{3'\text{-}UTR}$, used as control to confirm that the HA_{3x} tag did not impede BrlA function, and $brlA_{M1\text{-}Q319}::msnA_{T463\text{-}H515}::brlA_{S376\text{-}E432}::ha_{3x}::pyrG^{Afum}::\ brlA_{3'\text{-}UTR}$, were used to transform protoplasts of the reference strain TN02A3 (*pyrG89; argB2; pyroA4,* Δ*nkuA::argB; veA1*; (Nayak et al. 2006)), following the procedures previously described by (Tilburn et al. 1983; Szewczyk et al. 2007). Transformants hypothetically expressing BrlA::HA_{3x} showed a wild-type conidiating phenotype, while those hypothetically expressing the BrlA-MsnA chimera showed an aconidial phenotype. Genomic DNA of selected transformants was extracted as previously described by us (Garzia et al. 2009), and the maintenance of the reading frame in the *brlA-msnA* chimera was confirmed by Sanger sequencing.

Again, the Fusion-PCR procedure was used to generate synthetic, linear DNA cassettes for the generation of null mutants of *An10192* (strain BD1421), *An1500* (strain BD1423), *An2001/nsiD* (strain BD1425), *An4324* (strain BD1427) and *An4527* (strain BD1430). Three fragments were amplified and fused in each case: 1) 1.5 Kb of the 5´-UTR region of the targeted gene, using genomic DNA as template and oligonucleotides PP1 and PP2 (Table S10); 2) $pyrG^{Afum}$ as the selection marker (oligonucleotides SMP1 and GFP2, and a plasmid bearing the selection marker as template); and 3) the 3´-UTR region of the targeted gene, using oligonucleotides GSP3 and GSP4 and again genomic DNA as the template. Each cassette was used to transform protoplasts of strain TN02A3. After extraction of DNA of candidate transformants, diagnostic PCRs using oligonucleotide pairs sPP1/sGSP4, sPP1/GFP2 or SMP1/sGSP4 were carried out to confirm homologous recombination at the targeted *loci*.

Author Contribution

O.E. conceived and carried out the experimental part, wrote and reviewed the manuscript.

Acknowledgements

Work at O.E.´s lab was supported by the University of the Basque Country (GIU19/014 to O.E.) and the Basque Government (PIBA-PUE, PIBA_2020_1_0032, to O.E., and Elkartek, KK-2019/00076, to Prof. María Teresa Dueñas). I want to acknowledge the contribution of Drs. Eduardo A. Espeso (CIB-CSIC) and Marc S. Cortese to the improvement of this work.

Conflict of Interest

The author declares no conflict of interest.

References

Adams TH, Wieser JK, Yu J-H. 1998. Asexual sporulation in *Aspergillus nidulans*. *Microbiol. Mol. Biol. Rev.* 62:35–54.

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, et al. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46:W537–W544.

Alkhayyat F, Ni M, Kim SC, Yu J-H. 2015. The WOPR domain protein OsaA orchestrates development in *Aspergillus nidulans*. *PLoS One* 10:e0137554.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37:420–423.

Arst HN, Hondmann DHA, Visser J. 1990. A translocation activating the cryptic nitrogen regulation gene *areB* inactivates a previously unidentified gene involved in glycerol utilisation in *Aspergillus nidulans*. *Mol. Gen. Genet. MGG* 223:134–137.

Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS. 2016. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 44:W147–W153.

Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* 11:1005–1017.

Bao Y, Xu F, Shimeld SM. 2017. Phylogenetics of lophotrochozoan bHLH genes and the evolution of lineage-specific gene duplicates. *Genome Biol. Evol.* 9:869–886.

Basenko E, Pulman J, Shanmugasundram A, Harb O, Crouch K, Starns D, Warrenfeltz S, Aurrecoechea C, Stoeckert C, Kissinger J, et al. 2018. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. *J. Fungi* 4:39.

Bat-Ochir C, Kwak J-Y, Koh S-K, Jeon M-H, Chung D, Lee Y-W, Chae S-K. 2016.

The signal peptide peptidase SppA is involved in sterol regulatory element-binding protein cleavage and hypoxia adaptation in *Aspergillus nidulans*. *Mol. Microbiol.* 100:635–655.

Boylan MT, Mirabito PM, Willett CE, Zimmerman CR, Timberlake WE. 1987. Isolation and physical characterization of three essential conidiation genes from *Aspergillus nidulans*. *Mol. Cell. Biol.* 7:3113–3118.

Britton CS, Sorrells TR, Johnson AD. 2020. Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit. *Science (80-.).* 367:96 LP-100.

Brown NA, Ries LNA, Reis TF, Rajendran R, Corrêa dos Santos RA, Ramage G, Riaño-Pachón DM, Goldman GH. 2016. RNAseq reveals hydrophobins that are involved in the adaptation of *Aspergillus nidulans* to lignocellulose. *Biotechnol. Biofuels* 9:145.

Busby TM, Miller KY, Miller BL. 1996. Suppression and enhancement of the *Aspergillus nidulans medusa* mutation by altered dosage of the *bristle* and *stunted* Genes. *Genetics* 143:155 LP-163.

Caddick MX, Brownlee AG, Arst HN. 1986. Phosphatase regulation in *Aspergillus nidulans*: responses to nutritional starvation. *Genet. Res.* 47:93–102.

Caruso ML, Litzka O, Martic G, Lottspeich F, Brakhage AA. 2002. Novel Basic-region Helix–Loop–Helix transcription factor (AnBH1) of *Aspergillus nidulans* counteracts the CCAAT-binding complex AnCF in the promoter of a penicillin biosynthesis gene. *J. Mol. Biol.* 323:425–439.

Chang YC, Timberlake WE. 1993. Identification of *Aspergillus brlA* response elements (BREs) by genetic selection in yeast. *Genetics* 133:29 LP-38.

Coradetti ST, Xiong Y, Glass NL. 2013. Analysis of a conserved cellulase transcriptional regulator reveals inducer-independent production of cellulolytic enzymes in *Neurospora crassa*. *Microbiologyopen* 2:595–609.

Coradetti ST, Xiong Y, Glass NL. 2013. Analysis of a conserved cellulase

transcriptional regulator reveals inducer-independent production of cellulolytic enzymes in *Neurospora crassa*. *Microbiologyopen* 2:595–609.

Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, Villalobos-Escobedo JM, Grimwood J, Álvarez MI, Avalos J, Bauer D, et al. 2016. Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Dowzer CE, Kelly JM. 1991. Analysis of the *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. *Mol. Cell. Biol.* 11:5701 LP-5709.

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

Etxebeste O, Espeso EA. 2020. *Aspergillus nidulans* in the post-genomic era: a top-model filamentous fungus for the study of signaling and homeostasis mechanisms. *Int. Microbiol.* 23:5–22.

Etxebeste O, Markina-Iãrrairaegui A, Garzia A, Herrero-García E, Ugalde U, Espeso EA. 2009. Kapl, a non-essential member of the Pse1p/Imp5 karyopherin family, controls colonial and asexual development in *Aspergillus nidulans*. *Microbiology* 155:3934–3945.

Etxebeste O, Otamendi A, Garzia A, Espeso EA, Cortese MS. 2019. Rewiring of transcriptional networks as a major event leading to the diversity of asexual multicellularity in fungi. *Crit. Rev. Microbiol.* 45:548–563.

Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert Jr. CJ. 2011. Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. *Curr. Protoc. Bioinforma.* 35:6.12.1-6.12.19.

Garzia A, Etxebeste O, Herrero-Garcia E, Fischer R, Espeso EA, Ugalde U. 2009. *Aspergillus nidulans* FlbE is an upstream developmental activator of conidiation

functionally associated with the putative transcription factor FlbB. *Mol. Microbiol.* 71:172–184.

Garzia A, Etxebeste O, Herrero-García E, Ugalde U, Espeso EA. 2010. The concerted action of bZip and cMyb transcription factors FlbB and FlbD induces *brlA* expression and asexual development in *Aspergillus nidulans*. *Mol. Microbiol.* 75:1314–1324.

Garzia A, Etxebeste O, Rodríguez-Romero J, Fischer R, Espeso EA, Ugalde U. 2013. Transcriptional changes in the transition from vegetative cells to asexual development in the model fungus *Aspergillus nidulans*. *Eukaryot. Cell* 12:311–321.

Ge SX, Jung D, Yao R. 2019. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*.

Gerke J, Braus GH. 2014. Manipulation of fungal development as source of novel secondary metabolites for biotechnology. *Appl Microbiol Biotechnol* 98.

Han M V, Zmasek CM. 2009. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10:356.

Hawksworth DL, Lücking R. 2017. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol. Spectr*. 5.

Hynes MJ, Murray SL, Duncan A, Khew GS, Davis MA. 2006. Regulatory genes controlling fatty acid catabolism and peroxisomal functions in the filamentous fungus *Aspergillus nidulans*. *Eukaryot. Cell* 5:794 LP-805.

Igler C, Lagator M, Tkačik G, Bollback JP, Guet CC. 2018. Evolutionary potential of transcription factors for gene regulatory rewiring. *Nat. Ecol. Evol.* 2:1633–1643.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11:97–108.

Käll L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35:W429–W432.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20:1160–1166.

Khaldi N, Wolfe KH. 2008. Elusive origins of the extra genes in *Aspergillus oryzae*. *PLoS One* 3:e3036.

Kim M-J, Lee M-K, Pham HQ, Gu MJ, Zhu B, Son S-H, Hahn D, Shin J-H, Yu J-H, Park H-S, et al. 2020. The velvet Regulator VosA Governs Survival and Secondary Metabolism of Sexual Spores in *Aspergillus nidulans*. *Genes (Basel).* 11:103.

Kiss E, Hegedis B, Varga T, Merenyi Z, Koszo T, Balint B, Prasanna AN, Krizsan K, Riquelme M, Takeshita N, et al. 2019. Comparative genomics reveals the origin of fungal hyphae and multicellularity. *Nat. Commun.* 10:4080.

Kitano H. 2004. Biological robustness. *Nat. Rev. Genet.* 5:826–837.

Kowalczyk JE, Gruben BS, Battaglia E, Wiebenga A, Majoor E, de Vries RP. 2015. Genetic interaction of *Aspergillus nidulans galR*, *xlnR* and *araR* in regulating D-galactose and L-arabinose release and catabolism gene expression. *PLoS One* 10:e0143200.

Krizsán K, Almási É, Merényi Z, Sahu N, Virágh M, Kószó T, Mondo S, Kiss B, Bálint B, Kües U, et al. 2019. Transcriptomic atlas of mushroom development reveals conserved genes behind complex multicellularity in fungi. *Proc. Natl. Acad. Sci.*:201817822.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes Edited by F. Cohen. *J. Mol. Biol.* 305:567–580.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35:1547–1549.

Kwon N-J, Garzia A, Espeso EA, Ugalde U, Yu J-H. 2010. FlbC is a putative nuclear C2H2 transcription factor regulating development in *Aspergillus nidulans*. *Mol.*

Microbiol. 77:1203–1219.

Lallemand T, Leduc M, Landès C, Rizzon C, Lerat E. 2020. An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes (Basel).* 11:1046.

Lee M-K, Kwon N-J, Choi JM, Lee I-S, Jung S, Yu J-H. 2014. NsdD is a key repressor of asexual development in *Aspergillus nidulans*. *Genetics* 197:159–173.

Lee M-K, Kwon N-J, Lee I-S, Jung S, Kim S-C, Yu J-H. 2016. Negative regulation and developmental competence in *Aspergillus*. *Sci. Rep.* 6:28874.

Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47:W256–W259.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.

Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, Stajich JE, Spatafora JW, Groenewald M, Dunn CW, Hittinger CT, et al. 2020. A genome-scale phylogeny of Fungi; insights into early evolution, radiations, and the relationship between taxonomy and phylogeny. *bioRxiv*:2020.08.23.262857.

Lind AL, Wisecaver JH, Smith TD, Feng X, Calvo AM, Rokas A. 2015. Examining the evolution of the regulatory circuit controlling secondary metabolism and development in the fungal genus *Aspergillus*. *PLOS Genet.* 11:e1005096.

Marshall MA, Timberlake WE. 1991. *Aspergillus nidulans wetA* activates spore-specific gene expression. *Mol. Cell. Biol.* 11:55–62.

Mead ME, Borowsky AT, Joehnk B, Steenwyk JL, Shen X-X, Sil A, Rokas A. 2020. Recurrent loss of *abaA*, a master regulator of asexual development in filamentous fungi, correlates with changes in genomic and morphological Traits. *Genome Biol. Evol.* 12:1119–1130.

de Mendoza A, Sebé-Pedrós A. 2019. Origin and evolution of eukaryotic transcription factors. *Curr. Opin. Genet. Dev.* 58–59:25–32.

de Mendoza A, Sebé-Pedrós A, Šestak MS, Matejčić M, Torruella G, Domazet-Lošo T, Ruiz-Trillo I. 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc. Natl. Acad. Sci.* 110:E4858 LP-E4866.

Miller KY, Wu J, Miller BL. 1992. StuA is required for cell pattern formation in *Aspergillus*. *Genes Dev.* 6:1770–1782.

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, et al. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47:D351–D360.

Nayak T, Szewczyk E, Oakley CE, Osmani A, Ukil L, Murray SL, Hynes MJ, Osmani SA, Oakley BR. 2006. A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics* 172:1557–1566.

Nguyen M, Ekstrom A, Li X, Yin Y. 2015. HGT-Finder: A New Tool for Horizontal Gene Transfer Finding and Application to *Aspergillus genomes*. *Toxins (Basel).* 7:4035–4053.

Nocedal I, Johnson AD. 2015. How Transcription Networks Evolve and Produce Biological Novelty. *Cold Spring Harb. Symp. Quant. Biol.* 80:265–274.

Noman N, Monjo T, Moscato P, Iba H. 2015. Evolving robust gene regulatory networks. *PLoS One* 10:e0116258.

Nützmann H-W, Scazzocchio C, Osbourn A. 2018. Metabolic Gene Clusters in Eukaryotes. *Annu. Rev. Genet.* 52.

Oiartzabal-Arano E, Garzia A, Gorostidi A, Ugalde U, Espeso EA, Etxebeste O. 2015. Beyond asexual development: Modifications in the gene expression profile caused by the absence of the *Aspergillus nidulans* transcription factor FlbB. *Genetics* 199:1127–1142.

Ojeda-López M, Chen W, Eagle CE, Gutiérrez G, Jia WL, Swilaiman SS, Huang Z, Park H-S, Yu J-H, Cánovas D, et al. 2018. Evolution of asexual and sexual

reproduction in the Aspergilli. *Stud. Mycol.* 91:37–59.

Pandit SS, Lohmar MJ, Ahmed S, Etxebeste O, Espeso AE, Calvo MA. 2018. UrdA controls secondary metabolite production and the balance between asexual and sexual development in *Aspergillus nidulans*. *Genes (Basel).* 9.

Park J, Wang HH. 2018. Systematic and synthetic approaches to rewire regulatory networks. *Curr. Opin. Syst. Biol.* 8:90–96.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33:290–295.

Picazo I, Etxebeste O, Requena E, Garzia A, Espeso EA. 2020. Defining the transcriptional responses of *Aspergillus nidulans* to cation/alkaline pH stress and the role of the transcription factor SltA. *Microb. Genomics* 6:e000415.

Pires N, Dolan L. 2010a. Early evolution of bHLH proteins in plants. *Plant Signal. Behav.* 5:911–912.

Pires N, Dolan L. 2010b. Origin and diversification of Basic-Helix-Loop-Helix proteins in plants. *Mol. Biol. Evol.* 27:862–874.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. *Nucleic Acids Res.* 46:W200–W204.

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20:38.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29:24–26.

Rokas A, Wisecaver JH, Lind AL. 2018. The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.*

Salzberg SL. 2019. Next-generation genome annotation: we still struggle to get it right. *Genome Biol.* 20:92.

Shelest E. 2017. Transcription Factors in fungi: TFome dynamics, three major families, and dual-specificity TFs. *Front. Genet.* 8:53.

Sibthorp C, Wu H, Cowley G, Wong PWH, Palaima P, Morozov IY, Weedall GD, Caddick MX. 2013. Transcriptome analysis of the filamentous fungus *Aspergillus nidulans* directed to the global identification of promoters. *BMC Genomics* 14:847.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539.

Siggers T, Reddy J, Barron B, Bulyk ML. 2014. Diversification of Transcription Factor Paralogs via Noncanonical Modularity in C2H2 Zinc Finger DNA Binding. *Mol. Cell* 55:640–648.

Skromne I, Sanchez O, Aguirre J. 1995. Starvation stress modulates the expression of the *Aspergillus nidulans brlA* regulatory gene. *Microbiology* 141:21–28.

Son S-H, Son Y-E, Cho H-J, Chen W, Lee M-K, Kim L-H, Han D-M, Park H-S. 2020. Homeobox proteins are essential for fungal differentiation and secondary metabolism in *Aspergillus nidulans*. *Sci. Rep.* 10:6094.

Sorrells TR, Johnson AD. 2015. Making Sense of Transcription Networks. *Cell* 161:714–723.

Spatafora JW, Aime MC, Grigoriev I V., Martin F, Stajich JE, Blackwell M. 2017. The fungal tree of life: from molecular systematics to genome-scale phylogenies. *Microbiol. Spectr.* 5.

Strittmatter AW, Irniger S, Braus GH. 2001. Induction of *jlbA* mRNA synthesis for a putative bZIP protein of *Aspergillus nidulans* by amino acid starvation. *Curr. Genet.* 39:327–334.

Szewczyk E, Nayak T, Oakley CE, Edgerton H, Xiong Y, Taheri-Talesh N, Osmani SA, Oakley BR. 2007. Fusion PCR and gene targeting in *Aspergillus nidulans*. *Nat. Protoc.* 1:3111–3120.

Tamayo EN, Villanueva A, Hasper AA, Graaff LH de, Ramón D, Orejas M. 2008. CreA mediates repression of the regulatory gene xlnR which controls the production of xylanolytic enzymes in *Aspergillus nidulans*. *Fungal Genet. Biol.* 45:984–993.

Thieme KG, Gerke J, Sasse C, Valerius O, Thieme S, Karimi R, Heinrich AK, Finkernagel F, Smith K, Bode HB, et al. 2018. Velvet domain protein VosA represses the zinc cluster transcription factor SclB regulatory network for *Aspergillus nidulans* asexual development, oxidative stress response and secondary metabolism. *PLOS Genet.* 14:e1007511.

Tilburn J, Scazzocchio C, Taylor GG, Zabicky-Zissman JH, Lockington RA, Davies RW. 1983. Transformation by integration in *Aspergillus nidulans*. *Gene* 26:205–221.

Todd RB, Andrianopoulos A. 1997. Evolution of a fungal regulatory gene family: The Zn(II)2Cys6 binuclear cluster DNA binding motif. *Fungal Genet. Biol.* 21:388–405.

Todd RB, Zhou M, Ohm RA, Leeggangers HACF, Visser L, de Vries RP. 2014. Prevalence of transcription factors in ascomycete and basidiomycete fungi. *BMC Genomics* 15:214.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515.

Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* 43:W401–W407.

Tüncher A, Reinke H, Martic G, Caruso Maria L, Axe lA. B. 2004. A basic-region helix–loop–helix protein-encoding gene (*devR*) involved in the development of *Aspergillus nidulans*. *Mol. Microbiol.* 52:227–241.

Vallim MA, Miller KY, Miller BL. 2002. *Aspergillus* SteA (Sterile12-like) is a homeodomain-C2/H2-Zn+2 finger transcription factor required for sexual

reproduction. *Mol. Microbiol.* 36:290–301.

Verd B, Monk NAM, Jaeger J. 2019. Modularity, criticality, and evolvability of a developmental gene regulatory network.Altenberg L, Wittkopp PJ, editors. *Elife* 8:e42832.

Vienken K, Fischer R. 2006. The Zn(II)2Cys6 putative transcription factor NosA controls fruiting body formation in *Aspergillus nidulans*. *Mol. Microbiol.* 61:544–554.

Vienken K, Scherer M, Fischer R. 2005. The $Zn(II)_2Cys_6$ putative *Aspergillus nidulans* transcription factor repressor of sexual development inhibits sexual development under low-carbon conditions and in submersed culture. *Genetics* 169:619 LP-630.

Voordeckers K, Pougach K, Verstrepen KJ. 2015. How do regulatory networks evolve and expand throughout evolution? *Curr. Opin. Biotechnol.* 34:180–188.

de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, Anderluh G, Asadollahi M, Askin M, Barry K, et al. 2017. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol.* 18:28.

Wortman JR, Gilsenan JM, Joardar V, Deegan J, Clutterbuck J, Andersen MR, Archer D, Bencina M, Braus G, Coutinho P, et al. 2009. The 2008 update of the *Aspergillus nidulans* genome annotation: a community effort. *Fungal Genet Biol* 46.

Yang L, Ukil L, Osmani A, Nahm F, Davies J, De Souza CPC, Dou X, Perez-Balaguer A, Osmani SA. 2004. Rapid production of gene replacement constructs and generation of a green fluorescent protein-tagged centromeric marker in *Aspergillus nidulans*. *Eukaryot. Cell* 3:1359–1362.

Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18:292–298.

Figures

<u>Figure 1</u>: Potential TFome of the model filamentous ascomycete *A. nidulans*: A) Pie charts showing the number and percentage of TFs (left; of a total of 520) and IPDs (right; of a total of 923) of each family potentially encoded by the genome of *A. nidulans*. Color keys are included. Both tables are an update of the list published by Wortman and colleagues (Wortman et al. 2009). B) Draft tree generated by MAFFT and edited using iTOL showing the enrichment of the potential TFome of *A. nidulans* in zinc clusters (light orange clade) and Fungal-specific TF (dark orange clade) IPDs, followed by C2H2-type IPDs (dark grey clade). The three pie charts indicate the number of genes coding for TFs with a standard name (top right), the number of potential TFs predicted to include a signal peptide in their sequences (middle right), or those predicted to contain TM domains (bottom right). See also Tables S1, S2 and S3.

<u>Figure 2</u>: **Comparison of annotation of genes potentially encoding TFs and RNA-seq data**. A) Pie chart showing the approximate number and percentage of TF-coding genes correctly annotated according to RNA-seq data, the fraction of those that should be less or more extensively re-annotated and the fraction of those that could not be assessed (mostly due to their very low or null expression levels). Approximately 30.2% of the TF-coding genes would need some re-annotation. B) Null expression levels of *An3391*, predictably encoding a C2H2- plus a Zn2Cys6-type dual-specificity TF, in comparison to those of *An3390*, coding for a putative pectinesterase, as an example of genes the annotation of which could not be assessed. Data from our RNA-seq experiments (visualized with IGV; (Garzia et al. 2013; Oiartzabal-Arano et al. 2015; Picazo et al. 2020)) and those provided by the FungiDB database were analyzed (Sibthorp et al. 2013). C) Expression levels of *An2290/steA* as an example of a gene correctly annotated. RNA-seq experiments retrieved from the FungiDB database corresponded to complete and minimal media (Sibthorp et al. 2013), carbon starvation and a null *clrB* mutant (Coradetti *et al.*, 2013), and null *mtfA* or null *veA* backgrounds (Lind et al. 2015). D) Alternative splicing event, represented by the case of *An1569*. Annotation of the coding region is correct but there are intron-retention (IR) events in both predicted introns, and, furthermore, the non-processed form is the most abundant. Retention of the first intronic sequence leads to a premature stop codon that generates a peptide containing only the Zn2Cys6 IPD (amplification of the dotted orange square in the bottom panel). The concentration of the full-length form will be, in all probability,

very low. E-F) Examples of genes whose coding sequences should be re-annotated in 5´ (*An11098* in panel E) and 3´ (*An2782* in panel F) ends. The dotted orange squares indicate the regions amplified in bottom panels. There is also an IR event in the fourth predicted intron of *An11098*, adding 17 extra codons between those coding for amino acids D399 and G400 but without modifying the reading frame. G-I) Examples of genes which would need major re-annotation of their coding sequences. IR of the first intron of *An1812/jlbA* (G) leads to a premature stop. The probable start codon will be that for Met175, retaining only the sequence encoding the bZIP domain. Probable IR of the first two predicted introns of *An9221* (H) would lead to the loss of the region coding for a predicted signal peptide and six transmembrane domains. The coding region will probably begin with the codon for Met246. IR in the case of *An5252* (or *An9240*) (I), leading probably to two coding sequences (those covering M1-G90 plus 9 codons and M127-D168 plus 20 codons). The first C2H2 domain (dark grey square bold in red) would be lost while the second coding region would include the second C2H2 IPD. St: start codon. Sp: stop codon. pSp: predicted stop codon. RF: Reading frame. See also Table S4.

Figure 3: Expression pattern of each TF-coding gene of each IPD family under different culture conditions and in genetic backgrounds. Heatmaps were not clustered so as to observe if there could be any general expression pattern associated to a specific family of IPDs. Colors in the left bar indicate each IPD family and follow the same color-key as in the previous figures and tables. The extension of each color-bar reflects the predicted number of genes within each family of IPDs (each row in the heatmaps). Heatmaps 1 to 10 show variation of expression 1) with the use of sugarcane bagasse (FPKM values), 2) in a null *clrB* mutant (FPKM), 3) a null *flbB* mutant (FPKM) before (VG) and after (AD) the induction of conidiation, 4) a null *mtfA* background (FPKM), 5) nitrogen starvation (FPKM), 6) a null *osaA* background (FPKM), 7) a null *sltA* background (under standard culture conditions, pH 8, or addition of 1M NaCl) (TPM), 8) a null *sclB* background (TPM), 9) a null *veA* background (FPKM) and 10) a null *vosA* background (FPKM). See references in the Materials and Methods section. All FPKM/TPM values below 5 were removed from the heatmaps and are indicated in grey, in order to highlight those TF-coding genes showing low or null expression levels in most or all genetic backgrounds and culture conditions. Black bars below the heatmaps indicate genetic backgrounds and culture conditions in which a

change in the general expression pattern of the *A. nidulans* TFome might be induced. See also Figure S1 and Table S5.

Figure 4: **General modification of the expression pattern of TF-coding genes in response to specific culture conditions or in specific mutant backgrounds**. Clustered heatmaps (Pearson correlation) showing a general, qualitative change in the expression levels of a majority TF-coding genes in response to A) the addition of sugarcane bagasse as the main nutrient source, B) removal of the nitrogen source from the culture medium, and C) deletion of *sltA*. Panels D to F show expression profiles of specific genes selected from the groups highlighted in panels A to C, respectively: D) upregulation in a medium containing sugarcane bagasse; E) upregulation as a response to nitrogen starvation; and F) downregulation in a null *sltA* background in the three conditions tested (standard minimal medium or *unstressed*, alkalinity and sodium stress; up), only at pH 8 (middle) or both at pH 8 and sodium stress (bottom).

Figure 5: Evolution of the bHLH-type TF family of *A. nidulans*. A) Phylogenetic tree showing the conservation of bHLH-type TFs of *A. nidulans* in the family Trichocomaceae. Sequences of predicted bHLH-type TFs in multiple *Aspergilli*, *Penicillium rubens*, *Talaromyces stipitatus* and *T. marneffei* were retrieved from the FungiDB database and aligned using Clustal Omega. The tree was generated using Mega, version 7.0.26 (Maximum Likelihood method and JTT model, with 60 bootstrap replications), and edited using iTOL (see Materials and Methods). The dotted black squares in the An7170 clade highlight the probable existence of two paralogs in *T. marneffei*, *T. stipitatus* and *A. fumigatus*. B-E) Density of reads mapping to *An5078* (B), *An0396* (C), *An2392* (D) and *An2811* (E) *loci*. Images were exported using the IGV software (top; (Picazo *et al.*, 2020)) or from the JBrowse genome browser linked to the FungiDB database (*middle*). Bottom pictures correspond to BLAST searches at the NCBI website (panels B-D) or a Clustal alignment visualized with Genedoc (panel E). The positions of predicted introns and exons, predicted (pSt/pSp) and probable (St/Sp; based on RNA-seq data) start/stop codons, and the regions encoding the bHLH domains are indicated. See HMMER analyses in Table S6.

Figure 6: Duplication events within the zinc cluster family of TFs in *A. nidulans*. A) Phylogenetic tree generated with Mega, version 7.0.26 (Neighbor-joining method and 5,000 bootstrap replications), and edited using iTOL, showing the paralog clusters

predicted by the CD-Hit suite. Predicted paralog clusters of 2, 3, 4, 5 and 6 members are highlighted with brown, green, purple, red and blue lines, respectively. Some Zn2Cys6-type TFs excluded from these paralog clusters (clusters of 1 member in the CD-Hit suite) were also included. Circles indicate those TFs predictably involved or shown to be involved in the control of secondary metabolism while stars indicate TFs having a standard name. B-D) BLAST alignments (NCBI website) for each of the three paralog clusters of six members identified by CD-Hit, using as queries An3385 (B), An11169 (C) and An8590 (D), respectively. Score, expect and coverage values are included. The score color key is shown in panel B. The extension of Zn2Cys6 and Fungal specific regulatory domains is shown in each panel with the black squares. The alignments of the sequences in each panel (B-D) can be seen in Figures S2, S3 and S4, respectively.

Figure 7: Similarities of the C2H2 domains of BrlA with those of specific regulators of sex and stress-response. A) Phylogenetic tree built with the predicted sequences of *A. nidulans* C2H2-type TFs. The Mega software, version 7.0.26 (Neighbor-joining method and 5,000 bootstrap replications), was used to generate the tree, and iTOL to edit it. Clusters of possible paralogs predicted by the CD-Hit suite are highlighted with brown (clusters of 2 members) and green (3 members) lines, respectively. Circles indicate those TFs predictably involved or shown to be involved in the control of secondary metabolism while stars indicated those TFs having a standard name. B) Phylogenetic tree of a subset of C2H2-type Zn Finger domains of specific *A. nidulans* and *S.cerevisiae* (cer) TFs (Neighbor-joining method and 5,000 bootstrap replications). The DBD of BrlA (red arrow) group with Msn-like domains of *S. cerevisiae* (red square) and *A. nidulans* (red square), in a clade located next to that of SteA C2H2 DBD. C) Alignment of the C2H2 DBDs of specific *A. nidulans* and *S.cerevisiae* (cer) TFs. The position and nature of the predicted DNA-binding amino acids (based on (Siggers et al. 2014)) are indicated in green (conserved in all sequences analyzed, including BrlA) or red (not conserved in BrlA). The picture was exported from Genedoc, version 2.7.000, and edited. D) Phenotype of a strain expressing a BrlA::HA_{3x} chimera in which the C2H2 domains of BrlA (residues F320-H375) had been replaced by those of MsnA (residues T463-H515), after 72 hours of culture at 37 ºC in adequately supplemented AMM, and compared to the parental wild-type strain and a strain expressing BrlA::HA_{3x}. The diameter of plates is 5.5 cm.

<u>Figure 8:</u> Nitrogen starvation induces the expression of multiple genes encoding developmental regulators. A) Heatmap showing the expression pattern of genes with a FC ≥ 3 (lower clade) and a FC ≤ 3 (upper clade) when comparing the transcriptomes of samples collected in AMM with ammonium as the nitrogen source (third column) and 72 hours of nitrogen starvation (fifth column). The list of genes and the corresponding FPKM values in each of the five conditions tested were retrieved from the FungiDB database (see Table S7), based on RNA-seq data by Sibthorp and colleagues (Sibthorp et al. 2013). B) Number of genes with a FC ≥ 3 or a FC ≤ 3 under nitrogen starvation (72 h) compared to AMM containing ammonium as the nitrogen source. The ratio of up- versus downregulated genes is 2.84, suggesting that nitrogen starvation induces a transcriptional response. C) Bar plot showing the number of upregulated genes (FC ≥ 3) coding for proteins predicted to contain the listed Interpro domains. Zn2Cys6, Fungal specific (Fungal-sp. TF domain) and C2H2-type transcriptional regulatory domains are the top 1, top 4 and top 7 Interpro domains, respectively. D) Number of TF-coding genes with a FC ≥ 3 or a FC ≤ 3 under nitrogen starvation (72 h) compared to AMM containing ammonium as the nitrogen source (see Table S8). The ratio of up- versus downregulated TF-coding genes is 10.89. E) Bar-plot showing log (FPKM + 1) values for genes encoding known developmental regulators, in AMM with ammonium as the nitrogen source (pink bars), 4 hours (light orange bars) and 72 hours (dark orange bars) after nitrogen starvation was induced. FPKM values were taken from (Sibthorp et al. 2013), and in all the cases shown, expression is induced under nitrogen starvation. F) Same bar-plot but for the genes selected for deletion (due to their induction under nitrogen starvation; see the phenotypes in panel G) and under the five culture conditions analyzed by Sibthorp and colleagues (Sibthorp et al. 2013). G) Phenotypes of the deletion strains of the genes analyzed in F, after 72 (left block of pictures) and 144 (right block) hours of culture at 37 ºC and compared to the parental wild-type strain. Diameter of plates is 5,5 cm. The red squares indicate the colonies and colony regions amplified in the right column of each block of images. Red arrowheads highlight cleistothecia.

Supplementary Material

<u>Table S1</u>: Predicted TFome of *A. nidulans*. The table includes the systematic name of genes coding for those potential TFs, the information about their role given by the FungiDB database, information about their annotation based on RNA-seq data

(see also Table S5), the IPD or IPDs included in their sequence, the coordinates of those IPDs as well as IPD and full-length amino acid sequences.

Table S2: TF-coding genes of *A. nidulans* having a standard name.

Table S3: TFs of *A. nidulans* predicted to include a signal peptide (FungiDB or SignalP) or TM domains (FungiDB, TMHMM and Phobius) in their amino acid sequences.

Table S4: Comparison of gene-annotation data and RNA-seq data for each TF-coding gene. Gene-annotations were checked in the FungiDB database or downloaded from http://www.aspgd.org/download/gff/A_nidulans_FGSC_A4/ as a gff file (A_nidulans_FGSC_A4_version_s10-m04-r16_features_with_chromosome_sequences). RNA-seq data available also in the FungiDB database or *bam* files corresponding to our RNA-seq analyses were used for comparison with gene-annotation data. The color key in column C is the same as in Figure 2A (correct/incorrect annotation) while the color key in columns E and F is the same as in Figure 1 (family of IPD).

Table S5: Expression level of each *A. nidulans* gene potentially coding for a TF. Empty columns separate the different RNA-seq experiments checked in this work. See Materials and Methods and the legend of Figures S1 and 3 for more information and references.

Table S6: HMMER analyses of predicted bHLH-type TFs An5078, An0396, An2392 and An2811. Each sub-table includes the hits returned by HMMER, and in the case of the 50 sequences with the highest scores (20 in the case of An2811), the species they belong to, the scores, the query coverage values, e values, lengths and the first hit of the confirmatory reverse retrieval against the *A. nidulans* genome in the AspGD database. Corrected amino acid sequences were used based on mapping of RNA-seq reads (Picazo et al. 2020).

Table S7: List of genes with a FC ≥ 3 or a FC ≤ 3 when comparing nitrogen starvation conditions (72h) with AMM supplemented with ammonium as the nitrogen source. FPKM values in each of the five culture conditions analyzed by (Sibthorp et al. 2013) **are included.**

<u>Table S8</u>: List of TF-coding genes with a FC ≥ 3 or a FC ≤ 3 when comparing nitrogen starvation conditions (72h) with AMM supplemented with ammonium as the nitrogen source. FPKM values in each of the five culture conditions analyzed by (Sibthorp et al. 2013) **as well as the types of transcriptional regulatory domains encoded by these genes are included.**

<u>Table S9</u>: Genomic location of genes coding for the Zn2Cys6 family TFs identified by CD-Hit. Genes are ordered based on the hypothetic paralog clusters they belong to. Genomic locations of each gene were retrieved from the FungiDB database. Expression levels were retrieved from Table S5. See also Figure S6

<u>Table S10</u>: Oligonucleotides used in this study.

<u>Figure S1</u>: Heatmaps showing the expression patterns of each TF-coding gene of each IPD family under different culture conditions and in genetic backgrounds. Heatmaps were not clustered so as to observe if there could be any general expression pattern associated to a specific family of IPDs. Colors in the left bar indicate each IPD family and follow the same color-key as in the previous figures and tables. The extension of each color-bar reflects the predicted number of genes within each family of IPDs (each row in the heatmaps). Heatmaps 1 to 10 show variation of expression 1) with the use of sugarcane bagasse (FPKM values), 2) in a null *clrB* mutant (FPKM), 3) a null *flbB* mutant (FPKM) before (VG) and after (AD) the induction of conidiation, 4) a null *mtfA* background (FPKM), 5) nitrogen starvation (FPKM), 6) a null *osaA* background (FPKM), 7) a null *sltA* background (under standard culture conditions, pH 8, or addition of 1M NaCl) (TPM), 8) a null *sclB* background (TPM), 9) a null *veA* background (FPKM) and 10) a null *vosA* background (FPKM). See references in the Materials and Methods section. Black bars below heatmaps indicate genetic backgrounds and culture conditions in which a general change in the expression pattern of the *A. nidulans* TFome might be induced. See Table S5.

<u>Figure S2</u>: **Alignment of the predicted aminoacidic sequences of An3385 (used as query in Figure 6B), An7896, An2036, An6788, An1029 and An0533 as probable paralog TFs controlling secondary metabolism**. This cluster of probable paralog TFs

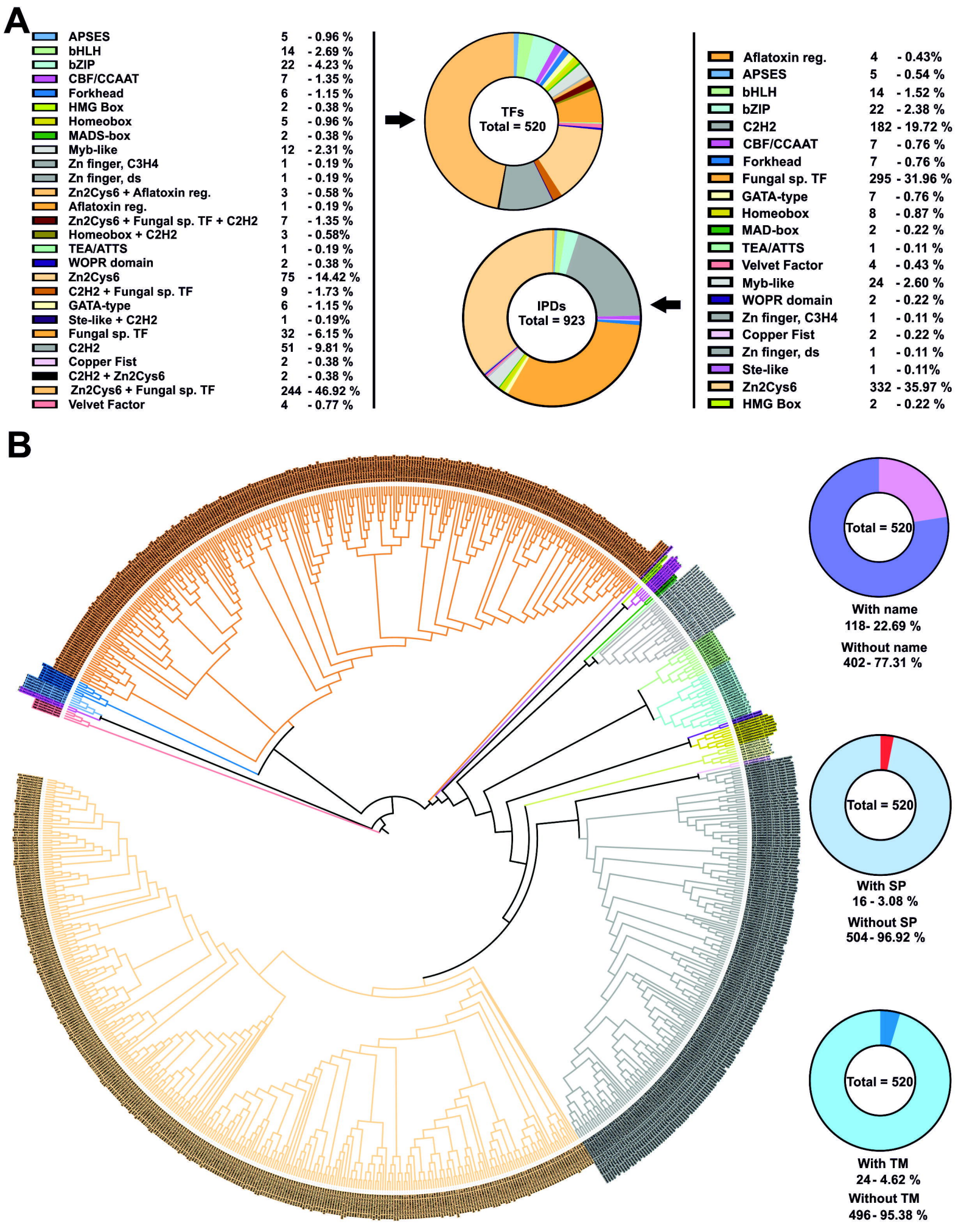
was identified using the CD-Hit suite. The alignment was carried out using Clustal Omega and visualized using Genedoc.

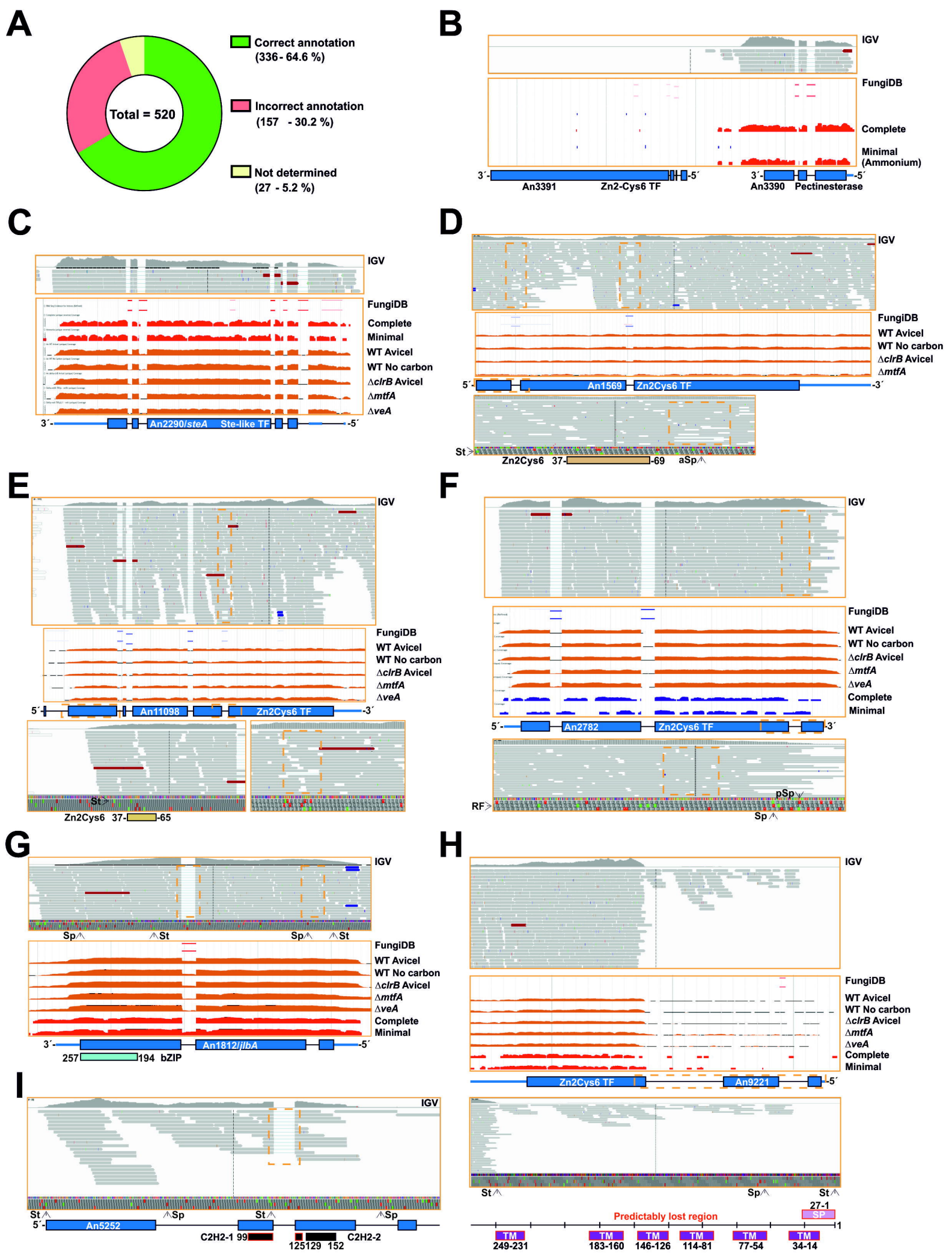
Figure S3: **Alignment of An11169 (used as query in Figure 6C), An10906, An8636, An5955, An2375 and An8535 as a probable paralog cluster of TFs identified using the CD-Hit suite**. Alignment was carried out using Clustal Omega and visualized using Genedoc.

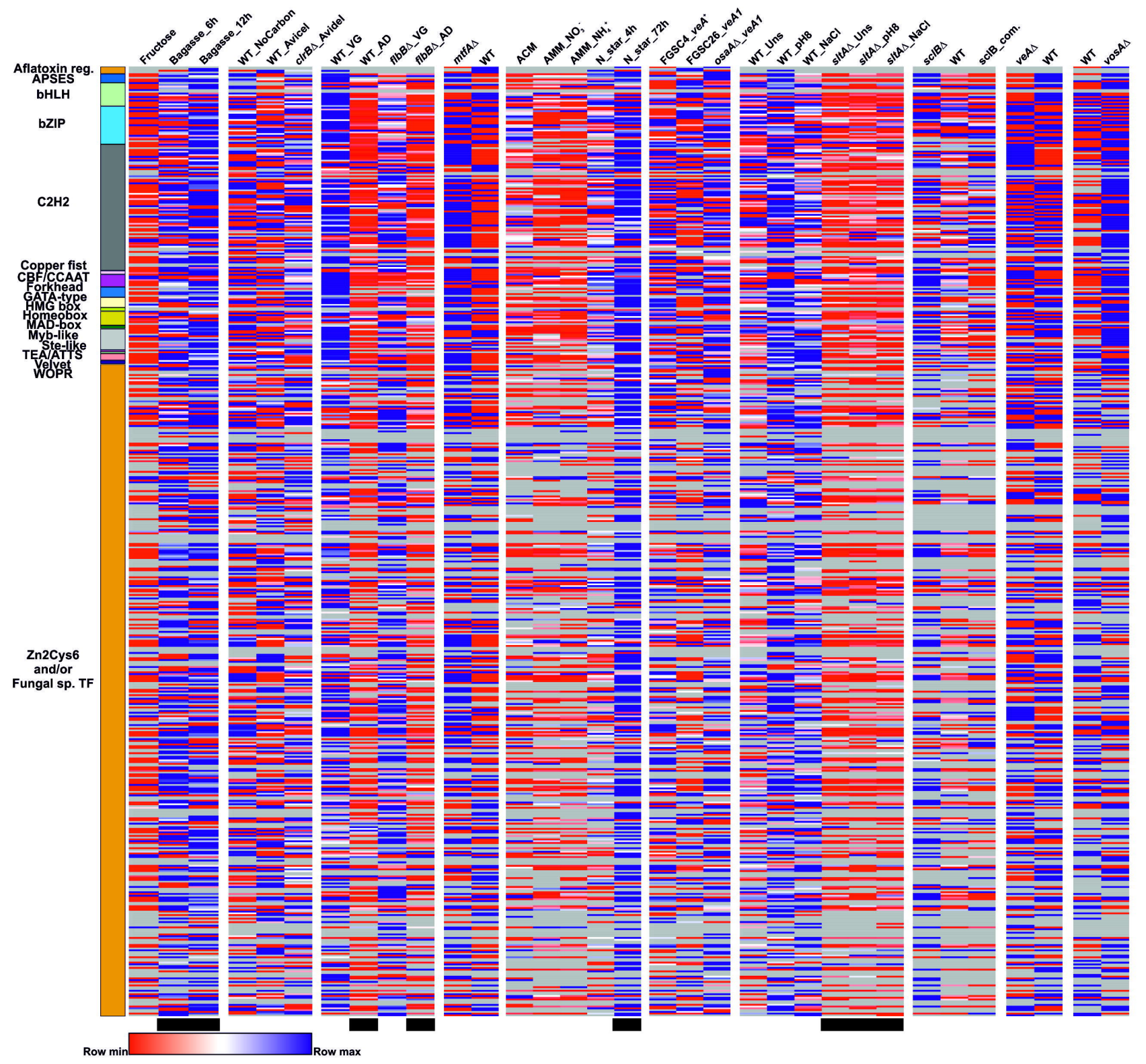
Figure S4: **Alignment of An8590 (used as query in Figure 6D), An7507, An4972, An11185, An8930 and An0742 as a probable paralog cluster of TFs identified using the CD-Hit suite**. Alignment was carried out using Clustal Omega and visualized using Genedoc.

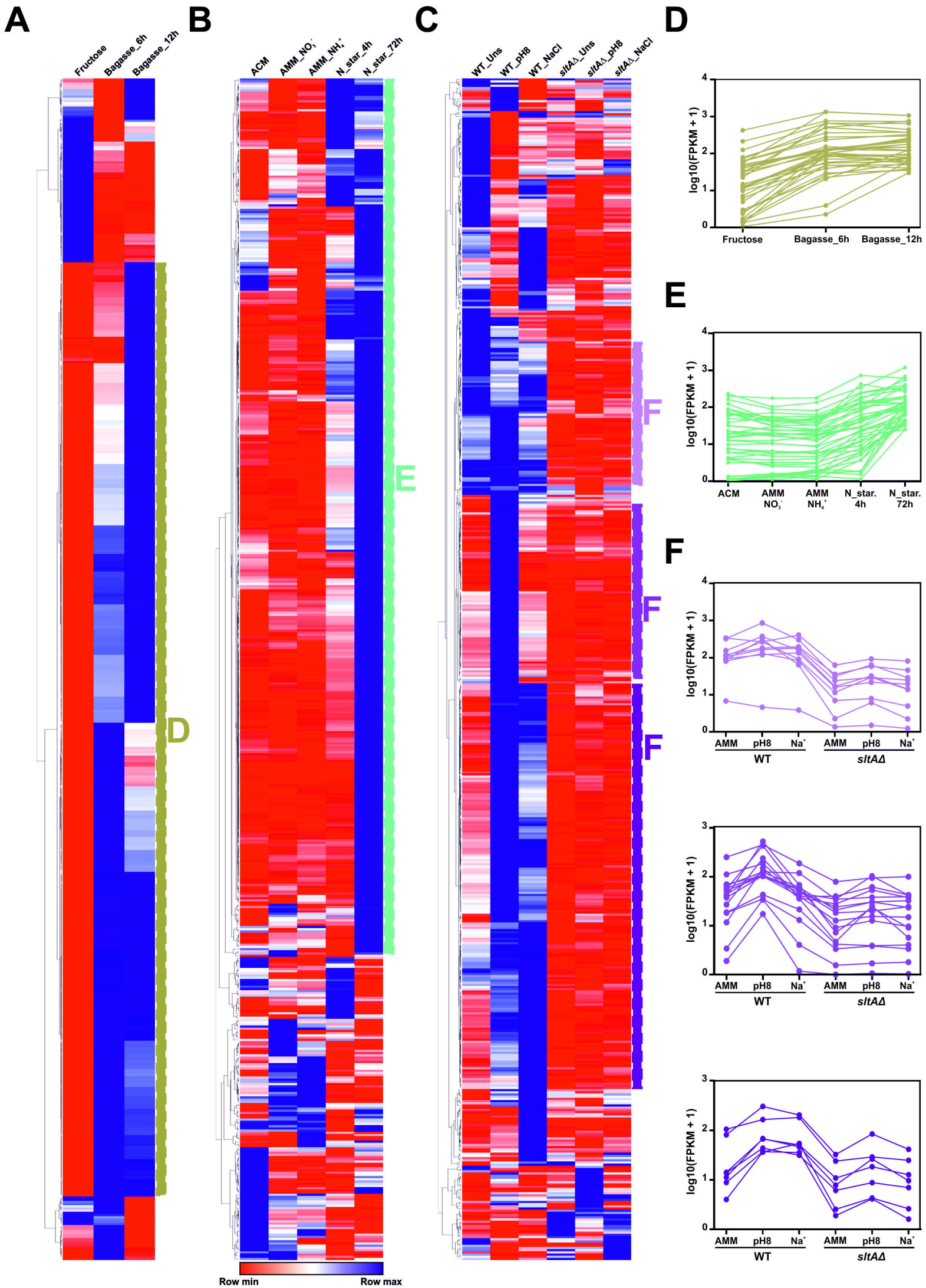
Figure S5: Phenotypes of null strains of TF-coding genes induced under nitrogen starvation. Phenotypes of the null mutants shown in Figure 8G, in ACM or AMM supplemented with ammonium (first and third columns, respectively) or nitrate (second and fourth columns, respectively), after 72 (A) or 144 hours (B) of culture at 37 ºC, and compared to the parental wild-type strain. The phenotypes of the same strains under nitrogen limitation (1/25 of ammonium or nitrate compared to AMM) are shown in columns five and six, respectively. Diameter of plates is 5,5 cm. Red squares indicate the colonies amplified in the images shown in panels C (72 hours of culture) and D (144 hours). Red arrowheads highlight cleistothecia.

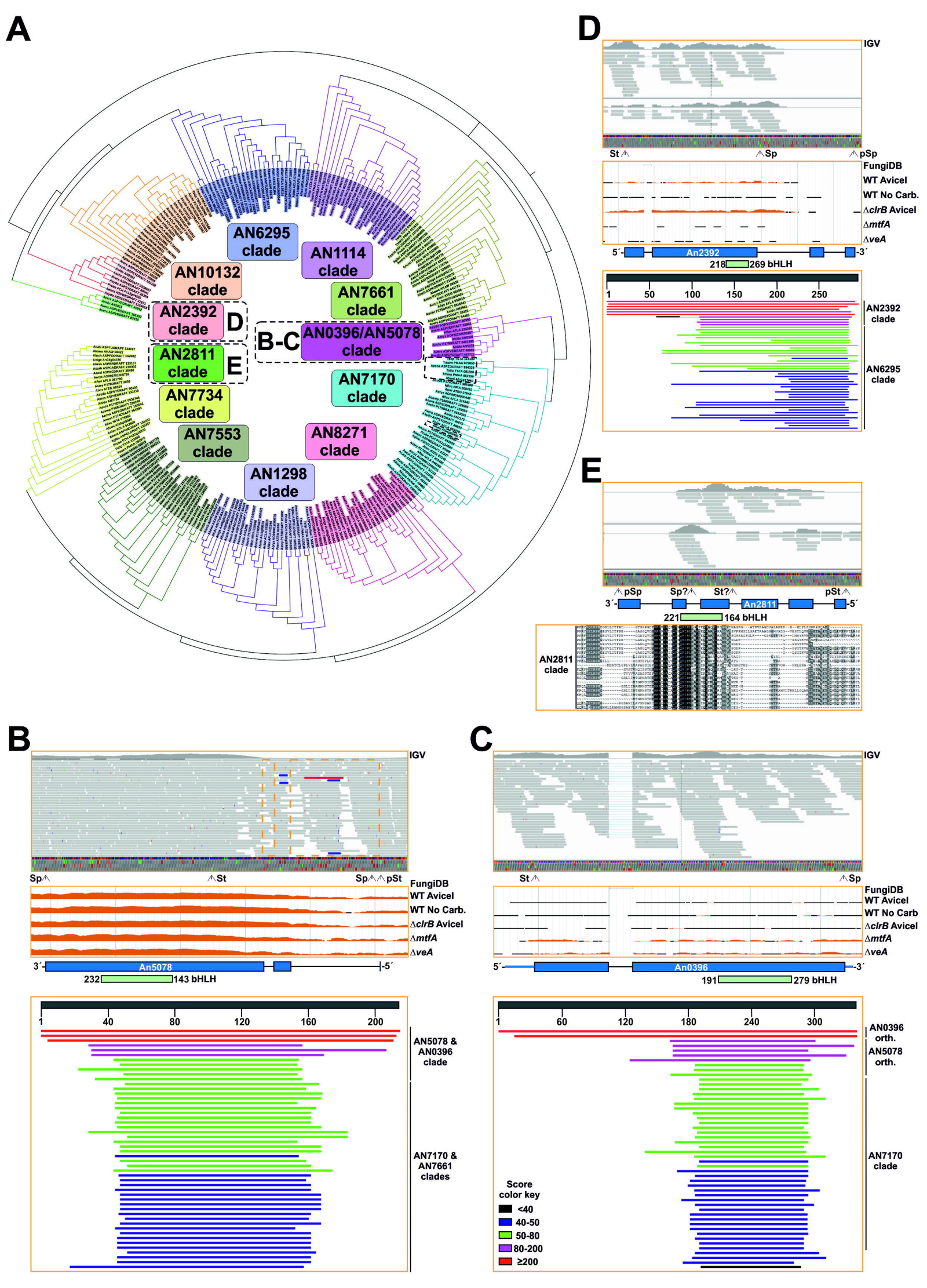
Figure S6: Distribution of genes coding for Zn2Cys6-type TFs in the genome of *A. nidulans*. Those genes coding for Zn2Cys6-type TFs and identified by the CD-Hit suite as members of probable paralog clusters were considered. The image was downloaded from the FungiDB database and edited. The blue color is for genes in the forward strand while the red color is for genes in the reverse strand. See also Table S10.

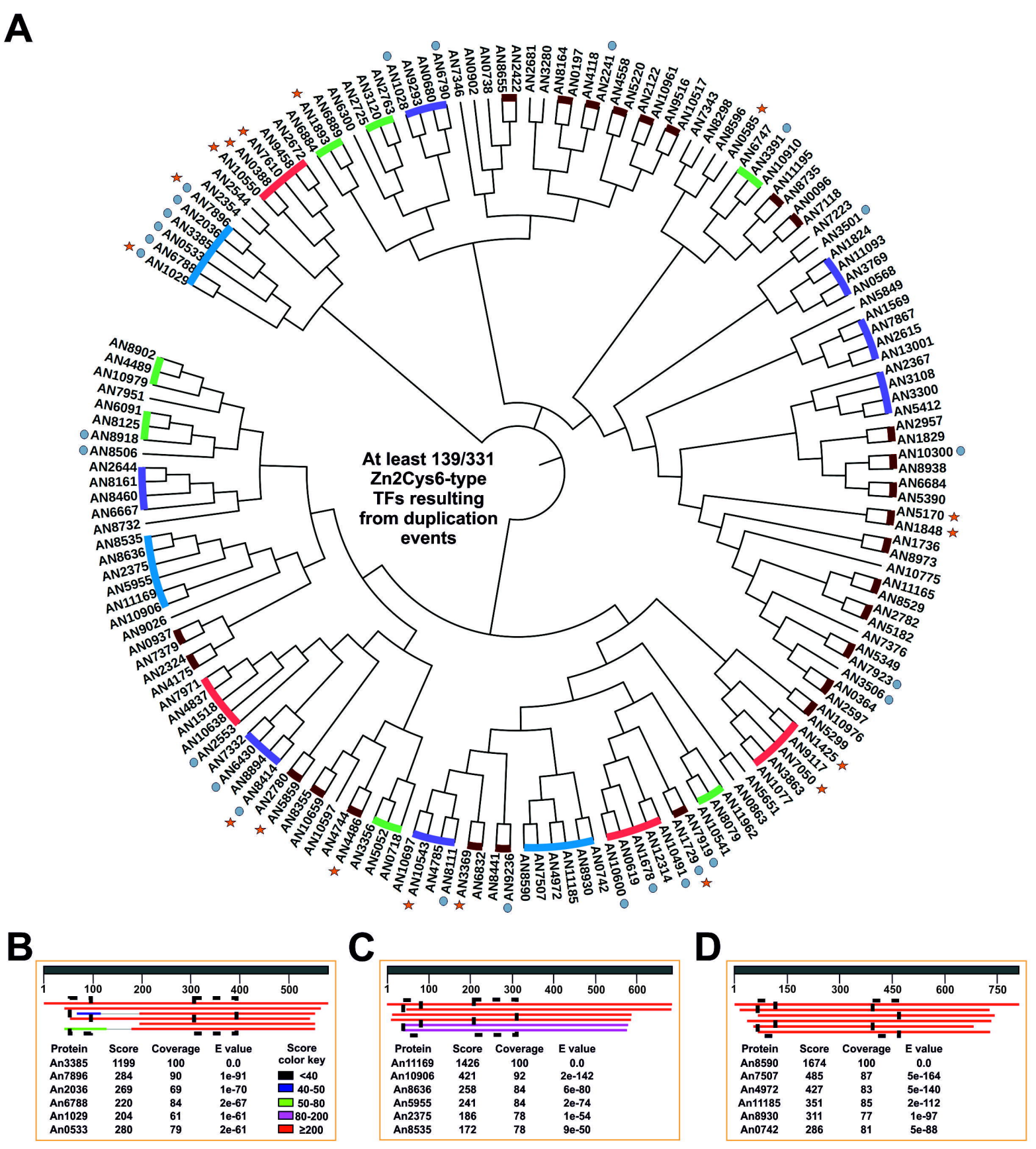

A
APSES 5 - 0.96 %
bHLH 14 - 2.69 %
bZIP 22 - 4.23 %
CBF/CCAAT 7 - 1.35 %
Forkhead 6 - 1.15 %
HMG Box 2 - 0.38 %
Homeobox 5 - 0.96 %
MADS-box 2 - 0.38 %
Myb-like 12 - 2.31 %
Zn finger, C3H4 1 - 0.19 %
Zn finger, ds 1 - 0.19 %
Zn2Cys6 + Aflatoxin reg. 3 - 0.58 %
Aflatoxin reg. 1 - 0.19 %
Zn2Cys6 + Fungal sp. TF + C2H2 7 - 1.35 %
Homeobox + C2H2 3 - 0.58%
TEA/ATTS 1 - 0.19 %
WOPR domain 2 - 0.38 %
Zn2Cys6 75 - 14.42 %
C2H2 + Fungal sp. TF 9 - 1.73 %
GATA-type 6 - 1.15 %
Ste-like + C2H2 1 - 0.19%
Fungal sp. TF 32 - 6.15 %
C2H2 51 - 9.81 %
Copper Fist 2 - 0.38 %
C2H2 + Zn2Cys6 2 - 0.38 %
Zn2Cys6 + Fungal sp. TF 244 - 46.92 %
Velvet Factor 4 - 0.77 %
TFs
Total = 520
IPDs
Total = 923
Aflatoxin reg. 4 - 0.43%
APSES 5 - 0.54 %
bHLH 14 - 1.52 %
bZIP 22 - 2.38 %
C2H2 182 - 19.72 %
CBF/CCAAT 7 - 0.76 %
Forkhead 7 - 0.76 %
Fungal sp. TF 295 - 31.96 %
GATA-type 7 - 0.76 %
Homeobox 8 - 0.87 %
MAD-box 2 - 0.22 %
TEA/ATTS 1 - 0.11 %
Velvet Factor 4 - 0.43 %
Myb-like 24 - 2.60 %
WOPR domain 2 - 0.22 %
Zn finger, C3H4 1 - 0.11 %
Copper Fist 2 - 0.22 %
Zn finger, ds 1 - 0.11 %
Ste-like 1 - 0.11%
Zn2Cys6 332 - 35.97 %
HMG Box 2 - 0.22 %
B
Total = 520
With name
118- 22.69 %
Without name
402- 77.31 %
Total = 520
With SP
16 - 3.08 %
Without SP
504 - 96.92 %
Total = 520
With TM
24 - 4.62 %
Without TM
496 - 95.38 %

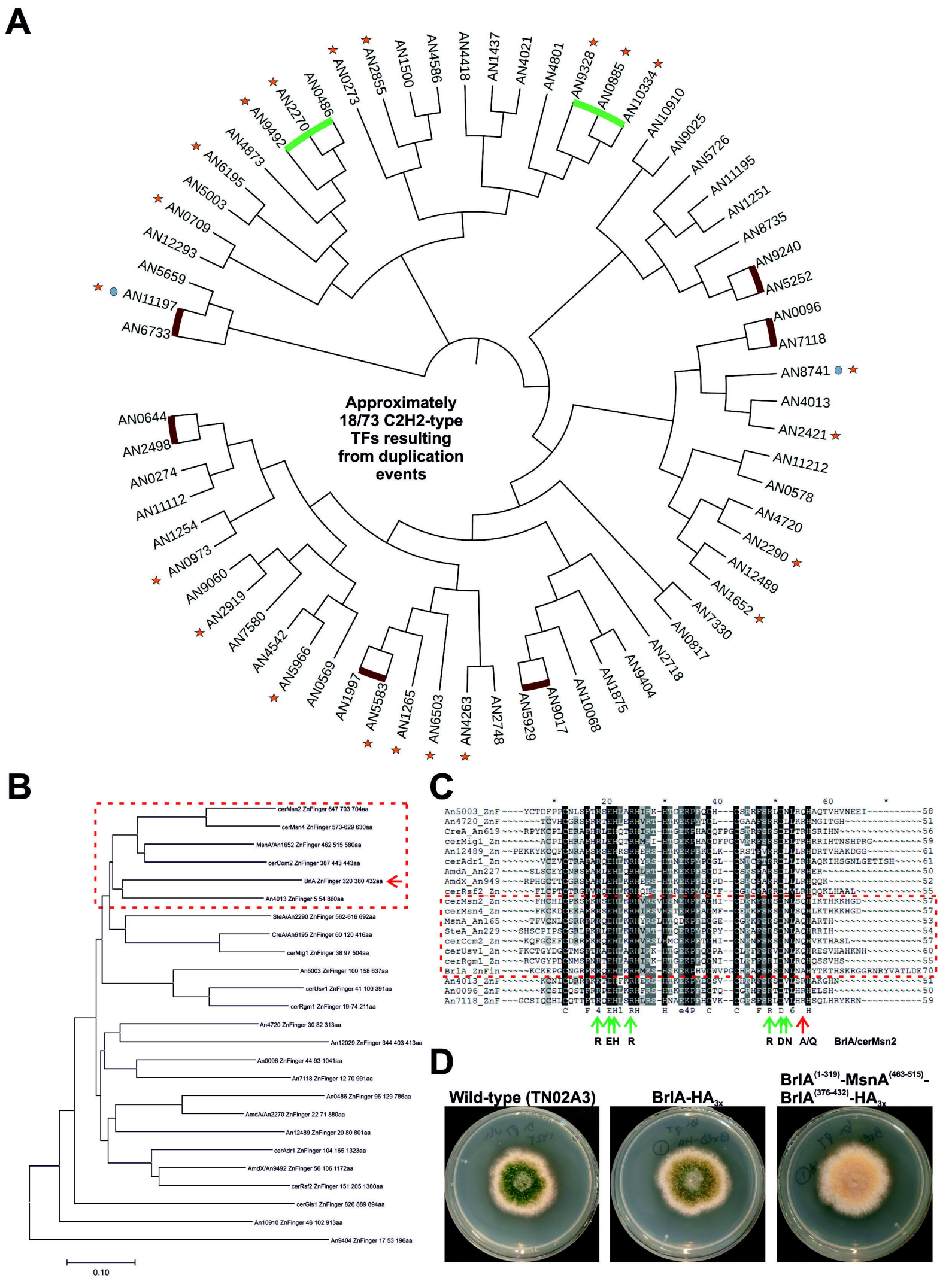

A
Total = 520
Correct annotation
(336 - 64.6 %)
Incorrect annotation
(157 - 30.2 %)
Not determined
(27 - 5.2 %)
B
IGV
FungiDB
Complete
Minimal
(Ammonium)
3´-
An3391
Zn2-Cys6 TF
-5´
3´-
An3390
Pectinesterase
-5´
C
IGV
FungiDB
Complete
Minimal
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
3´-
An2290/steA
Ste-like TF
-5´
D
IGV
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
5´-
An1569
Zn2Cys6 TF
-3´
St
Zn2Cys6 37-
-69
aSp
E
IGV
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
5´
An11098
Zn2Cys6 TF
-3´
St
Zn2Cys6 37-
-65
F
IGV
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
Complete
Minimal
5´-
An2782
Zn2Cys6 TF
-3´
RF
pSp
Sp
G
IGV
Sp
St
Sp
St
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
Complete
Minimal
3´-
An1812/jlbA
-5´
257
194
bZIP
H
IGV
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
Complete
Minimal
Zn2Cys6 TF
An9221
-5´
St
Sp
St
Predictably lost region
27-1
SP
1
TM
249-231
TM
183-160
TM
146-126
TM
114-81
TM
77-54
TM
34-14
I
IGV
St
Sp
St
Sp
5´-
An5252
C2H2-1
99
125
129
152
C2H2-2

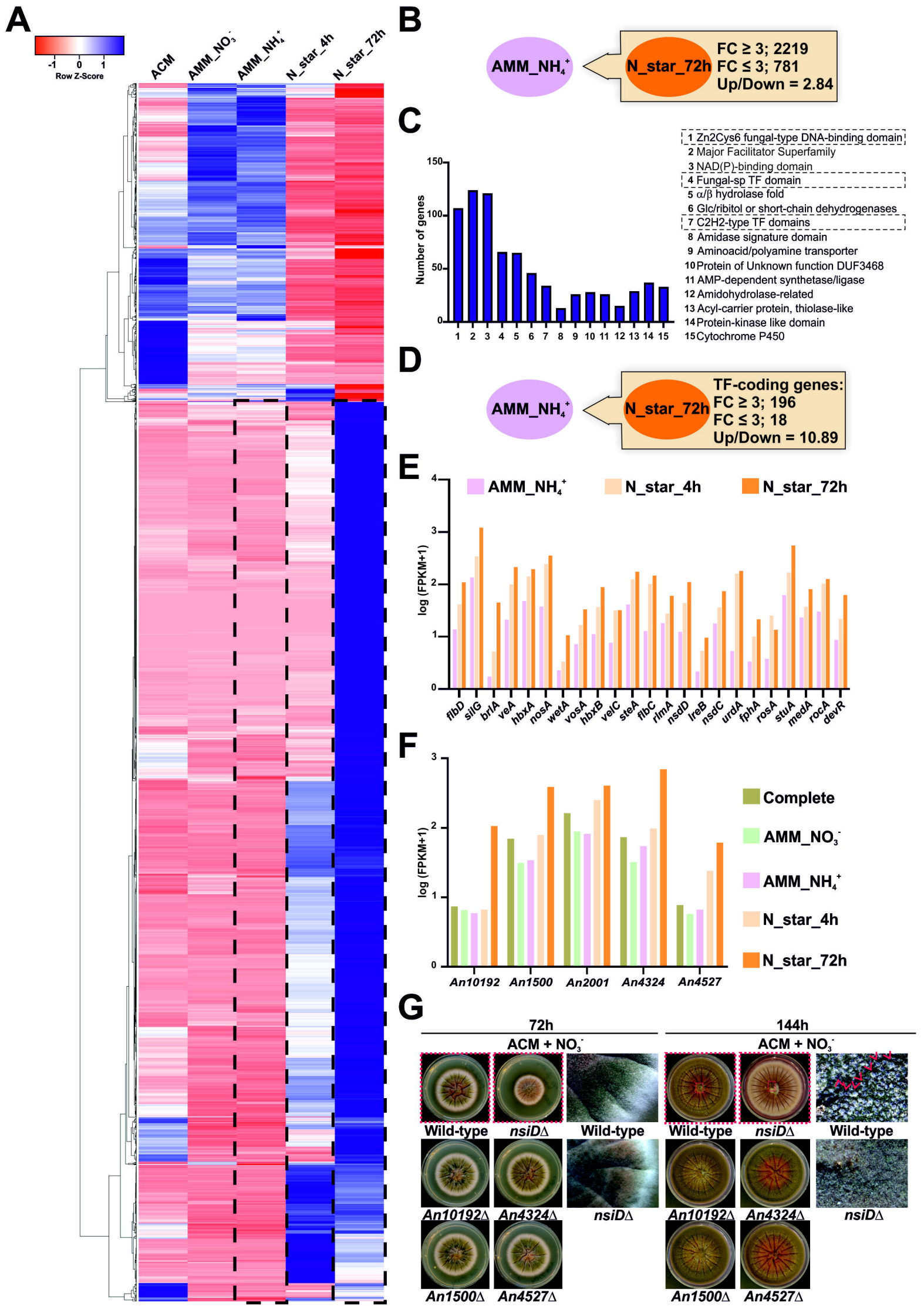

Fructose
Bagasse_6h
Bagasse_12h
WT_NoCarbon
WT_Avicel
cirBΔ_Avidel
WT_VG
WT_AD
flbBΔ_VG
flbBΔ_AD
mtfAΔ
WT
ACM
AMM_NO3-
AMM_NH4+
N_star_4h
N_star_72h
FGSC4_veA+
FGSC26_veA1
osaAΔ_veA1
WT_Uns
WT_pH8
WT_NaCl
sltAΔ_Uns
sltAΔ_pH8
sltAΔ_NaCl
sclBΔ
WT
sclB_com.
veAΔ
WT
WT
vosAΔ
Aflatoxin reg.
APSES
bHLH
bZIP
C2H2
Copper fist
CBF/CCAAT
Forkhead
GATA-type
HMG box
Homeobox
MAD-box
Myb-like
Ste-like
TEA/ATTS
Velvet
WOPR
Zn2Cys6
and/or
Fungal sp. TF
Row min
Row max


A
Fructose
Bagasse_6h
Bagasse_12h
D
B
ACM
AMM_NO3-
AMM_NH4+
N_star_4h
N_star_72h
E
Row min
Row max
C
WT_Uns
WT_pH8
WT_NaCl
sltAΔ_Uns
sltAΔ_pH8
sltAΔ_NaCl
F
F
F
D
log10(FPKM + 1)
Fructose
Bagasse_6h
Bagasse_12h
E
log10(FPKM + 1)
ACM
AMM NO3-
AMM NH4+
N_star. 4h
N_star. 72h
F
log10(FPKM + 1)
AMM
pH8
Na+
WT
sltAΔ
log10(FPKM + 1)
log10(FPKM + 1)


A
AN6295 clade
AN1114 clade
AN10132 clade
AN7661 clade
AN2392 clade
D
B-C
AN0396/AN5078 clade
AN2811 clade
E
AN7170 clade
AN7734 clade
AN7553 clade
AN8271 clade
AN1298 clade
D
IGV
St
Sp
pSp
FungiDB
WT Avicel
WT No Carb.
ΔclrB Avicel
ΔmtfA
ΔveA
5´-
An2392
-3´
218
269 bHLH
1
50
100
150
200
250
AN2392 clade
AN6295 clade
E
pSp
Sp?
St?
pSt
3´-
An2811
-5´
221
164 bHLH
AN2811 clade
B
IGV
Sp
St
Sp
pSt
FungiDB
WT Avicel
WT No Carb.
ΔclrB Avicel
ΔmtfA
ΔveA
3´-
An5078
-5´
232
143 bHLH
1
40
80
120
160
200
AN5078 & AN0396 clade
AN7170 & AN7661 clades
C
IGV
St
Sp
FungiDB
WT Avicel
WT No Carb
ΔclrB Avicel
ΔmtfA
ΔveA
5´-
An0396
-3´
191
279 bHLH
1
60
120
180
240
300
AN0396 orth.
AN5078 orth.
AN7170 clade
Score color key
<40
40-50
50-80
80-200
≥200


A
At least 139/331 Zn2Cys6-type TFs resulting from duplication events
B
Protein Score Coverage E value
An3385 1199 100 0.0
An7896 284 90 1e-91
An2036 269 69 1e-70
An6788 220 84 2e-67
An1029 204 61 1e-61
An0533 280 79 2e-61
Score color key
<40
40-50
50-80
80-200
≥200
C
Protein Score Coverage E value
An11169 1426 100 0.0
An10906 421 92 2e-142
An8636 258 84 6e-80
An5955 241 84 2e-74
An2375 186 78 1e-54
An8535 172 78 9e-50
D
Protein Score Coverage E value
An8590 1674 100 0.0
An7507 485 87 5e-164
An4972 427 83 5e-140
An11185 351 85 2e-112
An8930 311 77 1e-97
An0742 286 81 5e-88


A
Approximately 18/73 C2H2-type TFs resulting from duplication events
AN0486
AN2270
AN9492
AN0273
AN2855
AN1500
AN4586
AN4418
AN1437
AN4021
AN4801
AN9328
AN0885
AN10334
AN10910
AN9025
AN5726
AN11195
AN1251
AN8735
AN9240
AN5252
AN0096
AN7118
AN8741
AN4013
AN2421
AN11212
AN0578
AN4720
AN2290
AN12489
AN1652
AN7330
AN0817
AN2718
AN9404
AN1875
AN10068
AN9017
AN5929
AN2748
AN4263
AN6503
AN1265
AN5583
AN1997
AN0569
AN5966
AN4542
AN7580
AN2919
AN9060
AN0973
AN1254
AN1112
AN0274
AN2498
AN0644
AN6733
AN11197
AN5659
AN12293
AN0709
AN5003
AN6195
AN4873
B
cerMsn2 ZnFinger 647 703 704aa
cerMsn4 ZnFinger 573-629 630aa
MsnA/An1652 ZnFinger 462 515 580aa
cerCom2 ZnFinger 387 443 443aa
BrlA ZnFinger 320 380 432aa
An4013 ZnFinger 5 54 860aa
SteA/An2290 ZnFinger 562-616 692aa
CreA/An6195 ZnFinger 60 120 416aa
cerMig1 ZnFinger 38 97 504aa
An5003 ZnFinger 100 158 637aa
cerUsv1 ZnFinger 41 100 391aa
cerRgm1 ZnFinger 19-74 211aa
An4720 ZnFinger 30 82 313aa
An12029 ZnFinger 344 403 413aa
An0096 ZnFinger 44 93 1041aa
An7118 ZnFinger 12 70 991aa
An0486 ZnFinger 96 129 786aa
AmdA/An2270 ZnFinger 22 71 880aa
An12489 ZnFinger 20 80 801aa
cerAdr1 ZnFinger 104 165 1323aa
AmdX/An9492 ZnFinger 56 106 1172aa
cerRsf2 ZnFinger 151 205 1380aa
cerGis1 ZnFinger 826 889 894aa
An10910 ZnFinger 46 102 913aa
An9404 ZnFinger 17 53 196aa
0.10
C
R EH R
R DN A/Q BrlA/cerMsn2
D
Wild-type (TN02A3)
BrlA-HA3x
BrlA(1-319)-MsnA(463-515)-BrlA(376-432)-HA3x


A
Row Z-Score
-1
0
1
ACM
AMM_NO3-
AMM_NH4+
N_star_4h
N_star_72h
B
AMM_NH4+
N_star_72h
FC ≥ 3; 2219
FC ≤ 3; 781
Up/Down = 2.84
C
Number of genes
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 Zn2Cys6 fungal-type DNA-binding domain
2 Major Facilitator Superfamily
3 NAD(P)-binding domain
4 Fungal-sp TF domain
5 α/β hydrolase fold
6 Glc/ribitol or short-chain dehydrogenases
7 C2H2-type TF domains
8 Amidase signature domain
9 Aminoacid/polyamine transporter
10 Protein of Unknown function DUF3468
11 AMP-dependent synthetase/ligase
12 Amidohydrolase-related
13 Acyl-carrier protein, thiolase-like
14 Protein-kinase like domain
15 Cytochrome P450
D
AMM_NH4+
N_star_72h
TF-coding genes:
FC ≥ 3; 196
FC ≤ 3; 18
Up/Down = 10.89
E
AMM_NH4+
N_star_4h
N_star_72h
log (FPKM+1)
4
3
2
1
0
flbD
silG
brlA
veA
hbxA
nosA
wetA
vosA
hbxB
velC
steA
flbC
rlmA
nsdD
lreB
nsdC
urdA
fphA
rosA
stuA
medA
rocA
devR
F
log (FPKM+1)
3
2
1
0
An10192
An1500
An2001
An4324
An4527
Complete
AMM_NO3-
AMM_NH4+
N_star_4h
N_star_72h
G
72h
144h
ACM + NO3-
ACM + NO3-
Wild-type
nsiDΔ
Wild-type
Wild-type
nsiDΔ
Wild-type
An10192Δ
An4324Δ
nsiDΔ
An10192Δ
An4324Δ
nsiDΔ
An1500Δ
An4527Δ
An1500Δ
An4527Δ