Article

Transcription factors in the fungus *Aspergillus nidulans*: Markers of genetic innovation, network rewiring and conflict between genomics and transcriptomics.

Oier Etxebeste[1,≠] (ORCID iD: 0000-0002-9786-6091).

[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.

Keywords: Transcription factor, transcriptional regulation, transcriptomics, gene duplication, gene regulatory network, network rewiring.

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

One-sentence summary: A 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

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 TF-coding genes, their divergent sequence evolution and gain/loss/modification of CRSs are events rewiring GRNs. Fungi (mainly Saccharomycotina) have served as a reference kingdom for the study of GRN evolution. Here, I studied the genes predicted to encode TFs in the fungus *Aspergillus nidulans* (Pezizomycotina). The analysis of the expansion of different families of TFs suggests that duplication of TFs impacts the species level and that the expansion in Zn2Cys6 TFs is mainly due to dispersed duplication events. Comparison of genomic annotation and transcriptomic data suggests that a significant percentage of genes should be re-annotated while many others remain silent. Finally, a new regulator of growth and development is identified and characterized. Overall, this study establishes a novel theoretical framework in synthetic biology, since overexpression of silent TF forms would provide additional tools to assess how GRNs are rewired.

Impact Statement

GRNs controlling fungal sexual and asexual development, stress response, carbon/nitrogen metabolism or secondary metabolism, have been widely characterized in the fungus *A. nidulans* and, in several cases, were characterized for the first time in this fungus. The functional and genetic relationships among the TFs of these networks are known. By comparing the genomes/proteomes of hundreds of fungal genomes available in public repositories, researchers can track the emergence of regulators of interest, and identify groups of paralog TFs. The present work uses basic bioinformatics tools and standard RNA-seq analysis tools to compare genomic annotation and transcriptomics data, find incorrectly annotated TFs, identify paralog clusters of TFs, track the trajectories of duplication events and expansion of bHLH-type TFs, and identify TFs with a role in the control of growth and development. The study of TFs and GRNs of *A. nidulans* offers the opportunity to dissect and analyze sequentially the contribution of each cis and trans regulatory element to network rewiring and the modification of developmental patterns.

Data Summary

The data underlying this article are available in supplementary tables and supplementary figures. RNA-seq, genomic and proteomic data used is this article (and previously deposited by others; see references and accession numbers in Materials and Methods) were retrieved from sources (www.ebi.ac.uk/arrayexpress/; www.ncbi.nlm.nih.gov/bioproject/; https://www.ncbi.nlm.nih.gov/geo/) and databases (https://fungidb.org/fungidb/app) in the public domain. *Aspergillus nidulans* strains generated in this article will be shared on reasonable request to the corresponding author.

Abbreviations

AMM, Aspergillus minimal medium; ACM, Aspergillus complete medium; TF, Transcription factor; GRN, Gene regulatory network; CRS, Cis-regulatory sequence; DBD, DNA-binding domain; FC, Fold change; TM, Transmembrane domain; IPD, InterPro regulatory domain; UDA, Upstream developmental activation pathway; CDP, Central developmental pathway; VG, Vegetative growth; AD, Asexual development;

SD, Segmental duplication; DD, Dispersed duplication; TGD, Tandem gene duplication; HGT, Horizontal gene transfer.

Background

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 [1]. The architecture of GRNs ranges from simple, such as that controlling mating-type in yeasts [2] 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 [3]).

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 [1].

There are different mechanisms for modifying the structure of GRNs [4]. The emergence of new regulators (rather than the incorporation of individual target genes; “regulator-first” model; see references within [3]) is one of the major events, and occurs mainly through duplication of genes coding for TFs and divergent evolution of each paralog sequence. Amino acid changes in TFs can precede gains of CRSs [5], leading to neofunctionalization (the development of new functions) or subfunctionalization (retention of a subset of the functions of the original paralog) [3, 4].

The kingdom fungi is composed of 2-4 million species [6]. 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 [7]. 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 [2, 3, 5]. Recent comparative studies carried out in Basidiomycetes and Ascomycetes have shown that TFs controlling (asexual and/or sexual) developmental patterns emerged gradually in evolution [8–10]. 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 [11]. BrlA emerged much later than other upstream and downstream TFs of the same GRN [12, 13], suggesting that it caused rewiring events in pre-existent ancient networks [9].

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* [14]. 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. The analysis of several RNA-seq experiments carried out using *A. nidulans* as the reference system and their comparison with gene models in the FungiDB database suggest that the annotation of a high proportion of genes encoding TFs should be updated and that many others tend to remain silent. A new regulator of *A. nidulans* growth and sexual/asexual developmental cycles, AN2001, is also identified and the phenotype of the corresponding null mutant characterized. 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 silent TF forms, their overexpression

and their functional assessment, would provide additional tools for studying mechanisms of GRN rewiring.

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 [14] 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 [15]. 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 [16] and SignalP [17]. Presence of putative transmembrane (TM) domains within protein sequences was analyzed using Interpro, TopCons [18], TMHMM [19] and Phobius [20]. 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 [21] while clustering of IPDs (IPD: Interpro Regulatory Domains) was carried out using "*rough clustering*" at MAFFT, version 7 [22]. Phylogenetic trees were generated using MEGAX [23]. Phylogenetic (MEGAX) and rough trees (MAFFT) were edited using iTOL [24] and Archaeopteryx version 0.9921 [25]. 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 [26]. 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 [27].

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 (quantitative data in this case correspond to a single sample per condition and do not include biological replicates; see comments in the last section of results), 2) a null *veA* genetic background, 3) a null *mtfA* background, or 4) a null *clrB* genetic background were retrieved from the FungiDB database (Coradetti et al. 2013; Sibthorp 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 [32, 33]. Raw RNA-seq data deposited under accession numbers 7) E-MTAB-6996 (null *sclB* background) [34], 8) PRJNA588808 (null *vosA* background) [35], 9) GSE72316 (null *osaA* background) [36], and 10) PRJNA294437 (use of waste steam-exploded sugarcane bagasse as a nutrient source) [37], were retrieved and processed at the Galaxy website [38]. *bam* files were generated from *fastq* files using RNA Star [39], and Stringtie [40] or CuffNorm [41] were used to generate FPKM-based or TPM-based expression tables.

To check the annotation of genes potentially coding for TFs and the presence or not 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; [42]), in combination with the set of transcriptomics data available in the FungiDB database. GO enrichment analysis were carried out using ShinnyGO [43]. Heatmaps were generated using Heatmapper [44] 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 [45] 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) and starts of *brlAα* and *brlAβ* transcripts [46] were maintained as in the wild type. First, a *brlA::ha*$_{3x}$*::pyrG::brlA*$_{3´-UTR}$ 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 S1); 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 S1); 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*$_{3´-UTR}$, used as control to confirm that the HA_{3x} tag did not impede BrlA function, and *brlA*$_{M1-Q319}$*::msnA*$_{T463-H515}$*::brlA*$_{S376-E432}$*::ha*$_{3x}$*::pyrG*^{Afum}:: *brlA*$_{3´-UTR}$, were used to transform protoplasts of the reference strain TN02A3 (*pyrG89; argB2; pyroA4,* Δ*nkuA::argB; veA1*; [47]), following the procedures previously described by [48, 49]. BrlA::HA_{3x} transformants showed a wild-type conidiating phenotype, while BrlA-MsnA chimera transformants showed an aconidial phenotype. Genomic DNA of selected transformants was extracted as previously described by us [50], and the maintenance of the reading frame in the *brlA-msnA* chimera was confirmed by Sanger sequencing.

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* (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 S1); 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 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*.

Results

The set of potential TFs 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 [51]. Thus, a large number of TFs have been characterized functionally and their genetic/molecular inter-relationships determined. TFs and GRNs controlling in fungi, e.g., 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 [9, 12, 13]. 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 [52, 53]. 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 [14] and updated (Table S2) based on several works available in the literature and the FungiDB database [15]. 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 [16]. Table S2 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 [52, 53]. 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 annotation of specific genes (e.g., *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). 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 S1 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 S3; Figure 1B). 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 S4; Figure 1B), 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 S4) 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%) (Figure 1B). 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* [54]. 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 [55]. 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 [31–33] 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 S5 (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 S5; 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 338 genes/transcripts are processed as predicted (65.0%; 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 [56], is shown as an example in Figure 2C). In some 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 probably the most common alternative splicing form. 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 155 out of 520; 29.8 %; 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 the start codon

is probably located in the beginning of the predicted second exon (M23) (Figure 2E), and *AN2782*, also coding for a binuclear zinc cluster, in which the third predicted intron is absent, 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 differ from the predicted ones. Three examples are shown in Figure 2. The start of the *AN1812/jlbA* transcript, which is upregulated under nitrogen starvation [57], is as annotated. However, the predicted first intron is absent and the 5´-UTR region of *AN1812/jlbA* is probably longer than predicted (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 predicted exons and introns is very low compared those mapping to the predicted third exon. RNA-seq reads also suggest that there is no intron in the sequence, and that *AN9221* is two separate genes, *alnG* and *alnR* of the asperlin cluster, as proposed by Grau and colleagues [58] (bottom panel in Figure 2H). The probable start codon of *alnR* is that annotated in the FungiDB database as Met246 for AN9221, just before the region coding for 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 the gene contains only one intron (the predicted second intron). The 5´-UTR region is probably longer than predicted and the real start codon probably corresponds to that annotated as M117, implying the loss of the first predicted C2H2 domain (dark grey square bold in red). The predicted second C2H2 IPD (M127-D168 plus 20 codons out of frame; coordinates of the FungiDB version) would be included in the coding region of AN5252.

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 S6 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 [52], 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 S2 shows that it is not the case.

Since an important number of the transcripts in Table S6 showed low or null expression, all FPKM (heatmaps 1-6 and 9-10) or TPM (heatmaps 7 and 8) values below 5 were removed from the table to build Figure 3 (indicated in gray). The presence of larger gray areas (FPKM/TPM < 5) 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 or genetic backgrounds (others are silent in most but not all culture conditions/genetic backgrounds). That is expected, for example, 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 S6; Figure 3). As described above, annotation of others, such as *AN5252*/*AN9240*, is clearly incorrect (see Figure 2 and Table S5), suggesting that no functional TF forms are translated from some predicted genes in detectable levels.

Figure 3 and Figure S2 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 S2). Clustering of heatmaps corresponding to the effect of the addition of sugarcane bagasse as the main nutrient source (Figure 4A and 4C) [37], or a null *sltA* (salt sensitivity) background (Figure 4B and 4D) (Picazo et

al., 2020) supports this possibility. Qualitatively, a majority of TF-coding genes seem to be upregulated in the first culture conditions but downregulated in the *sltA*Δ mutant.

Duplication of TF-coding genes impacts the species level

In a previous work, and using 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 [9]. 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 [8]. 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 [59]. 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 predicted or shown to code for TFs of this family were identified by the FungiDB database. Five of them had a standard name (*AN1298/glcD*, *AN7553/devR*, *AN7661/srbA*, *AN7734/anbH1* and *AN8271/palcA*; [54, 60–63]), while the remaining eight had only a systematic name. By sequence similarity, AN4394/UrdA was initially described to belong to this family of TFs [33, 64]. However, the Interpro website did not identify a basic helix-loop-helix domain within its primary structure. Thus, it was included in Table S2 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 present, 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; [9]) returned AN5078 as the first hit in very few cases (Table S7), confirming a conservation pattern for AN5078 restricted to less than ten *Aspergillus* species..

The phylogenetic tree in Figure 5A included AN0396 in the AN5078 clade. RNA-seq data showed that the predicted intron of *AN0396* 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 apparently 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 present, leading to an earlier stop codon prior to the last two predicted exons (Figure 5D). BLAST and HMMER searches, together with confirmatory reverse retrievals, strongly suggest that AN2392 corresponds to a duplication of AN6295 or AN10132, and that its conservation pattern is highly restricted, a common feature of genes belonging to secondary metabolite gene clusters [65, 66]. Finally, despite the low expression levels, RNA-seq data suggested that the annotation of *AN2811* is incorrect and that the probable amino acid sequence is shorter than that predicted but probably 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 AN2811 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 family of TFs as a paradigm 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 [52, 53], 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 [67], 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/DbaA, AN2036, AN3385, AN0533, AN6788 and AN1029/AfoA 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 S3; the current version of AfoA in the FungiDB database does not include the Zn2Cys6 domain; a previous version including it, XP_658633.1, was used in the alignment based on [58, 68]). A similar pattern was observed in the remaining two clusters of six paralogs (Figure 6C and 6D; sequence alignments shown in Figures S4 and S5, respectively). In the case of Figure S4, two of the six Cys residues of the Zn2Cys6 domain of AN10906 seem to have mutated. Analysis of RNA-seq results strongly suggest that the annotation of the gene is correct. Thus, it would be of interest to assess how the divergent evolution of the Zn2Cys6 domain of AN10906 could affect its ability to bind DNA and its functionality.

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 two or three members only (1: AN0486, AN2270 and AN9492; 2: AN0885, AN9328 and AN10334; 3: AN6733 and AN11197; 4: AN0096 and AN7118; 5: AN0644 and AN2498; 6: AN1997 and AN5583; 7: AN5929 and AN9017). A total of 16 (considering that AN5252 and AN9240 are the same protein; see above) out of 73 potential C2H2-type TFs were predicted to be a result of duplication, a clearly lower percentage (21.9%) compared to that estimated in the case of Zn2Cys6 TFs (42.0%).

The case of the developmental regulator BrlA [69] 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 [9, 12, 13]. 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 [51]. 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 [9]. BLAST and HMM analyses carried out to determine the conservation pattern of BrlA in the proteomes of 503 fungal species [9] 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) [56].

Trying to obtain clues on the origin and the ancestor(s) of BrlA, I compared the C2H2 IPDs of BrlA with those of specific *Aspergillus nidulans* and *Saccharomyces cerevisiae* TFs of the same family (Figure 7A). The sequences aligned in Figure 7A conserved all but one of the residues of *S. cerevisiae* Msn paralogs (cerMsn2 and cerMsn4) predicted to contact DNA (green and red arrows, respectively) [70]. 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 7B).

Identification of a Zn2Cys6-type protein necessary for growth and development.

To understand the structure, function and evolution of complex GRNs, scientists hierarchically decompose these networks into manageable and intelligible subsystems (see references in [71]). 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 [72, 73]. 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 [51]). 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 [74].

Sibthorp and colleagues analyzed the effect of the nitrogen source (ammonium or nitrate) added to the culture medium or the effect of nitrogen starvation (4 and 72 hours) in the expression profiles of *A. nidulans* [28]. RNA-seq data of this experiment were based on the SOLiD sequencing technology and corresponded to a single experiment without biological replicates. Thus, they were not intended for use in comparative expression analyses but to provide expression information to assist genome annotation [28, 75]. This means that quantitative expression data must be handled with care. Here I used FPKM values provided by the FungiDB database as preliminary data to explore a hypothetic relationship between nitrogen starvation and the induction of development. For example, when comparing the sample collected under nitrogen starvation (72h) and

that in AMM that used ammonium as the nitrogen source (Table S8), most of the genes seemed to be upregulated (2,219 genes with a FC ≥ 3 compared to 781 genes with a FC ≤ 3; ratio ~ 2.8; black squares in Figure S6A; Figure S6B). Genes encoding TFs (Zn2Cys6-fungal specific and C2H2 families) apparently were among the top Interpro domains upregulated under nitrogen starvation (196 TF-coding genes were upregulated and 18 downregulated; ratio ~ 10.9; Figure S6C and S6D; Table S9). Genes encoding key regulators of nitrogen metabolism were found, as expected, in the list of upregulated TF-coding genes, but also several known regulators of asexual and sexual development (Figure S6E) [33, 61, 69, 76–81]. Thus, it was hypothesized that new developmental regulators could be identified in this preliminary list. Five genes were selected (Figure S6F) and their null mutants generated.

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 (e.g., 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) [32]. 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 *AN2001*Δ mutant showed partial inhibition of radial growth and conidia production, mainly in ACM supplemented with nitrate as the main nitrogen source (Figure 8A and 8C). After 144 hours of culture at 37 ºC, while the reference wild-type strain produced cleistothecia in ACM (Figure 8B and 8D), the *AN2001*Δ mutant was blocked in the development of sexual structures. Overall, results strongly suggest that, in the future, the analysis of the transcriptomic responses of *A. nidulans* to nitrogen starvation could help in the identification of hitherto uncharacterized regulators of development. Such analysis, however, should be based on RNA-seq data that include biological replicates. 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. In the case of AN2001, 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 S6), that its three introns are processed as predicted (Table S5), and that it is conserved in the subphylum Pezizomycotina (not shown).

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 [82]). Robustness is related to the ability of a system to sustain its functionality in the face of perturbation [83, 84], which, in the case of GRNs, could be understood as the extent a phenotypic trait persists after mutation [82]. 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 [71, 82]. Gene duplication and divergent sequence evolution of paralogs are 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 [3–5]. 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 [7, 85], gives fungal researchers the possibility to elucidate when each family of transcriptional regulators and each TF emerged [8, 9, 53, 86, 87]. *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* [12, 88]. 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 [8, 9]. 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 [89]. 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 [89, 90]. 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) [91]. 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* or *AN10132*. *AN0396* seems to be 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 [52, 53], 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 [52]. Zinc clusters grew much faster in Ascomycetes, suggesting a particular role and evolutionary

history of Zn2Cys6-type TFs [52]. 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 [92, 93]. Whole-genome duplication events have been described in fungi [94] but not in *Aspergilli* [95]. Figure S7 and Table S10 show the genomic distribution of those genes encoding the Zn2Cys6-type TFs shown in Figure 6A and which 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 [92]. 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 [96]. But, as mentioned previously, Khaldi and Wolfe described that they found no clues of SDs in *Aspergilli* [95]. Dispersed duplications (DDs) have been defined as the generation of two gene copies that are neither neighboring nor colinear [97], a definition which fits with the genomic distribution of the members of the hypothetical Zn2Cys6 paralog clusters of *A. nidulans* (Figure 6A, Figure S7 and Table S10). Despite the fact that the mechanisms enabling DDs remain unclear and have been defined as unpredictable and occurring through random patterns [97], 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/DbaA, AN2036, AN3385, AN0533, AN6788 and AN1029/AfoA; Figure 6B). A BLAST search against the genome of *A. nidulans* using the sequence of AN7903/pkeA (the polyketide synthase, PKS, of the *dba* cluster) as the query [98], suggested that the PKSs AN2032/PkhA, AN3386/PkiA, AN0523/PkdA and AN1034/AfoE (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/DbaE, AN2031, AN0529, AN6789 and AN1032/AfoC, or the FAD-binding monoxygenases AN7902/DbaH, AN2033, AN3382/SalA, AN0530 and AN1033/AfoD. Additional BLAST searches against the *A. nidulans* genome also related AN7893 with AN0526, AN11584/DbaC (located between AN7897/DbaB and AN7898/DbaD) with AN0527, or AN7898/DbaB 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* [33, 99], while the donor of the cluster defined by AN2036 probably was *Coccidioides immitis* [33] or *Penicillium italicum* [99]. The rest of the secondary metabolite gene clusters corresponding to the aforementioned Zn2Cys6 paralog group have not been linked with HGT events [99], 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/PkdA 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 the other mechanisms.

The most likely possibility after duplication is the loss of the new paralog. If the duplication event persists in the population, then pseudogenization would be a probable outcome [92]. 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 S2). However, multiple duplications are fixed and maintained in genomes, and different models have been proposed to explain how neofunctionalization, subfunctionalization or functional redundancy occur [92, 100]. 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 [101, 102], 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 [103, 104]. A third example could be the paralog cluster composed of AN1425/FarB and AN7050/FarA (fatty-acid utilization) [105] 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 [52]. 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, in order to induce their expression, by upstream TFs or, alternatively, by TFs with an autoregulatory activity, as could be the cases of AlcR, CpcA, FacB. PacC and others. In this scenario, Figure 3 and Table 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 S10). It remains to be determined if this observation reflects the requirement of the corresponding functions under specific environmental/culture conditions (as could be the case of regulators of secondary metabolite gene clusters or that of specific TFs regulating their own expression) 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 [70]. 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´)

[70]. As can be seen, similarities compared to BrlA Response Elements (BREs; 5´-CAAGGGG-3´), defined by [106] 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 [107]). On the other hand, the *fluffy* phenotype described after the replacement of the C2H2 domains of BrlA with those of MsnA suggests 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.

Author Contribution

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

Conflict of Interest

The author declares that there are no conflicts of interest.

Funding Information

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).

Acknowledgements

I want to acknowledge the contribution of Drs. Eduardo A. Espeso (CIB-CSIC) and Marc S. Cortese to the improvement of this work.

References

1. **Park J, Wang HH**. Systematic and synthetic approaches to rewire regulatory networks. *Current Opinion in Systems Biology* 2018;8:90–96.

2. **Nocedal I, Johnson AD**. How transcription networks evolve and produce biological novelty. *Cold Spring Harbor Symposia on Quantitative Biology* 2015;80:265–274.

3. **Sorrells TR, Johnson AD**. Making sense of transcription networks. *Cell* 2015;161:714–723.

4. **Voordeckers K, Pougach K, Verstrepen KJ**. How do regulatory networks evolve and expand throughout evolution? *Current Opinion in Biotechnology* 2015;34:180–188.

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

6. **Hawksworth DL, Lücking R**. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiology Spectrum*;5. Epub ahead of print 2017. DOI: 10.1128/microbiolspec.FUNK-0052-2016.

7. **Spatafora JW, Aime MC, Grigoriev I V., Martin F, Stajich JE, *et al.*** The fungal tree of life: from molecular systematics to genome-scale phylogenies. *Microbiology Spectrum*;5. Epub ahead of print 2017. DOI: 10.1128/microbiolspec.FUNK-0053-2016.

8. **Krizsán K, Almási É, Merényi Z, Sahu N, Virágh M, *et al.*** Transcriptomic atlas of mushroom development reveals conserved genes behind complex multicellularity in fungi. *Proceedings of the National Academy of Sciences* 2019;116:7409–7418.

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

10. **Kiss E, Hegedis B, Varga T, Merenyi Z, Koszo T, *et al.*** Comparative genomics reveals the origin of fungal hyphae and multicellularity. *Nature Communications* 2019;10:4080.

11. **Adams TH, Wieser JK, Yu J-H**. Asexual sporulation in *Aspergillus nidulans*. *Microbiology and Molecular Biology Reviews* 1998;62:35–54.

12. **Ojeda-López M, Chen W, Eagle CE, Gutiérrez G, Jia WL, *et al.*** Evolution of asexual and sexual reproduction in the aspergilli. *Studies in Mycology* 2018;91:37–59.

13. **de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, *et al.*** Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biology* 2017;18:28.

14. **Wortman JR, Gilsenan JM, Joardar V, Deegan J, Clutterbuck J, *et al.*** The 2008 update of the Aspergillus nidulans genome annotation: a community effort. *Fungal Genet Biol* 2009;46:S2–S13.

15. **Basenko E, Pulman J, Shanmugasundram A, Harb O, Crouch K, *et al.*** FungiDB: An integrated bioinformatic resource for fungi and oomycetes. *Journal of Fungi* 2018;4:39.

16. **Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, *et al.*** InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research* 2019;47:D351–D360.

17. **Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, *et al.*** SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* 2019;37:420–423.

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

19. **Krogh A, Larsson B, von Heijne G, Sonnhammer ELL**. Predicting

transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology* 2001;305:567–580.

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

21. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al.*** Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 2011;7:539.

22. **Katoh K, Rozewicki J, Yamada KD**. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 2019;20:1160–1166.

23. **Kumar S, Stecher G, Li M, Knyaz C, Tamura K**. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 2018;35:1547–1549.

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

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

26. **Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, *et al.*** HMMER web server: 2018 update. *Nucleic Acids Research* 2018;46:W200–W204.

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

28. **Sibthorp C, Wu H, Cowley G, Wong PWH, Palaima P, *et al.*** Transcriptome analysis of the filamentous fungus *Aspergillus nidulans* directed to the global identification of promoters. *BMC Genomics* 2013;14:847.

29. **Coradetti ST, Xiong Y, Glass NL**. Analysis of a conserved cellulase transcriptional regulator reveals inducer-independent production of cellulolytic

enzymes in *Neurospora crassa*. *MicrobiologyOpen* 2013;2:595–609.

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

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

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

33. **Oiartzabal-Arano E, Garzia A, Gorostidi A, Ugalde U, Espeso EA, *et al.*** Beyond asexual development: Modifications in the gene expression profile caused by the absence of the *Aspergillus nidulans* transcription factor FlbB. *Genetics* 2015;199:1127–42.

34. **Thieme KG, Gerke J, Sasse C, Valerius O, Thieme S, *et al.*** 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 Genetics* 2018;14:e1007511.

35. **Kim M-J, Lee M-K, Pham HQ, Gu MJ, Zhu B, *et al.*** The velvet regulator VosA governs survival and secondary metabolism of sexual spores in *Aspergillus nidulans*. *Genes* 2020;11:103.

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

37. **Brown NA, Ries LNA, Reis TF, Rajendran R, Corrêa dos Santos RA, *et al.*** RNAseq reveals hydrophobins that are involved in the adaptation of *Aspergillus nidulans* to lignocellulose. *Biotechnology for Biofuels* 2016;9:145.

38. **Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, *et al.*** The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* 2018;46:W537–W544.

39. **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, *et al.*** STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2012;29:15–21.

40. **Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, *et al.*** StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 2015;33:290–295.

41. **Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, *et al.*** Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28:511–515.

42. **Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, *et al.*** Integrative genomics viewer. *Nature Biotechnology* 2011;29:24–26.

43. **Ge SX, Jung D, Yao R**. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 2020;36:2628–2629.

44. **Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, *et al.*** Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* 2016;44:W147–W153.

45. **Yang L, Ukil L, Osmani A, Nahm F, Davies J, *et al.*** Rapid production of gene replacement constructs and generation of a green fluorescent protein-tagged centromeric marker in *Aspergillus nidulans*. *Eukaryotic cell* 2004;3:1359–62.

46. **Prade RA, Timberlake WE**. The *Aspergillus nidulans brlA* regulatory *locus* consists of overlapping transcription units that are individually required for conidiophore development. *The EMBO Journal* 1993;12:2439–2447.

47. **Nayak T, Szewczyk E, Oakley CE, Osmani A, Ukil L, *et al.*** A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics* 2006;172:1557–66.

48. **Tilburn J, Scazzocchio C, Taylor GG, Zabicky-Zissman JH, Lockington RA, *et al.*** Transformation by integration in *Aspergillus nidulans*. *Gene* 1983;26:205–

221.

49. **Szewczyk E, Nayak T, Oakley CE, Edgerton H, Xiong Y, *et al.*** Fusion PCR and gene targeting in *Aspergillus nidulans*. *Nature Protocols* 2007;1:3111–3120.

50. **Garzia A, Etxebeste O, Herrero-Garcia E, Fischer R, Espeso EA, *et al.*** *Aspergillus nidulans* FlbE is an upstream developmental activator of conidiation functionally associated with the putative transcription factor FlbB. *Molecular Microbiology* 2009;71:172–184.

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

52. **Shelest E**. Transcription Factors in fungi: TFome dynamics, three major families, and dual-specificity TFs. *Frontiers in Genetics* 2017;8:53.

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

54. **Bat-Ochir C, Kwak J-Y, Koh S-K, Jeon M-H, Chung D, *et al.*** The signal peptide peptidase SppA is involved in sterol regulatory element-binding protein cleavage and hypoxia adaptation in *Aspergillus nidulans*. *Molecular Microbiology* 2016;100:635–655.

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

56. **Vallim MA, Miller KY, Miller BL**. *Aspergillus* SteA (Sterile12□like) is a homeodomain□C2/H2□Zn+2 finger transcription factor required for sexual reproduction. *Molecular Microbiology* 2002;36:290–301.

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

58. **Grau MF, Entwistle R, Chiang Y-M, Ahuja M, Oakley CE, *et al.*** Hybrid

transcription factor engineering activates the silent secondary metabolite gene cluster for (+)-Asperlin in *Aspergillus nidulans*. *ACS Chemical Biology* 2018;13:3193–3205.

59. **Fischer S, Brunk BP, Chen F, Gao X, Harb OS, *et al.*** Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics* 2011;35:6.12.1-6.12.19.

60. **Arst HN, Hondmann DHA, Visser J**. A translocation activating the cryptic nitrogen regulation gene *areB* inactivates a previously unidentified gene involved in glycerol utilisation in *Aspergillus nidulans*. *Molecular and General Genetics MGG* 1990;223:134–137.

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

62. **Caruso ML, Litzka O, Martic G, Lottspeich F, Brakhage AA**. 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. *Journal of Molecular Biology* 2002;323:425–439.

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

64. **Pandit SS, Lohmar MJ, Ahmed S, Etxebeste O, Espeso AE, *et al.*** UrdA controls secondary metabolite production and the balance between asexual and sexual development in *Aspergillus nidulans*. *Genes* 2018;9:570.

65. **Nützmann H-W, Scazzocchio C, Osbourn A**. Metabolic Gene Clusters in eukaryotes. *Annual Review of Genetics* 2018;52:159–183.

66. **Rokas A, Wisecaver JH, Lind AL**. The birth, evolution and death of metabolic gene clusters in fungi. *Nature Reviews Microbiology* 2018;16:731–744.

67. **Todd RB, Andrianopoulos A**. Evolution of a fungal regulatory gene family: The

Zn(II)2Cys6 binuclear cluster DNA binding motif. *Fungal Genetics and Biology* 1997;21:388–405.

68. **Chiang Y-M, Szewczyk E, Davidson AD, Keller N, Oakley BR, *et al.*** A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, Asperfuranone, in *Aspergillus nidulans*. *Journal of the American Chemical Society* 2009;131:2965–2970.

69. **Boylan MT, Mirabito PM, Willett CE, Zimmerman CR, Timberlake WE**. Isolation and physical characterization of three essential conidiation genes from *Aspergillus nidulans*. *Molecular and Cellular Biology* 1987;7:3113–3118.

70. **Siggers T, Reddy J, Barron B, Bulyk ML**. Diversification of transcription factor paralogs via noncanonical modularity in C2H2 Zinc Finger DNA binding. *Molecular Cell* 2014;55:640–648.

71. **Verd B, Monk NAM, Jaeger J**. Modularity, criticality, and evolvability of a developmental gene regulatory network. *eLife* 2019;8:e42832.

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

73. **Lee M-K, Kwon N-J, Lee I-S, Jung S, Kim S-C, *et al.*** Negative regulation and developmental competence in *Aspergillus*. *Scientific Reports* 2016;6:28874.

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

75. **Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G, *et al.*** The *Aspergillus* Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic Acids Research* 2014;42:D705–D710.

76. **Kwon N-J, Garzia A, Espeso EA, Ugalde U, Yu J-H**. FlbC is a putative nuclear C2H2 transcription factor regulating development in *Aspergillus nidulans*. *Molecular Microbiology* 2010;77:1203–1219.

77. **Garzia A, Etxebeste O, Herrero-García E, Ugalde U, Espeso EA**. The

concerted action of bZip and cMyb transcription factors FlbB and FlbD induces *brlA* expression and asexual development in *Aspergillus nidulans*. *Molecular Microbiology* 2010;75:1314–1324.

78. **Son S-H, Son Y-E, Cho H-J, Chen W, Lee M-K, *et al.*** Homeobox proteins are essential for fungal differentiation and secondary metabolism in *Aspergillus nidulans*. *Scientific Reports* 2020;10:6094.

79. **Marshall MA, Timberlake WE**. *Aspergillus nidulans wetA* activates spore-specific gene expression. *Molecular and Cellular Biology* 1991;11:55–62.

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

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

82. **Igler C, Lagator M, Tkačik G, Bollback JP, Guet CC**. Evolutionary potential of transcription factors for gene regulatory rewiring. *Nature Ecology & Evolution* 2018;2:1633–1643.

83. **Noman N, Monjo T, Moscato P, Iba H**. Evolving robust gene regulatory networks. *PLOS ONE* 2015;10:e0116258.

84. **Kitano H**. Biological robustness. *Nature Reviews Genetics* 2004;5:826–837.

85. **Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, *et al.*** A genome-scale phylogeny of Fungi; insights into early evolution, radiations, and the relationship between taxonomy and phylogeny. *bioRxiv* 2020;2020.08.23.262857.

86. **de Mendoza A, Sebé-Pedrós A, Šestak MS, Matejčić M, Torruella G, *et al.*** Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proceedings of the National Academy of Sciences* 2013;110:E4858 LP-E4866.

87. **de Mendoza A, Sebé-Pedrós A**. Origin and evolution of eukaryotic transcription factors. *Current Opinion in Genetics & Development* 2019;58–59:25–32.

88. **Mead ME, Borowsky AT, Joehnk B, Steenwyk JL, Shen X-X, *et al.*** Recurrent loss of *abaA*, a master regulator of asexual development in filamentous fungi, correlates with changes in genomic and morphological Traits. *Genome Biology and Evolution* 2020;12:1119–1130.

89. **Pires N, Dolan L**. Early evolution of bHLH proteins in plants. *Plant Signaling & Behavior* 2010;5:911–912.

90. **Pires N, Dolan L**. Origin and diversification of Basic-Helix-Loop-Helix proteins in plants. *Molecular Biology and Evolution* 2010;27:862–874.

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

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

93. **Zhang J**. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* 2003;18:292–298.

94. **Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, *et al.*** *Expansion of signal transduction pathways in fungi by extensive genome duplication*. 2016. Epub ahead of print 2016. DOI: 10.1016/j.cub.2016.04.038.

95. **Khaldi N, Wolfe KH**. Elusive origins of the extra genes in *Aspergillus oryzae*. *PLOS ONE* 2008;3:e3036.

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

97. **Qiao X, Li Q, Yin H, Qi K, Li L, *et al.*** Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biology* 2019;20:38.

98. **Gerke J, Braus GH**. Manipulation of fungal development as source of novel

secondary metabolites for biotechnology. *Appl Microbiol Biotechnol* 2014;98:8443–8455.

99. **Nguyen M, Ekstrom A, Li X, Yin Y**. HGT-Finder: A new tool for horizontal gene transfer finding and application to Aspergillus genomes. *Toxins* 2015;7:4035–4053.

100. **Innan H, Kondrashov F**. The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics* 2010;11:97–108.

101. **Tamayo EN, Villanueva A, Hasper AA, Graaff LH de, Ramón D, *et al.*** CreA mediates repression of the regulatory gene *xlnR* which controls the production of xylanolytic enzymes in *Aspergillus nidulans*. *Fungal Genetics and Biology* 2008;45:984–993.

102. **Kowalczyk JE, Gruben BS, Battaglia E, Wiebenga A, Majoor E, *et al.*** Genetic interaction of *Aspergillus nidulans galR*, *xlnR* and *araR* in regulating D-galactose and L-arabinose release and catabolism gene expression. *PLOS ONE* 2015;10:e0143200.

103. **Vienken K, Scherer M, Fischer R**. The Zn(II)2Cys6 putative *Aspergillus nidulans* transcription factor repressor of sexual development inhibits sexual development under low-carbon conditions and in submersed culture. *Genetics* 2005;169:619 LP – 630.

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

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

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

107. **Dyer PS, O'Gorman CM**. Sexual development and cryptic sexuality in fungi:

insights from Aspergillus species. *FEMS Microbiology Reviews* 2012;36:165–192.

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 [14]. B) Pie charts showing the number of genes coding for TFs with a standard name (left), the number of potential TFs predicted to include a signal peptide in their sequences (middle), or those predicted to contain TM domains (right). See also Tables S2, S3 and S4, and Figure S1.

<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 (green), the fraction of those that should be less or more extensively re-annotated (red) and the fraction of those that could not be assessed (mostly due to their very low or null expression levels) (yellow). Approximately 30.2% of the TF-coding genes would need some re-annotation. B) Null expression levels of *AN3391*, predicted to encode a C2H2- plus a Zn2Cys6-type dual-specificity TF, in comparison to those of *AN3390*, encoding 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 [28]. 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 [28], carbon starvation and a null *clrB* mutant [29], and null *mtfA* or null *veA* backgrounds [30]. 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 would generate 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. The fifth predicted intron of *AN11098* is not present, adding, compared to the annotated sequence, 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. The first predicted intron of *AN1812/jlbA* (G) is not present, generating a long 5´-UTR region. The probable start codon will be that for Met175, retaining only the sequence encoding the bZIP domain. The predicted two introns of *AN9221* (H) are probably absent, generating two coding sequences, *alnG* and *alnR* of the asperlin cluster (see main text). The coding region of *alnR*, a TF of the asperlin metabolic cluster, will probably begin with the codon for Met246 (FungiDB coordinates). The first predicted intron of *AN5252* (or *An9240*) (I), is not present, 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. aSp: alternative stop codon. RF: Reading frame. See also Table S5.

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; see main text), 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 S2 and Table S6.

<u>Figure 4</u>: **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 of TF-coding genes in response to A) the addition of sugarcane bagasse as the main nutrient source, or B) deletion of *sltA*. Panels C and D show expression profiles of specific genes selected from the groups highlighted in panels A and B, respectively. C) Upregulation in a medium containing sugarcane bagasse. D) Downregulation in a null *sltA* background in the three conditions tested (standard minimal medium or *unstressed*, alkalinity and sodium stress; upper graph), downregulation only at pH 8 (middle) or downregulation both at pH 8 and sodium stress (bottom).

<u>Figure 5</u>: 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 square in the AN0396 and AN5078 clade highlights the proximity of both sequences. Bootstrap values are indicated by blue circles. 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; [31]) 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 S7.

<u>Figure 6</u>: 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. Gray 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. Bootstrap values are indicated by light purple circles. 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 S3, S4 and S5, respectively.

Figure 7: Similarities of the C2H2 domain of BrlA with those of specific regulators of sexual development and stress-response. A) 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 [70]) 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. B) Phenotype of a strain expressing a BrlA::HA_{3x} chimera in which the C2H2 domain of BrlA (residues F320-H375) had been replaced by that 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.

Figure 8: Phenotypes of single-null mutant strains of five TF-coding genes potentially necessary for *A. nidulans* development. Phenotypes of the null mutants in ACM (Aspergillus complete medium) 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 dilution of the nitrogen source 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.

Supplementary Material

Table S1: Oligonucleotides used in this study.

Table S2: 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 S3: TF-coding genes of *A. nidulans* having a standard name.

Table S4: 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 S5: 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 S6: 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 Figure S1 and Figure 3 for more information and references.

Table S7: HMMER analyses of predicted bHLH-type TFs ANN5078, 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*

proteome in the AspGD database. Corrected amino acid sequences were used based on mapping of RNA-seq reads [31]**.**

Table S8: Preliminary 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 [28] **are included. See main text. FPKM values correspond to a single sample per culture condition.**

Table S9: 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 [28] **as well as the types of transcriptional regulatory domains encoded by these genes are included. See main text. FPKM values correspond to a single sample per culture condition.**

Table S10: Genomic location of genes coding for the Zn2Cys6 family TFs identified using CD-Hit. Genes are ordered based on the hypothetical paralog clusters they belong to. Genomic locations of each gene were retrieved from the FungiDB database. Expression levels were retrieved from Table S6. See also Figure S7.

Figure S1: Draft tree generated by MAFFT and edited using iTOL of the TFome of *A. nidulans*. The tree shows 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).

Figure S2: 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; see main text), 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 S6.

Figure S3: **Alignment of the predicted amino acid sequences of AN3385 (used as query in Figure 6B), AN7896/DbaA, AN2036, AN6788, AN1029/AfoA 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.

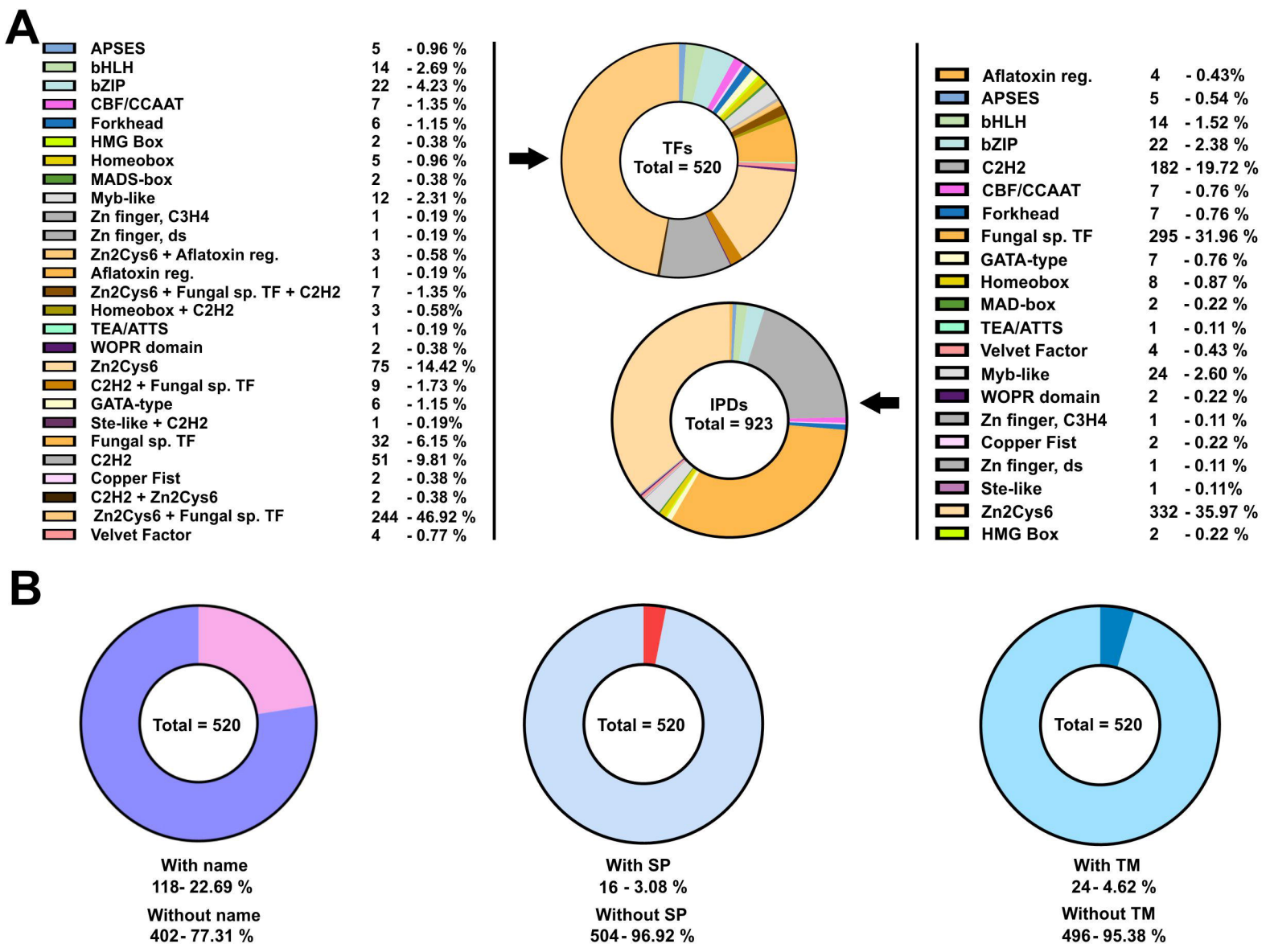
Figure S4: **Alignment of AN11169 (used as query in Figure 6C), AN10906, An8636, AN5955, AN2375 and AN8535 as a probable paralog cluster of TFs identified using the CD-Hit suite**. The alignment was carried out using Clustal Omega and visualized using Genedoc.

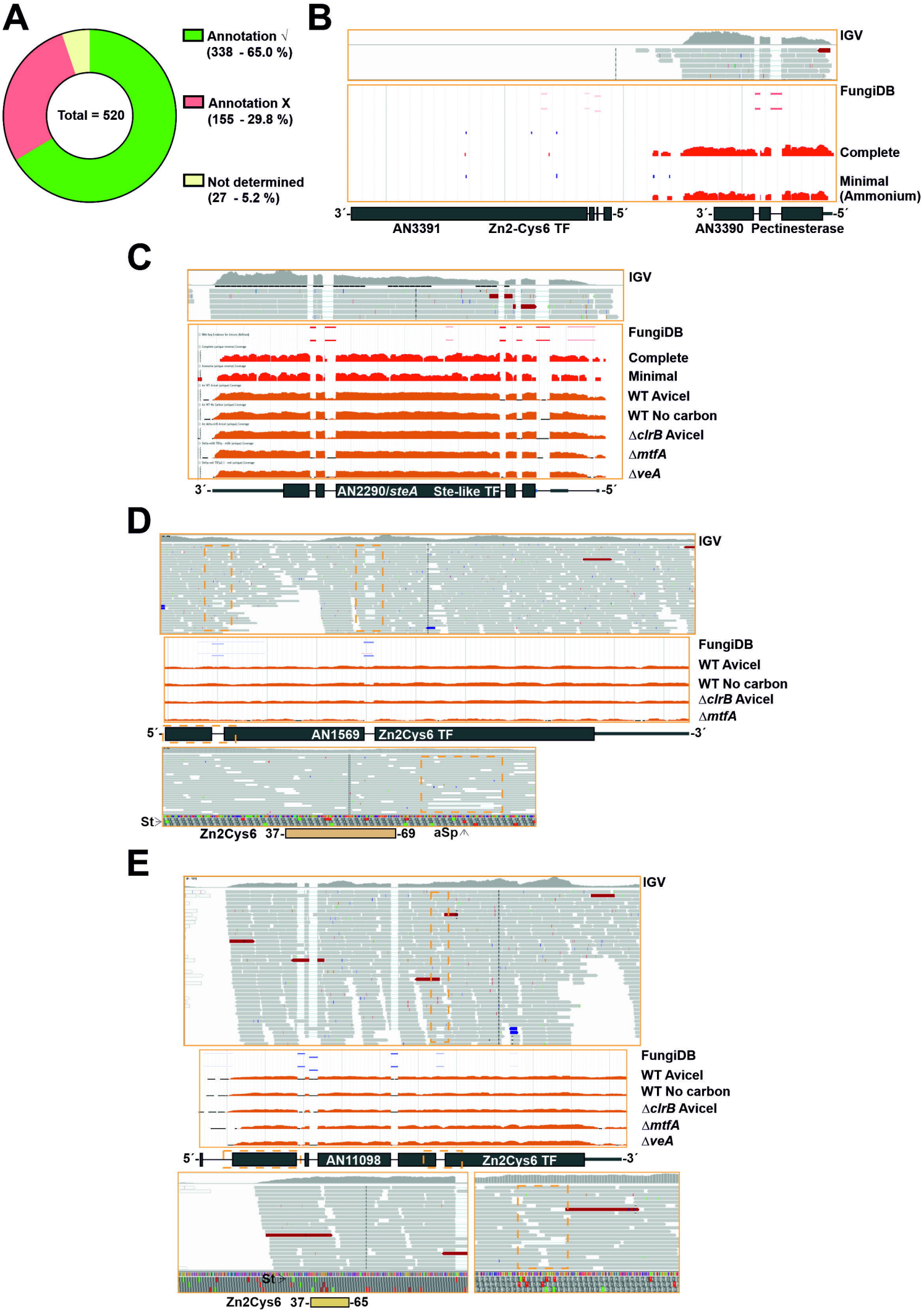
Figure S5: **Alignment of AN8590 (used as query in Figure 6D), AN7507, AN4972, AN11185, AN8930 and AN0742 as a probable paralog cluster of TFs identified using the CD-Hit suite**. The alignment was carried out using Clustal Omega and visualized using Genedoc.

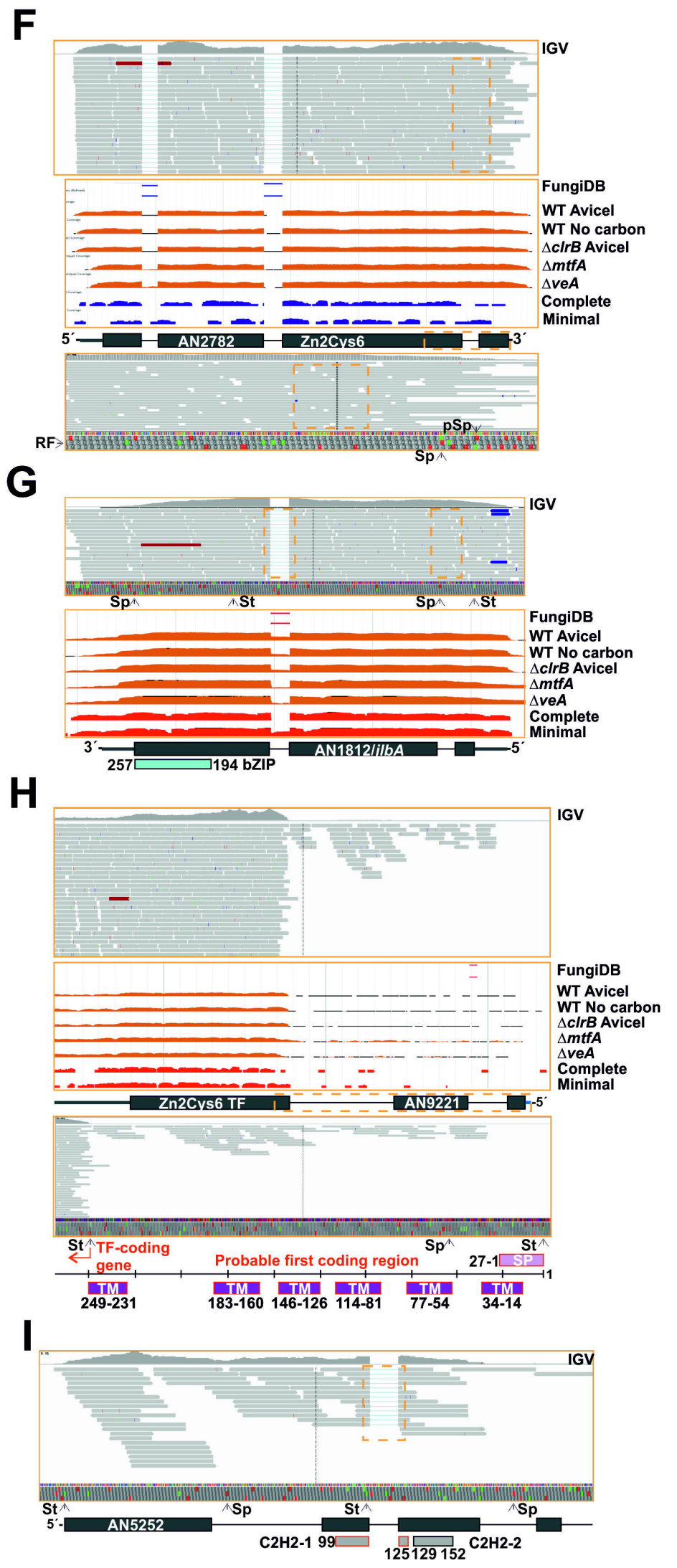
Figure S6: Nitrogen starvation apparently induces the expression of multiple genes encoding developmental regulators. A) Heatmap showing the expression pattern of genes with a FC ≥ 3 (lower clade) and a FC ≤ 3 (upper clade) when comparing the transcriptomes of samples collected in AMM with ammonium as the nitrogen source (third column) and 72 hours of nitrogen starvation (fifth column). The list of genes and the corresponding approximate FPKM values in each of the five conditions tested were retrieved from the FungiDB database (see Table S8), based on RNA-seq data by Sibthorp and colleagues [28]. These values correspond to a single sample per culture conditions and must be considered as preliminary data. B) Number of genes with a FC ≥

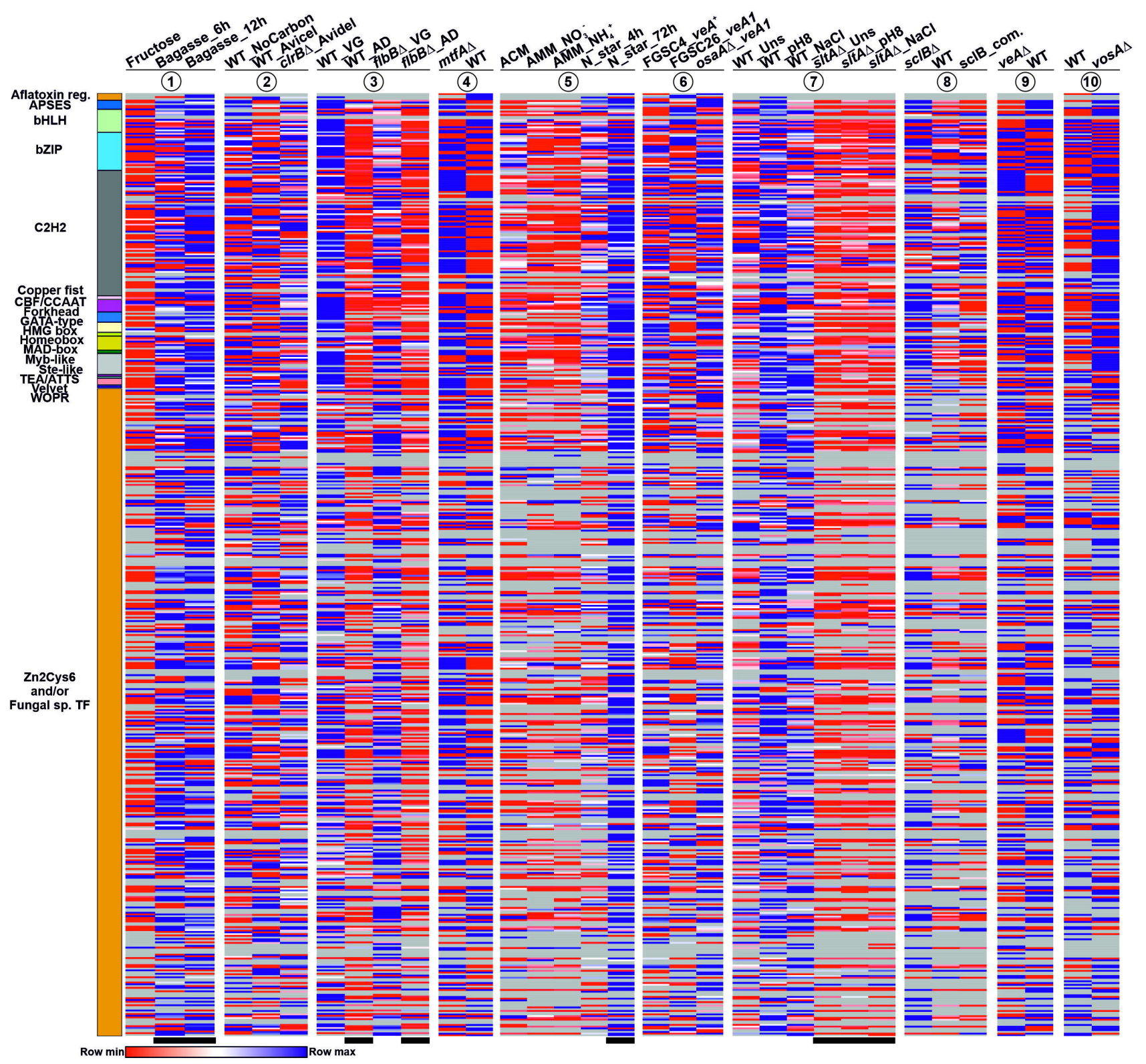
3 or a FC ≤ 3 under nitrogen starvation (72 h) compared to AMM containing ammonium as the nitrogen source. The ratio of up- versus downregulated genes is 2.84, suggesting that nitrogen starvation could induce a transcriptional response. C) Bar plot showing the number of upregulated genes (FC ≥ 3) coding 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 S9). 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 (orange bars) and 72 hours (green bars) after nitrogen starvation was induced. FPKM values were taken from [28], and in all the cases shown, expression seems to be induced under nitrogen starvation. F) Same bar-plot but for the genes selected for deletion as potential regulators of development (see the phenotypes in Figure 8).

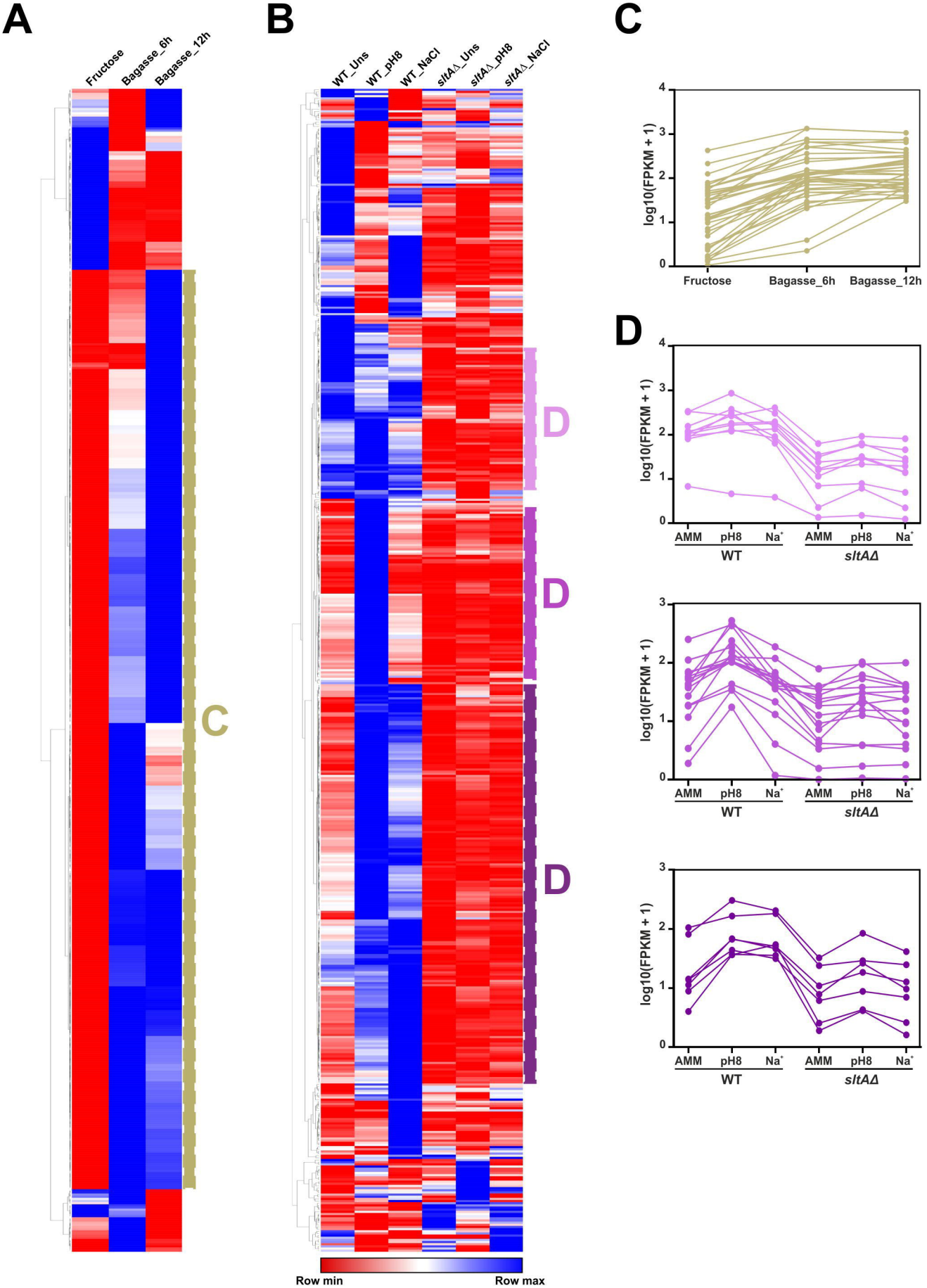
Figure S7: Distribution in the genome of *A. nidulans* of the genes coding for the Zn2Cys6-type TFs that probably belong to paralog clusters. Only those genes coding for Zn2Cys6-type TFs and identified by the CD-Hit suite as members of probable paralog clusters were considered. The image was downloaded 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.

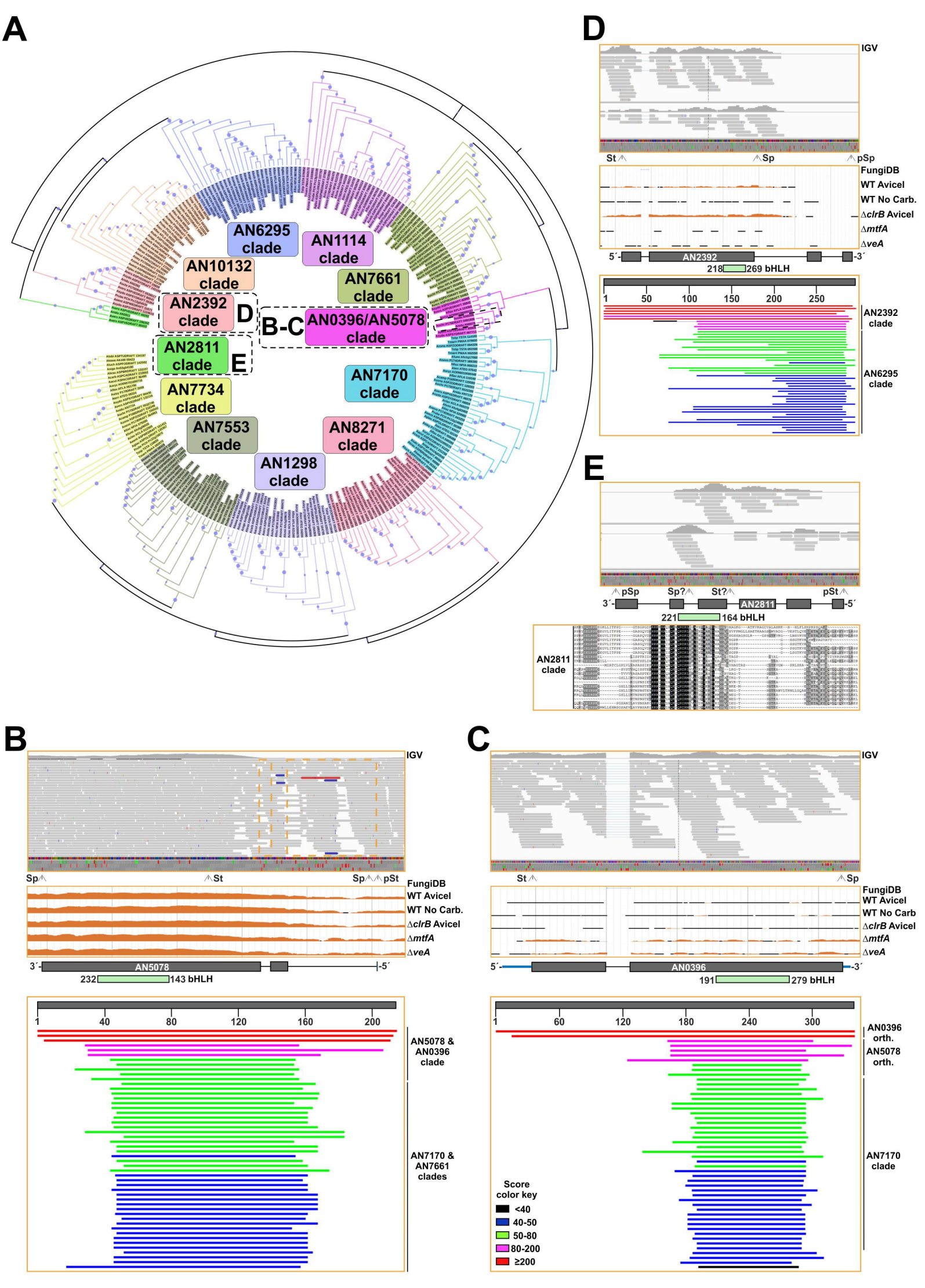

A
APSES 5 - 0.96 %
bHLH 14 - 2.69 %
bZIP 22 - 4.23 %
CBF/CCAAT 7 - 1.35 %
Forkhead 6 - 1.15 %
HMG Box 2 - 0.38 %
Homeobox 5 - 0.96 %
MADS-box 2 - 0.38 %
Myb-like 12 - 2.31 %
Zn finger, C3H4 1 - 0.19 %
Zn finger, ds 1 - 0.19 %
Zn2Cys6 + Aflatoxin reg. 3 - 0.58 %
Aflatoxin reg. 1 - 0.19 %
Zn2Cys6 + Fungal 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 %

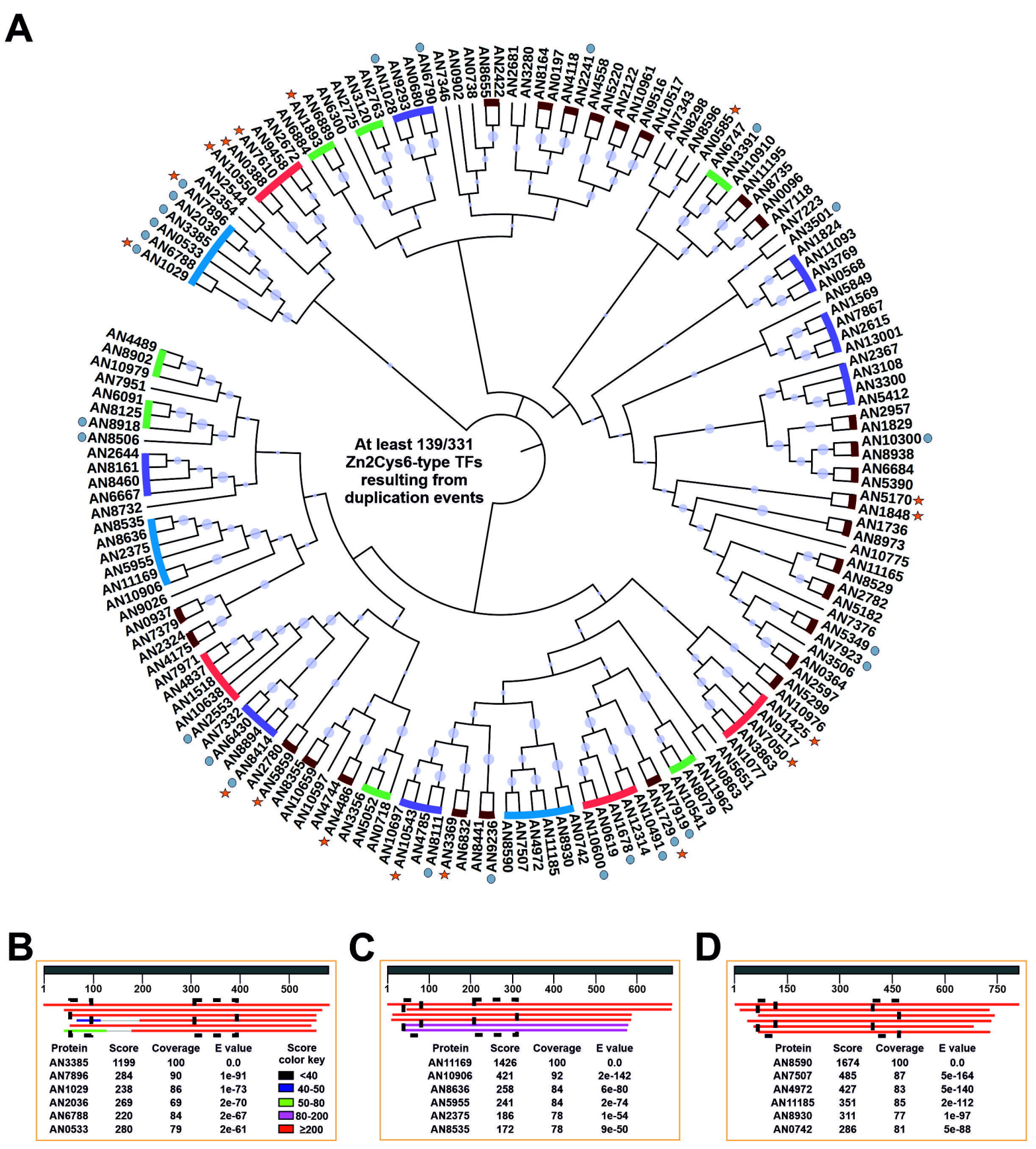

A
Total = 520
Annotation √
(338 - 65.0 %)
Annotation X
(155 - 29.8 %)
Not determined
(27 - 5.2 %)
B
IGV
FungiDB
Complete
Minimal
(Ammonium)
3´-
-5´
AN3391
Zn2-Cys6 TF
3´-
-5´
AN3390
Pectinesterase
C
IGV
FungiDB
Complete
Minimal
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
3´-
AN2290/steA
Ste-like TF
-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

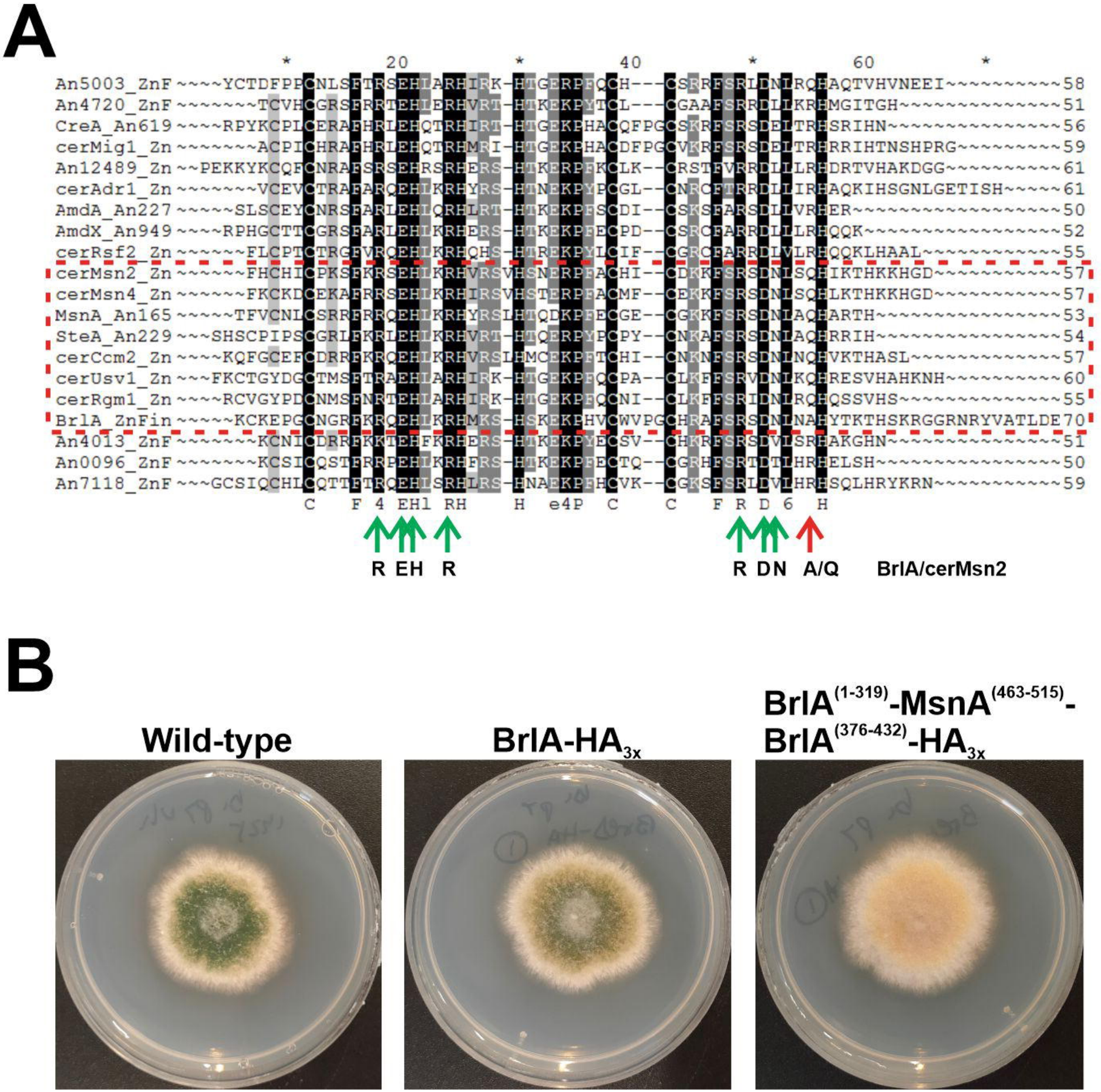

F
IGV
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
Complete
Minimal
5´
AN2782
Zn2Cys6
3´
pSp
RF
Sp
G
IGV
Sp
St
Sp
St
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
Complete
Minimal
3´
AN1812/ilbA
5´
257
194 bZIP
H
IGV
FungiDB
WT Avicel
WT No carbon
ΔclrB Avicel
ΔmtfA
ΔveA
Complete
Minimal
Zn2Cys6 TF
AN9221
5´
St
TF-coding gene
Probable first coding region
Sp
St
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

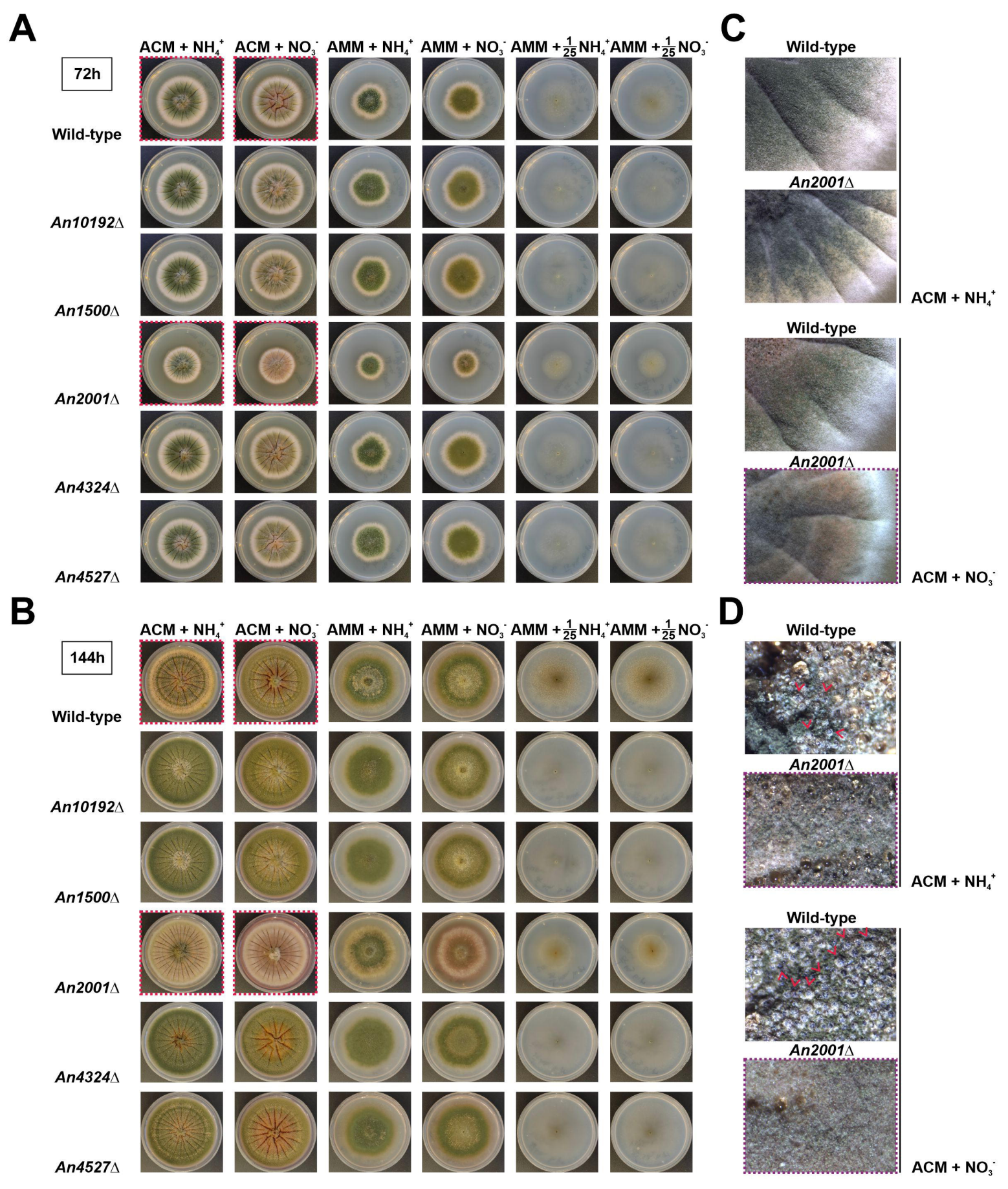

Fructose
Bagasse_6h
Bagasse_12h
WT_NoCarbon
WT_Avicel
clrBΔ_Avidel
WT_VG
WT_AD
flbBΔ_VG
flbBΔ_AD
mtfAΔ
WT
ACM
AMM_NO3-
AMM_NH4+
N_star_4h
N_star_72h
FGSC4_veA+
FGSC26_veA1
osaAΔ_veA1
WT_Uns
WT_pH8
WT_NaCl
sltAΔ_Uns
sltAΔ_pH8
sltAΔ_NaCl
sclBΔ
WT
sclB_com.
veAΔ
WT
WT
vosAΔ
1
2
3
4
5
6
7
8
9
10
Aflatoxin 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
C
B
WT_Uns
WT_pH8
WT_NaCl
sltAΔ_Uns
sltAΔ_pH8
sltAΔ_NaCl
D
D
D
Row min
Row max
C
log10(FPKM + 1)
Fructose
Bagasse_6h
Bagasse_12h
D
log10(FPKM + 1)
AMM
pH8
Na+
WT
sltAΔ


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
AN1029 238 86 1e-73
AN2036 269 69 2e-70
AN6788 220 84 2e-67
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
20
40
60
An5003_ZnF
An4720_ZnF
CreA_An619
cerMig1_Zn
An12489_Zn
cerAdr1_Zn
AmdA_An227
AmdX_An949
cerRsf2_Zn
cerMsn2_Zn
cerMsn4_Zn
MsnA_An165
SteA_An229
cerCcm2_Zn
cerUsv1_Zn
cerRgm1_Zn
BrlA_ZnFin
An4013_ZnF
An0096_ZnF
An7118_ZnF
R EH R
R DN A/Q
BrlA/cerMsn2
B
Wild-type
BrlA-HA3x
BrlA(1-319)-MsnA(463-515)-BrlA(376-432)-HA3x


A
ACM + NH4+
ACM + NO3-
AMM + NH4+
AMM + NO3-
AMM + 1/25 NH4+
AMM + 1/25 NO3-
72h
Wild-type
An10192Δ
An1500Δ
An2001Δ
An4324Δ
An4527Δ
B
144h
C
Wild-type
An2001Δ
ACM + NH4+
ACM + NO3-
D