

Rothamsted Repository Download

A - Papers appearing in refereed journals

Lawler, K., Hammond-Kosack, K. E., Brazma, A. and Coulson, R. M. R. 2013. Genomic clustering and co-regulation of transcriptional networks in the pathogenic fungus *Fusarium graminearum*. *BMC Systems Biology*. 7, p. 52.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.1186/1752-0509-7-52>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/8qx43>.

© 27 June 2013. Licensed under the Creative Commons CC BY.

RESEARCH ARTICLE

Open Access

Genomic clustering and co-regulation of transcriptional networks in the pathogenic fungus *Fusarium graminearum*

Katherine Lawler^{1,2*}, Kim Hammond-Kosack³, Alvis Brazma¹ and Richard MR Coulson^{1,4}

Abstract

Background: Genes for the production of a broad range of fungal secondary metabolites are frequently colinear. The prevalence of such gene clusters was systematically examined across the genome of the cereal pathogen *Fusarium graminearum*. The topological structure of transcriptional networks was also examined to investigate control mechanisms for mycotoxin biosynthesis and other processes.

Results: The genes associated with transcriptional processes were identified, and the genomic location of transcription-associated proteins (TAPs) analyzed in conjunction with the locations of genes exhibiting similar expression patterns. Highly conserved TAPs reside in regions of chromosomes with very low or no recombination, contrasting with putative regulator genes. Co-expression group profiles were used to define positionally clustered genes and a number of members of these clusters encode proteins participating in secondary metabolism. Gene expression profiles suggest there is an abundance of condition-specific transcriptional regulation. Analysis of the promoter regions of co-expressed genes showed enrichment for conserved DNA-sequence motifs. Potential global transcription factors recognising these motifs contain distinct sets of DNA-binding domains (DBDs) from those present in local regulators.

Conclusions: Proteins associated with basal transcriptional functions are encoded by genes enriched in regions of the genome with low recombination. Systematic searches revealed dispersed and compact clusters of co-expressed genes, often containing a transcription factor, and typically containing genes involved in biosynthetic pathways. Transcriptional networks exhibit a layered structure in which the position in the hierarchy of a regulator is closely linked to the DBD structural class.

Keywords: Transcriptional networks, DNA-binding domains, mycotoxin biosynthesis, filamentous fungi, gene clusters

Background

Bacterial genes are organised into co-transcribed operons sharing a common promoter, with coding sequences present within operons frequently producing polypeptides of related function [1]. The genomes of nematodes and trypanosomes also exhibit polycistronic transcription of gene clusters [2,3], though most eukaryotic genes are generally considered to be monocistronic, each with its own

promoter and transcription terminator [4]. This implies that eukaryotic genes do not have to be in close proximity to be co-expressed, and that their organization within a genome could be random. However, it appears that genes having similar and/or coordinated expression are often clustered [5]: the genes comprising the vertebrate β -globin cluster are organised according to the timing of their expression during development [6] with members of the mammalian homeobox transcriptional regulator loci arranged according to their spatial pattern of expression along developmental axes [7]. Additionally, in all crown-group eukaryotes (organisms found at the top of molecular phylogenetic trees) there is a significant tendency for genes from the same metabolic pathway to cluster [8].

* Correspondence: katherine.lawler@kcl.ac.uk

¹European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK

²Institute for Mathematical and Molecular Biomedicine, King's College London, Hodgkin Building, London SE1 1UL, UK

Full list of author information is available at the end of the article

This suggests that higher order genome organisation is linked to expression patterns.

Gene clusters have been described in fungi for phenotypes as varied as nutrient use, mating type and pathogenicity [9]. Further, genes for production of a broad range of secondary metabolites are located adjacent to one another; additionally a pathway-specific regulatory gene is often embedded within these clusters. The existence in budding yeast of a higher-order organization of genes across chromosomes is constrained by transcriptional regulation, with the target genes of most transcription factors positionally clustered within chromosomes [10]. The molecular mechanisms underpinning co-expression are not understood, though incidences seem to fall into two categories acting: (i) on a relatively local scale and dependant on *in cis* promoters in the immediate vicinity, and (ii) over broad genomic spans, possibly involving *in trans* chromatin state and positioning within the nucleus [5]. Price *et al.* [11] have hypothesized that as the amount of regulatory information required to specify an optimal expression pattern increases, evolving the optimal expression profile separately for each gene becomes more difficult, whilst creating an operon does not. Hence, co-expression in eukaryotes could reduce stochastic differences in gene expression and also synchronise fluctuations, or noise, in the levels of components of pathways and complexes [12].

The filamentous fungus *Fusarium graminearum* is a major cause of blight in cereal crops, resulting in heavy losses to grain yield and quality, which can be exacerbated by the contamination of grain with various mycotoxins that pose a serious threat to food and feed safety [13]. These secondary metabolites - not essential for survival - have been shown in a few cases to be synthesized by gene clusters. For example, the *TRI*-gene cluster contains up to 14 genes coding for proteins involved in production of harmful B-type trichothecene deoxynivalenol (DON) and its acetylated derivatives [14]. In *Saccharomyces cerevisiae*, clustering of

essential genes increases the robustness of populations to mutation, and may provide a significant selective force shaping meiotic crossover distribution [15]. Little or no recombination is observed over long sections (megabases) across *F. graminearum* chromosomes, followed by much shorter regions displaying considerably higher than average recombination rates [16-18]. This contrasts strikingly with *S. cerevisiae* where high densities of crossing over are often present within just a few kilobases [19] compared with the several hundred kb in *Fusarium*. Here, the impact of this pattern of recombination in *F. graminearum* on the composition of co-expressed gene clusters was investigated, in conjunction with the genomic locations and protein-domain composition of the genes controlling the transcriptional process.

Results

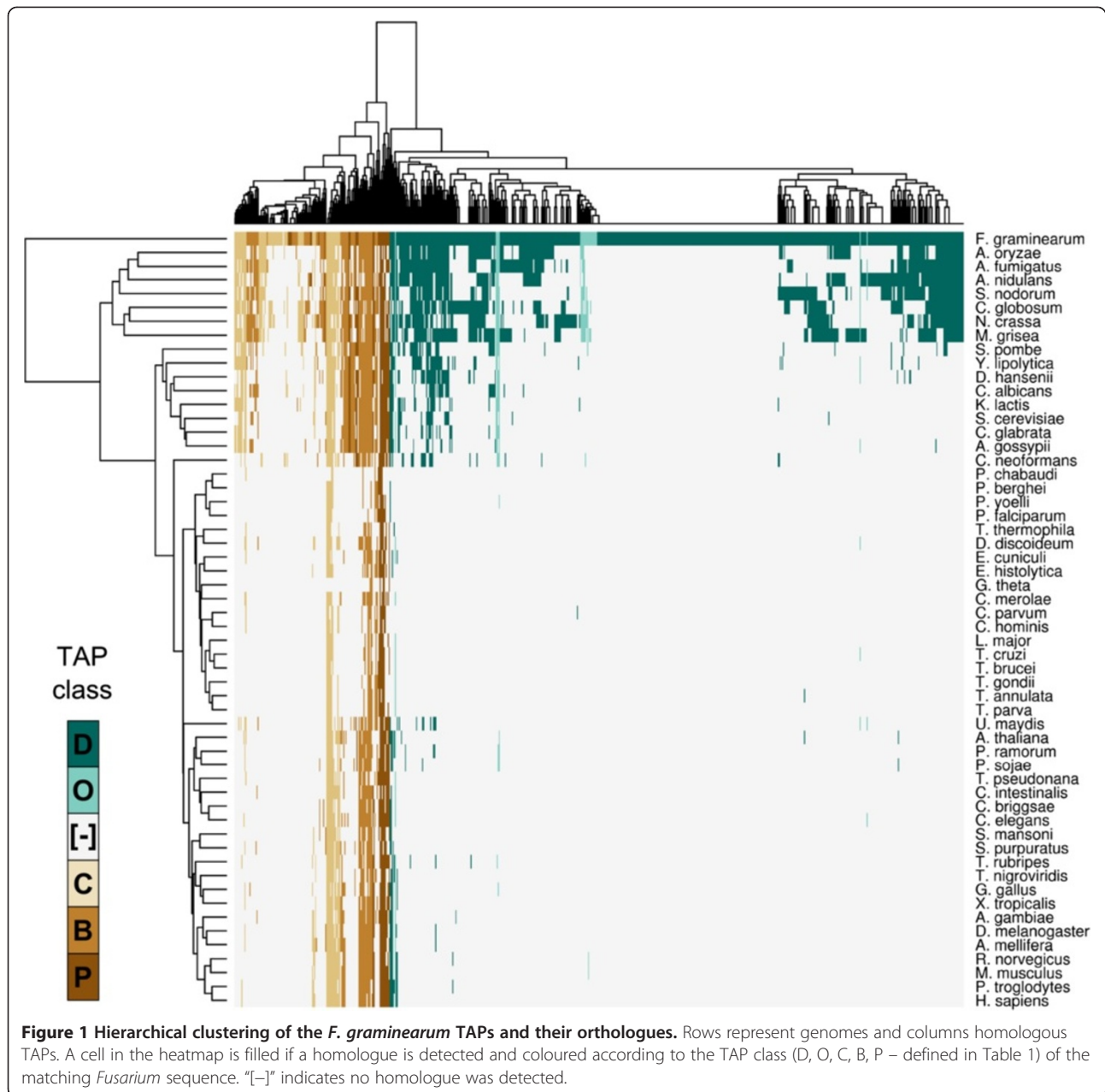
The *F. graminearum* genes encoding proteins associated with the transcriptional process were identified by protein family detection and profile matching (see Methods and Additional file 1: Table S1): of the 14,100 protein entries comprising the *F. graminearum* genome [17], 723 were linked to transcription (transcription-associated proteins, TAPs - Table 1). Sequences orthologous to these 723 TAPs were obtained from 56 complete eukaryotic genomes through sequence searching ('Detection of *F. graminearum* TAP orthologues' in Methods), and placed into one of five categories derived from the TAP reference set functional annotations, namely basal transcription factors and cofactors (B), RNA polymerase subunits (P), DNA binding (D), chromatin remodelling and histone modification factors (C) and others (O).

The phylogenetic distribution of the TAP orthologues displays a high correspondence with their TAP class (Figure 1): one-third of the *F. graminearum* TAPs encoding DNA-binding proteins are only observed within the species (Table 1), and nearly a half (258/546) just have orthologues in filamentous fungi (*Pezizomycotina*).

Table 1 Distribution of *F. graminearum* genes homologous to the TAP reference set and matching DNA-binding domain HMMs

TAP class	Description	Count	Taxonomic distribution			
			<i>Fgr</i>	<i>Pez.</i>	<i>Fungi</i>	<i>Euk</i>
B	Basal transcription factors and co-factors	63	4	10	15	34
C	Chromatin remodelling and histone modification	63	8	14	19	22
D	DNA-binding proteins	546	179	258	89	20
P	RNA polymerase subunits	27	4	4	4	15
O	Unclassified (CCR-NOT subunits, non-DNA-binding factors)	24	6	8	3	7
Total		723				

The TAP classes (B, C, D, P, O) were split into four taxonomic groups according to whether homologous genes were found only in *F. graminearum* (*Fgr*), only in *Pezizomycotina* (*Pez*) genomes, only in fungal genomes (*Fungi*), or widely detected within the eukaryotic genomes (*Euk*).



This contrasts with the four other categories, where nearly 60% of the B- and P-TAPs have orthologues in all the eukaryotic genomes analyzed.

Chromosomal distribution of TAP functional classes

The association between the degree of *F. graminearum* TAP conservation across eukaryotes and the recombination rate of the chromosomal region the TAP lies within was examined by delineating the genome into four groups by recombination rate (R , cM/27kb) [17,20]: these groups ranged in value from no or very low ($R < 1$) to very high

($R \geq 8$) recombination rates. TAP classes B, C and P are under-represented amongst genes in regions of very high recombination rate, contrasting with their significant enrichment in areas of low or no recombination (Figure 2A,B). This under-representation is also seen with DNA-binding TAPs (D-TAPs) having homologues in metazoan genomes (Figure 1), as none are seen in areas of high or very high recombination ($R \geq 3$). However, the percentage of D-TAPs increases in areas of high recombination as the clades become taxonomically less diverse i.e. twice the proportion of *Fusarium*-specific

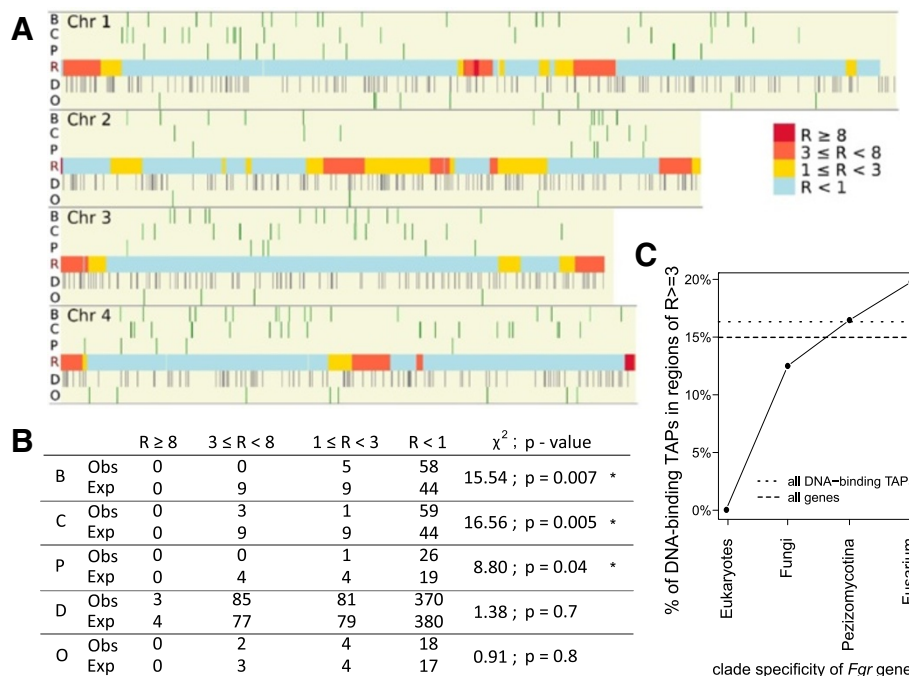


Figure 2 Chromosomal distribution of *F. graminearum* TAPs. (A) The genome was divided into four blocks representing high ($R \geq 8$) to low ($R < 1$) recombination rate (cM/27kb). Chromosomal locations of the B, C, P TAPs and D, O TAPs are shown above and below, respectively, of the recombination map. (B) The observed (Obs) and expected (Exp) TAP gene counts are shown per recombination block for each TAP class, compared with a uniform distribution of TAPs over *F. graminearum* gene positions (χ^2 -test; *: $p < 0.05$). (C) Graph of the percentage of clade-specific DNA-binding TAPs lying in regions of high recombination ($R \geq 3$). The percentage of all DNA-binding TAPs (short dotted line), and of all genes (long dotted line), lying in regions with high recombination rate are indicated.

D-TAPs lie in these regions compared with those that have homologues in the fungal species examined (Figure 2C). These observations imply that proteins highly conserved, and associated with the transcriptional process, reside within areas of minimal recombination.

Condition dependence of TAP gene expression

An analysis of transcriptome data was undertaken to define TAP expression patterns and co-expressed gene clusters under a variety of environmental conditions. Several microarray gene expression data sets were selected: three spanning the *F. graminearum* lifecycle (FG1, infection of susceptible barley ears; FG5 & FG6, sexual spore development *in vitro*) and one a comparison of mycelium growth under nutrient rich and two different nutrient limiting conditions (FG2) (Figure 3A and Methods). Overall, 4,477 of the 13,830 genes represented on the array (32%) were identified as differentially expressed within at least one data set (Figures 3B and C, Additional file 1: Table S2), of which over two-thirds were observed to be differentially expressed only in one experiment (3,056 – the sum of the unique portions of the four gene sets, Figure 3C). Similarly, one-third of the DNA-binding TAPs - 155 (the union of the four DNA-binding gene sets) of the 536 represented on the array - were found to be differentially

expressed (Figure 3D, E), again with the majority (80%) in only one experiment (Figure 3E). These analyses suggest there is an abundance of condition-specific control of *Fusarium* transcript levels. Within each of the four experiments, on average 15% of expressed non-TAPs, compared with 8.6% of expressed TAPs show differential expression; additionally, of all the differentially expressed DNA-binding TAPs, none exhibit altered expression levels in all four experiments i.e. the intersection of all four sets of these genes is zero (Figure 3E).

Gene expression patterns for probesets differentially expressed during the time course experiments were merged on the basis of similar temporal behaviours to create co-expression groups. This procedure compared subsequent timepoints with the earliest timepoint, to ascertain whether the expression level of the probeset was either stably (\uparrow or \downarrow , Figure 4) or transiently ($\uparrow\downarrow$ or $\downarrow\uparrow$, Figure 4) altered. For example in FG5, 806 and 426 display stable increased (\uparrow) and decreased (\downarrow) levels, respectively, with 316 and 247 exhibiting transient elevated ($\uparrow\downarrow$) and reduced ($\downarrow\uparrow$) levels (Figure 4). All other changes in behaviour are indicated by a tilde (~). In the co-expression groups observed in the FG1, FG5 and FG6 microarray experiments, stable alterations in transcript levels predominate over transient ones.

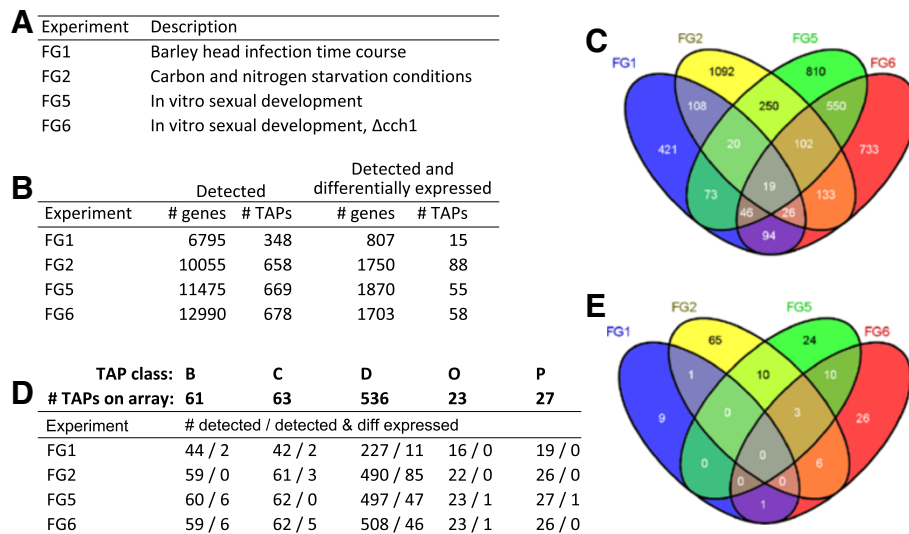


Figure 3 Genes expressed in transcriptomics experiments. (A) Gene expression data sets used in the study, listed by PlexDB identifier and description. (B) Number of genes expressed (Detected – ‘# genes’) and the number which are TAPs (‘# TAPs’) within each *F. graminearum* transcriptomics data set, and number that are also differentially expressed (‘Detected and differentially expressed’). (C) Venn diagram [21] summarizing the distribution of differentially expressed genes across the four sets of conditions. (D) Number of TAPs in each TAP class which were detected, or detected and differentially expressed, within each experiment (B, C, D, O, P; Table 1). (E) Venn diagram showing the overlap of DNA-binding TAPs differentially expressed in the microarray experiments.

Genomic clustering of co-expressed genes and TAPs

Target genes of transcription factors are positionally clustered within budding yeast chromosomes [10], and in filamentous fungi there are secondary metabolite gene clusters known to be co-regulated (e.g. aflatoxin/sterigmatocystin and trichothecene biosynthesis in *Aspergillus* and *Fusarium*, respectively [22,23], and reviewed in [9,24]). To examine whether the genes present in the co-expression groups defined from the *Fusarium* transcriptomics experiments (Figure 4) also exhibit close proximity within the genome, several clustering methods were employed (Figure 5).

Global tests of clustering of co-expressed genes amongst all genes on the genome would not reveal densely packed, localized clusters in a background of sparsely distributed co-expressed genes, i.e. they would not necessarily distinguish localised organisation in a noisy background. To look for initial evidence of such clusters, the distance (in gene number) between consecutive pairs of co-expressed genes was measured within each of the 22 co-expression groups (Figure 5A; see Methods). In 9 of the 22 co-expression groups, significantly more genes were observed to be in close proximity (up to 10 genes apart on the genome) than expected if they were randomly distributed (Additional file 2: Table S3). This is consistent with the existence of localized clusters, but does not distinguish between tightly packed co-expressed clusters or dispersed proximal pairs of co-expressed genes.

The manner of localized clustering in gene order within the co-expression groups was further investigated

by examining the expression patterns of genes in the vicinity of each differentially expressed TAP (Figure 3). For each differentially expressed TAP, the number of genes within the same co-expression group was counted for a variable window size of 2 to 40, i.e. 1 to 20 genes on each side of the TAP (Figure 5B). Fisher’s exact test was used to determine significance of enrichment of group members for a given window size, and resulted in the detection of seven TAP-centred clusters (TC; Figures 5D and 6). Five of the TAP-centred clusters (TCs) were observed in co-expression groups exhibiting non-random genomic distributions (Figure 5A). TC3 contains a gene encoding a polyketide synthase (PKS3/PGL1 – an enzyme involved in synthesizing black perithecial pigment) and it has been suggested that the sequences surrounding this PKS gene form a co-regulated group [18,25,26]. Furthermore, the members of TC3 show increasing expression during progression from vegetative mycelia to mature perithecia (FG5, FG6) reflecting an elevation in pigment synthesis during sexual development. Three out of the five genes present in the *F. graminearum* mating-type locus [27,28] were shown to comprise a co-expressed cluster (TC7), with this cluster containing an additional adjacent gene. Members of TC5 exhibit decreased levels of expression in nitrogen-minimal conditions and are highly homologous to the budding yeast MAL1 locus [29] which also contains a maltase, a Zn₂(II)Cys₆ DNA-binding factor and a maltose permease.

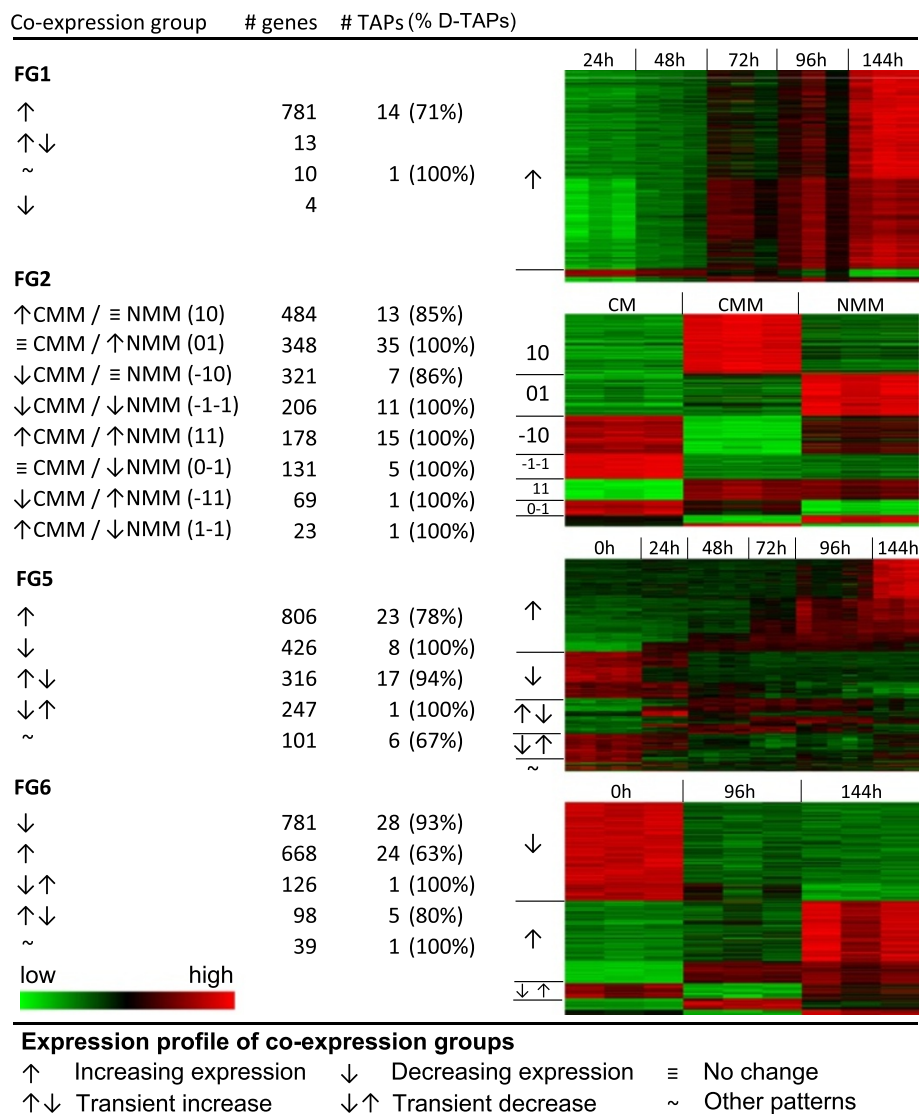


Figure 4 Co-expression groups derived from transcriptomics experiments. Column one shows the expression pattern observed in each co-expression group for a given data set, and columns two and three show the number of genes (# genes) and TAPs (# TAPs), respectively, with this expression profile. % D-TAPs indicates the percentage of DNA-binding TAPs in the co-expression group. The final column shows the heatmaps corresponding to the various co-expression groups; increasing (red) and decreasing (green) expression is shown relative to the initial time point for the developmental and infection studies, and relative to complete media for the nutrient-limited conditions.

The presence of localized clusters (LCs) independent of the presence of a proximal co-expressed TAP was investigated using the Positional Gene Enrichment (PGE) tool [30] (Figure 5; see Methods). Seventeen LCs were identified as significantly enriched for co-expressed genes (Figures 5 and 7; LC1-17). Four were identified as TAP-containing (Figures 6 and 7) as described above. Two further classes of cluster were observed: ones where there were no intervening genes that were not members of the same co-expression group (LCs 7, 11, 13, 14, 15 and 17), and having an average size of five genes. The other cluster type showed much less compactness with an average size of 64 genes, of which around a third were members of the co-expression

cluster. Several of these clusters contain sequences encoding nonribosomal peptide synthetases (NPS) – enzymes producing a wide range of mycotoxins and linked to *Fusarium* pathogenicity [31,32].

Putative protein functions for unannotated, co-expressed genes lying in TCs and LCs were derived from sequence and protein domain homology, and linking the results of these searches to Gene Ontology entries (Methods and Additional file 1: Table S5). Most of the 170 non-TAP genes appear to play a role in secondary metabolism (Figures 6 and 7), with 42% and 22% associated with metabolic and biosynthetic processes, respectively (Table 2). Hence, most of the co-expressed genes

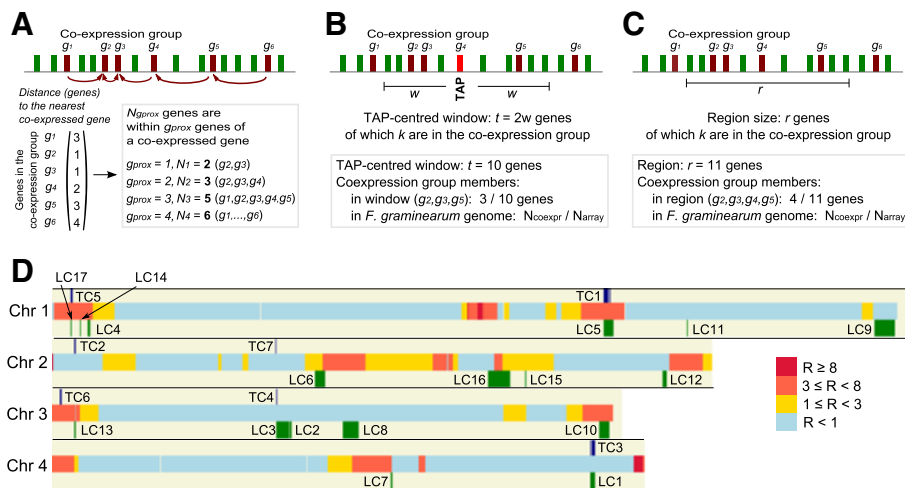


Figure 5 Methods testing for chromosomal clustering of co-expressed genes. (A) Genomic clustering of co-expressed genes. Within a given co-expression group (dark red), for each gene the distance to the nearest member was recorded. The total number (N_{gprox}) of genes lying within g_{prox} genes of a co-expressed gene was counted and compared to uniform sampling from the genome. (B) TAP-centred clusters of co-expressed genes. For each TAP (red), the genes lying within a window spanning t adjacent genes around the TAP were tested for enrichment for co-expressed genes (dark red). The window size t was increased incrementally ($t = 1$ to $t = 20$). (C) Localized clustering of co-expressed genes. Each genomic region was tested for enrichment of co-expressed genes using PGE. (D) Chromosomal locations of the seven TAP-centred (TC) and seventeen localized (LC) gene clusters detected by the TAP-window and PGE methods, respectively.

exhibiting positional clustering appear to encode proteins associated with secondary metabolism, indicating that co-regulation of gene clusters is primarily associated with controlling the biosynthesis of mycotoxins or other metabolites.

Defining global transcriptional regulators

Co-regulation of *Peizizomycotina* gene clusters encoding components of secondary metabolism pathways is partly coordinated through 'narrow'- and 'broad'-domain transcription factors [9], proteins primarily containing either

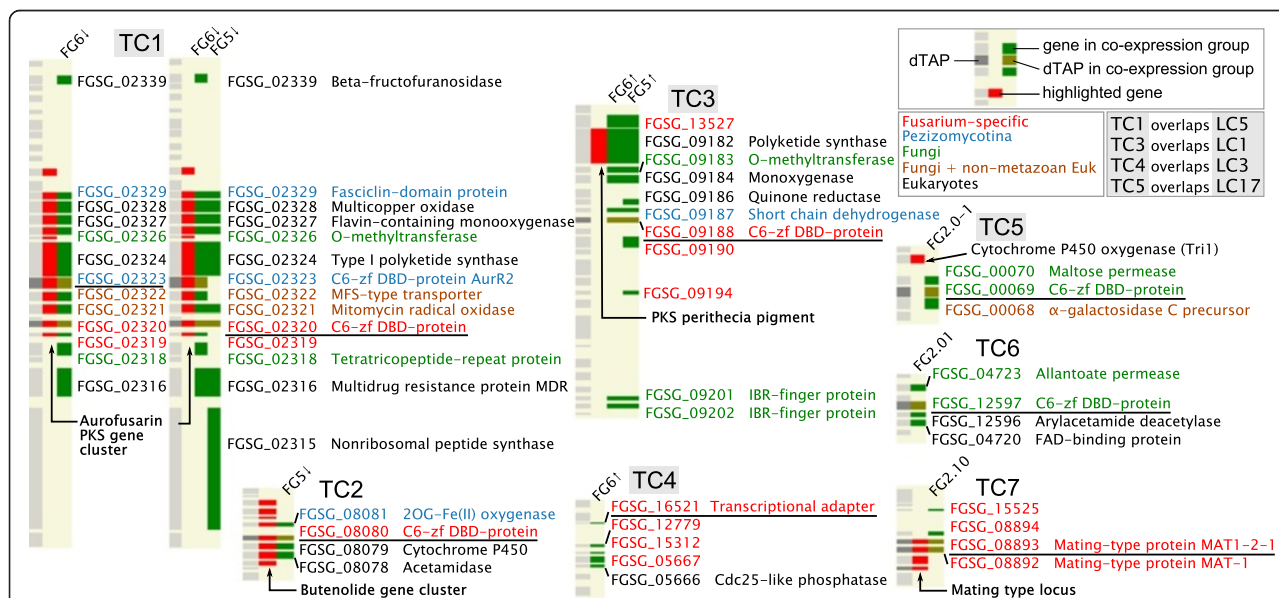


Figure 6 TAP-centred clusters. The eight TAPs (underlined) located within the seven regions significantly enriched for co-expressed genes are shown. Each cluster is displayed as three columns with the left showing the genes on the chromosome, the middle displaying genes of interest (red), and the right the members of the TC (non-DNA-binding TAPs shown in green, with D-TAPs in brown). The colour of the FG gene IDs (next to the chromosomal locations of the TC members) indicates taxonomic specificity, i.e. in which phylogenetic groups homologues are found. Alongside this identifier is given a functional annotation derived from the sequence searches and protein family detection. Abbreviations: C6-zf DBD-protein, Zn(II)₂Cys₆ DNA-binding protein.

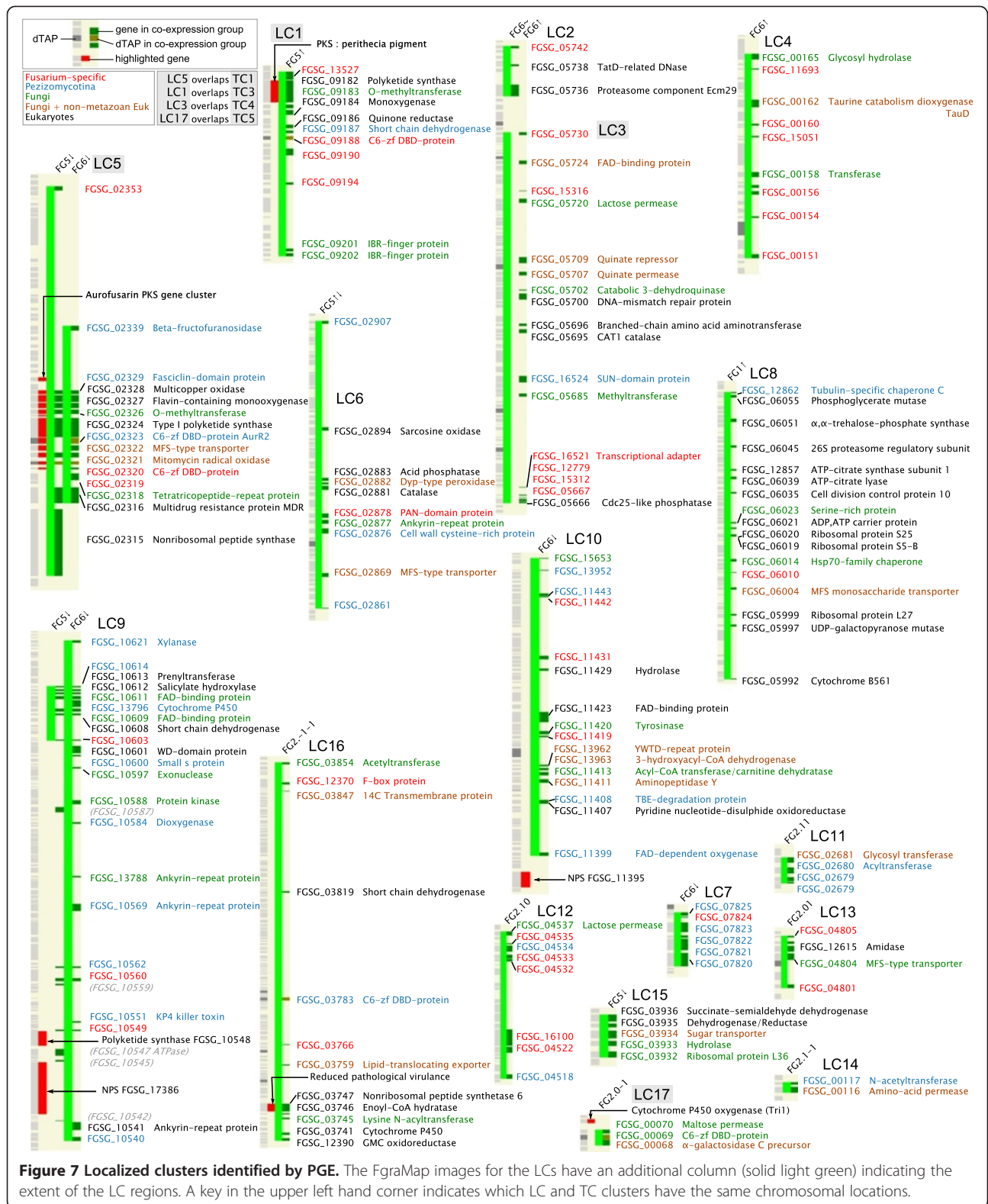


Figure 7 Localized clusters identified by PGE. The FgraMap images for the LCs have an additional column (solid light green) indicating the extent of the LC regions. A key in the upper left hand corner indicates which LC and TC clusters have the same chromosomal locations.

Zn(II)₂Cys₆ or Cys₂His₂ amino acid motifs, respectively. In FG1 there is an enrichment of genes expressed that code for proteins containing bZIP, GATA and Cys₂His₂

DNA-binding domains (DBDs): significantly more Cys₂His₂-containing genes are expressed than would be expected given the overall proportion of DNA-binding

Table 2 GO Slim assignments of the 170 non-TAP differentially expressed genes which are present in TCs and LCs (Figures 6 and 7)

GO Slim term	% of TC/LC genes	GO Slim ID
Metabolic process	42%	GO:0008152
Cellular process	39%	GO:0009987
Biosynthetic process	22%	GO:0009058
Macromolecular metabolic process	16%	GO:0043170
Regulation of biological process	16%	GO:0050789
Transport	14%	GO:0006810
Response to stimulus	12%	GO:0050896
Multicellular organismal development	11%	GO:0007275
Cell communication	10%	GO:0007154
Catabolic process	9%	GO:0009056
Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	8%	GO:0006139
Cellular amino acid and derivative metabolic process	6%	GO:0006519
Cell death	2%	GO:0008219
Cell differentiation	2%	GO:0030154
Behaviour	1%	GO:0007610
Pathogenesis	1%	GO:0009405

The percentage of these genes annotated with each term is shown.

TAPs expressed (Fisher's exact test, $p < 0.0005$; 55 of the 93 Cys₂His₂-D-TAPs are detected compared with 227 D-TAPs out of the 536 on the microarray (Figure 3D)). Conversely, a significant depletion in the expression of 'narrow'-domain Zn(II)₂Cys₆ transcription factors is observed (Fisher's exact test, $p < 10^{-6}$; 105 out of the 325 represented on the microarray).

The identification of genes controlling transcription of the members of co-expression groups was facilitated by Kumar *et al.* [33]: they reported the presence of 326 DNA-motifs located in upstream promoter regions of *F. graminearum* genes and conserved in *Fusarium* genomes. The enrichment of these DNA-motifs was tested in a region 600 bp upstream of the transcriptional start site of the genes in each co-expression group. Of the 326 motifs, 113 were enriched in at least one co-expression group (Additional file 1: Table S6). These enriched motifs could act as binding sites for global transcriptional regulators, i.e. transcription factors controlling the expression levels of a significant proportion of co-expressed genes (see Methods). To classify the DNA-binding domains present within these putative transcriptional regulators, *S. cerevisiae* motifs were searched and significant matches determined. *F. graminearum* proteins with significant homology to the associated budding yeast DNA-binding proteins were considered as global regulators (Additional file 2: Figure S2).

Putative *F. graminearum* DNA-binding proteins were assigned to 31 enriched motifs (Additional file 2: Table S7): two-thirds of these 18 sequences contain either Cys₂His₂ or helix-loop-helix domains (Table 3). This deviation in the distribution of DNA-binding domains (DBDs) amongst the global regulators from the background distribution of DBDs was found to be highly significant (Figure 8A; $p < 0.001$, Fisher's exact test). D-TAP-encoding genes present in co-expression groups and containing motifs potentially bound by global regulators in the upstream promoter region ('second-tier regulators') show different DBD-distributions from these top-tier (global) regulators. These second-tier regulators identified in FG2 and FG6 have similar DBD-distributions to all D-TAPs in the genome. In contrast, those in the FG1 infection experiment predominantly contain Cys₂His₂ and bZIP domains.

Son *et al.* [34] performed a systematic analysis of seven phenotypes of mutants in *F. graminearum* transcription factors. The DBDs present in these transcription factors was obtained from the HMM searches of the *F. graminearum* genome (see Methods). Comparison of the DBD-distributions with the background revealed enrichment of Cys₂His₂ and a depletion of Zn(II)₂Cys₆

Table 3 *F. graminearum* global regulators determined by homology to *S. cerevisiae* DNA-binding motifs

<i>F. gr.</i> gene ID	DBD	# Enriched co-expression groups	Enriched co-expression groups
FGSG_04220	APSES	1	FG2.11
FGSG_10384	APSES	1	FG2.11
FGSG_08634	GATA	1	FG2.01
FGSG_00750	HLH	2	FG1↑, FG2.11
FGSG_01307	HLH	2	FG1↑, FG6↓↑
FGSG_02814	HLH	4	FG1↑, FG2.11, FG6↓, FG6↓↑
FGSG_05567	HLH	4	FG1↑, FG2.11, FG6↓, FG6↓↑
FGSG_09043	Homeobox/zf-C2H2	1	FG2.11
FGSG_06359	HSF_DNA-bind	1	FG6↓
FGSG_13911	Myb_DNA-binding	1	FG2.01
FGSG_01298	zf-C2H2	1	FG6↓
FGSG_01341	zf-C2H2	2	FG1↑, FG2.10
FGSG_01350	zf-C2H2	2	FG1↑, FG2.10
FGSG_02743	zf-C2H2	2	FG1↑, FG2.10
FGSG_06311	zf-C2H2	1	FG6↓
FGSG_06871	zf-C2H2	4	FG1↑, FG2.-10, FG2.-11, FG6↓
FGSG_12970	zf-C2H2	1	FG6↓
FGSG_08010	Zn_clus	2	FG1↑, FG2.10

The gene identifier of the putative regulator (*F. gr.* gene ID) along with its DBD-composition ('DBD'), and the number and details of co-expression groups enriched for an associated DNA-motif are shown.

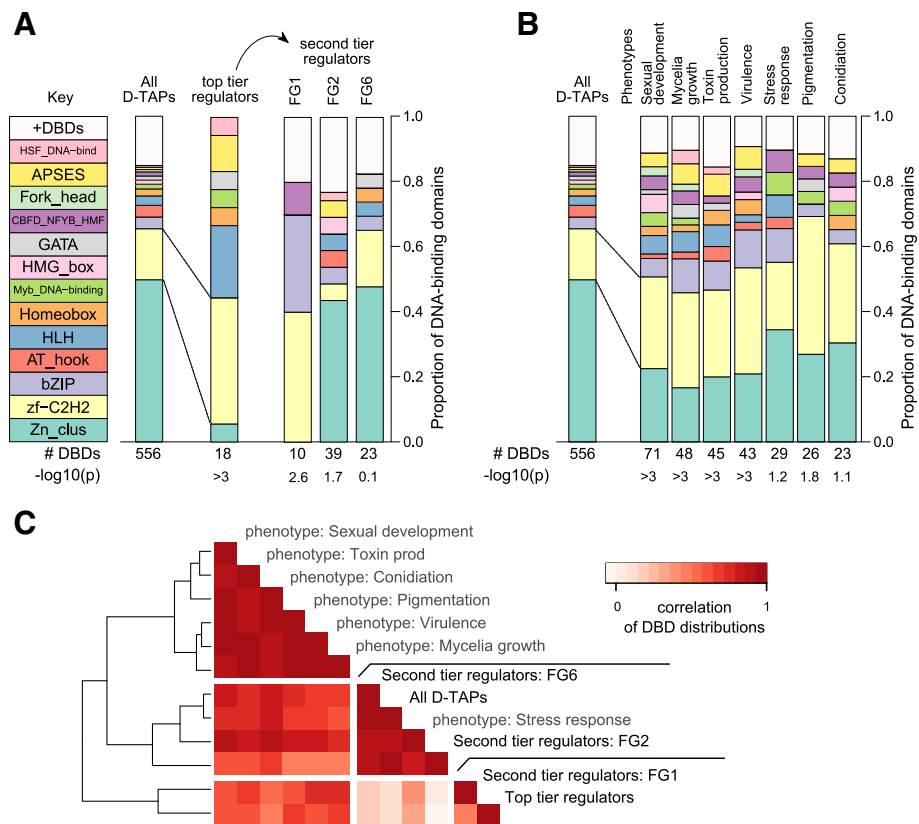


Figure 8 Distribution of DNA-binding domains among *F. graminearum* transcriptional regulators. **(A)** Barplots showing the distribution of DNA-binding domains present in D-TAPs ('All D-TAPs'), and top- and second-tier regulators. #DBDs: number of DNA-binding domains detected; p-value (displayed as $-\log_{10}(p)$) obtained from Fisher's exact test when comparing the top and second-tier distributions with the background. **(B)** Distributions of the pTF DBDs. **(C)** Heatmap visualizing the Pearson correlation coefficients of the DBD abundances – the pairwise correlation between the fraction of background, top and second-tier regulator and pTF DBDs.

domains (Figure 8B): 40% of the top-tier regulators contain Cys_2His_2 domains, with only 6% containing $Zn(II)_2Cys_6$ domains – an order of magnitude reduction compared with the background. This pattern of increased abundance in Cys_2His_2 -containing proteins is also observed with the phenotype-associated transcription factors (pTFs), with an average of 30% containing this DBD contrasting with a background level of 16%. Pairwise similarities in the DBDs present in global, second-tier and pTFs were visualized by hierarchical clustering of these correlations (Figure 8C). These analyses showed that FG2 and FG6 second-tier regulators, along with those regulating stress responses, are highly similar to the background distribution. The FG1 second-tier regulators and the global regulators appear to cluster together due to the high and low levels of TAPs containing Cys_2His_2 and $Zn(II)_2Cys_6$ domains, respectively. The DBD patterns amongst the pTFs exhibit a less pronounced deviation from the background but seem to form a distinct cluster of six phenotype groups with a more diverse range of DBDs.

Discussion

Gene function and phylogenetic conservation were found to be related constraints on gene positioning at the whole genome level in *Fusarium*. Rates of recombination were associated with levels of protein sequence conservation: conserved TAP categories - basal transcription factors and cofactors, RNA polymerase subunits, and chromatin remodelling/histone modification factors - are predominantly found in regions of very low or no recombination, possibly reflecting their fundamental role in the transcriptional process. However, highly diverged DNA-binding proteins (potentially with regulatory roles) are more often present in regions of high recombination. This organisation of transcription factors may increase the rate of adaptive evolution in *Fusarium* by more readily allowing the formation of transcriptional networks with superior adaptation to the habitat of the fungus.

Genome-wide searches for positionally-clustered genes revealed several categories: compact groups, with some containing putative transcriptional regulators, and more

dispersed groups with co-expressed members lying amongst non-coexpressed genes. This pattern of clustering is consistent with that described in other eukaryotic genomes and indicates a diversity of mechanisms for co-regulation [5]. The aurofusarin gene cluster was detected (TC1, Figure 6) around the TAP *AurRI/GIP2* which encodes a regulator required for cluster transcription [25,35,36], and the butenolide gene cluster containing a cytochrome P450 (TC2) around a gene encoding a Zn(II)₂Cys₆ zinc-finger protein thought to regulate this cluster [37]. Deoxynivalenol (DON) mycotoxin production only occurs during the infection of barley ears, and six of the fourteen *TRI* genes are upregulated in FG1: four lie within the *TRI8-TRI14* gene cluster [38] and two are disparate *TRI* genes [39], precluding their identification as a single gene cluster. The remaining members of this biosynthetic pathway do not exhibit significant differences in transcript levels, possibly reflecting the role of post-transcriptional mechanisms in controlling DON synthesis. Furthermore, Reyes-Dominguez *et al.* [40] show that chromatin modification plays a significant role in the regulation of the tightly-linked genes involved with mycelium pigment (aurofusarin) and DON biosynthesis. Together, these observations imply that *Fusarium* gene clusters are subject to multiple levels of co-ordinated regulation.

A hallmark of secondary metabolism genes – in contrast to genes involved in primary metabolism – is that they are clustered in fungal genomes [41]. Systematic annotation of genes within the TAP-centred and localized clusters (Figure 5D) shows a number with homology to polyketide synthases, non-ribosomal peptide synthetases and other types of enzymes that synthesize mycotoxins. Present in the TAP-centred cluster containing the polyketide synthase gene *PKS9* (TC3) is a sequence encoding a Zn(II)₂Cys₆ zinc-finger protein; this local transcription factor controls the production of novel fusarielins [42]. TC6 is a putative, novel co-regulated gene cluster, and its components could comprise a biosynthetic cluster: a transport protein (allantoate permease), a Zn(II)₂Cys₆ DNA-binding factor (transcriptional regulator), and putative modifying enzymes (a FAD-binding protein and a deacetylase). The majority of localised clusters have members with homologues either in *Fusarium* or other filamentous fungi only, and is suggestive of biosynthetic pathways producing *Fusarium*-specific mycotoxins. The close proximity of the genes encoding both enzymatic and regulatory functions, and comprising these positional clusters, may provide an evolutionary mechanism that facilitates adaptation to a wide variety of environments.

The majority of differentially expressed TAP genes encode DNA-binding proteins (D-TAPs), consistent with a role of such sequences in controlling developmental programs and responses to environmental fluctuations. D-TAP

differential expression was found to be predominantly condition-specific (Figure 3), suggesting that different sets of transcription factors orchestrate various regulatory events. Interestingly, on average within each transcriptomics experiment, an order of magnitude more non-TAPs than TAPs exhibited differential expression (Figure 3B), and consistent with the frequency of motifs identified in the promoter regions of functionally related genes [33]. This suggests that individual TAPs may control the expression of multiple genes.

A comparison with two yeasts showed conservation of transcription factors, their binding sites and the target genes regulated by these factors with *Fusarium* pathways known to respond to stress conditions or phosphate metabolism [33]. These observations were extended to identify the types of DNA-binding domains (DBDs) present in the putative transcriptional regulators defined from the co-expression groups. Most global regulators contain either Cys₂His₂ or HLH domains and may control expression across a number of conditions. Additionally, a Cys₂His₂ zinc-finger protein encoded within the trichothecene gene cluster (*TRI6*) has been shown to act as a global transcriptional regulator [43]. This increase of Cys₂His₂ and depletion of Zn(II)₂Cys₆ domains is also seen with transcription factors which when individually deleted produce mutant strains with a variety of phenotypes; however, the classes of DBDs present in these proteins are more complex, possibly reflecting greater diversity in the biological processes studied.

Two distinct patterns of DBDs are observed within the second-tier regulators. The barley ear infection (FG1) secondary regulators are enriched for transcription factors containing a bZIP domain, and their classes are similar in distribution to those of the top-tier regulators. This may indicate more elaborate transcriptional networks are employed during host infection as the bZIP containing transcription factor *ZEB2* can act as a local regulator: the zearalenone biosynthesis gene cluster consists of four members, three of which are regulated transcriptionally by the fourth - *ZEB2* [13]. The secondary regulators identified in nutrient-deprived conditions (FG2) and differentiation from mycelia to perithecia (FG6) contain Zn(II)₂Cys₆ domains predominately. Their DBD class distributions are highly correlated with those of the transcription factors linked to the stress response and the background distribution of D-TAPs; this suggests they may regulate directly the transcription of genes participating in the response to various stimuli and sexual development.

McCord and Bulyk [44] observed in yeast that bZIP, Cys₂His₂ and HLH-containing global regulators are enriched in regulatory hubs, contrasting with local Zn(II)₂Cys₆-containing transcription factors that

are depleted, and implying this global/local nature of a regulator is a feature of its structural class. Hence, top-tier regulators could contain DBDs (e.g. Cys₂His₂ or HLH structures) able to bind more degenerate DNA sequences and so control the transcription of many genes, whereas the Zn(II)₂Cys₆ domain may only recognize highly-specific DNA binding sites ensuring restricted regulation of a gene cluster. This use of different classes of DNA-binding proteins at certain levels within a transcriptional network, could thus allow the evolutionary diversification of mycotoxin production through the gain or loss of sequences from biosynthesis gene clusters [17,18]. These resultant phylogenetic distributions may provide further insights into the role, organization and regulation of gene clusters in *Fusarium* and other emerging fungal threats [45].

Conclusions

Proteins associated with the basal functions of transcription e.g. RNA synthesis, are encoded by genes lying in areas of the *Fusarium* genome with little or no recombination, contrasting with those performing roles in controlling gene activation. Systematic searches for gene clusters revealed compact groups usually containing DNA-binding proteins and more dispersed types; however, both seem to contain an abundance of genes whose products could partake in pathways synthesizing secondary metabolites, suggesting that this gene proximity is important to mycotoxin production.

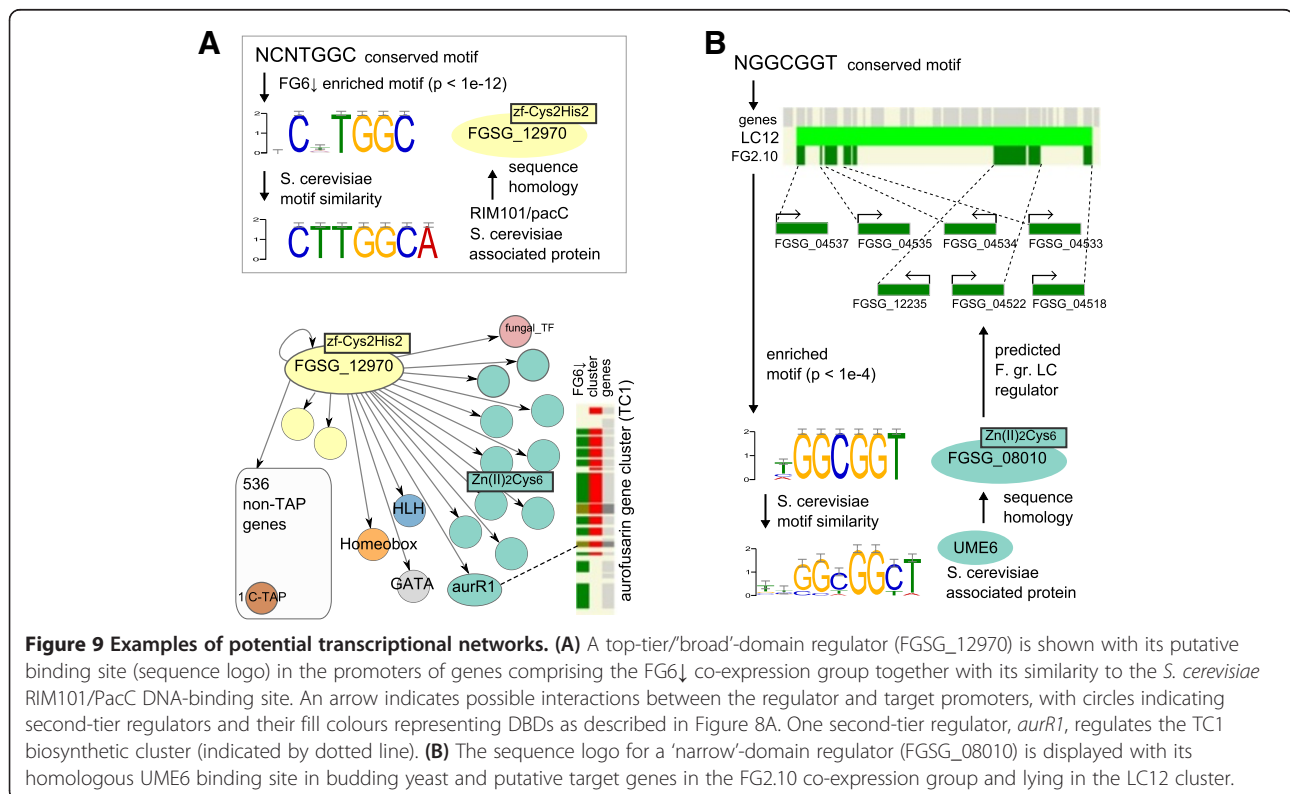
Garber et al. [46] propose that in animals, transcription factors exhibit a multilayered architecture: Pioneer factors initiate remodelling of the epigenome, allowing broad binders to prime lineage-specific genes, with dynamic factors facilitating the activation of environment-specific genes. This layering - though in a much less complex and more compressed manner - is observed with the transcription networks/co-expression groups studied in this analysis; global regulators (mostly containing Cys₂His₂ and HLH DBDs) could play an analogous role to Pioneer and broad binders (Figure 9A), with factors predominately containing the Zn(II)₂Cys₆ DBD (only found in Fungi) activating small subsets of genes functioning in metabolism and development (Figure 9B).

Methods

Identification of *Fusarium graminearum* transcription-associated proteins

Genes encoding proteins associated with all aspects of the transcriptional process in *F. graminearum* were identified by querying the protein sequence entries of its genome [17] with two different types of data set [47,48]:

- (i) A reference set of transcription-associated proteins (TAPs) was assembled from the UniProt database [49] by extracting entries whose GO terms are linked to transcription. The reference set sequences



were filtered for compositional bias using CAST [50], and then used to search the *F. graminearum* genome with BLASTp [51]; any sequence similarities with an E -value $\leq 10^{-6}$ were considered as significant. To identify *F. graminearum* TAPs, the *Fusarium* homologues and their matching reference-set TAPs were clustered using TRIBE-MCL [52] at an inflation value of 2.0, as described previously [53] - this parameter value was chosen to minimize cluster granularity and ensure maximum coverage of the corresponding protein families. Any *F. graminearum* sequences present in the detected protein families (that also contain reference TAPs), were placed into one of five TAP categories based on the functional annotation of the reference sequences.

(ii) Profile hidden-Markov models (HMMs) of domains present in the proteins that constitute the TRANSFAC database of well-characterised eukaryotic transcriptional factors [54], in addition to those of DNA-binding domains present in all three domains of life [55], were used to search the *F. graminearum* genome (with HMMER [56]); any sequences that matched a HMM with a score greater than the lowest score for sequences included in the Pfam [57] full alignment of the family, were considered as hits. Based on the Pfam database records of the HMMs, these hits (TAPs) were placed into one of the five categories defined by the TAP reference set.

These complimentary sequence-comparison approaches identified a total of 723 TAPs in the *F. graminearum* genome (Additional file 1: Table S1).

Detection of *F. graminearum* TAP orthologues

The 723 *F. graminearum* TAPs were used to query 56 eukaryotic genomes (BLASTp). Additionally, all the HMMs comprising the Pfam database were matched against the TAPs and these genomes. A sequence was considered to be orthologous to a *Fusarium* TAP if :-

- (i) its protein-domain structure (defined by multiply-matching HMMs) is conserved entirely with a TAP, and $\geq 60\%$ of this target sequence aligns with $\geq 60\%$ of the query TAP,
- (ii) $\geq 70\%$ of it aligns to $\geq 70\%$ of a TAP, both proteins match the same (single) HMM and neither match HMMs not present in both,
- (iii) $\geq 80\%$ of both TAP and query align and neither match an HMM.

Microarray data analysis

Raw expression data (CEL files) from selected Affymetrix *Fusarium*520094 GeneChip studies (FG1, FG2 [58], FG5

[59], FG6 [60]) were retrieved from PLEXdb [61]. These experiments were selected to represent a large part of the *Fusarium* life cycle. Each experiment was normalised independently to preclude batch effects which would obscure gene expression patterns [62]. Data sets FG2, FG5 and FG6 were quantile normalized using RMA (affy [63,64], R/Bioconductor [65]). To correct for increasing *Fusarium* hyphal biomass during the course of FG1 [58], a variance-stabilising model [66] was fitted to standardise the mean expression levels of RNA polymerase subunits. This procedure allows for both increases and decreases in gene expression to be detected using linear models of differential expression. Probeset detection calls were obtained using mas5calls [64,67]. Present, marginal and absent calls on replicate arrays were scored 1, 0.5, 0, respectively and called as detected if mean score across replicates > 0.6 [59]. Differential expression of probesets was determined using the limma package [68,69] with contrasts comparing each condition to the first time point or to complete media, using the minimum control probe p -value as the differential expression threshold [69].

Genomic clustering of co-expressed genes

The *F. graminearum* coding sequences were mapped to chromosome position using BLAT and displayed using FgrMap (OmniMapFree [20]). Chromosomal clustering methods for co-expressed genes were based on gene order with a background gene list comprising the 13,773 genes represented on the Affymetrix microarray. Within each co-expression group (Figure 4) the distance (number of genes) from each member to the nearest member was stored; this generated a vector of pair-wise distances whose values were then compared with g_{prox} , an integer parameter that was varied from 1 to 200 (Figure 5A). With each iteration of g_{prox} , the number of elements in the vector whose value is less than or equal to g_{prox} was calculated (N_{gprox}). The significance was evaluated using a p -value obtained from 1000 randomly sampled gene lists of the same size as the co-expression group. A p -value of less than 0.05 indicates that the observed N_{gprox} or greater was seen in fewer than 5% of randomly drawn gene lists, and a multiple testing correction was applied within each co-expression group [70]. For each co-expression group exhibiting a corrected significant value of N_{gprox} ($p < 0.05$) a Z-score was used to estimate conservatively how many of the observed proximal genes may be sufficient to explain the elevated value of N_{gprox} . A threshold value of three standard deviations (σ) from the mean (μ) was obtained for the N_{gprox} null distribution (with μ and σ obtained from the randomly sampled gene lists). The difference $N_{gprox} - (\mu + 3\sigma)$ provides an indication of the excess of proximal genes observed in the co-expression group compared with the resampled gene lists (Additional file 2: Table S3).

TAP-centred (TC) clusters of co-expressed genes

For each differentially expressed TAP (Figure 4) the presence of neighbouring genes in the same co-expression group as the TAP was investigated using a variable window size of 2 to 40 genes, i.e. 1 to 20 genes on each side of the TAP (Figure 5B). For a given window size, Fisher's exact test was used to determine the significance of enrichment for co-expression group members within the window. Within each co-expression group, p -values were corrected for multiple testing [71] and windows with $p < 0.05$ were considered significantly enriched. Where nested windows were significantly enriched, the largest such window size was identified for each TAP (Figures 5D and 6).

Localized clusters (LC) of co-expressed genes

The Positional Gene Enrichment (PGE) tool was used to detect regions of the genome enriched for co-expression groups, with the emphasis on identifying localized clusters [30]. PGE was used to estimate a null distribution for each of the 22 co-expression groups by generating 10,000 random gene lists of equal size to the group: for each random list the most significant region and its associated enrichment p -value ($\min-p_i$) was returned. A p -value threshold was determined for each of the 22 $\min-p_i$ distributions at the 5%-ile of the 10,000 $\min-p_i$ values. PGE was then run on each co-expression group gene list, and regions were reported as significantly enriched localized clusters (LCs) only if the associated enrichment p -value was smaller than the corresponding threshold; nested or overlapping regions were merged (Figures 5D and 7).

Functional annotation of genes present in TAP-centred and localized clusters

Putative functions for the proteins encoded by the 170 non-TAP genes present in the TCs and LCs were obtained by querying the 56 eukaryotic genomes, and clustering these sequences and their homologues as described above. The matches of the Pfam HMMs to the eukaryotic genomes enabled GO terms to be assigned to the TC and LC genes. Furthermore, querying the UniProtKB [49] with the gene identifiers of eukaryotic homologues present in the detected protein families allowed their functional annotations to be transferred to the *Fusarium* sequences. The clade specificity of each protein family was assigned as 'Fusarium', 'Peizomycotina', 'Fungi', 'non-metazoan Eukaryotes' or 'Eukaryotes' using the NCBI taxonomic descriptions of these UniProtKB entries [72]. The predicted functional annotation for the TC/LC genes is provided in Additional file 1: Table S5.

Enrichment of conserved upstream motifs and similarity to yeast motifs

Kumar *et al.* [33] reported 326 motifs located in upstream promoter regions of *F. graminearum* genes and conserved in *Fusarium* genomes, and summarized by motif similarity. Motif occurrence in upstream promoter regions (at least one forward or reverse-pair motif) was tested for enrichment amongst genes in each co-expression group, TC and LC. Following Kumar *et al.* [33], the upstream promoter region was defined as up to 600 bp upstream of each gene but excluding any overlapping upstream gene. 12,257 of the gene identifiers represented on the microarray were mapped uniquely to upstream promoter sequences (FG3 assembly [73]) and defined the universe of upstream regions for motif enrichments. The threshold for significant enrichment in co-expression groups was p -value $< 10^{-5}$ (Fisher's exact, one-tailed test) with an estimated false discovery rate of ~9% based on enrichment of permuted motifs (Additional file 2: Figure S3A). Upstream regions of genes in each TC and LC were tested for motif enrichment with significance threshold p -value $< 10^{-3}$ (Additional file 2: Figure S3B). Tomtom (v4.8.1) [74] was used to detect similarity to yeast motifs (databases MacIsaac_v1 [75] and SCPD [76]; $E < 1$ and motif length ≤ 9) and to identify *S. cerevisiae* motif-associated proteins (ScAPs). *F. graminearum* homologues of ScAPs were identified (BLASTp; $E < 10^{-6}$ and matched regions covering $\geq 30\%$ of the query or target sequence).

Additional files

Additional file 1: Contains Tables S1, S2, S5 and S6.

Additional file 2: Contains Tables S3, S4 and S7, Supplementary Table legends, and Figures S1, S2 and S3.

Abbreviations

DBD: DNA-binding domain; DON: Deoxynivalenol; D-TAP: DNA-binding transcription-associated protein; LC: Localized cluster; TAP: Transcription-associated protein; TC: TAP-centred Cluster; PGE: Positional Gene Enrichment; pTF: Phenotype-associated transcription factor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KL performed the transcriptomics, gene cluster and statistical analyses, contributed to the study design and helped to draft the manuscript. KH-K contributed to the design of the gene cluster analysis. AB participated in the study design and helped to draft the manuscript. RMRC performed the TAP identification and annotation, designed the study and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Roland Barriot (Katholieke Universiteit Leuven; currently Université Paul Sabatier, Toulouse) for providing source code for PGE minPi correction, Ulrich Güldener (MIPS, Helmholtz Zentrum München) for assistance with accessing GeneChip annotations, Ekaterina Pilicheva (EMBL-EBI) for help with retrieval of UniProt annotations, Martin Urban (Rothamsted Research (RRes)) for discussion on the selection of

microarray experiments, and John Antoniw (RRes) for use of OmniMapFree pre-publication. K.H.-K. would like to thank Sue Welham (RRes) for statistical advice. K.L. and R.M.R.C. would like to thank Mark Field (University of Cambridge) for critical reading of the manuscript. K.L. would like to thank EMBL for supporting this work; R.M.R.C. and A.B. would like to thank the Medical Research Council for support through a Special Training Fellowship in Bioinformatics to R.M.R.C. and the BioSapiens Network of Excellence funded by the European Commission FP6 Programme under the thematic area 'Life sciences, genomics and biotechnology for health', contract number LSHG-CT-2003-503265. K.H.-K. was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom through the Institute Strategic Programme 20:20 Wheat.

Author details

¹European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. ²Institute for Mathematical and Molecular Biomedicine, King's College London, Hodgkin Building, London SE1 1UL, UK. ³Department of Plant Biology and Crop Science, Rothamsted Research, Harpenden, Herts AL5 2JQ, UK. ⁴Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK.

Received: 30 November 2012 Accepted: 18 June 2013

Published: 27 June 2013

References

- Lawrence J: Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell* 2002, **110**:407–413.
- Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK: A global analysis of *Caenorhabditis elegans* operons. *Nature* 2002, **417**:851–854.
- Donelson JE, Gardner MJ, El-Sayed NM: More surprises from *Kinetoplastida*. *Proc Natl Acad Sci U S A* 1999, **96**:2579–2581.
- Blumenthal T: Operons in eukaryotes. *Brief Funct Genomic Proteomic* 2004, **3**:199–211.
- Hurst LD, Pal C, Lercher MJ: The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 2004, **5**:299–310.
- Levings PP, Bungert J: The human beta-globin locus control region. *Eur J Biochem* 2002, **269**:1589–1599.
- Chang HY: Anatomic demarcation of cells: genes to patterns. *Science* 2009, **326**:1206–1207.
- Lee JM, Sonnhammer EL: Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* 2003, **13**:875–882.
- Keller NP, Turner G, Bennett JW: Fungal secondary metabolism - from biochemistry to genomics. *Nat Rev Microbiol* 2005, **3**:937–947.
- Janga SC, Collado-Vides J, Babu MM: Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *Proc Natl Acad Sci U S A* 2008, **105**:15761–15766.
- Price MN, Huang KH, Arkin AP, Alm EJ: Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res* 2005, **15**:809–819.
- Sneppen K, Pedersen S, Krishna S, Dodd I, Semsey S: Economy of operon formation: cotranscription minimizes shortfall in protein complexes. *mBio* 2010, **1**:e00177–00110.
- Desjardins AE, Proctor RH: Molecular biology of *Fusarium* mycotoxins. *Int J Food Microbiol* 2007, **119**:47–50.
- Kimura M, Tokai T, O'Donnell K, Ward TJ, Fujimura M, Hamamoto H, Shibata T, Yamaguchi I: The trichothecene biosynthesis gene cluster of *Fusarium graminearum* F15 contains a limited number of essential pathway genes and expressed non-essential genes. *FEBS Lett* 2003, **539**:105–110.
- Keller PJ, Knop M: Evolution of mutational robustness in the yeast genome: a link to essential genes and meiotic recombination hotspots. *PLoS Genet* 2009, **5**:e1000533.
- Gale LR, Bryant JD, Calvo S, Giese H, Katan T, O'Donnell K, Suga H, Taga M, Usgaard TR, Ward TJ, Kistler HC: Chromosomal complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. *Genetics* 2005, **171**:985–1001.
- Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, et al: The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 2007, **317**:1400–1402.
- Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al: Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 2010, **464**:367–373.
- Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM: High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 2008, **454**:479–485.
- Antoniw J, Beacham AM, Baldwin TK, Urban M, Rudd JJ, Hammond-Kosack KE: OmniMapFree: a unified tool to visualise and explore sequenced genomes. *BMC Bioinforma* 2011, **12**:447.
- Oliveros JC: VENNY. An interactive tool for comparing lists with Venn Diagrams; 2007. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
- Fernandes M, Keller NP, Adams TH: Sequence-specific binding by *Aspergillus nidulans* AfIR, a C6 zinc cluster protein regulating mycotoxin biosynthesis. *Mol Microbiol* 1998, **28**:1355–1365.
- Proctor RH, Hohn TM, McCormick SP, Desjardins AE: TRI6 encodes an unusual zinc finger protein involved in regulation of trichothecene biosynthesis in *Fusarium sporotrichioides*. *Appl Environ Microbiol* 1995, **61**:1923–1930.
- Palmer JM, Keller NP: Secondary metabolism in fungi: does chromosomal location matter? *Curr Opin Microbiol* 2010, **13**:431–436.
- Gaffoor I, Brown DW, Plattner R, Proctor RH, Qi WH, Trail F: Functional analysis of the polyketide synthase genes in the filamentous fungus *Gibberella zeae* (anamorph *Fusarium graminearum*). *Eukaryot Cell* 2005, **4**:1926–1933.
- Proctor RH, Butchko RA, Brown DW, Moretti A: Functional characterization, sequence comparisons and distribution of a polyketide synthase gene required for perithecial pigmentation in some *Fusarium* species. *Food Addit Contam* 2007, **24**:1076–1087.
- Martin SH, Wingfield BD, Wingfield MJ, Steenkamp ET: Structure and evolution of the *Fusarium* mating type locus: new insights from the *Gibberella fujikuroi* complex. *Fungal Genet Biol* 2011, **48**:731–740.
- Yun SH, Arie T, Kaneko I, Yoder OC, Turgeon BG: Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella/Fusarium* species. *Fungal Genet Biol* 2000, **31**:7–20.
- Charron MJ, Dubin RA, Michels CA: Structural and functional analysis of the MAL1 locus of *Saccharomyces cerevisiae*. *Mol Cell Biol* 1986, **6**:3891–3899.
- De Preter K, Barriort R, Speleman F, Vandessompele J, Moreau Y: Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. *Nucleic Acids Res* 2008, **36**:e43.
- Varga J, Kocsube S, Toth B, Mesterhazy A: Nonribosomal peptide synthetase genes in the genome of *Fusarium graminearum*, causative agent of wheat head blight. *Acta Biol Hung* 2005, **56**:375–388.
- Tobiasen C, Aahman J, Ravnholt KS, Bjerrum MJ, Grell MN, Giese H: Nonribosomal peptide synthetase (NPS) genes in *Fusarium graminearum*, *F. culmorum* and *F. pseudograminearum* and identification of NPS2 as the producer of ferricrocin. *Curr Genet* 2007, **51**:43–58.
- Kumar L, Breakspear A, Kistler C, Ma LJ, Xie XH: Systematic discovery of regulatory motifs in *Fusarium graminearum* by comparing four *Fusarium* genomes. *BMC Genomics* 2010, **11**:208.
- Son H, Seo YS, Min K, Park AR, Lee J, Jin JM, Lin Y, Cao P, Hong SY, Kim EK, et al: A phenotype-based functional analysis of transcription factors in the cereal head blight fungus *Fusarium graminearum*. *PLoS Pathog* 2011, **7**:e1002310.
- Kim JE, Jin J, Kim H, Kim JC, Yun SH, Lee YW: GIP2, a putative transcription factor that regulates the aurofusarin biosynthetic gene cluster in *Gibberella zeae*. *Appl Environ Microbiol* 2006, **72**:1645–1652.
- Frandsen RJ, Nielsen NJ, Maolanon N, Sorensen JC, Olsson S, Nielsen J, Giese H: The biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the naphthoquinones and naphthopyrones. *Mol Microbiol* 2006, **61**:1069–1080.
- Harris LJ, Alexander NJ, Saparno A, Blackwell B, McCormick SP, Desjardins AE, Robert LS, Tinker N, Hattori J, Piche C, et al: A novel gene cluster in *Fusarium graminearum* contains a gene that contributes to butenolide synthesis. *Fungal Genet Biol* 2007, **44**:293–306.
- Brown DW, Dyer RB, McCormick SP, Kendra DF, Plattner RD: Functional demarcation of the *Fusarium* core trichothecene gene cluster. *Fungal Genet Biol* 2004, **41**:454–462.
- Proctor RH, McCormick SP, Alexander NJ, Desjardins AE: Evidence that a secondary metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the filamentous fungus *Fusarium*. *Mol Microbiol* 2009, **74**:1128–1142.
- Reyes-Dominguez Y, Boedi S, Sulyok M, Wiesenberger G, Stoppacher N, Krska R, Strauss J: Heterochromatin influences the secondary metabolite

- profile in the plant pathogen *Fusarium graminearum*. *Fungal Genet Biol* 2012, **49**:39–47.
41. Yu JH, Keller N: Regulation of secondary metabolism in filamentous fungi. *Annu Rev Phytopathol* 2005, **43**:437–458.
 42. Sorensen JL, Hansen FT, Sondergaard TE, Staerk D, Lee TV, Wimmer R, Klitgaard LG, Purup S, Giese H, Frandsen RJ: Production of novel fusarielins by ectopic activation of the polyketide synthase 9 cluster in *Fusarium graminearum*. *Environ Microbiol* 2012, **14**:1159–70.
 43. Nasmith CG, Walkowiak S, Wang L, Leung WY, Gong Y, Johnston A, Harris LJ, Guttman DS, Subramaniam R: *TRI6* is a global transcription regulator in the phytopathogen *Fusarium graminearum*. *PLoS Pathog* 2011, **7**:e1002266.
 44. McCord RP, Bulyk ML: Functional trends in structural classes of the DNA binding domains of regulatory transcription factors. *Pac Symp Biocomput* 2008, **13**:441–452.
 45. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ: Emerging fungal threats to animal, plant and ecosystem health. *Nature* 2012, **484**:186–194.
 46. Garber M, Yosef N, Goren A, Raychowdhury R, Thielke A, Guttman M, Robinson J, Minie B, Chevrier N, Itzhaki Z, et al: A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol Cell* 2012, **47**:810–822.
 47. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, et al: The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005, **309**:436–442.
 48. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, et al: Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 2008, **455**:757–763.
 49. Bairoch A, Bougueleret L, Altairac S, Amendolia V, Auchincloss A, Puy GA, Axelsen K, Baratin D, Blatter MC, Boeckmann B, et al: The universal protein resource (UniProt). *Nucleic Acids Res* 2008, **36**:D190–195.
 50. Promponas VJ, Enright AJ, Tsoka S, Kreil DP, Leroy C, Hamodrakas S, Sander C, Ouzounis CA: CAST: an iterative algorithm for the complexity analysis of sequence tracts. *Bioinformatics* 2000, **16**:915–922.
 51. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, **25**:3389–3402.
 52. Enright AJ, Van Dongen S, Ouzounis CA: An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 2002, **30**:1575–1584.
 53. Coulson RMR, Ouzounis CA: The phylogenetic diversity of eukaryotic transcription. *Nucleic Acids Res* 2003, **31**:653–660.
 54. Wingender E, Dietze P, Karas H, Knuppel R: TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 1996, **24**:238–241.
 55. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA: DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res* 2008, **36**:D88–92.
 56. Eddy SR: Profile hidden Markov models. *Bioinformatics* 1998, **14**:755–763.
 57. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al: The Pfam protein families database. *Nucleic Acids Res* 2004, **32**:D138–141.
 58. Guldener U, Seong KY, Boddu J, Cho S, Trail F, Xu JR, Adam G, Mewes HW, Muehlbauer GJ, Kistler HC: Development of a *Fusarium graminearum* Affymetrix GeneChip for profiling fungal gene expression *in vitro* and *in planta*. *Fungal Genet Biol* 2006, **43**:316–325.
 59. Hallen HE, Huebner M, Shiu SH, Guldener U, Trail F: Gene expression shifts during perithecial development in *Gibberella zeae* (anamorph *Fusarium graminearum*), with particular emphasis on ion transport proteins. *Fungal Genet Biol* 2007, **44**:1146–1156.
 60. Hallen HE, Trail F: The L-type calcium ion channel *CCH1* affects ascospore discharge and mycelial growth in the filamentous fungus *Gibberella zeae* (anamorph *Fusarium graminearum*). *Eukaryot Cell* 2008, **7**:415–424.
 61. Wise RP, Caldo RA, Hong L, Shen L, Cannon E, Dickerson JA: BarleyBase/ PLEXdb. *Methods Mol Biol* 2007, **406**:347–363.
 62. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA: Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010, **11**:733–739.
 63. Bolstad BM, Irizarry RA, Astrand M, Speed TP: A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003, **19**:185–193.
 64. Gautier L, Cope L, Bolstad BM, Irizarry RA: affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004, **20**:307–315.
 65. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al: Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004, **5**:R80.
 66. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M: Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002, **18**(Suppl 1):S96–104.
 67. Liu WM, Mei R, Di X, Ryder TB, Hubbell E, Dee S, Webster TA, Harrington CA, Ho MH, Baid J, Smeekens SP: Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics* 2002, **18**:1593–1599.
 68. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004, **3**:1. Article 3.
 69. Smyth GK: Limma: linear models for microarray data. In *Volume Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Edited by Gentleman VC R, Dudoit S, Irizarry R, Huber W. New York: Springer; 2005:397–420.
 70. Benjamini Y, Yekutieli D: The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 2001, **29**:1165–1188.
 71. Benjamini Y, Hochberg Y: Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995, **57**:289–300.
 72. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetverin V, Church DM, Diucchio M, Federhen S, et al: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2010, **38**:D5–16.
 73. *Fusarium Comparative Sequencing Project, Broad Institute of Harvard and MIT*. <http://www.broadinstitute.org/>.
 74. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS: Quantifying similarity between motifs. *Genome Biol* 2007, **8**:R24.
 75. Maclsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E: An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinforma* 2006, **7**:113.
 76. Zhu J, Zhang MQ: SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics* 1999, **15**:607–611.

doi:10.1186/1752-0509-7-52

Cite this article as: Lawler et al.: Genomic clustering and co-regulation of transcriptional networks in the pathogenic fungus *Fusarium graminearum*. *BMC Systems Biology* 2013 **7**:52.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

