



OPEN

Analysis of the chromosomal clustering of *Fusarium*-responsive wheat genes uncovers new players in the defence against head blight disease

Alexandre Perochon^{1,2}, Harriet R. Benbow^{1,2}, Katarzyna Ślęczka-Brady¹, Keshav B. Malla¹ & Fiona M. Doohan¹✉

There is increasing evidence that some functionally related, co-expressed genes cluster within eukaryotic genomes. We present a novel pipeline that delineates such eukaryotic gene clusters. Using this tool for bread wheat, we uncovered 44 clusters of genes that are responsive to the fungal pathogen *Fusarium graminearum*. As expected, these *Fusarium*-responsive gene clusters (FRGCs) included metabolic gene clusters, many of which are associated with disease resistance, but hitherto not described for wheat. However, the majority of the FRGCs are non-metabolic, many of which contain clusters of paralogues, including those implicated in plant disease responses, such as glutathione transferases, MAP kinases, and germin-like proteins. 20 of the FRGCs encode nonhomologous, non-metabolic genes (including defence-related genes). One of these clusters includes the characterised *Fusarium* resistance orphan gene, *TaFROG*. Eight of the FRGCs map within 6 FHB resistance loci. One small QTL on chromosome 7D (4.7 Mb) encodes eight *Fusarium*-responsive genes, five of which are within a FRGC. This study provides a new tool to identify genomic regions enriched in genes responsive to specific traits of interest and applied herein it highlighted gene families, genetic loci and biological pathways of importance in the response of wheat to disease.

Prokaryote genomes encode co-transcribed genes with related functions that cluster together within operons. Clusters of functionally related genes also exist in eukaryote genomes, including fungi, nematodes, mammals and plants¹. Gene clusters consist of two main classes; paralogous genes that most likely arose by gene duplication, and nonhomologous genes that are physically clustered and co-regulated². The yeast (*Saccharomyces cerevisiae*) genome encodes gene clusters that are important for the use of alternative sources of carbon³ and nitrogen⁴. Filamentous fungi contain metabolic gene clusters, especially those involved in secondary metabolic pathways that generate toxins, antibiotics, and other metabolites⁵. In animals, gene clusters play a role in development and defence, such as the major histocompatibility complex (MHC) that encodes diverse proteins involved in adaptive and innate immunity⁶.

Plant gene clusters include paralogous genes and functionally related nonhomologous genes coding for enzymes involved in secondary metabolic pathways⁷. Plant biosynthetic gene clusters produce a variety of secondary metabolites, such as benzoxazinoids, cyanogenic glucoside, polyketides, terpenes and benzylisoquinoline alkaloids⁸. The clustering of genes involved in plant metabolic pathways was first demonstrated in maize (*Zea mays*) for the benzoxazinoid (*Bx*) gene cluster that encodes enzymes involved in the biosynthesis of the cyclic hydroxamic acids involved in defence against microbial pathogens and herbivores⁹. Since then, tens of metabolic gene clusters have been characterised in eudicot and monocot plant species⁸. Paralogous gene clusters are shown to contribute to important agronomic traits such as adaptation to abiotic and biotic stresses¹⁰; plant disease resistance genes such as lectin receptor kinases (LecRK), wall-associated receptor kinases (WAK) and nucleotide-binding leucine-rich repeat (NLR) containing proteins are often organised into clusters^{11–13}. The major quantitative trait locus (QTL) *Submergence-1* (*Sub1*) provides an example of a paralogous plant gene cluster involved in abiotic

¹UCD School of Biology and Environmental Science and Earth Institute, College of Science, University College Dublin, Belfield, Dublin 4, Ireland. ²These authors contributed equally: Alexandre Perochon and Harriet R. Benbow. ✉email: fiona.doohan@ucd.ie

stress tolerance: it encodes a cluster of ethylene-responsive transcription factors induced after submergence, of which the *Sub1A* allele confers tolerance to prolonged submergence¹⁴. In tomato, a pathogen-responsive metabolic gene cluster is responsible for faltarindiol biosynthesis, a modified lipid involved in pathogen resistance¹⁵. In rice, the *Bph3* locus corresponds to a cluster of three *LecRKs* induced during brown planthopper infestation, which function together to confer insect resistance¹². In wheat, five paralogue salt-responsive cytochrome P450 genes (CYPs) confer salinity tolerance to wheat¹⁶.

The recently released annotated bread wheat *Triticum aestivum* genome was used as a model. Wheat is a very important crop worldwide with a large and recently annotated genome (RefSeq v1.1), detailing 107,891 high-confidence genes across 21 chromosomes¹⁷. Fusarium head blight (FHB) is a fungal disease of wheat caused by several *Fusarium* spp., most commonly *Fusarium graminearum*. Resistance to FHB is quantitative, and ~500 FHB resistance QTL have been reported in wheat¹⁸. However, the underpinning genes have only been elucidated for two QTL^{19–22}. Additionally, comparative transcriptomics analyses identified other FHB resistance genes but only few have been validated²³. Consequently, more studies are needed to decipher the molecular mechanisms underpinning wheat–*Fusarium* interactions, to optimise disease resistance breeding. Based on evidence that stress-responsive genes can be clustered in plant genomes, we designed a pipeline to determine if diverse functionally related genes cluster in the genome, using FHB of wheat as a model. Using this pipeline, we captured paralogous and nonhomologous clusters encoding metabolic and non-metabolic genes that were responsive to *F. graminearum* infection of wheat. These *Fusarium*-responsive gene clusters (FRGCs) were assessed for their biological significance using gene co-expression analysis and are discussed with respect to their homology with other plants, their links to defence mechanisms, and their proximity to published FHB QTL.

Results and discussion

Identification and characterisation of *Fusarium*-responsive gene clusters across the wheat genome.

We developed a pipeline to determine if wheat genes differentially expressed (DEGs) in response to *F. graminearum* are randomly distributed along the wheat genome, or if they form physical clusters. We used the publicly available transcriptomics dataset from the experiment E-MTAB-4222, wherein heads of wheat near-isogenic lines (NILs) were treated with *F. graminearum* and harvested at 3, 6, 12, 24, 36, and 48 h post inoculation²⁴. Differential expression analysis identified 10,567 high confidence (HC) genes that were differentially expressed in response to *F. graminearum* treatment in at least one of the timepoint x NIL combinations (fold change ≥ 1.5 ; false discovery rate < 0.05). Using these data, a sliding window of 10 genes was used to screen each wheat chromosome and calculate the *Fusarium*-responsive gene (FRG) density and FRG consecutiveness. The sliding window analysis was repeated on 10,000 random permutations of the data to generate a threshold at which density and consecutiveness of DEGs are considered non-random. Using normalised raw expression data for each gene within a cluster, co-expression was measured as the average absolute correlation coefficient between all genes within a cluster and compared to the co-expression of genes outside the cluster. A schematic of the pipeline can be seen in Fig. 1.

In this study, significant clusters were identified as those with a FRG density ≥ 0.7 (Fig. 2a) and/or 5 consecutive FRGs (Supplementary Fig. S1), as these values appeared $< 0.01\%$ of the time in the permutation test. We found a total of 44 gene clusters containing 297 FRGs (2.8% of FRGs) located on 16 of the 21 wheat chromosomes (Fig. 2a, Supplementary Fig. S1, Table S1). FRGCs contain between 5 and 11 FRGs, and 5–8 consecutive FRGs, with a physical size between 18 and 1268 kb (Supplementary Table S1), similar to the size of the metabolic gene clusters identified by Schlöpfer et al.²⁵ (33–284 kb in length, ranging from 4 to 18 genes). The average distance between genes within FRGC is 58 kb, smaller than the average distance between genes near the cluster (10 genes upstream and downstream; 103 kb), or elsewhere in the genome (132 kb), indicating that gene density within FRGCs is high.

Subgenome distribution, classification and co-regulation of FRGCs. Wheat is hexaploid (AA, BB and DD subgenomes) and the FRGCs were unevenly distributed between the three subgenomes; 12 (27%) were on the A subgenome, 9 (20%) on the B subgenome, 22 (50%) on the D subgenome, and 1 (2%) on an unassigned sequence scaffold (ChrUn). The D subgenome is reported to have an abundance of genes responsive to *F. graminearum*²⁶, and we found that it encodes more FRGs than either the A or B subgenomes, both in terms of absolute numbers (3,431, 3,308 and 3670, respectively, on the A, B and D subgenomes) and relative to total gene number per subgenome (9.7, 9.3 and 10.7%, respectively, on the A, B and D subgenomes). D genome enrichment of stress-responsive genes was also noted by Zhang et al.²⁷, who identified 143 gene cluster “hotspot regions” in the wheat genome that contain at least 3 consecutive salt-responsive genes. Interestingly, despite the fact that their study used an older genome annotation version and different cut-off criteria for selection of DEGs, 14 of the FRGCs contained the same genes as those identified in the salt-responsive gene clusters (Supplementary Table S1 and Supplementary dataset 3 from Zhang et al.²⁷). Thus, some genes contributing to a shared transcriptional response between salt stress and FHB are located within gene clusters.

We determined the chromosomal zone of each FRGC according to the International Wheat Genome Sequencing Consortium (IWGSC)¹⁷, which separates the chromosomes into three zones: distal telomeric (short arm (R1) and long arm (R3)), interstitial regions (short arm (R2a) and long arm (R2b)), and centromere regions (C). FRGCs were more abundant in telomeric regions compared to interstitial regions (56 and 44%, respectively) and were absent in centromeric region (Fig. 2b). Overall, this distribution correlates with the number of genes present in the different chromosomal zone, with 49,555 genes in the telomeric regions, 48,360 genes in the interstitial regions, and only 7282 genes in the centromeric region (Fig. 2c).

Twenty of the FRGCs have at least one reciprocal homoeologous cluster (Table 1), with 8 pairs and 1 triad. While the homoeologous regions of the remaining 24 FRGCs do contain FRGs, these reciprocal regions did

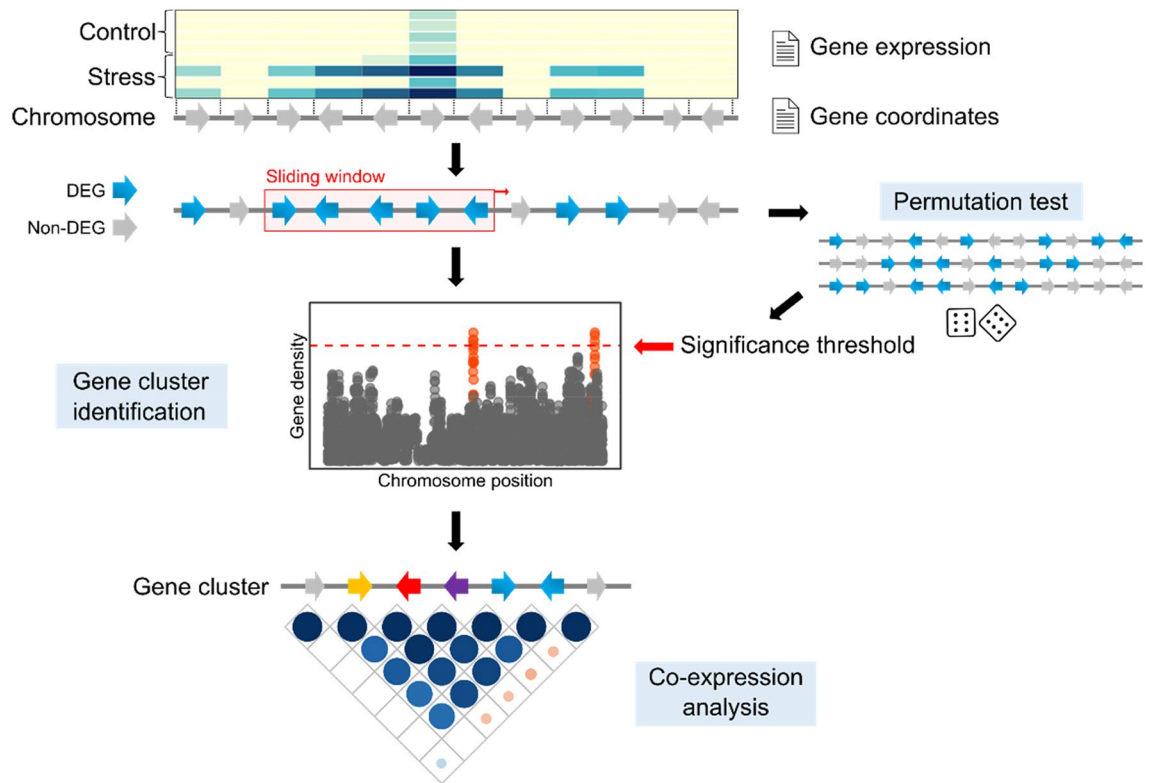


Figure 1. Genome mining of stress-responsive gene clusters. Differential expression analysis was used to identify stress-responsive genes from transcriptomics data. Using chromosomal coordinates of genes from a reference genome, the average number of differentially expressed genes (DEGs) within a sliding window of a user-defined size (hereby called gene density), and the length of a string of consecutive DEGs was calculated. Shuffled data sets were also analysed to determine what values of gene density and consecutiveness can be considered non-random. Clusters of genes that breach this threshold were designated as significantly non-random gene clusters, and their expression profiles are used for co-expression analysis.

not pass the threshold of density and/or consecutiveness and were therefore not considered FRGC (Table 1 and Supplementary Table S2).

In particular FRGCs 11, 16, 36, and 42 contained many subgenome-specific *Fusarium*-responsive genes, and therefore we concluded they were subgenome-specific FRGCs (Table 1 and Supplementary Table S2).

We categorised FRGCs based on the homology and the metabolic function of their encoded FRGs. FRGCs were classified into four classes on the basis that their encoded FRGs were either paralogous or nonhomologous, and either as metabolic or non-metabolic (Table 1). Most of the clusters were nonhomologous (64%, of which 46% were non-metabolic), and paralogous gene clusters represented 36% of all clusters (of which 27% were non-metabolic) (Table 1). Based on the ‘guilt-by-association’ principle, it is suggested that genes involved in the same biological process are generally co-regulated and co-expressed by a similar regulatory mechanism²⁸. Thus, we analysed the correlation between the expression profiles of the FRGs within FRGCs. With the exception of two FRGCs, we found that all clusters had an average Pearson’s correlation coefficient (PCC) ≥ 0.3 , and the PCC was significantly higher between the FRGC than the neighbouring genes ($P \leq 0.001$; Supplementary Fig. S2, Table S3). Half of the clusters had a strong PCC ($r \geq 0.7$), with 8 clusters having a very strong average correlation between the genes within the cluster ($r \geq 0.9$). Eleven clusters have a moderate ($0.5 \geq r < 0.7$) and 7 had a weak ($0.3 \geq r < 0.5$) correlation.

Nonhomologous metabolic gene clusters. PlantiSMASH is a bioinformatic tool for secondary metabolic gene cluster (MGC) prediction²⁹. We used it to test if the metabolic FRGCs identified were predicted as secondary MGC. We identified 239 MGCs within the wheat genome (Supplementary Dataset 1), of which 10 share common genes with 9 FRGCs, including all 8 of the nonhomologous metabolic FRGCs, with one FRGC matching with 2 different MGCs (Supplementary Dataset 2). From the 10 MGCs identified by PlantiSMASH, 6 encompass entire FRGCs, and include additional genes surrounding the FRGC that, in general, are not responsive to *Fusarium*. Some FRGCs shared some homologies with previously characterised plant metabolic gene clusters, containing strongly co-expressed metabolic genes, and present some link with plant defence mechanisms associated to *Fusarium* diseases. These correspond to a benzoxazinoid-like gene cluster (FRGC 11), momilactone-like gene clusters (FRGC 8, 13, 15 and 44) and terpene metabolic gene clusters (FRGC 27 and 34).

FRGC 11 is a subgenome specific cluster located on chromosome 2B that shares all FRGs contained within MGC 7 (Supplementary Dataset 2), of which six are cytochrome P450 (*CYP450*) genes and one is a wall-associated kinase (*WAK*) gene (Fig. 3a). Two of the *CYP450*s belong to the *CYP79A* subfamily and four to the *CYP71C*

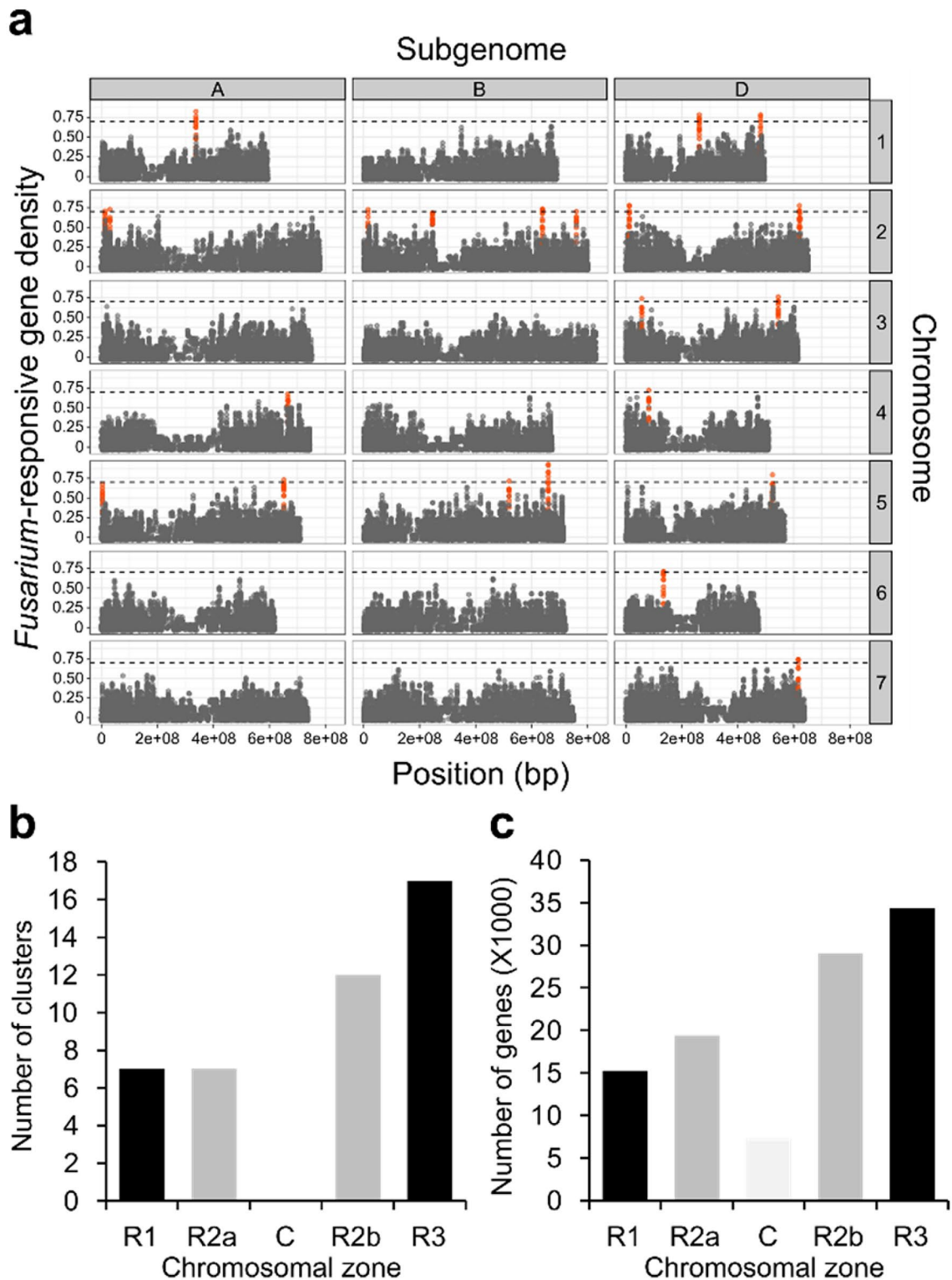


Figure 2. Chromosomal distribution of *Fusarium*-responsive gene clusters (FRGCs). **(a)** *Fusarium*-responsive gene density along each wheat chromosome. A sliding window of 10 genes was used to screen each chromosome and calculate the *Fusarium*-responsive gene density. A significance threshold (dashed line) was calculated by 10,000 random permutations of the data and enriched *Fusarium*-responsive gene loci (those that passed the threshold) were identified and are represented in red vermillion. **(b)** Number of FRGCs and **(c)** total number of high confidence genes are shown in bar diagrams with their distribution in different chromosomal zone such as distal telomeric (short arm (R1) and long arm (R3)), interstitial regions (short arm (R2a) and long arm (R2b)) and centromere regions (C), according to¹⁷.

Gene cluster category	Chromosome	Proteins within the cluster	Cluster ID
Nonhomologous, non-metabolic	1A	Embryogenesis-associated protein EMB8, subtilisin-like protease, kinase family protein, transmembrane protein, putative DUF1218 proteins	2
	1B, 1D	Receptor-like kinases, glucan endo-1,3-beta-glucosidase, cysteine-rich receptor kinase	3, 6
	2B	Receptor-like kinases, glucan endo-1,3-beta-glucosidase, receptor kinase, cysteine-rich receptor kinase	12
	2D	Nucleotide-binding leucine-rich repeats, receptor kinase, plastid transcriptionally active 5	16
	2D	B3 domain-containing protein, RING/U-box superfamily proteins, 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	17
	2D	Blue copper proteins, FAD-binding Berberine family protein-putative, Ethylene-responsive transcription factor-putative	19
	3D	Lectin receptor kinases, FMN-dependent NADPH-azoreductases, pyruvate decarboxylase, transcription factor, sugar transporter-putative	20
	3D	Glutathione S-transferases, DNase I-like superfamily protein, SAUR-like auxin-responsive family proteins, nucleotide-binding leucine-rich repeat	21
	4A, 7D	Lectin receptor kinases, cytochrome P450, ABC transporter (Leaf rust resistance gene 34 (LR34)), sugar transporter-putative	23, 40
	4A	Carbonic anhydrases, Coatomer beta subunit, putative, Fatty acyl-CoA reductase, Peroxidase	24
	4D	Fusarium resistance orphan protein (TaFROG), RZ53, lectin receptor kinase, alcohol dehydrogenases, SWAP (Suppressor-of-White-APricot)/surp domain-containing protein, carboxypeptidase	26
	5A, 5B	MYB transcription factor, histone mono-ubiquitination 2, receptor kinase, alcohol dehydrogenases, putative, TSL-kinase interacting protein 1	28, 31
	5D	Kinases putative, cysteine-rich receptor-kinase-like protein, wall-associated receptor kinase-like proteins	35
	6D	Receptor-like kinases, TMV response-related proteins	39
	7D	Protein EXORDIUM-like 1, glycosyltransferase, RNA-binding protein	41
7D	J-type co-chaperone jac1, mitochondrial, processing peptidase, TMV response-related proteins	42	
7D	Magnesium transporter NIPA, eukaryotic aspartyl protease proteins, 1-amino-cyclopropane-1-carboxylate oxidase 1, DUF668 family protein, chaperone protein dnaJ5	43	
Nonhomologous, metabolic	2A, 2B, 2D, Un	Cytochrome P450s, kaurene synthases, copalyl diphosphate synthases, UDP-dependent glycosyltransferases, vacuolar sorting receptor family protein, wall-associated receptor kinase-like protein, receptor-like protein kinase	8, 13, 15, 44
	2B	Cytochrome P450s, wall-associated receptor kinase	11
	5A, 5D	Cytochrome P450s, hydroxysteroid dehydrogenases, terpene cyclase/mutases family member, receptor-like protein kinase, always EARLY-like protein, obtusifoliol 14-alpha demethylase	27, 34
	5D	O-methyltransferases, cytochrome P450, chalcone synthases, chalcone-flavonone isomerase family, leucine rich repeat family protein	36
Paralogous, metabolic	1D	ATP-dependent zinc metalloproteases FtSHs	7
	2A	Ribulose biphosphate carboxylase small chains , alcohol dehydrogenase, putative	9
	2A	Phenylalanine ammonia-lyases , glucuronoxylan 4-O-methyltransferase	10
	5B	Agmatine coumaroyltransferases , ornithine decarboxylase	32
Paralogous, non-metabolic	1A, 1D	Glutathione S-transferases , RNA binding protein, putative, phospholipase D, alpha/beta-hydrolase superfamily protein	1, 4
	1D	Mitogen-activated protein kinase kinase kinases	5
	2B, 2D	Heavy metal transport/detoxification superfamily proteins , basic helix-loop-helix (bHLH) DNA-binding superfamily protein, nucleotide-binding leucine-rich repeat	14, 18
	3D	RZ53s	22
	4D	Germin-like protein 1 s	25
	5A	Protein of unknown function-DUF538s , 1-pyrroline-5-carboxylate dehydrogenase	29
	5A, 5B	Zinc finger proteins	30, 33
6A, 6B	MYB transcription factors , hexosyltransferase	37, 38	

Table 1. *Fusarium*-responsive gene cluster categories and contents. Encoded proteins by *Fusarium*-responsive gene within each cluster are listed. Paralogous proteins that represented $\geq 60\%$ of the total protein content of the cluster are highlighted in bold.

subfamily. The *CYP450s* are strongly co-expressed ($r=0.7$), while the *WAK* is not (Fig. 3a). Recently, a different *WAK* on chromosome 2AS was proposed to be the gene underlying *Fusarium* resistance *QFhb.mgb-2A*³⁰. The *CYP79A* subfamily is associated with the production of defence molecules, such as cyanogenic glucosides beta- and gamma-hydroxynitrile glucosides and glucosinolates³¹. The *CYP71C* subfamily proteins are associated with the biosynthesis of the phytoalexin benzoxazinoid (Bx)³², and FRGC 11 *CYP71Cs* are orthologues of a cluster of *CYP71C* genes (36, 56 and 5) on the maize chromosome 2 named “Bx2-like” genes based on their homology to the Bx biosynthetic gene³³. Benzoxazinoid metabolites have been associated with FHB resistance in Danish wheat varieties³⁴ and their degradation by another FHB-causal species, *F. pseudograminearum*, is important for its virulence towards wheat in head blight inoculation assays³⁵. Two other wheat *Bx2-like* genes were previously

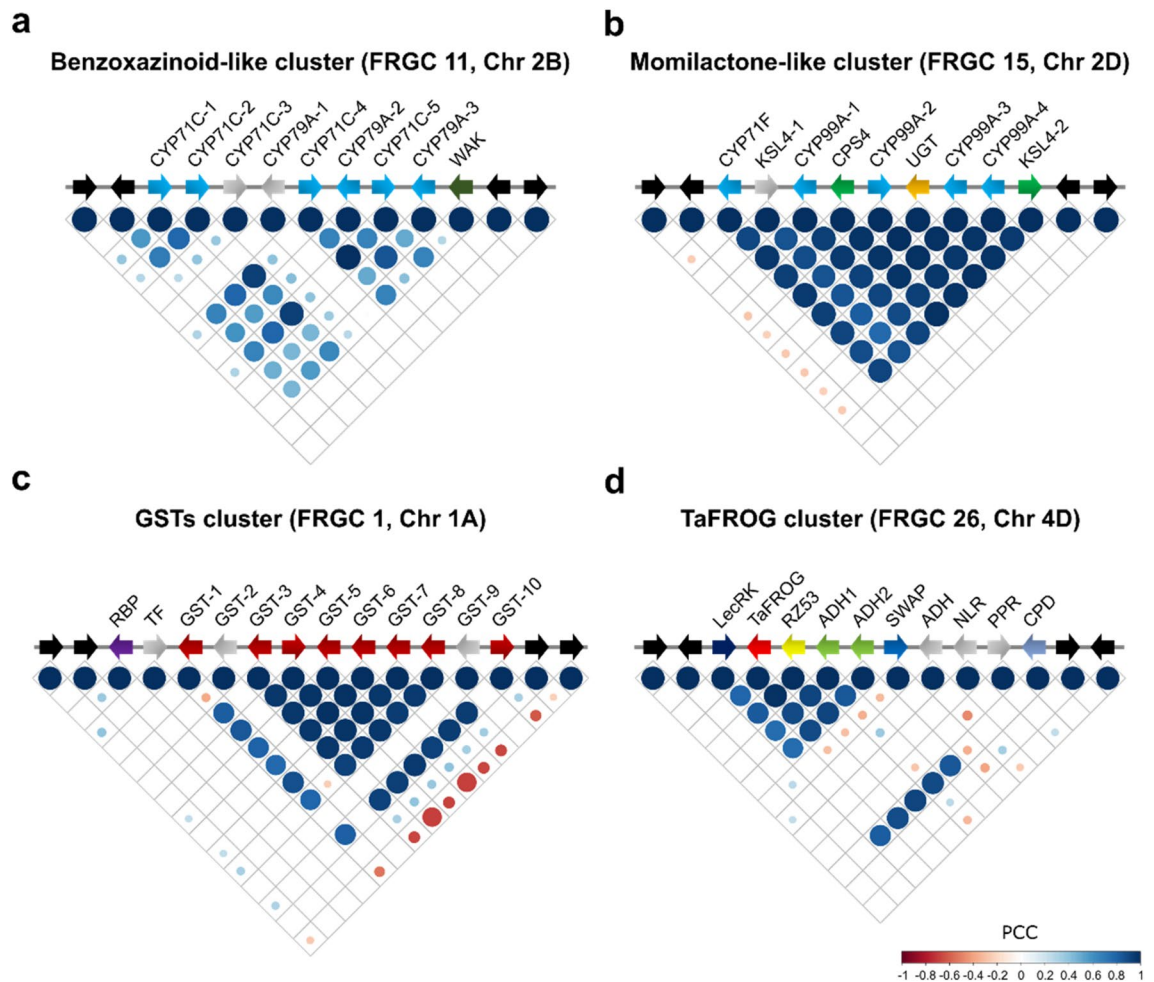


Figure 3. Content, arrangement, and co-expression profiles of representative FRGCs. Schematic representations of *Fusarium*-responsive gene clusters are represented as: *Fusarium*-responsive (coloured arrows), non-responsive genes (grey arrows) and two neighbouring genes (black arrows). Different colours indicate distinct classes of proteins. Co-expression matrices are shown below each cluster. Circle characteristics represent the direction (colour, blue = positive, red = negative) and the strength (size) of the correlation. Four examples of FRGCs are illustrated and correspond to nonhomologous metabolic ((a) benzoxazinoid-like and (b) momilactone-like cluster), paralogous non-metabolic ((c) GSTs cluster), and nonhomologous non-metabolic ((d) TaFROG cluster) categories. FRGC ID and chromosome number are indicated between brackets. (a) The benzoxazinoid-like cluster contains genes coding for a wall-Associated kinase (WAK) and two classes of CYP450s genes with four benzoxazinoid-like genes coding for CYP71C, and two CYP79A genes. (b) The momilactone-like cluster is similar to the characterised rice momilactone gene cluster³⁷, with CYP450s (CYP) monooxygenases (CYP99A2/A3), two terpene synthase kaurene synthase (KS) and copalyl diphosphate synthase (CPS) genes, but without the momilactone A synthase gene absent in the wheat cluster. (c) The GSTs cluster contains an RNA binding protein (RBP) and 8 glutathione-S-transferases (GST) genes. (d) The TaFROG cluster contains genes coding for a lectin receptor kinase (LecRK), the *Fusarium* resistant orphan protein (TaFROG), a poaceae-specific protein (RZ53), a SWAP/surp domain-containing protein (SWAP), two alcohol dehydrogenase (ADH) and a carboxypeptidase (CDP). Cluster representation is not to scale.

reported to be highly induced by *F. pseudograminearum* infection³⁶ but they are not yet functionally characterised and may be involved in alternative secondary metabolic pathways³³.

Four MGCs predicted by PlantiSMASH match and encompass 4 homoeologous FRGCs (FRGC 8, 13, 15 and 44, Table 1 and Supplementary Table S1) found on chromosome 2A, 2B, 2D and Un, respectively. In these FRGCs, we identified wheat genes coding for core enzymes of a rice momilactone gene cluster³⁷, such as the orthologues of the diterpene synthases copalyl diphosphate synthase (CPS) *OsCPS4* and kaurene synthase-like (KSL) *OsKSL4*, and two CYP450s (*CYP99A2*, 3) genes (as illustrated for FRGC 15 in Fig. 3b). The rice momilactone cluster contains an additional dehydrogenase gene (*OsMAS*; *Os Momilactone A Synthase*)³⁷ that is absent in the wheat clusters. Therefore, we named these clusters momilactone-like. The momilactone-like FRGCs contain two additional genes identified as a CYP71F and a UDP-dependent glycosyltransferase (UGT) for FRGC 8 and 15 (chromosome 2A and 2D), and two CYP701A for FRGC 13 (chromosome 2B). All metabolic genes (*CPS*, *KSL*, *CYP99A*, *CYP71F*, *CYP701A*, *UGT*) within the 4 homoeologous FRGCs were induced by *Fusarium*

(Supplementary Table S4) and were strongly co-expressed ($r > 0.85$, Supplementary Fig. S2, Table S3), with the FRGs within FRGC 15 showing the strongest correlation ($r = 0.95$, Fig. 3b). Momilactone and related metabolites belong to the diterpenoid phytoalexin group of secondary metabolites, which are involved in antimicrobial activities, allelopathic activities against competing rice weeds, and insect pest antifeedant activities³⁸. While rice and maize diterpenoid functions and biosynthesis have been widely studied, key diterpene synthases CPS and KSL remain uncharacterised in wheat³⁹. To our knowledge, this is the first time that a diterpenoid gene cluster has been described in wheat. Momilactone and related diterpenoids were shown to have *in vitro* activities against fungal pathogens in maize (*F. graminearum*, *Cochliobolus heterostrophus* and *F. verticillioides*) and in rice (*Xanthomonas oryzae* and *Magnaporthe poae*)^{40,41}. Furthermore, some genetic evidence highlights the role of diterpenoid and diterpene synthases in disease defence. Momilactone gene cluster *OsCPS4* knock-out and knock down rice plants showed enhanced resistance to *X. oryzae* and conversely decreased resistance to the non-host pathogen *M. poae*⁴². Knock-out and knock down of diterpenoid metabolism gene *OsCPS2* increased susceptibility to both *Magnaporthe oryzae* and *X. oryzae*⁴². Maize *ZmKSL2* mutant lines lacking kauralexins were more susceptible to *F. graminearum* stalk rot⁴¹. Hence, diterpenoid biosynthetic pathways play an important role in plant disease resistance against fungal pathogens, and further studies are needed to examine potential involvement of the wheat momilactone-like clusters in defence mechanisms.

Two homoeologous clusters FRGC 27 & 34 (Table 1) match with terpene metabolic gene clusters (MGC 1 on chromosome 5A and MGC 2 on chromosome 5D, Supplementary Dataset 2). Except for an *F-box* gene included in the MGC, all genes are responsive to *Fusarium*. FRGC 27 & 34 contain three *CYP450s* from the subfamily *CYP51H*, a 3- β -hydroxysteroid dehydrogenase/isomerase, and a terpene cyclase with 85% protein similarity to the rice cycloartenol synthase *OsOSC2*⁴³. The *CYP51Hs* may function in triterpene synthesis, as has been seen for *SAD2*, a *CYP51H* from *Avena strigosa*⁴³. Interestingly, *SAD1* codes for a cycloartenol synthase and forms a metabolic cluster with *SAD2*. *SAD2* is involved in the biosynthesis of the avenacin, a triterpenoid saponin that is involved in oat resistance to a variety of fungal pathogens, including the FHB causal agent and DON producer *Fusarium culmorum*⁴⁴. The metabolic cluster FRGC 36 on chromosome 5D matched with MGC 11 and is a polyketide type (Table 1, Supplementary Dataset 2), containing a *CYP450* subfamily *CYP71C* and flavonoid biosynthetic genes coding for four caffeic acid/5-hydroxyferulic acid O-methyltransferases (OMTs), two chalcone synthases (CHSs) and one chalcone isomerase (CHI).

Paralogous gene clusters. Four of the metabolic and twelve of the non-metabolic FRGCs contain paralogous genes (Table 1). Two of the four metabolic FRGCs contain strongly co-expressed genes (Table S3) and contain genes annotated as ATP-dependent zinc metalloproteases FtSHs (FRGC 7), and phenylalanine ammonia-lyases (*PAL*) (FRGC 10). Ten of the twelve non-metabolic paralogous FRGCs contain strongly co-expressed genes (Table S3). These clusters code for glutathione S-transferases (*GSTs*) (FRGC 1, 4), Mitogen Activated Protein kinase kinase kinases (*MAPKKKs*) (FRGC 5), *RZ53s* (FRGC 22), germin-like proteins (FRGC 25), protein of unknown function-DUF538s (FRGC 29), zinc finger proteins (FRGC 30, 33) and MYBs (FRGC 37, 38).

The *GST* clusters correspond to three homoeologous regions found on the group 1 chromosomes and contain 10 *GST* genes (Supplementary Table S2). Only the *GST* clusters on chromosome 1A (FRGC 1) and 1D (FRGC 4) passed our significance threshold for *Fusarium* response enrichment. As illustrated on Fig. 3c for the chromosome 1A cluster, *GST* clusters are made up with 8 *GSTs* induced for at least two consecutive time points from 12 to 48 h post inoculation with *Fusarium* (Supplementary Table S4) and are highly co-expressed ($|PCC| = 0.87-0.92$) (Fig. 3c, Supplementary Fig. S2). Non-*GST* genes in both clusters were not, or were slightly, co-expressed with the *GSTs*. *GST* gene clusters are common in plants, and in bread wheat *GSTs* form 37 gene clusters⁴⁵. During plant-pathogen interactions, *GSTs* can function in diverse biological processes and recently, it has been shown that *Fhb7*, a major FHB resistance gene, encodes for a *GST* that confers resistance to *Fusarium* by detoxifying trichothecene mycotoxins via glutathione conjugation²². While the *Fhb7* *GST* belongs to the fungal *GST* etherase-related class²², *GSTs* from FRGC 1 and 4 correspond to the *GST* class Tau (*GSTU*)⁴⁵. Interestingly, a *GSTU* cluster in cotton (*Gossypium hirsutum*) that plays an important role in *Verticillium* wilt resistance⁴⁶, codes for *GSTs* with high similarity (> 60% identity) to some of the wheat *GSTs* within FRGC1 and 4. In cotton, over-expression or silencing of *GSTs* enhanced resistance and susceptibility to *Verticillium dahlia*, respectively⁴⁶. The authors suggested that the cotton *GST* cluster might enhance plant resistance to *V. dahlia* by fine tuning the balance between production and scavenging of H_2O_2 .

Case study: the nonhomologous, non-metabolic *TaFROG* cluster. As stipulated by Nutzmann et al.⁸, it is interesting to consider that nonhomologous gene clusters, which are not secondary metabolism clusters, might be present in plant genomes. It is possible that nonhomologous, non-metabolic co-expressed genes may be involved in the same biological process, triggered by *Fusarium* treatment. Six nonhomologous non-metabolic gene clusters (FRGCs 3, 12, 26, 35, 39 & 40) contain genes that are strongly co-expressed (Supplementary Fig. S2, Table S3). One of the characteristics of these clusters is that they all contain at least one receptor-like kinase (*RLK*) gene (Table 1, Supplementary Table S1). Interestingly, FRGC 26 encompasses the D genome homoeologue of the recently characterised Pooideae-specific *Fusarium* resistance orphan gene (*TaFROG-A*), an orphan gene known to contribute to resistance to *F. graminearum* and the mycotoxin DON⁴⁷. Other *Fusarium*-responsive genes within this cluster include the well-conserved alcohol dehydrogenase (*ADH*) genes *Adh1* and *Adh2*, the Poaceae-specific uncharacterised gene *RZ53*, a *LecRK*, a SWAP/surp domain-containing protein (*SWAP*) and a carboxypeptidase (*CDP*) gene (Fig. 3d). Three other genes are present but not responsive to *Fusarium*. Genes within this locus are previously shown to be important in the plant stress response. *RZ53* was used to delineate QTL for aluminium and cold tolerance, and brown planthopper resistance in rice⁴⁸⁻⁵⁰. The closest wheat *LecRK* (L-type lectin receptor kinases) in Arabidopsis, *LecRK-IX.1*, regulates *Phytophthora* resist-

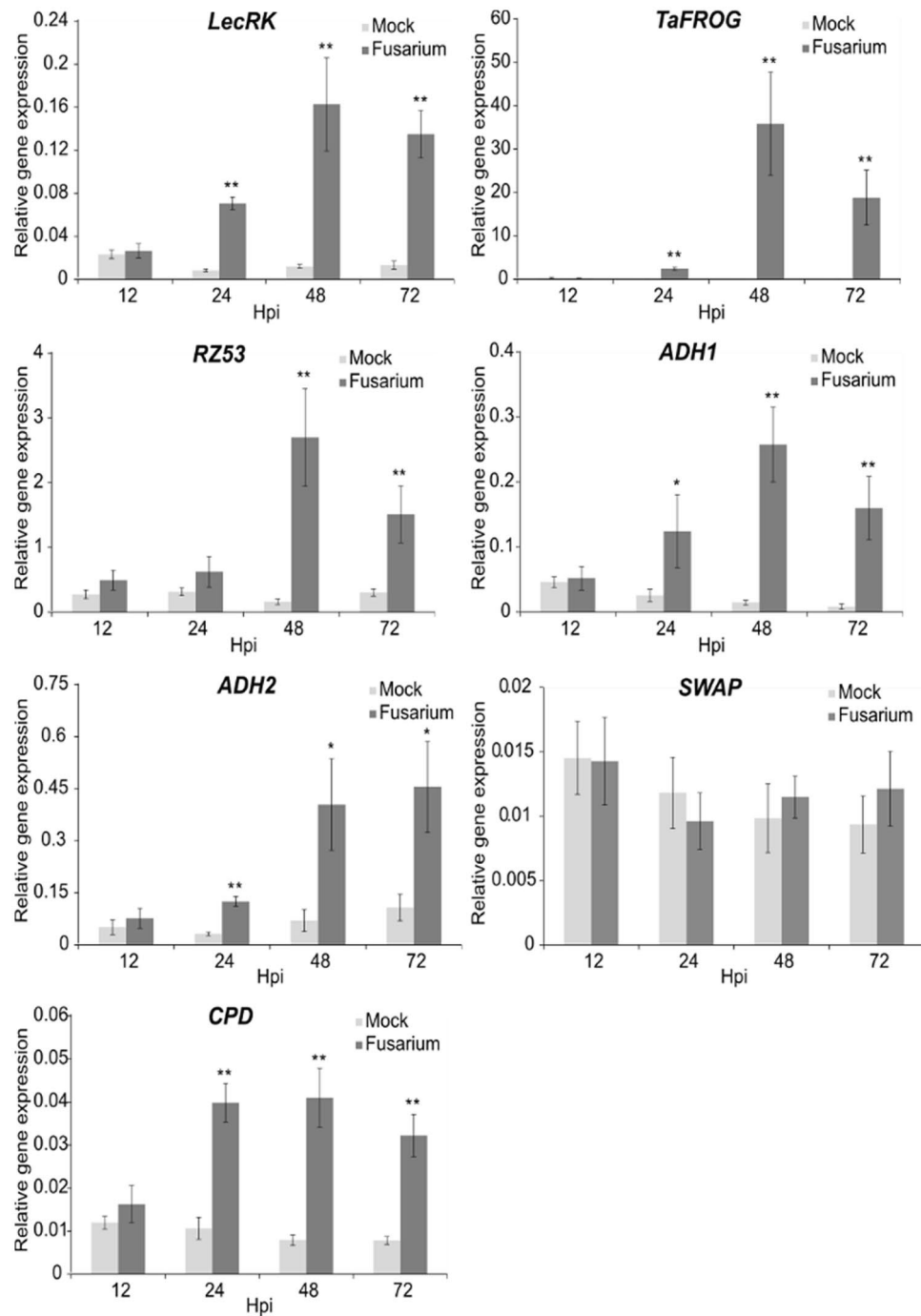


Figure 4. Expression profiles of *Fusarium*-responsive genes within the TaFROG cluster. Transcript levels in wheat (cv. CM-82036) heads after treatment with *Fusarium graminearum* was measured by qRT-PCR using the formula $2^{-(Ct \text{ target gene} - Ct \text{ average housekeeping genes})}$. Spikelets were harvested at various hours post-inoculation (hpi) as indicated. Results represent the mean of three independent trials (each include two technical replicates per treatment from a pooled of 4 biological sample) and error bars represent \pm S.E.M. Asterisks show significant differences (* $P < 0.05$, ** $P < 0.01$).

ance and controls cell death in *Nicotiana benthamiana*^{51, 52}. *HvADH1*, the barley orthologue of *ADH1*, reduces powdery mildew penetration when silenced and increases fungal penetration success when overexpressed^{53, 54}. The carboxypeptidase (*CPD*) gene codes for an orthologue of the Arabidopsis protein SCPL42 (AT5G42240), which belong to the serine carboxypeptidase clade 2⁵⁵. The expression of the genes within FRGC 26 was confirmed by quantitative reverse transcriptase PCR (qRT-PCR) in an independent experiment using the wheat FHB-resistant cultivar CM-82036, with all the genes being induced by *Fusarium* treatment at different time points post-inoculation, except for *SWAP* (Fig. 4). To conclude, it appears that except for *SWAP*, all of the genes

within the *TaFROG-D* cluster are directly or indirectly linked to stress resistance and are highly co-expressed upon *Fusarium* treatment. It is known for metabolic gene clusters that co-expression is an indicator of genes functioning in the same biological process⁵⁶. Therefore, we speculate that like *TaFROG-A*, *TaFROG-D* and the other *Fusarium*-responsive genes within this cluster may be involved in a similar biological process and may contribute to *Fusarium* resistance. We note that the *TaFROG-D* cluster genes are conserved on chromosomes A and B, but fewer of them are *Fusarium*-responsive (Supplementary Table S2) and were therefore not detected as significant clusters. We notice in particular that the Poaceae-specific gene *RZ53* is adjacent to, and highly co-expressed with *TaFROG-D* ($r=0.99$) (Figs. 3d, 4). In addition to *TaFROG-D* and *RZ53*, the cluster also contains a *LecRK*. Interestingly, orphans and poplar-specific genes were found to be frequently co-located with leucine-rich repeat (LRR) domain receptor-like proteins within poplar⁵⁷. Based on expression data, the authors suggested that a pair of physically associated poplar-specific and receptor-like gene products might be functionally linked in response to wounding stress. Similarly, we proposed that the LRR might be functionally linked to Poaceae-specific genes in response to *Fusarium*. Future studies will need to explore potential functional linked between *TaFROG* and the other genes in the cluster, and their impact on *Fusarium* resistance.

Potential *Fusarium*-responsive gene clusters linked with FHB QTL. As the wheat genotypes used in this study segregate for FHB QTL, we compared gene expression data from FRGCs in NIL38 (carrying QTL *Fhb1* and *Qfhs.ifa-5A*) and NIL51 wheat (which lacks these two QTL). In most of the cases, expression of *Fusarium*-responsive genes was similar across genotypes. 30 of the 44 FRGCs (68%) appeared independently in both NILs, and components of most FRGCs were represented in both NILs (Supplementary Table S4). FRGCs 2, 9, 16 and 33 (on chromosomes 1A, 2A, 2D and 5B, respectively) passed the significance threshold as a cluster in one but not both NILs. Overall, this indicates that the expression of the genes within the FRGCs is not dependent on the presence/absence of the FHB QTL *Fhb1* and *Qfhs.ifa-5A*.

To explore proximity of FRGCs to other known FHB QTL, we physically positioned 216 molecular markers linked to FHB QTL, from a recent compilation of the 556 QTL related to FHB resistance⁵⁸. The physical position of each marker was compared to the position of the 44 FRGCs to identify if any FRGCs are close to FHB QTL (Fig. 5). We found 6 FHB QTL less than 150 Mb wide that each encompass one to two FRGCs (Fig. 5, Table 2). Although 5 of these QTL regions were large (44.2–118 Mb) and contain many genes (472–1071 genes), we noted that QTL-1A, QTL-2B and QTL-4D regions each include one of the FRGCs previously described in this study; a GST cluster (FRGC 1), the benzoxazinoid-like gene cluster (FRGC 11) and the *TaFROG* gene cluster (FRGC 26), respectively. On chromosome 7D, QTL-7D has the smallest length (4.7 Mb) and contains only 43 high confidence genes based on the wheat reference gene annotation (Supplementary Table S5). Eight of these genes were *Fusarium*-responsive, of which five form the cluster FRGC 41. These genes code for three paralogue EXORDIUM-like 1 proteins, a glycosyltransferase, and an RNA-binding protein. All five genes were induced at 36 and 48 hpi in the FHB resistant NIL-38, whereas in FHB susceptible NIL-51, two EXORDIUM-like 1 and the RNA-binding genes were induced at only 48 h, and one EXORDIUM-like gene was not differentially expressed (Supplementary Table S4). FRGC 41 genes had a weak overall co-expression correlation ($r=0.49$) but a high co-expression of the three EXORDIUM-like genes ($r=0.87$) (Supplementary Fig. S2, Table S3). FRGC 41 contains genes potentially linked to FHB resistance mechanisms. Glycosyltransferases belong to the uridine diphosphate (UDP)-glycosyltransferases (UGTs) family, other members of which contribute to DON detoxification and FHB resistance in Poaceae species^{59, 60}. The three EXORDIUM-like 1 proteins share 55% identity with the *A. thaliana* EXORDIUM (EXO), a regulator of brassinosteroid (BR)-responsive genes⁶¹. EXO is a gene marker of BR-responses in *A. thaliana* vegetative tissue and, based on mutant studies, it is suggested that EXO contributes to a signalling process modulating BR-responses⁶². Mutation of (BR)-insensitive 1 (*BRI1*), the main receptor of BR, enhanced barley resistance to FHB and Fusarium crown rot (FCR) caused by *F. culmorum*^{63, 64}. Additionally, a T-DNA insertion in the 5' UTR of *Brachypodium distachyon* *BRI1* impaired BR signalling and reduced susceptibility to FHB and Fusarium root rot (FRR)⁶³. Thus, based on small number of genes within the QTL-7D physical region, their *Fusarium* responsiveness, and potential function of genes within the cluster in response to FHB, further studies are worthwhile in order to uncover potential role of FRGC 41 genes in FHB resistance.

Of all the markers reported by Venske et al.⁵⁸, 10 molecular markers that are either peak or flanking markers of FHB QTL are positioned less than 5 Mb from FRGC genes (Fig. 5, Supplementary Table S6). In total, 5 clusters (FRGC 5, 9, 15, 22 and 25) were found to be less than 5 Mb from QTL markers. Four of these FRGCs are paralogous clusters, with three containing strongly co-expressed genes, such as the MAPKKKs (FRGC 5), *RZ53s* (FRGC 22) and germin-like protein 1 s (FRGC 25) clusters, and a ribulose biphosphate carboxylase small chain (RBCS) cluster (FRGC 9), from which a marker is physically positioned within the cluster region. Finally, the momilactone-like cluster on chromosome 2D (FRGC 15) discussed above, is close to three different markers from three studies using three independent resistant sources (Supplementary Table S6). Thus, we suggest that these FRGCs might be considered in future studies as potential regions containing genes underpinning FHB QTL. It is important to note that our method is limited to genome availabilities and gene expression dataset attributes (e.g. time points). For example, *PFT*, one of the genes underpinning the *Fhb1* QTL, is *Fusarium*-responsive²¹, but is not present in the wheat cv. Chinese spring reference genome. Therefore, we were unable to detect it in this study.

Conclusion

In this study, we present a bioinformatics pipeline capable of identifying potentially functionally related gene clusters. This pipeline is adaptable to any eukaryotic genome, and only requires a reference gene annotation and whole genome expression data. Using this pipeline, we identified 44 physical clusters of *F. graminearum* responsive genes within the bread wheat genome. While clusters of functionally related metabolic genes are common through genomes, we found clusters of non-metabolic and nonhomologous genes that are co-expressed under

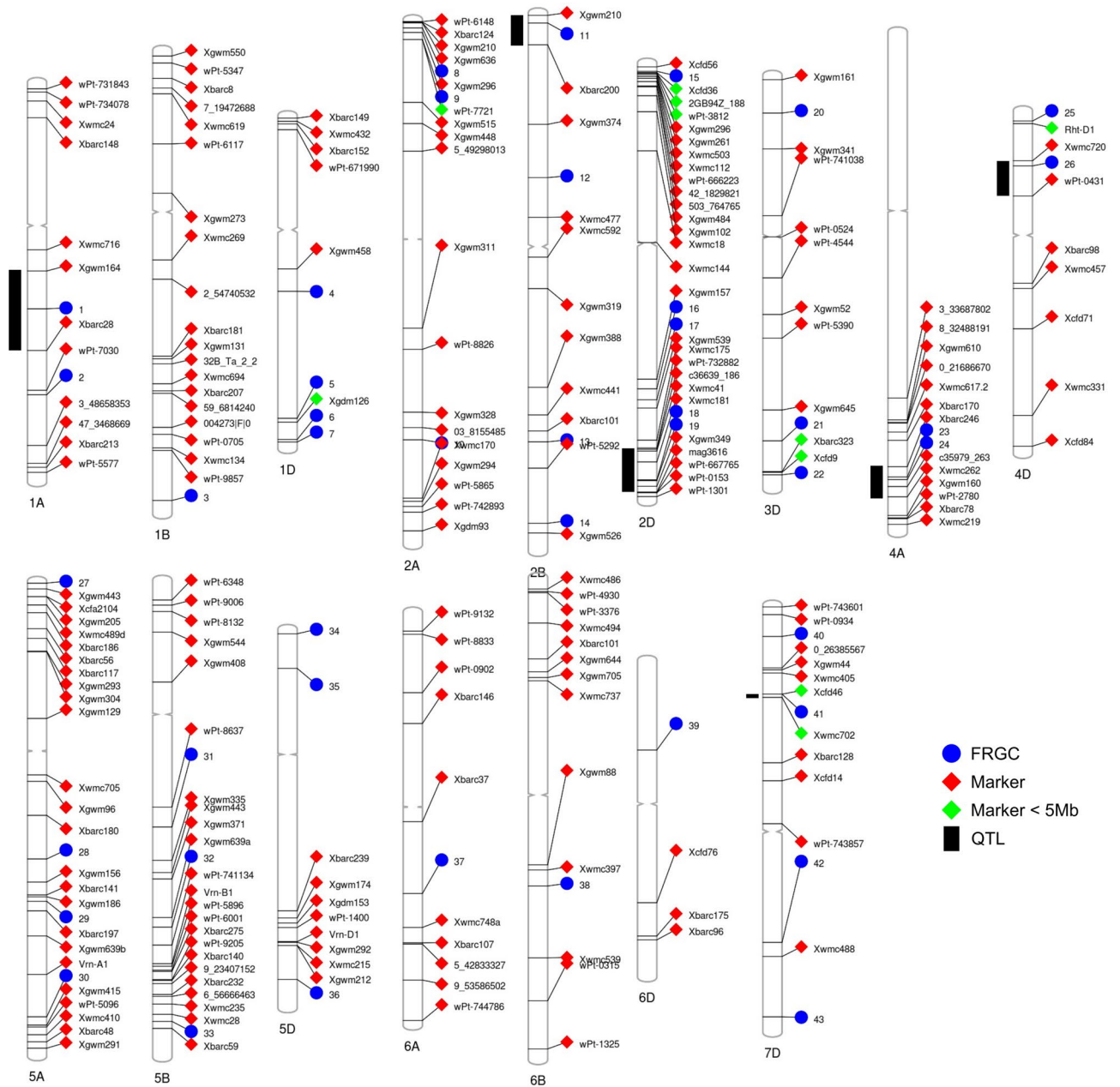


Figure 5. Chromosomal distribution of FRGCs, FHB QTL and associated markers. *Fusarium*-responsive gene clusters and 216 FHB QTL molecular markers were mapped based on their genome positions on the 21 wheat chromosomes. Six QTL regions (Chromosome 1A, 2B, 2D, 4A, 4D and 7D), in which the distance between flanking markers is less than 150 Mb and encompass FRGCs, are indicated with a black vertical bar. FRGCs are represented with a blue circle, and markers with a square, either red if the marker is physically > 5 Mb from a cluster, or green if the marker is physically < 5 Mb from a cluster. Centromeric region are represented as the restricted central region of the chromosomes.

Fusarium infection, and therefore may have co-evolved in response to stress. Additionally, we found that many of the gene clusters are physically close to and/or are within known FHB QTL, and therefore this analysis may serve as an additional layer of genetic information when searching for candidate genes underpinning important QTL.

Materials and methods

Wheat genome data. Wheat high confidence gene annotations (IWGSC RefSeq v1.1)¹⁷ and their coordinates (.BED files) were obtained from the IWGSC sequence repository at URGI (<https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations>) using the JBrowse tool (<https://wheat-urgi.versailles.inra.fr/Tools/JBrowse>). The wheat genome assembly (IWGSC RefSeq v1.0) as a .fasta file and high confidence genes annotation as GFF3 file (IWGSC RefSeq v1.1) were downloaded using the FTP server from Ensembl plant version 44 (<https://>

QTL	Chromosomes	Sources	Flanking markers		Distance between markers (Mb)	Cluster(s) within QTL	References
QTL-1A	1A	Pelikan	Xgwm164	Xbarc28	118	1	Häberle et al. ⁶⁵
QTL-2B	2B	Goldfield	Xbarc200	Xgwm210-2B	44.2	11	Gilsinger et al. ⁶⁶
QTL-2D	2D	Frontana	wPt-732882	wPt-667765	67.4	18, 19	Szabó-Hevér et al. ⁶⁷
QTL-4A	4A	86ISMN_2137	Xbarc246	Xwmc262	44.4	23, 24	McCartney et al. ⁶⁸
QTL-4D	4D	86ISMN_2137	Xwmc720	wPt-0431	52.8	26	McCartney et al. ⁶⁸
QTL-7D	7D	Haiyanzhong	Xcfd46	Xwmc702	4.7	41	Li et al. ⁶⁹

Table 2. FHB QTL that are smaller than 150 Mb and encompass FRGC.

plants.ensembl.org/Triticum_aestivum/Info/Index). Wheat gene homoeologues, paralogues, and orthologues in *Arabidopsis thaliana*, *Oryza sativa* and *Hordeum vulgare* were identified via Ensembl Plants BioMart tool.

RNAseq data collection and expression analysis. The gene expression dataset used in this study details RNAseq of FHB treatment of the near isogenic (NIL) wheat lines NIL-38 and NIL-51²⁴. Briefly, these NILs are derived from cross between FHB-resistant line CM-82036 and FHB susceptible cultivar Remus, with CM-82036 used as the recurring parent in the backcross. NIL-38 is homozygous for both resistant alleles at the *Fhb1* and *Qhfs.ifa-5A* QTL, and NIL-51 is homozygous for both susceptible alleles at these QTL²⁴. A gene count matrix of this RNAseq dataset was downloaded from www.wheat-expression.com^{70,71} and differential expression analysis was done as per Benbow et al.⁷², where differential expression was conducted per NIL per timepoint.

Development of a gene cluster and co-expression analysis pipeline. The gene cluster analysis pipeline was written in the R programming language⁷³, and determines density and consecutiveness of disease responsive genes. Gene density is defined as the number of disease responsive genes in the given window (i.e. 5 disease responsive genes in a window size of 10 = a density of 0.5). Consecutiveness is the number of disease responsive genes in a continuous string within a window, i.e. 5 disease responsive genes in a row, with no non-disease responsive genes between them. The pipeline consists of three main elements, (1) a permutation test to determine a significance threshold λ , (2) a cluster analysis to determine density and consecutiveness of disease-responsive genes within a sliding window of size χ (user defined, in this study we used a window size of 10 genes) using the pre-determined threshold λ , and (3) determining the correlation coefficients of genes within a cluster versus those outside the cluster. The data inputs are: a .BED file of gene coordinates, with columns “Chromosome”, “Start position (bp)”, “End position (bp)”, and “Gene ID”, and an expression data file, in which column 1 is “GeneID”, and subsequent columns are binary designations of 0 = non differentially expressed, and 1 = differentially expressed for each treatment (i.e. *Fusarium graminearum*). The input files used in this study are supplied as supplementary dataset 3. The permutation test takes a user-defined number of permutations n ($n = 10,000$, in this study), and creates n random datasets, where the number of differentially expressed genes and the genome coordinates of all genes remains constant, but the binary designation of disease-responsiveness is scrambled across the genome. The gene cluster analysis pipeline is then run on all n of these random datasets, and the frequency distribution of gene density and consecutiveness are calculated as a percentage. This allows the user to pick λ based on the frequency of λ occurring by chance. The sliding window analysis takes the .BED file, the expression matrix, and the predetermined λ value and the window size χ and calculates the average number of disease-responsive genes within the window, the gene density. Subsequently, the number of consecutive DEGs is calculated. The pipeline returns a .csv file per chromosome with the gene density and number of consecutive DEGs appended to it, and a dot plot showing gene density across the chromosomes. The final part of the pipeline is the correlation analysis. Within this, raw expression data (the count matrix of gene expression across all samples) is extracted for all DEGs within each gene cluster, and Pearson's correlation coefficient (PCC; r) is calculated for every pair of genes. This is repeated on the cluster ‘flanking region’ of y genes either side of the cluster (y is user-defined; $y = 20$ in these study). The average of absolute PCCs is calculated, and a Kruskal-Wallis test is conducted to determine if the absolute PCC within a cluster is significantly different to the absolute PCC outside the cluster. The analysis pipeline scripts are all available at https://github.com/hbenbow/Gene_clusters. All plots created by the analysis pipeline are generated with the R package ‘ggplot2’⁷⁴.

Gene and metabolic cluster annotation. The wheat gene annotation corresponds to the “Human-Readable-Description” from the IWGSC RefSeq v1.0 annotation. Gene descriptions were updated for the MAP kinase kinase kinase genes utilising homology with *Arabidopsis* MAP kinases, for NLR with the NLR recent annotation⁷⁵, and for the CYP genes using homology and nomenclature of rice CYPs⁷⁶ combined with the wheat CYPs nomenclature curated on the wheat TGACv1 annotation (personal communication by Professor David Nelson, University of Tennessee). Cluster genes were manually classified as metabolic/non-metabolic based on their gene annotation, KEGG pathway ID and Plant reactome ID⁷⁷ retrieved via Ensembl Plants BioMart tool. Additionally secondary metabolic clusters were predicted for each chromosome of the wheat genome with the PlantSMASH online tool (<http://plantismash.secondarymetabolites.org/>²⁹) using default parameters. PlantSMASH output results for each chromosome are reported in Supplementary Dataset 1. Finally, clusters were

classified as metabolic if $\geq 60\%$ of genes were metabolic or if more than 50% of the genes within the cluster were recognised as secondary metabolic genes by the PlantSMASH tool.

Plant material for gene expression. Cultivated plant bread wheat (*Triticum aestivum*) cultivar CM-82036 (kindly provided by Hermann Buerstmayr, BOKU) was used in this study⁷⁸. CM-82036 emerges from the cross Sumai-3/Thornbird-S and was developed by the International Maize and Wheat Improvement Center (CIMMYT, <https://www.cimmyt.org>). This cultivar can be obtained from the CIMMYT wheat germplasm bank (1.10.6) under the identifier BW 45,200 (plant name CM82036-ITP-10Y-0SJ-10Y-10M-0FC). This cultivar is resistant to FHB and carries two QTL for FHB resistance (*Fhb1* and *Qhfs.ifa-5A*) (as does its' NIL38 derivative used to generate the RNAseq data analysed in this study). Plant growth conditions were similar to Perochon et al.⁷⁹, and carried out under contained glasshouse conditions. *Fusarium graminearum* wild type strain GZ3639 was used in this study⁸⁰. Fungal mycelium stored at $-80\text{ }^{\circ}\text{C}$ was subcultured on potato dextrose agar (PDA) (Difco, UK) and plates were incubated 5 days at $25\text{ }^{\circ}\text{C}$. Fungal spores were produced in mung bean broth⁸¹, harvested, washed and adjusted to 10^6 conidia ml^{-1} in 0.02% Tween 20, as previously described⁸². At wheat mid-anthesis, two central spikelets of the heads were inoculated with either 20 μl 0.02% (v/v) Tween 20 (mock) or this solution added with 10^6 conidia of *F. graminearum*. Treated heads were covered with plastic bags for 2 d before being harvested at various timepoints post-inoculation and were flash frozen in liquid N_2 prior to RNA extraction. The experiment comprised three independent trials, each including eight heads from four individual plants (two heads per plant) per treatment combination. RNA was extracted from one pooled sample per treatment (representing a pool of 8 heads). RNA extractions and cDNA synthesis were performed as described previously by Perochon et al.⁷⁹. Experimental research on plants (glasshouse studies on cultivated wheat, including obtaining the wheat material from BOKU University) complied with relevant UCD, Irish, EU and international guidelines and legislation.

Quantitative reverse transcriptase PCR analysis. qRT-PCR analysis was conducted using the Mx3000 Real-Time PCR (Stratagene, Germany). Homoeologue-specific PCR primers used in this study (listed in Supplementary Table S7) were designed using Primer3web (<http://primer3.ut.ee/>⁸³) and Genome Specific Primer (GSP) online tools⁸⁴. Specificity of the primers was verified by PCR of DNA extracts from nullisomic-tetrasomic lines of cv. Chinese Spring (obtained from Germplasm Resources Unit, JIC, Norwich <http://www.jic.ac.uk/germplasm/>), analysis of the dissociation curve, electrophoretic profile of PCR products and sequence analysis. Housekeeping genes used were phosphatase 2A subunit A3 (*TaPP2AA3*) and Yellow-leaf specific gene 8 (*TaYLS8*) and were previously described to be not differentially expressed in FHB experiments⁸⁵. Each qRT-PCR reaction contained 1.25 μL of cDNA, 0.2 μm of each of the primers (except for *LecRK* and *CPD*, concentrations were 0.15 and 0.1 μm , respectively) and $1 \times$ SYBR Premix Ex Taq (Tli RNase H plus, RR420A; Takara) in a total reaction volume of 12.5 μL . PCR conditions were as follows: 1 cycle of 1 min at $95\text{ }^{\circ}\text{C}$; 40 cycles of 5 s at $95\text{ }^{\circ}\text{C}$ and 20 s at $60\text{ }^{\circ}\text{C}$; and a final cycle of 1 min at $95\text{ }^{\circ}\text{C}$, 30 s, at $55\text{ }^{\circ}\text{C}$, and 30 s at $95\text{ }^{\circ}\text{C}$ for the dissociation curve. All qPCR analyses were conducted in duplicate using two cDNA generated from independent reverse transcriptions. The threshold cycle (Ct) values obtained by qPCR were used to calculate the relative gene expression using the formula $2^{-(\text{Ct target gene} - \text{Ct housekeeping gene})}$ as described previously⁸⁶.

Identification of FRGCs that lie within FHB QTL intervals. To identify if any of the FRGCs were physically located within a previously identified QTL region for *Fusarium* resistance, we used wheat FHB QTL information documented by a systematic review of the wheat FHB QTLome⁵⁸. We extracted marker information linked to QTL and downloaded their physical positions from wheat JBrowse (<https://wheat-urgi.versailles.inra.fr/Tools/JBrowse>). Using a bespoke R script (available at https://github.com/hbenbow/Gene_clusters), we retrieved the left and right position (cM) of each QTL. The physical position of each of these markers (in bp) was extracted, and all markers for which the chromosome designation for the physical map and the genetic map did not agree were removed. A subset of the genome .BED file was then extracted, retrieving only the coordinates within the physical QTL interval. For each QTL interval, FRGCs were identified within the region, and the physical size of the QTL region was identified. For QTL where information on flanking markers were reported, we identified their physical size, chromosomal position and if they contained FRGCs. For every marker we calculated their distance from a FRGC. The figure of the chromosomal distribution of QTL, markers and FRGCs was generated using the visualisation tool from Ritchie lab (<http://visualization.ritchielab.org/phenograms/plot>).

Statistical analysis and figure drawing . The normality of the data distribution was evaluated with the Shapiro–Wilk test using R⁷³ and SPSS statistics version 24 software (IBM). Statistical analyses were performed with R for cluster gene co-expression using Kruskal–Wallis, with SPSS for qRT-PCR gene expression analyses using Mann–Whitney U test and comparing individual treatments. All figures were draw using the free vector graphics software Inkscape 1.0 (<https://inkscape.org/>).

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files). Datasets are also available from the corresponding author on reasonable request and the wheat cultivar used in this study can be obtained from the CIMMYT wheat germplasm bank (1.10.6) under the identifier BW 45,200 (plant name CM82036-ITP-10Y-0SJ-10Y-10M-0FC).

Received: 1 October 2020; Accepted: 8 March 2021

Published online: 02 April 2021

References

- Hurst, L. D., Pal, C. & Lercher, M. J. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299–310. <https://doi.org/10.1038/nrg1319> (2004).
- Osbourne, A. E. & Field, B. Operons. *Cell Mol. Life Sci.* **66**, 3755–3775. <https://doi.org/10.1007/s00018-009-0114-3> (2009).
- Hittinger, C. T., Rokas, A. & Carroll, S. B. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl. Acad. Sci. USA* **101**, 14144–14149. <https://doi.org/10.1073/pnas.0404319101> (2004).
- Wong, S. & Wolfe, K. H. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat. Genet.* **37**, 777–782. <https://doi.org/10.1038/ng1584> (2005).
- Rokas, A., Wisecaver, J. H. & Lind, A. L. The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.* **16**, 731–744. <https://doi.org/10.1038/s41579-018-0075-3> (2018).
- Horton, R. *et al.* Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**, 889–899. <https://doi.org/10.1038/nrg1489> (2004).
- Boycheva, S., Daviet, L., Wolfender, J. L. & Fitzpatrick, T. B. The rise of operon-like gene clusters in plants. *Trends Plant. Sci.* **19**, 447–459. <https://doi.org/10.1016/j.tplants.2014.01.013> (2014).
- Nutzmann, H. W., Huang, A. & Osbourne, A. Plant metabolic clusters—from genetics to genomics. *New Phytol.* **211**, 771–789. <https://doi.org/10.1111/nph.13981> (2016).
- Frey, M. *et al.* Analysis of a chemical plant defense mechanism in grasses. *Science* **277**, 696–699. <https://doi.org/10.1126/science.277.5326.696> (1997).
- Panchy, N., Lehti-Shiu, M. & Shiu, S. H. Evolution of gene duplication in plants. *Plant Physiol.* **171**, 2294–2316. <https://doi.org/10.1104/pp.16.00523> (2016).
- Saintenac, C. *et al.* Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to fungal pathogen *Zymoseptoria tritici*. *Nat. Genet.* **50**, 368–374. <https://doi.org/10.1038/s41588-018-0051-x> (2018).
- Liu, Y. *et al.* A gene cluster encoding lectin receptor kinases confers broad-spectrum and durable insect resistance in rice. *Nat. Biotechnol.* **33**, 301–305. <https://doi.org/10.1038/nbt.3069> (2015).
- van Wersch, S. & Li, X. Stronger when together: clustering of plant NLR disease resistance genes. *Trends Plant. Sci.* **24**, 688–699. <https://doi.org/10.1016/j.tplants.2019.05.005> (2019).
- Xu, K. *et al.* Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**, 705–708. <https://doi.org/10.1038/nature04920> (2006).
- Jeon, J. E. *et al.* A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* **180**, 176–187.e119. <https://doi.org/10.1016/j.cell.2019.11.037> (2020).
- Wang, M. *et al.* TaCYP81D5, one member in a wheat cytochrome P450 gene cluster, confers salinity tolerance via reactive oxygen species scavenging. *Plant Biotechnol. J.* **18**, 791–804. <https://doi.org/10.1111/pbi.13247> (2020).
- IWGSC. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, 661. <https://doi.org/10.1126/science.aar7191> (2018).
- Buerstmayr, M., Steiner, B., Buerstmayr, H. & Léon, J. Breeding for Fusarium head blight resistance in wheat—progress and challenges. *Plant Breed.* **139**, 429–454. <https://doi.org/10.1111/pbr.12797> (2019).
- Su, Z. *et al.* A deletion mutation in TaHRC confers Fhb1 resistance to Fusarium head blight in wheat. *Nat. Genet.* **51**, 1099–1105. <https://doi.org/10.1038/s41588-019-0425-8> (2019).
- Li, G. *et al.* Mutation of a histidine-rich calcium-binding-protein gene in wheat confers resistance to Fusarium head blight. *Nat. Genet.* **51**, 1106–1112. <https://doi.org/10.1038/s41588-019-0426-7> (2019).
- Rawat, N. *et al.* Wheat Fhb1 encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. *Nat. Genet.* **48**, 1576–1580. <https://doi.org/10.1038/ng.3706> (2016).
- Wang, H. *et al.* Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. *Science* **368**, eaba5435. <https://doi.org/10.1126/science.aba5435> (2020).
- Kazan, K. & Gardiner, D. M. Transcriptomics of cereal-Fusarium *graminearum* interactions: what we have learned so far. *Mol. Plant Pathol.* <https://doi.org/10.1111/mpp.12561> (2017).
- Schweiger, W. *et al.* Suppressed recombination and unique candidate genes in the divergent haplotype encoding Fhb1, a major Fusarium head blight resistance locus in wheat. *Theor. Appl. Genet.* **129**, 1607–1623. <https://doi.org/10.1007/s00122-016-2727-x> (2016).
- Schlapfer, P. *et al.* Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol.* **173**, 2041–2059. <https://doi.org/10.1104/pp.16.01942> (2017).
- Nussbaumer, T. *et al.* Joint transcriptomic and metabolomic analyses reveal changes in the primary metabolism and imbalances in the subgenome orchestration in the bread wheat molecular response to *Fusarium graminearum*. *G3 (Bethesda)* **5**, 2579–2592. <https://doi.org/10.1534/g3.115.021550> (2015).
- Zhang, Y. *et al.* Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*Triticum aestivum* L.). *Sci. Rep.* **6**, 21476. <https://doi.org/10.1038/srep21476> (2016).
- Saito, K., Hirai, M. Y. & Yonekura-Sakakibara, K. Decoding genes with coexpression networks and metabolomics: “majority report by precogs”. *Trends Plant. Sci.* **13**, 36–43. <https://doi.org/10.1016/j.tplants.2007.10.006> (2008).
- Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourne, A. & Medema, M. H. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* **45**, W55–W63. <https://doi.org/10.1093/nar/gkx305> (2017).
- Gadaleta, A., Colasuonno, P., Giove, S. L., Blanco, A. & Giancaspro, A. Map-based cloning of QFhb.mgb-2A identifies a WAK2 gene responsible for Fusarium Head Blight resistance in wheat. *Sci. Rep.* **9**, 6929. <https://doi.org/10.1038/s41598-019-43334-z> (2019).
- Nelson, D. & Werck-Reichhart, D. A P450-centric view of plant evolution. *Plant J.* **66**, 194–211. <https://doi.org/10.1111/j.1365-3113.2011.04529.x> (2011).
- Nomura, T., Ishihara, A., Yanagita, R. C., Endo, T. R. & Iwamura, H. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proc. Natl. Acad. Sci. USA* **102**, 16490–16495. <https://doi.org/10.1073/pnas.0505156102> (2005).
- Dutartre, L., Hilliou, F. & Feyerisen, R. Phylogenomics of the benzoxazinoid biosynthetic pathway of Poaceae: gene duplications and origin of the Bx cluster. *BMC Evol. Biol.* **12**, 64. <https://doi.org/10.1186/1471-2148-12-64> (2012).
- Solttoft, M., Jørgensen, L. N., Svensmark, B. & Fomsgaard, I. S. Benzoxazinoid concentrations show correlation with Fusarium Head Blight resistance in Danish wheat varieties. *Biochem. Syst. Ecol.* **36**, 245–259 (2008).
- Kettle, A. J. *et al.* Degradation of the benzoxazinone class of phytoalexins is important for virulence of *Fusarium pseudograminearum* towards wheat. *Mol. Plant Pathol.* **16**, 946–962. <https://doi.org/10.1111/mpp.12250> (2015).
- Powell, J. J. *et al.* The Fusarium crown rot pathogen *Fusarium pseudograminearum* triggers a suite of transcriptional and metabolic changes in bread wheat (*Triticum aestivum* L.). *Ann. Bot.* **119**, 853–867. <https://doi.org/10.1093/aob/mcw207> (2017).
- Shimura, K. *et al.* Identification of a biosynthetic gene cluster in rice for momilactones. *J. Biol. Chem.* **282**, 34013–34018. <https://doi.org/10.1074/jbc.M703344200> (2007).

38. Schmelz, E. A. *et al.* Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J.* **79**, 659–678. <https://doi.org/10.1111/tj.12436> (2014).
39. Murphy, K. M. & Zerbe, P. Specialized diterpenoid metabolism in monocot crops: biosynthesis and chemical diversity. *Phytochemistry* **172**, 112289. <https://doi.org/10.1016/j.phytochem.2020.112289> (2020).
40. Zhao, M., Cheng, J., Guo, B., Duan, J. & Che, C. T. Momilactone and related diterpenoids as potential agricultural chemicals. *J. Agric. Food Chem.* **66**, 7859–7872. <https://doi.org/10.1021/acs.jafc.8b02602> (2018).
41. Ding, Y. *et al.* Multiple genes recruited from hormone pathways partition maize diterpenoid defences. *Nat. Plants* **5**, 1043–1056. <https://doi.org/10.1038/s41477-019-0509-6> (2019).
42. Lu, X. *et al.* Inferring roles in defense from metabolic allocation of rice diterpenoids. *Plant Cell* **30**, 1119–1131. <https://doi.org/10.1105/tpc.18.00205> (2018).
43. Inagaki, Y. S. *et al.* Investigation of the potential for triterpene synthesis in rice through genome mining and metabolic engineering. *New Phytol.* **191**, 432–448. <https://doi.org/10.1111/j.1469-8137.2011.03712.x> (2011).
44. Papadopoulou, K., Melton, R. E., Leggett, M., Daniels, M. J. & Osbourn, A. E. Compromised disease resistance in saponin-deficient plants. *Proc. Natl. Acad. Sci. USA* **96**, 12923–12928. <https://doi.org/10.1073/pnas.96.22.12923> (1999).
45. Wang, R. *et al.* Genome-wide identification and expression profiling of glutathione transferase gene family under multiple stresses and hormone treatments in wheat (*Triticum aestivum* L.). *BMC Genom.* **20**, 986. <https://doi.org/10.1186/s12864-019-6374-x> (2019).
46. Li, Z. K. *et al.* A newly identified cluster of glutathione S-transferase genes provides Verticillium wilt resistance in cotton. *Plant J.* **98**, 213–227. <https://doi.org/10.1111/tj.14206> (2019).
47. Perochon, A. *et al.* TaFROG encodes a pooidae orphan protein that interacts with SnRK1 and enhances resistance to the myco-toxicogenic fungus *Fusarium graminearum*. *Plant Physiol.* **169**, 2895–2906. <https://doi.org/10.1104/pp.15.01056> (2015).
48. Nguyen, V. T. *et al.* Molecular mapping of genes conferring aluminum tolerance in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **102**, 1002–1010. <https://doi.org/10.1007/s001220000472> (2001).
49. Zhi-Hong, Z., Li, S., Wei, L., Wei, C. & Ying-Guo, Z. A major QTL conferring cold tolerance at the early seedling stage using recombinant inbred lines of rice (*Oryza sativa* L.). *Plant Sci.* **168**, 527–534. <https://doi.org/10.1016/j.plantsci.2004.09.021> (2005).
50. Xu, X. F., Mei, H. W., Luo, L. J., Cheng, X. N. & Li, Z. K. RFLP-facilitated investigation of the quantitative resistance of rice to brown planthopper (*Nilaparvata lugens*). *Theor. Appl. Genet.* **104**, 248–253. <https://doi.org/10.1007/s00122-001-0777-0> (2002).
51. Wang, Y., Nsibo, D. L., Juhar, H. M., Govers, F. & Bouwmeester, K. Ectopic expression of Arabidopsis L-type lectin receptor kinase genes LecRK-I9 and LecRK-IX.1 in *Nicotiana benthamiana* confers Phytophthora resistance. *Plant Cell Rep.* **35**, 845–855. <https://doi.org/10.1007/s00299-015-1926-2> (2016).
52. Wang, Y. *et al.* Arabidopsis lectin receptor kinases LecRK-IX1 and LecRK-IX2 are functional analogs in regulating phytophthora resistance and plant cell death. *Mol. Plant Microbe Interact.* **28**, 1032–1048. <https://doi.org/10.1094/MPMI-02-15-0025-R> (2015).
53. Pathuri, I. P., Reitberger, I. E., Huckelhoven, R. & Proels, R. K. Alcohol dehydrogenase 1 of barley modulates susceptibility to the parasitic fungus *Blumeria graminis* sp. hordei. *J. Exp. Bot.* **62**, 3449–3457. <https://doi.org/10.1093/jxb/err017> (2011).
54. Kasbauer, C. L. *et al.* Barley ADH-1 modulates susceptibility to Bgh and is involved in chitin-induced systemic resistance. *Plant Physiol. Biochem.* **123**, 281–287. <https://doi.org/10.1016/j.plaphy.2017.12.029> (2018).
55. Mugford, S. T. *et al.* A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell* **21**, 2473–2484. <https://doi.org/10.1105/tpc.109.065870> (2009).
56. Wei, H. *et al.* Transcriptional coordination of the metabolic network in Arabidopsis. *Plant Physiol.* **142**, 762–774. <https://doi.org/10.1104/pp.106.080358> (2006).
57. Petre, B., Hacquard, S., Duplessis, S. & Rouhier, N. Genome analysis of poplar LRR-RLP gene clusters reveals RISP, a defense-related gene coding a candidate endogenous peptide elicitor. *Front. Plant Sci.* **5**, 111. <https://doi.org/10.3389/fpls.2014.00111> (2014).
58. Venske, E. *et al.* Meta-analysis of the QTLome of Fusarium Head Blight resistance in bread wheat: refining the current puzzle. *Front. Plant Sci.* **10**, 727. <https://doi.org/10.3389/fpls.2019.00727> (2019).
59. Li, X. *et al.* Transgenic wheat expressing a barley UDP-glucosyltransferase detoxifies deoxynivalenol and provides high levels of resistance to *Fusarium graminearum*. *Mol. Plant Microbe Interact.* **28**, 1237–1246. <https://doi.org/10.1094/MPMI-03-15-0062-R> (2015).
60. Pasquet, J. C. *et al.* A Brachypodium UDP-Glycosyltransferase Confers Root Tolerance to Deoxynivalenol and Resistance to Fusarium Infection. *Plant Physiol.* **172**, 559–574. <https://doi.org/10.1104/pp.16.00371> (2016).
61. Coll-Garcia, D., Mazuch, J., Altmann, T. & Mussig, C. EXORDIUM regulates brassinosteroid-responsive genes. *FEBS Lett* **563**, 82–86. [https://doi.org/10.1016/S0014-5793\(04\)00255-8](https://doi.org/10.1016/S0014-5793(04)00255-8) (2004).
62. Schroder, F., Lisso, J., Lange, P. & Mussig, C. The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. *BMC Plant Biol* **9**, 20. <https://doi.org/10.1186/1471-2229-9-20> (2009).
63. Goddard, R., Peraldi, A., Ridout, C. & Nicholson, P. Enhanced disease resistance caused by BRI1 mutation is conserved between Brachypodium distachyon and barley (*Hordeum vulgare*). *Mol Plant Microbe Interact* **27**, 1095–1106. <https://doi.org/10.1094/MPMI-03-14-0069-R> (2014).
64. Ali, S. S. *et al.* Plant disease resistance is augmented in uzu barley lines modified in the brassinosteroid receptor BRI1. *BMC Plant Biol* **14**, 227. <https://doi.org/10.1186/s12870-014-0227-1> (2014).
65. Häberle, J., Schweizer, G., Schöndelmaier, J., Zimmermann, G. & Hartl, L. Mapping of QTL for resistance against Fusarium head blight in the winter wheat population Pelikan//Bussard/Ning8026. *Plant Breeding* **128**, 27–35. <https://doi.org/10.1111/j.1439-0523.2008.01540.x> (2009).
66. Gilsinger, J., Kong, L., Shen, X. & Ohm, H. DNA markers associated with low Fusarium head blight incidence and narrow flower opening in wheat. *Theor Appl Genet* **110**, 1218–1225. <https://doi.org/10.1007/s00122-005-1953-4> (2005).
67. Szabó-Hevér, Á. *et al.* Erratum to: Differential influence of QTL linked to Fusarium head blight, Fusarium-damaged kernel, deoxynivalenol contents and associated morphological traits in a Frontana-derived wheat population. *Euphytica* **200**, 27–27. <https://doi.org/10.1007/s10681-014-1229-7> (2014).
68. McCartney, C. A. *et al.* Fusarium Head Blight Resistance QTL in the Spring Wheat Cross Kenyon/86ISMN 2137. *Front Microbiol* **7**, 1542. <https://doi.org/10.3389/fmicb.2016.01542> (2016).
69. Li, T., Bai, G., Wu, S. & Gu, S. Quantitative trait loci for resistance to fusarium head blight in a Chinese wheat landrace Haiyanzhong. *Theor Appl Genet* **122**, 1497–1502. <https://doi.org/10.1007/s00122-011-1549-0> (2011).
70. Borrill, P., Ramirez-Gonzalez, R. & Uauy, C. expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiol.* **170**, 2172–2186. <https://doi.org/10.1104/pp.15.01667> (2016).
71. Ramirez-Gonzalez, R. H. *et al.* The transcriptional landscape of polyploid wheat. *Science* **361**, 662–+, doi:<https://doi.org/10.1126/science.aar6089> (2018).
72. Benbow, H. R., Jermini, L. S. & Doohan, F. M. Serpins: Genome-Wide Characterisation and Expression Analysis of the Serine Protease Inhibitor Family in *Triticum aestivum*. *G3-Genes Genomes Genetics* **9**, 2709–2722, doi:<https://doi.org/10.1534/g3.119.400444> (2019).
73. R Core Team. (R Foundation for Statistical Computing, Vienna, Austria 2018).
74. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Ggplot2: Elegant Graphics for Data Analysis*, 1–212, doi:<https://doi.org/10.1007/978-0-387-98141-3> (2009).
75. Steuernagel, B. *et al.* The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol* **183**, 468–482. <https://doi.org/10.1104/pp.19.01273> (2020).

76. Wei, K. & Chen, H. Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice. *BMC Genomics* **19**, 35. <https://doi.org/10.1186/s12864-017-4425-8> (2018).
77. Naithani, S. *et al.* Plant Reactome: a resource for plant pathways and comparative analysis. *Nucleic Acids Res* **45**, D1029–D1039. <https://doi.org/10.1093/nar/gkw932> (2017).
78. Buerstmayr, H. *et al.* Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). *Theor Appl Genet* **104**, 84–91. doi:<https://doi.org/10.1007/s001220200009> (2002).
79. Perochon, A. *et al.* The wheat SnRK1alpha family and its contribution to Fusarium toxin tolerance. *Plant Sci* **288**, 110217. <https://doi.org/10.1016/j.plantsci.2019.110217> (2019).
80. Proctor, R., Hohn, T. & McCormick, S. Main content area Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. *Mol Plant Microbe Interact* **8**, 593–601 (1995).
81. Bai, G.-H. & Shaner, G. Variation in *Fusarium graminearum* and cultivar resistance to wheat scab. *Plant Dis.* **80**, 975–979 (1996).
82. Brennan, J. M., Egan, D., Cooke, B. M. & Doohan, F. M. Effect of temperature on head blight of wheat caused by *Fusarium culmorum* and *F. graminearum*. *Plant Pathol.* **54**, 156–160. <https://doi.org/10.1111/j.1365-3059.2005.01157.x> (2005).
83. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115–e115 (2012).
84. Wang, Y. *et al.* GSP: a web-based platform for designing genome-specific primers in polyploids. *Bioinformatics* **32**, 2382–2383. <https://doi.org/10.1093/bioinformatics/btw134> (2016).
85. Perochon, A. *et al.* A wheat NAC interacts with an orphan protein and enhances resistance to Fusarium head blight disease. *Plant Biotechnol. J.* **17**, 1892–1904. <https://doi.org/10.1111/pbi.13105> (2019).
86. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* **25**, 402–408. <https://doi.org/10.1006/meth.2001.1262> (2001).

Acknowledgements

This work received funding from Science Foundation Ireland Project No. 14/1A/2508, the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 818144, and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 674964. A.P. and K.S. were funded by Science Foundation Ireland Project 14/1A/2508; H.B. was funded by the European Union's Horizon 2020 project INNOVAR under Grant Agreement No. 818144; K.M. was funded by the European Union's Horizon 2020 research and innovation programme Marie Skłodowska-Curie grant CEREALPATH under Grant Agreement No. 674964. The authors thank Prof. David Nelson of the University of Tennessee and Thalia Christodoulou of University College Dublin for their advice and technical support.

Author contributions

A.P., H.B. and F.D. designed the experiments. H.B. wrote the scripts. A.P. analysed the data, A.P., K.S. and K.M. carried out lab work. A.P., H.B. and F.D. wrote the manuscript. All authors commented and approved the final version.

Competing interests

F.D. and A.P. hold a patent on the *TaFROG* gene located within one of the *Fusarium*-responsive gene clusters. H.B., K.S. and K.M. have no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-86362-4>.

Correspondence and requests for materials should be addressed to F.M.D.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021