


RESEARCH ARTICLE

Open Access



# Whole genome sequencing and comparative genomics of closely related *Fusarium* Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*

Sean Walkowiak<sup>1,2</sup>, Owen Rowland<sup>1</sup>, Nicolas Rodrigue<sup>1</sup> and Rajagopal Subramaniam<sup>2\*</sup> 

## Abstract

**Background:** The *Fusarium graminearum* species complex is composed of many distinct fungal species that cause several diseases in economically important crops, including Fusarium Head Blight of wheat. Despite being closely related, these species and individuals within species have distinct phenotypic differences in toxin production and pathogenicity, with some isolates reported as non-pathogenic on certain hosts. In this report, we compare genomes and gene content of six new isolates from the species complex, including the first available genomes of *F. asiaticum* and *F. meridionale*, with four other genomes reported in previous studies.

**Results:** A comparison of genome structure and gene content revealed a 93–99% overlap across all ten genomes. We identified more than 700 k base pairs (kb) of single nucleotide polymorphisms (SNPs), insertions, and deletions (indels) within common regions of the genome, which validated the species and genetic populations reported within species. We constructed a non-redundant pan gene list containing 15,297 genes from the ten genomes and among them 1827 genes or 12% were absent in at least one genome. These genes were co-localized in telomeric regions and select regions within chromosomes with a corresponding increase in SNPs and indels. Many are also predicted to encode for proteins involved in secondary metabolism and other functions associated with disease. Genes that were common between isolates contained high levels of nucleotide variation and may be pseudogenes, allelic, or under diversifying selection.

**Conclusions:** The genomic resources we have contributed will be useful for the identification of genes that contribute to the phenotypic variation and niche specialization that have been reported among members of the *F. graminearum* species complex.

**Keywords:** Fusarium Head Blight, Cereal pathogens, *Fusarium graminearum* species complex, Bioinformatics, Phylogenomics, Pan genome, Accessory genome

\* Correspondence: rajagopal.subramaniam@agr.gc.ca

<sup>2</sup>Agriculture and Agri-Food Canada, Government of Canada, 960 Carling Ave, Ottawa, Canada

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



## Background

Fungi belonging to the genus *Fusarium* are very diverse and occupy a wide variety of ecological niches. The genus is divided into several distinct species complexes, each with many species [1]. For example, the *Fusarium oxysporum* species complex is composed of a variety of species including plant and animal pathogens, decomposers and soil fungi [2–5]. The genomes of at least ten *F. oxysporum* isolates were sequenced in an attempt to delineate the genetic basis of niche specialization, and lineage specific regions within the genome were demonstrated to be important for pathogenicity in plants. For example, a supernumerary chromosome identified in *F. oxysporum* harbors virulence genes that confer pathogenicity to tomato [4].

In this study, we focused our attention on closely related individuals from the *Fusarium graminearum* species complex, which cause Fusarium Head Blight (FHB) disease of wheat. The species complex composed of several distinct species, including *F. graminearum*, *F. asiaticum*, and *F. meridionale* [6]. The genome of *F. graminearum* isolate PH-1 was the first genome available from the species complex, and has been re-examined as additional information on gene and genome structure has become available [7–10]. There have also been some large-scale genomics studies that have used this genome as a reference to study different biological phenomenon, such as genome wide association analyses of SNPs, as well as analyses of recombination events by restriction site associated DNA sequencing [11, 12]. Few studies have explored the genetic diversity of members of the *F. graminearum* species complex by performing whole genome assembly and comparative genetics; studies were largely focused on *F. graminearum sensu stricto* – a single lineage within the complex [7, 13, 14].

As part of the infection process, these fungi produce trichothecene mycotoxins that contribute to disease progression and accumulate in infected plant tissues [15]. The toxicity of *Fusarium* trichothecenes makes contaminated crops unsafe for consumption by animals [16]. The economic losses resulting from reductions in crop yield and grain quality has prompted significant amounts of research into understanding the genetic basis of toxin production by these fungi, which has been largely facilitated by the available genomic sequence of *F. graminearum* isolate PH-1 [7]. Phenotypic variations in toxin production, pathology on different hosts, development and morphology, and growth differences have been reported both among and between species of the *F. graminearum* species complex [17–19]. Currently, the only well-established virulence factor contributing to disease in wheat by *F. graminearum* and closely related species are the trichothecene mycotoxins [15, 20]. Allelic differences in genes as well as large-scale changes have been reported for the trichothecene gene cluster, with

some genes occurring outside of the main cluster in separate regions of the genome [20]. For example, allelic differences in the biosynthetic gene *Tri8* are attributed to structural variation in the trichothecene deoxynivalenol (DON), namely 3-acetyldeoxynivalenol (3-ADON) and 15-acetyldeoxynivalenol (15-ADON) [21]. Similarly, allelic variation in *Tri1* is associated with the production of the alternate trichothecene NX-2, while *Tri13* and *Tri7* are associated with production of the trichothecene nivalenol [22, 23]. Although differences in genes that contribute to trichothecene variation are known, our current knowledge of other factors that lead to variations in disease profiles remains largely unknown. Information gathered from sequences of additional members of the *F. graminearum* species complex will lead to a better understanding of the genetic variations between isolates that are associated with reported phenotypic diversity.

This study reports on genomic and genetic differences among ten closely related members of the *F. graminearum* species complex that cause FHB. Six isolates were newly sequenced as part of this study (Accessions: NRRL 6101, NRRL 28720, NRRL 28721, NRRL 28723, NRRL 28336 and DAOM 180378) and four were sequenced previously (PH-1, DAOM 233423 or GZ 3639, DAOM 241165, and CS 3005) [7, 13, 14]. NRRL 6101, isolated from barley in Japan, was reported to be non-pathogenic in wheat and was categorized as *F. asiaticum* [17, 24]. NRRL 28720, NRRL 28721, and NRRL 28723 were isolated from maize in Nepal. While NRRL 28720 belongs to the species *F. asiaticum*, NRRL 28723 is of the related species *F. meridionale*, and the species classification for NRRL 28721 is still unclear and has been suggested to be a species hybrid (*F. asiaticum* × *F. meridionale*) [6, 17, 18, 25–27]. Further investigation with additional markers suggested that this isolate may be in a basal group within the *F. asiaticum* clade or belong within *F. meridionale* [6, 18]. The isolates PH-1, DAOM 233423, DAOM 241165, DAOM 180378, NRRL 28336, and CS 3005 belong to *F. graminearum sensu stricto*. With the exception of CS 3005, which was isolated in Australia, all other isolates were isolated in North America [6, 7, 13, 14]. PH-1 and DAOM 180378 were isolated from maize, NRRL 28336, DAOM 233423, and DAOM 241165 were isolated from wheat, and CS 3005 was isolated from barley. Both the trichothecene chemotype and pathology on wheat for each isolate has been reported (Table 1) [14, 17, 19, 21, 24, 28–30].

In this study, we compared the genome sequences of these ten closely related *Fusarium* isolates and identified genes that are part of the accessory genome and potentially involved in niche specialization within and between species.

**Table 1** Description of *Fusarium* Isolates Used in this Study

Accession	Species	Disease on Wheat <sup>a, b</sup>	Trichothecene Toxin <sup>b</sup>	Origin	
				Host	Location
PH-1	<i>F. graminearum</i>	High	15-ADON, DON	Maize	USA
CS 3005	<i>F. graminearum</i>	High	15-ADON, DON	Barley	Australia
DAOM 233423 (GZ 3639)	<i>F. graminearum</i>	High	15-ADON, DON	Wheat	USA
DAOM 241165	<i>F. graminearum</i>	High	3-ADON, DON	Wheat	Canada
DAOM 180378	<i>F. graminearum</i>	High	15-ADON, DON	Maize	Canada
NRRL 28336	<i>F. graminearum</i>	Medium	3-ADON, DON	Wheat	USA
NRRL 6101	<i>F. asiaticum</i>	None	3-ADON, DON	Barley	Japan
NRRL 28720	<i>F. asiaticum</i>	Medium	DON	Maize	Nepal
NRRL 28721	<i>F. meridionale</i>	Low	4A-NIV, NIV	Maize	Nepal
NRRL 28723	<i>F. meridionale</i>	Low	NIV	Maize	Nepal

<sup>a</sup>High = more than six florets infected, Medium = three to six florets infected, Low = infection on florets adjacent to inoculation site, None = no disease beyond the inoculation site

<sup>b</sup>Toxin analyses and pathology testing were performed in culture and/or on wheat [14, 17, 19, 21, 24, 28–30]

## Methods

### *Fusarium* isolates, sequencing, and genome assembly

The isolates NRRL 28336, NRRL 28721, and NRRL 6101 were obtained from the collection at the Centraalbureau voor Schimmelcultures (CBS), in the Netherlands, while isolates NRRL 28720, NRRL 28723, and DAOM180378 were obtained from the collection at the Department of Agriculture Mycology (DAOM), in Canada. NRRL accession numbers are from the United States Department of Agriculture's Agriculture Research Service Culture Collection. Total DNA was isolated from all the NRRL *Fusarium* isolates: 6101, 28336, 28720, 28721, 28723 and the DAOM isolate 180378 by the E.Z.N.A. Fungal DNA Mini Kit (Omega). DNA was sequenced according to manufacturer instructions using 74 bp reads from paired-end libraries by Illumina GAI sequencing. The sequencing reactions were performed by the Center for the Analysis of Genome Evolution and Function (CAGEF) at the University of Toronto, Canada. Sequence reads were imported into

CLC Genomics Workbench. Prior to *de novo* assembly, reads were trimmed and short reads <20 bp were removed. Assemblies were optimized for contig length by adjusting the 'word size' parameter and keeping all other parameters at default settings. The optimal word sizes were determined to be 34, 41, 39, 38, 34, and 39 for NRRL 6101, NRRL 28720, NRRL 28723, NRRL 28336, and DAOM 180378, respectively. Average contig coverage was determined for each genome assembly and contigs that was 75% or lower than the average contig coverage and contigs shorter than 200 bp were removed from the assembly. The contigs were then reordered to the genome sequence of PH-1 from the MIPS database v3.2 by AB-ACAS [31]. The ordered contig sequences were uploaded to the GenBank database at NCBI (Table 2).

### Genomic alignments

The newly sequenced genomes, genome for PH-1 (MIPS v3.2), and sequences for CS 3005, DAOM 241165 and

**Table 2** Genome and Assembly Statistics of *Fusarium* Isolates Used in this Study

Isolate	Genome Size (Mb)	N50	Coverage (fold)	NCBI <sup>a</sup> Accession	Genome Reference <sup>b</sup>
PH-1	36.6	5.37	10	AACM	Cuomo 2007 [7]
CS 3005	36.7	0.46	40	JATU	Gardiner 2014 [13]
DAOM 233423	36.5	0.17	86	LAJZO	Walkowiak 2015 [14]
DAOM 241165	36.6	0.44	144	LAKA	Walkowiak 2015 [14]
DAOM 180378	36.4	0.16	53	LHUC	This study
NRRL 28336	36.7	0.40	36	LHUD	This study
NRRL 6101	36.5	1.21	111	LHTY	This study
NRRL 28720	36.4	0.35	36	LHTZ	This study
NRRL 28721	36.5	0.11	51	LHUA	This study
NRRL 28723	36.4	0.41	32	LHUB	This study

<sup>a</sup>National Center for Biotechnology Information

<sup>b</sup>Genomes were sequenced using Sanger, Illumina HiSeq 2000, and Illumina GAI sequencing for Cuomo 2007 [7], Gardiner 2014 [13], and Walkowiak 2015 [14], respectively

DAOM 233423 that were obtained from GenBank at NCBI were all used in separate pairwise genome alignments using the NUCmer and PROmer algorithms in the software package MUMmer [32, 33]. NUCmer is designed for aligning highly conserved regions of DNA, while PROmer aligns more divergent sequences. Since all matching and alignment routines are performed on the six frame amino acid translation of the DNA input sequence PROmer is more sensitive than NUCmer.

NUCMER was also used to extract the SNP and indel data, which was obtained by default parameters of the 1-to-1 algorithm in dnadiff [32, 33]. For each reference genome, the number of SNPs in each of the genomes was summed and averaged for each of the species to represent an average for each species. SNPs within conserved regions of the genome were obtained from PANSEQ with the following parameters: run mode = pan, fragment size = 500, percent identity cut-off = 90 and core genome threshold = 10 (i.e. sequences were present in all genomes) [34]. The resultant SNP alignment was imported into CLC Genomics Workbench for phylogenetic analyses using the maximum likelihood phylogeny tool, and the phylogeny constructed by the neighbor joining method and Jukes Cantor substitution model using 100 replicates; the bootstrap values at all nodes were 100%.

#### Construction of pan and accessory genomes

Pan genome consists of full complement of genes from all ten genomes. The list was started by including all the genes from PH-1, then, genes from other genomes were compared to the current pan genome with BLASTn and added if the e-value was  $> 1E-10$ . The software program AUGUSTUS was used to predict genes from the six newly sequenced genomes [35]. The percentage alignment to PH-1 by MUMmer was used to select the order in which genomes were selected to contribute to the pan genome list going from greatest percent alignment to lowest. Genes in the pan genome that did not originate from PH-1 were annotated by Blast-2-Go using default parameters, which identified similar sequences in databases at NCBI, gene families/domains (InterPro), and gene ontology (GO) [36]. The accessory genome was created by comparing genes from the pan genome to each genome and included if they had a BLASTn e-value  $> 1E-10$  in at least one genome.

#### Analysis of sequence variation within genes

For the analysis of sequence variation within genes, the number of positions that contained SNPs and indels from the dnadiff analysis in MUMmer were tabulated for the full length of the sequence for each gene. Orthologues of select PH-1 genes in the other genomes were obtained by BLASTn. The orthologues were aligned by MUSCLE using default parameters [37]. Genes containing a single

orthologue in each genome were examined for nonsynonymous and synonymous substitution rates by Prank with the -codon -F options [38]. We then used CodeML to fit the M0, M7, and M8 models to study the overall nonsynonymous to synonymous substitution rate ratio (dN/dS) of these genes [39]. M0 is used to estimate global dN/dS ratios, while M7 and M8 are separate models that are compared and a likelihood ratio test showing the preference of M8 constitutes evidence of adaptive evolution [39].

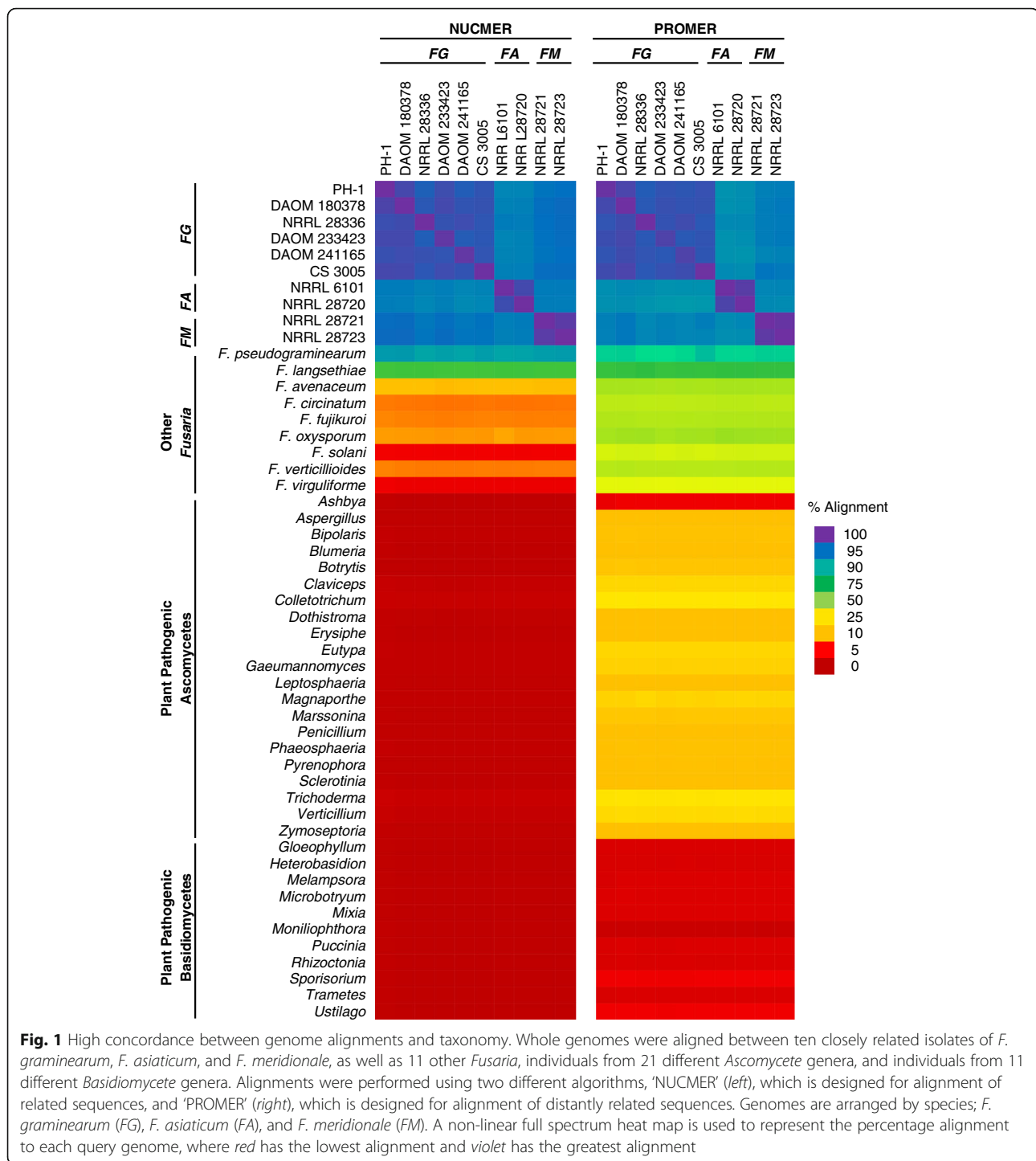
## Results

### Genome assembly information and statistics

We performed whole genome sequencing of six fungal isolates (NRRL 6101, NRRL 28336, NRRL 28720, NRRL 28721, NRRL 28723 and DAOM 180378) from the *F. graminearum* species complex that are associated with FHB disease in plants (Table 1). The newly sequenced genomes, along with four previously sequenced isolates, were selected based on their incidence in various geographical locations, taxonomy/species, host isolation, toxin production and pathology [6, 17, 18, 26, 27] (Table 1). Overall, the ten sequenced genomes were comparable in size and structure when compared to the four previously characterized chromosomes of isolate PH-1 (Table 2) [7]. The sizes of the ten genomes ranged from approximately 36.4 to 36.7 Mb. The N50 values is defined as the contig length such that using equal or longer contigs produces half the bases of the genome, which ranged from 0.11 to 1.21 Mb for the six genomes sequenced by Illumina platforms; the genome coverages ranged from 32x to 144x (Table 2).

### Whole genome alignments of ten *Fusarium* pathogens within the *F. graminearum* complex

To determine the degree of overlap among the ten genomes, we performed genome alignments of the ten genomes with one another (Table 2) and with 41 genomes of other plant pathogenic fungi using the NUCmer and PROmer algorithms in the software package MUMmer, (Fig. 1) [32]. A direct pairwise comparison with the sequence aligner NUCmer, which is designed for the alignment of closely related sequences, revealed at least 96.4% sequence alignment among the six *F. graminearum* genomes (PH-1, CS 3005, DAOM 233423, DAOM 241165, DAOM 180378, and NRRL 28336). Similarly, when the two *F. asiaticum* genomes (NRRL 6101 and NRRL 28720) were compared, 97.7% of the genomes aligned. Lastly, a comparison between the two *F. meridionale* genomes (NRRL 28721 and NRRL 28723) revealed a 99.0% alignment between them (Fig. 1). When we aligned the genomes of different species, *F. graminearum* and *F. asiaticum* had at least 93.1% sequence alignment, *F. graminearum* and *F. meridionale* had at



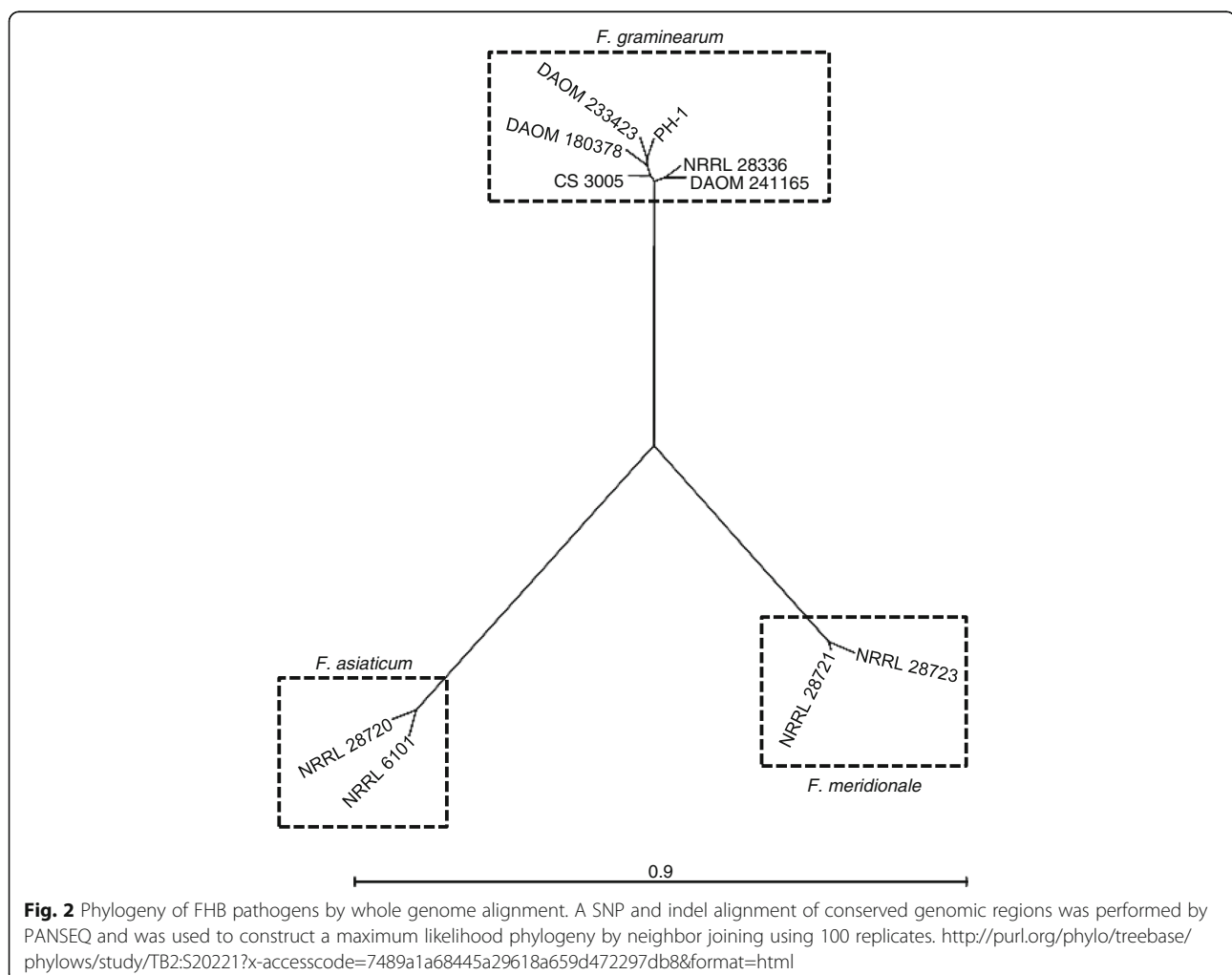
least 94.4% sequence alignment, and *F. asiaticum* and *F. meridionale* had at least 93.8% sequence alignment. The ten genomes had between 90.9 and 92.0% alignment to *F. pseudograminearum* and 64.9–65.5% alignment to *F. langsethiae*, but only 3.5–12.8% aligned to other *Fusarium* genomes, and < 1% alignment to genomes from other plant pathogenic fungi (Fig. 1).

Analysis with PROMer, which is designed for the alignment of distantly related sequences, yielded results similar to those of NUCmer in pairwise comparisons of the ten genomes to one another and to *F. pseudograminearum* and *F. langsethiae* (Fig. 1). However, the alignment to other *Fusarium* species increased from 3.5–12.8 to 39.0–50.8%. Alignment to other phytopathogenic *Ascomycetes*

increased from <1% to 3.4–26.1%, and alignment to phytopathogenic Basidiomycetes remained lower at 0.8–3.9% (Fig. 1). The concordance observed between genomes from our pairwise alignments agrees with taxonomic phylogenies suggested that *F. graminearum*, *F. asiaticum*, and *F. meridionale* are more closely related, while *F. pseudograminearum* and *F. langsethiae* are more distantly related, and that other *Fusaria* belong to even more distant taxonomic groups, followed by other genera and other phyla (Fig. 1). This agrees with the established evolutionary relationships suggested by other research studies [1, 6, 18, 40].

In addition to analyzing genetic overlap by direct pairwise alignments, we also identified SNPs and indels in all genomes, which were used to construct a maximum likelihood phylogenetic tree (Fig. 2). We used PANSEQ to identify a total of 704,566 SNPs and indels that differed among the ten genomes [34]. As observed, the ten isolates arranged according to their previously reported species assignments and were in agreement with the

pairwise alignment data (Fig. 1). Together, the data indicate that PH-1, DAOM 233423, DAOM 180378, DAOM 241165, CS 3005, and NRRL 28336 are *F. graminearum sensu stricto*, while NRRL 6101 and NRRL 29720 are *F. asiaticum* [1, 6, 18, 40] (Fig. 2). The phylogeny also supports our pairwise genome alignment data which indicated that NRRL 28721 is more closely related to NRRL 28723 and may correspond to *F. meridionale* (Fig. 1). Interestingly, the two *F. graminearum* isolates that produce the trichothecene 3-ADON (DAOM 241165 and NRRL 28336) form a distinct genetic group from the four isolates that produce 15-ADON (PH-1, DAOM 233423, DAOM 180378, and CS 3005) (Fig. 2). This supports claims that individuals from these two chemotypes (3-ADON and 15-ADON) belong to distinctly separate genetic populations [41]. The *F. graminearum* isolate CS 3005, originating from Australia grouped in the same phylogenetic group as the other 15-ADON producers from North America, implying that it may belong to the same genetic population (Fig. 2).



**Fig. 2** Phylogeny of FHB pathogens by whole genome alignment. A SNP and indel alignment of conserved genomic regions was performed by PANSEQ and was used to construct a maximum likelihood phylogeny by neighbor joining using 100 replicates. <http://purl.org/phylo/treebase/phylows/study/TB2:S20221?x-accesscode=7489a1a68445a29618a659d472297db8&format=html>

### Regions of genomic variability in FHB fungi

To visualize the genomic regions with sequence differences, the genomes were organized into 100 kb sections and the average number of SNPs and indels in the genomes of each species were plotted across the length of the four chromosomes using PH-1 as a reference (black line, Fig. 3). Generally, we observed that the average number of SNPs and indels were fewer within the species (black line) than genomes from *F. asiaticum* (red line) and *F. meridionale* (blue line) (Fig. 3). When a genome from a separate species was used as a reference, we observed that SNPs and indels were generally fewer in genomes within the same species than between genomes from different species (Additional file 1: Figure S1). Overall, we observed averages of 105–189 kb of SNPs and indels when comparing genomes from the same species and 796–893 kb of SNPs and indels when comparing genomes from different species. Sequence variation was co-localized at the ends of chromosomes and select regions within chromosomes, similar to what has been reported previously [14]. In addition, we also observed islands of increased nucleotide variation in specific genomes. For example, in the genomes of CS 3005, NRRL 28336, and NRRL 28720, we observed regions where the nucleotide variations increased compared to most other genomes (boxed, Additional file 1: Figure S1).

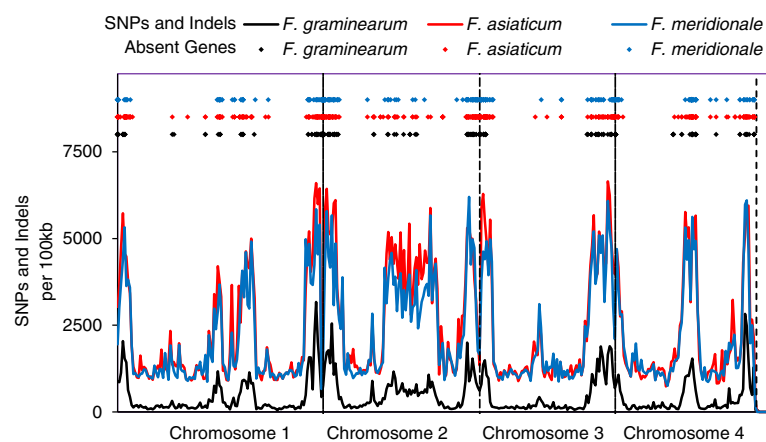
There is preponderance of evidence to suggest that niche adaptation, toxin production, and disease outcomes are attributed to differences in gene content and variation within genes [4, 14, 20–22, 42, 43]. We were interested to know if nucleotide variations were associated with genes absent among the ten genomes. AUGUSTUS was used to predict genes from the six genomes and gene sequences

for isolates PH-1, DAOM 233423, DAOM 241165 and CS 3005 were obtained from previous studies [9, 13, 14, 35]. Gene conservation was determined by BLASTn using similar approaches to other studies [14]; genes were considered absent if they had the expected value  $> 1E-10$  compared to other genomes. Our analysis showed that regardless of the reference genome, genes that are absent in *F. graminearum* (black symbols), *F. meridionale* (blue symbols), and *F. asiaticum* (red symbols) were all located in regions with high SNPs and indels (Fig. 3). We also identified genes that were unique to each isolate (purple symbols), which were also localized to regions of increased nucleotide variability (Additional file 1: Figure S1).

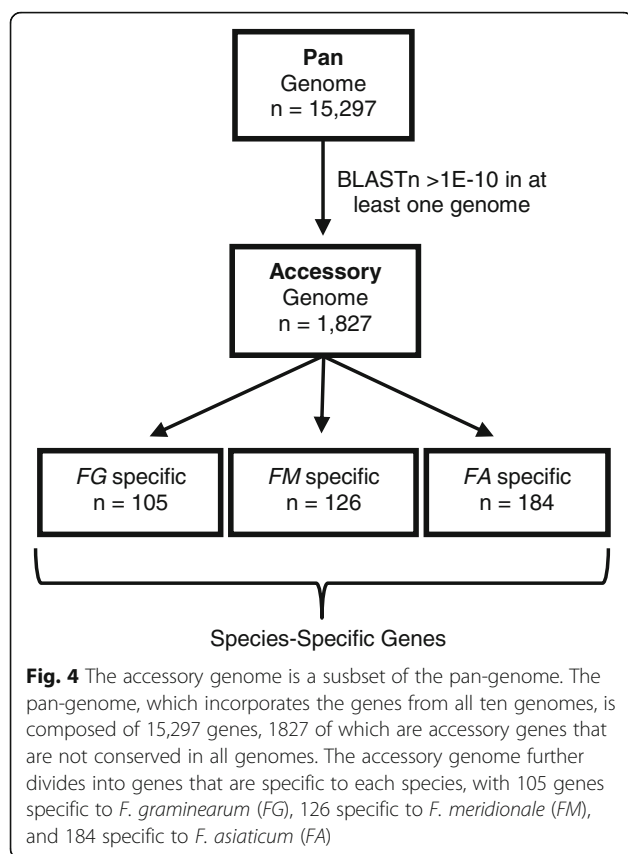
### Construction of the pan and accessory genomes of the *Fusarium* species complex

We constructed a pan genome by reiteratively appending genes of the PH-1 isolate from other genomes based on BLASTn expected values ( $e\text{-value} > 1E-10$ ). This non-redundant gene list contained a total of 15,297 genes (Fig. 4). Of the 15,297 genes in the pan-genome 12% or 1827 genes had BLASTn hits with expected values  $> 1E-10$  in at least one genome and was considered to be part of the accessory genome (Fig. 4, Additional file 2: Table S1). The accessory genome contributes to the genetic variability within and among species and might encode biochemical functions that are not essential for *Fusarium* growth, but confer selective advantages, such as adaptation to different niches or colonization of a new host [34].

The accessory genome was further parsed into genes that are unique to a species or to each isolate (Additional file 2: Table S2). Analyses identified a total of 105 species-



**Fig. 3** Genome sequence and gene content variability in PH-1 compared to other isolates. SNPs/indels and absent genes between genomes were determined separately, by whole genome alignments and BLASTn using PH-1 as a reference. The number of SNPs and indels per 100 kb were averaged for each species and are presented separately for *F. graminearum* (black lines), *F. asiaticum* (red lines), and *F. meridionale* (blue lines). Positions of absent genes whose sequences had a lowest expected value  $> 1E-10$  by BLASTn for at least one member of the species are presented as sySNPsols; *F. graminearum* (black diamonds), *F. asiaticum* (red diamonds), and *F. meridionale* (blue diamonds). Chromosomes are demarked by vertical dashed lines



specific genes that were present in all six *F. graminearum* isolates, but were absent in all genomes of the other two species (Fig. 4). Similarly, a total of 126 and 184 species-specific genes were found in *F. meridionale* and *F. asiaticum*, respectively (Fig. 4). Analysis of the accessory genome also revealed genes that are present in a single isolate, but absent in others (Uniquely Present, Table 3, Additional file 2: Table S2). We observed between 12 and 92 unique genes in each of the ten isolates (Uniquely Present, Table 3, Additional file 2: Table S2). In contrast, the *F. asiaticum* isolate NRRL 6101 had the most number of “uniquely absent” genes (34) and this isolate has been reported to be unable to infect wheat [17].

#### The accessory genome reveals species and isolate-specific genes potentially involved in secondary metabolism and disease

Functional enrichment analyses of 105 *F. graminearum*-specific genes, 90 of which were originally identified in PH-1 by the MIPS FunCat database revealed 72 genes or 80% as unclassified. The remaining 18 genes were enriched in metabolic and defense pathways (Additional file 2: Table S3). The genes involved in defense included two aldehyde dehydrogenases, *FGSG\_17538* and *FGSG\_01759*, as well as *FGSG\_17130* encoding a protein similar to pisatin demethylase cytochrome P450. Pisatin is a

**Table 3** Isolate-Specific Genes Identified from Accessory Genome

Isolate	Uniquely Present Genes <sup>a</sup>	Uniquely Absent Genes <sup>b</sup>
PH-1	14	13
DAOM 180378	18	8
NRRL 28336	92	2
DAOM 233423	18	10
DAOM 241165	12	1
CS 3005	14	6
NRRL 6101	65	34
NRRL 28720	72	29
NRRL 28721	18	2
NRRL 28723	22	7

<sup>a</sup>Uniquely present genes were considered present in the genome of the isolate and absent in genomes of all other isolates; absence was determined by having a BLASTn hit with the lowest expected value > 1E-10 in the genome of the isolate

<sup>b</sup>Uniquely absent genes were considered absent in the genomes of the isolate and present in the genomes of all other isolates; absence was determined by having a BLASTn hit with the lowest expected value > 1E-10 in the genome of the isolate

phytoalexin produced by plants as a defense molecule; detoxification of pisatin has been demonstrated by the action of demethylase genes from *F. solani* [44]. Analysis of 184 genes specific to *F. asiaticum* included secondary metabolic genes such as a non-ribosomal peptide synthetase (NRPS) APS1, which is involved in apicidin biosynthesis, and a polyketide synthase (PKS) PKS40 associated with W493 biosynthesis (Table 4) [42, 45, 46]. There were also 126 genes specific to *F. meridionale* and a gene (*g3644*) is annotated as a ‘pectin lyase fold virulence factor’ implicated in cell wall catabolism (Additional file 2: Table S4) [47]. Twelve heterokaryon or vegetative incompatibility genes were also identified as being species-specific; these gene classes are known to diversify in fungi and may have roles in speciation [48, 49].

Our functional analysis was extended to isolate-specific genes. Of the 1827 genes from the accessory genome, 12–92 genes were present in a single isolate and were considered unique to that isolate (Table 3). The function of unique genes spanned a wide range of categories, including primary, secondary metabolism, as well as signal transduction. Examples include a glutamine synthase (*g9634*) from NRRL 28336 and a non-ribosomal peptide synthase (*g11745*) from NRRL 6101 (Table 4; Additional file 2: Tables S2 and S4). Among the 92 genes unique to NRRL 28336, nine are predicted to have kinase functions and could be involved in signal transduction (Additional file 2: Tables S2 and S4). A kinase unique to PH-1, *FGSG\_11614*, has been disrupted and resulted in reduced DON production [50].



**Table 4** Secondary Metabolic Genes are Part of Accessory Genome

Gene Name	Source	Possible Function	Alternate Name(s)	Conservation <sup>a</sup>
Polyketide metabolism				
<i>FGSG_04694</i>	PH-1	polyketide synthase	PKS2	FG, FM
<i>FGSG_06540</i>	PH-1	polyketide cyclase		FG, FA
<i>g11843</i>	NRRL 28723	chalcone synthase b		FM, FA
<i>g6531</i>	NRRL 6101	polyketide enoyl_reductase		FA
<i>g1989</i>	NRRL 6101	polyketide synthase	PKS40	FA
<i>g21</i>	NRRL 28723	polyketide cyclase dehydrase protein		FM
<i>FG05_30424</i>	CS 3005	polyketide synthase	PKS52	FG, FA
<i>FG05_30565</i>	CS 3005	snoal-like polyketide cyclase family protein		FG, FM
<i>FG05_30492</i>	CS 3005	polyketide synthase		FG, FM, FA
<i>FG05_30491</i>	CS 3005	polyketide synthase		FG, FM, FA
<i>FG05_30490</i>	CS 3005	polyketide synthase		FG, FM, FA
<i>FG05_30489</i>	CS 3005	polyketide synthase	PKS43	FG, FM, FA
Non-ribosomal peptide metabolism				
<i>FGSG_10702</i>	PH-1	non-ribosomal peptide synthetase	NRPS17	FG
<i>g11745</i>	NRRL 6101	non-ribosomal peptide synthetase		FA
<i>g6644</i>	NRRL 6101	non-ribosomal peptide synthetase	APS1, NRPS31	FA
Terpene metabolism				
<i>FGSG_08181</i>	PH-1	terpene synthase		FG, FA
<i>g8968</i>	NRRL 28721	Tri5-like terpene synthase		FM, FA

<sup>a</sup>Gene was present in the genome of at least one isolate of *F. graminearum* (FG), *F. asiaticum* (FA), or *F. meridionale* (FM)

Genes that were uniquely absent in an isolate were also analysed (uniquely absent, Table 3). The isolate NRRL 6101 was reported to have limited virulence in wheat and contained the most number of uniquely absent genes (34, Table 3). These included two genes categorized by MIPS to be involved in cell rescue, defense and virulence, namely *FGSG\_16080* predicted to encode a chitinase, and *FGSG\_00132* with potential kinase function. The kinase *FGSG\_00132* was previously target by reverse genetics, but no phenotype was observed in the mutant [50]. Another gene, *FGSG\_08005* that was absent in the NRRL 28720 genome was predicted to be involved in isoprenoid metabolism, potentially contributing to secondary metabolic pathways (Additional file 2: Table S4).

Our analysis also showed that some of the genes in the accessory genome had irregular patterns of gene conservation across the ten genomes; many of these genes were associated with secondary metabolism (Table 4). Examples include *FGSG\_04694*, a gene encoding polyketide synthase (PKS2), which was absent in both *F. asiaticum* genomes but was present in all *F. graminearum* and *F. meridionale* isolates (Table 4). Mutation of *PKS2* in *F. graminearum* showed reduced mycelial growth and a defect in virulence [51]. A polyketide cyclase, *FGSG\_06540* showed poor conservation in both *F. meridionale* isolates, but was present in *F. graminearum* and *F. asiaticum* isolates [43]. Another example includes gene *g11843* from NRRL 28723 (*F.*

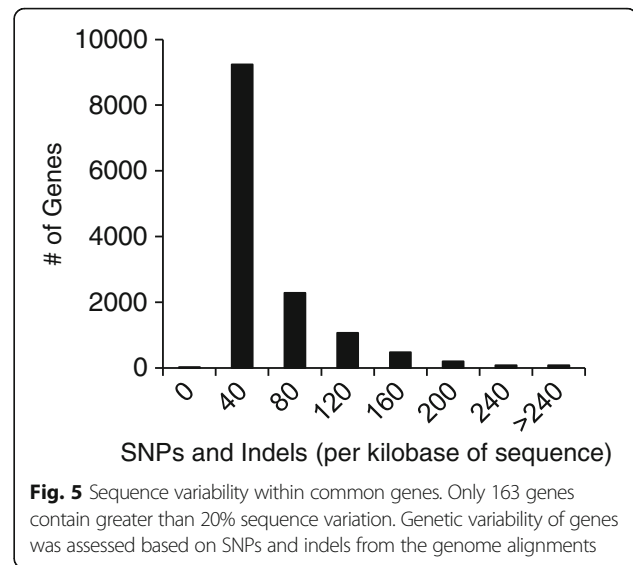
*meridionale*), which shares similarity with *FAVG1\_07768*, a gene from *F. avenaceum* that is present in the genomes of both *F. meridionale* isolates and the *F. asiaticum* isolate NRRL 6101, but is absent in all *F. graminearum* isolates. This gene was predicted to encode a chalcone synthase, which has been characterized in plants and is involved in the biosynthesis of defensive flavonoid compounds [52]. In addition to polyketide synthases, a metabolic gene *g8968* from NRRL 28721 (*F. meridionale*) shares identity with a terpene synthase gene, *Tri5*, which catalyzes the first step of trichothecene biosynthesis in *F. graminearum* and is essential for trichothecene biosynthesis and infection of wheat [15]. Interestingly, *g8968* did not have a positive BLASTn hit to the other *F. meridionale* genome NRRL 28723 but had positive BLASTn hits to the *F. asiaticum* genomes NRRL 6101 and NRRL 28720. Upon closer examination, *g8968* showed only 21.8% protein identity to *Tri5* from the *F. graminearum* isolate PH-1, but 95% protein identity to *FAVG1\_13210* from *F. avenaceum* [53]. A true homologue of *Tri5* (g5480) with >97% protein identity was identified in the *F. meridionale* isolate NRRL 28721. Despite the low identity to *Tri5*, *g8968* was predicted to contain a *Tri5* domain by Interpro and Pfam (e-value 5.3E-15). It is unclear if the protein from this gene acts on substrates from the trichothecene biosynthetic pathway and contributes to their production, or if the protein is involved in the production of other molecules.

In addition to individual genes involved in secondary metabolism, accessory genes with irregular patterns of conservation also extended to gene clusters. For example, a gene cluster composed of five genes was absent in the *F. meridionale* genomes, but is present in *F. graminearum* and *F. asiaticum* isolates (Table 4) [42, 43]. The genes within this cluster include *FGSG\_08181*, which is predicted to encode a terpene synthase, *FGSG\_08182*, which encodes a putative transcription factor and three putative cytochrome P450 genes *FGSG\_17088*, *FGSG\_08183*, and *FGSG\_08187* [42, 43]. Genes from this cluster have been reported to be expressed during infection in different cereal hosts [54]. Similarly, another gene cluster (*FG05\_30489* - *FG05\_30492*) from the *F. graminearum* isolate CS 3005 from Australia was absent in the 15-ADON producing isolates of *F. graminearum* from North America, but was present in all other genomes (Table 4). The gene *FG05\_30489* with similarity to a polyketide synthase gene *PKS43* was also present in the *F. avenaceum* and *F. equiseti* genomes [42].

In addition to the genes potentially involved in secondary metabolism, there were other genes from the accessory genome with irregular conservation patterns that have been shown to contribute to fecundity and disease by the fungi (Table 4). A study that targeted transcription factors determined that accessory genes *FGSG\_08182* and *FGSG\_11561* contributed to disease and spore production, respectively [55]. A separate study that focused on kinases determined that the accessory genes *FGSG\_07812*, *FGSG\_11614*, and *FGSG\_02153* all contributed to toxin production while *FGSG\_03146* contributed to spore production and disease [50]. Disruption of a possible phosphatase regulatory subunit *FGSG\_11641* has also been shown to reduce disease accumulation [56]. Together, the accessory genome appears to contain genes that contribute to metabolic processes, as well as other processes potentially involved in development or disease.

#### Highly polymorphic genes in common genome include alleles, pseudogenes and genes under diversifying selection

Genes not considered part of the accessory genome was identified to contain high levels of sequence variation. We identified a total of 163 genes that were highly variable with >200 SNPs or indel sites per kb (Fig. 5, Additional file 2: Table S5). Since 110 of the 163 genes originated from PH-1, the MIPS FunCat database (<http://mips.helmholtz-muenchen.de/funecatDB/>) was used to detect functional enrichment of gene categories (Additional file 2: Table S6). The analysis revealed that the majority of the genes (73.6%) are functionally unclassified, compared to 52.3% for all genes from PH-1. Functional categories that were determined to be enriched compared to the rest of the genome ( $p$ -value < 0.05) included motility, energy,



metabolism, and cell migration (Additional file 2: Table S6). Select genes that exhibited high frequency of SNPs and indels were inspected in greater detail, these included *Tri8*, three putative transcription factors, two proteins with nodulin-like domains and three putative heterokaryon incompatibility genes.

Allelic variation in *Tri8* is attributed to the production of either 15-ADON or 3-ADON [21]. Multiple sequence alignment by MUSCLE indicated that within a specific chemotype, such as 3-ADON producing isolates, *Tri8* had greater than 95% sequence identity in the coding regions. Comparison of the coding sequences between chemotypes showed only 84–89% sequence identity. Although, no acetylated form of DON has been shown to be produced by NRRL 28720 isolate, *Tri8* from this isolate showed 97% identity to other 15-ADON producing isolates [17]. Thus, high alignment of *Tri8* suggests that the high variation in this gene may be associated with allelic variation within chemotypes.

Three genes *FGSG\_13457*, *FGSG\_08954*, and *FGSG\_10508* that encode transcription factors have been previously targeted for functional analysis [55]. Disruption of *FGSG\_13457* resulted in no apparent phenotypic difference in toxin production or disease, while no mutants were obtained for *FGSG\_08954* and *FGSG\_10508* [55]. Closer inspection of *FGSG\_08954* in the ten genomes revealed that the gene may have a small deletion and a premature stop codon in NRRL 28720 and duplicate copies in DAOM 241165. The two other transcription factors *FGSG\_13457* and *FGSG\_10508* were subjected to diversifying selection analysis by CodeML to assess overall nonsynonymous to synonymous substitution ( $dN/dS$ ) rate of these genes. With the M0 model, the  $dN/dS$  ratios for *FGSG\_13457* and *FGSG\_10508* were estimated to be 0.221 and 0.408, respectively, and application of the two site

model (M7 and M8) did not provide evidence of diversifying selection (Additional file 2: Table S7).

Two genes encoding putative nodulin-like domains, *FGSG\_03381* and *FGSG\_03550*, were also found to have high sequence variation; nodulin-like domains are often found in plant proteins and are implicated in plant-microbe interaction [57]. The orthologue of *FGSG\_03550* spanned a region with ambiguous nucleotides for four of the genomes or was at the very end of a contig, the gene also contained a 24 bp insertion in the two *F. asiaticum* genomes. The *dN/dS* ratio of *FGSG\_03381* was 0.355, and the two site model indicated no evidence for adaptive evolution (Additional file 2: Table S7).

Lastly, three putative heterokaryon incompatibility genes *FGSG\_10601*, *FGSG\_08120* and *FGSG\_08144* with potential roles in non-self-recognition were identified to have high sequence variation [48, 49]. *FGSG\_08120* was determined to have a premature stop in NRRL 6101, while the closest orthologue of *FGSG\_08144* in NRRL 6101 had an identity <50 to the orthologues. Analysis of diversifying selection for *FGSG\_10601* showed a *dN/dS* ratio of 0.708 and was the only gene from the four investigated that showed significant evidence of adaptive evolution (with a *p*-value < 0.01) by the two site model (Additional file 2: Table S7).

Altogether, the increased sequence variation in select common genes may be associated with allelic variation (*Tri8*) or mutations resulting in premature stop codons resulting in a pseudogene (*FGSG\_08120* and *FGSG\_08954*). In addition, *FGSG\_10601* was determined to be under adaptive evolutionary pressures that may have contributed to diversification of this gene.

## Discussion

*Fusarium* Head Blight fungi are globally distributed and isolates are known to exhibit phenotypic diversity in their morphology, growth and development, production of toxins and other secondary metabolites, and pathogenicity towards specific hosts [17]. DNA-based phylogenetic analyses using gene markers have allowed us to identify distinct species and genetic populations of these fungi. However, some isolates have proven difficult to classify using marker analyses and the biological mechanisms underlying phenotypic diversity remain unsettled [6, 17, 18, 25–27]. Genome sequencing of additional isolates provided the opportunity to better understand and confirm the genetic relationship between fungal isolates, and identified gene candidates that could contribute to phenotypic differences and niche adaptation. Similar approaches were successfully applied to the *F. oxysporum* species complex; for example, genomic comparison of *F. oxysporum* isolates identified gene expansion that may be associated with pathogenicity towards banana, as well as genes associated with infection in melon [2, 5].

A comparative genomics approach used in this study was effective at better resolving the genetic relationship among fungal species and isolates. In pairwise alignments, the genomes of the ten closely related *Fusaria* exhibited high levels of identity, particularly within species. The genomes from isolates of the *F. graminearum* species complex also had greater alignment to *F. pseudograminearum* and *F. langsethiae*, than to genomes from other *Fusaria*, or other genera altogether. These findings confirm the previously established phylogenetic relationships among these closely related *Fusarium* to each other and other fungi [4, 17, 18]. Interestingly, on average, *F. meridionale* shared more genes and showed high levels of identity than *F. asiaticum* to *F. graminearum*. Based on these findings, *F. meridionale* and *F. graminearum* may be more closely related than *F. asiaticum* and *F. graminearum*. In addition to providing insights into the relationships between species, we were also able to better resolve species placement of NRRL 28721, which was difficult to determine using marker analyses. Marker analyses had suggested that NRRL 28721 is either *F. asiaticum* or *F. meridionale* or a species hybrid [6, 17, 18, 25–27]. Analyses of NRRL 28723 and NRRL 28721 genomes showed high level of identity (99.0%). In combination of with SNPs and indels analyses, we suggest that NRRL 28721 and NRRL 28723 isolates are closely related and may be of the same species, *F. meridionale* (Fig. 2). Similar analyses also enabled us to separate 3-ADON or 15-ADON genetic populations within *F. graminearum* species complex that have been reported by others [41].

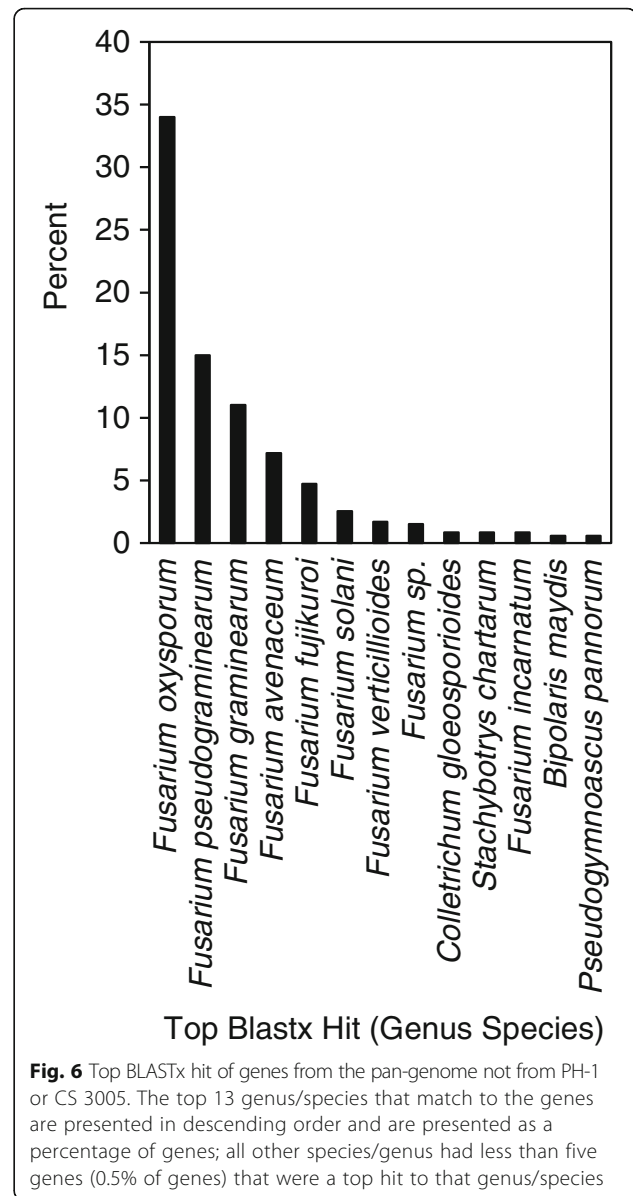
This study provided insights into the relationships between isolates with known differences in geographical distribution, species and population backgrounds. For example, CS 3005, an isolate from Australia grouped within the same population as the 15-ADON producers in North America. This suggested that this Australian isolate may be more related to the 15-ADON producers rather than the 3-ADON producers from North America. Furthermore, the study suggests that genetic populations are globally distributed, and that geographical barriers may not play a significant role in the spread of these fungi. The presence of different species such as *F. asiaticum* and *F. meridionale* in Nepal, or genetic populations of the same species, such as 3-ADON and 15-ADON populations of *F. graminearum* in North America that cause similar diseases indicates that different populations and species can coexist and may have distinguishable traits that allow them to occupy slightly different niches than other populations or species. For example, environmental and host conditions have been demonstrated to affect disease outcomes by 3-ADON and 15-ADON producers of *F. graminearum* [58]. Further monitoring will better resolve the geographical distribution of these populations/species and the stability of

the population and community structure over time. This may be particularly important when we consider the environmental consequences of climate change.

In addition to providing insights into the genetic relationship between isolates, our comparative genomics analyses also provided perspective into the conservation and evolution of genes that may contribute to phenotypic differences among these pathogens. We observed a high degree of conservation (95 to 99%) of genes among the ten genomes; highest within species than between species (Additional file 3: Figure S2). Analyses of the pan genome indicated that the two 3-ADON producing strains of *F. graminearum* (NRRL 28336 and DAOM 241165) had the fewest number of absent genes (914 and 990, respectively). In contrast, the two *F. asiaticum* isolates NRRLs 6101 and 28720 had the greatest number of absent genes, 1126 and 1146, respectively, when compared to the pan genome (Additional file 2: Table S2). Further study across additional isolates would be required to determine if there is gene expansion or loss in populations or species of these fungi.

Many of the genes from the accessory genome were predicted to be involved in metabolism. This is particularly important because *Fusarium* species are well known for their ability to make a wide array of secondary metabolites, including DON, which has been studied for its role as a virulence factor during plant infection and for its toxic effects in humans [15]. Based on the sequence, *F. graminearum* isolate PH-1 is reported to have at least 16 known PKS genes and 19 known NRPS and other metabolic genes [42, 43]. We discovered that some of these genes, such as *PKS2*, *PKS52* and *NRPS17* that have described previously in *F. graminearum* are absent in some of the genomes [42]. We also identified genes in some of the ten genomes, such as *NRPS31*, that were previously not reported in the *F. graminearum* species complex (Table 4) [42]. These analyses underscore the potential importance of secondary metabolites in host and niche adaptation and future studies should focus on the discovery of metabolites produced by these enzymes.

Genes to be species or isolate specific were annotated by Blast-2-Go based on sequence similarity to genes from other *Fusaria* (Fig. 6). A total of 1061 gene sequences were annotated from the new genomes (excluding CS 3005 and PH-1) and 361 or 34% of these had a top tBLASTx hit in *F. oxysporum*, while the next top hits were in *F. pseudograminearum* (15%), *F. graminearum* (11%), *F. avenaceum* (7%), *F. fujikuroi* (5%), *F. solani* (3%), and *F. verticillioides* (2%) (Fig. 6). All other genus/species that had more than 0.5% of genes BLASTx hits were from other *Fusaria* or other Ascomycete genera, including *Colletotrichum*, *Stachybotrys*, *Bipolaris*, and *Pseudogymnoascus*. This indicates that many genes from the accessory genome are conserved across the genus,



**Fig. 6** Top BLASTx hit of genes from the pan-genome not from PH-1 or CS 3005. The top 13 genus/species that match to the genes are presented in descending order and are presented as a percentage of genes; all other species/genus had less than five genes (0.5% of genes) that were a top hit to that genus/species

but may be lost or acquired, depending on selective pressures from the environment or plant host. This has been previously described as an attribute of pathogen effectors that are involved in the evolutionary interactions between pathogen and host.

Such conservation of accessory genes within the genus also suggest that horizontal gene transfer may play a significant role in the biology of *Fusarium*. This is supported by our observations of genes that were absent across isolates and were co-localized in regions of the genome with increased nucleotide variability (Fig. 3). The co-localization of divergent genetic regions and genes that are absent among related genomes is consistent with what has been reported previously for *Fusarium* species [4, 14, 59]. Co-localization may suggest that genes were acquired or lost as

clusters; multiple gains or losses of gene clusters have been suggested for the secondary metabolite gene cluster responsible for fumonisin biosynthesis in *Fusarium* [60]. Lineage specific chromosomes and large scale horizontal gene transfer events, such as the transfer of an entire chromosome that have been reported in *F. oxysporum*, are yet to be reported for members of the *F. graminearum* species complex [4]. However, small scale horizontal transfer events have been suggested between *F. graminearum* and other fungi [43]. Horizontal gene transfer has also been suggested between bacteria and *F. pseudograminearum*, which is also within the *graminearum* species complex and is reported to cause disease on the roots of cereals [59]. Closer inspection of the presence, location, and orientation within the genomes of other organisms could shed light on their history and origin.

In addition to accessory genes contributing to genetic differences amongst the genomes, we also identified common genes that are variable and could contribute to niche adaptation and disease. Common genes involved in plant-pathogen interaction that are recognized by the host are selected against but in some cases cannot be shed; instead, these genes modify their sequence in order to avoid recognition by the host [61]. Our analyses identified 163 common genes with more than 20% sequence variability present among genomes. This included *Tri8*, which has reported allelic variation and contributes to the structural variation that exists between 3-ADON and 15-ADON toxins [21]. Other genes, such as *FGSG\_08120* and *FGSG\_08954* may have acquired mutations that resulted in premature stop codons, resulting in pseudogenes. In addition, analysis of synonymous and nonsynonymous mutation rates of *FGSG\_10601* indicated that this gene may be under diversifying selection. A comparative analysis of three *Fusarium* species identified genes from *F. graminearum* that may be under diversifying selection [62]. These included *FGSG\_03550*, *FGSG\_03859*, *FGSG\_05819*, *FGSG\_10508*, *FGSG\_11067*, *FGSG\_12487*, *FGSG\_13517*, *FGSG\_04520*, *FGSG\_10636*, and *FGSG\_15331*, which were also identified to have >20% sequence variation between genomes in our study [62]. In contrast, our analyses of *FGSG\_10508* among the *F. graminearum* species complex did not find significant evidence for diversifying selection, possibly due to the shorter evolutionary distance between the genomes we investigated. Nevertheless, further investigation of these highly variable common genes as well as the accessory genes could implicate these genes with phenotypic differences or known adaptive processes.

## Conclusions

Comparative genomics has enabled us to place ten isolates from the *F. graminearum* species complex into their respective species or genetic populations. A construction of a pan and accessory genome has also given us insights into

the genetic diversity among members of the *F. graminearum* species complex. Further investigation of the roles of both accessory genes and highly variable common genes with respect to niche adaptation is required. As more genomes of this species complex become available, a larger comparative study will allow for further characterization of the genetic diversity amongst these economically important pathogenic fungi.

## Additional files

**Additional file 1:** Sequence and gene content variability between genomes. Description: SNPs/indels and absent genes between genomes were determined separately, by whole genome alignments and BLASTn, using each genome as a reference. a, *F. graminearum* genomes were used as a reference. b, *F. meridionale* genomes are used as a reference. c, *F. asiaticum* genomes are used as a reference. The number of SNPs and indels per 100 kb between the reference genome and the average for each species is presented separately for *F. graminearum* (black lines), *F. asiaticum* (red lines), and *F. meridionale* (blue lines). The minimum and maximum for each plot is 0 and 10,000 nucleotide positions. Positions of absent genes whose sequences had a lowest expected value > 1E-10 by BLASTn for at least one member of the other species are presented as symbols; *F. graminearum* (black diamonds), *F. asiaticum* (red diamonds), and *F. meridionale* (blue diamonds). Absent genes in all other members of all species were considered unique genes (purple squares). Some regions of SNPs/indels that are greater in a specific reference genome than in other reference genomes are highlighted (dotted box). (DOCX 635 kb)

**Additional file 2:** Summary tables for gene conservation, variation, annotation, and functional enrichment analyses. Description: **Table S1**, the accessory genome is a subset of the pan genome. **Table S2**, components of the accessory genome that are species and isolate specific. **Table S3**, functional categories of *F. graminearum* species-specific accessory genes. **Table S4**, genes from the pan genome not from PH-1 were annotated by Blast-2-Go to be involved in a wide array of processes. **Table S5**, common genes with >20% sequence variation. **Table S6**, common genes that are variable are mostly of unknown function. **Table S7**, parameter values and log likelihood scores of codon substitution models. (XLSX 2162 kb)

**Additional file 3:** Percent gene overlap between genomes. Description: The percentage of genes from the query genome in the reference genome are presented in a greyscale heatmap, where white represents 95% conservation and black represents 100% conservation. Genes were considered absent if they had a lowest BLASTn e-value > 1E-10. Genomes are arranged by species; *F. graminearum* (FG), *F. asiaticum* (FA), and *F. meridionale* (FM). (DOCX 28 kb)

## Abbreviations

15-ADON: 15-acetyldeoxynivalenol; 3-ADON: 3-acetyldeoxynivalenol; CAGEF: Center for the Analysis of Genome Evolution and Function; CBS: Centraalbureau voor Schimmelcultures; DAOM: Department of Agriculture Mycology; DON: Deoxynivalenol; FA: *Fusarium asiaticum*; FG: *Fusarium graminearum*; FHB: Fusarium Head Blight; FM: *Fusarium meridionale*; GO: Gene ontology; Indel: Insertion and/or deletion; kb: Kilo base pairs; Mb: Mega base pairs; NRPS: Non-ribosomal peptide synthase; ORDC: Ottawa Research and Development Centre; PKS: Polyketide synthase; SNP: Single nucleotide polymorphism

## Acknowledgements

The authors would like to acknowledge Li Wang (Agriculture and Agri-Food Canada, Ottawa Research and Development Centre (ORDC, Ottawa) for technical assistance.

## Funding

This work was supported by funds from Agriculture and Agri-Food Canada and the National Sciences and Engineering Research Council of Canada.

### Availability of data and materials

Isolate	NCBI Accession
PH-1	AACM
CS 3005	JATU
DAOM 233423	LAJZO
DAOM 241165	LAKA
DAOM 180378	LHUC
NRRL 28336	LHUD
NRRL 6101	LHTY
NRRL 28720	LHTZ
NRRL 28721	LHUA
NRRL 28723	LHUB

<http://purl.org/phylo/treebase/phyloids/study/TB2:S20221?x-accesscode=7489a1a68445a29618a659d472297db8&format=html>

### Authors' contributions

SW, OR and RS designed the project and wrote the manuscript. SW carried out data analysis and data interpretation. Adaptive evolution analyses were conducted by NR. All authors have read and approved the final version of the manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Consent for publication

Not applicable.

### Ethics approval and consent to participate

We declare that all relevant licenses and permissions were obtained for using the strains used in this study.

### Author details

<sup>1</sup>Department of Biology, Carleton University, 1125 Colonel By Dr, Ottawa, Canada. <sup>2</sup>Agriculture and Agri-Food Canada, Government of Canada, 960 Carling Ave, Ottawa, Canada.

Received: 17 June 2016 Accepted: 2 December 2016

Published online: 09 December 2016

### References

- Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MK, Brandt ME, et al. One fungus, one name: defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology*. 2013;103:400–8.
- Guo L, Han L, Yang L, Zeng H, Fan D, Zhu Y, et al. Genome and transcriptome analysis of the fungal pathogen *Fusarium oxysporum* f. sp. cubense causing banana vascular wilt disease. *PLoS ONE*. 2014;9:e95543.
- Ma L-J, Shea T, Young S, Zeng Q, Kistler HC. Genome Sequence of *Fusarium oxysporum* f. sp. melonis Strain NRRL 26406, a Fungus Causing Wilt Disease on Melon. *Genome Announc*. 2014; doi:10.1128/genomeA.00730-14.
- Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. 2010;464:367–73.
- Schmidt SM, Lukasiewicz J, Farrer R, van Dam P, Bertoldo C, Rep M. Comparative genomics of *Fusarium oxysporum* f. sp. melonis reveals the secreted protein recognized by the For-2 resistance gene in melon. *New Phytol*. 2016;209:307–18.
- O'Donnell K, Ward TJ, Geiser DM, Corby Kistler H, Aoki T. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. *Fungal Genet Biol*. 2004;41:600–23.
- Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*. 2007;317:1400–2.
- King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-Kosack KE. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics*. 2015;16:544.
- Wong P, Walter M, Lee W, Mannhaupt G, Münsterkötter M, Mewes H-W, et al. FGDB: revisiting the genome annotation of the plant pathogen *Fusarium graminearum*. *Nucleic Acids Res*. 2011;39:D637–9.
- Zhao C, Waalwijk C, de Wit PJGM, Tang D, van der Lee T. RNA-Seq analysis reveals new gene models and alternative splicing in the fungal pathogen *Fusarium graminearum*. *BMC Genomics*. 2013;14:21.
- Talas F, Kalih R, Miedaner T, McDonald BA. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of *Fusarium graminearum*. *Mol Plant Microbe Interact*. 2016;29:417–30.
- Talas F, McDonald BA. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genomics*. 2015;16:996.
- Gardiner DM, Stiller J, Kazan K. Genome sequence of *Fusarium graminearum* isolate CS3005. *Genome Announc*. 2014. doi:10.1128/genomeA.00227-14.
- Walkowiak S, Bonner CT, Wang L, Blackwell B, Rowland O, Subramaniam R. Intraspecific interaction of *Fusarium graminearum* contributes to reduced toxin production and virulence. *Mol Plant Microbe Interact*. 2015;28:1256–67.
- Jansen C, von Wettstein D, Schäfer W, Kogel K-H, Felk A, Maier FJ. Infection patterns in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted *Fusarium graminearum*. *Proc Natl Acad Sci U S A*. 2005;102:16892–7.
- Sobrova P, Adam V, Vasatkova A, Beklova M, Zeman L, Kizek R. Deoxynivalenol and its toxicity. *Interdiscip Toxicol*. 2010;3:94–9.
- O'Donnell K, Kistler HC, Tacke BK, Casper HH. Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. *Proc Natl Acad Sci U S A*. 2000;97:7905–10.
- Starkey DE, Ward TJ, Aoki T, Gale LR, Kistler HC, Geiser DM, et al. Global molecular surveillance reveals novel *Fusarium* head blight species and trichothecene toxin diversity. *Fungal Genet Biol*. 2007;44:1191–204.
- Ward TJ, Clear RM, Rooney AP, O'Donnell K, Gaba D, Patrick S, et al. An adaptive evolutionary shift in *Fusarium* head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. *Fungal Genet Biol*. 2008;45:473–84.
- 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;74:1128–42.
- Alexander NJ, McCormick SP, Waalwijk C, van der Lee T, Proctor RH. The genetic basis for 3-ADON and 15-ADON trichothecene chemotypes in *Fusarium*. *Fungal Genet Biol*. 2011;48:485–95.
- Lee T, Han Y-K, Kim K-H, Yun S-H, Lee Y-W. *Tri13* and *Tri7* determine deoxynivalenol- and nivalenol- producing chemotypes of *Gibberella zeae*. *Appl Environ Microbiol*. 2002;68:2148–54.
- Varga E, Wiesenberger G, Hametner C, Ward TJ, Dong Y, Schöfbeck D, et al. New tricks of an old enemy: isolates of *Fusarium graminearum* produce a type A trichothecene mycotoxin. *Environ Microbiol*. 2015;17:2588–600.
- Vesonder RF, Ellis JJ, Kwolek WF, DeMarini DJ. Production of vomitoxin on corn by *Fusarium graminearum* NRRL 5883 and *Fusarium roseum* NRRL 6101. *Appl Environ Microbiol*. 1982;43:967–70.
- Carter JP, Rezanoor HN, Holden D, Desjardins AE, Plattner RD, Nicholson P. Variation in pathogenicity associated with the genetic diversity of *Fusarium graminearum*. *Eur J Plant Pathol*. 2002;108:573–83.
- Ward TJ, Bielawski JP, Kistler HC, Sullivan E, O'Donnell K. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. *Proc Natl Acad Sci U S A*. 2002;99:9278–83.
- Yang L, van der Lee T, Yang X, Yu D, Waalwijk C. *Fusarium* populations on Chinese barley show a dramatic gradient in mycotoxin profiles. *Phytopathology*. 2008;98:719–27.
- Gardiner DM, Kazan K, Manners JM. Novel genes of *Fusarium graminearum* that negatively regulate deoxynivalenol production and virulence. *Mol Plant Microbe Interact*. 2009;22:1588–600.
- Gilbert J, Abramson D, McCallum B, Clear R. Comparison of Canadian *Fusarium graminearum* isolates for aggressiveness, vegetative compatibility, and production of ergosterol and mycotoxins. *Mycopathologia*. 2002;153:209–15.
- Miller JD, Taylor A, Greenhalgh R. Production of deoxynivalenol and related compounds in liquid culture by *Fusarium graminearum*. *Can J Microbiol*. 1983;29:1171–8.

31. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*. 2009;25:1968–9.
32. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5:R12.
33. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. Alignment of whole genomes. *Nucleic Acids Res*. 1999;27:2369–76.
34. Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, et al. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics*. 2010;11:461.
35. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003;19:ii215–25.
36. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21:3674–6.
37. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
38. Loytynoja A, Goldman N. A model of evolution and structure for multiple sequence alignment. *Philos Trans R Soc Lond B Biol Sci*. 2008;363:3913–9.
39. Yang Z. Paml 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91.
40. Kristensen R, Torp M, Kosiak B, Holt-Jensen A. Phylogeny and toxigenic potential is correlated in *Fusarium* species as revealed by partial translation elongation factor 1 alpha gene sequences. *Mycol Res*. 2005;109:173–86.
41. Liang J, Lofgren L, Ma Z, Ward TJ, Kistler HC. Population subdivision of *Fusarium graminearum* from barley and wheat in the Upper Midwestern United States at the turn of the century. *Phytopathology*. 2015;105:1466–74.
42. Hansen FT, Gardiner DM, Lysøe E, Fuertes PR, Tudzynski B, Wiemann P, et al. An update to polyketide synthase and non-ribosomal synthetase genes and nomenclature in *Fusarium*. *Fungal Genet Biol*. 2015;75:20–9.
43. Sieber CMK, Lee W, Wong P, Münsterkötter M, Mewes H-W, Schmeitzl C, et al. The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS ONE*. 2014;9: doi:10.1371/journal.pone.0110311.
44. Milani NA, Lawrence DP, Elizabeth Arnold A, VanEtten HD. Origin of pisatin demethylase (PDA) in the genus *Fusarium*. *Fungal Genet Biol*. 2012;49:933–42.
45. Jin J-M, Lee S, Lee J, Baek S-R, Kim J-C, Yun S-H, et al. Functional characterization and manipulation of the apicidin biosynthetic pathway in *Fusarium semitectum*. *Mol Microbiol*. 2010;76:456–66.
46. Sørensen JL, Sondergaard TE, Covarelli L, Fuertes PR, Hansen FT, Frandsen RJN, et al. Identification of the biosynthetic gene clusters for the lipopeptides fusaristatin A and W493 B in *Fusarium graminearum* and *F. pseudograminearum*. *J Nat Prod*. 2014;77:2619–25.
47. Mayans O, Scott M, Connerton I, Gravesen T, Benen J, Visser J, et al. Two crystal structures of pectin lyase A from *Aspergillus* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. *Structure*. 1997;5:677–89.
48. Glass NL, Kaneko I. Fatal attraction: nonself recognition and heterokaryon incompatibility in filamentous fungi. *Eukaryot Cell*. 2003;2:1–8.
49. Paoletti M, Saupe SJ, Clavé C. Genesis of a fungal non-self recognition repertoire. *PLoS ONE*. 2007;2: doi:10.1371/journal.pone.0000283.
50. Wang C, Zhang S, Hou R, Zhao Z, Zheng Q, Xu Q, et al. Functional analysis of the kinome of the wheat scab fungus *Fusarium graminearum*. *PLoS Pathog*. 2011;7:e1002460.
51. Gaffoor I, Brown DW, Trail F. Functional analysis of the polyketide synthase genes in the filamentous fungus *Gibberella zeae* (anamorph *Fusarium graminearum*). *Eukaryotic Cell*. 2005;4:1926–33.
52. Dao TTH, Linthorst HJM, Verpoorte R. Chalcone synthase and its functions in plant resistance. *Phytochem Rev*. 2011;10:397–412.
53. Lysøe E, Harris LJ, Walkowiak S, Subramaniam R, Divon HH, Riiser ES, et al. The genome of the generalist plant pathogen *Fusarium avenaceum* is enriched with genes involved in redox, signaling and secondary metabolism. *PLoS ONE*. 2014;9:e112703.
54. Harris LJ, Balcerzak M, Johnston A, Schneiderman D, Ouellet T. Host-preferential *Fusarium graminearum* gene expression during infection of wheat, barley, and maize. *Fungal Biol*. 2016;120:111–23.
55. Son H, Seo Y-S, Min K, Park AR, Lee J, Jin J-M, et al. A phenome-based functional analysis of transcription factors in the cereal head blight fungus, *Fusarium graminearum*. *PLoS Pathog*. 2011;7: doi:10.1371/journal.ppat.1002310.
56. Yun Y, Liu Z, Yin Y, Jiang J, Chen Y, Xu J-R, et al. Functional analysis of the *Fusarium graminearum* phosphatome. *New Phytol*. 2015;207:119–34.
57. Denancé N, Szurek B, Noël LD. Emerging functions of nodulin-like proteins in non-nodulating plant species. *Plant Cell Physiol*. 2014;55:469–74.
58. Gilbert J, Brûlé-Babel A, Guerrieri AT, Clear RM, Patrick S, Slusarenko K, et al. Ratio of 3-ADON and 15-ADON isolates of *Fusarium graminearum* recovered from wheat kernels in Manitoba from 2008 to 2012. *Can J Plant Pathol*. 2014;36:54–63.
59. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, et al. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathog*. 2012;8:e1002952.
60. Proctor RH, Van Hove F, Susca A, Stea G, Busman M, van der Lee T, et al. Birth, death and horizontal transfer of the fumonisin biosynthetic gene cluster during the evolutionary diversification of *Fusarium*. *Mol Microbiol*. 2013;90:290–306.
61. Jones JDG, Dangl JL. The plant immune system. *Nature*. 2006;444:323–9.
62. Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, et al. Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biol Evol*. 2015;7:1613–27.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

