A high-resolution genetic map of the cereal crown rot pathogen *Fusarium pseudograminearum* provides a near-complete genome assembly

DONALD M. GARDINER*, AURELIE H. BENFIELD, JIRI STILLER, STUART STEPHEN, KAREN AITKEN, CHUNJI LIU AND KEMAL KAZAN

Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, 306 Carmody Road, St Lucia, Brisbane, Qld, Australia

SUMMARY

Fusarium pseudograminearum is an important pathogen of wheat and barley, particularly in semi-arid environments. Previous genome assemblies for this organism were based entirely on short read data and are highly fragmented. In this work, a genetic map of *F. pseudograminearum* has been constructed for the first time based on a mapping population of 178 individuals. The genetic map, together with long read scaffolding of a short read-based genome assembly, was used to give a near-complete assembly of the four *F. pseudograminearum* chromosomes. Large regions of synteny between *F. pseudograminearum* and *F. graminearum*, the related pathogen that is the primary causal agent of cereal head blight disease, were previously proposed in the core conserved genome, but the construction of a genetic map to order and orient contigs is critical to the validation of synteny and the placing of species-specific regions. Indeed, our comparative analyses of the genomes of these two related pathogens suggest that rearrangements in the *F. pseudograminearum* genome have occurred in the chromosome ends. One of these rearrangements includes the transposition of an entire gene cluster involved in the detoxification of the benzoxazolinone (BOA) class of plant phytoalexins. This work provides an important genomic and genetic resource for *F. pseudograminearum*, which is less well characterized than *F. graminearum*. In addition, this study provides new insights into a better understanding of the sexual reproduction process in *F. pseudograminearum*, which informs us of the potential of this pathogen to evolve.

Keywords: *Fusarium* crown rot, genotype-by-sequencing, PacBio.

INTRODUCTION

Emerging insights from comparative and functional pathogenomic studies indicate that, from an evolutionary perspective, some pathogenic ascomycetous fungal genomes are subdivided into what can be conceptually considered as two different sub-

genomic compartments. These two compartments, given various names in the literature, such as 'two-speed genome' or 'core and accessory' genome, consist of a basal sub-genome, which is thought to encode components necessary for basic life functions, and a species- or strain-specific sub-genome. The species/strain-specific genome is thought to provide many of the species/strain-defining attributes, such as niche specialization or virulence towards a particular host. The physical arrangement of these two sub-genomes is observed to be different from one species to another. For example, some species, such as *Zymoseptoria tritici*, have separate chromosomes to encode the different components (Goodwin *et al.*, 2011), whereas others, such as *Aspergillus*, have specific regions of chromosomes, such as sub-telomeric regions (Fedorova *et al.*, 2008), that appear to house the genes that encode species-specific attributes. Regardless of the observed physical arrangement of different sub-genomic compartments, the species-specific regions tend to show higher levels of inter-strain diversity, including small nucleotide changes, the presence/absence of multi-gene regions and often higher densities of transposable elements or the remnants thereof. In sexual species, higher recombination rates are often observed in these regions, suggesting that these different sub-genomes evolve at different speeds. Thus, the dissection of genome structures can be highly informative to understand fungal biology. In the genus *Fusarium*, contrasting arrangements of the sub-genomes have been observed. In some species, sub-genomes are confined to entire chromosomes that are dispensable, whereas, in other species, they are confined to sub-telomeric regions (Ma *et al.*, 2013; Trail and Gardiner, 2014).

Fusarium pseudograminearum is an important pathogen of wheat and other small grain cereals in many regions of the world (Burgess *et al.*, 2001; Leslie and Summerell, 2006). Recent sequencing and comparative analyses of the *F. pseudograminearum* genome have provided new insights into the evolution of virulence in this pathogen (Gardiner *et al.*, 2012; Moolhuijzen *et al.*, 2013). However, currently available genome assemblies of *F. pseudograminearum* (Gardiner *et al.*, 2012; Moolhuijzen *et al.*, 2013) are based entirely on Illumina paired-end data, which, in this species, are insufficient for the assessment of genome structure. Each of the five *F. pseudograminearum* genome assemblies

*Correspondence: Email: Donald.Gardiner@csiro.au

generated so far varies in the number of scaffolds, from 182 (isolate CS3427) to 656 (isolate CS3096). We have hypothesized previously, based on alignments to *F. graminearum*, that genes that quantitatively contribute to virulence in *F. pseudograminearum* are located near chromosome ends (Gardiner *et al.*, 2012). However, in *F. pseudograminearum*, there is no direct evidence for this hypothesis because of the limitations of the available genome assemblies. Indeed, even the chromosome number in this species is not known, and, in the Fusaria, this can range from four (*F. graminearum*) to at least 17 (*F. solani*) (Coleman *et al.*, 2009; Gale *et al.*, 2005). Even within a species, the chromosome number can vary considerably, as exemplified by *F. oxysporum* f. sp. *cubense*, in which the chromosome number is in the range 9–14 (Boehm *et al.*, 1994). The chromosome number of *F. pseudograminearum* is expected to be four, based on its close relationship to *F. graminearum*, which also has four chromosomes. However, it is estimated that these species diverged at 1.7 Mya with a range of 0.5–3.4 Mya (O'Donnell *et al.*, 2013), and therefore the chromosome number in *F. pseudograminearum* may well be different from that in *F. graminearum*.

Genetic maps have been critical tools in the higher order assembly of the *F. verticillioides* (Jurgenson *et al.*, 2002b; Xu and Leslie, 1996) and *F. graminearum* (Gale *et al.*, 2005) genomes. A genetic map also provides an independent means to validate or improve sequence assemblies (Foulongne-Oriol, 2012), as shown for *F. graminearum* (Lee *et al.*, 2008). Our objective in this study was to develop a high-resolution genetic map of *F. pseudograminearum* and to anchor the previously published whole-genome shotgun (WGS) sequence for this species to this map. Our working hypothesis was that *F. pseudograminearum* would share a similar chromosomal organization and genetic attributes to *F. graminearum*. Despite *F. pseudograminearum* being heterothallic and *F. graminearum* being homothallic, the results obtained in this study largely confirm this hypothesis. The study also advances our understanding of this pathogen's genome architecture, for example, with respect to the placement of species-specific genes, and how in some cases *F. pseudograminearum* and *F. graminearum* have diverged in their genomic organization.

RESULTS AND DISCUSSION

Establishment of sexual crosses in *F. pseudograminearum*

Fusarium pseudograminearum strains are often infertile under laboratory conditions (Aoki and O'Donnell, 1999), and being a heterothallic species means that fertile strains are required for sexual reproduction in this pathogen. Typically, fertility issues relate to the ability of strains to differentiate the female ascogonium structure (Leslie and Klein, 1996). However, an all versus all co-culture approach was taken to identify productive crosses circumventing

the need to test all strains for female fertility. To this end, 25 different *F. pseudograminearum* strain combinations, using five strains each of MAT1-1 and MAT1-2 isolates, were tested for their ability to complete a sexual cycle. Only two combinations produced perithecia and ascospores. These were isolates CS3270 and CS3096 in combination with RBG5266. The cross between CS3270 and RBG5266 produced the greatest number of ascospores, and these strains served as the parents of the mapping population used in genetic map construction.

The genome sequences of parental isolates reveal a modified mating-type locus in *F. pseudograminearum*

Although the isolates CS3270 and RBG5266 could be crossed and provide progeny for the construction of a genetic map, the genome of neither parental strain was available. To assist in the calling of markers, a shotgun whole-genome sequence was assembled for both isolates using 75-bp paired-end MiSeq reads (see Experimental procedures). Using CLC genomics workbench, the nuclear genome of CS3270 was initially assembled into 745 contigs, with a total assembly size of 36.6 Mbp. The RBG5266 genome was assembled into 735 contigs with a length of 36.3 Mbp. The 745 CS3270 contigs were scaffolded using PBJelly software and Pacific Bioscience (PacBio) long reads from a 20-kbp library sequenced on a single-molecule real-time (SMRT) cell. After manual curation of the assembly, the final number of scaffolds was 89 with a total length of 37.1 Mbp.

The availability of genomes for isolates of both mating types allowed a complete analysis of the *F. pseudograminearum* mating-type idiomorphs (Fig. 1). As defined by Ma *et al.* (2013), similar to other *Fusarium* mating-type loci, the *F. pseudograminearum* idiomorphs were flanked by the conserved genes *APN2* (*FPSE_08225*) and *SLA2* (*FPSE_08218*). The *MAT1-1* idiomorph of *F. pseudograminearum* has also been described previously to contain, in addition to the expected three *MAT1-1* genes, a gene fragment with homology to *MAT1-2-3* in the *F. graminearum* mating-type locus (Ma *et al.*, 2013). Analysis of RNA sequencing (RNA-seq) data for *F. pseudograminearum* indicated that this fragment was not expressed in isolate CS3096 (data not shown), suggesting that it was a pseudogene. The *MAT1-1* idiomorph in *F. pseudograminearum* also appears to contain two additional genes, *FPSE_08219* and *FPSE_08220*, for which a small number of RNA-seq reads were observed (data not shown). Neither gene has sequence homologues in GenBank, but both are predicted to encode proteins of less than 100 amino acids. The *MAT1-2* idiomorph in isolate RBG5266 contains two genes between the conserved flanking genes which are divergently transcribed, in contrast with the arrangement in *F. graminearum*. Thus, although the mating-type idiomorphs of *F. pseudograminearum* contain the expected genes, there are additional genes in the *MAT1-1* locus that may or may not be involved in the mating process.

Fig. 1 The mating-type idiomorphs of *Fusarium pseudograminearum* (CS3270/CS3096 and RBG5266) in comparison with the *F. graminearum* locus. *Fusarium pseudograminearum* has a heterothallic mating system, whereas *F. graminearum* is homothallic. The mating-type loci in *Fusarium* spp. are flanked by the conserved genes *APN2* (FPSE_08225) and *SLA2* (FPSE_08218), which are depicted in grey. Genes shown in black represent those that have not been observed in mating-type loci previously and appear to be unique in each species. Ψ MAT1-2-3 represents the gene fragment in the *F. pseudograminearum* MAT1-1 locus with homology to the *F. graminearum* MAT1-2-3 gene.

The genetic map of *F. pseudograminearum* resolves four linkage groups most probably corresponding to four chromosomes

A genotyping-by-sequencing approach was used to identify single nucleotide polymorphism (SNP) markers in 178 individuals from the cross between CS3270 and RBG5266. Four linkage groups were resolved with 1512 markers at a logarithm of the odds (LOD) of 4. Marker grouping was maintained in all cases to a LOD of 10. An iterative approach was applied to validate the CS3270 genome assembly with the genetic map, and discrepancies between the map and assembly resulted in the splitting of one sequence scaffold. This split was also supported by the read alignments across the region. The map had a total length of 898 cM and an average marker density of 0.59 cM (Fig. 2). This map length is approximately 400 cM shorter than those reported for *F. graminearum* in two separate studies (Gale *et al.*, 2005; Jurgenson *et al.*, 2002a), although it is unknown whether these differences are a result of technical (e.g. different marker systems or software used in mapping) or inherent biological reasons. The linkage group corresponding to chromosome 2 was the most recombinationally active, accounting for 32% of the map length, although only comprising 24% of the sequence length (see below). The map data are presented in File S1 (see Supporting Information).

One hundred and thirty-five markers showed significant ($P < 0.1$) segregation distortion. These markers were arranged in five regions (Fig. 2). In four of the five cases, the CS3270 alleles were observed in higher frequency than the RBG5266 alleles. The underlying biological reason for segregation distortion of these regions is unknown, but presumably these alleles provide some quantitative advantage during the crossing and/or single ascospore isolation procedures. The one region that was observed to contain RBG5266 alleles at higher frequency corresponds to the final 230 kbp of chromosome 2 containing both the benzoxazolinone detoxification cluster and the *FpDLH1-FpAMD1* cluster previously implicated in pathogen virulence (Gardiner *et al.*, 2012; Kettle *et al.*, 2015). However, there is no evidence that any of these regions are linked to pathogen aggressiveness, as this has not been assessed across the mapping population. Similarly,

segregation distortion has been observed during the construction of a genetic map for *F. verticillioides*. In this case, at least one region showing such unexpected segregation could be attributed to a spore killer locus (Xu and Leslie, 1996). However, it is unlikely that these regions are a result of spore killer loci in *F. pseudograminearum* as the degree of segregation distortion observed was never more than 60% of alleles being derived from either parent. These regions in *F. pseudograminearum* are therefore more similar to the loci that showed less extreme segregation distortion in *F. verticillioides* (Xu and Leslie, 1996).

Integration of genetic map and physical scaffolds of *F. pseudograminearum*

Based on the genetic map, 57 scaffolds representing 35.7 Mbp of the genome could be unequivocally ordered with respect to each other. An additional 11 scaffolds representing 1.1 Mbp of sequence could be placed in the assembly based on alignment to *F. graminearum*. Twenty-two scaffolds representing 262 kbp of sequence could not be assigned to a chromosome.

Assembly of the genome into four chromosomes allowed an analysis of recombination rates across the chromosomes. With the exception of a few regions (Fig. 3), the polymorphic markers available in the mapping population from the cross between CS3270 and RBG5266 allowed recombination rates to be calculated across most of the genome. Similarly to *F. graminearum* (Gale *et al.*, 2005), extremes of high and low recombination rates were observed across each chromosome (Fig. 3). Chromosome ends and discrete regions within these chromosomes, thought to be ancient chromosome fusion sites (Cuomo *et al.*, 2007), showed high recombination rates. The density of SNPs between the two parents was variable across the chromosomes (Fig. 3). Similarly to *F. graminearum*, a large internal section of chromosome 2 showed relatively high recombination rates and very high SNP density. This region is thought to be undergoing a higher rate of adaptive evolution than the rest of the genome in *F. pseudograminearum* and *F. graminearum* (Sperschneider *et al.*, 2015).

Fig. 2 A genetic map of *Fusarium pseudograminearum* and alignment to the physical sequence map. The map was based on 1512 single nucleotide polymorphism (SNP) markers and the positions of these SNPs were used to order the sequence scaffolds into their chromosomal order and orientation. Each of the linkage groups is labelled 'LG' with marker positions indicated as horizontal bars on the linkage group diagram and distance (cM) on the left. Representations of the physical sequences are labelled Chr. Scaffolds present in the physical sequence assembly in forward orientation are depicted in red and scaffolds in the reverse orientation are depicted in green. The relationship between each of the linkage groups and the corresponding scaffolds is shown with lines connecting the first and last markers on each scaffold. All scaffolds are represented and those lacking links to the map have been placed in the chromosomal assembly via alignment to *F. graminearum*. Regions in which haplotype blocks of one parent were observed at higher frequency are highlighted in blue or purple with the isolate number of the parent of origin for which markers were over-represented labelled.

Location of known virulence and species-specific genes

One of the initial motivations for the construction of a genetic map and alignment of the physical sequence contigs into chromosomes for *F. pseudograminearum* was to determine where genes identified in functional studies, including those associated with virulence or secondary metabolite production functions, are located in the genome. Of 12 447 predicted genes in the original

F. pseudograminearum (CS3096) assembly, 12 350 could be aligned to the CS3270 genome. Based on the overall synteny between *F. pseudograminearum* and *F. graminearum*, most genes were expected to be located in similar genomic regions in both species (Fig. 4). Indeed, genomic locations of only 80 of a total of 11 547 orthologous genes in each species appear to have been transposed between different chromosomes. There are, however, some important exceptions to the maintenance of synteny between the two organisms. These exceptions are biased towards

Fig. 3 Recombination rates, single nucleotide polymorphisms (SNPs) and informative markers in the CS3270 × RBG5266 cross across the *Fusarium pseudograminearum* chromosomes. The recombination rate was calculated by summing the genetic distance in centimorgans (cM) for markers in 100-kbp windows from the map generated from the cross between isolates CS3270 and RBG5266. The availability of markers in the mapping population in any one 100-kbp window is indicated by the green dots. This represents windows in which there was at least one polymorphic marker called in the population in 173 or more individuals. For regions in which no polymorphic markers were observed, it is impossible to detect recombination in this mapping population. The SNP density is based on differences between isolates CS3270 and RBG5266, which were used as parents of the mapping population. Not all SNPs between these parents were converted to markers using the genotype-by-sequencing approach which employs reduced representational libraries in the sequencing process. The assembled sizes of chromosomes 1–4 are 11.8, 8.9, 7.9 and 8.2 Mbp.

the ends of the *F. pseudograminearum* chromosomes or internal regions of high SNP density and recombination rate. For species-specific genes, synteny cannot be established, and most of these are located in regions of high SNP density and recombination rates. For example, the gene cluster for the lipopeptide W493, produced by *F. pseudograminearum* but not by *F. graminearum* (Hansen *et al.*, 2015; Sørensen *et al.*, 2014), is located at approximately 6.5 Mbp of chromosome 1 in a region of reasonably high SNP density. As described by Sørensen *et al.* (2014), the W493 cluster appears to be inserted into the *F. pseudograminearum* genome in a region of microsynteny between the two species, with the deletion of only one gene from *F. graminearum*.

We have identified previously one genomic region in *F. pseudograminearum* that contains the *FpDLH1* virulence gene, which shares synteny with both chromosome 1 and chromosome 2 of *F. graminearum*, originally making it impossible to assign to a chromosomal position (Gardiner *et al.*, 2012). The same sequence contig from *F. pseudograminearum* also contains the benzoxazolinone detoxification cluster which, in *F. graminearum*, is located 300 kbp from the start of chromosome 1 (Kettle *et al.*, 2015). The chromosome-scale assembly of *F. pseudograminearum* achieved in this work places this group of genes about 230 kbp from the end of chromosome 2. This suggests that, although the genes in this region are largely conserved between *F. graminearum* and *F. pseudograminearum* (with the exception of *FpDLH1*), there has been a transposition of 14 genes into either *F. graminearum* chromosome 1 or *F. pseudograminearum* chromosome 2. Shuffling in this region is even more evident when *F. pseudograminearum* is compared with *F. verticillioides*, as the 'benzoxazolinone cluster' is present not as a single locus, but as two separate clusters on different chromosomes in *F. verticillioides* (Glenn *et al.*, 2002; Kettle *et al.*, 2015).

In summary, the establishment of crossing in *F. pseudograminearum* is an important step, as it opens up opportunities for the undertaking of new approaches towards a better understanding of the biology of this important pathogen, such as point mutation-based genetic screens similar to that recently applied to *F. graminearum* (Blum *et al.*, 2016). The generation of a chromosome-scale assembly has allowed, for the first time, the determination of the physical location of various genes involved in virulence towards plants and secondary metabolite production in *F. pseudograminearum*. The assembly and genetic map will serve as an important tool for work with this pathogen into the future.

EXPERIMENTAL PROCEDURES

Strains and nucleotide extraction

Strains were either from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) collection held in Brisbane, Australia or those described from the Sydney Royal Botanic Gardens (Australia) to

Fig. 4 Gene order is largely conserved between *Fusarium pseudograminearum* and *F. graminearum*. Dot plot of the gene order of genes with reciprocal best blast hits between *F. pseudograminearum* and *F. graminearum*. Coloured bars on the x and y axes delineate the four chromosomes of *F. pseudograminearum* and *F. graminearum*.

have female fertility (Bentley *et al.*, 2008). MAT1-1 strains were CS3096, CS3270, CS5834, RBG5263 and RBG5264. MAT1-2 strains were CS3220, CS3427, CS3487, RBG5265 and RBG5266.

DNA for Illumina-based sequencing was extracted from fungal mycelia grown in 20 mL of potato dextrose broth in deep 14-cm Petri dishes in the dark at 28°C for 4–5 days. Like other ascomycetes, mycelia of *F. pseudograminearum* contain haploid nuclei. DNA was extracted using a QIAgen (Melbourne, Victoria, Australia) Plant DNeasy gDNA spin kit according to the manufacturer's instructions. DNA for PacBio sequencing was extracted as follows. *Fusarium pseudograminearum* cultures were grown from spores in three 15-cm Petri dishes containing half-strength potato dextrose broth at 28°C in the dark. Mycelia were collected in miracloth [Calbiochem (Millipore), Melbourne, Victoria, Australia], the remaining liquid was removed by squeezing the cloth, and the mycelia were snap frozen in liquid nitrogen and lyophilized overnight. The dried mycelia were ground to a powder in 2-mL microcentrifuge tubes using a Retsch bead beater (MEP Instruments, Sydney, New South Wales, Australia) and a single 4-mm ball bearing, and resuspended in 15 mL of extraction buffer [1% sodium sarcosyl (w/v), 150 mM ethylenediaminetetraacetic acid (EDTA), 50 mM tris(hydroxymethyl)aminomethane (Tris), pH 8]. The suspension was centrifuged at 5000 *g* for 10 min at 4°C, the supernatant was transferred to a 50-mL tube, proteinase K (final concentration, 300 µg/mL) was added and the mixture was incubated at 50°C for 2 h. Two phenol–chloroform–iso-amyl alcohol (25 : 24 : 1) extractions were performed with gentle mixing only and centrifugation at 5000 *g* for 10 min at 4°C. DNA was ethanol precipitated (2 vol of 100% ethanol and 0.1 vol of 1 M NaCl) for 1 h at room temperature and pelleted at 5000 *g* for 20 min at 4°C. The pellet was air dried and resuspended in 2 mL of ultrapure (type I) water (produced using an Elga UltraPure Classic, Sydney, New South

Wales, Australia), 250 µL RNase A (10 mg/mL) (Sigma, Sydney, New South Wales, Australia) were added and incubated at room temperature for 1 h. RNase A was removed by phenol–chloroform–iso-amyl alcohol extraction followed by ethanol precipitation, as described above. The pellet was washed with 70% ethanol and centrifuged for 10 min at 5000 *g* at 4°C. The pellet was air dried and resuspended in 500 µL of ultrapure water by swirling without pipetting. The gDNA was cleaned up using a QIAgen genomic tip 100/G according to the manufacturer's instructions with the DNA sample diluted in 10 vol of QIAgen QBT buffer prior to loading on the column. All pipetting was performed as slowly as possible without vortexing the sample. The final pellet was dissolved in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0 with HCl).

Crossing methodology

Crossing was conducted as described for *F. graminearum* (Cavinder *et al.*, 2012) on carrot agar using a co-culture of wild-type strains of *F. pseudograminearum*. Ascospores were collected from plate lids (replaced the day before ascospore collection) by washing in 1 mL of sterile water and checked under a microscope. The suspension was plated out on water agar and ascospores were allowed to germinate overnight at room temperature. From the final selected cross between isolates CS3270 and RBG5266, 330 single germinating ascospores were picked onto half-strength potato dextrose agar (PDA) plates using an Olympus SZ51 (Brisbane, Queensland, Australia) stereo microscope to ensure that single germinating ascospores were selected. Isolates were grown at room temperature under fluorescent lamps prior to further analysis and storage.

Assembly of the genomes of isolate CS3270 and RBG5266

gDNA of isolate CS3270 was sequenced using half an Illumina MiSeq flow cell. The sequencing library was prepared using the Nextera XT kit from Illumina with Pippin Prep size selection (Sage Science, Boston, MA, USA). Paired-end reads of 75 bp in length were generated. Genome assembly was carried out on a desktop personal computer running Ubuntu version 15.04. MiSeq reads were quality trimmed using SolexaQA++ (Cox *et al.*, 2010) with a minimum read length of 40 bp and a minimum quality of Phred 30. For isolate CS3270, reads were imported into CLC genomics workbench version 8.5 (QIAGEN, Aarhus, Denmark). An initial *de novo* assembly was carried out (minimum contig size, 1000 bp; no scaffolding; length fraction, 1; similarity fraction, 0.99; automatic paired-end distance detection) and three high-coverage contigs representing the mitochondrial genome (two contigs) and PhiX spike in the sequencing control library were saved and used to filter the reads prior to the final assembly. The remaining reads were re-assembled (using the same parameters) into 747 contigs. One contig (7.8 kbp) representing the rDNA repeat and another of 1.2 kbp in length of unknown origin with coverage of more than two-fold that of the rest of the assembly were removed, leaving a final assembly of 745 contigs. Isolate RBG5266 was assembled in a similar manner, except that CLC Genomics Workbench 9.01 was used. After PhiX filtering, the RBG5266 genome was assembled into 737 contigs, three of which represented the mitochondrial genome and one the rRNA encoding repeats, which were removed. Two contigs were split following manual inspection of the assembly highlighted by discrepancies

in the Mauve-guided alignment to the scaffolded CS3270 genome (see below).

The resulting assembled contigs of CS3270 were scaffolded using PacBio reads derived from a single 20-kbp SMRT bell library sequenced in a single SMRT cell at the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia). One hundred and thirty-two scaffolds were constructed using the PBJelly part of PBSuite version 15.8.24 (English *et al.*, 2012) with default parameters.

The CS3270 assembly was polished using manual curation guided by both PBHoney (English *et al.*, 2014) and the genetic map. PBHoney was run with default parameters using the PacBio reads and the alignment of both Illumina and PacBio reads to regions highlighted as potential rearrangements inspected manually in CLC Genomics Workbench. The assembly was also inspected manually where the genetic map and scaffolds were discordant (see details in 'Map determination' section below). Using this combined approach, 13 scaffolds were split to give a scaffold number of 144. This number was reduced to 89 final scaffolds after manual inspection of PacBio read alignments to the scaffolds as follows. Reads were aligned to the scaffolds using the blasr program in PBSuite retaining the single best hit. Fifty-five short contigs (total length, 109 kbp) were removed in this way as they did not have PacBio read support. In most cases, these short sequences appeared to belong to junctions between contigs originally scaffolded and filled in with PBJelly that were not incorporated into the larger scaffolds, but the filling in process had none the less already incorporated their sequences.

The genome assembly for CS3270 has been deposited at DDBJ/EMBL/GenBank under the accession JTG000000000, and that of isolate RBG5266 under JTG000000000.

Marker calling

Markers segregating amongst the mapping population were identified using a genotyping-by-sequencing approach. Library construction for the mapping population followed the methods described previously for maize (Elshire *et al.*, 2011). Library preparation and sequencing were carried out by BGI (Hong Kong). Briefly, DNA samples were mixed with barcoded and common adaptors and dried down. The DNA–adaptor mixes were digested with *ApeKI* and ligated using T4 DNA ligase. An aliquot of the ligated DNA and adaptors was pooled (96 per pool) and size exclusion purified to remove unligated adaptors. Pooled samples were polymerase chain reaction (PCR) amplified according to Elshire *et al.* (2011), and the library size was selected on an agarose gel to include DNA fragments of approximately 300–600 bp in length. The library was analysed on a Bioanalyser 2100 and quantified via quantitative PCR prior to sequencing on an Illumina HiSeq2000 (91 nucleotide paired-end reads).

Sequence analysis was performed on high-performance computing resources available at CSIRO. Fastq files were assessed for quality and processed using SolexaQA++ (Cox *et al.*, 2010) with a quality threshold of Q30 and a minimum length of 40 nucleotides. An additional two nucleotides were removed from the 3' end of each read using the unix shell command `sed`, as individual reads that spanned an entire *ApeKI* restriction fragment were often observed with a terminal nucleotide derived from the 3' adaptor. Reads from the parental strains were processed in the same manner, but derived from 75-bp paired-end MiSeq reads used in the assembly of the parental strain genomes. Processed reads were then

aligned to the CS3270 genome assembly using Biokanga align version 3.7.8 (Stephen and Taylor, 2015), run without paired-end read processing, to allow read pairs for which one of the two reads were removed during quality filtering to be included. Alignment of reads with up to three mismatches per 100 bp was allowed. SNPs were not considered for processing unless there was a minimum of two reads covering that locus. SNPs were called using Biokanga snpmarkers 3.7.8 (Stephen and Taylor, 2015), again with a minimum read coverage of two at any locus before an SNP was considered. The parental isolate read alignments and progeny read alignments against the CS3270 reference sequence were all processed in a single implementation of Biokanga snpmarkers.

Map determination

The output of Biokanga snpmarkers was edited using custom shell and Perl scripts to remove unnecessary columns, leaving one column per isolate (available on request). SNPs called by Biokanga snpmarkers were further filtered to include only SNPs that had agreement in at least 90% of reads covering that locus and were also present in one of the parental strain alignments. Twenty-two of the progeny (isolates 6, 7, 10, 12, 31, 33, 36, 42, 64, 86, 90, 93, 112, 129, 136, 148, 149, 153, 164, 170, 188, 192) were also removed because of low coverage via shell scripts. The data were then encoded using awk in the unix shell interface, such that the data were represented in a format suitable for import into JoinMap 4.0 (Van Ooijen, 2006). The CS3270 derived alleles were all encoded as 'a' and the RBG5266 alleles were all encoded as 'b'. Missing data were encoded as '-'. SNPs that were called in at least 173 of the remaining 178 progeny were retained using a custom awk script which left 2435 SNPs. Adjacent markers (i.e. those next to each other within a scaffold) in the CS3270 assembly that segregated in an identical manner across the mapping population were also removed using the uniq function in the shell environment, leaving 1514 loci.

Map construction was performed using JoinMap version 4.0 build 31jul13.4feb11 employing the haploid (HAP1) population type. An additional two markers were removed from the mapping population in JoinMap as all progeny had the CS3270 allele. Although 255 markers were considered to be identical by JoinMap, these were maintained in the groups to facilitate assignment of a small number of scaffolds to chromosomal regions which had markers only identical to those on adjacent scaffolds. Markers were assigned to groups using two-point analysis at a LOD cut-off of 4. Maps of individual groups were calculated using maximum-likelihood mapping with default parameters. A small number of discrepancies between marker order and nucleotide coordinates were identified in the original round of mapping and, together with the manual curation based on PacBio data, scaffolds were split and the read alignment, marker calling and map generation repeated.

A golden path (agp) file was created manually from the order and orientation derived from the genetic map and, in some cases, alignment to the *F. graminearum* genome. This alignment was performed using progressiveMauve (Darling *et al.*, 2010) where map position or orientation was unavailable. A custom Perl script was used to generate whole chromosome sequences based on this agp file. Twenty-two sequences remained unplaced in the chromosomes, representing 262 kbp, less than 1% of the total assembly. Mapping was repeated against the chromosomal assembly, and the validity of this assembly was checked by ensuring that the

total length of the genetic map was the same when the map was generated without marker order constraints and when the map was generated using a fixed marker order based on the chromosomal position.

SNP density and recombination rate determination

SNPs were detected by aligning the filtered reads from each parent against the CS3270 assembly. As WGS sequence reads were used in this process, as opposed to genotyping-by-sequencing libraries, slightly greater read coverage cut-offs were implemented because of the more even coverage of the WGS libraries. To align the reads, Biokanga align version 3.7.8 (Stephen and Taylor, 2015) was used with a maximum substitution rate of three nucleotides per 100 bases of read length and with SNPs being putatively assigned only when at least five reads were covering a single locus. Biokanga snpmarkers (Stephen and Taylor, 2015) was then used to identify the SNPs with a maximum *P* value of 0.05 and additionally filtered using a custom Perl script (available on request). SNPs were only called if both isolates had a minimum coverage of five reads at the locus and a minimum of 90% of the base calls at the SNP locus were in agreement. In general, most of the called SNPs had much higher coverage as the genome average coverage was ~30, and a higher percentage of bases were the same (typically 100%). For SNP density determination, Microsoft Excel was used to count SNP calls in 100-kbp bins with logical functions employing the map generated using markers called against the chromosomal assemblies. Similarly, the genetic distance between adjacent markers was also determined in Microsoft Excel and summed for each 100-kbp window along the chromosomes using logical functions.

Gene order comparisons between *F. pseudograminearum* and *F. graminearum*

Gene order in the CS3270 genome was determined by mapping protein sequences from the existing annotation of isolate CS3096 to the chromosomal assembly of CS3270. Exonerate version 2.2.0 (Slater and Birney, 2005) was used to perform the mapping employing the protein2genome model with the maximum intron size set to 500 and only the single best mapping reported. Of the 12 447 genes predicted in CS3096, 12 350 could be aligned to the CS3270 genome.

Update of the existing *F. pseudograminearum* genome assembly and scaffolding of the RBG5266 genome

The original *F. pseudograminearum* genome assembly of isolate CS3096 (Gardiner *et al.*, 2012) and the RBG5266 assembly were aligned to the CS3270 chromosomal assembly to order and orient the contigs using the 'order contigs' feature in Mauve version 2.3.1 (Darling *et al.*, 2010). The reordering process used six and 15 iterations of alignments for CS3096 and RBG5266, respectively. The mauve alignments were converted to a golden path file using custom Perl scripts. The agp file was provided to GenBank and is available under the original accession for CS3096 (AFNW01000000).

Accessions

The genome assemblies for CS3270 and RBG5266 have been deposited at DDBJ/EMBL/GenBank under the accessions JTG000000000 and JTG000000000, respectively. Sequence reads for the parental isolates and mapping population are available at the National Center for Biotechnology Information (NCBI) via the BioProject accessions PRJNA267062, PRJNA267063 and PRJNA306180 and sequence read archive accessions SRP051011, SRP051012, SRR3098978 and SRP067696. The chromosomal assembly of isolate CS3096 is available at GenBank under accession AFNW01000000.

ACKNOWLEDGEMENTS

This work was funded by the Grains Research and Development Corporation, an Australian Government statutory authority, through project CSP00154. The authors declare no conflicts of interest relevant to this work.

REFERENCES

- Aoki, T. and O'Donnell, K. (1999) Morphological characterization of *Gibberella coronicola* sp. nov., obtained through mating experiments of *Fusarium pseudograminearum*. *Mycoscience*, **40**, 443–453.
- Bentley, A.R., Summerell, B.A. and Burgess, L.W. (2008) Sexual compatibility in *Fusarium pseudograminearum* (*Gibberella coronicola*). *Mycol. Res.* **112**, 1101–1106.
- Blum, A., Benfield, A.H., Stiller, J., Kazan, K., Batley, J. and Gardiner, D.M. (2016) High-throughput FACS-based mutant screen identifies a gain-of-function allele of the *Fusarium graminearum* adenyl cyclase causing deoxynivalenol overproduction. *Fungal Genet. Biol.* **90**, 1–11.
- Boehm, E.W.A., Ploetz, R.C. and Kistler, H.C. (1994) Statistical analysis of electrophoretic karyotype variation among vegetative compatibility groups of *Fusarium oxysporum* f. sp. *cubense*. *Mol. Plant–Microbe Interact.* **7**, 196–197.
- Burgess, L.W., Backhouse, D., Summerell, B.A. and Swan, L.J. (2001) Crown rot of wheat. In: *Fusarium* (Summerell, B.A., Leslie, J.F., Backhouse, D., Bryden, W.L. and Burgess, L.W., eds), pp. 271–294. St Paul, MN: APS Press.
- Cavinder, B., Sikhakolli, U., Fellows, K.M. and Trail, F. (2012) Sexual development and ascospore discharge in *Fusarium graminearum*. *J. Vis. Exp.* **61**, e3895.
- Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., Schmutz, J., Taga, M., White, G.J., Zhou, S., Schwartz, D.C., Freitag, M., Ma, L.J., Danchin, E.G., Henrissat, B., Coutinho, P.M., Nelson, D.R., Straney, D., Napoli, C.A., Barker, B.M., Gribskov, M., Rep, M., Kroken, S., Molnár, I., Rensing, C., Kennell, J.C., Zamora, J., Farman, M.L., Selker, E.U., Salamov, A., Shapiro, H., Pangilinan, J., Lindquist, E., Lamers, C., Grigoriev, I.V., Geiser, D.M., Covert, S.F., Temporini, E. and Vanetten, H.D. (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet.* **5**, e1000618.
- Cox, M., Peterson, D. and Biggs, P. (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, **11**, 485.
- Cuomo, C.A., Güldener, U., Xu, J.R., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L.J., Baker, S.E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.L., Decaprio, D., Gale, L.R., Gnerre, S., Goswami, R.S., Hammond-Kosack, K., Harris, L.J., Hilburn, K., Kennell, J.C., Kroken, S., Magnuson, J.K., Mannhaupt, G., Mauceli, E., Mewes, H.W., Mitterbauer, R., Muehlbauer, G., Münsterkötter, M., Nelson, D., O'Donnell, K., Ouellet, T., Qi, W., Quesneville, H., Roncero, M.I., Seong, K.Y., Tetko, I.V., Urban, M., Waalwijk, C., Ward, T.J., Yao, J., Birren, B.W. and Kistler, H.C. (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*, **317**, 1400–1402.
- Darling, A.E., Mau, B. and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, **5**, e11147.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.
- English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C. and Gibbs, R.A. (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One*, **7**, e47768.
- English, A.C., Salerno, W.J. and Reid, J.G. (2014) PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics*, **15**, 7.
- Fedorova, N.D., Khaldi, N., Joardar, V.S., Maiti, R., Amedeo, P., Anderson, M.J., Crabtree, J., Silva, J.C., Badger, J.H., Albarraq, A., Angiuoli, S., Bussey, H., Bowyer, P., Cotty, P.J., Dyer, P.S., Egan, A., Galens, K., Fraser-Liggett, C.M., Haas, B.J., Inman, J.M., Kent, R., Lemieux, S., Malavazi, I., Orvis, J., Roemer, T., Ronning, C.M., Sundaram, J.P., Sutton, G., Turner, G., Venter, J.C., White, O.R., Whitty, B.R., Youngman, P., Wolfe, K.H., Goldman, G.H., Wortman, J.R., Jiang, B., Denning, D.W. and Nierman, W.C. (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet.* **4**, e1000046.
- Foulongne-Oriol, M. (2012) Genetic linkage mapping in fungi: current state, applications, and future trends. *Appl. Microbiol. Biotechnol.* **95**, 891–904.
- Gale, L.R., Bryant, J.D., Calvo, S., Giese, H., Katan, T., O'Donnell, K., Suga, H., Taga, M., Usgaard, T.R., Ward, T.J. and Kistler, H.C. (2005) Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. *Genetics*, **171**, 985–1001.
- Gardiner, D.M., McDonald, M.C., Covarelli, L., Solomon, P.S., Rusu, A.G., Marshall, M., Kazan, K., Chakraborty, S., McDonald, B.A. and Manners, J.M. (2012) Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathog.* **8**, e1002952.
- Glenn, A.E., Gold, S.E. and Bacon, C.W. (2002) *Fdb1* and *Fdb2*, *Fusarium verticillioides* loci necessary for detoxification of preformed antimicrobials from corn. *Mol. Plant–Microbe Interact.* **15**, 91–101.
- Goodwin, S.B., Ben M'barek, S., Dhillon, B., Wittenberg, A.H.J., Crane, C.F., Hane, J.K., Foster, A.J., Van der Lee, T.A., Grimwood, J., Aerts, A., Antoniw, J., Bailey, A., Bluhm, B., Bowler, J., Bristow, J., van der Burgt, A., Canto-Canche, B., Churchill, A.C., Conde-Ferraz, L., Cools, H.J., Coutinho, P.M., Csukai, M., Dehal, P., De Wit, P., Donzelli, B., van de Geest, H.C., van Ham, R.C., Hammond-Kosack, K.E., Henrissat, B., Kilian, A., Kobayashi, A.K., Koopmann, E., Kourmpetis, Y., Kuzniar, A., Lindquist, E., Lombard, V., Maliepaard, C., Martins, N., Mehrabi, R., Nap, J.P., Ponomarenko, A., Rudd, J.J., Salamov, A., Schmutz, J., Schouten, H.J., Shapiro, H., Stergiopoulos, I., Torriani, S.F., Tu, H., de Vries, R.P., Waalwijk, C., Ware, S.B., Wiebenga, A., Zwieters, L.H., Oliver, R.P., Grigoriev, I.V. and Kema, G.H. (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet.* **7**, e1002070.
- Hansen, F.T., Gardiner, D.M., Lysøe, E., Fuertes, P.R., Tudzynski, B., Wiemann, P., Sondergaard, T.E., Giese, H., Brodersen, D.E. and Sørensen, J.L. (2015) An update to polyketide synthase and non-ribosomal synthetase genes and nomenclature in *Fusarium*. *Fungal Genet. Biol.* **75**, 20–29.
- Jurgenson, J.E., Bowden, R.L., Zeller, K.A., Leslie, J.F., Alexander, N.J. and Plattner, R.D. (2002a) A genetic map of *Gibberella zeae* (*Fusarium graminearum*). *Genetics*, **160**, 1451–1460.
- Jurgenson, J.E., Zeller, K.A. and Leslie, J.F. (2002b) Expanded genetic map of *Gibberella moniliformis* (*Fusarium verticillioides*). *Appl. Environ. Microbiol.* **68**, 1972–1979.
- Kettle, A.J., Batley, J., Benfield, A.H., Manners, J.M., Kazan, K. and Gardiner, D.M. (2015) Degradation of the benzoxazolinone class of phytoalexins is important for virulence of *Fusarium pseudograminearum* towards wheat. *Mol. Plant Pathol.* **16**, 946–962.
- Lee, J., Jurgenson, J.E., Leslie, J.F. and Bowden, R.L. (2008) Alignment of genetic and physical maps of *Gibberella zeae*. *Appl. Environ. Microbiol.* **74**, 2349–2359.
- Leslie, J.F. and Klein, K.K. (1996) Female fertility and mating type effects on effective population size and evolution in filamentous fungi. *Genetics*, **144**, 557–567.
- Leslie, J.F. and Summerell, B.A. (2006) *The Fusarium Laboratory Manual*. Ames, IA: Blackwell Publishing.
- Ma, L.J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., Gardiner, D.M., Manners, J.M. and Kazan, K. (2013) *Fusarium* pathogenomics. *Annu. Rev. Microbiol.* **67**, 399–416.
- Moolhuijzen, P.M., Manners, J.M., Wilcox, S.A., Bellgard, M.I. and Gardiner, D.M. (2013) Genome sequences of six wheat-infecting *Fusarium* species isolates. *Genome Announc.* **1**, 00613–e00670.
- O'Donnell, K., Rooney, A.P., Proctor, R.H., Brown, D.W., McCormick, S.P., Ward, T.J., Frandsen, R.J., Lysøe, E., Rehner, S.A., Aoki, T., Robert, V.A., Crous, P.W., Groenewald, J.Z., Kang, S. and Geiser, D.M. (2013) Phylogenetic analyses of

- RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genet. Biol.* **52**, 20–31.
- Slater, G.S.C. and Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.
- Sørensen, J.L., Sondergaard, T.E., Covarelli, L., Fuertes, P.R., Hansen, F.T., Frandsen, R.J.N., Saei, W., Lukassen, M.B., Wimmer, R., Nielsen, K.F., Gardiner, D.M. and Giese, H. (2014) Identification of the biosynthetic gene clusters for the lipopeptides Fusaristatin A and W493 B in *Fusarium graminearum* and *F. pseudograminearum*. *J. Nat. Prod.* **77**, 2619–2625.
- Sperschneider, J., Gardiner, D.M., Thatcher, L.F., Lyons, R., Singh, K.B., Manners, J.M. and Taylor, J.M. (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biol. Evol.* **7**, 1613–1627.
- Stephen, S. and Taylor, J.M. (2015) *BioKanga: A Suite of High Performance Bioinformatics Applications*. Canberra, Australia: CSIRO.
- Trail, F. and Gardiner, D.M. (2014) Application of genomics to the study of pathogenicity and development in *Fusarium*. In: *The Mycota Vol. XIII: Fungal Genomics* (Nowrousian, M., ed.), pp. 267–300. Berlin-Heidelberg: Springer.
- Van Ooijen, J.W. (2006) *JoinMap® 4.0: Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma B.V.
- Xu, J. and Leslie, J.F. (1996) A genetic map of *Gibberella fujikuroi* mating population A (*Fusarium moniliforme*). *Genetics*, **143**, 175–189.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

File S1 *Fusarium pseudograminearum* genetic map. The map was derived from 178 progeny from a cross between isolates CS3270 and RBG5266. Maps for the four linkage groups are presented in the 'Map' tab, where the markers are given names corresponding to the sequence contig names and nucleotide positions. A genetic map was also generated, whereby the positions of all markers were set based on their chromosomal nucleotide positions, and this is presented in the 'Map_FixedOrderChrom' tab.