Comparative genomic analysis revealed rapid differentiation in the pathogenicity-related gene repertoires between *Pyricularia oryzae* and *Pyricularia penniseti* isolated from a *Pennisetum* grass

Huakun Zheng[a,b#], Zhenhui Zhong[b,c#], Mingyue Shi[b,c], Limei Zhang[b,c], Lianyu Lin[b,d], Yonghe Hong[b,d], Tian Fang[b,c], Yangyan Zhu[b,c], Jiayuan Guo[b,c], Limin Zhang[b,c], Jie Fang[b,d], Hui Lin[a], Justice Norvienyeku[b,d], Xiaofeng Chen[e], Guodong Lu[a,b*], Hongli Hu[b,c*], Zonghua Wang[b,c,e*]

[a] National Engineering Research Center of JUNCAO Technology, College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

[b] State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

[c] College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

[d] College of life science, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

[e] Institute of Oceanography, Minjiang University, Fuzhou 350108, China.

[#] These authors contributed to this work equally.

* To whom correspondence should be addressed. Email: wangzh@fafu.edu.cn, huhongli7905@gmail.com and gdlufafu@163.com.

Running title: Whole-genome sequencing of a *Pennisetum*-infecting blast fungus

Key words: *Pennisetum*, *Pyricularia*, PacBio sequencing, Pathogenesis-related genes, Comparative genomic analysis

Corresponding author:

Zonghua Wang

Mailing address: Institute of Oceanography, Minjiang University, Fuzhou 350108, China

Email: wangzh@fafu.edu.cn

Phone number: 86-13706948783

Hongli Hu

Email: huhongli7905@gmail.com

Phone number : 86-13599082338

Guodong Lu

National Engineering Research Center of JUNCAO Technology, College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

Email: gdlufafu@163.com

Phone number: 86-13950480067

Abstract

Backgrounds: *Pyricularia* is a multispecies complex that could infect and cause severe blast disease on diverse hosts, including rice, wheat and many other grasses. Although the genome size of this fungal complex is small [~40 Mbp for *Pyricularia oryzae* (syn. *Magnaporthe oryzae*), and ~45 Mbp for *P. grisea*], the genome plasticity allows the fungus to jump and adapt to new hosts. Therefore, deciphering the genome basis of individual species could facilitate the evolutionary and genetic study of this fungus. However, except for the *P. oryzae* subgroup, many other species isolated from diverse hosts, such as the *Pennisetum* grasses, remain largely uncovered genetically.

Results: Here, we report the genome sequence of a pyriform-shaped fungal strain *P. penniseti* P1609 isolated from a *Pennisetum* grass (JUJUNCAO) using PacBio SMRT sequencing technology. We performed a phylogenomic analysis of 28 Magnaporthales species and 5 non-Magnaporthales species and addressed P1609 into a *Pyricularia* subclade that is distant from *P. oryzae*. Comparative genomic analysis revealed that the pathogenicity-related gene repertoires were fairly different between P1609 and the *P. oryzae* strain 70-15, including the cloned avirulence genes, other putative secreted proteins, as well as some other predicted *Pathogen-Host Interaction* (*PHI*) genes. Genomic sequence comparison also identified many genomic rearrangements.

Conclusion: Taken together, our results suggested that the genomic sequence of the *P. penniseti* P1609 could be a useful resource for the genetic

67 study of the *Pennisetum*-infecting *Pyricularia* species.

68

69

Introduction

Pyricularia was established by Saccardo to accommodate a type of fungal species based on the pyriform conidia when the first species of this pathogen, *Pyricularia grisea*, was isolated from crabgrass (*Digitaria sanguinalis* L.) [1]. What raised the concern of this *Pyricularia* fungus was the notorious blast disease on rice and wheat caused by one of its species, *Pyricularia oryzae* (syn. *Magnaporthe oryzae*) [2-4]. To date, 100 plant genera comprising of 256 species have been documented as the hosts of the *Pyricularia* species (https://nt.ars-grin.gov/fungaldatabases/fungushost/fungushost.cfm), among which 54 genera belong to the Poaceae family. Seven *Pyricularia* species (including one unidentified species) have been isolated from *Pennisetum* spp., a widespread genus in the Poaceae family, and more than one *Pyricularia* species can be found on the same *Pennisetum* species. For instance, 4 *Pyricularia* species, namely, *P. penniseti*, *P. penniseticola*, *P. setariae* and *Pyricularia sp*. have been found on *P. typhoides* [5, 6].

The genome sequence of *P. oryzae* strain 70-15, a hybrid clone of rice-infecting isolate 104-3 and the weeping love grass isolate AR4 [7] has facilitated the revelation of developmental and pathogenic mechanisms of the blast fungus and rendered it to become one of the most important model fungus [8, 9]. Since the publication of the *P. oryzae* strain 70-15 genome, more field blast isolates were sequenced and assembled, including the Ina168, HN9311, FJ81278, Y34, P131, 98-06 and Guy11 [10-14]. Comparative

92 genomic analyses and functional studies of these strains revealed genome
93 plasticity and the involvement of the lineage specific genes in pathogenicity [12,
94 14]. More recently, facilitated by the fast developing of sequencing
95 technologies, a number of field isolates from rice, as well as isolates from
96 different grass and cereal hosts, were sequenced and subjected to
97 population-level analyses, revealing host immunity as the major force driving
98 specialization after host shift [2, 15-18]. However, genomes of most of the
99 species of the *Pyricularia* complex remain unexplored. For example, among
100 the 7 identified *Pyricularia* species isolated from *Pennisetum* grasses [5], only
101 *P. pennisetigena* was recently sequenced [2].

102 Here, we reported the whole-genome sequence of *P. penniseti* [19]
103 isolated from a *Pennisetum* grass JUJUNCAO (*Pennisetum giganteum* z. x.
104 Lin). JUJUNCAO was originally developed as culture matrix for the cultivation
105 of edible mushrooms by Lin *et al*, and later became a versatile grass that are
106 used as forage for cattle and sheep, material for the biofuel production, and
107 tool for the remedy of soil erosion [20-23]. We have recently isolated a
108 pyriform-shaped fungus, P1609, from leaf spots of JUJUNCAO. P1609 caused
109 a typical blast fungal disease symptom on JUJUNCAO, showing small, round
110 or elliptical lesions as an initial symptom and spindle shaped, grayish to tan
111 necrotic with yellow halos at a later disease stage. Since the morphologic and
112 phylogenetic analyses distinguished P1609 from other identified *Pyricularia*
113 species, but undistinguished from the *P. penniseti* reported in 1970 India [5, 24],

we therefore termed this fungus *P. penniseti* [19]. In this study, we performed genome sequencing of this strain, aiming for a proper classification of this fungus in the *Pyricularia* population and identification of genes that may be involved in the adaptation of this fungus to JUJUNCAO.

Results

Genome sequencing and assembly

We sequenced the P1609 genome with the long-read PacBio technology. In total, 312,061 reads with 8.6 Kb average lengths were obtained, representing about 60-fold coverage of the genome (Fig. 1A). The genome sequence was assembled with the HAGP pipeline, resulting in the total assembly space of 41.82 Mb (Table 1), similar to assemblies of isolates sequenced by PacBio [10]. The assembly contains 53 contigs, with the N50 of 3.4 Mb and the largest contig of 7.56 Mb (Fig. 1B). Contigs > 1Mb cover 89.7% and contigs > 100 Kb cover 98.5% of the genome (Fig. 1B; Table 1), indicating long-continuity of the assembly. The GC content of the assembly is 50.3%, similar to genomes of *Pyricularia* isolates from different host plants, which range from 48.6% to 51% [18]. Genome annotation identified 13,102 genes with average gene size of 1,758 bp, fairly evenly dispersed on contigs (Table 2, Fig. 1C, track b).

De novo repeat sequence analysis identified 7.67% of repeat sequences, among which 4.27% were Gypsy and Copia, the two long terminal repeats (LTR)-type retrotransposons. This finding is consistent with former

studies, suggesting that LTR-type retrotransposons are the most expanded transposable elements (TEs) in P1609 genome [25]. The TEs are not evenly distributed along the contigs. While some contigs are enriched with TEs, other contigs have less TE rich regions, suggesting that the P1609 genome also underwent transposon expansion as observed in other plant pathogens [26].

Comparative and phylogenetic analysis

To understand the genetic relationship of P1609 with other fungal phytopathogens, we generated a phylogenetic tree of P1609 with *Botrytis cinerea* (Bc), *Colletotrichum gloeosporioides* (Cg), *Fusarium graminearum* (Fg), *Neurospora crassa* (Nc), *Pyricularia oryzae* (Po), *Sclerotinia sclerotiorum* (Ss), *Trichoderma reesei* (Tr) and *Ustilago maydis* (Um). Since P1609 showed a close morphological relationship with *P. oryzae* isolates, we included *Pyricularia* isolates collected from different host plants, including *Oryza sativa* (PoOs), *Triticum aestivum* (PoTa), *Digitaria sanguinalis* (PoDs), *Setaria viridis* (PoSv), and *Eleusine indica* (MoEi). In total, 2,051 single-copy genes shared by all the examined genomes were selected to infer phylogeny [27]. The resulting phylogenetic tree indicated that P1609 is more divergent with PoOs, PoTa, PoSv, and PoEi (*P. oryzae*) in contrast to PoDs (*P. grisea*) (Fig. 2A). We then estimated divergence time of P1609 and *Pyricularia* isolates by assuming a constant molecular clock calibrated in a previous study, which estimated the divergence of *Neurospora* and *Pyricularia* at about 200 million years ago (MYA). The estimation indicated that P1609 and *Pyricularia* isolates diverged

158 at about 31 MYA, earlier than the divergent time of rice- and *S. viridis*-infecting
159 isolates (about 10,000 years ago) [28, 29]. The phylogenetic tree also
160 indicated that P1609 is a member of Magnaporthales, but is distantly related to
161 the *Pyriculaira* isolates collected from PoOs, PoTa, PoDs, PoSv, and PoEi. To
162 further determine exactly where P1609 localized in Magnaporthales, we also
163 generated a phylogenomic tree of P1609 with 28 Magnaporthales species and
164 5 non-Magnaporthales species using amino acid sequences of 226 conserved
165 orthologous genes as described [30]. The result showed that P1609 was
166 localized in the *Pyricularia* subclade, and was closer to *P. oryzae* than
167 *Xenopyricularia zizaniicola* (Fig. 2B), indicated that P1609 is a *Pyricularia*
168 species.

169 We next conducted a comparative genomic study of P1609 with 70-15
170 (the reference isolate of *P. oryzae*), *N. crassa*, *F. graminearum* and *C.*
171 *gloeosporioides*. Venn diagram showed that 5 organisms share 5,454 of gene
172 families with each other, which covers 50.2% of the gene set of P1609 (Fig.
173 2C). Notably, the number of unique genes (no homologs in the selected
174 organisms) in P1609 was 2,210, which is about twice more than the number of
175 unique genes recorded for 70-15. Although most of these unique genes had no
176 functional annotations, Pfam annotation indicated that some of the unique
177 genes encode carbohydrate-active enzymes involved in polysaccharides
178 metabolism pathways. The high representation of carbohydrate-active
179 enzymes may be related to the adaptation to the host *P. giganteum* (Fig. S1).

We therefore scanned natural selection of 5,991 pairs of orthologs of *P. oryzae* and P1609, identifying 6 genes with Ka/Ks > 1 (Table 3). These genes are involves in different secondary metabolic pathways.

Gene Categories Involved in Pathogenicity

Plant pathogenic fungi employed diverse gene repertoires to invade host plants and subvert host immune systems, which include effectors, carbohydrate-active enzymes (CAZymes), other secreted enzymes and fungal secondary metabolisms [31, 32]. P1609 and *P. oryzae* isolates are divergent and no cross-infectivity exists between them. To understand their secretomic difference, we first compared predicted secretomes between P1609 and the *P. oryzae* isolates. Among 1,409 putative secreted proteins of P1609, 236 were unique in P1609 (Table S1). By contrast, 165 putative effectors in the 70-15 genome were absent in the P1609 genome (Table S2). Notably, all known avirulence genes (*AVR*s) from the rice-infecting isolate genomes were absent in P1609 genome.

We next compared CAZymes of P1609, groups of enzymes enable plant pathogens to break down the plant cell wall [33]. Our BLASTp search results showed that P1609 contains more predicted CAZyme-coding genes than 70-15 (Fig. S2). Detailed analysis showed that the P1609 genome encodes 5 unique CAZyme-coding genes, namely CBM61, GH117, GH35, GH65 and PL24. While it has 6 copies of GH28 pectinases (3 copies in

Pyricularia species; Fig. S2). We then further analyzed the distribution of annotated PHI genes in P1609. In total, we identified 1,692 potential PHI genes belonging to 1,154 gene families (Table S3). Interestingly, we found that several PHI genes exhibited great expansion in P1609 genome. For instance, *MGG_12656*, a gene involved in virulence in *P. oryzae*, has 107 homologs in P1609, and ChLae1 contributing to toxin production and virulence in a maize pathogen *Cochliobolus heterostrophus* has 17 homologs in P1609 [34, 35]. Compared with 70-15, 35 PHI genes were unique in P1609 (Table 4), most of which had highly similar homologs in *Fusarium* (*Gibberella*).

Chromosome rearrangements

To explore genome collinearity and rearrangement between P1609 and 70-15, since 70-15 was assembled into chromosome level and showed a closer phylogenetic relationship with the P1609. The identified collinear gene blocks that linked with different chromosomes in 70-15 (Fig. 3A) were visualized in the contigs (ctgs) of P1609 (Fig. 3B). Generally, P1609 and 70-15 displayed high genome collinearity. Ctg2, ctg3, ctg5 and ctg7 in P1609 correspond to chr.2, chr.6, chr.3 and chr.1 of 70-15, respectively. We found that the second largest contig in the P1609, ctg2, is a recombinant of chr.4 and chr.7 of 70-15 (Fig. 3B). The joining region was spanned by multiple single PacBio long reads (Fig. 4), excluding the possibility that the rearrangement was an artifact due to assembly errors. Meanwhile, notably, contig end region of P1609 showed a higher level of chromosome deletion and rearrangement (Fig. 3c). For instance,

ctg4 and ctg8 were merged by blocks from chr.1 and chr.3 at the end of the contig, while ctg9 was merged by blocks from chr.3 and chr.5 at the end of the contig (Fig. 3B).

Discussion

Although the genome size of *Pyricularia* species is small, it is well-known for its complicated genomic plasticity. Here we sequence the genome of *P. penniseti* using the PacBio SMRT technology. Comparative genomic analysis revealed several chromosome rearrangement events and difference in pathogenicity-related gene repertoires between the *P. penniseti* strain P1609 and the *P. oryzae* strain 70-15.

Chromosome fission and fusion

Long-read sequencing greatly improved genome assembly and thus provided more valuable details on genome structures at a chromosome level. By comparing the chromosome structure between P1609 and 70-15, we found several chromosome splitting and rearrangement events. Chromosomal rearrangements have been reported to be associated with virulence evolution in several pathogens by losing the *AVR* gene(s) [26, 36, 37]. Our results indicated frequent chromosome rearrangement and splitting in the blast fungi, and the telomere regions are very unstable in its haploid genome during the adaptation to different host species. Further study is required to investigate the role of the chromosome recombination during the adaptation to JUJUNCAO.

Unique genes and positively selected genes

In this study, we found 2,210 unique genes that were absent in the 70-15 genome. The P1609 genome encodes more CAZymes than 70-15 genome. For example, P1609 contains twice the number of genes encoding glycosyl hydrolase family 28 (GH28) pectinases in 70-15. Generally, necrotrophic plant pathogens possess more GH28 enzymes than biotrophic and non-pathogens fungi [38]. These results suggested that P1609 might heavily rely on CAZymes in the interaction with JUJUNCAO.

To identify the positively selected genes, we detected with Ka>Ks from 5,991 orthologous pairs between *P. oryzae* and P1609 and identified 6 genes with Ka>Ks in P1609. Functional annotation showed that most of these genes involved in secondary metabolic pathways. For example, P1609_5032 encodes an isoamyl alcohol oxidase that turns isoamyl alcohol into isovaleraldehyde [39]. In *Saccharomyces cerevisiae*, isoamyl alcohol could induce filament formation [40]. P1609_5032 might play a role in fungal development through controlling the level of isoamyl alcohol. P1609_791 encodes a folylpolyglutamate synthase involved in biosynthesis of folate, required for protein synthesis of bacterial, mitochondrial, and chloroplast, including purines, dTMP, methionine, and formyl-methionyl-tRNA [41]. P1609_3360 encodes a glycerol uptake protein, which is an important intracellular osmolyte participating in osmotic stress response [42] and critical for the function of appressorium in *P. oryzae* [43]. P1609_7869 encodes a

spore surface glycoprotein. Its homologs have been proved to be involved in spore adhesion to hydrophobic surface in several *Colletotrichum* species [44-46] rather than spore tip mucilage as in *P. oryzae* [47]. P1609_1006 encodes a BclB glycoprotein (collagen-like protein). Its homologs in *Bacillus anthracis* were important components of the infection-associated structure exosporium [48-50], although its role in filamentous pathogens remains unknown. The positive selection on these genes suggested that they might play roles in the interaction between P1609 and JUJUNCAO.

Putative secreted proteins

There are two layers of plant immunity: the pathogen-associated molecular patterns (PAMPs)-triggered immunity (PTI), and the effetor-triggered immunity (ETI). It was previously proposed by Schulze-Lefert and Panstruga that ETI is the major force driving the host specificity of pathogens [51]. Our previously study focusing on *AVR* gene evolution of the *Pyricularia* species also revealed that directional selection exerted by host plants is the direct force driving host specificity in *Pyricularia* species [18]. Recent studies on the *P. oryzae* populations revealed that divergent host immunity systems (both PTI and ETI) in japonica (*Oryza sativa* subsp. xian) and indica (*Oryza sativa* subsp. geng), determining the deposition of effector repertoires and specialization to the two subspecies [15-17]. Our comparative genomic analysis showed that P1609 contains a large number of unique effector candidates by comparing with 70-15, but lost many putative effectors in the 70-15 genome, including all

known AVR effectors. One possible explanation is that the JUJUNCAO harbors a high level of basal immunity, as well as an arsenal of resistance genes, which driven P1609 to gain a lot of effectors to overcome the robust basal immunity posed by JUJUNCAO, and at the meanwhile, abandoned the *AVR* genes that could be recognized by the *R* genes from JUJUNCAO.

Conclusion

Pyricularia species are pathogens of either food- or forage grasses. The model fungus *P. oryzae* had been well studied. However, there are only a few whole-genome sequences available of other species, such as those from *Pennisetum* grasses. Here, we generated long-read PacBio reads and produced a assemblage with long-continuity contig sequences. Phylogenomic and comparative genomic analysis showed that P1609 is a *Pyricularia* species genetically distant with *P. oryzae*, and the two genomes vary substantially in their pathogenicity-related gene repertoires. In summary, the P1609 assembly and genome annotation represents the few available *Pyricularia* genome resources for studying the pathogenic mechanism of this fungus towards *Pennisetum* grasses.

MATERIALS AND METHODS

Isolation of the fungal strain

The *Pennisetum*-infecting strain P1609 was isolated from the leaf spot lesion of JUJUNCAO (*Pennisetum giganteum* z. x. Lin), in the nursery of National

Engineering Research Center of JUNCAO Technology, Fujian Agriculture and Forestry University located at No. 63 Xiyuangong Road, Minhou County, Fuzhou, Fujian Province, China.

DNA extraction, amplification and sequencing

To prepare the genomic DNA for sequencing, the P1609 isolate was cultured in the liquid complete medium (CM) in a 110-rpm shaker at 25 °C for 3 to 4 days. The mycelia were then collected for the preparation of genomic DNA using a CTAB method as previously described [18]. Sequencing libraries were prepared using the SMRTbellTM Template Prep Kit 1.0 (PACBIO) and sequenced using PacBio Sequel platform (NovoGene, China).

Assembly and annotation

De novo sequence assembly was conducted by SMRTLink v. 5.0.1.10424, HGAP 4 pipeline provided by Pacific Bioscience Company. In HGAP 4 pipeline, the expected genome size was set as 45 Mb based on the reported size of *Pyricularia* genomes, and default settings were used for other parameters. Gene prediction was conducted using Fgenesh from SoftBerry (MolQuest II v2.4.5.1135, http://linux1.softberry.com/berry.phtml) with *Pyricularia* additional variants as training organism. Gene functional domain annotation was conducted by InterproScan (version 4.8, http://www.ebi.ac.uk/ interpro/), and PfamScan [52]. Pathogen-Host Interaction (PHI) genes were predicted by performing a whole genome blastp analysis against the PHI database ($E<10^{-10}$) [53, 54]. Putative carbohydrate-active enzymes (CAZymes) were identified

using the HMMER 3.1b1 by searching annotated HMM profiles of CAZymes downloaded from the dbCAN database in protein sequences of P1609 [55].

Repeat analysis

De novo repeat sequence identification was analyzed by using RepeatModeler (version 1.0.8) with default settings. Repeat sequences obtained from RepeatModeler have been used to search for repeat sequences in the P1609 genome by RepeatMasker (version 3.3.0) (http://www.repeatmasker.org/) [56].

Phylogenetic analysis and comparative genomic analysis

Phylogenomic tree of P1609 and *B. cinerea* [57], *C. gloeosporioides* [58], *F. graminearum* [59], *N. crassa* [60], *P. oryzae* [8], *S. sclerotiorum* [57],*T. reesei* [61] and *U. maydis* [62] was built based on single copy orthologs from clustering result of OrthoFinder (v0.6.1) [27]. 2,051 single copy genes have been selected out from 13 organisms in total (see Fig. 1A) and aligned with MAFFT (mafft-linsi-anysymbol) [63]. The phylogenetic tree was constructed using FastTree based on the alignments of single-copy orthologs with approximately-maximum-likelihood model and bootstrap 100 [64]. For divergence time estimation, the phylogenetic analysis was conducted using r8s (version 1.81), and the divergence time of *Pyricularia* and *Neurospora* (200 MYA) was used as a reference [29, 65]. Clustering result of 13 genomes was also used for unique gene identification and comparative genomic analysis. We used MCScanX to identified syntenic blocks between P1609 and 70-15. To detect the conserved synteny blocks, the reciprocal best-match paralogs of

P1609 and 70-15 were conducted by all-against-all BLASTP comparison, with E-value $<10^{-10}$ [66].

List of abbreviations

Bc: *Botrytis cinerea* (*B. cinerea*); **Cg:** *Colletotrichum gloeosporioides* (*C. gloeosporioides*); **Fg:** *Fusarium graminearum* (*F. graminearum*); **Nc:** *Neurospora crassa* (*N. crassa*); ***P. grisea*:** *Pyricularia grisea*; **Po:** *Pyricularia oryzae* (*P. oryzae*); **Ss:** *Sclerotinia sclerotiorum* (*S. sclerotiorum*); **Tr:** *Trichoderma reesei* (*T. reesei*); **Um:** *Ustilago maydis* (*U. maydis*); **PoDs:** *Pyricularia* strain isolated from *Digitaria sanguinalis*; **PoSv:** *Pyricularia* strain isolated from *Setaria viridis*; **MoEi:** *Pyricularia* strain isolated from *Eleusine indica*; **PoOs:** *Pyricularia* strain isolated from *Oryza sativa*; **PoTa:** *Pyricularia* strain isolated from *Triticum aestivum*; ***P. penniseti*:** *Pyrucularia penniseti*; ***P. giganteum*:** *Pennisetum giganteum*; ***P. penniseticola*:** *Pyricularia penniseticola*; ***P. setariae*:** *Pyricularia setariae*; ***P. typhoides*:** *Pennisetum typhoides*; ***AVR* gene:** avirulence gene; **CAZymes:** carbohydrate-active enzymes; **CBM:** Carbohydrate-Binding Module Family; chr: chromosome; **CM:** complete medium; **CTAB:** Cetyltrimethyl Ammonium Bromide; **ctg:** contig; **ETI:** effetor-triggered immunity; **GH:** glycosyl hydrolase family pectinases; **LTR:** long terminal repeats; **MYA:** million years ago; **PHI:** Pathogen-Host Interaction; **PL:** Polysaccharide Lyase Family; **PTI:** pathogen-associated molecular pattern (PAMPs)-triggered immunity; ***R* genes:** resistance gene; **SMRT:**

Single-Molecule Real-Time; **TE:** transposable elements.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Genome assembly and PacBio reads are available in GenBank under BioProject PRJNA416656. The Whole Genome sequence has been deposited at GenBank under the accession PELF00000000.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Natural Science Foundations of China to Z.W (31770156), the Scientific Research Foundation of the Graduate School of FAFU to Z.Z, and the State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops to H.Z (SKB2017002).

Authors' contributions

The study was conceived and designed by ZW and GL. The initial collection and culturing of the strain was performed by HZ, XC, HH, MS, LZ, TF, YZ, JG, LZ, JF and HL. Bioinformatics was performed by ZZ, and LL. ZZ, HZ, JN, GL and ZW wrote, revised and approved the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Dr. Sanzhen Liu at Kansas State University for critically reading the manuscript.

Additional files

Additional files 1: Table S1. P1609_vs_7015_unique_secreted

Additional files 2: Table S2. 70-15 VS P1609 unique secreted proteins

Additional files 3: Table S3. Predicted PHI in P1609

Additional files 4: Figure S1. Gene copy numbers of CAZYmes in *Botrytis cinereal* (Bc), *Colletotrichum gloeosporioides* (Cg), *Fusarium graminearum* (Fg), *Neurospora crassa* (Nc), *Sclerotinia sclerotiorum* (Ss), *Trichoderma reesei* (Tr) and *Ustilago maydis* (Um) as well as *Pyricularia* isolates collected from *O. sativa* (PoOs), *T. aestivum* (PoTa), *D. sanguinalis* (PoDs), *S. viridis* (PoSv), and *E. indica* (PoEi). Log2 Copy Number presents variation of copy number with increased red color means increased number of CAZymes.

Additional files 5: Figure S2. GH28 of P1609 (P1609_11576, P1609_5879, P1609_2497, P1609_5781, P1609_680 and P1609_5514)) and PoOs (MGG_09608, MGG_08752 and MGG_08938), PoDs (Ds0505_9820). Extra copies of GH28 in P1609 is marked by blue.

References

1. Saccardo P. Fungorum Extra-Europaeorum pugillus. Michelia. 1880; 2(6):136-149.
2. Gladieux P, Condon B, Ravel S, Soanes D, Leodato Nunes Maciel J, Nhani J, Antonio. , Chen L, Terauchi R, Lebrun M-H, Tharreau D et al. Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. mBio. 2017:9:e01219-01217.
3. Igarashi S, Utiamada C, Igarashi L, Kazuma A, Lopes R. *Pyricularia* em trigo. 1. Ocorrência de Pyricularia sp. no estado do Paraná. Fitopatologia

444 Brasileira. 1986; 11:351-352.

445 4. Wilson RA, Talbot NJ. Under pressure: investigating the biology of plant
446 infection by *Magnaporthe oryzae*. Nat Rev Microbiol. 2009; 7(3):185-195.

447 5. Klaubauf S, Tharreau D, Fournier E, Groenewald JZ, Crous PW, de Vries
448 RP, Lebrun MH. Resolving the polyphyletic nature of *Pyricularia*
449 (Pyriculariaceae). Stud Mycol. 2014; 79:85-120.

450 6. Lenne JM. A world list of fungal diseases of tropical pasture species.
451 Australasian Plant Pathology. 1991; 20(3):122-124.

452 7. Chao CCT, Ellingboe AH. Selection for mating competence in
453 *Magnaporthe grisea* pathogenic to rice. Canadian Journal of Botany. 1991;
454 69(10):2130-2134.

455 8. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ,
456 Thon M, Kulkarni R, Xu JR, Pan H et al. The genome sequence of the rice
457 blast fungus *Magnaporthe grisea*. Nature. 2005; 434(7036):980-986.

458 9. Yan X, Talbot NJ. Investigating the cell biology of plant infection by the rice
459 blast fungus *Magnaporthe oryzae*. Curr Opin Microbiol. 2016; 34:147-153.

460 10. Bao J, Chen M, Zhong Z, Tang W, Lin L, Zhang X, Jiang H, Zhang D, Miao
461 C, Tang H. PacBio Sequencing Reveals Transposable Element as a Key
462 Contributor to Genomic Plasticity and Virulence Variation in *Magnaporthe*
463 *oryzae*. Mol Plant. 2017:DOI: 10.1016/j.molp.2017.1008.1008.

464 11. Chen C, Lian B, Hu J, Zhai H, Wang X, Venu RC, Liu E, Wang Z, Chen M,
465 Wang B et al. Genome comparison of two *Magnaporthe oryzae* field

466 isolates reveals genome variations and potential virulence effectors. BMC
467 Genomics. 2013; 14:887.
468 12. Dong Y, Li Y, Zhao M, Jing M, Liu X, Liu M, Guo X, Zhang X, Chen Y, Liu Y
469 et al. Global genome and transcriptome analyses of *Magnaporthe oryzae*
470 epidemic isolate 98-06 uncover novel effectors and pathogenicity-related
471 genes, revealing gene gain and lose dynamics in genome evolution. PLoS
472 Pathog. 2015; 11(4):e1004801.
473 13. Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, Zhao W, Shen M, Zhang H, Li
474 C et al. Comparative analysis of the genomes of two field isolates of the
475 rice blast fungus *Magnaporthe oryzae*. PLoS Genet. 2012; 8(8):e1002869.
476 14. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Tosa Y,
477 Chuma I, Takano Y, Win J, Kamoun S et al. Association genetics reveals
478 three novel avirulence genes from the rice blast fungal pathogen
479 *Magnaporthe oryzae*. Plant Cell. 2009; 21(5):1573-1591.
480 15. Gladieux P, Ravel S, Rieux A, Cros-Arteil S, Adreit H, Milazzo J, Thierry M,
481 Fournier E, Terauchi R, Tharreau D. Coexistence of Multiple Endemic and
482 Pandemic Lineages of the Rice Blast Pathogen. mBio. 2018; 9(2).
483 16. Liao J, Huang H, Meusnier I, Adreit H, Ducasse A, Bonnot F, Pan L, He X,
484 Kroj T, Fournier E et al. Pathogen effectors and plant immunity determine
485 specialization of the blast fungus to rice subspecies. Elife. 2016; 5:DOI:
486 10.7554/eLife.19377.
487 17. Zhong Z, Chen M, Lin L, Han Y, Bao J, Tang W, Lin Y, Somai R, Lu L,

Zhang W et al. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. ISME J. 2018.

18. Zhong Z, Justice N, Chen M, Bao J, Lin L, Chen L, Lin Y, Wu X, Cai Z, Zhang Q. Directional selection from host plants is a major force driving host specificity in *Magnaporthe species*. Sc Rep. 2016; 6:25591. DOI: 25510.21038/srep25591.

19. Hu H, Zheng H, Yang T, Chen X, Ye W, Lu G, Lin Z, Wang Z. First report of *Pyricularia* leaf spot on *Pennisetum giganteum* (JUJUNCAO) in China. Plant disease. 2018:Submitted.

20. Lin X, Lin Z, Lin D, Lin H, Luo H, Hu Y, Lin C, Zhu C. Effects of Planting *Pennisetum* sp. (Giant Juncao) on Soil Microbial Functional Diversity and Fertility in the Barren Hillside. Acta Ecologica Sinica. 2014; 34(15):4304-4312.

21. Lin ZX, Lin DM, Su DW, Lin H, Jing L, Zheng D, Yu S-K. Effect of Different Salt-affected Soils on Biological Characteristics of *Pennisetum* sp.. Southwest Chin J Agr Sci. 2015; 28(2):675-680.

22. Shi J, Lin ZX, Lin DM, De-Wei SU, Luo HL, Lin XS, Lin ZS, Zheng D, Chen JH, Yao JX. Enzymolysis conditions of *Pennisetum* sp. cellulose. Pratacultural Sci. 2014; 31(4):760-765.

23. Zhu D, Wang P, Lin X, Lin H, Dewei SU, Lin Z. Karyotype analysis of *Pennisetum purpureum* and *Pennisetum giganteum*. Guizhou Agr Sci.

2015; 43(5):14-18.

24. Prasada R, Goyal JP. A new species of *Pyricularia* on Bajra. Cur Sci. 1970; 39(12).

25. Muszewska A, Hoffmansommer M, Grynberg M. LTR Retrotransposons in Fungi. Plos One. 2011; 6(12):e29425. DOI: 29410.21371/journal.pone. 0029425.

26. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 2012; 10(6):417-430.

27. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015; 16:157. DOI: 110.1186/s13059-13015-10721-13052.

28. Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, Notteghem JL, Kohn LM. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics. 2005; 170(2):613-630.

29. Hedges SB. The origin and evolution of model organisms. Nat Rev Genet. 2002; 3(11):838-849.

30. Luo J, Qiu H, Cai G, Wagner NE, Bhattacharya D, Zhang N. Phylogenomic analysis uncovers the evolutionary history of nutrition and infection mode in rice blast fungus and other Magnaporthales. Sci Rep. 2015; 5:9448.

31. Chiapello H, Mallet L, Guerin C, Aguileta G, Amselem J, Kroj T, Ortega-Abboud E, Lebrun MH, Henrissat B, Gendrault A et al. Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. Genome Biol Evol. 2015; 7(10):2896-2912.

32. Kubicek CP, Starr TL, Glass NL. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. Annu Rev Phytopathol. 2014; 52(1):427.

33. Zhao Z, Liu H, Wang C, Xu JR. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 2013; 14:274.

34. Jeon J, Park SY, Chi MH, Choi J, Park J, Rho HS, Kim S, Goh J, Yoo S, Park JY et al. Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. Nat Genet. 2007; 39(4):561-565.

35. Wu D, Oide S, Zhang N, Choi MY, Turgeon BG. ChLae1 and ChVel1 regulate T-toxin production, virulence, oxidative stress response, and development of the maize pathogen *Cochliobolus heterostrophus*. PLoS Pathog. 2012; 8(2):e1002542.

36. de Jonge R, Bolton MD, Kombrink A, van den Berg GC, Yadeta KA, Thomma BP. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 2013; 23(8):1271-1282.

37. Hartmann FE, Sanchez-Vallet A, McDonald BA, Croll D. A fungal wheat

554 pathogen evolved host specialization by extensive chromosomal
555 rearrangements. ISME J. 2017; 11(5):1189-1204.
556 38. Sprockett DD, Piontkivska H, Blackwood CB. Evolutionary analysis of
557 glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions
558 in necrotrophic fungal pathogens. Gene. 2011; 479(1-2):29-36.
559 39. Yamashita N, Motoyoshi T, Nishimura A. Purification and Characterization
560 of Isoamyl Alcohol Oxidase ("Mureka"-Forming Enzyme). Biosci
561 Biotechnol Biochem. 1999; 63(7):1216-1222.
562 40. Kern K, Nunn CD, Pichova A, Dickinson JR. Isoamyl alcohol-induced
563 morphological change in *Saccharomyces cerevisiae* involves increases in
564 mitochondria and cell wall chitin content. FEMS Yeast Res. 2004;
565 5(1):43-49.
566 41. DeSouza L, Shen Y, Bognar AL. Disruption of cytoplasmic and
567 mitochondrial folylpolyglutamate synthetase activity in *Saccharomyces*
568 *cerevisiae*. Arch Biochem Biophys. 2000; 376(2):299-312.
569 42. Varela JC, Mager WH. Response of *Saccharomyces cerevisiae* to
570 changes in external osmolarity. Microbiology. 1996; 142 (Pt 4):721-731.
571 43. Foster AJ, Ryder LS, Kershaw MJ, Talbot NJ. The role of glycerol in the
572 pathogenic lifestyle of the rice blast fungus *Magnaporthe oryzae*. Environ
573 Microbiol. 2017; 19(3):1008-1016.
574 44. Mercure EW, Kunoh H, Nicholson RL. Adhesion of *Colletotrichum*
575 *graminicola* conidia to corn leaves, a requirement for disease development.

576 Physiological & Molecular Plant Pathology. 1995; 45(6):407-420.

577 45. Sela-Buurlage MB, Epstein L, Rodriguez RJ. Adhesion of ungerminated
578 *Colletotrichum musae* conidia. Physiological & Molecular Plant Pathology.
579 1991; 39(5):345-352.

580 46. Young DH, Kauss H. Adhesion of *Colletotrichum lindemuthianum* spores to
581 *Phaseolus vulgaris* hypocotyls and to polystyrene. Applied &
582 Environmental Microbiology. 1984; 47(4):616-619.

583 47. Hamer JE, Howard RJ, Chumley FG, Valent B. A mechanism for surface
584 attachment in spores of a plant pathogenic fungus. Science. 1988;
585 239(4837):288-290.

586 48. Thompson BM, Hoelscher BC, Driks A, Stewart GC. Assembly of the BclB
587 glycoprotein into the exosporium and evidence for its role in the formation
588 of the exosporium 'cap' structure in *Bacillus anthracis*. Molecular
589 Microbiology. 2012; 86(5):1073.

590 49. Thompson BMWLN, Fox KF. The BdB glycoprotein of *Bacillus anthracis* is
591 involved in exosporium integrity. Journal of Bacteriology. 2007;
592 189(18):6704.

593 50. Wang Y, Jenkins SA, Gu C, Shree A, Martinez-Moczygemba M, Herold J,
594 Botto M, Wetsel RA, Xu Y. *Bacillus anthracis* spore surface protein BclA
595 mediates complement factor H binding to spores and promotes spore
596 persistence. PLoS Pathog. 2016; 12(6):e1005678.

597 51. Schulze-Lefert P, Panstruga R. A molecular evolutionary concept

connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends Plant Sci. 2011; 16(3):117-125.

52. Mistry J, Bateman A, Finn RD. Predicting active site residue annotations in the Pfam database. BMC Bioinformatics. 2007; 8:298. DOI: 210.1186/1471-2105-1188-1298.

53. Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, Sadanadan V, Khamari L, Billal S, Mohanty S et al. PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database. Nucleic Acids Res. 2017; 45(D1):D604-D610.

54. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Kohler J, Hammond-Kosack KE. PHI-base: a new database for pathogen host interactions. Nucleic Acids Res. 2006; 34:D459-D464.

55. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012; 40(Web Server issue):W445-451.

56. Tarailo-Graovac M, Chen N. UNIT 4.10Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. 2009; Chapter 4(Unit 4):4.10.11-14.10.14.

57. Amselem J, Cuomo CA, van Kan JA, Viaud M, Benito EP, Couloux A, Coutinho PM, de Vries RP, Dyer PS, Fillinger S et al. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. 2011; 7(8):e1002230. DOI:

1002210.1001371/journal. pgen.1002230.

58. Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of Colletotrichum fungi. New Phytol. 2013; 197(4):1236-1249.

59. Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Pietro AD, Walton JD, Ma LJ, Baker SE, Rep M. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science. 2007; 317(5843):1400-1402.

60. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S et al. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature. 2003; 422(6934):859-868.

61. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol. 2008; 26(5):553-560.

62. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature. 2006; 444(7115):97-101.

63. Kato H, Yamamoto M, Yamaguchiozaki T, Kadouchi H, Iwamoto Y,

642 Nakayashiki H, Tosa Y, Mayama S, Mori N. Pathogenicity, mating ability
643 and DNA restriction fragment length polymorphisms of *Pyricularia*
644 populations isolated from Gramineae, Bambusideae and Zingiberaceae
645 plants. J Gen Plant Pathol. 2000; 66(1):30-47.
646 64. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular
647 evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;
648 30(12):2725-2729.
649 65. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and
650 divergence times in the absence of a molecular clock. Bioinformatics. 2003;
651 19(2):301-302.
652 66. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B,
653 Guo H et al. MCScanX: a toolkit for detection and evolutionary analysis of
654 gene synteny and collinearity. Nucleic Acids Res. 2012; 40(7):e49. DOI:
655 10.1093/nar/gkr1293.

656

657 **FIGURE LEGENDS**

658 Table 1 Details of sequencing reads and genome assembly of P1609

659

660 Table 2 Details of genome annotation of P1609

661

662 Table 3 Positively selected genes in P1609

663

664 Table 4 Unique Pathogen Host Interaction (PHI) genes in P1609.

665

666 Fig. 1 PacBio sequencing and genome assembly of P1609.

667 (A) Reads length distribution.

668 (B) Contig length of assembled contig > 100 Kb.

669 (C) Overview of P1609 genome. (Track a) Contig1 to contig16 of P1609, (track

670 b) gene density, (track c) Transposon elements density, (track d) secreted

671 proteins density, (track e) unique gene (compared with *P. oryzae*, *N. crassa*, *F.*

672 *graminearum* and *C. gloeosporioides*) density of P1609 per 50 Kb.

673

674 Fig. 2 Phylogenetic and comparative genomic study of P1609.

675 (A) Phylogenomic tree of P1609 with *Botrytis cinereal* (Bc), *Colletotrichum*

676 *gloeosporioides* (Cg), *Fusarium graminearum* (Fg), *Neurospora crassa* (Nc),

677 *Sclerotinia sclerotiorum* (Ss), *Trichoderma reesei* (Tr) and *Ustilago maydis*

678 (Um) as well as *Pyricularia* isolates collected from *O. sativa* (PoOs), *T.*

679 *aestivum* (PoTa), *D. sanguinalis* (PoDs), *S. viridis* (PoSv), and *E. indica* (PoEi)

680 based on 2,051 single copy genes. The values of all of the branches are 100.

681 (B) Maximum likelihood tree of P1609 and 28 Magnaporthales species, as well

682 as 5 Sordariomycetes used as outgroup species based on 82,715 amino acid

683 positions derived from 226 genes.

684 (C) Venn diagram showed an overlap of gene families among P1609,

685 *Pyricularia* rice isolates (PoOs), *C. gloeosporioides* (Cg), *F. graminearum* (Fg)

686 and *N. crassa* (Nc).

687

688 Fig. 3 Chromosome rearrangement and splitting between P1609 and 70-15.

689 (A) Bar plot showing the chromosomes in 70-15.

690 (B) Bar plot showing the assembled contigs of P1609. Colinear chromosomes

691 of 70-15 and contigs of P1609 are indicated by the same color.

692 (C) Dual synteny plot showing splitting of Ctg2 of P1609 into chr. 4 and chr. 7 in

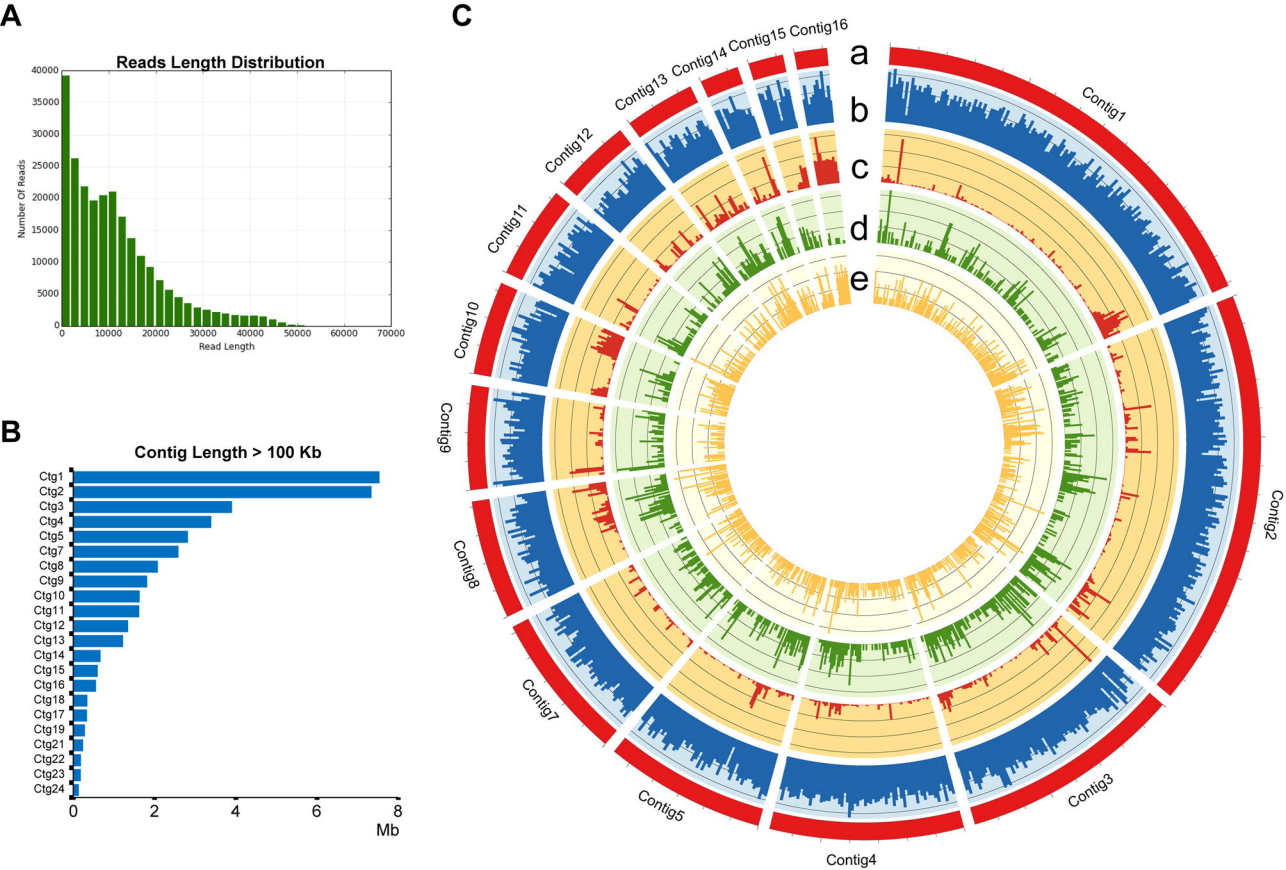
693 70-15.

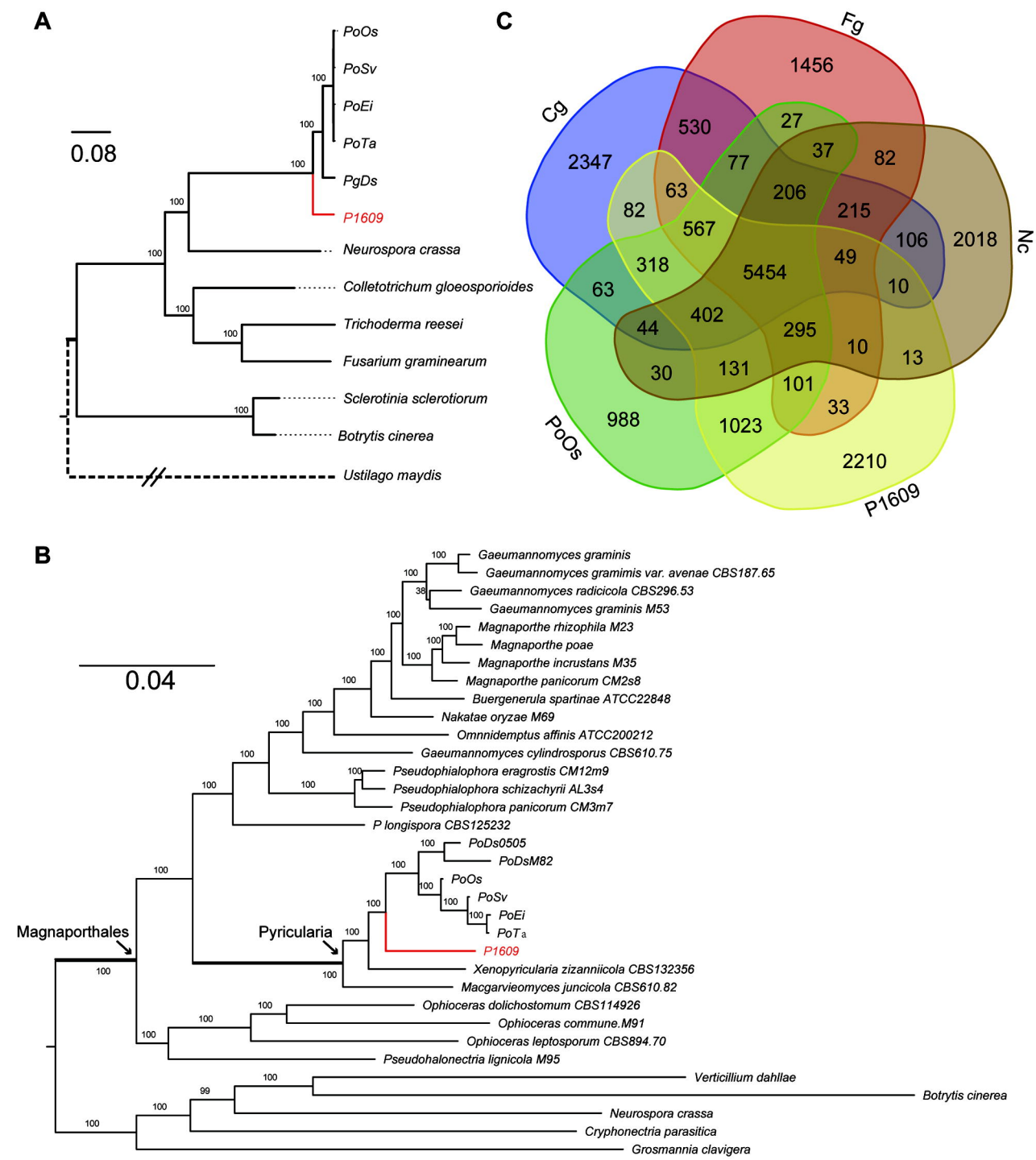
694

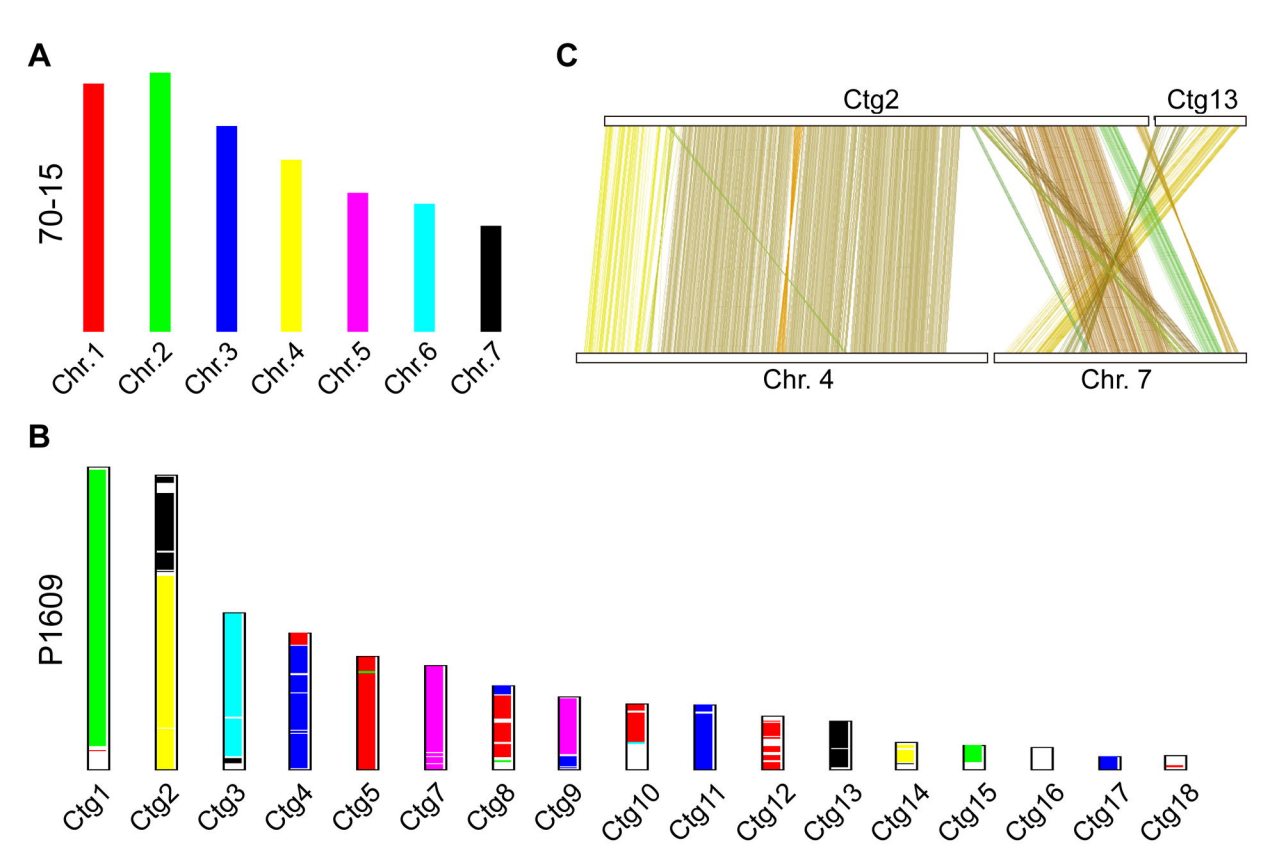
695 Fig. 4 PacBio long-read coverage from 2.02 to 2.94 Mb of Ctg 2. Color of reads

696 indicate different read lengths.

697


A
Reads Length Distribution
Number Of Reads
Read Length
B
Contig Length > 100 Kb
Ctg1
Ctg2
Ctg3
Ctg4
Ctg5
Ctg7
Ctg8
Ctg9
Ctg10
Ctg11
Ctg12
Ctg13
Ctg14
Ctg15
Ctg16
Ctg18
Ctg17
Ctg19
Ctg21
Ctg22
Ctg23
Ctg24
Mb
C
a
b
c
d
e
Contig1
Contig2
Contig3
Contig4
Contig5
Contig7
Contig8
Contig9
Contig10
Contig11
Contig12
Contig13
Contig14
Contig15
Contig16


A
PoOs
PoSv
PoEi
PoTa
PgDs
P1609
Neurospora crassa
Colletotrichum gloeosporioides
Trichoderma reesei
Fusarium graminearum
Sclerotinia sclerotiorum
Botrytis cinerea
Ustilago maydis
0.08
C
Fg
Cg
Nc
PoOs
P1609
1456
2347
2018
2210
988
1023
5454
530
27
37
82
77
63
206
215
82
567
106
318
49
63
10
402
44
295
10
13
30
131
101
33
B
0.04
Gaeumannomyces graminis
Gaeumannomyces gramimis var. avenae CBS187.65
Gaeumannomyces radicicola CBS296.53
Gaeumannomyces graminis M53
Magnaporthe rhizophila M23
Magnaporthe poae
Magnaporthe incrustans M35
Magnaporthe panicorum CM2s8
Buergenerula spartinae ATCC22848
Nakatae oryzae M69
Omnnidemptus affinis ATCC200212
Gaeumannomyces cylindrosporus CBS610.75
Pseudophialophora eragrostis CM12m9
Pseudophialophora schizachyrii AL3s4
Pseudophialophora panicorum CM3m7
P longispora CBS125232
PoDs0505
PoDsM82
PoOs
PoSv
PoEi
PoTa
P1609
Xenopyricularia zizanniicola CBS132356
Macgarvieomyces juncicola CBS610.82
Ophioceras dolichostomum CBS114926
Ophioceras commune.M91
Ophioceras leptosporum CBS894.70
Pseudohalonectria lignicola M95
Verticillium dahliae
Botrytis cinerea
Neurospora crassa
Cryphonectria parasitica
Grosmannia clavigera
Magnaporthales
Pyricularia


A
70-15
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
C
Ctg2
Ctg13
Chr. 4
Chr. 7
B
P1609
Ctg1
Ctg2
Ctg3
Ctg4
Ctg5
Ctg7
Ctg8
Ctg9
Ctg10
Ctg11
Ctg12
Ctg13
Ctg14
Ctg15
Ctg16
Ctg17
Ctg18

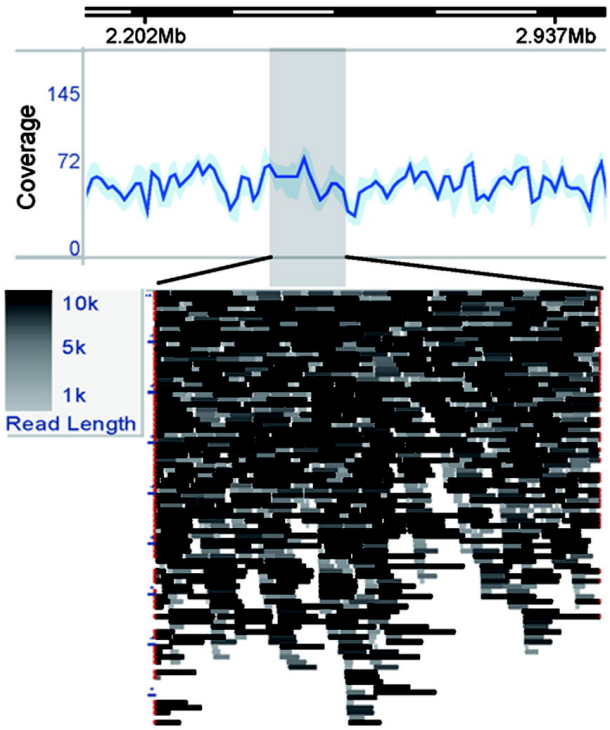

2.202Mb
2.937Mb
Coverage
145
72
0
10k
5k
1k
Read Length

Table 1 Details of sequencing reads and genome assembly of P1609

Feature	**Value**
Number of Subreads	312,061
Total Length of Subreads (bp)	2,688,966,115
Mean Coverage	59
Polished Contigs	53
Maximum Contig Length (bp)	7,555,856
N50 Contig Length (bp)	3,411,241
Sum of Contig Lengths (Mb)	41.82
GC level	50.30%
length > 1 Mb	89.70%
length > 100 Kb	98.50%

Table 2 Details of genome annotation of P1609

Category	Value
Total TE	7.67%
LINEs	1.50%
LTR elements	4.27%
DNA elements	0.63%
Unclassified	1.27%
Gene Number	13102
Average Gene Length (bp)	1758
Secreted Protein Number	1409

Table 3 Positively selected genes in P1609

Protein ID	Ka	Ks	Ka/Ks	P-Value (Fisher)	Description
P1609_7184	0.707979	0.440748	1.60631	0.00697146	Unknown
P1609_5032	0.550905	0.343609	1.60329	1.98443E-06	Isoamyl alcohol oxidase
P1609_3360	0.183224	0.118517	1.54597	0.00293936	Glycerol uptake protein 1
P1609_791	0.302231	0.199786	1.51277	9.75093E-06	Folylpolyglutamate synthase
P1609_7869	0.491873	0.328109	1.49911	0.0296442	Thioredoxin reductase
P1609_1006	0.40206	0.275456	1.45962	0.0288585	Spore surface glycoprotein BclB

Table 4 Unique Pathogen Host Interaction (PHI) genes in P1609

Protein ID	PHI ID	Reference Organism	Phenotypes
P1609_2497_494aa	PHI:115	*Cochliobolus carbonum*	Unaffected pathogenicity
P1609_11506_262aa, P1609_12781_266aa, P1609_4501_233aa, P1609_683_218aa	PHI:698	*Vibrio cholerae*	Reduced virulence
P1609_11592_161aa, P1609_4614_892aa	PHI:1284	*Fusarium graminearum*	Unaffected pathogenicity
P1609_11335_610aa	PHI:1803	*Fusarium graminearum*	Unaffected pathogenicity
P1609_9374_326aa	PHI:2394	*Fusarium graminearum*	Increased virulence
P1609_1007_600aa	PHI:225	*Fusarium solani*	Reduced virulence
P1609_13007_833aa	PHI:1714	*Fusarium graminearum*	Lethal
P1609_3609_435aa, P1609_5044_426aa, P1609_8181_460aa, P1609_960_144aa	PHI:1455	*Fusarium graminearum*	Unaffected pathogenicity
P1609_11548_383aa, P1609_5843_397aa	PHI:1823	*Fusarium graminearum*	Unaffected pathogenicity
P1609_2665_64aa	PHI:1421	*Fusarium graminearum*	Lethal
P1609_2607_206aa, P1609_2608_820aa	PHI:3418	*Staphylococcus saprophyticus*	Mixed outcome
P1609_11380_688aa	PHI:1809	*Fusarium graminearum*	Lethal
P1609_7767_257aa	PHI:317	*Aspergillus fumigatus*	Reduced virulence
P1609_6736_447aa	PHI:323	*Verticillium fungicola*	Reduced virulence
P1609_13_169aa	PHI:1147	*Fusarium pseudograminearum*	Unaffected pathogenicity
P1609_2376_825aa	PHI:1871	*Fusarium graminearum*	Lethal
P1609_10320_326aa	PHI:1522	*Fusarium graminearum*	Lethal
P1609_11485_355aa, P1609_2130_433aa	PHI:3116	*Pseudomonas syringae*	Mixed outcome
P1609_10203_738aa	PHI:1815	*Fusarium graminearum*	Unaffected pathogenicity
P1609_11600_2486aa	PHI:2628	*Salmonella enterica*	Reduced virulence, 663
P1609_11388_1000aa	PHI:3076	*Candida parapsilosis*	Mixed outcome
P1609_11869_369aa	PHI:2375	*Alternaria brassicicola*	Mixed outcome
P1609_5035_383aa	PHI:233	*Cochliobolus carbonum*	Reduced virulence
P1609_5702_488aa	PHI:1271	*Fusarium graminearum*	Unaffected pathogenicity
P1609_11864_596aa	PHI:2380	*Alternaria brassicicola*	Mixed outcome