

RESEARCH ARTICLE

Open Access



Comparative genome analyses of four rice-infecting *Rhizoctonia solani* isolates reveal extensive enrichment of homogalacturonan modification genes

Da-Young Lee^{1†}, Jongbum Jeon^{2,3†}, Ki-Tae Kim⁴, Kyeongchae Cheong^{2,3}, Hyeunjeong Song^{2,3}, Gobong Choi^{2,3}, Jaeho Ko^{2,5}, Stephen O. Opiyo⁶, James C. Correll⁷, Shimin Zuo⁸, Sheshu Madhav⁹, Guo-Liang Wang^{1*}  and Yong-Hwan Lee^{2,3,5,10*}

Abstract

Background: Plant pathogenic isolates of *Rhizoctonia solani* anastomosis group 1-intraspecific group IA (AG1-IA) infect a wide range of crops causing diseases such as rice sheath blight (ShB). ShB has become a serious disease in rice production worldwide. Additional genome sequences of the rice-infecting *R. solani* isolates from different geographical regions will facilitate the identification of important pathogenicity-related genes in the fungus.

Results: Rice-infecting *R. solani* isolates B2 (USA), ADB (India), WGL (India), and YN-7 (China) were selected for whole-genome sequencing. Single-Molecule Real-Time (SMRT) and Illumina sequencing were used for de novo sequencing of the B2 genome. The genomes of the other three isolates were then sequenced with Illumina technology and assembled using the B2 genome as a reference. The four genomes ranged from 38.9 to 45.0 Mbp in size, contained 9715 to 11,505 protein-coding genes, and shared 5812 conserved orthogroups. The proportion of transposable elements (TEs) and average length of TE sequences in the B2 genome was nearly 3 times and 2 times greater, respectively, than those of ADB, WGL and YN-7. Although 818 to 888 putative secreted proteins were identified in the four isolates, only 30% of them were predicted to be small secreted proteins, which is a smaller proportion than what is usually found in the genomes of cereal necrotrophic fungi. Despite a lack of putative secondary metabolite biosynthesis gene clusters, the rice-infecting *R. solani* genomes were predicted to contain the most carbohydrate-active enzyme (CAZyme) genes among all 27 fungal genomes used in the comparative analysis. Specifically, extensive enrichment of pectin/homogalacturonan modification genes were found in all four rice-infecting *R. solani* genomes.

(Continued on next page)

* Correspondence: wang.620@osu.edu; yonglee@snu.ac.kr

[†]Da-Young Lee and Jongbum Jeon contributed equally to this work.

¹Department of Plant Pathology, The Ohio State University, Columbus, OH 43210, USA

²Fungal Bioinformatics Laboratory, Seoul National University, Seoul 08826, South Korea

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



© The Author(s). 2021, corrected publication 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusion: Four *R. solani* genomes were sequenced, annotated, and compared to other fungal genomes to identify distinctive genomic features that may contribute to the pathogenicity of rice-infecting *R. solani*. Our analyses provided evidence that genomic conservation of *R. solani* genomes among neighboring AGs was more diversified than among AG1-IA isolates and the presence of numerous predicted pectin modification genes in the rice-infecting *R. solani* genomes that may contribute to the wide host range and virulence of this necrotrophic fungal pathogen.

Keywords: *Rhizoctonia solani* AG1-IA, Rice sheath blight, Plant cell wall degrading enzymes, Homogalacturonan/pectin modification genes

Background

First reported in Japan in 1910 [1, 2], rice sheath blight (ShB) is one of the most devastating fungal diseases threatening rice production worldwide [3–5]. The causal agent of ShB is the soil-borne, necrotrophic fungi *Rhizoctonia solani* Kühn [teleomorph: *Thanatephorus cucumeris* (A.B. Frank) Donk], which belongs to the division Basidiomycota and subdivision Agaricomycotina. *R. solani* is a species complex that is classified into 13 anastomosis groups (AGs) based on the ability of genetically similar isolates to undergo hyphal fusion (anastomosis) [6–8]. Isolates in each AG are further categorized into intraspecific groups (ISGs) based on differences in host range, pathogenicity, cultural morphology, and biochemical characteristics [6, 8–10]. For instance, isolates of ISG IA belonging to AG1 (AG1-IA) can infect members of the Poaceae family including rice, maize, and turfgrass [11–13] to cause ShB, banded leaf sheath blight, and brown patch disease, respectively [3, 6, 14, 15]. Multiple studies have used genomic [13, 16–23], transcriptomic [16, 24] and proteomic approaches [25, 26] to examine the molecular basis for *R. solani* pathogenesis. A 36.94-Mbp draft genome sequence for *R. solani* AG1-IA isolated from infected rice in South China was assembled into 2648 scaffolds (for an N₅₀ scaffold size of 474.5 Kb) in which 6156 genes were annotated [16]. However, more high-quality genome sequences from multiple rice-infecting isolates are needed to robustly identify conserved genomic signatures of rice-infecting *R. solani* AG1-IA.

Pathogenic fungi secrete various proteins to promote successful infection by suppressing host defenses and/or manipulating the physiology of host cells [27, 28]. Accordingly, this suite of proteins determines both the lifestyle and host ranges of these fungi [27]. A subset of these proteins is cysteine-rich, small (≤ 300 amino acids) secreted proteins (SSPs) called effectors [29, 30] that can form disulfide bridges to stabilize their tertiary structure, making them more resistant to degradation [31–33]. Plant pathogenic fungi also secrete carbohydrate-active enzymes (CAZymes) that allow them to breach the plant cell wall and enter their hosts [34–36]. These CAZymes

are categorized into modules: glycoside hydrolases (GHs), carbohydrate esterases (CEs), auxiliary activities (AAs), glycosyltransferases (GTs), polysaccharide lyases (PLs), and (non-catalytic) carbohydrate-binding modules (CBMs) [36, 37].

This study aimed to sequence four genomes of *R. solani* isolated from ShB-infected rice and conducted comparative genome analyses amongst them (*R. solani* AG1-IA) as well as to 4 *R. solani* genomes belonging to AGs aside from AG1-IA (AG1-IB, AG2, AG3 and AG8). We hypothesized that rice-infecting *R. solani* would possess a large arsenal of cell wall-degrading genes to support its necrotrophic lifestyle and host range. To test this hypothesis, we assembled high-quality genome sequences for four rice-infecting *R. solani* strains that were isolated from rice grown in diverse geographic regions of the world (USA, China, and India) and compared them to publicly available genomes belonging to *R. solani* AG1-IA, AG2-IB, AG2-2IIIB, AG3, and AG8 [13, 16, 17, 19, 20]. We also selected representative genomes encompassing different nutritional lifestyles and hosts from Basidiomycota (9 genomes) and Ascomycota (9 genomes) into our comparative analyses. In this study, pairwise whole-genome alignments suggest that macrosynteny exists among the rice-infecting *R. solani* genomes. This phylogenetic proximity is supported by a phylogenetic tree constructed using the maximum likelihood method as well as the existence of a larger set of core-orthogroups among rice-infecting AG1-IA genomes (5812 orthogroups) compared to core orthogroups of *R. solani* genomes from diverse *R. solani* AGs (3635 orthogroups). Comparative genome analyses also revealed that rice-infecting *R. solani* have a smaller set of SSPs compared to biotrophs and other necrotrophs (cereal). Conversely, rice-infecting *R. solani* genomes code for the highest number of CAZymes, which are predicted to be involved in plant cell wall modification and degradation. Specifically, all *R. solani* genomes used in this study, regardless of AG, were highly enriched in pectin-degrading genes, containing even more than the well-known pectin-degrading, necrotrophic fungus *Verticillium dahliae*. The high-quality genome

sequencing data and comparative genomic results from this study are useful resources for functional analysis of pathogenicity genes in this important fungal pathogen of rice.

Results

High-quality genome sequences for rice-infecting *R. solani* AG1-IA isolates

Four rice-infecting *R. solani* isolates were collected from rice grown in USA (B2), India (ADB, WGL), and China (YN-7) (Table 1). De novo sequencing of the B2 genome was achieved using a Single-Molecule Real-Time (SMRT; Pacific Biosciences); the WGL, ADB, and YN-7 genomes were sequenced using Illumina technology and subsequently assembled using the B2 genome as reference (Additional file 1: Fig. S1). B2 had the largest genome of the four isolates (45.01 Mbp; 96 scaffolds), while YN-7 had the smallest (38.92 Mbp; 413 scaffolds). ADB (39.90 Mbp; 811 scaffolds) and WGL (39.98 Mbp; 724 scaffolds) were intermediate in size. Accordingly, B2 had the highest number of protein-coding genes (11,505 genes), followed by WGL (10,044 genes), ADB (10,010 genes), and YN-7 (9715 genes). Despite these slight variations in genome size and gene numbers, all four genomes had similar total GC contents (47.3 to 47.7%). Based on scaffold numbers, which ranged from 96 to 811, and N_{50} values, which ranged from 1.22 Mbp to 1.56 Mbp, the newly sequenced *R. solani* AG1-IA B2 genome is of higher quality than the previously sequenced *R. solani* AG1-IA genome (2648 scaffolds; 36.94 Mbp) [16]. Thus, all comparative genomic analyses hereafter used the B2 genome as the representative for rice-infecting *R. solani* AG1-IA isolates.

Phylogenetic proximity of *R. solani* isolates

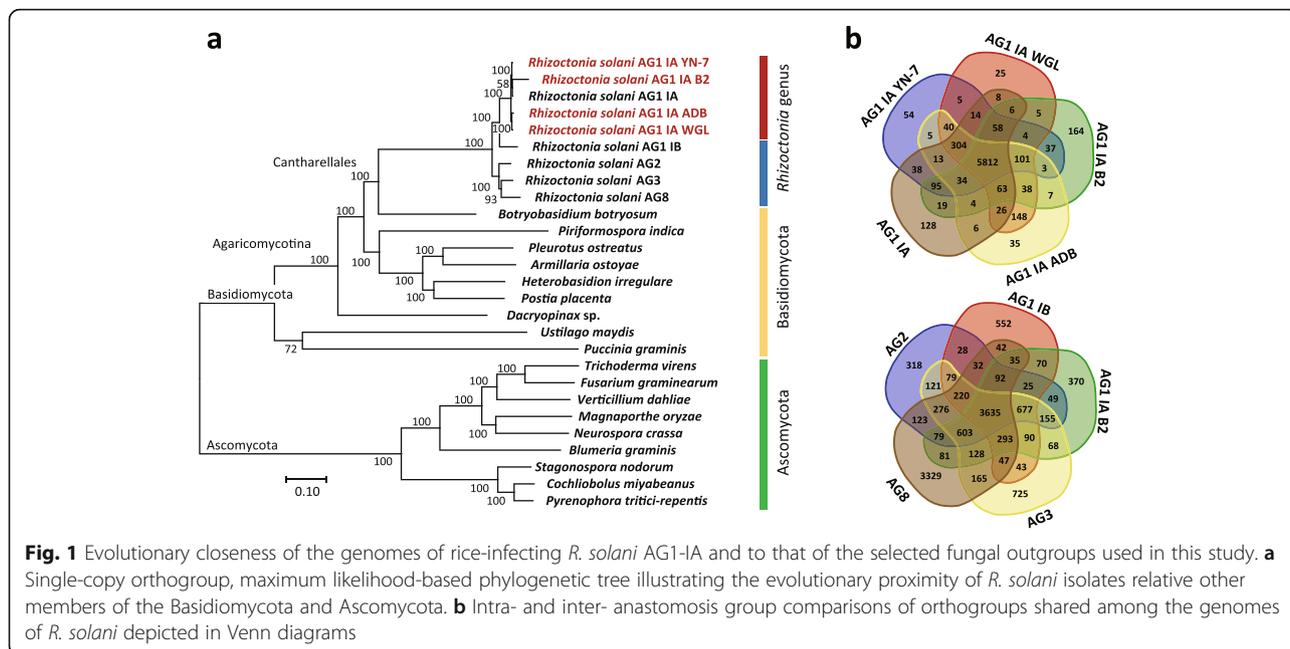
To evaluate the phylogenetic relationships of all 27 fungal genomes used in this study (Additional file 2: Table S1), we constructed a maximum likelihood-based phylogenetic tree using single-copy orthogroups (Fig. 1a). We confirmed that the four rice-infecting *R. solani* isolates were most phylogenetically related to each other and to the previously sequenced Chinese *R. solani* isolate, followed by AG1-IB, and finally the remaining *R. solani* AG isolates used in this study. In addition, comparison of the *R. solani* genomes indicated that the rice-infecting *R. solani* genomes share 5812 orthogroups of protein-coding genes while the genomes of different AG-ISG groups (AG1-IA B2, AG1-IB, AG2-2IIIB, AG3, and AG8) share only 3635 orthogroups (Fig. 1b). Of these orthogroups, 25 to 164 were specific to the genomes of rice-infecting *R. solani* AG1-IA while 318 to 3329 are AG-ISG specific.

Genome synteny between *R. solani* isolates

To determine the synteny between *R. solani* genomes, we performed pairwise whole-genome alignments using PROmer [38], a script-based pipeline to align multiple, divergent sequences and identify similar genomic regions based on the translation of all six reading frames. Large syntenic size (Fig. 2a) and diagonal dot plots (Additional file 3: Fig. S2) suggested that the five rice-infecting *R. solani* genomes (the four from this study and the previously sequenced Chinese isolate) share a high degree of genome conservation, ranging from 66 to 70.9% (Additional file 4: Table S2) when the B2 genome was used as reference for comparison. However, the degree of genome conservation between the B2 genome and those of different *R. solani* AGs (AG1-IB, AG2-

Table 1 Genome statistics of the four rice-infecting *Rhizoctonia solani* AG1-IA (B2, ADB, WGL and YN-7), AG1-IB, AG2-2IIIB, AG3 and AG8 isolates

<i>Rhizoctonia solani</i> AG-ISG	Isolate	Host	Origin of Isolate	Sequencing method	Genome size (Mbp)	Scaffold number	Scaffold N_{50} (bp)	G-C content (%)	Protein number	Reference
AG1-IA	B2	Rice	USA (Arkansas)	PacBio de novo	45.0095	96	1,561,158	47.32	11,505	This study
	ADB	Rice	India	Reference-based Illumina	39.9044	811	1,218,423	47.53	10,010	
	WGL	Rice	India	Reference-based Illumina	39.9757	724	1,240,390	47.54	10,044	
	YN-7	Rice	China	Reference-based Illumina	38.9167	413	1,349,706	47.71	9722	
AG1-1A		Rice	China	Illumina GA II	36.9381	2648	474,500	47.6	10,489	[16]
AG1-1B	7-3-14	Lettuce	Germany	GenomeSequencer (GS) FLX	48.6737	23,356	N/A	48.1	12,268	[17]
AG2-2IIIB	BBA69670	Sugar beet	Germany	Illumina MiSeq	56.0285	2065	81,152	35.88	11,897	[19]
AG3	Rhs1AP	Potato	USA (Maine)	Sanger and GS-FLX 454	51.7059	326	N/A	48.4	12,726	[20]
AG8	WAC10335	Lupin	Australia	Illumina	39.8229	857	160,500	48.7	13,952	[13]

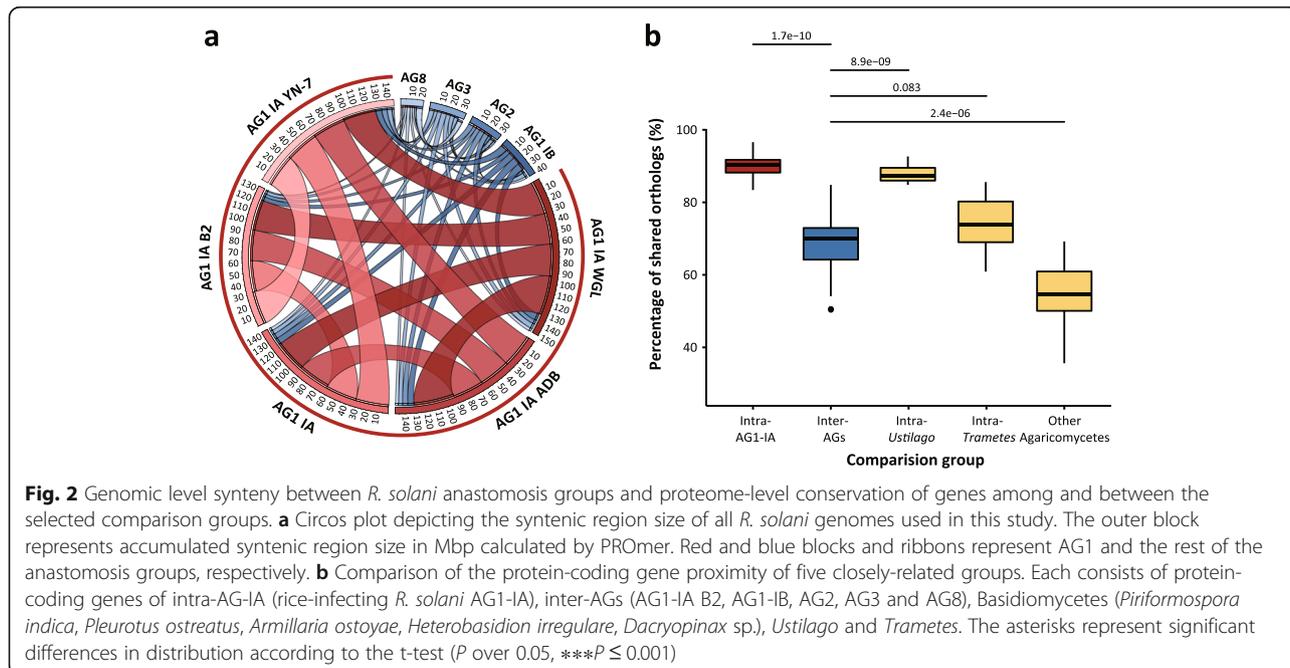


2IIIB, AG3, and AG8) dramatically decreases to 3.3 to 9.6%, suggesting that the AG1-IA genome is quite divergent from other AGs.

Protein-coding gene conservation among *R. solani* isolates

To determine the degree of proteome similarity in AGs, we performed pairwise ortholog clustering to compare each protein sequence of protein-coding genes. The

number of shared orthologs ranged from 8798 to 9723 in the protein-coding genes of *R. solani* AG1–1A isolates (Additional file 5: Table S3). The average proteome similarity of intra-AGs of *R. solani* was 89.69%. In contrast, the protein-coding gene similarity of inter-AGs of *R. solani* was more diversified, wherein the percentage of shared predicted proteomes in inter-AGs was averagely 68.36%. Moreover, in order to determine protein-coding gene similarity of genomes belonging to Basidiomycota,



we added four *Ustilago* and four *Trametes* genomes along with five Agaricomycetes genomes used for phylogenetic analyses (Additional file 6: Table S4). *R. solani* genomes of inter-AGs (compared with B2 strain) were larger than that of different Basidiomycota genus-group (54.53%). On the contrary, intra-genus group under Basidiomycota showed high protein-coding gene similarity compared to inter-AG (*Ustilago* intra-genus; 88.08%, *Trametes* intra-genus; 83.41%) (Fig. 2b). In comparison of single copy orthologs, discrepancy of inter AGs were increased (intra-AG1 IA; 63.89%, inter-AGs; 33.27%).

Transposable element profiles of the rice-infecting *R. solani* isolates

Transposable elements (TEs), such as class I retrotransposons and class II DNA transposons, can create temporary or permanent genomic rearrangements and modifications [39], and the abundance and frequency of these genetic elements can significantly influence the size of eukaryotic genomes [40]. To define the repetitive element profiles for the rice-infecting *R. solani* genomes, we analyzed the type and proportion of repetitive elements in each newly sequenced genome. The B2 genome contains the largest proportion of TEs (26.74%) compared to three other *R. solani* AG1-IA genomes; ADB (8.89%), WGL (9.16%), and YN-7 (6.18%), respectively (Additional file 7: Table S5). The number of total TEs for the ADB and WGL genomes are comparable (ADB: 10,421, WGL: 10,248), yet less than that of the B2 genome (17,123) and more than that of YN-7 (8030). Specifically, the B2 genome contains the highest proportion of DNA transposons, Long Terminal Repeats (LTRs) and Long Interspersed Nuclear Elements (LINEs) among the *R. solani* AG1-IA genomes, wherein the proportion of LTRs in B2 (20.14%) was identified to be more than 3 times of that of ADB (5.47%), WGL (5.65%) and YN-7 (3.84%). However, all four newly sequenced AG1-IA genomes as well as the previously sequenced AG1-IA genome possess lower numbers of TEs compared to AG1-IB, AG2-2IIIB, AG3 and AG8. In terms of average length of repetitive sequences, the PacBio-sequenced B2 genome contains approximately twice as long (797.4 bp) compared to the Illumina sequenced ADB (370 bp), WGL (385.6 bp), YN-7 (345.2 bp) in this study and other previously sequenced genomes of AG1-IA, AG1-IB, AG2-2IIIB, AG3 and AG8.

Predicted secretome of rice-infecting *R. solani* isolates

The putative secretomes of the rice-infecting *R. solani* isolates were analyzed, and 818 to 888 predicted secreted protein genes were identified (Fig. 3a, Additional file 8: Table S6). This suggests that rice-infecting *R. solani* isolates have a secretome that is intermediate in size smaller than those of necrotrophic ascomycetes (cereal)

but larger than those of brown-rot fungi *Postia placenta* and *Dacryopinax* sp. as well as biotrophs *Ustilago maydis* and *Blumeria graminis*. The number of small secreted proteins (SSPs; putative effectors) in the four genomes ranges from 263 to 279, accounting for 30–33% of each isolate's individual predicted secretome. We identified 367 *R. solani* specific orthogroups of SSPs. Among these specific SSPs, 12 (AG1-IB) to 105 (AG8) AG-specific SSPs were identified through ortholog comparison analysis of putative SSP gene sets (Additional File 9: Table S7), and the greatest number of specific SSPs were identified in AG8. We also observed that *Rhizoctonia* AG1-IA genomes have relatively small predicted protein-coding genes and SSPs are shown to be necrotrophic fungal groups (Fig. 3b).

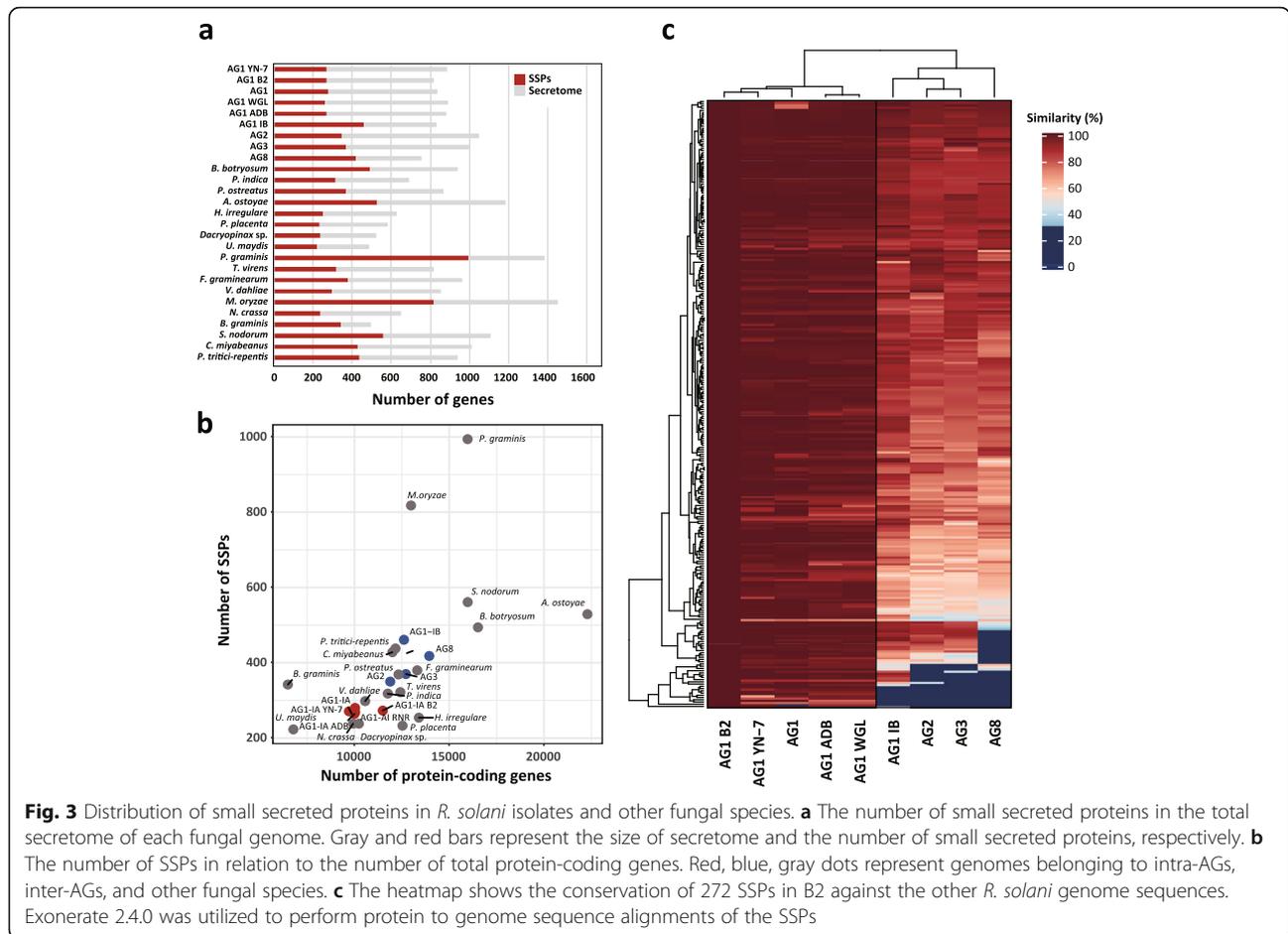
Furthermore, using the 272 SSPs identified in B2, we performed alignment of their protein sequence to the genomes of *R. solani* AG1-IA, AG1-IB, AG2-2IIIB, AG3, and AG8. The B2 SSPs possess a high degree of homology amongst rice-infecting AG1-IA genomes, whereas they showed decreased homology to genomes of other *R. solani* AGs (Fig. 3c, Additional file 10: Table S8).

Predicted CAZyme genes of the rice infecting *R. solani* isolates

We identified cell wall degrading enzymes by searching for each of the different CAZyme gene families across all 27 fungal genomes (Fig. 4). These fungal genomes were then categorized into 11 groups considering their nutritional lifestyle and type of host. A chi-square test of proportions was then used to determine whether gene frequency variations between genomes of each grouping were significant (Additional file 11: Table S9 and Additional file 12: Table S10). Our analyses indicated that there was significant variation across all 11 groups for all CAZyme gene families except GTs. Rice-infecting *R. solani* genomes had the highest enrichment of CAZyme genes (725 genes) while the genomes of other *R. solani* AGs, necrotrophs (cereal and dicot) showed only moderate enrichment for these genes. In contrast, biotrophs and brown rot genomes contain a relatively low number of CAZymes compared to rice-infecting *R. solani* genomes.

Lignocellulose-degrading genes in rice-infecting *R. solani* isolates

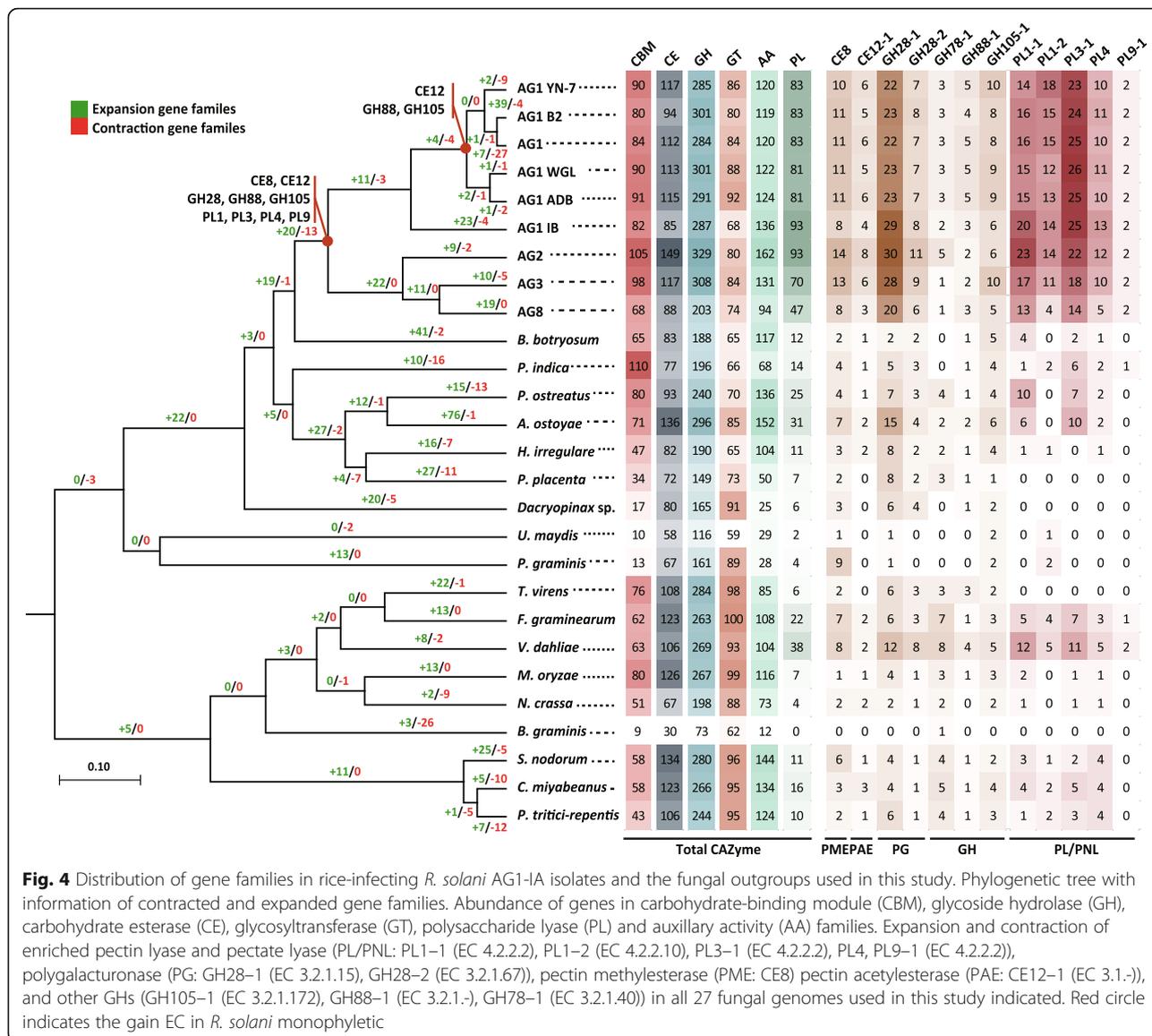
To ascertain whether rice-infecting *R. solani* isolates can degrade lignocellulose in a similar fashion to fungi in the subdivision Agaricomycotina, we specifically searched for AA family-encoding genes (Additional file 13: Table S11). Genomes of white-rot fungi and other *R. solani* AGs had the highest total number of CAZyme genes (131 genes), followed by necrotrophs (cereal) (128 genes) and rice-



infecting *R. solani* (121 genes). In contrast, brown-rot and biotroph genomes completely lacked these lignin depolymerization genes. We also examined the genomes to identify genes belonging to individual AA subfamilies. Genes belonging to the AA1 subfamily (EC 1.10.3.2) were most abundant in white-rot fungi but could also be found in rice-infecting *R. solani* isolates, hemibiotrophs, and necrotrophs (cereal); however, the presence of these AA1 genes was significantly lower in brown-rot fungi and biotrophs. In addition, while brown-rot and rice-infecting *R. solani* genomes did not contain any representatives of the AA2 subfamily, these genes were present in the genomes of necrotrophs (cereal) and white-rot fungi; the latter were highly enriched in manganese peroxidase (MnPs; 1.11.1.13) and versatile peroxidase (VPs; 1.11.1.16) genes. We also observed enrichment of AA8 subfamily genes in the genomes of rice-infecting *R. solani* isolates, while white-rot genomes had only 1 or 2 AA8 genes and cereal necrotroph genomes had none (except for *Septoria nodorum*, which had 1). Finally, the AA5 subfamily, which was absent in brown-rot fungi, was

significantly enriched in other *R. solani* AGs and to a lesser extent in rice-infecting *R. solani* and white-rot fungi.

We assessed the abundance of cellulose-degrading genes across all genomes in a similar fashion (Additional file 13: Table S11). Endoglucanase genes (GH3) were equally abundant in rice-infecting *R. solani* isolates and necrotrophs (cereal). However, exo-1,3-β-glucanase (3.2.1.58, GH5), endo-1,6-β-D-glucanase (3.2.1.75, GH5), and cellulose 1,4-β-cellobiosidase (3.2.1.176, GH7) genes were enriched in rice-infecting *R. solani* isolates but absent in necrotrophic fungi (cereal). Strikingly, rice-infecting *R. solani* genomes were highly enriched in genes encoding starch-degrading α-amylase (3.2.1.1, GH13). Moreover, genes encoding endo-1,4-β-D-glucanohydrolase (3.2.1.4, GH5), which causes endohydrolysis of (1 → 4)-β-D-glucosidic linkages in cellulose, lichenin, and cereal β-D-glucans, were enriched in rice-infecting *R. solani* and *S. nodorum*. Finally, genes encoding lytic polysaccharide monoxygenases (LPMOs) of the AA9 subfamily were significantly enriched in other *R. solani* AGs, though they were also present in rice-infecting *R. solani*. In



contrast, biotrophs and brown-rot fungi had, at most, one AA9 gene.

Pectin-degrading and modifying genes in rice-infecting *R. solani*

Extensive enrichment of genes belonging to the PL family was observed in both rice-infecting (82 genes) and other *R. solani* AGs (76 genes) (Additional file 13: Table S11). Otherwise, only necrotrophic (dicot) *V. dahliae* was predicted to have an enriched number of PL genes, though it only had approximately half as many PL genes as *R. solani*. In-depth analyses of the PL subfamilies revealed that genes encoding members such as pectate lyases (EC 4.2.2.2; PL1, 3, 9), pectin lyases (EC 4.2.2.10; PL1), and rhamnogalacturonan endolyases (EC

4.2.2.-, PL4) were significantly enriched in rice- and other *R. solani* AGs.

GH genes that encode pectin-degrading enzymes like unsaturated rhamnogalacturonyl hydrolase (EC 3.2.1.172, GH105), unsaturated β-glucuronyl hydrolase (3.2.1.-, GH88), and α-L-rhamnosidase (3.2.1.40, GH78) were analyzed in a similar fashion. GH28 genes were more enriched in rice-infecting *R. solani* than in necrotrophs (cereal) and *V. dahliae*. Specifically, 13 to 14 polygalacturonase (PGs) (EC 3.2.1.15, GH28) genes were found in AG1-IA genomes and other *R. solani* AG genomes while *V. dahliae* only had 2 PG genes. Exo-PGs (EC 3.2.1.67, GH28) were similarly abundant and enriched in *R. solani* and *V. dahliae*. Meanwhile, CE genes that encode pectin-modifying enzyme genes, including pectin methylsterases (PMEs; CE8) and pectin acetylsterases (CE12; EC 3.1.1.-), were most

abundant in all *R. solani* genomes. All of the enriched genes belonging to the homogalacturonan modification genes (PAE, PME, PG, and PL/PNL) have largely expanded when the *R. solani* divergence occurred (Additional file 14: Fig. S3).

Genes for monocot-specific cell wall degrading enzymes

To determine whether rice-infecting *R. solani* isolates have any CAZyme genes that allow them to infect their monocot host, we searched for CAZyme genes that degrade arabinoxylans, ferulic acids, and mixed linked glucans (MLGs), such as α -L-arabinofuranosidases (EC_3.2.1.55), feruloyl esterases (EC_3.1.1.73), and (1,3;1,4)- β -D-glucan endohydrolases/licheninase (EC 3.2.1.73), respectively (Additional file 15: Table S12). α -L-arabinofuranosidase genes were enriched in necrotrophs, including *R. solani* AG1-IA and other *R. solani* AGs, but not in white- and brown-rot fungi. However, while feruloyl esterase genes were enriched in necrotrophs (cereal), symbionts, and hemibiotrophs, none were identified in rice-infecting *R. solani*. While the raw data suggested that there was an enrichment in genes encoding (1,3;1,4)- β -D-glucan endohydrolases/licheninases in rice-infecting *R. solani* genomes, our chi-square test failed to reject the null hypothesis, so the proportion of (1,3;1,4)- β -D-glucan endohydrolases/licheninases genes across the different fungal genomes is likely similar.

Prediction of secondary metabolite biosynthesis gene clusters

antiSMASH [41] was used to identify putative secondary metabolite biosynthesis gene clusters for polyketide synthase (PKS), terpene synthase (TS), non-ribosomal peptide synthetase (NRPS), and other accessory enzymes in rice-infecting *R. solani* isolates (Additional file 15: Table S12). However, none of the secondary metabolite biosynthesis gene clusters predicted for rice-infecting *R. solani* isolates or members of related AGs contained PKS genes. In contrast, there was an abundance of secondary metabolite-producing enzymes predicted for necrotrophic ascomycetes (cereal), including type 1 PKS, type 2 PKS, NRPS, and TS. Despite the low abundance and diversity of putative secondary metabolite biosynthesis gene clusters in *R. solani* genomes, *R. solani* AGs and necrotrophic ascomycetes (cereal) have similar levels of terpenes, which suggests that secondary metabolites may not be important for *R. solani* virulence.

Discussion

De novo and reference-based genome assemblies of rice-infecting *R. solani* isolates

Both SMRT and Illumina sequencing technologies were used in the de novo sequencing of the B2 genome, which

allowed us to utilize (1) the ability of SMRT sequencing to generate long read lengths (5 to 20 kb) and precisely capture genomic regions containing repetitive elements and novel gene isoforms [42]. Utilizing these sequencing approaches facilitated assembly of the 45-Mbp B2 genome, which is much larger than a previously published 36.94-Mbp *R. solani* AG1-IA genome and has a relatively high proportion of repetitive sequences. In addition, it also provided an opportunity to accurately annotate the TE content of the B2 genome. Upon comparing to both newly and previously sequenced *R. solani* AG1-IA genomes as well as to genomes of other AGs, the B2 genome was found to possess the highest proportion of TEs and the longest average length of TEs. Thus, the higher quality B2 genome generated from the two sequencing methods allowed us to more accurately annotate the *R. solani* AG1-IA genomes for detailed comparative genome analyses of this important fungal pathogen.

Genomic differences among *R. solani* isolates

In previous studies, *R. solani* isolates have been classified based on their ability to hyphal fuse or through sequence analysis of phylogenetic markers [9, 43, 44] but whole-genome comparisons were not available. Here, we compared five *R. solani* AG1-IA genomes and four neighboring AG representative genomes and found out that genomic drastically decreased as the comparisons were made from among AG1-IA genomes to among different AG group genomes. Moreover, the similarity of inter-AGs was lower than other Basidiomycota genus-groups. This phylogenomic result shows *R. solani* species complex reflects multi-species feature in genome contents.

Small set of predicted putative effectors in rice-infecting *R. solani* isolates

It has been shown that necrotrophic fungi have fewer effectors than biotrophs [45]. Along this line, we found that the newly sequenced rice-infecting *R. solani* AG1-IA genomes had relatively small set of effectors among fungal genomes analyzed in this study. Previous reports suggest that the broad host range of *R. solani* is not dependent on the size of its secretome [46] but rather on the secretion of specific effectors than can infect a variety of different hosts, as is the case with necrotroph *Sclerotinia sclerotiorum* [47]. Furthermore, it has been reported that *Colletotrichum* pathogens from different clades have tailored suites of CAZymes that are specific to their individual host range and infection lifestyle [48]. We speculate that the relatively low number of SSPs (putative effectors) may be compensated by the diverse and large arsenal of CAZymes of rice-infecting *R. solani*, allowing them to be a competitive pathogen of broad host range. Additional bioinformatic and functional

genomics analyses must be conducted in order to dissect the role of SSPs in rice-infecting *R. solani* genomes. However, we have provided evidence that the SSPs among the rice-infecting *R. solani* genomes share high homology and decrease homology of these SSPs among genomes of different *R. solani* AGs suggest that the SSPs of from *R. solani* species complex as a whole may be diverse that previously expected.

Lignocellulose-degrading CAZyme genes in rice-infecting *R. solani* isolates

Essential for cell growth and differentiation, cell walls are primarily comprised of cellulose, hemicellulose, pectin, and lignin [49–53], which also provide plants with resistance to and protection from biotic and abiotic stresses [54–56]. Plant biomass-degrading fungi that inhabit diverse ecological niches such as forest litters, trees, crops, and grasses of the subdivision Agaricomycotina [34] are classified as either white- or brown-rot [57] based on their ability to degrade lignin. White-rot fungi use oxidative enzymes [58, 59] like glyoxal oxidases to efficiently depolymerize lignin and class II heme peroxidases of the AA2 subfamily, such as MnPs and VPs, to degrade the lignin matrix and expose the embedded cellulose [60, 61]. While we observed enrichment of glyoxal oxidase-encoding genes in rice-infecting *R. solani* genomes, we could not identify any peroxidase genes in the rice-infecting *R. solani* or brown-rot genomes. This apparent lack of peroxidases may explain why brown-rot fungi rapidly degrade cellulose but leave behind a chemically modified lignin matrix for long-term degradation by other microbes [62–64].

While a relationship between modes of wood decomposition and CAZyme families exists, some wood-decaying fungi cannot be strictly categorized as white- or brown-rot, suggesting that a continuum may exist between these two types of wood-decay fungi [57]. Rice-infecting *R. solani* isolates were enriched in lignocellulose-degrading CAZyme genes as well as strong cellulose-degrading enzymes, which are hallmarks of white- and brown-rot fungi, respectively. Hence, we hypothesize that *R. solani* may fall along the continuum between these two types of wood-decay fungi. Furthermore, the abundance of oxidoreductase and iron reductase genes in the *R. solani* genomes may indicate that these enzymes are involved in the production of hydroxyl radicals to drive lignocellulose attack, thereby working with crystalline cellulose-degrading LPMOs to expose the complex lignocellulose structure for further cell wall degradation by other CAZymes.

Enrichment of pectin-degrading enzymes in rice-infecting *R. solani* isolates

Pectin is a structural heteropolysaccharide with a 1,4- α -D-galacturonic acid (GalA) backbone that contributes to the mechanical strength of plants [65–67] by forming a gel-like matrix that interacts with cellulose and hemicellulose in the primary plant cell wall [51, 68]. It is present in higher proportions in dicot (type I) cell walls than in monocot (type II) cell walls [50], and some reports suggest that dicot-specific fungal pathogens have higher amounts of pectin-degrading enzymes than monocot-specific pathogens [69, 70]. Pectin is classified based on the degree of methoxylation, methylesterification, and/or acetylation of its backbone [71, 72]. Homogalacturonan (HG), a polymer of 1,4-linked α -D-galactopyranosyluronic acid, exists in methylesterified or acetylated form in the primary cell walls of plants [65], and pectin methylsterases (PMEs; EC 3.1.1.11) and pectin acetylsterases (PAE; EC 3.1.1.6) catalyze the demethylesterification and deacetylation of HG, respectively [73–75]. This process yields substrates for PGs [76], pectin lyases (PNLs; EC 4.2.2.10), and pectate lyases (PLs; EC 4.2.2.2), which loosen the cell wall [77]. Demethylesterification of HGs can also lead to ‘egg-box formation’ and cell wall stiffening caused by the interaction of negatively-charged demethylesterified HG and divalent cations such as calcium ions [77], which may explain the stiff, hollow-textured stem phenotypes observed in ShB-susceptible rice cultivars but not in moderately ShB-resistant rice cultivars (Lee et al., unpublished data). The significant enrichment of PMEs, PAEs, PGs, PNLs, and PLs in *R. solani* genomes suggests that these pathogens may have evolved a diverse suite of pectin depolymerization enzymes that allow them to efficiently breach host cell walls. These expanded homogalacturonan modification genes are known to have the enzymatic activity of cell wall loosening roles in the infection process. Similarly, mutations in *Arabidopsis thaliana* pectin methylsterase 35 (PME35) lead to the suppression of HG demethylesterification and a concomitant increased stem deformation rate, supporting the essential role of pectin in maintaining the integrity of the plant cell wall and supporting the plant’s mechanical properties [65].

Gene family expansions and contractions are the signatures of an organism’s adaptation to new ecological niches [78], as exemplified by the presence of numerous pectin-degrading genes in rice-infecting *R. solani* and neighboring *R. solani* AGs. However, dicot-specific fungal pathogens may not necessarily have more specialized pectin-depolymerizing enzyme suites than monocot-specific pathogens, as genes encoding HG-modifying enzymes are more highly enriched in rice-infecting *R. solani* isolates than in the dicot-specific pathogen *V. dahliae*. This

enrichment in pectin degrading genes indicates that *R. solani* can degrade a wide range of pectic substrates and polysaccharide linkages, allowing it to use multiple virulence mechanisms to invade a variety of hosts. Moreover, the large and diverse suite of pectin-degrading enzymes in rice-infecting *R. solani* isolates may not have evolved in response to the amount of pectin in host plant cell walls but rather as an efficient mechanism for loosening plant cell walls, breaking crosslinks with other cell wall components, and dissolving plant tissue.

Monocot-specific degrading genes in rice-infecting *R. solani* isolates

Previous studies have shown a correlation between cell wall composition and host specificity, suggesting that plant pathogens produce host-specific cell wall degrading enzymes [79, 80]. Accordingly, monocot-specific fungal pathogens are adept at hydrolyzing monocot cell walls while dicot-specific fungal pathogens are better at degrading dicot cell walls [81]. Non-cellulosic polysaccharides such as arabinoxylans, MLGs, and hydroxycinnamates such as ferulic acids, which are enriched in grass cell walls but either limited or absent in dicot cell walls [50, 82–84], are degraded by α -L-arabinofuranosidases (EC 3.2.1.55, GH51) [81, 85], licheninases [86], and feruloyl esterases (EC 3.1.1.73) [34]. However, we did not observe significant differences in the number of monocot-specific cell wall degrading CAZyme genes present in the different rice-infecting *R. solani* genomes, which may indicate the broad host range of this necrotrophic fungus [13].

Secondary metabolite biosynthesis clusters in *R. solani* isolates

Previous studies suggest the association of loss of secondary metabolite genes with biotrophy [87, 88]. However, most of the pathways for secondary metabolite synthesis in the biotrophic fungus *Cladosporium fulvum* were revealed to be cryptic [89]. Despite our results suggest that *R. solani* possess limited number of secondary metabolite biosynthesis clusters, further research on expression analyses of the putative secondary metabolite genes and those identified along with metabolite extraction and chromatography will be needed. These analyses will provide conclusive evidence about the extent of involvement of secondary metabolite genes in the life-style and pathogenicity of *R. solani*.

Conclusion

In this study, we analyzed the cell wall degrading enzyme profiles of four newly sequenced rice-infecting *R. solani* genomes. Comparative analyses of these rice-infecting *R. solani* genomes can help identify cell wall degrading mechanisms, such as homogalacturonan

modification, that are utilized by this necrotrophic, rice-infecting ShB pathogen. With more and more *R. solani* genomes are sequenced in the future, reclassification of this fungal pathogen should be discussed and implemented. Moreover, our findings, along with the high-quality genome sequences of rice-infecting isolates of *R. solani* AG1 IA, provide additional genomic resources that can be used to further our understanding of the pathobiology of this necrotrophic fungal pathogen.

Methods

Sources of *R. solani* AG1-IA isolates

Four *R. solani* AG1-IA isolates were collected from rice cultivars grown in the USA, India and China with famers' permission. B2 was recovered from Jerry Bogard Farms, Stuttgart, Arkansas, USA. ADB and WGL were recovered from Srinivasa Rao Farms in Adilabad and Bose Reddy Farms in Warangal in Telangana State, India, respectively. YN-7 was recovered from Zongliang Chen Farms in Yangzhou, China.

Fungal DNA extraction of *R. solani* isolates

Hyphal tip isolation and culture maintenance of *R. solani* isolates were conducted using Potato Dextrose Agar (PDA). *R. solani* hyphae-containing agar blocks were isolated from the actively growing mycelial portion of the fungus and cultured in the dark in liquid Potato Dextrose Broth (PDB) at 25 °C on an orbital shaker (150 rpm) for 4–5 days. Mycelia were filtered using sterile Miracloth (Millipore, Sigma, Burlington, MA, USA), rinsed with sterile distilled water, and frozen in liquid nitrogen. Genomic DNA (gDNA) was extracted using a DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA), and the resulting DNA pellet was resuspended in 10 mM Tris-HCl (pH 8.0) buffer. The quality of the isolated gDNA was assessed using agarose gel-based electrophoresis while the total DNA concentration was calculated based on UV-Vis measurements on a Nanodrop spectrophotometer. Isolated gDNAs were sent to the National Instrumentation Center for Environmental Management (NICEM), Seoul National University, Korea, for SMRT and Illumina-based sequencing.

Genome sequencing, assembly, and annotation

A PacBio sequencing assemblage strategy was used to assemble the B2 genome. Raw PacBio RSII sequence reads were assembled and corrected using Canu v2.0 [90] and trimmed using Circlator v0.14.0 [91]. All resulting contigs were then joined before the Redundans assembly pipeline [92] and Pilon v1.22 [93] was used to improve the final draft genome.

For Illumina-based sequencing and assembly of the ADB, WGL, and YN-7 genomes, contigs were aligned to the PacBio RSII-sequenced B2 genome. As with B2, the

Redundans assembly pipeline was used to process Illumina sequencing reads, but with an additional step: filtered Illumina reads by FASTQC v0.11.6 [94] were assembled using Jellyfish v1.1.5 [95]. Alignment and assembly were performed using SOAPdenovo2 [96] and Velvet v1.2.10 [97]. Gap-filling within scaffolds was achieved using GapCloser v1.1.2-r6 [98]. The resulting sequences were joined and corrected using Pilon v1.22. AUGUSTUS v3.2.2 [98] was used to make gene annotation predictions based on deposited six different protein-coding genes of *R. solani* in NCBI (AG1-IA, AG1-IB, AG2-2IIB, AG3 Rhs1AP, AG3 123E, and AG8 WAC10335). To assess the repetitive element content of the newly sequenced *R. solani* AG1-IA isolates, RepeatScout [99] was used to predict de novo consensus repetitive element families in all the *R. solani* genomes used in this study. Through this approach, all low-complexity sequences, tandem repeats as well as repeat elements which contain less than 10 repeats sequences were filtered out. The resulting consensus repeat elements were then classified using TEclass [100] and mapped using RepeatMasker v4.0.7 [101].

Comparative analyses and ortholog clustering

For pairwise genomic comparisons, MUMmer v3.23 [38] was used to align and compare the whole genome sequences of *R. solani* isolates. PROMer, a built-in MUMmer package that generates and aligns translations of all six reading frames for genome sequences of interest, was used to determine the extent of synteny between the genomes used in this study. OrthoFinder v 2.2.7 [102] was used for ortholog clustering to sort out single-copy gene families that would be the most phylogenetically informative. Single-copy ortholog genes in all fungal species were then aligned using ClustalW v2.1 [103], and poorly aligned regions were removed using trimAl v1.2 with the strict method [104]. RAXML v8.2.8 [105] and a bootstrap value of 1000 was used to construct a maximum likelihood-based phylogenetic tree. Ortholog genes were annotated with Gene Ontology (GO) annotation using Interproscan v5.20 [106].

Gene family analyses

Genes encoding plant cell wall degrading enzymes were predicted and categorized using dbCAN HMMER v6 [107]. Each EC gene was collected from classification of CAZyDB-ec-info.txt.07-20-2017. Each classified group from dbCAN was subdivided using EC classification using BLAST 2.2.26. Aligning EC classified protein sequences using ClustalW 2.1 and removal of poorly aligned regions by trimAl v1.2 were preceded before phylogenetic analysis. Phylogeny trees were constructed using RAXML version 8.2.9 with a bootstrap value of 1000. We reconciled the gene tree resulting from this

analysis with the species tree using NOTUNG 2.6 [108]. The secretome data of selected species were obtained from the Fungal Secretome Database (FSD) [109]. The database detects all possible secreted proteins by eliminating proteins with transmembrane or endoplasmic reticulum domains and using SignalP 3.0 [110]. The SSPs were then selected from each fungal secretome, considering proteins with a length shorter than 300 amino acids, as previously described [45]. Exonerate 2.4.0 was utilized to perform protein to genome sequence alignments of the effectors among *R. solani* genomes [111]. Genes encoding laccases and peroxidases were predicted using fPoxDB [112], while putative secondary metabolite biosynthesis gene clusters were identified using antiSMASH v3.0 [41], and the P450 database [113] was searched to predict cytochrome P450 genes in each genome. Transcription factors were identified using the Fungal Transcription Factor Database (FTFD) pipeline [114], which utilizes data from Interpro v12 [115].

Statistical analysis

Chi-square tests of proportions

Chi-square tests of proportions for comparative analyses of CAZyme secondary metabolite biosynthesis clusters were performed using R [116].

Abbreviations

AG: Anastomosis group; ISG: Intraspecific group; ShB: Sheath blight; HG: Homogalacturonan; SMRT: Single Molecule Real-Time; CAZyme: Carbohydrate-active enzyme; AA: Auxiliary activity; MnP: Manganese peroxidase; VP: Versatile peroxidase; LPMO: Lytic polysaccharide monooxygenase; PKS: Polyketide synthase; TS: Terpene synthase; NRPS: Non-ribosomal peptide synthetase; SSP: Small secreted protein; GH: Glycoside hydrolase; CE: Carbohydrate esterase; GT: Glycosyltransferase; PL: Polysaccharide lyase; PG: Polygalacturonase; PME: Pectin methylesterase; PAE: Pectin acetylerase; PNL: Pectin lyase

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07549-7>.

Additional file 1: Figure S1. Genome sequence assembly pipeline *R. solani* AG1-IA isolates B2, ADB, WGL and YN-7.

Additional file 2: Table S1. List of genomes of fungal species/genomes utilized in this study for comparative genome analyses (excluding the *R. solani* AG1-IA B2, ADB, WGL and YN-7).

Additional file 3: Figure S2. Synteny dot plots of *R. solani* genomes and reference B2 genome.

Additional file 4: Table S2. Pairwise genome alignments in terms of absolute and percent genome similarity among rice-infecting *R. solani* and different anastomosis groups (AG1-IB, AG2-2IIB, AG3, AG8) using PROMer.

Additional file 5: Table S3. Absolute number and proportion of orthologs upon pairwise comparisons of 27 fungal genomes.

Additional file 6: Table S4. List of additional Basidiomycota genomes included for comparative protein-coding gene analysis.

Additional file 7: Table S5. Repetitive element profiles for *R. solani* AG1-IA, AG1-IB, AG2-2IIB, AG3, and AG8 isolates.

Additional file 8: Table S6. The predicted number of protein-coding genes, secreted protein-coding genes, and small secreted protein-coding genes (SSPs) in 27 fungal genomes were used in this study.

Additional file 9: Table S7. List of conserved and unique SSPs found in all *R. solani* genomes used in this study.

Additional file 10: Table S8. Percentage of homology shared by 272 SSPs of *R. solani* B2 with *R. solani* AG1-IA (YN7, ADB, RNR), AG1-IB, AG2-2IIIB, AG3, and AG8.

Additional file 11: Table S9. Chi-square analysis of number of predicted Carbohydrate Active Enzymes (CAZymes) belonging carbohydrate-binding module (CBM), glycoside hydrolase (GH), carbohydrate esterase (CE), glycosyltransferase (GT), polysaccharide lyase (PL), and auxiliary activity (AA) families observed in 11 fungal groupings based on lifestyle and host of genomes used in this study ($p < 0.001$).

Additional file 12: Table S10. Analysis of chi-square tests of selected CAZymes and secondary metabolites.

Additional file 13: Table S11. Chi-square analysis of the number of predicted lignocellulose, pectin, and monocot cell wall degrading genes observed across the 11 fungal groupings based on lifestyle and host of genomes used in this study ($p < 0.001$).

Additional file 14: Figure S3. Duplication-loss model of enriched EC gene families related to pectin modification in 27 fungal genomes utilized in this study.

Additional file 15: Table S12. Putative secondary metabolite biosynthesis gene clusters were observed in 11 fungal groupings based on lifestyle and host of genomes used in this study ($p < 0.001$).

Acknowledgements

We thank Stella Lai for language editing.

Authors' contributions

YHL (Fungal Bioinformatics Laboratory, Seoul National University), GLW (The Ohio State University), and DYL (The Ohio State University) designed the study. JJ, KTK, KC, GC, JK, and HS (Fungal Bioinformatics Laboratory, Seoul National University) performed the bioinformatics analyses. SO (The Ohio State University) conducted the statistical analyses. SZ, JCC and SM collected and cultured the strains and isolated the DNA samples for sequencing. DYL and JJ interpreted the data and wrote the manuscript. All authors read and approved the final version of the manuscript.

Funding

This research was supported by a Ph. D fellowship awarded to D.-Y. Lee by the Monsanto Beachell-Borlaug International Scholarship Program (MBBISP) as well as grants from the National Research Foundation of Korea to YHL (NRF-2020R1A2B5B03096402, NRF-2015M3A9B8028679, and NRF-2018R1A5A1023599), the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry through Agricultural Microbiome Program to YHL (918017-04) and the USDA Hatch Project to GLW. KTK and JK is grateful for a graduate fellowship through the Brain Korea 21 Plus Program.

Availability of data and materials

The genome assemblies of *R. solani* AG1-IA isolates B2, YN-7, ADB, and WGL generated during the current study are available in the NCBI repository, [accession numbers JACYCC000000000-JACYCF000000000 in PRJNA645424]. For fungal genomes used for comparative analyses in this study were downloaded from the NCBI, JGI and Broad Institute database; their accession numbers are listed in Additional file 2: Table S1 and Additional file 6: Table S4. The datasets supporting the conclusions of this article are included within the article and its additional files.

Declarations

Ethics approval and consent to participate

Consent to participate was obtained verbally from each of the farmers.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Plant Pathology, The Ohio State University, Columbus, OH 43210, USA. ²Fungal Bioinformatics Laboratory, Seoul National University, Seoul 08826, South Korea. ³Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul 08826, South Korea. ⁴Department of Agricultural Life Science, Suncheon National University, Suncheon 57922, South Korea. ⁵Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, South Korea. ⁶Ohio Agricultural Research and Development Center (OARDC) Molecular & Cellular Imaging Center (MCIC)-Columbus, The Ohio State University, Columbus, OH 43210, USA. ⁷Department of Entomology & Plant Pathology, University of Arkansas, Fayetteville, AK 72701, USA. ⁸Key Laboratory of Plant Functional Genomics of the Ministry of Education/ Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Agricultural College of Yangzhou University, Yangzhou, China. ⁹Indian Council of Agricultural Research-Indian Institute of Rice Research (ICAR-IIRR), Hyderabad 500030, Telangana, India. ¹⁰Center for Fungal Genetic Resources, Plant Immunity Research Center, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, Seoul National University, 08826 Seoul, South Korea.

Received: 22 July 2020 Accepted: 23 March 2021

Published online: 07 April 2021

References

- Miyake I. Studien über die Pilze der Reis pflanze in Japan. J Coll Agric Imp Univ Tokyo. 1910;2:237–76.
- Hashiba T, Kobayashi T. Rice diseases incited by *Rhizoctonia* species. In: Sneh B, Jabaji-Hare S, Neate S, Dijst G, editors. *Rhizoctonia* species: taxonomy, molecular biology, ecology, pathology and disease control. Dordrecht: Springer; 1996. p. 331–40.
- Banniza S, Holderness M. Rice sheath blight—pathogen biology and diversity. In: Sreenivasaprasad S, Johnson R, editors. Major fungal diseases of rice. Dordrecht: Springer; 2001. p. 201–11. https://doi.org/10.1007/978-94-01-7-2157-8_14.
- Willcoquet L, Fernandez L, Savary SJPP. Effect of various crop establishment methods practised by Asian farmers on epidemics of rice sheath blight caused by *Rhizoctonia solani*. Plant Pathol. 2000;49(3):346–54. <https://doi.org/10.1046/j.1365-3059.2000.00454.x>.
- Kim K-H, Cho J, Lee YH, Lee W-SJA. Predicting potential epidemics of rice leaf blast and sheath blight in South Korea under the RCP 4.5 and RCP 8.5 climate change scenarios using a rice disease epidemiology model, EPIRICE. Agric For Meteorol. 2015;203:191–207. <https://doi.org/10.1016/j.jagformet.2015.01.011>.
- Ogoshi A. Ecology and pathogenicity of anastomosis and intraspecific groups of *Rhizoctonia solani* Kühn. Annu Rev Phytopathol. 1987;25(1):125–43. <https://doi.org/10.1146/annurev.py.25.090187.001013>.
- Carling D, Baird R, Gitaitis R, Brainard K, Kuninaga S. Characterization of AG-13, a newly reported anastomosis group of *Rhizoctonia solani*. Phytopathology. 2002;92(8):893–9. <https://doi.org/10.1094/PHYTO.2002.92.8.893>.
- Ajayi-Oyetunde O, Bradley C. *Rhizoctonia solani*: taxonomy, population biology and management of *Rhizoctonia* seedling disease of soybean. Plant Pathol. 2018;67(1):3–17. <https://doi.org/10.1111/ppa.12733>.
- Gonzalez D, Rodriguez-Carres M, Boekhout T, Stalpers J, Kuramae EE, Nakatani AK, et al. Phylogenetic relationships of *Rhizoctonia* fungi within the Cantharellales. Fungal Biol. 2016;120(4):603–19. <https://doi.org/10.1016/j.funbio.2016.01.012>.
- Priyatmojo A, Escopalao VE, Tangonan NG, Pascual CB, Suga H, Kageyama K, et al. Characterization of a new subgroup of *Rhizoctonia solani* anastomosis group 1 (AG-1-ID), causal agent of a necrotic leaf spot on coffee. Phytopathology. 2001;91(11):1054–61. <https://doi.org/10.1094/PHYTO.2001.91.11.1054>.
- Jones R, Belmar S. Characterization and pathogenicity of *Rhizoctonia* spp. isolated from rice, soybean, and other crops grown in rotation with rice in Texas. Plant Dis. 1989;73(12):1004–10. <https://doi.org/10.1094/PD-73-1004>.
- Pascual C, Toda T, Raymondo A, Hyakumachi M. Characterization by conventional techniques and PCR of *Rhizoctonia solani* isolates causing banded leaf sheath blight in maize. Plant Pathol. 2000;49(1):108–18. <https://doi.org/10.1046/j.1365-3059.2000.00429.x>.

13. Hane JK, Anderson JP, Williams AH, Sperschneider J, Singh KB. Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLoS Genet.* 2014;10(5):e1004281. <https://doi.org/10.1371/journal.pgen.1004281>.
14. Ahuja S, Payak MJ. Symptoms and signs of banded leaf and sheath blight of maize. *Phytoparasitica.* 1982;10(1):41–9. <https://doi.org/10.1007/BF02981891>.
15. Giesler LJ, Yuen GY, Horst GLJP. The microclimate in tall fescue turf as affected by canopy density and its influence on brown patch disease. *Plant Dis.* 1996;80(4):389–94. <https://doi.org/10.1094/PD-80-0389>.
16. Zheng A, Lin R, Zhang D, Qin P, Xu L, Ai P, et al. The evolution and pathogenic mechanisms of the rice sheath blight pathogen. *Nat Commun.* 2013;4(1):1424. <https://doi.org/10.1038/ncomms2427>.
17. Wibberg D, Jelonek L, Rupp O, Hennig M, Eikmeyer F, Goesmann A, et al. Establishment and interpretation of the genome sequence of the phytopathogenic fungus *Rhizoctonia solani* AG1-IB isolate 7/3/14. *J Biotechnol.* 2013;167(2):142–55. <https://doi.org/10.1016/j.jbiotec.2012.12.010>.
18. Wibberg D, Rupp O, Jelonek L, Krober M, Verwaaijen B, Blom J, et al. Improved genome sequence of the phytopathogenic fungus *Rhizoctonia solani* AG1-IB 7/3/14 as established by deep mate-pair sequencing on the MiSeq (Illumina) system. *J Biotechnol.* 2015;203:19–21. <https://doi.org/10.1016/j.jbiotec.2015.03.005>.
19. Wibberg D, Andersson L, Tzelepis G, Rupp O, Blom J, Jelonek L, et al. Genome analysis of the sugar beet pathogen *Rhizoctonia solani* AG2-2IIIb revealed high numbers in secreted proteins and cell wall degrading enzymes. *BMC Genomics.* 2016;17(1):245. <https://doi.org/10.1186/s12864-016-2561-1>.
20. Cubeta MA, Thomas E, Dean RA, Jabaji S, Neate SM, Tavantzis S, et al. Draft genome sequence of the plant-pathogenic soil fungus *Rhizoctonia solani* anastomosis group 3 strain Rhs1AP. *Genome Announc.* 2014;2(5):e01072–14.
21. Wibberg D, Genzel F, Verwaaijen B, Blom J, Rupp O, Goesmann A, et al. Draft genome sequence of the potato pathogen *Rhizoctonia solani* AG3-PT isolate Ben3. *Arch Microbiol.* 2017;199(7):1065–8. <https://doi.org/10.1007/s00203-017-1394-x>.
22. Patil VU, Girmalla V, Sagar V, Bhardwaj V, Chakrabarti SK. Draft genome sequencing of *Rhizoctonia solani* anastomosis group 3 (AG3-PT) causing stem canker and black scurf of potato. *Am J Potato Res.* 2017;95:87–91.
23. Losada L, Pakala SB, Fedorova ND, Joardar V, Shabalina SA, Hostetler J, et al. Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. *FEMS Microbiol Lett.* 2014; 352(2):165–73. <https://doi.org/10.1111/1574-6968.12387>.
24. Xia Y, Fei B, He J, Zhou M, Zhang D, Pan L, et al. Transcriptome analysis reveals the host selection fitness mechanisms of the *Rhizoctonia solani* AG1IA pathogen. *Sci Rep.* 2017;7(1):10120. <https://doi.org/10.1038/s41598-017-10804-1>.
25. Kwon YS, Kim SG, Chung WS, Bae H, Jeong SW, Shin SC, et al. Proteomic analysis of *Rhizoctonia solani* AG-1 sclerotia maturation. *Fungal Biol.* 2014; 118(5-6):433–43. <https://doi.org/10.1016/j.funbio.2014.02.001>.
26. Anderson JP, Hane JK, Stoll T, Pain N, Hastie ML, Kaur P, et al. Proteomic analysis of *Rhizoctonia solani* identifies infection-specific, redox associated proteins and insight into adaptation to different plant hosts. *Mol Cell Proteomics.* 2016;15(4):1188–203. <https://doi.org/10.1074/mcp.M115.054502>.
27. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol.* 2015;66(1):513–45. <https://doi.org/10.1146/annurev-arplant-043014-114623>.
28. Girard V, Dieryckx C, Job C, Job DJP. Secretomes: the fungal strike force. *Proteomics.* 2013;13(3-4):597–608. <https://doi.org/10.1002/pmic.201200282>.
29. Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, et al. Metabolic priming by a secreted fungal effector. *Nature.* 2011;478(7369):395–8. <https://doi.org/10.1038/nature10454>.
30. Khang CH, Berruyer R, Giraldo MC, Kankanala P, Park SY, Czymmek K, et al. Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. *Plant Cell.* 2010;22(4):1388–403. <https://doi.org/10.1105/tpc.109.069666>.
31. De Wit PJ, Mehrabi R, Van den Burg HA, Stergiopoulos I. Fungal effector proteins: past, present and future. *Mol Plant Pathol.* 2009;10(6):735–47. <https://doi.org/10.1111/j.1364-3703.2009.00591.x>.
32. Ciuffetti LM, Manning VA, Pandelova I, Betts MF, Martinez JP. Host-selective toxins, Ptr ToxA and Ptr ToxB, as necrotrophic effectors in the *Pyrenophora tritici-repentis*-wheat interaction. *New Phytol.* 2010;187(4):911–9. <https://doi.org/10.1111/j.1469-8137.2010.03362.x>.
33. Stergiopoulos I, Collemare J, Mehrabi R, De Wit PJGM. Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. *FEMS Microbiol Rev.* 2013;37(1):67–93. <https://doi.org/10.1111/j.1574-6976.2012.00349.x>.
34. Rytioja J, Hilden K, Yuzon J, Hatakka A, de Vries RP, Makela MR. Plant-polysaccharide-degrading enzymes from Basidiomycetes. *Microbiol Mol Biol Rev.* 2014;78(4):614–49. <https://doi.org/10.1128/MMBR.00035-14>.
35. Annis SL, Goodwin PH. Recent advances in the molecular genetics of plant cell wall-degrading enzymes produced by plant pathogenic fungi. *Eur J Plant Pathol.* 1997;103(1):1–14. <https://doi.org/10.1023/A:1008656013255>.
36. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 2009;37:D233–D238, Database, DOI: <https://doi.org/10.1093/nar/gkn663>.
37. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42(D1):D490–5. <https://doi.org/10.1093/nar/gkt1178>.
38. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5(2):R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
39. Castanera R, Borgognone A, Pisabarro AG, Ramirez L. Biology, dynamics, and applications of transposable elements in basidiomycete fungi. *Appl Microbiol Biotechnol.* 2017;101(4):1337–50. <https://doi.org/10.1007/s00253-017-8097-8>.
40. Kidwell MGJG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica.* 2002;115(1):49–63. <https://doi.org/10.1023/A:1016072014259>.
41. Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, et al. antiSMASH 3.0: a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 2015;43(W1):W237–43. <https://doi.org/10.1093/nar/gkv437>.
42. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinform.* 2015;13(5):278–89. <https://doi.org/10.1016/j.gpb.2015.08.002>.
43. Arakawa M, Inagaki K. Molecular markers for genotyping anastomosis groups and understanding the population biology of *Rhizoctonia* species. *J Gen Plant Pathol.* 2014;80(5):401–7. <https://doi.org/10.1007/s10327-014-0536-0>.
44. Sharma M, Gupta S, Sharma TR. Characterization of variability in *Rhizoctonia solani* by using morphological and molecular markers. *J Phytopathol.* 2005; 153(7-8):449–56. <https://doi.org/10.1111/j.1439-0434.2005.01000.x>.
45. Kim KT, Jeon J, Choi J, Cheong K, Song H, Choi G, et al. Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Front Plant Sci.* 2016;7:186.
46. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol Evol.* 2004;19(10):535–44. <https://doi.org/10.1016/j.tree.2004.07.021>.
47. Guyon K, Balagué C, Roby D, Raffaele S. Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. *BMC Genomics.* 2014;15(1):336. <https://doi.org/10.1186/1471-2164-15-336>.
48. Gan P, Narusaka M, Kumakura N, Tsuchida A, Takano Y, Narusaka Y, et al. Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol Evol.* 2016;8(5):1467–81. <https://doi.org/10.1093/gbe/eww089>.
49. Pauly M, Keegstra K. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 2008;54(4):559–68. <https://doi.org/10.1111/j.1365-3113.2008.03463.x>.
50. Vogel J. Unique aspects of the grass cell wall. *Curr Opin Plant Biol.* 2008; 11(3):301–7. <https://doi.org/10.1016/j.pbi.2008.03.002>.
51. Carpita NC, Gibeaut DM. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 1993;3(1):1–30. <https://doi.org/10.1111/j.1365-3113.1993.tb00007.x>.
52. Northcote D. Chemistry of the plant cell wall. *Annu Rev Plant Physiol.* 1972; 23(1):113–32. <https://doi.org/10.1146/annurev.pp.23.060172.000553>.
53. Zhong R, Ye ZH. Regulation of cell wall biosynthesis. *Curr Opin Plant Biol.* 2007;10(6):564–72. <https://doi.org/10.1016/j.pbi.2007.09.001>.

54. Underwood W. The plant cell wall: a dynamic barrier against pathogen invasion. *Front Plant Sci.* 2012;3:85.
55. Bacete L, Mérida H, Miedes E, Molina A. Plant cell wall-mediated immunity: cell wall changes trigger disease resistance responses. *Plant J.* 2018;93(4):614–36. <https://doi.org/10.1111/tpj.13807>.
56. Zhao Q, Dixon RA. Altering the cell wall and its impact on plant disease: from forage to bioenergy. *Annu Rev Phytopathol.* 2014;52(1):69–91. <https://doi.org/10.1146/annurev-phyto-082712-102237>.
57. Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, et al. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc Natl Acad Sci.* 2014;111(27):9923–8. <https://doi.org/10.1073/pnas.1400592111>.
58. Kirk TK, Farrell RL. Enzymatic "combustion": the microbial degradation of lignin. *Annu Rev Microbiol.* 1987;41(1):465–501. <https://doi.org/10.1146/annurev.mi.41.100187.002341>.
59. Hatakka A. Lignin-modifying enzymes from selected white-rot fungi: production and role from in lignin degradation. *FEMS Microbiol Rev.* 1994;13(2-3):125–35. <https://doi.org/10.1111/j.1574-6976.1994.tb00039.x>.
60. Makela MR, Donofrio N, de Vries RP. Plant biomass degradation by fungi. *Fungal Genet Biol.* 2014;72:2–9. <https://doi.org/10.1016/j.fgb.2014.08.010>.
61. Eastwood DC, Floudas D, Binder M, Majcherzyk A, Schneider P, Aerts A, et al. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science.* 2011;333(6043):762–5. <https://doi.org/10.1126/science.1205411>.
62. Blanchette RA. Delignification by wood-decay fungi. *Annu Rev Phytopathol.* 1991;29(1):381–403. <https://doi.org/10.1146/annurev.py.29.090191.002121>.
63. Worrall JJ, Anagnost SE, Zabel RA. Comparison of wood decay among diverse lignicolous fungi. *Mycologia.* 1997;89(2):199–219. <https://doi.org/10.080/00275514.1997.12026772>.
64. Yelle DJ, Ralph J, Lu F, Hammel KE. Evidence for cleavage of lignin by a brown rot basidiomycete. *Environ Microbiol.* 2008;10(7):1844–9. <https://doi.org/10.1111/j.1462-2920.2008.01605.x>.
65. Hongo S, Sato K, Yokoyama R, Nishitani K. Demethylesterification of the primary wall by PECTIN METHYLESTERASE35 provides mechanical support to the *Arabidopsis* stem. *Plant Cell.* 2012;24(6):2624–34. <https://doi.org/10.1105/tpc.112.099325>.
66. Bouton S. QUASIMODO1 encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in *Arabidopsis*. *Plant Cell.* 2002;14(10):2577–90. <https://doi.org/10.1105/tpc.004259>.
67. Liu H, Ma Y, Chen N, Guo S, Liu H, Guo X, et al. Overexpression of stress-inducible OsBURP16, the β subunit of polygalacturonase 1, decreases pectin content and cell adhesion and increases abiotic stress sensitivity in rice. *Plant Cell Environ.* 2014;37(5):1144–58. <https://doi.org/10.1111/pce.12223>.
68. Levesque-Tremblay G, Pelloux J, Braybrook SA, Müller KJP. Tuning of pectin methylesterification: consequences for cell wall biomechanics and development. *Planta.* 2015;242(4):791–811. <https://doi.org/10.1007/s00425-015-2358-5>.
69. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, et al. Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. *Nat Genet.* 2012;44(9):1060–5. <https://doi.org/10.1038/ng.2372>.
70. 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(1):274. <https://doi.org/10.1186/1471-2164-14-274>.
71. Mohnen D. Pectin structure and biosynthesis. *Curr Opin Plant Biol.* 2008;11(3):266–77. <https://doi.org/10.1016/j.pbi.2008.03.006>.
72. Harholt J, Suttangkakul A, Vibe SH. Biosynthesis of pectin. *Plant Physiol.* 2010;153(2):384–95. <https://doi.org/10.1104/pp.110.156588>.
73. Wolf S, Mouille G, Pelloux J. Homogalacturonan methyl-esterification and plant development. *Mol Plant.* 2009;2(5):851–60. <https://doi.org/10.1093/mp/ssp066>.
74. Micheli F. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci.* 2001;6(9):414–9. [https://doi.org/10.1016/S1360-1385\(01\)02045-3](https://doi.org/10.1016/S1360-1385(01)02045-3).
75. Pelloux J, Rusterucci C, Mellerowicz EJ. New insights into pectin methylesterase structure and function. *Trends Plant Sci.* 2007;12(6):267–77. <https://doi.org/10.1016/j.tplants.2007.04.001>.
76. Albersheim P, Darvill A, Roberts K, Sederoff R, Staehelin A. *Plant cell walls*. New York: Garland Science. Taylor and Francis Group; 2011.
77. Hocq L, Pelloux J, Lefebvre V. Connecting homogalacturonan-type pectin remodeling to acid growth. *Trends Plant Sci.* 2017;22(1):20–9. <https://doi.org/10.1016/j.tplants.2016.10.009>.
78. Lespinet O, Wolf YI, Koonin EV, Aravind L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 2002;12(7):1048–59. <https://doi.org/10.1101/gr.174302>.
79. Cooper RM, Longman D, Campbell A, Henry M, Lees PJP, Pathology MP. Enzymic adaptation of cereal pathogens to the monocotyledonous primary wall. *Physiol Mol Plant P.* 1988;32(1):33–47. [https://doi.org/10.1016/S0885-5765\(88\)80004-3](https://doi.org/10.1016/S0885-5765(88)80004-3).
80. Zalewska-Sobczak J. Sequential secretion of cell wall degrading enzymes by *Botrytis fabae* and *Fusarium avenaceum* during growth on host and non-host plants. *Biochem Physiol Pflanz.* 1985;180(3):169–75. [https://doi.org/10.1016/S0015-3796\(85\)80012-3](https://doi.org/10.1016/S0015-3796(85)80012-3).
81. King BC, Waxman KD, Nenni NV, Walker LP, Bergstrom GC, Gibson DM. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. *Biotechnol Biofuels.* 2011;4(1):4. <https://doi.org/10.1186/1754-6834-4-4>.
82. Buanafina MM. Feruloylation in grasses: current and future perspectives. *Mol Plant.* 2009;2(5):861–72. <https://doi.org/10.1093/mp/ssp067>.
83. Vega-Sanchez ME, Verherbruggen Y, Scheller HV, Ronald PC. Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant Signal Behav.* 2013;8(2):e23143. <https://doi.org/10.4161/psb.23143>.
84. Hatfield RD, Rancour DM, Marita JM. Grass cell walls: a story of cross-linking. *Front Plant Sci.* 2016;7:2056.
85. Sumiyoshi M, Nakamura A, Nakamura H, Hakata M, Ichikawa H, Hirochika H, et al. Increase in cellulose accumulation and improvement of saccharification by overexpression of arabinofuranosidase in rice. *PLoS One.* 2013;8(11):e78269. <https://doi.org/10.1371/journal.pone.0078269>.
86. Vega-Sanchez ME, Verherbruggen Y, Christensen U, Chen X, Sharma V, Varanasi P, et al. Loss of cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. *Plant Physiol.* 2012;159(1):56–69. <https://doi.org/10.1104/pp.112.195495>.
87. Collemare J, Lebrun MH. Fungal secondary metabolites: ancient toxins and novel effectors in plant–microbe interactions. In: Martin F, Kamoun S, editors. *Effectors in plant–microbe interactions*. vol. 2011. Chichester: Wiley; 2011. p. 377–400. <https://doi.org/10.1002/9781119949138.ch15>.
88. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science.* 2010;330(6010):1543–6. <https://doi.org/10.1126/science.1194573>.
89. Collemare J, Griffiths S, Iida Y, Jashni MK, Battaglia E, Cox RJ, et al. Secondary metabolism and biotrophic lifestyle in the tomato pathogen *Cladosporium fulvum*. *PLoS One.* 2014;9(1):e85877. <https://doi.org/10.1371/journal.pone.0085877>.
90. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–36. <https://doi.org/10.1101/gr.215087.116>.
91. Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 2015;16(1):294. <https://doi.org/10.1186/s13059-015-0849-0>.
92. Pryszcz LP, Gabaldon T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 2016;44(12):e113. <https://doi.org/10.1093/nar/gkw294>.
93. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014;9(11):e112963. <https://doi.org/10.1371/journal.pone.0112963>.
94. Andrews S. *FastQC: a quality control tool for high throughput sequence data*. Cambridge: Babraham Bioinformatics, Babraham Institute; 2010.
95. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 2011;27(6):764–70. <https://doi.org/10.1093/bioinformatics/btr011>.
96. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience.* 2012;1(1):18. <https://doi.org/10.1186/2047-217X-1-18>.
97. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18(5):821–9. <https://doi.org/10.1101/gr.074492.107>.
98. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 2005;33:W465-W477, Web Server, DOI: <https://doi.org/10.1093/nar/gki458>.

99. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*. 2005;21(Suppl 1):i351–8. <https://doi.org/10.1093/bioinformatics/bti1018>.
100. Abrusán G, Grundmann N, DeMester L, Makalowski W. TEclass - a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics*. 2009;25(10):1329–30. <https://doi.org/10.1093/bioinformatics/btp084>.
101. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013. <https://www.repeatmasker.org>.
102. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 2015;16(1):157. <https://doi.org/10.1186/s13059-015-0721-2>.
103. Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–8. <https://doi.org/10.1093/bioinformatics/btm404>.
104. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
105. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
106. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res*. 2016;45:D190–D9.
107. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, et al. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res*. 2017;46:D516–D21.
108. Chen K, Durand D, Farach-Colton M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol*. 2000;7(3-4):429–47. <https://doi.org/10.1089/106652700750050871>.
109. Choi J, Park J, Kim D, Jung K, Kang S, Lee Y-HJ. Fungal secretome database: integrated platform for annotation of fungal secretomes. *BMC Genomics*. 2010;11(1):105. <https://doi.org/10.1186/1471-2164-11-105>.
110. Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 2004;340(4):783–95. <https://doi.org/10.1016/j.jmb.2004.05.028>.
111. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform*. 2005;6(1):31. <https://doi.org/10.1186/1471-2105-6-31>.
112. Choi J, Détry N, Kim K-T, Asiegbu FO, Valkonen JP, Lee Y-H. fPoxDB: fungal peroxidase database for comparative genomics. *BMC Microbiol*. 2014;14(1):117. <https://doi.org/10.1186/1471-2180-14-117>.
113. Park J, Lee S, Choi J, Ahn K, Park B, Park J, et al. Fungal cytochrome P450 database. *BMC Genomics*. 2008;9(1):402. <https://doi.org/10.1186/1471-2164-9-402>.
114. Park J, Park J, Jang S, Kim S, Kong S, Choi J, et al. FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics*. 2008;24(7):1024–5. <https://doi.org/10.1093/bioinformatics/btn058>.
115. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2008;37:D211–D5.
116. R Core Team R. R: a language and environment for statistical computing. Vienna: R Foundation for statistical Computing; 2013. <https://www.R-project.org/>

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

