3 Biotech (2020) 10:342
https://doi.org/10.1007/s13205-020-02336-9

ORIGINAL ARTICLE

Comparative genomics of rice false smut fungi *Ustilaginoidea virens* Uv-Gvt strain from India reveals genetic diversity and phylogenetic divergence

Devanna Pramesh[1] · Muthukapalli K. Prasannakumar[2] · Kondarajanahally M. Muniraju[1] · H. B. Mahesh[2] · H. D. Pushpa[3] · Channappa Manjunatha[4] · Alase Saddamhusen[1] · E. Chidanandappa[1] · Manoj K. Yadav[5] · Masalavada K. Kumara[1] · Huded Sharanabasav[1] · B. S. Rohith[6] · Gaurab Banerjee[6] · Anupam J. Das[7]

Received: 5 November 2019 / Accepted: 12 July 2020
© King Abdulaziz City for Science and Technology 2020

Abstract

False smut disease of rice caused by *Ustilaginoidea virens,* is an emerging threat to rice cultivation worldwide due to its detrimental effects on grain yield and quality. False smut disease severity was 4.44−17.22% during a roving survey in *Kharif* 2016 in the four different rice ecosystems of Karnataka, India. Further, 15 pathogen isolates representing four different ecosystems were studied for their virulence and morphometric diversity. Among the 15 strains studied, most virulent strains Uv-Gvt was selected for whole genome sequencing in Illumina NextSeq 500 platform using 2 × 150 bp sequencing chemistry. The total assembled genome of Uv-Gvt was 26.96 Mb, which comprised of 9157 scaffolds with an N50 value of 15,934 bp and 6628 protein-coding genes. Next, the comparative genomic study revealed a similar gene inventory as UV-8b and MAFF 236576 strains reported from China and Japan, respectively. But, 1756 genes were unique to Uv-Gvt strain. The Uv-Gvt genome harbors 422 putative host–pathogen interacting genes compared to 359 and 520 genes in UV-8b and MAFF 236576 strains, respectively. The variant analysis revealed low genetic diversity (0.073−0.088%) among *U. virens* strains. Further, phylogenetic analysis using 250 single copy orthologs genes of *U. virens* revealed a distinct phylogeny and an approximate divergence time. Our study, report the genomic resource of rice false smut pathogen from India, where the disease originated, and this information will have broader applicability in understanding the pathogen population diversity.

Keywords Rice false smut · Genome · Comparative genomics · Sequencing · Phylogeny · Divergence

Introduction

In rice, biotic stress is a major limiting factor for productivity leading to significant economic losses to rice producers. Rice false smut is caused by the pathogenic fungus *Ustilaginoidea virens* (Cooke) Takahashi (teleomorph: *Villosiclava virens*) was first reported from India in the 1870s (Cooke

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-020-02336-9) contains supplementary material, which is available to authorized users.

✉ Devanna Pramesh
parmi.iari@gmail.com

[1] Rice Pathology Laboratory, All India Co-ordinated Rice Improvement Program, Gangavathi, University of Agricultural Sciences, Raichur 584 104, India

[2] University of Agricultural Sciences, Bangalore 560 065, India

[3] ICAR-Indian Institute of Oilseed Research, Hyderabad 500 030, India

[4] ICAR-Indian Agricultural Research Institute, Regional Station, Wellington 643 231, India

[5] ICAR- National Rice Research Institute, Cuttack 753 006, India

[6] Molsys Pvt. Ltd., Bangalore 560 064, India

[7] School of Biotechnology, REVA University, Bangalore 560 064, India

Published online: 19 July 2020

1878). In the rice field, this disease can be observed only on the panicle after grain filling stage. In the panicle, individual grains were transformed into dense white mycelia which later convert into a whitish smut ball. Subsequently, the colouration is altered from yellowish-orange, green, olive green and finally to greenish-black colored smut ball (Online Resource 1). The prevalence of false smut disease in major rice-growing regions globally is of concern due to significant grain yield reduction and the depreciation in market value resulting from grain contamination (Rush et al. 2000; Singh and Pophaly 2010). In the field conditions, high disease severity is attributed to the widespread cultivation of high-yielding cultivars and hybrids, excessive use of chemical fertilizers and an apparent shift in the regional and global climatic pattern (Lu et al. 2009). Besides, the production of toxins in the infected grains has the potential to cause health hazards in humans and animals (Zhang et al. 2014; Tsukui et al. 2014). In India, since the year 2001, the number of disease incidences are on the rise (Dodan and Singh 1996; Mandhare et al. 2008; Anonymous 2016), leading to the grain yield loss ranging from 1.01 to 10.91% (Atia 2004). While the disease incidence upto 85% has been recorded in different rice growing regions (Ladhalakshmi et al. 2012).

To develop a viable management strategy, it is essential to know the pathogen population composition along with its diversity and virulence, in addition to pathogenomics (genes involved in pathogenicity, effectors diversity, genes involved in the production of secondary metabolites like toxins and the pathogen-host interaction). Prior studies have provided useful insights into disease incidence, diagnosis, mycotoxins characterization, life cycle and management of the pathogen. However, to date there are limited reports on the pathogenomics. A complete genome sequence of *U. virens* was first reported by Zhang et al. (2014).Their study provided fundamental insights into the different genomic components associated with the disease symptoms and toxin production on rice grains. Subsequently, sequencing efforts by Kumagai et al. (2016) revealed diverse pathways in the production of mycotoxins. Although the disease and pathogen has been reported from various countries, the availability of pathogenomic information is limited to two origins (China and Japan). Furthermore, Sun et al. (2013) reported the existence of genetic diversity among geographically distinct isolates. Even though the disease was first reported in India (Cooke 1878), lacking the genomic resources of an Indian strain is critical to understand the phylogeny and genetic diversity at the global level. Our study seeks to contribute to the existing knowledge and establish a new genomic resource for the *U. virens* strain from India.

Our preliminary work included a survey to assess the disease status in different rice growing ecosystems followed by morphometric and virulence characterization of various isolates. The results depict significant diversity in virulence and morphometric features among the isolates. For further genomic characterization, the most virulent strain was selected.

Here, we report the whole genome sequence of a virulent Indian strain. To the best of our knowledge, this is the first complete genome sequence of the rice false smut pathogen strain from India. Our results may pave the way for the development of rapid and sensitive diagnostic techniques to be used in disease forecasting as well as disease resistance breeding program.

Materials and methods

Survey and collection of isolate

A roving survey was conducted during *Kharif* 2016 in four different rice growing ecosystems North-eastern Karnataka ecosystem (NEK) (includes Tungh-Bhadra Project [TBP] and Upper Krishna Project [UKP]), Cauvery ecosystem, Bhadra ecosystem and Hilly upland ecosystems of Karnataka, India. Rice panicles showing characteristics symptoms of false smut disease were collected from all four ecosystems for isolation and characterization of pathogen. Fifteen isolates were recovered from the diseased samples representing all ecosystems surveyed and were used for morphometric, virulence and genomic characterization. During the survey, disease incidence and severity on different rice cultivars in all four ecosystems was also estimated as suggested by Muniraju et al. (2017a; b).

Isolation and morphometric characterization

For isolation of pathogen, smut balls were sterilized in 70% (v/v) ethyl alcohol, and then suspended in 2% (w/v) sodium hypochlorite for one minute followed by re-suspending in 0.1 per cent aqueous solution of mercuric chloride for 30 s and finally washed thrice with sterilized distilled water. Later, chlamydospores were scraped and streaked on the Petri-plates containing potato sucrose agar (PSA) medium supplemented with streptomycin antibiotic (100 ppm). The inoculated Petri-plates were incubated at 25 ± 2 °C for 15 days to induce the germination of chlamydospores. Single white colony developed from the individual chlamydospore was sub-cultured to obtain a pure culture of the fungus. The cultures were identified as per the descriptions are given previously (Sharma and Joshi 1975; Verma and Singh 1988). Pure cultures of all 15 isolates were recovered and cultural characters on PSA medium were recorded. Morphometric characteristics of chlamydospores were recorded under a bright-field microscope fitted with the Image Analyzer Software (EVOS™ M7000 Imaging System, Thermo Fisher Scientific, Waltham, MA, USA).

Pathogenicity and virulence test

All 15 isolates were mass multiplied on Potato Sucrose Broth (PSB) and inoculated to two rice cultivars BPT5204 (susceptible) and IR28 (resistant) at booting stage. About 2 ml of spore suspension comprised secondary conidia (1×10^6 conidia/ml) and fragmented mycelium was inoculated to boot leaf sheath using a syringe (Ashizawa et al. 2011). Later, inoculated plants were incubated in the growth chamber at 28 °C with 90% relative humidity for two days and later transferred to the glass house. The pathogen was re-isolated from the symptomatic plant and its identity was confirmed based on microscopic observation. For virulence analysis, observation on per cent disease incidence and number of smut ball per panicles was recorded for each isolate-cultivar combination. Based on the disease reaction on BPT5204 and IR28, Uv-Gvt strain was identified as most virulent (data not shown) and was selected for further genomic characterization.

DNA isolation

Monoconidial derived pure culture of the Uv-Gvt strain was grown in PSB medium in an incubator shaker at 150 rpm at 28 °C for 10 days and the mycelial mat was harvested for DNA isolation. Genomic DNA was isolated using CTAB (cetyl trimethyl ammonium bromide) and phenol–chloroform extraction method (Rogers and Bendich 1994).

Illumina library preparation and sequencing

The paired-end sequencing library was prepared using the TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA). Approximately 200 ng of DNA was fragmented by Covaris M220 to generate a mean fragment distribution of 400 bp. The product was subjected to end repair, followed by adapter ligation to the fragments. The ligated products were size-selected using AMPure XP beads. The size selected product was PCR amplified with index primers following the manufacturer's protocol. PCR amplified library was analyzed on Tape Station 4200 (Agilent Technologies) using high sensitivity D1000 Screen Tape assay kit following manufactures instructions. The paired-end (PE) library was sequenced on Illumina NextSeq 500 platform using 2×150 bp sequencing chemistry at Eurofins Genomics India Pvt. Ltd, Bengaluru, India.

Data processing and genome assembly

High quality reads ($Q > 30$) were obtained after trimming the raw data using TrimGalore (v6.0) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), unpaired reads were also retained. Kmergenie was employed to estimate the best kmer size for the assembly. Reads were assembled with assemblers such as SPAdes: St. Petersburg genome assembler (auto kmer selection) (Nurk et al. 2013), Velvet ($K = 37$) (Zerbino and Birney 2008), ABySS: Assembly By Short Sequences (K within a range of 21–71) (Jackman et al. 2017), MaSuRCA: Maryland Super-Read Celera Assembler(Auto) (Zimin et al. 2013), IDBA-UD: Iterative De Bruijn Graph De Novo Assembler (K within a range of 21–71) (Peng et al. 2012). Assembled genomes were compared among each other with metrics like N50, number of gaps and number of scaffolds and total length.

Assembly QC and comparison

The quality of genome assembly was assessed using QUAST (Quality Assessment Tool) (Gurevich et al. 2013) by accounting for the number of scaffolds and the N50. The scaffold level assembly of UV-8b strain from China (GCA_000687475.1) and MAFF 236576 strain from Japan (GCA_000965225.2) was used as references to assess the genome quality of Uv-Gvt strain (Fig. 1). The BUSCO analysis (Simao et al. 2015) of Uv-Gvt genome determines the completeness of genome assembly and also RNA's were predicted using with QUAST.

Gene identification and annotation

Gene prediction was carried out using GeneMark-ES (Ter-Hovhannisyan et al. 2008). First, predicted protein sequences were annotated using PANNZER: Protein ANNotation with Z-scoRE powered by SANS: Suffix Array Neighborhood Search (https://ekhidna2.biocenter.helsinki.fi/sanspanz/). Second, the predicted transcripts and proteins were annotated against the NCBI non-redundant database using DIAMONDBLASTX and BLASTP (Basic Local Alignment Search Tool) programs, respectively, with parameters like "—max-target-seqs 1-e 1e-5—id 80—more-sensitive—subject-cover 100" also a lenient annotation was carried out using parameters such as "—max-target-seqs 1-e 1e-5—id 80". Furthermore, protein description among Uv-Gvt, Uv-8b and MAFF 236576, were compared to identify and report the unique genes found in Uv-Gvt. MAFF 236576 genes were all available as hypothetical proteins in the NCBI genome database. Thus we inferred the annotation from protein sequences using PANNZER, merged with available annotations and carried out the comparisons.

Identification of genes involved in host–pathogen interaction and effector protein coding

Ustilaginoidea virens genes involved in interaction with the rice plant during disease development were identified using PHIbase (v-4.2) database. BLASTX was performed

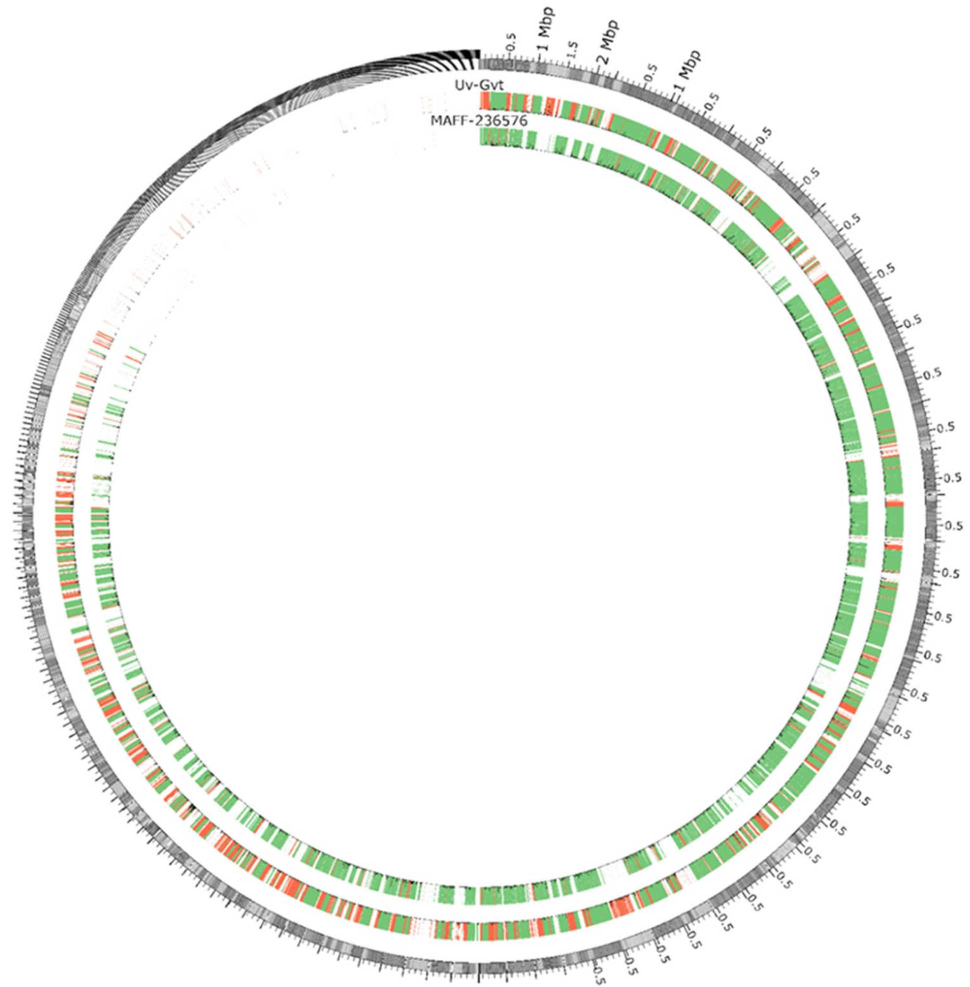

Fig. 1 Comparison of assembly quality of Uv-Gvt genome with other available genome of *Ustilaginoidea virens* in Genbank. GC (%) heat map from 0% (white) to 77% (black) and plot corresponds to a relative scale. Outer ring is UV-8b, inner rings are Uv-Gvt and MAFF_236576, respectively. Assembly tracks are combined with mismatches visualization, higher columns indicates larger mismatch rate. Green indicates a match and red indicate mismatch

with *e* value of 1e-5 for significant homology. Similarly, genes involved in coding effector proteins were also identified using PHIbase database.

KEGG pathway analysis

All the predicted genes of the Uv-Gvt strain were mapped to reference canonical pathways in KEGG. All mapped genes were classified under five categories: Metabolism, Cellular processes, Genetic information processing, Environmental information processing, and Organism systems. The output of KEGG analysis includes KEGG Orthology (KO) assignments and corresponding EC numbers and metabolic pathways of predicted genes using a KEGG automatic annotation server (KASS) (https://www.genome.jp/kass-bin/kaas_main).

Repeat and transposon identification

Genomes from all three strains were subjected to repeat analysis using RepeatMasker (Smit et al. 2008) configured with NCBI rmblast. With custom library built using RepeatModeler (Smit et al. 2008) assembled genome as input, the percentage of bases, giving rise to simple repeats and low complexity regions were identified.

Variant analysis

To identify the genetic variations in the Uv-Gvt strain compared to UV-8b and MAFF 236576 strains, Uv-Gvt raw reads were aligned to the UV-8b and MAFF 236576 masked references using Bowtie2 (Langmead and Salzberg 2012). Later, conversions and indexing of the alignment, results, and references were carried out using samtools (Heng et al. 2009). Further, variant detection, haplotype calling, and genotyping were performed using the Genome Analysis Toolkit (GATK) (DePristo et al. 2011) with recommended best practices for germline SNPs and InDel calling (https://software.broadinstitute.org/gatk/best-practices/workflow?id=11145). Some changes like utilization of Haplotype Caller at first to avail the input for Base Quality Recalibration were applied. "QD < 2.0 || FS > 200.0 || Read Pos

Rank Sum < −20.0 || Inbreeding Coeff < −0.8", "SB ≥ 0.10 || QD < 5.0 || HRun ≥ 4" and other parameters like cluster size of 3, mask extension 5, cluster window size 10 were applied, also a bed file obtained from the results of RepeatMasker (Smit et al. 2008) on the respective genome reference used in the analysis to filter the variants. Further SNPs in and around the InDels were removed, and effect predictions were made using SnpEff (Cingolani et al. 2012) with default parameters, but UV-8b and MAFF 236576 custom effects databases built using the respective references and annotations (https://snpeff.sourceforge.net/SnpEff_manual.html#databases).

Phylogenetic analysis and divergence dating

Phylogenetic position for *U. virens* strains (Uv-Gvt, UV-8b and MAFF 236576) was studied in relations to seven true smut fungi (basidiomycetes) and 17 ascomycetes fungi (13 belonging to family *Clavicipitaceae*, three belonging to Hypocreales-incertae-sedis and one belonging to family *Bionectriaceae*) along with one species of zygomycetes as outgroup. The representative genome sequences of each organism were obtained from the NCBI genome database (Online Resource 2). Each genome was then subjected to gene prediction using GeneMark-ES. The predicted genes were annotated using DIAMOND (Buchfink et al. 2014) program against NCBI-nr database with an *e* value cut-off of 1e-5. Protein sequences were obtained from the results. Two hundred and fifty-five single copy orthologs genes were clustered and extracted with the help of OrthoMCL (Li et al. 2003). T-Coffee was run in the mcoffee mode for each cluster, where it implements five multiple sequence aligners namely, clustalw2, t_coffee, POA: Partial Order Alignment, MUSCLE: MUltiple Sequence Comparison by Log-Expectation, MAFFT: Multiple Alignment using Fast Fourier Transform, DIALIGN-T, PCMA: Profile Consistency Multiple Sequence Alignment and ProbCons: Probabilistic Consistency-based Multiple Alignment of Amino Acid Sequences, and thus provides a combined multiple sequence alignment (msa) (Wallace et al. 2006). Obtained protein msa was converted to CDS msa using PAL2NAL (Suyama et al. 2006), poorly aligned regions in the msa were removed using Gblocks (Talavera and Castresana 2007) four-fold degenerate sequence alignments were extracted with the help of R package named RPHAST: Phylogenetic Analysis with Space/Time Models in R (Hubisz et al. 2010). 227 successfully converted CDS alignments were concatenated to form a supergene, which was utilized for phylogenetic reconstruction. The evolutionary relationships were calculated based on UPGMA method (Sneath and Sokal 1973) with bootstrap consensus tree (Felsenstein 1985) inferred from 500 replicates, distance calculated based on the number of substitutions per site using Maximum Composite Likelihood method in MEGAX: Molecular Evolutionary Genetics Analysis (Kumar et al. 2018). Thus, the obtained phylogenetic tree was subjected to PAML: Phylogenetic Analysis by Maximum Likelihood MCMCTree analysis (Yang 2007). The available fossil calibration for the divergence of ascomycetes and basidiomycetes (608 Mya, 549–701.5 Mya for 95% HPD) was coded on to the user tree. The output tree was read into R using the function read MCMC (Markov Chain Monte Carlo) Tree provided by the package MCMCTreeR (https://github.com/PuttickMacroevolution/MCMCTreeR) and plotted using the function geoscalePhylo provided by the package "strap" (Bell and Lloyd 2015). The edge length, node labels were set as mean node ages multiplied by 100 and rounded to 1 decimal point, root time was set as 100, and the root edge value was set as the difference between root time and maximum of the edge lengths. Age estimation is represented in million years ago (Mya) followed by their highest posterior density (HPD) in parentheses, which is based on a 95% confidence interval of all sample values. Divergence time estimation calculates stem and node ages for each clade. Subsequently, the stem ages are considered as the time a group originated-or differentiated from its sister clade, and the node ages as the time a group started diversification (Hedges et al. 2006).

Results

Disease status and morphometric characterization of pathogen isolates

During the survey, mean disease severity was ranged between 4.44−17.2% in different rice ecosystems (Online Resource 3). Observations on per cent infected tiller, percent grains infected and disease severity are presented in an additional file (Online Resource 3). The highest mean disease severity was observed in the Bhadra ecosystem (17.12%) followed by the Hilly ecosystem (10.73%), NEK ecosystem (6.71%), and Cauvery ecosystem (4.44%).

Following the single chlamydospore isolation technique, we regenerated pure cultures for all the 15 isolates, which were designated as Uv-1 to Uv-15 and were characterized morphologically. On PSA medium, all isolates exhibited significant variability in colony color, growth pattern and chlamydospore morphology (color, shape and size) (Online Resource 4). Isolates from Hilly ecosystems produced light-brown to brown colored chlamydospores, whereas, those from the irrigated ecosystem (NEK and Cauvery) produced brown to dark-brown chlamydospores (Online Resource 4). A significant variability was recorded in size of the chlamydospores (18.60 −104.29 μm^2) wherein, isolates from NEK and Cauvery irrigated ecosystem produced smaller chlamydospores (18.60−50.87 μm^2), compared to isolates other regions (50.87−104.29 μm^2) (Online Resource 4). No

correlation was observed between isolates and ecosystems for colony growth pattern, colony color, mycelial color and chlamydospore shape.

Genome sequencing and assembly comparison

Based on the virulence analysis, a virulent strain Uv-Gvt was sequenced to generate 10,771,481 high quality reads after trimming (low quality reads and adapter sequences) which corresponds to ~ 75 × sequencing depth. Next the high-quality reads of Uv-Gvt were assembled multiple times with different kmer sizes, the assembled genome length of ~ 26.96 Mb (un-gapped length ~ 26.85 Mb) constructed from kmers of size 59 was considered for the downstream analysis as it yielded an N50 of 15,534 Kb. A total number of 46,897 contigs (N50 6586 bp and L50 1253 bp) were scaffolded into 9157 scaffolds (Online resource 5). The genome constituted a G + C content of 54.92% with 110,401 assembly gap length. Complete genome assembly statistics of Uv-Gvt is shown in Table 1. The BUSCO analysis reported 96.2% complete genes, comprised 96.2% being single copy genes and 0% duplicated genes, 3.1% of genes were suggested to be fragmented and 0.7% of genes were suggested to be missing out of a total 290 predicted genes. BUSCO simple notation of the same results, *C*: 96.2% [*S*: 96.2%, *D*: 0.0%], *F*: 3.1%, *M*: 0.7%, *n*: 290. The genomic resources for *U. virens* strain Uv-Gvt was deposited in DDBJ/EMBL/GenBank under the accession number PGGP00000000 (Bio Project Id: PRJNA414696 and BioSample ID: SAMN07807409). The version described in this paper is version PGGP00000000.2

Table 1 Genome statistics of Uv-Gvt strain

Genomic features	Predicted quantity/value
Total sequence length (bp)	26,967,352
Un-gapped length (bp)	26,856,951
Total assembly gap length (bp)	110,401
G + C (%)	54.92
No of protein coding genes	6627
Average gene length (bp)	1322.55
Gene density (per Mb)	254.74
Number of scaffolds	9157
Scaffold N50 (bp)	15,534
Scaffold L50 (bp)	495
Number of contigs	46,897
Contig N50	6586
Contig L50	1253
Depth of sequencing	75×

Statistics were assessed after filtering for minimum scaffold length of 200 bp and for contaminations

Gene identification and annotation

Gene prediction using GeneMark-ES identified 6628 protein-coding genes in the Uv-Gvt strain with an average gene length of 132.552 and gene density of 254.74 genes per 1 Mb (Online Resource 6). However, UV-8b strain from China harbors 8426 protein-coding genes with an average gene length of 1627 bp and gene density of 214 genes per 1 Mb. Similarly, the genome of MAFF 236576 strain harbors 6451 protein-coding genes with an average gene length of 764 bp and gene density of 192 (Kumagai et al. 2016).The comparative genomics analysis of all three *U. virens* genomes are shown in Table 2. Comparative genome analysis of three strains Uv-Gvt, UV-8b and MAFF 236576 revealed the presence of 1756 unique genes in Uv-Gvt whereas, 5990 and 1975 genes are unique to Uv-8b and MAFF 236576 strains, respectively (Fig. 2). Further, annotations of identified unique genes are reported separately (Online Resource 7).

Functional annotation of predicted proteins

KEGG pathway analysis of all the predicted genes of Uv-Gvt genome identified diverse protein families (Pfam) which are associated with diverse pathways. The pathways were categorized into three types viz., cellular components, molecular processing and biological processing (Fig. 3 and see Online Resource 8). Functional annotation of predicted proteins revealed that the Uv-Gvt strain encodes 42 lipases, one pectinesterase, 161 dehydrogenases, 7 cutinases and 14glycoside hydrolases which are known to assist during pathogenesis. In the Uv-Gvt, Carbohydrate Active Enzymes (CAZyme) (glycoside hydrolases) are almost similar in number as that of UV-8b (Table 2). Whereas, MAFF 236576 is more abundant in carbohydrate degrading enzyme. The Uv-Gvt genome encodes two pectin lyases and a pectate lyase, which are essential for cell wall degradation during infection. Interestingly, UV-8b genome does not encode any pectate lyases whereas; MAFF 236576 strain encodes two pectate lyases. Phytopathogenic fungi utilize many signaling pathways during infection process. Uv-Gvt genome harbors five mitogen-activated protein kinase (MAPK) pathways and 17 GPCRs. In the Uv-Gvt strain, 73major facilitator superfamily (MFS) transporters were identified as compared to 80 and 26 MFS present in the UV-8b and MAFF 236576 genome, respectively (Table 2). Sugar/glucose-sensing receptor such as GPR1 protein is absent in Uv-Gvt as in UV-8b strain.

Genes involved during host–pathogen interaction and effector proteins

PHIbase (v-4.2) database search for de novo annotated genes of Uv-Gvt genome hits 422 homologous proteins which are

Table 2 Comparative genomics of different *U. virens* strains

S. no.	Genomic features	Uv-Gvt[a]	UV-8b	MAFF 236576
1	Genome size	26.967	39.397	33.567
2	Coverage	75 ×	142 ×	107 ×
3	GC content	54.92	49.9	51.3
4	% Repeat[b]	19.72	6.81+	6.40+
5	Total number of genes	6627	8426	6451
6	Average gene length	1322.55	1627	764
7	Gene density	245.74	213.87	192.17
10	Fungal specific transcription factors	26	116	23
11	Zinc finger transcription factor	2	25	3
12	Major facilitator family (MFS)	73	80	26
13	Protein kinases	227	229	221
14	Subtilisin	4	12	10
15	Aspartic proteases	10	8	9
16	Lipase	42	39	39
17	Pectinesterases	1	1	1
18	Glycoside hydrolases	14	14	19
19	Dehydrogenases	161	146	163
20	Cutinases	7	8	6
21	Pathogen–host interaction proteins[c]	422	359[a]	520[a]
22	Effector proteins	2	2	1
23	Pectin lyases	1	0	2
24	Pectate lyases	0	0	0
25	Polyketide synthases	4	11	2
26	Geranylgeranyl diphsphate synthases	0	1	0
27	RNA silencing components	5	5	10
28	Transposable Elements	1838	2056	1365
27	SNP Profile in UV-Gvt	NA	34,459	25,700

[a]Statistics inferred based on the predictions

[b]Hits having more than 60% identity

[c]Statistics were assessed before filtering for minimum scaffold length of 200 bp and for contaminations

putatively involved in host infection and post-infection host modification (Online Resource 9). Interestingly, we found only 359 HPI proteins in the UV-8b strain against 1103 proteins originally reported by Zhang et al. (2014) whereas, 520 HPI proteins are found in MAFF 236576 strain.

Repetitive elements and variant analysis

Comparative genome analysis of three strains of *U. virens* revealed the presence of 19.72% repetitive DNA, whereas, UV-8b contains less repetitive DNA (6.81%), followed by MAFF 236576 (6.4%) (Table 2). In all three strains, most of the repeats (approximately 75 −77%) are transposable elements (TEs) (Table 3). The most abundant repeat type was a simple repeat (2.66−3.09%) followed by low complexity repeat (0.3−0.4%) (Table 3). In addition, the most abundant retroelements were LTR elements (n = 1186−1855) followed by LINEs/Non-LTR elements (n = 24−184) (Table 3).

Genome-wide SNPs profile of the Uv-Gvt genome was examined and compared with UV-8b and MAFF 236576 genome. The number of SNPs ranged between 25,700 (between Uv-Gvt and MAFF 236576, 0.073% nucleotide diversity) to 34, 459 (between Uv-Gvt and UV-8b, 0.088% nucleotide diversity) (Table 4). We also observed higher SNP density in the intergenic regions as compared to the genic region.

Gene silencing components

RNA silencing pathways are well conserved in the Uv-Gvt genome, which includes one Argonaute-like protein (Gene Id: 5323_g), three RNA dependent RNA polymerase (Gene Id: 3027_g, 410_g and 4207_g) and one RecQ DNA helicase (Gene Id: 2461_g) (Online Resource 10). No Dicer protein was predicted in Uv-Gvt, whereas; UV-8b and MAFF 236576 strains contain two and three dicer proteins, respectively (Online Resource 10).

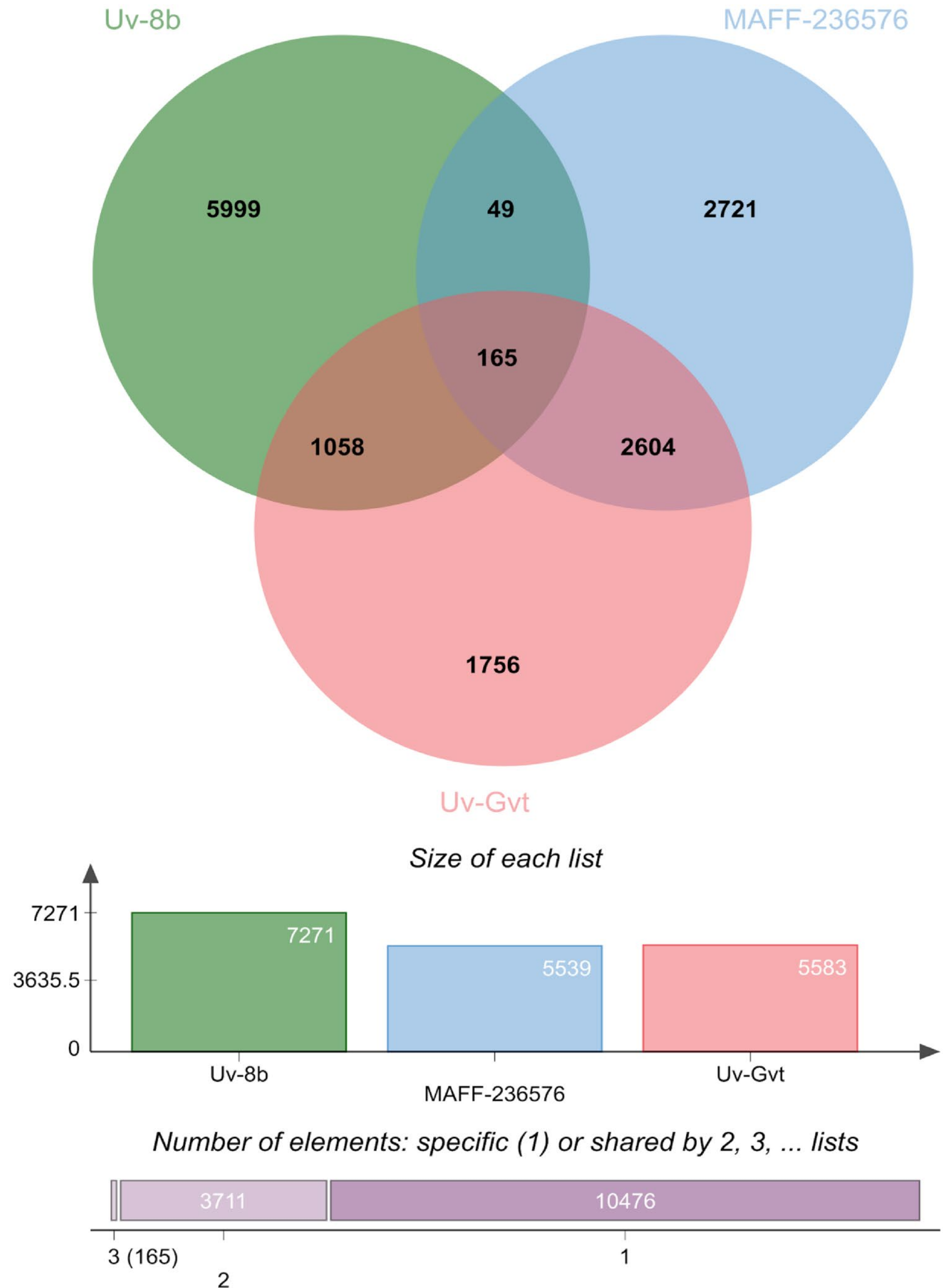

Fig. 2 Venn diagram showing common orthologous and unique sets of genes across three strains

Phylogeny and divergent time analysis

The genome of 22 fungi (includes 17 ascomycetes, 7 basidiomycetes and 1 zygomycetes as outgroup) which are taxonomically or symptomatically related to *U. virens* was considered for phylogenetic study (Online Resource 2). It revealed that all members of the *Clavicipitaceae* family (including three strains of the *U. virens*) are grouped within the ascomycetes cluster, whereas, other four fungi (three Hypocreales-incertae-sedis and one *Bionectriaceae*) form a separate clade within the ascomycetes cluster (Fig. 4). The clustering pattern also revealed that all *U. virens* strains formed a sub-clade within the *Clavicipitaceae* clade and showed clear separation from true smut causing fungi (i.e. basidiomycetes). Further, *U. virens* is more closely related to *Moelleriella libera* (*Clavicipitaceae*, Ascomycetes) an entomopathogenic fungus that lives parasitically on arthropods.

Furthermore, divergence time analysis using fossil calibration showed that the ascomycetes (includes false smut fungi) and basidiomycetes (true smut fungi) diverged from each other at a mean time of ~ 718.6 Mya

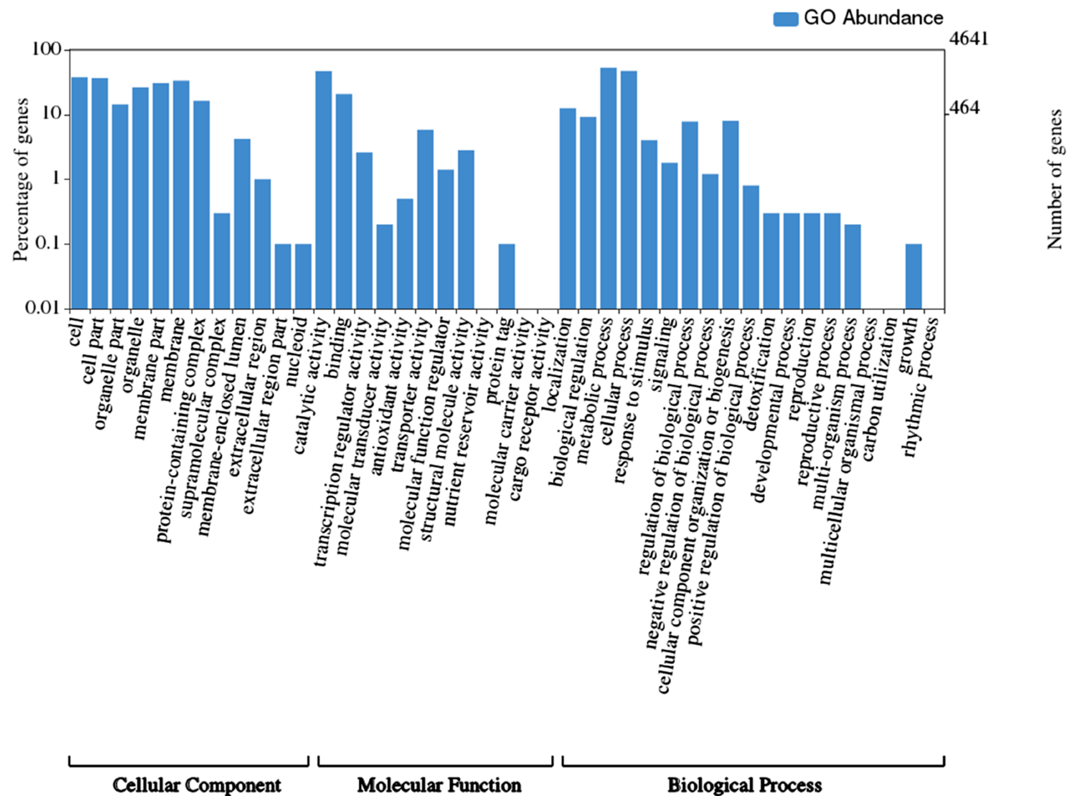

Fig. 3 Gene ontology (GO) annotation of predicted genes of *Ustilaginoidea virens* (Uv-Gvt) genome

Table 3 Repeat elements and transposable elements (TEs) identification in different *U. virens* strains

Strain	Repeat elements (%)			Transposable elements (TEs) (family count)		
	Simple repeats	Low complexity repeats	Total repeats	DNA transposon	LTR retrotransposon	Non-LTR retrotransposon
Uv-Gvt	2.7	0.3	3	25	1789	24
UV-8b	2.66	0.34	3	17	1855	184
MAFF 236576	3.09	0.4	3.49	15	1184	164

Statistics inferred based on the predictions

Table 4 Variant profile of Uv-Gvt compared to Uv-8b and MAFF 236576

Strain	Total mutations	SNPs	Insertions	Deletions	SNPs in intergenic region	SNPs in intronic region	# Missense mutations
UV-8b	40,649	34,459	3595	2595	27,868	1285	2638
MAFF 236576	30,621	25,700	2235	2686	20,778	943	1871

at 95% HPD. Later, *Clavicipitaceae* (includes *U. virens*) diverged from non-*Clavicipitaceae* members at a mean time of ~ 246.8 Mya at 95% HPD. It was also observed that Uv-Gvt (Indian strain) diverged from UV-8b (Chinese strain) and MAFF 236576 (Japanese strain) at a mean time of 22.6 Mya at 95% HPD whereas, UV-8b and MAFF 236576 diverged from each other at a mean time of 8.3 Mya at 95% HPD (Fig. 4).

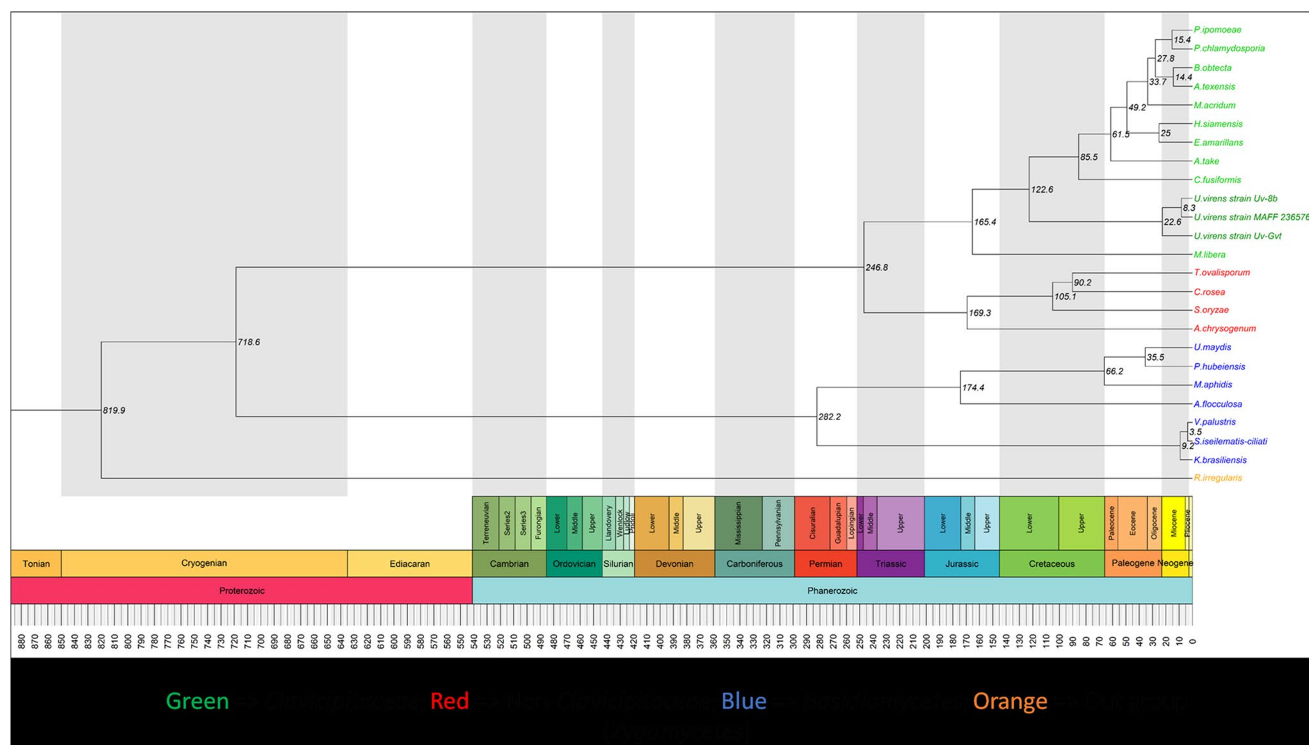

Fig. 4 Phylogenetic tree depicting the estimated divergence time. The node ages is represented in sMya (million year ago). UV-8b and MAFF 236576 were diverged from Uv-Gvt at a mean of 22.6 Mya and UV-8b and MAFF_236576 diverged from each other at a mean of 8.3 Mya. True smut fungi were diverged from false smut (ascomycetes) at a mean time of 718.6 Mya

Discussion

Rice false smut disease first reported from India (Cooke 1878) is an emerging global threat (Rush et al. 2000; Singh and Pophaly 2010; Zhang et al. 2014; Tsukui et al. 2014; Kumagai et al. 2016, Li et al. 1995). Although, the pathogen was shown to possess genetic diversity among its geographically distinct populations (Sun et al. 2013), limited genomic information is a matter of concern across rice-growing countries where, the incidence of false smut is prevalent. In the Karnataka state of India, major rice-producing regions include NEK Cauvery, Bhadra and Hilly upland ecosystems, which are characterized by the distinct soil, weather, crop-variety, cropping sequence and, diverse disease and pest profile. Therefore, a better understanding of the disease incidence across all eco-regions is essential prior to pathogen characterization.

In our study, we observed highest severity in Bhadra ecosystem (17.12%) and it could be attributed to extensive cultivation of high yielding cultivars in two consecutive seasons (*Kharif* and *Rabi*) and more frequent rains during *Kharif* (data not shown). Conversely, in the hilly upland ecosystem, despite the high prevalence of rainfall, the disease severity was low (10.73%), and it could be attributed to single cropping in *Kharif*, cultivation of traditional rice cultivars and less fertilizer use. Previous investigations have also reported high incidences of false smut disease under intensive rice cultivation system characterized by the higher use of chemical fertilizers (Ikegami 1962; Hegde and Anahosur 2000; Ladhalakshmi et al. 2012).

Ecologically distinct isolates of *U. virens* collected during the survey exhibited significant morphometric diversity in their mycelial, cultural and spore characters. We observed a correlation between the isolates and ecosystem for their spore color and size, whereas, no such grouping was observed between the other morphological traits and ecosystem. In India, many research groups have reported the morphological diversity among Indian isolates (Ladhalakshmi et al. 2012; Baite et al. 2014). However, no attempts were made to correlate the morphological traits with isolate's geographical origin. Interestingly, all isolates exhibited a difference in their virulence when artificially inoculated to different rice cultivars (data not shown).

Virulence study identified Uv-Gvt strain as highly virulent and therefore, was selected for whole-genome sequencing. Due to the non-availability of EST data, gene prediction in fungi was based on either de novo or comparative prediction models (Kellis et al. 2003; Galagan et al. 2005). The genome of many rice infecting fungal pathogens such as *M. oryzae* (Gowda et al. 2015), *Sarocladium oryzae*

(Hittalmani et al. 2016), *Rhizoctonia solani* (Nadarajah et al. 2017) are available in the database, which enables gene prediction easier. We predicted 6628 protein-coding genes in Uv-Gvt strain, whereas, 8426 and 6451 coding genes have been reported in UV-8b and MAFF 236576 strains, respectively (Table 2). The reduced number of genes in the Uv-Gvt and MAFF 236576 strain could be due to the difference in sequencing platforms and annotation tools. Moreover, about 183 genes predicted in UV-8b were not supported by RNAseq data (Zhang et al. 2014). Gene density in Uv-Gvt is higher compared to UV-8b and MAFF 236576 strains, which is still less than most of the other sequenced ascomycetes (Zhang et al. 2014; Hittalmani et al. 2016; Nowrousian et al. 2010; Dean et al. 2005).

Comparative genomics of three *U. virens* strains revealed the presence of 1756 unique genes in Uv-Gvt strain, whereas, 5999 and 2721 genes are unique in UV-8b and MAFF 236576 strains, respectively. Further, gene ontology database suggested that these unique genes are involved in many biological, molecular and cellular pathways. Functional annotation of predicted proteins in Uv-Gvt genome revealed the reduced inventory of CAZyme similar to that of UV-8b and MAFF 236576 strains. In *U. virens* a lower CAZyme may be responsible for a biotrophic way of pathogenesis in the floral niche of rice as reported by Zhang et al. (2014). Interestingly, our study predicted the pectate lyases in Uv-Gvt and MAFF 236576 genome, whereas, Uv-8b strain from China does not harbor any pectate lyases. The pectin and pectate lyases are essential for degrading host cell wall during infection and are abundant in hemibiotrophic and necrotrophic fungi, such as *Fusarium graminearum* (Zhang et al. 2014) but their role in biotrophic *U. virens* is not clear.

Phytopathogenic fungi utilize many signaling pathways such as GPCRs, MAPK, and cyclic-adenosine monophosphate (c-AMP) to recognize their host and further proceeds with infection process (Zhang et al. 2014; Lee et al. 2003; Xue et al. 2008; Jiang et al. 2018). The Uv-Gvt genome harbors five MAPK pathways similar to the UV-8b strain and *Saccharomyces cerevisiae,* suggesting the presence of similar signaling pathways as that of other phytopathogenic fungi. Similarly, Uv-Gvt encodes only 17 GPCRs and it has been reported previously that *U. virens* encodes fewer GPCRs than other hemibiotrophic fungi such as *M. oryzae* (up to 50) (Zhang et al. 2014). Sugar/glucose-sensing receptor such as GPR1 protein is absent and only fewer cellular transporters are present in Uv-Gvt genome. Reduced number of cellular transporters and lack of GPR1 protein has also been reported in UV-8b strain (Zhang et al. 2014). This finding supports the earlier findings that *U. virens* can adapt only to a narrow host range such as rice and within rice; it can only colonize floral organs (Lorenz et al. 2000; Zhang et al. 2014).

Host–pathogen interaction is a multifaceted process wherein all phytopathogenic fungi encode diverse proteins to modify their host during infection and disease development (Boyd et al. 2013; Gupta et al. 2015). The *U. virens* colonize floral organs of rice encodes diverse protein families to facilitate its infection process (Zhang et al. 2014). We identified 422 homologous proteins in the Uv-Gvt genome, which putatively involved in host infection and post-infection host modification. Interestingly, we could predict 359 PHI genes in UV-8b strain against 1103 of them previously reported (Zhang et al. 2014). Most of the PHI protein search hits matched with different plant pathogenic fungi (*F. graminearum*, *M. oryzae*, *Colletotrichum gloeosporioides*, *C. lagenarium*, *U. maydis,* etc.,), human pathogens (*Candida albicans*and*Aspergillus fumigates*), phytopathogenic bacteria (*Agrobacterium tumefaciens*, and *Ralstonia solanacearum*) and entomopathogenic fungi (*Metarhizium robertsii*, and *M. anisopliae*). All the predicted proteins are involved in host penetration, effectors transportation, synthesis, modification of secondary metabolites, transcription factors and other essential cellular processes. Prior studies have demonstrated the involvement of host genes during *U. virens* infection and disease pathogenesis through a transcriptomics approach (Chao et al. 2014; Han et al. 2015). However, there is limited information on genes involved in pathogenesis (Zhang et al. 2014). Furthermore, there is a need for more functional studies to validate the role of PHI genes in infection and disease development in rice.

Repetitive DNA in a fungal genome plays a vital role in genome evolution by providing an opportunity for acquiring genetic diversity, but it is also detrimental to the genome in terms of genetic stability (Galagan et al. 2005; Hittalmani et al. 2016). Comparative genome analysis of three strains of *U. virens* revealed the presence of 6.40−19.72% of repetitive DNA. Pathogenic ascomycetes are known to harbor more repetitive DNA in their genome (Hittalmani et al. 2016). The percentage of repetitive DNA of *U. virens* is higher as compared to other available ascomycetes genomes wherein only 3−10% repeats have been reported (Galagan et al. 2005; Gowda et al. 2015; Dean et al. 2005). A higher proportion of repetitive DNA has also been reported previously for UV-8b strain (Zhang et al. 2014). DNA transposons are more abundant in *U. virens* and we could predict 0.23−0.28% of total repetitive DNA as DNA transposons. The adverse effects of a large proportion of repetitive sequences in *U. virens* have been kept under control through RNA silencing and repeat-induced point mutations (Zhang et al. 2014). Genome-wide SNPs profile indicated the presence of low genetic diversity (0.073−0.088%) among the three geographical strains. However, the existence of significant genetic diversity among the geographical isolates from China has been reported previously (Sun et al. 2013). Our results corroborated with the previous report, where 0.0728−0.0926% diversity was

reported for different geographical strains of China (Zhang et al. 2014).Future studies, including more strains globally will shed light on the genetic diversity of *U. virens* populations.

Phylogenetic analysis of *U. virens* strains (Uv-Gvt, UV-8b and MAFF 236576) revealed that all three strains clustered together as sub-clade within the *Clavicipitaceae* clade (Fig. 4). The current study also revealed that *U. virens* is more closely related to entomopathogenic fungi such as *M. libera*. Previously, it was shown that *U. virens* is more closely related to entomopathogenic fungi, *M. anisopliae* and *M. acridum*, than other plant pathogenic fungi (Zhang et al. 2014). Molecular clock hypothesis has been useful in estimating divergence time among different fungal taxon (Boutin and Coineau 2000; Bromham and Penny 2003). The fossil calibration for the divergence of ascomycetes and basidiomycetes has been updated recently as Sinian Ediacaran to Early Cambrian (608 Mya, 549–701.5 Mya for 95% HPD) (Hedges et al. 2006) and the same can be coded on to the phylogenetic tree to ascertain the divergence time of any taxon under study. Our study found that basidiomycetes diverged from ascomycetes at a mean time of '~18.6 Mya and later, *Clavicipitaceae* diverged from other ascomycetes at a mean time of ~246.8 Mya. The present study also revealed that strains from Japan and China, diverged from the Indian strain at a mean time of ~22.6 Mya. The divergence of *U. virens* strains based on the phylogeny of merely three strains may be insufficient; nevertheless, this study provides maiden evidence. Further studies using a larger set of isolates from different parts of the world will provide insights on the strain evolution and divergence.

False smut disease of rice is assuming an epidemic state in most of the rice-growing regions of the world contributing to significant yield reduction as well as toxin contamination of food grains. Presently, fungicide sprays during flowering and post-flowering stages are the only recommended strategy to manage this disease (Muniraju et al. 2017a, b). However, the application of fungicides anytime close to the crop harvest stage is not advisable due to potential risk of grain and straw contamination with the fungicide residues. Therefore, the development of disease-resistant varieties is an alternative eco-friendly strategy that requires a clear knowledge of pathogen biology, population structure, genetic diversity, and effector profile. Therefore, the information generated in this study may facilitate further understanding of host–pathogen interactions and pathogen diversity, which in turn could expedite the resistance breeding programs in rice.

Conclusion

False smut disease of rice was first reported from India, caused by ascomycetes fungi (unlike true smut fungi belonging to basidiomycetes) as an emerging threat to rice cultivation in India and beyond. Although genomic resources have been developed for strains reported from China and Japan, no attempts were made to study an Indian strain. Based on the virulence profile of 15 distinct isolates, a virulent strain (Uv-Gvt) yielded 26.96 Mb genome after de novo assembly. Comparative genomics among three Asian strains revealed the 6.40−19.72% of repetitive DNA in the *U. virens* genome and reduced inventory of CAZyme. A total of 422 host–pathogen interacting genes were predicted and further, functional characterization of these genes would aid in better understanding the *Ustilaginoidea*-rice interaction during disease development. Also, the information on molecular divergence time revealed the evolution of *U. virens* from other fungal taxons and the intra-species divergence among different strains. Overall, this study provides insights into pathogen population diversity at the global level and also contribute to the development of species/strain-specific assays.

Acknowledgements This project was funded by the Early Career Research (ECR) Grant (Project Number: ECR/2017/000246) of Science and Engineering Research Board (SERB), Government of India to Pramesh D. We are thankful to Dr. Raju Soolanayakanahally, Saskatoon Research Center, Agriculture and Agri-food Canada, Canada and Dr. C. T. Manjunath Prasad, Wageningen Campus, Wageningen University and Research, The Netherlands for critically reviewing the manuscript.

Author contributions DP and MKP conceived the project, arranged the funds, designed the experiments and wrote the manuscript. KMM, AS, EC, MKK and HS conducted the survey, collected isolates, cultured and performed morphometric study. HBM, PHD, CM, MKY, BSR, GB and AJD submitted the raw data, carried out assembly, annotation, pathway analysis, phylogenetic study, SNP profiling, prepared tables and figures and wrote the manuscript. All authors read and approved the final manuscript for publication.

Availability of data and materials The Whole-Genome Shotgun project for *U. virens* strain Uv-Gvt was deposited in DDBJ/EMBL/GenBank under the accession number PGGP00000000 (Bio Project Id: PRJNA414696 and Bio-Sample ID: SAMN07807409). The version described in this paper is version PGGP00000000.2.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest in the publication.

References

Anonymous (2016) Production oriented survey. ICAR-Indian Institute of Rice Research, Hyderabad

Ashizawa T, Takahashi M, Moriwaki J, Hirayae K (2011) A refined inoculation method to evaluate false smut resistance in rice. J Gen Plant Pathol 77(1):10–16

Atia MMM (2004) Rice false smut (*Ustilaginoidea virens*) in Egypt. J Plant Dis Prot 111(1):71–82

Baite MS, Sharma RK, Devi TP, Sharma P, Kamil D (2014) Morphological and molecular characterization of *Ustilaginoidea virens* isolates causing false smut of rice in India. Indian Phytopathol 67(3):222–227

Bell MA, Lloyd GT (2015) strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. Palaeontology 58:379–389

Boutin C, Coineau N (2000) Evolutionary rates and phylogenetic age of some stygobiontic species. In: Wilkens H, Culver DC, Humphreys WF (eds) Subterranean ecosystems—ecosystems of the world 30. Elsevier, Amsterdam, pp 433–451

Boyd LA, Ridout C, O'Sullivan DM, Leach JE, Leung H (2013) Plant–pathogen interactions: disease resistance in modern agriculture. Trends Genet 29(4):233–240

BromhamL PD (2003) The modern molecular clock. Nat Rev Genet 4:216–224. https://doi.org/10.1038/nrg1020

Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12(1):59

Chao J, Jin J, Wang D, Han R, Zhu R, Zhu Y, Li S (2014) Cytological and transcriptional dynamics analysis of host plant revealed stage-specific biological processes related to compatible rice-*Ustilaginoidea virens* interaction. PLoS ONE 9(3):0091391

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Ruden DM, Lu X (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6(2):80–92

Cooke MC (1878) Some extra–European fungi. Grevillea 7:13–15

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434(7036):980–986

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hart C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5):491

Dodan DS, Singh R (1996) False smut of rice: present status. Agric Res 17(4):227–240

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4):783–791

Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B (2005) Genomics of the fungal kingdom: insights into eukaryotic biology. Genome Res 15(12):1620–1631

Gowda M, Shirkem MD, Mahesh HB, Chandaranam P, Rajamanim A, Chattoo BB (2015) Genome analysis of rice-blast fungus *Magnaportheoryzae* field isolates from southern India. Genomics Data 5:284–291

Gupta R, Lee SE, Agrawal GK, Rakwal R, Park S, Wang Y, Kim ST (2015) Understanding the plant-pathogen interactions in the context of proteomics-generated apoplastic proteins inventory. Front Plant Sci 6:00352

GurevichA SV, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075

Han Y, Zhang K, Yang J, Zhang N, Fang A, Zhang Y, Liu Y, Chen Z, Hsiang T, Sun W (2015) Differential expression profiling of the early response to *Ustilaginoidea virens* between false smut resistant and susceptible rice varieties. BMC Genomics 16(1):955

Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22(23):2971–2972

Hegde YR, Anahosur KH (2000) Effect of false smut of rice on yield components and growth parameters. Indian Phytopathol 53(2):181–184

Heng L, Bob H, Alec W, Tim F, Jue R, Nils H, Gabor M, Goncalo A, Richard D (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079

Hittalmani S, Mahesh HB, Mahadevaiah C, Prasannakumar MK (2016) De novo genome assembly and annotation of rice sheath rot fungus *Sarocladium oryzae* reveals genes involved in Helvolic acid and Cerulenin biosynthesis pathways. BMC Genomics 17(1):271

Hubisz MJ, Pollard KS, Siepel A (2010) PHAST and RPHAST: phylogenetic analysis with space/time models. Brief Bioinform 12(1):41–51

Ikegami H (1962) Studies on the false smut of rice: seedling inoculation with the chlamydospores of the false smut fungus. Ann Phytopathol Soc Jpn 1:27

Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I (2017) ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res 27(5):768–777

Jiang C, Zhang X, Liu H, Xu JR (2018) Mitogen-activated protein kinase signaling in plant pathogenic fungi. PLoS Pathog 14(3):1006875

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423(6937):241–254

Kumagai T, Ishii T, Terai G, Umemura M, Machida M, Asai K (2016) Genome sequence of *Ustilaginoidea virens* IPU010, a rice pathogenic fungus causing false smut. Genome Announc 4(3):00306–316

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing plat forms. Mol Biol Evol 35(6):1547–1549

Ladhalakshmi D, Laha GS, Singh R, Karthikeyan A, Mangrauthia SK, Sundaram RM, Thukkaiyannan P, Viraktamath BC (2012) Isolation and characterization of *Ustilaginoidea virens* and survey of false smut disease of rice in India. Phytoparasitica 40(2):171–176

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357

Lee N, D'Souza CA, Kronstad JW (2003) Of smuts, blasts, mildews, and blights: cAMP signaling in phytopathogenic fungi. Annu Rev Phytopathol 41(1):399–427

Li Y, Koiso Y, Kobayashi H, Hashimoto Y, Iwasaki S (1995) Ustiloxins, new antimitotic cyclic peptides: interaction with porcine brain tubulin. Biochem Pharmacol 49(10):1367–1372

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13(9):2178–2189

Lorenz MC, PanX HT, Cardenas ME, Xue Y, Hirsch JP, Heitman J (2000) The G protein-coupled receptor Gpr1 is a nutrient sensor that regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. Genetics 154(2):609–622

Lu DH, Yang XQ, Mao JH, Ye HL, Wang P, Chen YP, He ZQ, Chen F (2009) Characterising the pathogenicity diversity of *Ustilaginoidea virens* in hybrid rice in china. J Plant Pathol 91(2):443–451

Mandhare VK, Gawade SB, Game BC, Padule DN (2008) Prevalence and incidence of bunt and false smut in paddy (*Oryza sativa* L.) seeds in Maharashtra. Agric Sci Digest 28(4):292–294

Muniraju KM, Pramesh D, Mallesh SB, Mallikarjun K, Guruprasad GS (2017a) Disease severity and yield losses caused by false smut disease of rice in different rice ecosystems of Karnataka. Int J Microbiol Res 9(10):955–958

Muniraju KM, Pramesh D, Mallesh SB, Mallikarjun K, Guruprasad GS (2017b) Novel fungicides for the management of false smut disease of rice caused by *Ustilaginoidea virens*. Int J Curr Microbiol Appl Sci 6(11):2664–2669

Nadarajah K, Razali NM, Cheah BH, Sahruna NS, Ismail I, Tathode M, Bankar K (2017) Draft genome sequence of *Rhizoctonia solani* anastomosis group 1 subgroup 1A strain 1802/KB isolated from rice. Genome Announc 5(43):01188–1217

Nowrousian M, Stajich JE, Chu M, Engh I, Espagne E, Halliday K, Kamerewerd J, Kempken F, Knab B, Kuo HC, Osiewacz HD (2010) De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. PLoS Genet 6(4):1000891

Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R (2013) Assembling genomes and mini-metagenomes from highly chimeric reads. In: Annual international conference on research in computational molecular biology. Springer, Berlin, pp 158–170

Peng Y, Leung HC, Yiu SM, Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28(11):1420–1428

Rogers SO, Bendich AJ (1994) Extraction of DNA from plant, fungal and algal tissues. In: Gelvin SB, Schilperoort RA (eds) Plant molecular biology manual, vol 1. Kluwer Academic Publishers, Boston, MA, pp 1–8

Rush MC, Shahjahan AKM, Jones JP (2000) Outbreak of false smut of rice in Louisiana. Plant Dis 84(1):100

Sharma ND, Joshi LK (1975) False smut disease of rice. Curr Sci 44(10):352–354

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212

Singh AK, Pophaly DJ (2010) An unusual rice false smut epidemic reported in Raigarh District, Chhattisgarh, India. Int Rice Res Notes 35:1–3

Smit AFA, Hubley R, Green P (2008–2015) Repeat Modeler Open 1.0

Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. WF Freeman & Co., San Francisco, 573p

Sun X, Kang S, Zhang Y, Tan X, Yu Y, He H, Zhang X, Liu Y, Wang S, Sun W, Cai L (2013) Genetic diversity and population structure of rice pathogen *Ustilaginoidea virens* in China. PLoS ONE 8(9):e76879

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34(2):609–612

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56(4):564–577

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18(12):1979–1990

Tsukui T, Nagano N, Umemura M, Kumagai T, Terai G, Machida M, Asai K (2014) Ustiloxins, fungal cyclic peptides, are ribosomally synthesized in *Ustilaginoidea virens*. Bioinformatics 31(7):981–985

Verma RK, Singh RA (1988) Variations in *Claviceps oryzae-sativae* the incitant of false smut of rice. Indian Phytopathol 41(1):48–50

Wallace IM, O'sullivan O, Higgins DG, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res 34(6):1692–1699

Xue C, Hsueh YP, Heitman J (2008) Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. FEMS Microbiol Rev 32(6):1010–1032

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586–1591

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):821–829

Zhang Y, Zhang K, Fang A, Han Y, Yang J, Xue M, Bao J, Hu D, Zhou B, Sun X, Li S (2014) Specific adaptation of *Ustilaginoidea virens* in occupying host florets revealed by comparative and functional genomics. Nat Commun 5:s3849

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. Bioinformatics 29(21):2669–2677