

ARTICLE

Received 27 Jan 2014 | Accepted 10 Apr 2014 | Published 20 May 2014

DOI: 10.1038/ncomms4849

OPEN

# Specific adaptation of *Ustilaginoidea virens* in occupying host florets revealed by comparative and functional genomics

Yong Zhang<sup>1,\*</sup>, Kang Zhang<sup>1,\*</sup>, Anfei Fang<sup>1,\*</sup>, Yanqing Han<sup>1</sup>, Jun Yang<sup>1,2</sup>, Minfeng Xue<sup>1,2</sup>, Jiandong Bao<sup>3</sup>, Dongwei Hu<sup>4</sup>, Bo Zhou<sup>3,5</sup>, Xianyun Sun<sup>6</sup>, Shaojie Li<sup>6</sup>, Ming Wen<sup>7</sup>, Nan Yao<sup>7</sup>, Li-Jun Ma<sup>8</sup>, Yongfeng Liu<sup>9</sup>, Min Zhang<sup>10</sup>, Fu Huang<sup>11</sup>, Chaoxi Luo<sup>12</sup>, Ligang Zhou<sup>1</sup>, Jianqiang Li<sup>1</sup>, Zhiyi Chen<sup>9</sup>, Jiankun Miao<sup>13</sup>, Shu Wang<sup>13</sup>, Jinsheng Lai<sup>14</sup>, Jin-Rong Xu<sup>15</sup>, Tom Hsiang<sup>16</sup>, You-Liang Peng<sup>1,2</sup> & Wenxian Sun<sup>1</sup>

*Ustilaginoidea virens* (Cooke) Takah is an ascomycetous fungus that causes rice false smut, a devastating emerging disease worldwide. Here we report a 39.4 Mb draft genome sequence of *U. virens* that encodes 8,426 predicted genes. The genome has ~25% repetitive sequences that have been affected by repeat-induced point mutations. Evolutionarily, *U. virens* is close to the entomopathogenic *Metarhizium* spp., suggesting potential host jumping across kingdoms. *U. virens* possesses reduced gene inventories for polysaccharide degradation, nutrient uptake and secondary metabolism, which may result from adaptations to the specific floret infection and biotrophic lifestyles. Consistent with their potential roles in pathogenicity, genes for secreted proteins and secondary metabolism and the pathogen-host interaction database genes are highly enriched in the transcriptome during early infection. We further show that 18 candidate effectors can suppress plant hypersensitive responses. Together, our analyses offer new insights into molecular mechanisms of evolution, biotrophy and pathogenesis of *U. virens*.

<sup>1</sup> Department of Plant Pathology and the Ministry of Agriculture Key Laboratory for Plant Pathology, China Agricultural University, Beijing 100193, China. <sup>2</sup> State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, China. <sup>3</sup> State Key Laboratory Breeding Base for Zhejiang Sustainable Pest and Disease Control, Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China. <sup>4</sup> Institute of Biotechnology, Zhejiang University, Hangzhou 310058, China. <sup>5</sup> International Rice Research Institute, Metro Manila 1301, Philippines. <sup>6</sup> Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. <sup>7</sup> Guangdong Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China. <sup>8</sup> Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, Massachusetts 01003, USA. <sup>9</sup> Institute of Plant Protection, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China. <sup>10</sup> College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha 410128, China. <sup>11</sup> College of Agronomy, Sichuan Agricultural University, Chengdu 611130, China. <sup>12</sup> College of Plant Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, China. <sup>13</sup> Institute of Plant Protection, Liaoning Academy of Agricultural Sciences, Shenyang 110161, China. <sup>14</sup> Department of Plant Breeding and Genetics, China Agricultural University, Beijing 100193, China. <sup>15</sup> Department of Botany and Plant Pathology, Purdue University, West Lafayette IN47907, USA. <sup>16</sup> School of Environmental Sciences, University of Guelph, Ontario, Guelph, Canada N1G 2W1. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to W.S. (email: wxs@cau.edu.cn) or to Y.-L.P. (email: pengyl@cau.edu.cn).

Rice false smut (RFS) caused by the pathogenic ascomycete fungus *Ustilagoideae virens* (Cooke) Takah (teleomorph: *Villosiclava virens*) was first reported from Tirunelveli district of Tamil Nadu State of India in the 1870s<sup>1,2</sup>. RFS was once categorized as a minor disease with sporadic occurrence in rice-growing areas. However, it has recently become one of the most devastating grain diseases in the majority of rice-growing areas of the world due to extensive planting of high-yielding cultivars and hybrids, overuse of chemical fertilizers and an apparent change in global and regional climates<sup>3–5</sup>. In particular, RFS has been estimated to occur in one third of the rice cultivation areas in China<sup>5</sup>.

RFS has an intriguing disease cycle. *U. virens* initiates infection through rice floral organs producing white hyphae, and then develops powdery dark green chlamydo-spore balls in spikelets during the late phase of infection<sup>6</sup> (Fig. 1). As diurnal temperatures fluctuate more in late autumn, sclerotia appear on the surface of spore balls<sup>6</sup>. A key characteristic of RFS is that the pathogen produces two types of mycotoxins, ustiloxins and ustilaginoindins. Ustiloxins are antimetabolic cyclic depsipeptides with a 13-membered core structure<sup>7</sup>. Ustiloxins inhibit cell division, especially by inhibiting microtubule assembly and cell skeleton formation<sup>8</sup>. Ustilaginoidins are bi(naphtho- $\gamma$ -pyrone) derivatives, which exhibit weak antitumor cytotoxicity to human epidermoid carcinoma<sup>9,10</sup>. RFS thus not only results in severe yield losses in rice<sup>4</sup> but also contaminates rice grain and straw to potentially cause human or animal poisoning<sup>11</sup>. Many previous studies have focused on isolation and characterization of ustiloxins and ustilaginoindins, but the pathways and genes responsible for their biosynthesis remain largely unknown<sup>12</sup>.

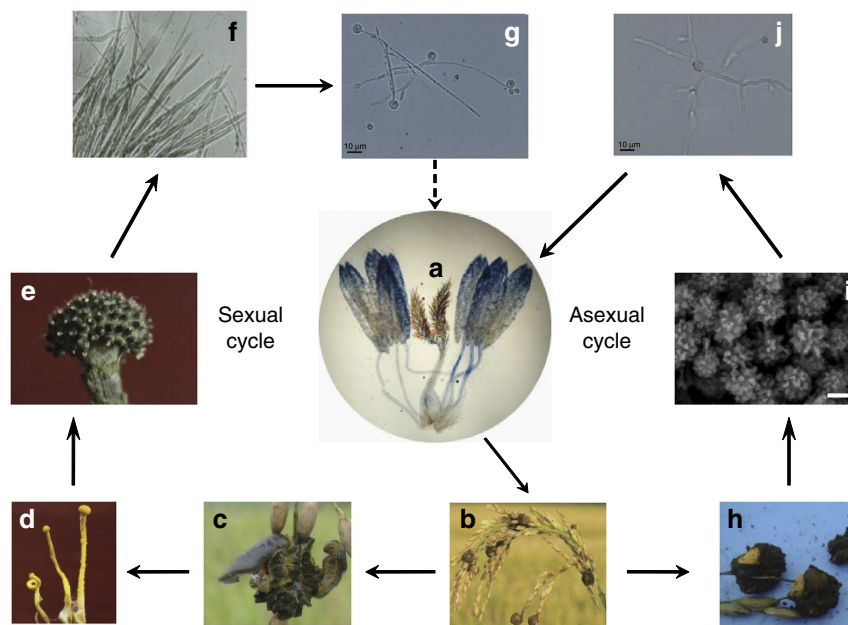
*U. virens* produces both sexual (ascospores) and asexual (chlamydo-spores) stages in its life cycle (Fig. 1), and thick-walled chlamydo-spores can survive up to 4 months in the field with sclerotia presumed to survive much longer<sup>13</sup>. Sclerotia on or under soil surface can germinate and form fruiting bodies, and ascospores that contribute to primary infections are then produced after mating and meiosis (Fig. 1)<sup>14</sup>. Some rice weeds

and grass species may act as alternative hosts for the pathogen and may be important sources of inoculum between seasons<sup>15</sup>. Previous results suggested that the fungus is both seed borne and soil borne, and can invade seedlings even if infection symptoms are not observed until the flowering stage in rice<sup>16</sup>. However, our knowledge of the disease cycle and infection processes of *U. virens* is incomplete and sometimes contradictory<sup>5</sup>.

The infection process of *U. virens* was recently investigated through extensive histological examinations<sup>17</sup>. *U. virens* hyphae enter floral organs primarily at upper parts of the stamen filaments between the ovary and lodicules<sup>17</sup>. *U. virens* occasionally infects the stigma and lodicules, but is not known to infect the ovaries. Primarily, the pathogen hyphae extend into the central vascular tissues without penetrating rice cell walls directly<sup>17</sup>. Therefore, the pathogen is considered to be a biotrophic parasite<sup>17</sup>.

*U. virens* is the only member of the tribe Ustilagoideae of the family Clavicipitaceae with a known sexual state, which is rarely observed in nature<sup>13</sup>. Its teleomorph was originally classified as *Claviceps virens* Sakurai ex Nakata and has been recently re-categorized as *Villosiclava virens* since it is morphologically distinct from *Claviceps*<sup>1</sup>. The Clavicipitaceae family consists of 43 genera such as *Claviceps*, *Cordyceps*, *Epichiloe*, *Metarhizium* and *Ustilagoideae*, which include pathogens across animal, plant and fungal kingdoms. Such a broad host range makes this family unique for exploring evolution and host adaptation<sup>18</sup>. It has been speculated that the tribe Ustilagoideae may have a different origin from other clavicipitaceous fungi<sup>19</sup>.

This project takes genomic approaches to dissect molecular mechanisms of RFS for development of effective disease management strategies. Our analyses provide solid phylogenetic placement for *U. virens* to be closely related to the entomopathogenic *Metarhizium* spp., Genomic content supports a specific adaptation of this pathogen in occupying host florets, which are rich in sugar content but deficient in cellulose, pectin and other cell wall materials. Genome analyses facilitate the isolation and characterization of genes and regulatory mechanisms that control



**Figure 1 | Major stages in the infection cycle of *Ustilagoideae virens*.** (a) Stamen filaments in the florets infected by *U. virens*. Hyphae were stained with trypan blue to show primary infection sites. (b) False smut balls formed in the rice spikelets. (c) Sclerotia formed on the surface of spore balls. (d) Stroma produced by a germinating sclerotium. (e) Ascocarp formed on the stroma. (f) Asci. (g) Ascospore germination, Scale bar, 10  $\mu$ m. (h) Spore balls. (i) Chlamydo-spores under scanning electron microscopy. Scale bar, 3  $\mu$ m. (j) Chlamydo-spore germination. Scale bar, 10  $\mu$ m.

mycotoxin production. The *in planta* upregulation of candidate effectors, genes involved in secondary metabolite biosynthesis, as well as the pathogen–host interaction database (PHI-base)<sup>20</sup> genes suggests a list of candidate virulence factors that may play important roles in pathogenicity. Overall, this genome project establishes foundation for the rice *Ustilaginoidea* pathosystem as a primary target for elucidating the interaction between hosts and biotrophic fungal pathogens.

## Results

**Genome assembly with lower gene density.** The genome of the *U. virens* strain UV-8b was assembled from sequencing data generated by Illumina and 454 sequencing technologies (Supplementary Table 1). Data from 454 sequencing were assembled into contigs using the Newbler assembler. Subsequently, these contigs were used as single ends and combined with trimmed paired-end Illumina data for assembly with ABySS<sup>21</sup>, resulting in longer sequences with a total assembly size of 38.8 Mb and an N50 contig length of about 29.3 kb (Supplementary Table 2). Contigs were further linked with SSPACE scaffolder<sup>22</sup> into 449 scaffolds (>2 kb) with an N50 length of 533.6 kb and a total assembly size of 39.4 Mb (Table 1, Supplementary Table 2) that accounts for 97.5% of the total genome size (~40.4 Mb) estimated by JELLYFISH<sup>23</sup>. High consistency between 454 and Illumina data sets and depth distribution analyses confirm the high quality of the final assembly (Supplementary Fig. 1, Supplementary Table 3, Supplementary Note 1). Furthermore, all RNA-seq reads were assembled into 17,168 contigs longer than 300 bp, and 99.2% of that are aligned with the genomic sequence, implying high coverage of coding regions in the assembly (Supplementary Table 4).

A total of 8,426 protein-coding genes were predicted from the genome assembly, 8,243 of which were supported by the RNA-seq data. The coding regions from the predicted genes constitute 32.7% of the genome with an average gene density of 214 genes per 1 Mb, which is fewer than most of other sequenced ascomycetes<sup>24,25</sup>. The average 1,627 bp gene length includes an average 1,414 bp of coding region and 213 bp of non-coding region (Table 1). The GC contents of the genome, coding sequences and repetitive elements are 49.96, 58.42 and 39.77%, respectively. A total of 204 tRNA genes were predicted from the assembly (Table 1).

**Repetitive elements affected by RIP and RNA silencing.** A total of 160 repetitive families including 10 families of DNA transposons, 45 long terminal repeat (LTR) and 6 non-LTR retrotransposons were identified in *U. virens* (Supplementary

Table 5, Supplementary Note 2). These repeat families account for ~25% of the genome (Fig. 2a). The majority of repetitive sequences (~72%) are transposable elements (TEs), and the remaining are simple sequence repeats. LTR retrotransposons represent about 83% of TEs. DNA transposons include Helitrons that account for ~50% of DNA transposons, one terminal inverted repeat family, and some undefined families (Fig. 2a).

RIP (repeat-induced point mutation) is a genome defence mechanism in fungi that might prevent gene creation through genomic duplication<sup>26</sup>. A high ratio of TpA/ApT (1.48) and a low value for (CpA + TpG)/(ApC + GpT) (<1.0) were observed, implying strong RIP in *U. virens* (Supplementary Note 2). The high average ratio of di-nucleotide preference also provides genome-level evidence for the efficiency of RIP. *U. virens* has a di-nucleotide ratio index similar to *N. crassa* and higher than other sequenced ascomycetes (Fig. 2b), which indicates that RIP, similar to that found with *N. crassa* genome, is a significant characteristic of *U. virens* genome. Furthermore, alignment of homologous TEs revealed numerous point mutations representing C-to-T (G-to-A) transitions in the TE sequences, implying that RIP can also control TE activity in *U. virens* (Supplementary Fig. 2). The active RIP defence mechanism in *U. virens* is also supported by the presence of a DNA methyltransferase RID (for RIP defective) homologue, a key protein in the RIP machinery in *N. crassa*.

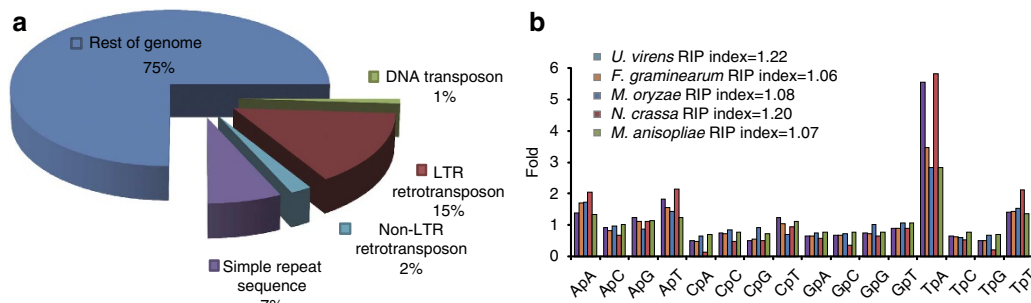
RNA silencing is well recognized as a principal regulatory mechanism in eukaryotes to control gene expression<sup>27</sup>. The key genes necessary for RNA silencing pathways are highly conserved in *U. virens* including three Argonaute-like proteins required for forming RNA-induced silencing complex, three RNA-dependent RNA polymerases, three ReqQ helicases and two Dicer genes (Supplementary Table 6). These genes are speculated to function in two fungus-specific RNA-silencing pathways: meiotic silencing of unpaired DNA and quelling (Supplementary Table 7).

**Comparative genomics and evolution.** Four additional strains collected from different geological regions in China (Supplementary Table 8) were subjected to low-depth sequencing for comparative genomics. Microsynteny analyses at the mating-type loci among the five strains indicated that each *U. virens* genotype has only a single mating-type locus and that *U. virens* is heterothallic (Supplementary Fig. 3, Supplementary Note 3). Comparison with the reference genome of UV-8b revealed low rate of single nucleotide polymorphisms among all these strains. The relatively low nucleotide diversity between genotypes indicates a low level of intraspecific sequence variation, suggesting a recent expansion of the *U. virens* population (Supplementary Table 9, Supplementary Note 4). The ratio of nonsynonymous to synonymous substitutions (Ka/Ks) was calculated to assess selection pressures on *U. virens* evolution from the host interaction. In the UV-8b strain versus the four other strains, 2,957 and 1,451 genes were found to have nonsynonymous and synonymous substitutions, respectively; 286 genes have both nonsynonymous and synonymous substitutions, of which 277 genes have a Ka/Ks < 1 and only nine genes have a Ka/Ks > 1 ( $P < 0.05$ ) (Supplementary Data 1). Among these are five effector-like genes and three with putative functions in DNA repair and cell cycle regulation (Supplementary Note 4).

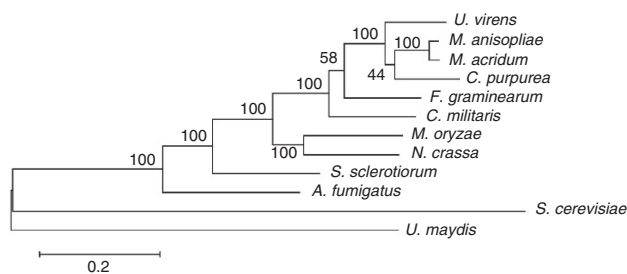
The phylogenetic position of *U. virens* was evaluated among other eleven selected fungal species (10 ascomycota and one basidiomycota outgroup) using a set of highly conserved single-copy genes suggested for the Fungal Tree of Life<sup>28</sup> (Supplementary Table 10, Supplementary Note 5). The analysis revealed that *U. virens* is more closely related to two entomopathogenic *Metarhizium* spp. (Hypocreales, Clavicipitaceae) than

**Table 1 | Genome features of *U. virens* UV-8b.**

Features	<i>U. virens</i>
Size (Mb)	39.4
Coverage	142 ×
% G + C content	49.96
% Repeat	25
Protein-coding genes	8,426
Average gene length (bp)	1,627
Gene density (# gene per Mb)	214
Exons per gene	2.6
Average exon length (bp)	544
Introns per gene	1.6
Average intron length (bp)	133
tRNA	204



**Figure 2 | Repetitive sequences have been affected by repeat-induced point mutations in *U. virens*.** (a) The proportion (%) of different types of repetitive sequences in the *U. virens* UV-8b genome. (b) Estimation of RIP for five sequenced ascomycete genomes. Average RIP indices in *U. virens*, *F. graminearum*, *M. oryzae*, *N. crassa* and *M. anisopliae* and the ratio of each type of di-nucleotide (ratio of frequency in repetitive elements to that in non-TEs) in different fungal genomes are shown.



**Figure 3 | Phylogenetic analyses for evolutionary relationships among different fungi.** The fungi are *U. virens*, *C. purpurea*, *M. anisopliae*, *M. acridum*, *C. militaris*, *M. oryzae*, *F. graminearum*, *N. crassa*, *A. fumigatus*, *S. cerevisiae*, *S. sclerotiorum* and *U. maydis*. A maximum likelihood phylogenetic tree constructed based on derived protein sequences of 21 single-copy genes common to these 12 fungal genomes that were identified from a set of 74 single-copy orthologous genes across the Kingdom Fungi showing evolutionary relationships among these fungi.

to other species including the plant pathogen *Claviceps purpurea* (Hypocreales, Clavicipitaceae) and the insect pathogenic fungus *Cordyceps militaris* (Hypocreales, Clavicipitaceae) (Fig. 3). Furthermore, *U. virens* shares more homologous proteins with >60% identity and more identical amino-acid positions with *Metarhizium* spp. than with the other tested fungi. Proteome comparisons also revealed a 10.7–11.1% greater amino-acid similarity of *U. virens* to *Metarhizium* spp. than to *C. purpurea* (Supplementary Table 11). Comparative genomic analyses revealed that *U. virens* shares more reciprocal best hit orthologous genes with *M. acridum* (5,702) and *M. anisopliae* (5,644) than with the other species, although by this metric *Fusarium graminearum* shared a surprisingly high 5,639 orthologous genes with *U. virens* (Supplementary Table 11). Synteny analysis also showed that *U. virens* and *M. anisopliae* genome structures have many large areas of synteny (Supplementary Fig. 4).

Comparisons among *U. virens*, *Magnaporthe oryzae*, *F. graminearum* and *M. anisopliae* genomes were performed to identify genes that are absent from the *U. virens* genome since it has a smaller gene inventory. A total of 703 orthogroups were identified, 486 of which have gene ontology (GO) annotations. Interestingly, the majority of missing orthogroups belong to the GO terms in molecular functions (4) and in biological processes (11; see Supplementary Table 12). A significant reduction was observed in proteins for oxidoreductase activity, proteins related to toxin and secondary metabolism, energy metabolism and catabolism and responses to toxin and chemical stimuli (Supplementary Table 12).

**Weakened polysaccharide degradation machinery.** For successful infection, phytopathogenic fungi often have to break down plant cell walls by carbohydrate-active enzymes including glycoside hydrolases, glycosyl transferases, polysaccharide lyases (PL) and cutinases. Interspecific comparisons showed that *U. virens* encodes significantly fewer carbohydrate-active enzymes than plant hemi-biotrophic fungi, such as *M. oryzae* and *F. graminearum*. Particularly, the proteins in such families as GH6, GH7, GH39 and GH51 that are involved in degrading cellulose, hemi-cellulose and pectin of plant cell walls are significantly reduced in comparison with *M. oryzae* and *F. graminearum* ( $P < 0.01$ ; Supplementary Tables 13, 14, Supplementary Note 6). This characteristic is similar to the obligate biotrophic powdery mildew and entomopathogenic fungi that have no canonical plant cell wall-degrading enzymes<sup>29</sup>. Corresponding to these findings, putative carbohydrate-binding proteins are significantly reduced in *U. virens* (Supplementary Table 15). Distinctively, the fungus seems to lack PLs including pectin lyases and pectate lyases (Table 2). Consistently, other enzymes involved in pectin degradation are less frequent or absent, such as the d-4,5 unsaturated  $\alpha$ -glucuronidase GH105 family that can sequentially degrade the products generated by PLs (Supplementary Table 13)<sup>30</sup>. Furthermore, *U. virens* contains fewer putative cutinases needed for plant cuticle degradation (Table 2). Collectively, these analyses suggest that polysaccharide degradation machinery is compromised in *U. virens* and degrading plant cell wall is probably not essential in establishing infection.

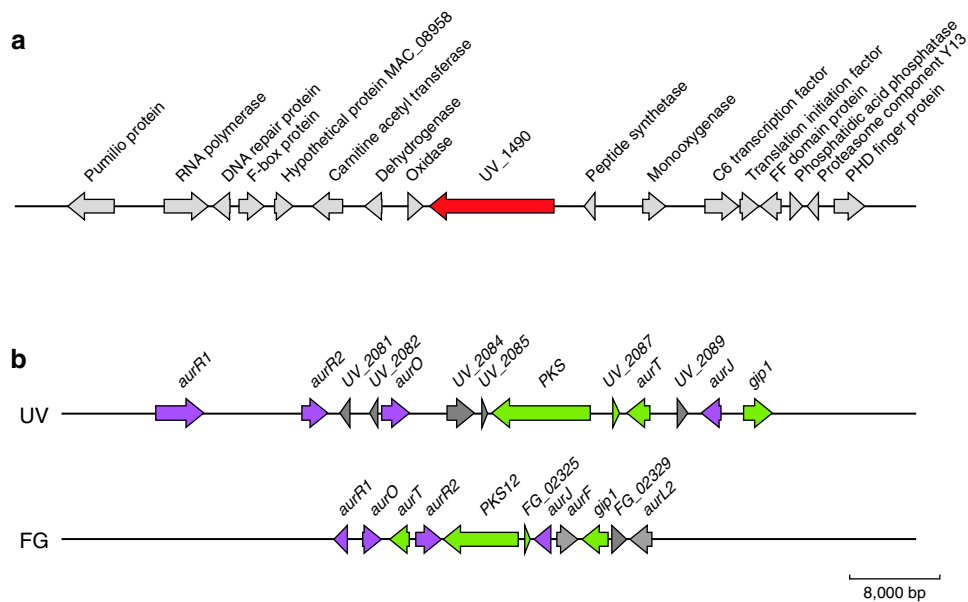
To verify the results from bioinformatic analyses, the pectinolytic activities of *U. virens* were evaluated in comparison with *M. oryzae*, *F. graminearum* and *Colletotrichum fragariae* (Supplementary Note 6). As predicted by genomic data that *U. virens* has very limited ability to secrete pectin-degrading enzymes, no translucent area indicating of pectin degradation was observed for *U. virens*, while a clear transparent halo was formed in the *C. fragariae*-cultured plates (Supplementary Fig. 5a). In addition, *U. virens* grew well in the medium using glucose or sucrose as sole carbon, while it did not grow at all on pectin and grew poorly on xylan and cellulose. In contrast, *C. fragariae* and *Rhizoctonia solani* grew rapidly on pectin plates (Supplementary Fig. 5b). Pectinolytic activity assays and carbon growth profiles of *U. virens* supported the bioinformatic prediction.

**Reduced G-protein-coupled receptors and transporters.** In phytopathogenic fungi, G-protein-coupled receptors (GPCRs) and G-proteins are important components of signalling pathways for response to environmental stimuli<sup>31</sup>. *U. virens*, similar to the non-pathogenic saprophyte *N. crassa*, encodes fewer GPCRs than

**Table 2 | Selected protein families involved in fungal pathogenesis in *U. virens* and other ascomycetes.**

Protein family	UV	FG	MO	BC	SS	CCM	MAA	MAC	NC	AN
Fungal-specific transcription factors	116	253	95	114	95	157	165	161	88	258
Zinc-finger transcription factors	25	85	58	48	54	63	86	53	63	58
Major facilitator superfamily	80	246	180	190	144	207	215	177	106	261
Cytochrome P450	38	112	137	129	93	57	123	95	40	116
Pth11-like G-protein-coupled receptor	17	51	50	23	25	21	29	22	20	44
Protein kinases	149	149	135	127	169	167	161	192	119	139
Trypsin	4	3	3	1	1	12	32	17	2	2
Subtilisin	11	24	29	7	6	35	55	43	10	4
Aspartic protease	14	18	19	14	21	21	33	25	19	16
Lipase	14	31	23	28	25	38	47	41	16	27
Pectinesterase	2	3	1	5	5	0	0	0	1	3
Glycoside hydrolase	117	204	228	158	162	126	151	123	168	231
Glycosyltransferase	66	91	102	93	79	82	109	94	76	90
Dehydrogenase	176	439	315	342	289	288	351	309	235	458
Cutinase	4	12	18	11	8	4	2	2	3	4
Pathogen-host interaction proteins	1103 (13.1%)	1979 (14.9%)	1645 (14.9%)	1741 (10.6%)	1531 (10.6%)	1483 (15.4%)	1769 (16.7%)	1586 (16.1%)	1257 (12.8%)	1965 (18.6%)
Pectin lyase	0	25	9	25	20	10	7	7	5	24

AN, *Aspergillus nidulans*; BC, *Botrytis cinerea*; CCM, *Cordyceps militaris*; FG, *Fusarium graminearum*; MO, *Magnaporthe oryzae*; MAA, *Metarhizium anisopliae*; MAC, *Metarhizium acridum*; NC, *Neurospora crassa*; SS, *Sclerotinia sclerotiorum*; and UV, *Ustilagoidea virens*.



**Figure 4 | Putative NRPS and PKS gene clusters for ustiloxin and ustilaginoidin biosynthetic pathways. (a)** The putative NRPS gene (central long red arrow) cluster for ustiloxin biosynthesis. UV\_1490 containing three adenylation domains may be responsible for the biosynthesis of the core structure of ustiloxins. The NRPS gene is clustered around genes encoding dehydrogenase, oxidase, monoxygenase and transcription factor that may be involved in the modification of secondary metabolites. **(b)** Comparison between the putative PKS gene cluster for ustilaginoidin biosynthesis in *U. virens* (upper) and the aurofusarin biosynthesis gene cluster in *F. graminearum* (bottom). The UV\_2086 (PKS), UV\_2087, aurT and gip1 genes (green) in *U. virens* were identified by BLASTing against the aurofusarin biosynthesis genes while other genes (purple) were predicted via Pfam domain search. The aurF and aurL2 genes (grey) in the aurofusarin biosynthesis gene cluster in *F. graminearum* are absent in *U. virens* genome.

the hemi-biotrophic pathogens *M. oryzae* and *F. graminearum* (Supplementary Table 16, Supplementary Note 7). In particular, *U. virens* contains many fewer Pth11 homologues that are required for appressorium differentiation in response to host surface stimuli and pathogenesis<sup>32</sup>. No homologues of *Saccharomyces cerevisiae* GPR1, the sucrose/glucose sensing receptor, were predicted in *U. virens*<sup>33</sup>. Significant loss of GPCRs suggests that the biotrophic *U. virens*, adapted to a narrow niche for host infection, might use specialized receptors for extracellular signals compared with the hemi-biotrophic

*M. oryzae* and *F. graminearum*. In addition, *U. virens* has more  $\beta$ -subunits of G-proteins (Supplementary Table 17).

*U. virens* also encodes significantly fewer transporters, including members in the amino-acid-polyamine-organocation, ATP-binding cassette and major facilitator superfamilies (MFS). In contrast, the proteins in the oxidoreduction-driven transporter subclass are expanded (Supplementary Table 18, Supplementary Note 7). Many MFS transporters are involved in the transport of a wide range of nutrient substrates, such as sucrose, galactoside, amino acid and peptides<sup>34</sup>. Reduced MFS transporters and lack of

the nutrient sensor GPR1 suggest that *U. virens* might have a narrow specificity to utilize nutrients from host plants.

**Secondary metabolism genes.** *U. virens* genome encodes 17 non-ribosomal peptide synthetases (NRPS), 14 polyketide synthases (PKS), 6 geranylgeranyl diphosphate synthases, 6 terpene synthases, and 6 terpenoid cyclases. These enzymes are responsible for fundamental steps in the biosynthesis of secondary metabolites, such as mycotoxins, alkaloids and pigments (Supplementary Table 19, Supplementary Note 8). However, *U. virens* lacks dimethylallyl tryptophan synthases, which catalyse the first committed step of ergot alkaloid biosynthesis<sup>35</sup>.

In *M. robertsii*, the NRPS protein DtxS1 with six adenylation domains is responsible for the biosynthesis of the cyclic hexadepsipeptide destruxins, which are structurally similar to ustiloxins<sup>36</sup>. Of the predicted NRPS/NRPS-like proteins, UV\_1490 and other five have adenylation domain(s) that recognize the cognate amino-acid substrate and activate it as its aminoacyl adenylate for peptide synthesis (Supplementary Table 20). These NRPS/NRPS-like genes are within clusters with genes encoding acyltransferases, dehydrogenases (DHGs), oxidoreductases and/or cytochrome P450s (CYPs) involved in the modification of secondary metabolites (Fig. 4a and Supplementary Fig. 6), suggesting that one or more of them may be responsible for the synthesis of ustiloxins. The core structure of ustilaginoidins is formed by dimerization of nor-rubrofusarin that is an intermediate during aurofusarin biosynthesis in *F. graminearum*<sup>12,37</sup>. Through comparisons with the identified aurofusarin biosynthesis gene cluster<sup>37</sup>, the PKS gene cluster that may be responsible for the biosynthesis of ustilaginoidins was also found in *U. virens* (Fig. 4b, Supplementary Table 21 and Supplementary Note 8).

Molybdoprotein biosynthesis and nutrient uptake pathways, such as thiamine biosynthesis play vital roles in lifestyle of pathogenic fungi<sup>38</sup>. The genes involved in these biosynthesis pathways were compared among *U. virens* and some obligate biotrophs, hemibiotrophs and necrotrophs to better understand evolution to biotrophy (Supplementary Note 8). Components of these pathways were partially found in the genomes of the biotrophs *U. virens*, *U. maydis*, *Blumeria graminis* and *Puccinia striiformis* (Supplementary Table 22), suggesting that loss of thiamine and molybdoprotein biosynthetic pathways is likely the result, but not the reason for biotrophy in fungi<sup>38</sup>.

**Important gene families involved in pathogenicity.** To investigate pathogenicity genes in *U. virens*, virulence-associated mitogen-activated protein kinase (MAPK) cascades were first surveyed. Multiple analyses indicated that *U. virens* might contain five MAPK pathways similar to *S. cerevisiae* and that MAPK signalling pathways were highly conserved in phytopathogenic fungi (Supplementary Figs 7, 8, Supplementary Table 23, Supplementary Note 9). Furthermore, genome-wide BLAST analyses against PHI-base<sup>20</sup>, which includes experimentally verified genes involved in pathogen–host interactions revealed that 1,103 protein-coding genes in *U. virens* were putatively involved in virulence and pathogenicity, which account for 13.1% of the total predicted genes (Table 2, Supplementary Table 24, Supplementary Note 9). Proteins identified in PHI-base as involved in secondary metabolism, degradation enzymes for large molecules that might be involved in breaking host physical barriers, MFS and ATP-binding cassette transporters, DHGs and CYPs, and some transcription factors in *U. virens* were less abundant than those in *M. oryzae* and *F. graminearum* (Supplementary Tables 24,25). In contrast, many protein families in PHI-base that are considered essential for basic cellular

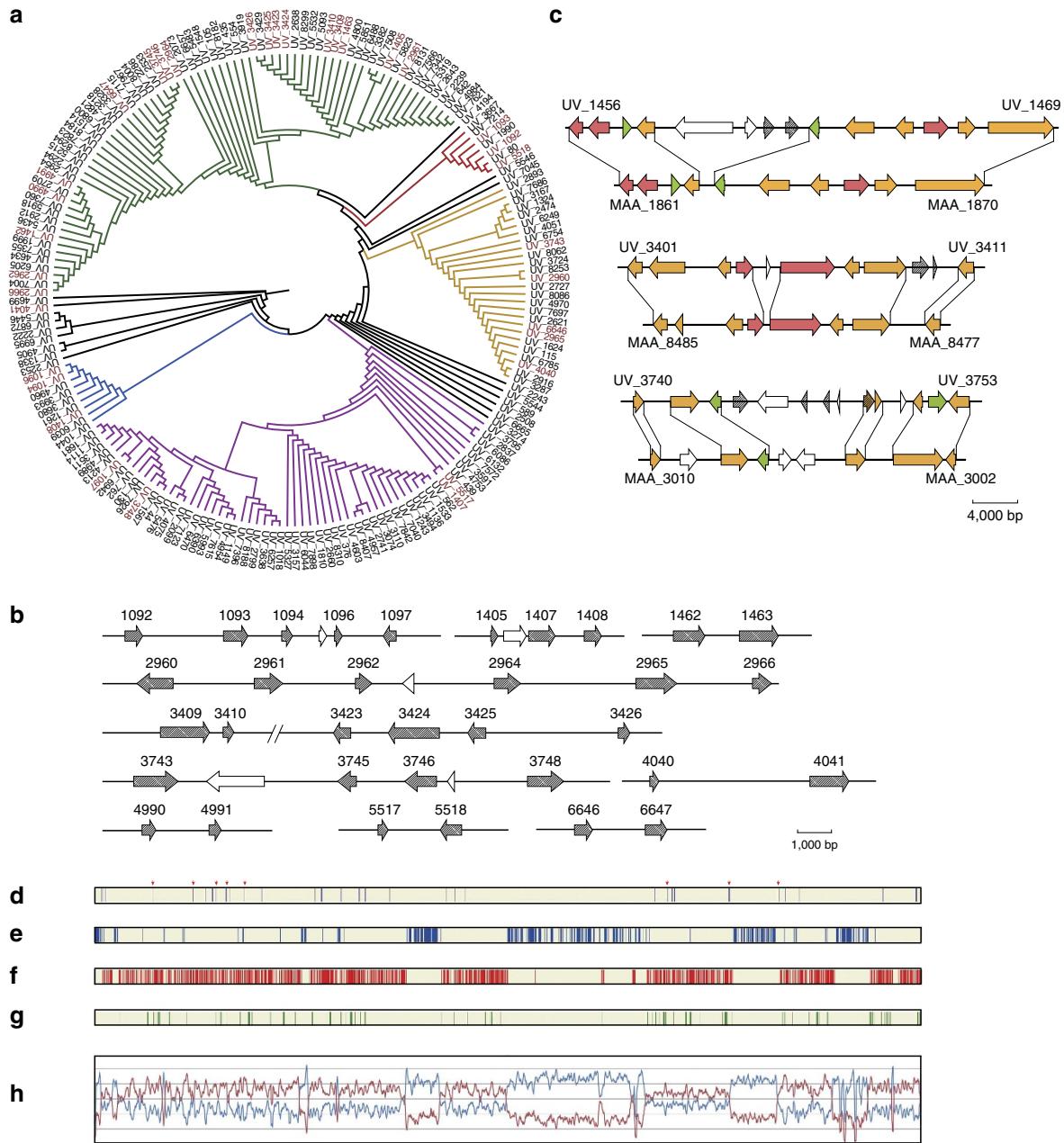
processes, such as DNA/RNA helicases, DNA topoisomerase and DNA-directed DNA polymerases, are conserved in *U. virens* (Supplementary Table 24). GLEYA adhesin, CheY-like proteins and proteins secreted in xylem are even expanded in PHI-base of *U. virens* as compared with *F. graminearum* or *M. oryzae* (Table 2, Supplementary Table 24, Supplementary Note 9).

***U. virens* secretome and effector prediction.** The proteins secreted by pathogenic fungi, particularly effectors, are among essential components for successful infection, especially for the biotrophic fungi that have intimate host–fungal interactions<sup>39</sup>. *U. virens* genome encodes 628 (527 without transmembrane domains) potentially secreted proteins, including 193 small (<400 amino acids), cysteine-rich ( $\geq 4$  cysteine residues) proteins that are putative effectors (Fig. 5a, Supplementary Note 10).

Among them, 34 predicted effector genes are organized into clusters/families containing two to six related genes, suggesting that local duplications might be involved in expansion of effectors in *U. virens* (Fig. 5a,b). In the maize pathogen *U. maydis*, some effector gene clusters play decisive roles through the course of infection<sup>40</sup>. Among these recently expanded *U. virens* effectors, we observed surprisingly low Ka/Ks ratios and low SNP frequency among five sequenced strains, suggesting a low intraspecific genetic diversity among the pathogen population, even for the fast evolving effectors (Supplementary Table 9 and Supplementary Data 1). However, when comparing with the close evolutionarily related fungal species *Metarhizium* spp. (Fig. 3), we observed accelerated interspecific diversification. For instance, pairwise comparisons showed that over 63% of non-secreted proteins in *U. virens* shared >60% amino-acid identity with those in either *M. anisopliae* or *M. acridum*, respectively. In contrast, predicted secreted proteins share significantly lower amino-acid identity between these species. More evidently, only 43 and 40 putative effector proteins (22.3% and 20.7%) in *U. virens* share high amino-acid identity (>60%) with the protein inventory in *M. anisopliae* and *M. acridum*, respectively (Supplementary Table 26). Microsynteny analyses between *U. virens* and *M. anisopliae* were performed for 11 putative effector gene clusters. Seven clusters containing 24 putative effectors were found in highly diverse genomic islands flanked by conserved regions (Fig. 5c and Supplementary Fig. 9). These results imply that secreted proteins, especially effector proteins, have evolved more rapidly than other proteins in *U. virens* and that these proteins may play important roles in host adaptation.

Repetitive elements often play important roles in genome evolution, including birth/death and diversification of functional genes such as effector genes, secondary metabolite biosynthesis genes<sup>35,41</sup>. However, no direct correlation between candidate effectors and TE-rich regions (low GC content) was evident in the *U. virens* genome (Fig. 5). This result suggests that TE-driven evolution might have less of an effect on the synergistic interactions of *U. virens* and rice than on some other host–pathogen interactions, such as *M. oryzae* and rice<sup>35,41</sup>.

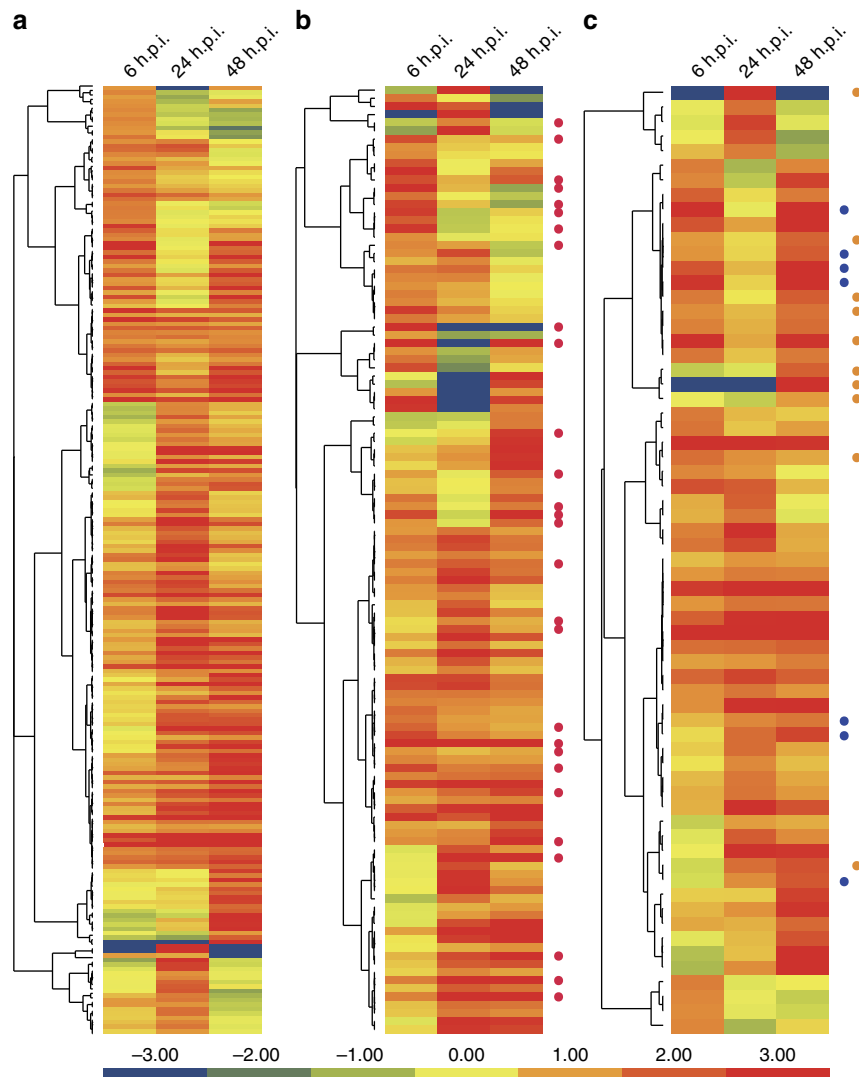
**Transcriptome profiles during early infection.** To understand pathogenicity and *in planta* transcriptional reprogramming, gene-expression profiles of *U. virens* were studied during early infection. As compared with axenic cultures, 830, 497 and 484 genes were upregulated while 452, 224 and 266 genes were downregulated at 6, 24 and 48 h post inoculation (h.p.i.), respectively (Supplementary Fig. 10). A total of 1,319 upregulated genes and 687 downregulated genes were identified during early stages of *U. virens* infection (Supplementary Fig. 11 and



**Figure 5 | Genes and gene clusters encoding putative effectors and their locations in relation to repetitive sequences in the *U. virens* genome.** (a) A phylogenetic tree of 193 putative effectors indicates that most effectors are categorized into several super clades. (b) Thirty-three putative effector genes are aggregated into 11 clusters with two to six effector genes. Shaded arrows represent putative effector genes while white arrows are for other genes. (c) Microsynteny analyses between *U. virens* and *M. anisopliae* genomes revealed that most of effector gene clusters lie in the highly diverse genomic regions flanked by conserved sequences. Shaded arrows are putative effector; brick red represents homologous proteins with >80% amino-acid identity; yellow, 60~80% amino-acid identity; green, <60% identity; and white, no homologous proteins revealed by genome comparisons using SyMAP v4.0. Three effector gene clusters are shown here, others are shown in Supplementary Fig. 9. (d-h) Physical locations of predicted secreted protein genes, all protein-coding genes and PHI-base genes in relation to regions of repetitive sequences and GC content distribution in the assembled genome of *U. virens* (using scaffold #1 as an example). (d) Locations of genes encoding predicted secreted proteins (purple) including putative effectors (indicated by red arrows) in the assembled genome. (e) The distribution of transposable elements (blue) in the *U. virens* genome. (f) Single-copy DNA regions (red) of the *U. virens* genome. (g) Locations of the PHI-base gene homologues (green) involved in pathogen-host interactions. (h) Graphs of GC (red) and AT (blue) contents. Areas of low GC correspond well to the regions of repetitive DNA. The maps from d-g were drawn with OmniMapFree.

Supplementary Data 2). GO enrichment analyses revealed that the majority of the very early specifically induced genes (at 6 h.p.i.) were enriched in the cellular component terms while the upregulated genes at slightly later stages (at 24 and 48 h.p.i.) were mainly enriched in the biological process and biological function terms (Supplementary Table 27). Four categories of important

virulence-associated genes were identified to be highly induced during infection. First, 213 PHI-base genes are upregulated in infected rice panicles (Fig. 6a). Particularly, 21 PHI-base genes are among the top 100 most highly induced genes (Supplementary Table 28), indicating an important role of these PHI-base genes in the interaction with host rice. Second, 15 secreted protein genes



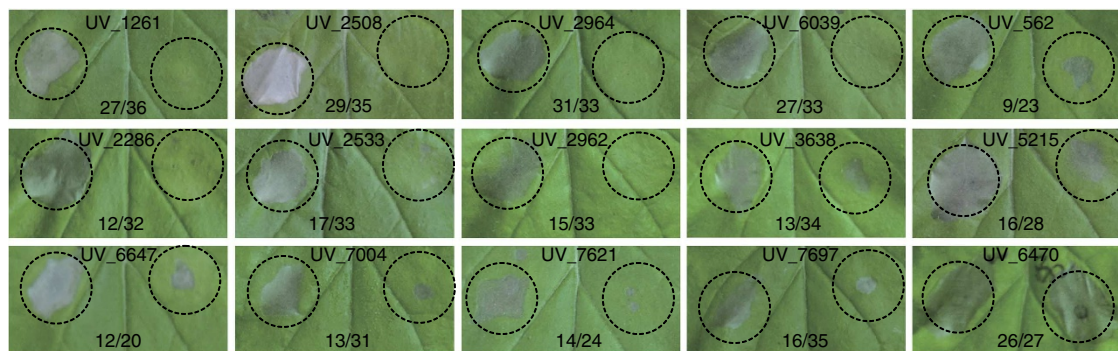
**Figure 6 | The expression pattern of upregulated PHI-base genes and genes for secreted proteins and secondary metabolism during *U. virens* infection.** (a) A total of 213 genes involved in pathogen–host interactions were upregulated at 6 h, 24 h and 48 h post inoculation (h.p.i.). (b) A total of 116 genes encoding predicted secreted proteins including 28 putative effectors were upregulated at 6 h.p.i., 24 h.p.i. and 48 h.p.i. Red dots indicate putative effector proteins. (c) A total of 65 genes involved in secondary metabolic processes were transcriptionally induced at 6 h.p.i., 24 h.p.i. and 48 h.p.i. Blue dots indicate predicted PKS/PKS-like proteins; yellow dots indicate predicted NRPS/NRPS-like proteins. These genes are significantly enriched in the upregulated genes during early stages of *U. virens* infection. Hierarchical clustering analysis was performed for these transcripts at 6 h.p.i., 24 h.p.i. and 48 h.p.i. Each column represents the fold change in transcript levels *in planta* at the indicated times, relative to the levels *in vitro*. The vertical dimension represents genes that exhibited changes in transcript level (cutoff:  $|\log_2[\text{fold change}]| \geq 1$  and false discovery rate  $\leq 0.001$ ). The colour scale indicates transcript abundance relative to the mycelium before inoculation: red, increase in relative transcript abundance; blue, decrease in relative transcript abundance.

including 5 putative effector genes are significantly enriched in the top 100 most highly induced genes ( $P < 0.01$ ; Supplementary Table 28 and Fig. 6b). Among the induced secreted proteins, some putative effectors with conserved functional domains such as fungal-specific extracellular EGF-like (CFEM), WSC, necrosis-inducing protein, ribonuclease, LysM, glycoside hydrolase or peptidase domains might play significant roles in fungal pathogenesis. Third, the secondary metabolism genes including 10 NRPS/NRPS-like, 7 PKS/PKS-like genes and many methyltransferase, acetyltransferase, carboxylesterase and CYP genes are significantly enriched in the transcriptome ( $P < 0.01$ ; Fig. 6c, Supplementary Table 27), implying essential roles of mycotoxins during early infection of *U. virens*. Fourth, a set of plant specifically expressed genes were identified, among which putative effector genes are enriched (Supplementary Table 29).

Interestingly, most of the plant-specific expressed genes have no known function(s) with characteristics of effectors.

**Functional studies on putative effectors.** Plant pathogenic fungi often overcome host immunity through the action of effector proteins<sup>42</sup>. To experimentally identify novel *U. virens* effectors, a pEDV-based screen was performed for 30 randomly selected putative effectors in which a type III secretion system of the emerging rice pathogen *Burkholderia glumae* was used to translocate these putative secreted proteins into plant cells to identify effectors that suppress hypersensitive responses (HR) in *Nicotiana benthamiana*<sup>43</sup>. The screen showed that 4 out of the 30 putative effectors such as UV\_2964 had very strong HR suppressive ability, 4 including UV\_5215 had a strong





**Figure 7 | Putative *U. virens* effectors suppressed *B. glumae*-induced HR in *N. benthamiana*.** Representative cell death symptoms were photographed at 2–3 days after *B. glumae* inoculation. The left half leaf sections were injected with *B. glumae* with the pEDV empty vector and the right half sections were injected with *B. glumae* with effector gene constructs. Numbers, for instance 27/36, indicate that the 27 right leaf sections out of 36 inoculated leaves showed no or significantly less severe symptoms as compared with left sections.

suppressive effect, 10 effectors had a weak effect, and others, for example, UV\_6470 had little or no effect on inhibiting HR in *N. benthamiana* (Fig. 7, Supplementary Table 30). Among them, UV\_6647 and UV\_2799 are upregulated during early infection, UV\_5215 is a plant-specific expression gene, and UV\_2962 and UV\_2964 are within the same gene cluster (Fig. 5b). Interestingly, UV\_2286 encodes a peptide highly similar to Hirsutellin A, an insecticidal ribotoxin produced by *Hirsutella thompsonii*<sup>44</sup>. UV\_6647 encodes a homologue to phosphohistidine phosphatase SixA from *Escherichia coli* that modulates the ArcB phosphorelay signal transduction<sup>45</sup>.

## Discussion

*U. virens* has been recognized as the fastest spreading and economically important pathogen on rice<sup>5,17</sup>. In this study, we sequenced, assembled and analysed the draft genome of this important pathogen. *U. virens* has a genome size of 40.4 Mb, which is in the midrange of 15 plant-associated Clavicipitaceae fungal genomes from 28.9–58.7 Mb, with large variations in repeat sequences<sup>35</sup>. The genome contains ~25% repeat sequences, most of which are TEs. Multiple analyses indicate that *U. virens*, similar to *N. crassa*, uses RIP, a fungus-specific genome defence mechanism<sup>46</sup> to control TE activity (Fig. 2, Supplementary Fig. 2). *U. virens* also possesses orthologues of all the *N. crassa* genes postulated to be involved in RNA silencing (Supplementary Table 6). This suggested that both RNA silencing and RIP are involved in defence against adverse effects of proliferating repetitive sequences in *U. virens*.

As the only sequenced fungal species in the tribe Ustilaginoideae, the *U. virens* genome provides useful information on host adaptation and evolutionary relationships in the Hypocreales. Comparison of *U. virens* with several Hypocreales and some other important plant pathogens showed that *U. virens* is evolutionarily close to the entomopathogenic *Metarhizium* spp., as assessed by multiple criteria, suggesting a host shift across the plant and animal kingdoms. Recently, *Metarhizium* spp. was found to be phylogenetically closer to plant pathogens than to the animal pathogens *A. fumigatus* and *C. albicans*<sup>34</sup>. Spatafora *et al.*<sup>47</sup> showed phylogenetic relationships among members of tribe Ustilaginoideae and hypothesized an animal pathogen origin for the ergot and grass endophytes. According to the host habitat hypothesis<sup>18</sup>, *U. virens* might have evolved from an entomopathogenic fungus because of their co-occurrence in the same environment.

Our comparative genomics analyses found that *U. virens* contains many fewer genes than hemi-biotrophic fungal

pathogens including *F. graminearum* and *M. oryzae* or the necrotrophic *S. sclerotiorum*<sup>48</sup>. In particular, *U. virens* is dramatically reduced in at least three categories of functional proteins. First, *U. virens* has a smaller inventory of carbohydrate-active enzymes with almost no ability to degrade pectin and a very low capability to break down cellulose and xylan (Table 2, Supplementary Fig. 5). These results indicate that *U. virens* with its biotrophic lifestyle may minimize the release of cell wall fragments by carbohydrate-active enzymes, whose products are often recognized as endogenous signals to induce plant immunity<sup>38,40</sup>. The smaller polysaccharide degradation machinery and other genomic features also reflect the infection pattern of *U. virens* that infects plants through stamen filaments where cell wall materials such as cellulose and pectin are deficient<sup>17</sup>. Second, *U. virens* contains many fewer MFS transporters that have been reported to be mainly associated with nutrient uptake and antifungal drug resistance<sup>49</sup>. Furthermore, the genome was not found to encode GPR1 nutrient sensors and has many fewer proteins involved in energy metabolism (Supplementary Tables 12,16). During floral infection, the fungus does not form haustoria<sup>17</sup>, the structure involved in nutrient uptake for many biotrophs. Taken together, these characters indicated that *U. virens* may have a limited ability to acquire nutrients from plant tissues and organs. Abundant and easily obtained nutrients in the male flower parts including pollens may contribute to primary colonization of *U. virens* in the rice florets. Third, DHGs, CYPs and other enzymes involved in secondary metabolism are significantly fewer (Table 2, Supplementary Table 24). CYPs on both flanks of NRPSs and PKSs are also much less abundant than those in *M. anisopliae*. DHGs and CYPs in phytopathogenic fungi contribute to detoxification of phytoalexin repertoires of host plants<sup>50</sup>. The smaller inventory of CYPs and DHGs indicates that *U. virens* may have a lower capacity for adaptation to different plants, which may explain the relatively narrow host range of *U. virens*.

Despite a reduced secondary metabolism gene inventory, many genes involved in secondary metabolism including more than half of NRPS and PKS genes were greatly induced *in planta* during infection (Fig. 6c and Supplementary Data 2). This is consistent with the observation that ustiloxin is easily extracted from spore balls, but not from axenic cultures. Therefore, we hypothesize that secondary metabolites, particularly mycotoxins might act as virulence factors during infection. The predicted ustiloxin and ustilaginoidin biosynthesis gene clusters provide the foundation to test such hypotheses experimentally.

*In planta* expression profiles also indicated that a set of secreted protein and PHI-base genes are important for

pathogenicity of *U. virens*. In particular, many effector candidates were highly upregulated during infection (Fig. 6b). They include secreted in xylem proteins, peptidases, ribonucleases, glycosyl hydrolases and other proteins with the typical effector domains, such as NPP1, CFEM and WSC, which have been found to play significant roles in fungal pathogenesis (Supplementary Data 2). We experimentally demonstrated that more than half of randomly selected putative effectors could suppress *B. glumae*-triggered HR in *N. benthamiana* to different degrees (Fig. 7). The results indicated that these effectors may play essential roles in shaping the interaction of *U. virens* with its host. Predicted secreted proteins, particularly putative effector proteins, are highly diverse between *U. virens* and *M. anisopliae* despite their close evolutionary relationship (Supplementary Table 26). Consistently, the effector gene clusters in *U. virens* are predominantly located in highly diverse genome regions when aligned with the *M. anisopliae* genome (Fig. 5c and Supplementary Fig. 9). These analyses suggest that adaptation to distinct hosts across kingdoms might result in divergent effector repertoires. However, within species these putative *U. virens* effector genes are highly conserved. We thus hypothesize that a recent founder effect has occurred in *U. virens*, and perhaps a virulent *U. virens* strain, which emerged and established intimate interactions with rice rather recently, spread rapidly thereafter. Over a short co-evolutionary history, effector-triggered host immunity has not evolved to suppress pathogen invasion and therefore effectors have been subjected to little selection pressure. No gene-for-gene resistance has been reported for the rice–*U. virens* pathosystem yet<sup>5</sup>, which supports the hypothesis.

In summary, we reveal the close evolutionary relationship between *U. virens* and *Metarhizium* spp., implying that the Clavicipitaceae fungi exhibit host shifting across plant and animal kingdoms. Besides, we hypothesize that low intraspecific genetic diversity in *U. virens* is perhaps attributable to a recent founder effect. Significant fewer gene inventories involved in cell wall degrading, nutrient uptaking and secondary metabolism reflect a biotrophic lifestyle and flower filament infection style of *U. virens*. Functional studies together with transcriptome and microsynteny analyses indicate that putative effectors play important roles in virulence, pathogenicity and host adaptation. Together, our comparative and functional genomics analyses offer new insights into molecular mechanisms of evolution, biotrophy and pathogenesis of *U. virens*, and lay the groundwork for future discoveries with this increasingly important plant pathogen.

## Methods

**DNA isolation and rice inoculation.** *Ustilagoideae virens* strains UV-8b, FJ1-1a, HB3-1a, LN10-16-1 and LN2010-1-1 were collected from different rice cultivars at different provinces in China, and used for sequencing (Supplementary Table 8). Total DNA was isolated from the mycelium of *U. virens* using a CTAB (cetyl trimethyl ammonium bromide) method<sup>51</sup>. Nuclear DNA was purified through a CsCl density-gradient centrifugation to remove mitochondrial DNA using standard procedures<sup>52</sup>. Rice plants of the cultivar LYP9 that is highly susceptible to RFS were inoculated with conidial suspensions as described with some modifications<sup>53</sup>. Briefly, the *U. virens* monoconidial cultures were grown in potato sucrose broth in an incubator shaker at 150 r.p.m. at 28 °C for 5 days. About 1 ml of conidial suspension ( $2 \times 10^5$  conidia ml<sup>-1</sup>) was injected into the leaf sheaths of field grown rice plants at the booting stage 5–7 days before heading. This was done in the late afternoon with a suitable temperature for better infection. The panicles were harvested at 6 h, 24 h and 48 h post inoculation (h.p.i.), frozen immediately in liquid nitrogen, and then kept at –70 °C for RNA isolation.

**Genome sequencing and assembly.** The UV-8b genome was sequenced with the Roche 454 sequencing platform at Sun Yat-sen University, Guangzhou, China. DNA libraries with 500 bp and 5 kb inserts were subsequently constructed for UV-8b and libraries with 500 bp inserts were constructed for the four other *U. virens* strains. These DNA libraries were 100 bp paired-end sequenced using the Illumina HiSeq2000 at the Beijing Genomics Institute (BGI) in Shenzhen, China. Genome assembly for 454 data was done with Newbler (version 2.5, Roche) using

the default parameters. Data from the Illumina libraries were first trimmed by removing bases with quality score below 20 at both ends and discarding trimmed reads with lengths less than 70 bp. The Newbler contigs were used as single-end data and fed into the ABySS<sup>21</sup> version 1.2.7 assembly pipeline along with the paired-end Illumina data for strain UV-8b. The contigs generated by ABySS were processed with SSPACE scaffolder to construct scaffolds<sup>22</sup>. All of the available paired-end reads were mapped to previously generated contigs to connect adjacent contigs through paired-end data from the 500 bp and 5 kb libraries. Larger scaffolds were then made using GapCloser<sup>54</sup> to complete the whole-genome sequence of *U. virens*. The genome size of *U. virens* was estimated from the UV-8b sequencing data using JELLYFISH (Version 2.0)<sup>23</sup>.

**Gene annotation.** Protein-coding genes in *U. virens* were predicted independently with three *ab initio* predictors SNAP, GeneMark + ES and AUGUSTUS<sup>55–57</sup>. The accuracy of *ab initio* annotation was then improved using RNA-seq data from *U. virens* and from the proteomes of nine sequenced fungi. The RNA-seq data were aligned to the *U. virens* UV-8b genome through TopHat<sup>58</sup>, and potential transcripts in *U. virens* genome were predicted using Cufflinks<sup>59</sup>. The nine sequenced fungal proteomes were aligned to the *U. virens* genome by TBLASTN<sup>60</sup>. Proteins with sequence identity more than 70% were noted and used as templates to predict the exons of *U. virens* genes with Genewise<sup>61</sup>. Subsequently, the outputs from these three methods were transformed to GFF file with local PERL scripts. The resultant GFF files from each of the prediction programs were used as input into the EVM program to integrate the three sources of gene predictions<sup>62</sup>.

**SNP identification and Ka/Ks calculation.** Genome sequencing reads of the four *U. virens* strains were mapped onto the *U. virens* UV-8b reference genome using SOAP2 (<http://soap.genomics.org.cn/soapaligner.html>) with default settings. SNPs were identified through the comparison with the reference by SOAPsnp (<http://soap.genomics.org.cn/soapsnp.html>). For Ka/Ks calculation, sequence reads were mapped to the coding sequences of the 8,426 predicted genes of the UV-8b reference strain and thus obtaining consensus CDS sequences that were aligned with ClustalW2 (ref. 63). The fasta files with aligned sequences were converted into AXT files by parseFastaIntoAXT.pl (<http://code.google.com/p/kaks-calculator/>) and ratios of Ka and Ks were calculated using the Modified YN method<sup>64</sup>. The Fisher's exact test was used to obtain *P*-values in this calculation.

**RNA-Seq and transcriptome analyses.** Total RNA was isolated from *in vitro* cultures and from inoculated rice panicles using RNeasy RNA isolation kit according to the manufacturer's instructions (Aidlab Biotechnologies, Beijing). RNA integrity was confirmed using the 2100 Bioanalyzer (Agilent Technologies) with a minimum RNA integrity number (RIN) value of 7. Total RNA from three biological replicates at each time point (6 h.p.i., 24 h.p.i. and 48 h.p.i.) were combined and used for library construction. All procedures, including mRNA purification, cDNA preparation, end repair of cDNA, adaptor ligation and cDNA amplification were carried out following BGI standard methods for preparing Illumina RNA-seq libraries. Each library had an insert size of ~200 bp. The reads of 49 bp (for *in planta* expression profiling) or 100 bp (for transcriptome analysis of *in vitro* cultures) sequences were generated via Illumina HiSeq2000 at BGI. Raw image data were transformed into sequence data by base calling. All sequences were filtered to remove adaptor sequences and low-quality sequences (where the percentage of the low-quality bases with quality value  $\leq 5$  was greater than 50% in a read). The quality-trimmed reads were mapped to *U. virens* genome using SOAP aligner/SOAP2, allowing up to two base mismatches per read. The RPKM (reads per kb per million reads) was calculated to reflect the expression level of *U. virens* transcripts *in planta*<sup>65</sup>. To compare the differences in gene expression, differentially expressed genes were identified with the criterion of the absolute log<sub>2</sub> ratio value  $\geq 1$  and false discovery rate  $\leq 0.001$  (ref. 66).

**Secreted proteins and potential effector analysis.** Potential secreted proteins in *U. virens* were predicted using SignalP 4.0 (ref. 67). Putative effectors in the *U. virens* secretomes were predicted based on the protein size ( $\leq 400$  amino-acid residues) and the number of cysteine residues ( $\geq 4$ )<sup>68</sup>. The prediction of transmembrane proteins and domain calling were described in the Supplementary Methods.

**Pectinolytic activity assays.** The pectinolytic activities of phytopathogenic fungi were evaluated using pectin plate assays as described previously with minor modifications<sup>69</sup>. Different fungi were grown on the pectin-containing media (1 l medium: 5.0 g pectin, 1.0 g yeast extract, 0.5 g glucose, 2.0 g KH<sub>2</sub>PO<sub>4</sub>, 1 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.9 g Na<sub>2</sub>HPO<sub>4</sub>, trace of MnSO<sub>4</sub>, 15 g agar, pH 7.0). After colonies grew up to 4 cm in diameter, a 1% (w/v) CTAB solution was added to the plates. CTAB precipitates pectin present in the medium, resulting in opaque media where pectin was still present, and a transparent halo where pectin was broken down. The experiment was repeated three times.

**Growth profiles of *U. virens* on single carbon source.** Citrus pectin, glucose, xylan, cellulose and gum guar (all from Sigma) were used as single carbon source

separately in agar medium, to evaluate the growth of different fungi. Inocula of *U. virens* and other fungi (*M. oryzae*, *R. solani*, *F. graminearum* and *C. fragariae*) were placed on these media and incubated at 28 °C with different cultivation times (21, 7, 5, 3 and 2 days, respectively). These tests were repeated three times, and the results were analysed to evaluate fungal mycelium growth, and graphed to visualize differences.

**Functional studies on putative effector proteins.** Putative effector genes (without signal peptide regions) of *U. virens* were amplified with the primer sets listed in Supplementary Table 31. The PCR products were digested with the corresponding restriction enzymes and subcloned into pEDV<sup>70</sup>. *Burkholderia glumae* competent cells were prepared as described with minor modifications<sup>43</sup>. These pEDV constructs were transformed into *B. glumae* by electroporation. Leaves from 3-week-old *Nicotiana benthamiana* plants were infiltrated with *B. glumae* with different effector gene constructs or empty vector (OD600 = 0.4) using needleless syringes. Cell death symptoms were evaluated and photographed at 2–3 days after inoculation.

## References

- Tanaka, E., Ashizawa, T., Sonoda, R. & Tanaka, C. *Villosiclava virens* gen. nov., comb. nov., the teleomorph of *Ustilagoidea virens*, the causal agent of rice false smut. *Mycotaxon* **106**, 491–501 (2008).
- Cooke, M. C. Some extra-European fungi. *Grevillea* **7**, 13–15 (1878).
- Rush, M. C., Shahjahan, A., Jones, J. P. & Groth, D. E. Outbreak of false smut of rice in Louisiana. *Plant Dis.* **84**, 100–100 (2000).
- Singh, A. K. & Pophaly, D. J. An unusual rice false smut epidemic reported in Raigarh District, Chhattisgarh, India. *Int. Rice Res. Notes* **35**, 1–3 (2010).
- Guo, X. *et al.* Progress in the study of false smut disease in rice. *J. Agr. Sci. Tech.* **2**, 1211–1217 (2012).
- Ikegami, H. Occurrence and development of sclerotia of the rice false smut fungus. *Res. Bull. Fac. Agr.* **18**, 47–53 (1963).
- Shan, T. *et al.* Determination and analysis of Ustiloxins A and B by LC-ESI-MS and HPLC in false smut balls of rice. *Int. J. Mol. Sci.* **13**, 11275–11287 (2012).
- Koiso, Y. *et al.* Ustiloxins, antimitotic cyclic peptides from false smut balls on rice panicles caused by *Ustilagoidea virens*. *J. Antibiot.* **47**, 765–773 (1994).
- Koyama, K. & Natori, S. Further characterization of seven bis (naphtho- $\gamma$ -pyrone) congeners of ustilaginoidins, coloring matters of *Claviceps virens* (*Ustilagoidea virens*). *Chem. Pharm. Bull.* **36**, 146–152 (1988).
- Koyama, K., Ominato, K., Natori, S., Tashiro, T. & Tsuruo, T. Cytotoxicity and antitumor activities of fungal bis(naphtho-gamma-pyrone) derivatives. *J. Pharmacobiodyn.* **11**, 630–635 (1998).
- Li, Y., Koiso, Y., Kobayashi, H., Hashimoto, Y. & Iwasaki, S. Ustiloxins, new antimitotic cyclic peptides: interaction with porcine brain tubulin. *Biochem. Pharmacol.* **49**, 1367–1372 (1995).
- Zhou, L. *et al.* in *Mycotoxins: Properties, Applications and Hazards* (eds Melborn, B. & Greene, J. C.) 109–130 (Nova Science Publishers, 2012).
- Biswas, A. False smut disease of rice: a review. *Environ. Ecol.* **19**, 67–83 (2001).
- Singh, R. A. & Dubey, K. S. Sclerotial germination and ascospore formation of *Claviceps oryzae sativae* in India. *Indian Phytopathol.* **37**, 168–170 (1984).
- Shetty, S. A. & Shetty, H. S. Role of *Panicum trypheron* in annual recurrence of false smut of rice. *Transact. Brit. Mycol. Soc.* **88**, 409–411 (1987).
- TeBeest, D. O., Jecmen, A. & Ditmore, M. Infection of rice by the false smut fungus, *Ustilagoidea virens*. *B.R. Wells Rice Res. Studies* **70**–81 (2010).
- Tang, Y. X. *et al.* Elucidation of the infection process of *Ustilagoidea virens* (teleomorph: *Villosiclava virens*) in rice spikelets. *Plant Pathol.* **62**, 1–8 (2013).
- Kepler, R. M. *et al.* Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and *Ustilagoidea* (Clavicipitaceae). *Am. J. Bot.* **99**, 552–561 (2012).
- Bischoff, J. F., Sullivan, R. F., Kjer, K. M. & White, J. F. Phylogenetic placement of the anamorphic tribe *Ustilagoidea* (Hypocreales, Ascomycota). *Mycologia* **96**, 1088–1094 (2004).
- Winnenburg, R. *et al.* PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* **34**, D459–D464 (2006).
- Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
- Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
- Nowrousian, M. *et al.* De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. *PLoS Genet.* **6**, e1000891 (2010).
- Dean, R. A. *et al.* The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986 (2005).
- Hane, J. & Oliver, R. *In silico* reversal of repeat-induced point mutation (RIP) identifies the origins of repeat families and uncovers obscured duplicated genes. *BMC Genomics* **11**, 655 (2010).
- Nakayashiki, H. RNA silencing in fungi: mechanisms and applications. *FEBS Lett.* **579**, 5950–5957 (2005).
- Taylor, J. W. & Berbee, M. L. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* **98**, 838–849 (2006).
- Spanu, P. D. *et al.* Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**, 1543–1546 (2010).
- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. & Warren, R. A. Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiol. Mol. Biol. Rev.* **55**, 303–315 (1991).
- Xue, C., Hsueh, Y. P. & Heitman, J. Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. *FEMS Microbiol. Rev.* **32**, 1010–1032 (2008).
- DeZwaan, T. M., Carroll, A. M., Valent, B. & Sweigard, J. A. *Magnaporthe grisea* Pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. *Plant Cell* **11**, 2013–2030 (1999).
- Lorenz, M. C. *et al.* The G protein-coupled receptor Gpr1 is a nutrient sensor that regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. *Genetics* **154**, 609–622 (2000).
- Gao, Q. *et al.* Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet.* **7**, e1001264 (2011).
- Schardl, C. L. *et al.* Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genet.* **9**, e1003323 (2013).
- Wang, B., Kang, Q., Lu, Y., Bai, L. & Wang, C. Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species. *Proc. Natl Acad. Sci. USA* **109**, 1287–1292 (2012).
- Frandsen, R. J. *et al.* The biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the naphthoquinones and naphthopyrones. *Mol. Microbiol.* **61**, 1069–1080 (2006).
- Kemen, E. *et al.* Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol.* **9**, e1001094 (2011).
- Brown, N. A., Antoniw, J. & Hammond-Kosack, K. E. The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLoS ONE* **7**, e33731 (2012).
- Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101 (2006).
- Xue, M. *et al.* Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet.* **8**, e1002869 (2012).
- Giraldo, M. C. & Valent, B. Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* **11**, 800–814 (2013).
- Sharma, S. *et al.* Deployment of the *Burkholderia glumae* type III secretion system as an efficient tool for translocating pathogen effectors to monocot cells. *Plant J.* **74**, 701–712 (2013).
- Mazet, I. & Vey, A. Hirsutellin A, a toxic protein produced *in vitro* by *Hirsutella thompsonii*. *Microbiology* **141**, 1343–1348 (1995).
- Matsubara, M. & Mizuno, T. The SixA phospho-histidine phosphatase modulates the ArcB phosphorelay signal transduction in *Escherichia coli*. *FEBS Lett.* **470**, 118–124 (2000).
- Singer, M. J., Marcotte, B. A. & Selker, E. U. DNA methylation associated with repeat-induced point mutation in *Neurospora crassa*. *Mol. Cell Biol.* **15**, 5586–5597 (1995).
- Spatafora, J. W., Sung, G. H., Sung, J. M., Hywel-Jones, N. L. & White, J. F. Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes. *Mol. Ecol.* **16**, 1701–1711 (2007).
- Amselm, J. *et al.* Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* **7**, e1002230 (2011).
- Pao, S. S., Paulsen, I. T. & Saier, M. H. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
- VanEtten, H. D., Matthews, D. E. & Matthews, P. S. Phytoalexin detoxification: importance for pathogenicity and practical implications. *Annu. Rev. Phytopathol.* **27**, 143–164 (1989).
- Rogers, S. O. & Bendich, A. J. Extraction of total cellular DNA from plants, algae and fungi. *Plant Mol. Biol. Manual* **D1**, 1–8 (1994).
- Garber, R. C. & Yoder, O. C. Isolation of DNA from filamentous fungi and separation into nuclear, mitochondrial, ribosomal, and plasmid components. *Anal. Biochem.* **135**, 416–422 (1983).
- Wang, S. *et al.* Sporulation, inoculation methods and pathogenicity of *Ustilagoidea albicans*, the cause of white rice false smut in China. *J. Phytopathol.* **156**, 755–757 (2008).
- Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
- Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
- Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004).

57. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990 (2008).
58. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
59. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
60. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
61. Birney, E., Clamp, M. & Durbin, R. Genewise and genomewise. *Genome Res.* **14**, 988–995 (2004).
62. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
63. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
64. Zhang, Z., Li, J. & Yu, J. Computing Ka and Ks with a consideration of unequal transitional substitutions. *BMC Evol. Biol.* **6**, 44 (2006).
65. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
66. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
67. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
68. Klosterman, S. J. *et al.* Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog.* **7**, e1002137 (2011).
69. Molina, S. M., Pelissari, F. A. & Vitorello, C. Screening and genetic improvement of pectinolytic fungi for degumming of textile fibers. *Bra. J. Microbiol.* **32**, 320–326 (2001).
70. Fabro, G. *et al.* Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. *PLoS Pathog.* **7**, e1002348 (2011).

## Acknowledgements

We are grateful to Andrew Bent at University of Wisconsin–Madison for comments on the manuscript and Chengshu Wang at Institute of Plant Physiology & Ecology, Chinese Academy of Science and Guozhen Zhang at China Agricultural University for providing

the *M. anisopliae* and *C. fragariae* strains. The work is supported by the Special Fund for Agro Scientific Research in the Public Interest of China (200903039) and Key Projects in the National Science & Technology Pillar Program 2012BAD19B03 to W.S. and the 973 program 2012CB114000, the Transgenic Project 2012ZX08009003-003, the 111 project B13006 and PCSIRT Program IRT1042 to Y.-L.P.

## Author contributions

Y.Z., K.Z. and A.F. contributed equally to this work as first authors. W.S. and Y.-L.P. managed the project and designed the analysis. Y.H. prepared the DNA and RNA samples. Y.Z. and J.L. performed the genome assembly and annotation. K.Z. performed the transcriptome analyses. A.F. and J.M. performed functional studies on the effectors. J.Y. investigated the pectinolytic activities. M.X., J.B. and B.Z. performed SNP analysis and Ka/Ks calculation. X.S., S.L., M.W. and N.Y. performed genome sequencing and analyses. D.H., Y.-L.P., S.W., M.Z., Z.C., L.Z., J.L., F. H. and C. L. collected strains and performed life cycle studies. W.S., Y.-L.P., Y.Z., L.-J.M., T.H. and J.X. wrote the paper. All of the other authors contributed annotation, analyses or data collection.

## Additional information

**Accession codes:** The Whole-Genome Shotgun project for *Ustilaginoidea virens* has been deposited in DDBJ/EMBL/GenBank under the accession code JHTR00000000.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Zhang, Y. *et al.* Specific adaptation of *Ustilaginoidea virens* in occupying host florets revealed by comparative and functional genomics. *Nat. Commun.* **5**:3849 doi: 10.1038/ncomms4849 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>