Dna Research, 2017, 0(0), 1–14
doi: 10.1093/dnares/dsx031
Full Paper

OXFORD

Full Paper

Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome

Zihong Ye[1,†], Yao Pan[2,3,†], Yafen Zhang[1], Haifeng Cui[1], Gulei Jin[4], Alice C. McHardy[2,3], Longjiang Fan[4,*], and Xiaoping Yu[1,*]

[1]Department of Zhejiang Provincial Key Laboratory of Biometrology and Inspection & Quarantine, China Jiliang University, Hangzhou, China, [2]Department of Algorithmic Bioinformatics, Heinrich Heine University, Düsseldorf, Germany, [3]Cluster of Excellence on Plant Sciences (CEPLAS), Düsseldorf, Germany, and [4]Department of Agronomy & Zhejiang Key Laboratory of Crop Germplasm Resources, Zhejiang University, Hangzhou, China

*To whom correspondence should be addressed. Tel. +86-571-86836006. Fax. +86-571-86914401. Email: yxp@cjlu.edu.cn (X.Y.); Tel. +86-571-88982730. Email: fanlj@zju.edu.cn (L.F.)

†These authors contributed equally to this work.

Edited by Dr. Sachiko Isobe

Received 26 December 2016; Editorial decision 17 June 2017; Accepted 21 June 2017

Abstract

Ustilago esculenta, infects *Zizania latifolia*, and induced host stem swollen to be a popular vegetable called *Jiaobai* in China. It is the long-standing artificial selection that maximizes the occurrence of favourable *Jiaobai*, and thus maintaining the plant–fungi interaction and modulating the fungus evolving from plant pathogen to entophyte. In this study, whole genome of *U. esculenta* was sequenced and transcriptomes of the fungi and its host were analysed. The 20.2 Mb *U. esculenta* draft genome of 6,654 predicted genes including mating, primary metabolism, secreted proteins, shared a high similarity to related Smut fungi. But *U. esculenta* prefers RNA silencing not repeat-induced point in defence and has more introns per gene, indicating relatively slow evolution rate. The fungus also lacks some genes in amino acid biosynthesis pathway which were filled by up-regulated host genes and developed distinct amino acid response mechanism to balance the infection–resistance interaction. Besides, *U. esculenta* lost some surface sensors, important virulence factors and host range-related effectors to maintain the economic endophytic life. The elucidation of the *U. esculenta* genomic information as well as expression profiles can not only contribute to more comprehensive insights into the molecular mechanism underlying artificial selection but also into smut fungi–host interactions.

Key words: *Ustilago esculenta*, whole-genome analysis, fungi–host interaction, *Zizania latifolia*

1. Introduction

Ustilago esculenta is a basidiomycete fungus, which infects ancient wild rice, *Zizania latifolia*.[1] *Zizania latifolia* belongs to the tribe Oryzeae within the Gramineae[2] and was one of the six most important cereal crops in ancient China.[3] After infection with *U. esculenta* around 2000 yrs ago,[4] *Z. latifolia* formed a shuttle-like gall and was

gradually domesticated as an aquatic vegetable called *Jiaobai* in East and Southeast Asia[5] particularly in China, which has antioxidant properties and may prevent hypertension, cardiovascular disease.[6] Similar to other crops domestication, *Jiaobai* plants were artificially selected to increase their suitability to human requirements: taste, yield, storage and cultivation practices. In some unfavourable conditions, like high-dose radiation, water deficiency, unsuitable application of fungicide or unfavourable growing temperature, or if the plants were abandoned, the plants may produce galls being full of dark teliospores (called grey *Jiaobai*) or escape from fungal infection (called male *Jiaobai*).[7,8] It was the long-term artificial selection that effectively maximized the occurrence of favourable *Jiaobai* and enabled *U. esculenta* consecutively to be maintained in host plants. Since *U. esculenta* was the cause of the vegetable, the process of crop domestication actually was the process of artificially selection on interaction complex of fungus and host. A key question, then, is to understand how domestication has shaped *U. esculenta* from the plant pathogen to entophyte and how, in turn, these changes may influence the interaction between the fungus and the host.

It has been shown that host domestication has driven the emergence of the rice blast pathogen *Magnaporthe oryzae*[9] and the wheat pathogen *Zymoseptoria tritici*.[10] In contrast to natural evolution, domestication is due to humans as selective agents; but, just like natural selection, domestication fully depends on genetic variation, mutations, inheritance and demography.[11] The emerging consensus is that domestication causes a distinct imprint on genomes.[12] Integrated and fully analyses of the genome will give insight into the underlying effect of domestication. The comparative genomic study on *Mycosphaerella graminicola* and its wild sister species showed that speciation of *M. graminicola* was associated with adaptation to domesticated wheat and its associated agro-ecosystem.[13] And indeed, comparative genomics has revealed underlying mechanisms of genome evolution, for example, transposon-mediated gene loss driven the rapid evolution of *M. oryzae*,[14,15] and gene loss rather than gene gain resulted in the host jump in the smut fungus *Melanopsichium pennsylvanicum*.[16]

Smut fungi are biotrophic pathogens causing characteristic sympotoms, the replacement of plant organs by black masses of teliospores, in a number of agriculturally important crop plants, mostly grass family (Grmineae). These basidiomycete fungi belong to the order Ustilaginales containing over 50 genera. Several different smut diseases in cereal crops are mostly caused not only by species of the genus Ustilago but also by other genera such as *Sporisorium*.[17] *Ustilago esculenta* together with *Ustilago maydis*, *Ustilago hordei* and *Sporisorium reilianum* are belonging to Ustilaginaceae family but owning some distinct characteristics. For instance, *U. esculenta* spend its entire life cycle in the host plant (Fig. 1),[3,18] while most smut fungi infect and colonize the plant with the dikaryotic mycelium and are released from the host after spore formation. There has been found that the phytopathogenic fungi developed many infection mechanisms to be successfully colonized, including surface sensors like Sho1 and Msb2,[19] intracellular signalling cascades like Cyclic Adenosine monophosphate (cAMP) and mitogen-activated protein kinase (MAPK),[20] secreted proteins to penetrate host plant and establish biotrophy[21] and transcription factors acting as key regulators of differentiation.[22] Besides, for dimorphic fungi, mating reaction and filament formation are essentials of the morphogenetic switch, which should be prepared already before penetration. Although for *U. esculenta*, there might be some special characters considering its unique entophytic lifecycle evolving from a pathogen. In addition, *U. esculenta* inhibits host inflorescence formation and cause stem enlargement, and *Z. latifolia* is its only known host[1] until now. Although many smut fungi are known to be non-host and -tissue specificity. For example, *U. maydis* cause symptom at any aboveground parts of host plant, which is coupled to secreted protein effectors[23] and their organ-specific expression.[24,25] *Ustilago hordei* and *S. reilianum* cause the grain, inflorescence develop into smut sori, respectively.[25] Covered smut of barley and oats is caused by *U. hordei*, and head smut caused by *S. reilianum* is found in both maize and sorghum. Differentiated effectors being targeting different host molecules were responsible for varied infecting strategy of *U. maydis* and *S. reilianum* even they parasitized the same host.[26] And a larger repeat content at important loci, including mating-type and effector loci, were identified for *U. hordei* to explain varied genome evolution on its host when comparing with *U. maydis* and *S. reilianum*.[26]

The availability of full genome sequences of over 50 basidiomycetes, including *U. maydis*, *U. hordei* and *S. reilianum* has accelerated research into basidiomycete genomics. *Ustilago esculenta*, which evolved as the result of crop domestication from plant pathogen to entophyte, offers the possibility to address several major questions in plant pathogen evolution. These include what general changes can be observed in genomes after a long-term domestication? To what extent are pathogenicity and the associated effectors affected by the crop domestication? What helps the smut fungi to sustain the suitable growth condition in the host with delayed teliospores formation? To address these questions, we determine the genome sequence of *U. esculenta*, highlight genomic comparison among *U. esculenta*, *U. maydis*, *U. hordei* and *S. reilianum* and perform transcriptome analysis to characterize the different stages of the *U. esculenta* life cycle in association with phenotype of host plant.

2. Materials and methods

2.1. Strains and growth conditions

Ustilago esculenta strain was isolated from edible *Jiaobai* of Longjiao No 2 in 2011 in Zhejiang province of China. Longjiao No 2 is one of the most widely cultivated *Jiaobai* varieties in east China. The galls were harvested at 160 days after planting. The *U. esculenta* strain was isolated from non-sporulating gall by slice separation method and cultured on potato dextrose agar (PDA) medium. After 6–10 days, the white mycelia that grew out from the tissue were transferred to potato dextrose broth (PDB) medium. The mycelia (M) for further analysis were collected after shaking in PDB medium for 3 days at 28°C. Teliospores were isolated from sporulating gall of grey *Jiaobai* and old gall of edible *Jiaobai* and cultured on PDA medium. After 3–5 days, haploid strains were obtained by single sporidia selection with capillary pipet under microscopy and transferred to PDB medium for additional 3 days shake-culture at 28°C. Enriched sporidia were collected for further analysis. Haploid strains from edible *Jiaobai* called MT strain with deposited number CGMCC Nos. 11841 and11842 in China Center for Type Culture Collection, haploid strains from grey *Jiaobai* called T strain with deposited number CGMCC Nos. 11843 and11844.

2.2. Mating assay

Two haploid strains with heterogametic types from *U. esculenta* were streak cultured on YEPS solid medium (1% yeast extract, 2% peptone, 2% sucrose and 1.5% agar) separately, at 28°C. After 3 days, single colonies were picked into YEPS medium for further culture (28°C and 180 r.p.m). The sporidia cells were collected when OD_{600} reached 1 and re-suspended to OD_{600} of 1.5. The two cultures were mixed 1:1 Vol/Vol and spotted onto YEPS solid medium

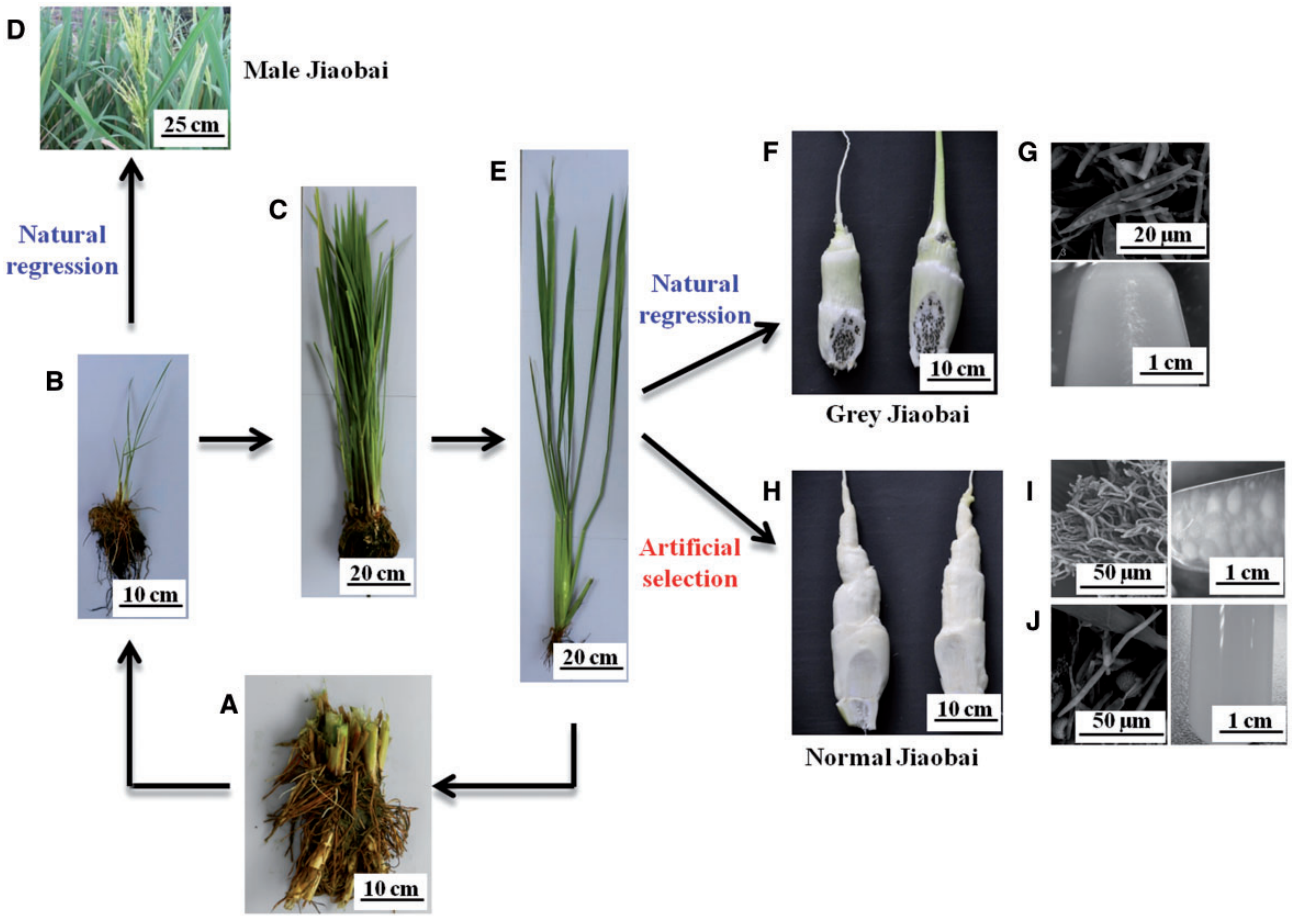

Figure 1. Life cycle of *U. esculenta* in *Z. latifolia*. (A) The fungus overwinters as hypha and teliospores in the rhizomes of host plant. (B) Young seedling sprouts from rhizomes. (C) Hypha invades young seedlings and inflorescence prevented in the infected plant. (D) Young seedling failed to be invaded and becomes male *Jiaobai* with normal inflorescence. (E) The fungus incites gall development at the basal stem. (F) Stems are full of black teliospores and called grey *Jiaobai*. (G) Morphology of fungus isolated from grey *Jiaobai*, which was T sample, observed from liquid culture (Down) and by scanning electron microscope (SEM) (Up). (H) Stems of edible *Jiaobai*. (I) Morphology of fungus isolated from edible *Jiaobai*, which was M sample, observed from liquid culture (Down) and by SEM (Up). (J) Morphology of fungus isolated from edible *Jiaobai*, which was MT sample, observed from liquid culture (Down) and by SEM (Up).

and cultured at 28°C. Microscopic observation and PCR confirmation of the mating loci were performed 4 days after incubation.

2.3. Effect of amino acids on mating and hyphal growth

Mating assays were carried out and the mixed culture of two mating strains were spotted onto the basic medium (K_2HPO_4 1 g/l, $MgSO_4{\cdot}7H_2O$ 0.5 g/l, $FeSO_4{\cdot}7H_2O$ 0.01 g/l, KCl 0.5 g/l, glucose and agar 15 g/l) supplement with different amino acids (arginine, histidine, isoluceine, leucine, lysine, methionine, phenylalanine, proline, tyrosine or valine) as nitrogen source (20 mM) or as trace elements (0.2 mM) when KNO_3 as nitrogen source (20 mM/l). With an interval 12 h, mating colonies were mounted for microscopy and after 3 days, hyphal length were measured.

2.4. Effect of arginine on strain T and MT hyphal growth

Two haploid T strains and MT strains with heterogametic types were cultured and mixed as described in 'Mating assay' in method. The 3 μl 1:1 Vol/Vol mixed cultures were inoculated on YEPS solid medium and Arg medium (K_2HPO_4 1 g/l, $MgSO_4{\cdot}7H_2O$ 0.5 g/l, $FeSO_4{\cdot}7H_2O$ 0.01 g/l, KCl 0.5 g/l, Sucrose 68 g/l and arginine 13.9 g/l) individually. The cultures were incubated at 28°C. With an interval 12 h, hyphal length was measured. The measurement was repeated for three times.

2.5. DNA, RNA preparation, genome sequencing and assembly

Mycelia (M), sporidia from grey *Jiaobai* (T) and from edible *Jiaobai* (MT) were used to isolate DNA and RNA. All the materials were washed with sterile water for three times and centrifuged. After grounding in liquid nitrogen, genomic DNA was extracted using the CTAB method[18] and total RNA was extracted using Trizol (Invitrogen). The isolated RNA was then treated by RNAse-Free DNase and then subsequently treated using Illumina mRNA-Seq Prep Kit (Illumina, San Diego, CA) following the manufacturer's instruction. Four DNA sequencing libraries with different insert lengths (170 bp, 500 bp, 6 Kb and 10 Kb) and 7.7 Gb raw sequence data (Supplementary Table S1) were generated using Illumina HiSeq2000 platform (BGI, Shenzhen, China).

After removing low quality and adapter sequences, the reads were *de novo* assembled using the CLC workbench 5.5.1 (CLC bio).

SSPACE BASIC (version 2.0)[27] was used for scaffold construction and GapFiller (version 1.10)[28] was applied for gap closure.

Genome heterozygosity was first analysed by K-mer analysis and then confirmed by alignment between assembled scaffolds and original reads. Genome Analysis TK (Version 1.6) was used to identify Single Nucleotide Polymorphisms (SNPs).

Genome sequences of *U. maydis* were downloaded from the Broad institute of Harvard and MIT (http://www.broadinstitute.org/annotation/genome/ustilago_maydis/ (6 July 2017, date last accessed)). *Sporisorium reilianum* and *U. hordei* genomic sequences were downloaded and retrieved from the Munich information center for protein sequences (ftp://ftpmips.gsf.de/fungi/Sporisorium_reilianum/ and ftp://ftpmips.gsf.de/fungi/MUHDB/).

2.6. Repeat annotation

A *de novo* repeat database of *U. esculenta* was generated using RepeatModeler (Version 1.0.7, http://www.repeatmasker.org/ (6 July 2017, date last accessed)). RepeatMasker (Version 3.2.7, http://www.repeatmasker.org/ (6 July 2017, date last accessed)) was used to identify repeats from our *de novo* database and Repbase database (http://www.girinst.org/ (6 July 2017, date last accessed)). LTR_FINDER (version 1.0.5)[29] was applied to identify the long terminal repeat (LTR) elements.

2.7. Gene annotation

We predicted genes as follows: (i) *de novo* prediction. Genes were predicted using Augustus (Version 2.03)[30] with training gene sets from the *U. maydis* genome. (ii) Homologue-based prediction. We mapped the protein sequences of *U. maydis*, *S. reilianum* and *U. hordei* to the *U. esculenta* genome using tBLASTn, with a cut-off *E*-value of 10^{-5}. (iii) RNA-seq based prediction. All contigs from RNA-seq were mapped to the *U. esculenta* genome by TopHat (Version 2.0, http://tophat.cbcb.umd.edu/). (iv) All the gene predictions were combined using GLEAN (version 1.0.1)[31] to produce consensus gene sets.

Gene functions were assigned according to the best match of the alignments using BLASTp (*E*-value $< 10^{-5}$)[32] searching against the nucleotide database of NCBI. The motifs and domains of genes were determined by InterproScan (Version 4.5)[33] against UniProt/Swiss-Prot protein database. All genes were classified according to Gene Ontology (GO) and KEGG (Release 48.2) pathways. If the best hit of the genes in any of these processes was 'function unknown', the second best hit was used to assign the function until there were no more hits that met the alignment criteria, then this gene is determined as functionally unknown. OrthoMCL (version 2.0.9)[34] was used to group *U. esculenta* genes into orthologue clusters.

Genes were also analysed for carbohydrate-active enzymes (CAZymes, http://www.cazy.org (6 July 2017, date last accessed)) using dbCAN.[35] Genes related to the biosynthesis of secondary metabolites were analysed using the JVI Secondary Metabolite Unique Regions Finder Web server.[36] SignalP (version 3.0)[37] was used for the prediction of secreted proteins. The candidates were analysed with ProtComp 6.0 (http://www.softberry.com (6 July 2017, date last accessed)) and TargetP (version 1.1)[38] were used to predict protein localization. Information regarding the effects of virulence was retrieved from PHI-base (http://www.phi-base.org/ (6 July 2017, date last accessed)).

The tRNA genes were identified by tRNAScan-SE (version 1.21).[39] For rRNAs identification, the rRNAs from *U. maydis* were aligned against the *U. esculenta* genome using BLASTn (*E*-value $< 10^{-5}$) to identify possible rRNAs. Other non-coding RNAs, including miRNA and snRNA, were identified using INFERNAL (version 1.1)[40] by searching against the Rfam database.[41]

Genome syntenic analysis of smut fungi were performed with MUMmer (version 3.0),[42] the schematic representation of the genome synteny was generated with CIRCOS (http://mkweb.bcgsc.ca/circos/).

2.8. Transcriptome sequencing and analysis

Mycelia (M) and sporidia (MT) of MT strain and sporidia (T) of T strain were collected for transcriptome sequencing. Three samples of same culture were collected for sequencing company to guarantee the quality of RNA extracting and sequencing. The transcriptomes were sequenced using Illumina HiSeq2000 platform (BGI, Shenzhen, China). The reads from different sequenced samples (M, T and MT) were mapped to the whole-genome assembly using SOAP2.[43] The expression quantification was calculated based on RPKM measurement.[44] Only if the gene's expression profile differs between two samples with a fold change > 2 and a *P*-value < 0.01 by edgeR,[45] with estimateDispersions method as 'blind', this gene was considered to be significantly differentially expressed. The transcriptomic data have been deposited at GenBank under accession SRR1611140, SRR5234219 and SRR5234220 for M, T and MT, respectively. To understand the effect of long-term interaction between the host plant and *U. esculenta*, the transcriptomic data of stems from edible *Jiaobai* with *U. esculenta* in the plant and wild *Z. latifolia* (GenBank accession number, PRJNA187578)[46] were downloaded and used.

2.9. Phylogenetic analysis

The phylogenetic relationships of *U. esculenta*, *U. maydis*, *U. hordei*, *S. reilianum* and *Cryptococcus neoformans* were constructed using 45 single copy gene families (Supplementary File S3) by OrthoMCL. The alignments were performed using Clustal omega (version 1.2.0)[47] with default parameters. One hundred bootstrap replicates were generated for each gene family alignment using the Seqboot package in PHYLIP (version 3.69, http://evolution.genetics.washington.edu/phylip/). For each replicate, a maximum likelyhood phylogenetic tree was constructed using Promlk package of PHYLIP under Joes–Taylor–Thornton model with testing of a molecular clock. The consensus tree was built based on 4,500 trees from the bootstrap experiments by the Consense package in PHYLIP with an extended majority rule. The percentage of support by the 4,500 individual trees was indicated on each branch of the consensus tree.

3. Results/discussion

3.1. Sequencing of the *U. esculenta* genome

The collected mycelia of *U. esculenta* were sequenced using Illumina Higseq2000 platform. Four different insert libraries (170 bp–10 kb) were sequenced to generate a total of 4.54 Mb trimmed data with a genome coverage of ×139 (Supplementary Table S1). The reads were *de novo* assembled into 1,869 contigs with a total size of 20.2 Mb (Table 1). Since the strain was diploid, we checked the heterozygosity by k-mer (Fig. 2) and SNP analysis (Supplementary File S1). Although some random sequencing errors exist, only one distribution was found by K-mers with volume peak at 85, indicating the high homozygosity of genome. To confirm the result, we aligned the assembled scaffolds with original reads. Totally, 2,118 SNPs including 228 homologous SNPs were found. Such a low SNP frequency and density further validated the genome homozygosity. The N50

lengths of contigs and scaffolds of the assembly were 101.8 and 404.8 kb, respectively. The assembled draft genome size of *U. esculenta* is moderately larger than *U. maydis* and *S. reilianum* but slightly smaller than *U. hordei*. In total, 6,654 candidates gene were identified based on homology to known proteins and the transcriptomic data (detailed annotations see Supplementary File S2), indicating that *U. esculenta* has similar proportion of protein-coding regions as the other three smut fungi. The proportion of genes encoding secreted proteins was around 10% (663 out of 6,654 genes) of the annotated genes. This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession JTLW00000000. The version described in this article is version JTLW01000000.

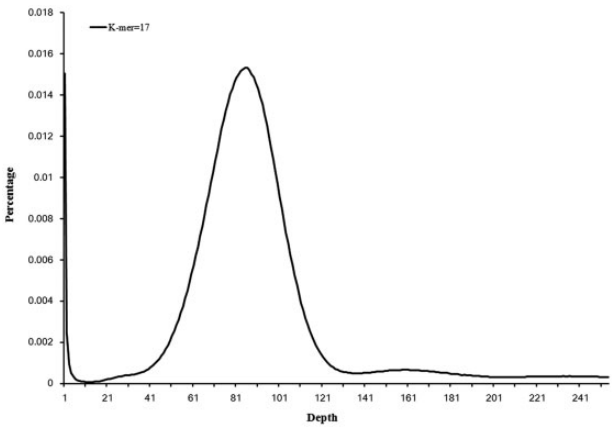

Figure 2. The volume of K-mers is plotted against the frequency at which they occur. The left, truncated peak at low frequency and high volume represents K-mers containing essentially random sequencing errors, whereas the right distribution represents proper (putatively error-free) data. The total K-mer number is 2,424,289,422 and the volume peak is 85. The genome size can be estimated as (total K-mer number)/(volume peak), which is 28.5 Mb.

3.2. Higher intron number and varied genome defence were found in *U. esculenta*

We performed comparative genomic analysis on the four Ustilaginales, *U. maydis*, *U. hordei*, *S. reilianum* and *U. esculenta*. *Ustilago esculenta* shares a high degree of genomic synteny with *U. hordei*, *U. maydis* and *S. reilianum* (Fig. 3). Orthologue comparison of *U. esculenta*, *U. maydis*, *U. hordei* and *S. reilianum* identified 3,602 genes clusters that were common to all four species (Fig. 4). *Ustilago esculenta* had the highest orthologue similarity to *U. hordei* (88.6%, 4,166 out of 4,699). It shared 87.2% (4,102 out of 4,699) and 82.1% (3,860 out of 4,699) gene orthologue similarity with *U. maydis* and *S. reilianum*, respectively. The results are consistent with the phylogenetic analysis, which showed that *U. esculenta* is more closely related to *U. hordei* than to *U. maydis* and *S. reilianum* (Fig. 5). Specific *U. esculenta* gene cluster (number: 399) with classifications (number: 103) were enriched in genes associated with secreted proteins (number: 62).

Unconserved regions still existed in the *U. esculenta* genome (white regions in the red Ue genome in Fig. 3). The length of these non-syntenic regions was 3.1 Mb, 15.1% of the whole genome. GC content of these regions was 55.63%, a little higher than that of syntenic regions (51.3%). Totally, 1,589 candidate genes were predicted on these regions (Supplementary File S5), which were spreading in 59 pieces (each piece containing at least 4 genes). Based on the synteny analysis of orthologous genes, 243 genes were found to be orphan genes, and 761 out of 1,346 orthologous genes were annotated as hypothetical protein or uncharacterized protein. No enriched function was found according to GO analysis.

Table 1. The genome statistics of *U. esculenta* compared with other sequenced smut fungi

Genome specifics	*U. esculenta*	*U. hordei*	*U. maydis*	*S. reilianum*
Assembly statistics				
Total contig length (Mb)	20.2	20.7	19.7	18.2
Total scaffold length (Mb)	21.4	21.2	19.8	18.4
Average base coverage	×100	×20	×10	×20
N50 contig (kb)	101.8	48.7	127.4	50.3
N50 scaffold (kb)	404.8	307.7	817.8	738.5
Chromosome no.		23	23	23
GC content (%)	51.3	52	54	59.7
Coding	56.7	54.3	56.3	62.6
Non-coding	44.2	43.4	50.5	54.3
Gene annotation				
Percentage coding (%)	61	57.5	61.1	65.9
Average gene size (bp)	1,964	1,708	1,836	1,858
Average gene density (kb/gene)	3.2	3.0	2.9	2.8
Protein-coding genes	6,655	7,113	6,786	6,648
Of homologous proteins (%)	6,045 (91)	6,325 (89)	6,214 (92)	6,394 (96)
Exons	12,237	10,907	9,783	9,776
Exons/gene	1.84	1.53	1.44	1.47
Average exon size (bp)	994	1,107	1,230	1,221
Intron	5,563	3,161	2,997	3,103
Introns/gene	0.84	0.44	0.44	0.46
Average intron size (bp)	185	141	142	144
Average intergenic distance (bp)	1,052	1,186	1,127	929
tRNA genes	152	110	111	96
TEs and repeats (%)	7.5	2.0	8.0	0.8

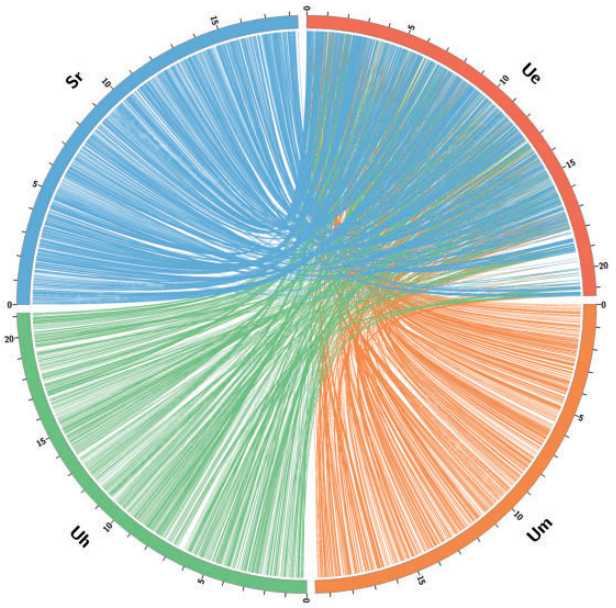

Figure 3. Genomic synteny of *U. esculenta* with three other Ustilagoinaceae species. The genomes of four fungi are depicted by different colours: *U. esculenta* (Ue) in red, *U. maydis* (Um) in orange, *U. hordei* (Uh) in green and *S. reilianum* (Sr) in blue. The scales and labels represented the genome size (Mb) are displayed outside. The synteny relationships between *U. esculenta* genome and the remaining three species were demonstrated by connected lines.

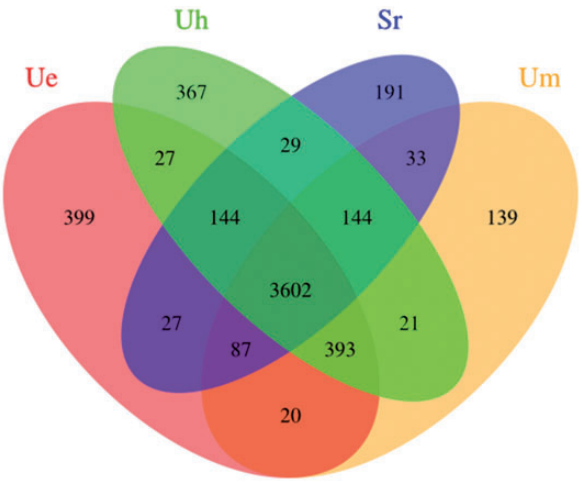

Figure 4. Gene orthologue comparisons between four species of Ustilagoinaceae family. The four fungi are represented by different colours: *U. esculenta* (Ue) in red, *U. maydis* (Um) in orange, *U. hordei* (Uh) in green and *S. reilianum* (Sr) in blue.

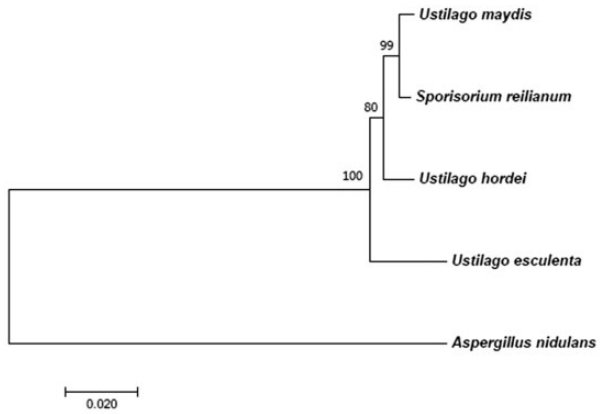

Figure 5. Molecular phylogeny of four species in Ustilagoinaceae family. Consensus phylogenetic tree was constructed from 4,500 trees (45 gene family × 100 bootstrap replicates). Trees were mid-point rooted. The numbers on the branches indicate the percentage of the 4,500 trees support the consensus tree. The fungus *A. Nidulans* is used as an out group.

These four smut fungi have similar genome sizes, GC content and gene-coding percentages. The most obvious differences between our sequenced genome and the other three Ustilaginales are the number and average length of introns and exons. *Ustilago esculenta* has ~2,000 more exons and introns individually compared with the other three genomes. This results in higher ratio of exons or introns to genes in *U. esculenta*. The number of intron-free genes is the lowest and the number of introns per gene (0.84 intron per gene) is the highest in *U. esculenta* among all four genomes. However in comparison to *Aspergillus nidulans* (2–3 introns per gene) or *C. neoformans* (5 introns per gene), *U. esculenta* has fewer introns. Similar to other Ustilaginales, introns of *U. esculenta* were primarily distributed at the 5′-end of genes, likely due to the selective pressure during evolution.[5] The number of introns may also be determined by the rates of evolution. Roy and Gilbert[48] suggest that a large number of introns are often found in relatively earlier ancestors based on maximum likelihood reconstructions. Slow-evolving species may retain more introns than the fast-evolving species. This may explain the higher intron numbers in *U. esculenta* compared with other three Ustilaginales. Farmers keep selecting on the Ustilago–Zizania interaction complex every planting season to retain the vegetables that maintain the desired morphology of the variety. Only a few plants with the potential to produce high quality *Jiaobai* are kept for propagation. The artificial selection may reduce the selective pressure from nature, and consequently, slow down the evolution rate and result in higher intron numbers in *U. esculenta*.

Analysis for repeating sequences showed that the repeated sequences did not cluster in a specific region but were randomly distributed across the draft genome of *U. esculenta*. The coverage of repeats (7%) and transposable elements (TEs) in the assembled contigs are similar to that in *U. hordei* (7.8%).[25] It is nearly 3 and 10 times higher than in *U. maydis* and *S. reilianum*, respectively. Surprisingly, the coverage of unclassified repeats in the *U. esculenta* assembled scaffolds (2.9%) was 10 times higher than that reported in *U. hordei* (0.23%).[25] TEs expansion in fungi may increase the size of the genomes but it could also help the species adapt to a new or challenging environment.[49] However, two-thirds of the *U. esculenta* TEs have no known functions. The *U. esculenta* TEs showed low similarity with their counterparts found in *U. hordei*.

Unlike *U. hordei*, we did not find a repeat-induced point (RIP) mutation mechanism in *U. esculenta*. The RIP mutation is a fungal-specific genome defence that repetitive elements mutated from CpN to TpN (GpN to ApN).[50] RIP may be a limiting factor for TEs replication[51] and it can control TEs activity. *Ustilago maydis* lacks genes responsible for RNA interference (RNAi) machinery. Similar to *U. hordei* and *S. reilianum*, genes related with RNAi mechanism were annotated in *U. esculenta*. Those included one sequence of *Ago1* (g3793), one sequence encoding *Dcl1* (g4220) and three sequences

encoding RNA-directed RNA polymerases (*RdRP1*, g3276; *RdRP2*, g1610 and *RdRP3*, g4076). Chromodomain-coding HP1-like (*Chp1*, g238) and C5-cytosine methyltransferase (*DNAme*, g3112) that were detected in *U. hordei* and *S. reilianum* were found in *U. esculenta* as well. *Ustilago esculenta* showed highly conserved synteny with *S. reilianum* and *U. hordei* in *Dcl1*, DNAme and *RdRP3*. Only *U. hordei* carried additional genes with functions associated with retrotransposition in the *Chp1* (UHOR_15241) and *RdRP1* (UHOR_13400, UHOR_08875 and UHOR_13402). The *RdRP1* locus of *U. esculenta* showed a distinct pattern compared with *U. hordei* and *S. reilianum*. There was one gene missing (homologue of UHOR_05573) in *U. esculenta*. It indicates that these four smut fungi have evolved differently regarding genome defence when they were under different selection pressure. The possession of RNAi-related genes and the lack of RIP may suggest that *U. esculenta* utilizes RNAi for genome defence to control TE proliferating and activity under high selection pressure. Although *U. hordei* balanced TE activity by combination of RNAi, methylation and RIP mutagenesis,[49] *U. maydis* may maintain genome stability through highly efficient recombination system for lack of universal mechanisms for TE control and heterochromatin formation.[49] Genome defence mechanism of *S. reilianum*, an intermediate between *U. hordei* and *U. maydis*, remains to be determined although silencing pathway components being represented in its genome.

3.3. DEGs between different strains

To further determine changes of gene expression between different life stages and different strains, digital gene expression profiling of M, T and MT were investigated (Supplementary File S2 and Table S2). Totally, 6,344, 6,307 and 6,335 genes were expressed at M, T and MT, respectively, and 97.7% (6,250 out of 6,394) were constitutively expressed, while only 52 genes were strain-specific and 92 genes were expressed at two strains (Fig. 6). Four hundred and seventy-seven differentially expressed genes (DEGs) (P-value < 0.01 and fold change of expression > 2.0) between mycelia and sporidia or between sporidia of T strain and MT strain were identified (Supplementary File S2).

GO term enrichment analysis of the 477 DEGs indicated that the DEGs correlated with transport (transmembrane transport and transporter activity), localization, oxidation reduction, oxidoreductase activity, ion binding, cation binding, cofactor binding, co-enzyme binding, membrane and so on were enriched. Interestingly, when compared with sporidial cells, most of DEGs (394 out of 441) were up-regulated in mycelia cells. And similar pattern was observed for sporidia of T strain when compared with that of MT strain (Supplementary Table S3), indicating a more active status of T strain cells.

3.4. Structure and function of mating-type loci

Genes that have a strong effect on speciation are those involved in mating, hyphal fusion or dikaryon formation, and those associated with ecological adaptation.[15] The *a* mating locus comprises a pheromone-receptor system, which is conserved in the sequenced smuts.[52] *Ustilago maydis* and *S. reilianum* have tetrapolar mating systems and their mating loci *a* and *b* locate in two different chromosomes. Whereas the mating loci *a* and *b* in *U. hordei* are in the same choromosome (bipolar).

A gene cluster of *a2*-locus is found in *U. esculenta* on Scaffold 34 containing the left border gene *lba* (g4094), one mating pheromone gene *mfa2.1* (g4096), the right border gene *rba* (g4097) and the *panC* (g4098) by bioinformatic analysis. Also the gene cluster

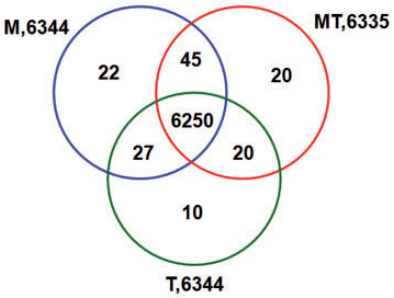

Figure 6. Clusters of expressed genes in M, MT and T strains. A expressed gene number was listed for each components and strains.

of *a3*-locus is found on Scaffold 161, containing one mating pheromone gene *mfa3.2* (g6433) and one pheromone receptor gene *pra3* (g6434). Comparing to other smut fungi, the *a* locus should have at least two genes, pheromone gene and pheromone receptor gene. So we did the homologous comparison to the sequence from *lba* gene to *rba* gene on Scaffold 34, finding the pheromone receptor gene *pra2*, another mating pheromone gene *mfa2.3* and the mitochondrial protein gene *rga2* in the *a2*-locus, which was confirmed by PCR with the primers designed from *lba* and *rba*. Same works have been done in the *a3*-locus. However, the right border gene could not be found and, instead, a long sequence mainly encoding transposase were found at the end of the Scaffold 161. Additional PCR were carried out with the primers designed from *lba*, *rba* and the intermediates, and another mating pheromone gene *mfa3.1* was found nearby the *rba* gene. The two *a* loci of *U. esculenta* were syntenic to that of *S. reilianum*, respectively, with conserved gene content, order and position, except for two remarkable characters (Fig. 7).

First, *a3* locus was distinctively enlarged in *U. esculenta*. The *a2* locus encompassed 10-kb regions, while *a3* locus extended to over 22-kb, with a transposase flanking to *panC* and a ~14-kb segment inserted between *pra3* and *mfa3.1* (Fig. 7). A comprehensive search of *a3* locus for repetitive elements identified seven LTR (two Gypsy, three Copia and two DIRS) occupied 31.6% of this large region. And 76.9% of LTR located at a 14-kb insertion region. This was similar to *MAT-1* of *U. hordei*, which accumulated repetitive elements (many were LTR) covering more than 50% of this region.[25] TEs, having been proved to be implicated in the mating-type rearrangement in basidiomycete[53,54] and yeast,[55] are associated with steps in the evolution of sexual types.[49] Selection pressure from the co-evolving hosts makes Ustilago and other obligate parasites continue adaptation and favour the presence of higher frequencies of TEs,[49,56] which may explain the extended segment of mating locus in *U. esculenta* and *U. hordei*. Recent research found that Gypsy-like elements were related to ovule development in sexual but not apomictic (asexual) genotypes of plants[57] and transposons drove sex chromosome evolution with the evidences from *Drosophila Miranda*.[58] Therefore, the high frequency of TEs in mating locimay help *U. esculenta* to maintain adaptive potential with regulation of the mating-type rearrangement and the sexual reproduction stage but more in-depth analysis is needed.

Second, the gene *lga2*, which ensuring uniparental mitochondrial DNA inheritance[59] and being major component interfering with pathogenic,[60] is missing from *a2* locus and also is not present in the genome. Uniparental inheritance of mitochondria dominates among sexual eukaryotes and is influenced by the mating-type loci in many species, such as slime mold *Physarum polycephalum*, basidiomycete

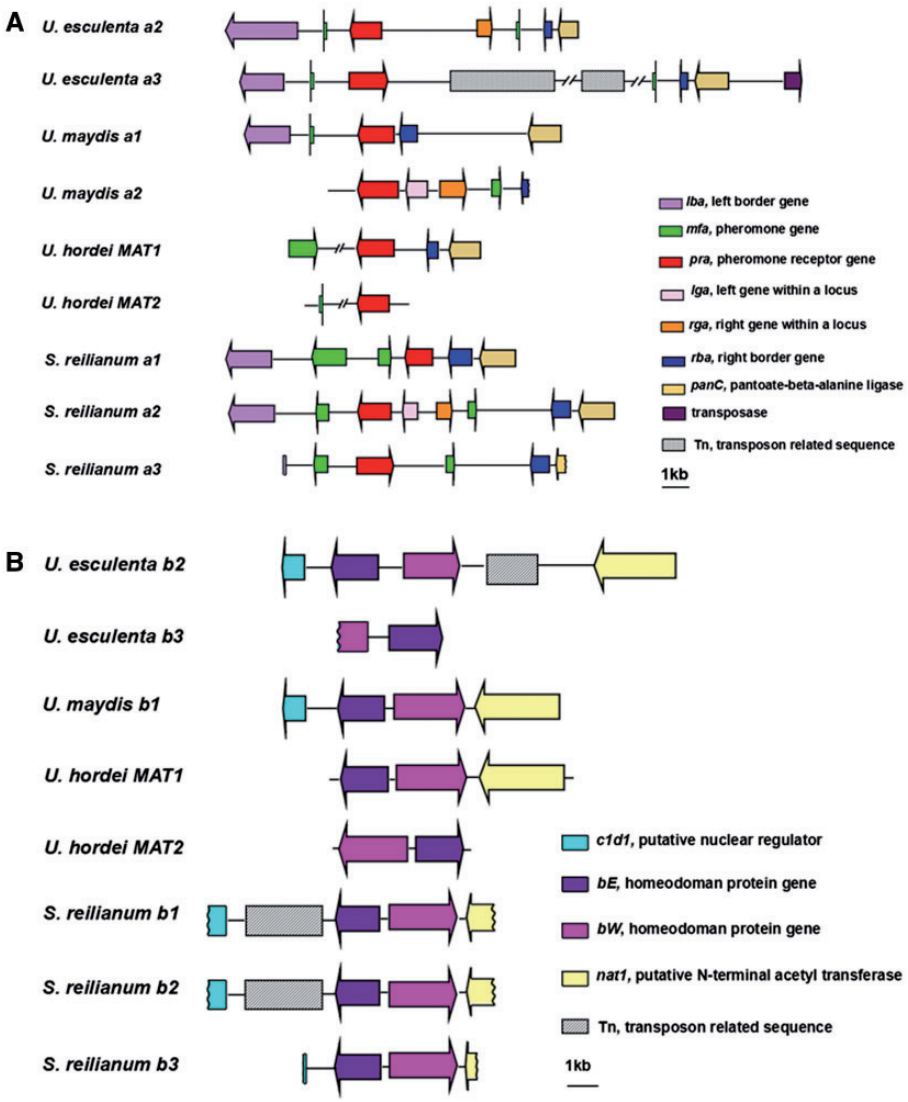

Figure 7. Mating-type loci structures of Ustilagoinaceae species. Genes are indicated by arrows. Arrows directions represent the genes orientations. The same colour indicates the same gene family and the detailed gene functions are shown at the right corners of the figures. The figures were drawn based on the real length of gene loci. (A) Indicates the loci a structures from the *lba* gene to the *panC* gene. The broken lines show the actual length is longer than it draws. (B) Indicates the loci b structures which include the two homeodomain proteins bE and bW.

fungus *C. neoformans* and *U. maydis*, green alga *Chlamydomonas reinhardtii*.[61] The *lga2* and *rga2* genes are specific to the *a2* mating-type locus and direct uniparental mtDNA inheriance by mediating elimination of the *a1*-associated mtDNA in *U. maydis*. Lga2 has a negative role on mitochondrial fusion and may functions on mtDNA elimination, whereas Rga2 can counteract the mtDNA elimination from Lga2-dependent.[59] Besides *rga2*, important genes involved in mitochondria inheritance were found in *U. esculenta*, for example, *dnm1* (g6223, a central component of mitochondrial fission), *fis1* (g2409, involved in mitochondrial fission), *fzo1* (g3102, key component of mitochondrial fusion) and *mrb1* (g5377, mitochondrial p32 family protein). It was implied that *U. esculenta* might mediate its mitochondrial genome integrity and inheritance in a differentiated way compared with wild *U. maydis* strain, but a similar way to *lga2* deletion strain, which leads to biparental inheritance. This may be benefit the fungi to maintain its entophytic life during artificial selection, for biparental inheritance, coupled with mitochondrial recombination, has been speculated to be a potentially adaptive strategy under challenging environmental situations.[62]

In addition to the *a* locus, Ustilaginales have the *b* locus encoding two transcription factors that can form an active heterodimer, if they come from compatible mating partners. Scaffold 22 contained a cluster of the *b* mating-type genes *bE2* (g2853) and *bW2* (g2854), which were significantly up-regulated in sporidia. In *S. reilianum*, *bW* is separated from the neighbouring gene *nat1* by a transposon. Similarly a transposon-related sequence was also found between *bE* and *c1d1* in *U. esculenta*. As for the *a* locus, a second copy of the *b* locus was present in the mycelial culture containing *bE3* and *bW3*. The main target of bW/bE is the transcription factor *rbf1*,[63] which also has an orthologue in *U. esculenta*. To test if mating and filament induction occurs, we generated haploid sporidial cultures from *U. esculenta* teliospores. Mixing of two sporidial cultures led to induction of filaments showed compatible mating types recognized each other (Supplementary Fig. S1). This indicated that the genetic

program for mating and induction of filaments was maintained in *U. esculenta*.

In our genome assembly, the *a* and *b* mating-type genes were present in different scaffolds (Scaffolds 34 and 22). Without further experiments, it is difficult to conclude if *U. esculenta* has a bipolar or tetrapolar mating system. However, we found the left border gene *lba* existed in *a2* and *a3* loci and *c1d1* gene in *b2* locus, indicating the complete border genes existed in *a* and *b* loci in *U. esculenta*, similar to that of *S. reilianum*. Besides, two pheromone genes were detected in both *a2* and *a3* locus, indicating additional *a* locus existed to provide pheromone receptor to sense the pheromone (mfa3.1 and mfa2.1). So there should be three *a* loci in *U. esculenta*, although the third one were not found in our genome draft. Base on the mating loci comparison, we assume that the mating system of *U. esculenta* is more similar to *S. reilianum* than *U. hordei*.

Recognition of the opposite mating partner leads to activation intracellular signal transduction network of a MAPK module and cAMP signalling pathway. Both have been shown to influence mating and later stages of the pathogenic development.[64,65] Most genes reported to be involved in meiosis, MAPK and cAMP signalling cascades in pathogenic Ustilaginales were found in the *U. esculenta* genome. We also found orthologues of the transcription factors *prf1*[66] and *clp1*,[67] which are required for cell fusion, filamentous growth and pathogenic development. The genes involved in the signalling network of *U. maydis* (Fig. 8) were also found in *U. esculenta* except *sho1*. All of these genes expressed and up-regulated expression were only detected for in strain M.

Based on the RNA-Seq data, we found that all genes involved in MAPK, cAMP and transcription factors expressed in sprodial cells and mycelia but only *kpp6* (g4166), *biz1* (g3925) significantly up-regulated in mycelia (M). And in sporidial cells (both in T and MT) the pheromone-receptor system *mfa2.3*, *pra2* was higher expressed, while it was expressed lower in mycelia (M). Higher expression of the pheromone-receptor system in sporidial cells would be consistent with capacity of sporidial cells to mate resulting in mycelial growth.

3.5. Distinct amino acids responding mechanism on balancing interaction between *U. esculenta* and *Z. latifolia*

Nutrition is essential prerequisite for the onset and manifestation of an infection by pathogenic microorganisms, and the plant pathogenic fungi usually delayed to digest host cells to obtain a range of nutrients.[68] Biotrophic fungi, including *U. maydis*, appear to absorb the nutrients available in the apoplast surrounding living host cells for much of its pathogenic growth phase, which helped to avoid plant defence systems throughout the infection process.[69] It makes the fungi have to maintain sets of genes encoding biosynthetic and degradative enzymes for primary metabolism. To fully understand relationship between *U. esculenta* and its host, we checked the genes involved in central metabolic pathways, such as glycolysis, gluconeogenesis, *β*-oxidation, glyoxylate cycle, tricarboxylic acid cycle and *γ*-aminobutyricacid (GABA) shunt (Supplementary File S6). Nearly all enzymes can find orthologues in *U. esculenta*; however, *ilvC* (important for all branched amino acids) and certain genes for biosynthesis of tyrosine, phenylalanine, histidine, alanine and serine were missing in *U. esculenta* (Fig. 9). *Ustilago maydis* also lacks a couple of these genes but contains *ilvC*. Since *U. esculenta* was an entophyte, we checked the candidate genes of amino acid biosynthesis and its expression in the host plant, by RNA-seq analysis of stem tissues from edible *Jiaobai* and wild *Z. latifolia* (which had no *U. esculenta*

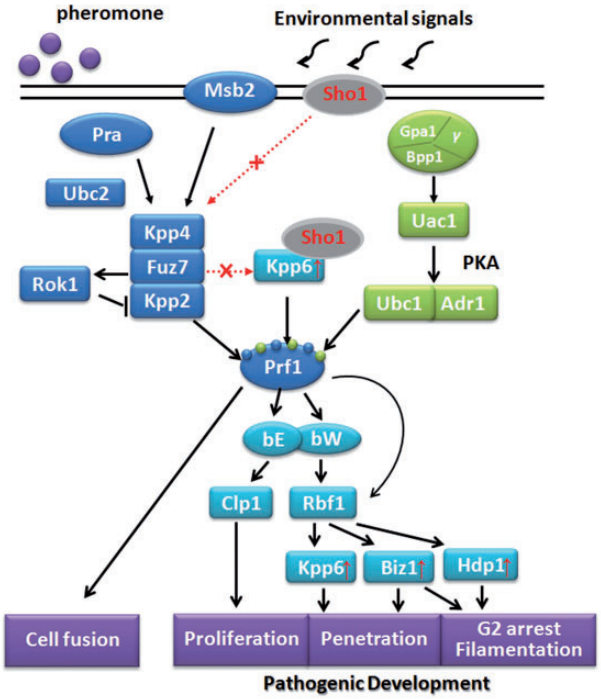

Figure 8. Networks of signalling and transcription factors regulate the pheromone response. Key components of cAMP-mediated Protein Kinase A (PKA) signalling are heterotrimeric G protein a subunit Gpa3, b subunit Bpp1, adenylate cyclase Uac1 and regulatory and catalytic PKA subunits Ubc1 and Adr1, respectively.[64] Lipopeptide pheromone activates its cognate seven-transmembrane domain receptor (Pra). The activated receptor triggers a signalling pathway containing the scaffold protein Ubc2 and a pheromone responsive MAPK module (MAPKKK Kpp4, MAPKK Fuz7 and MAPK Kpp2).[67] The MAPK regulates Prf1 via phosphorylation and at the transcriptional level. At least four different transcription factors, Rop1, Crk1, Hap2 and Prf1, recognize defined cis-active elements in the prf1 promoter (RRS1-3: Rop1-responsive sequence; CCAAT: CAT box; PREs: pheromone response elements; UAS: upstream activating sequence; bent arrow: transcriptional start site). In addition, MAPK activity is counteracted by phosphatase Rok1, which is activated at the transcriptional level by the MAPK module, resulting in negative feedback regulation. Prf1 leads to expression of the b-mating type locus gene bE and bW. After cell fusion, sexual and pathogenic development is orchestrated by the bE/bW-heterodimer, coordinating the key players of important developmental steps: Clp1 is required for in planta proliferation; Rbf1 is the central transcriptional regulator, controlling appressoria formation and penetration of the host plant via the transcription factor Biz1 and the MAPK Kpp6, as well as filamentous growth and the G2 cell cycle arrest by concerted action of Biz1 and Hdp1.[62]

inside). Remarkably, all the missing genes of amino acid biosynthesis in *U. esculenta* were found in *Z. latifolia* and most of them showed a higher expression level in edible *Jiaobai* (Fig. 9). This kind of cooperation was also found in symbiotic system of aphid host and its symbiont[70]; and in aphid, host gene expression and symbiont capabilities are closely integrated within bacteriocytes, whose function as specialized organs of amino acid production. Here, we deduced that *U. esculenta* is a strict biotrophic fungus, and a partial complementary mechanism had formed after 1000 yrs of symbiosis. However unlike insects and their symbionts, it was the host plant providing the intermediate amino acid production.

The amino acid effect on mating and *in vitro* mycelium growth of *U. esculenta* and *U. maydis* are very attracting. We found when using arginine as main nitrogen source, it inhibited the fusion of yeast

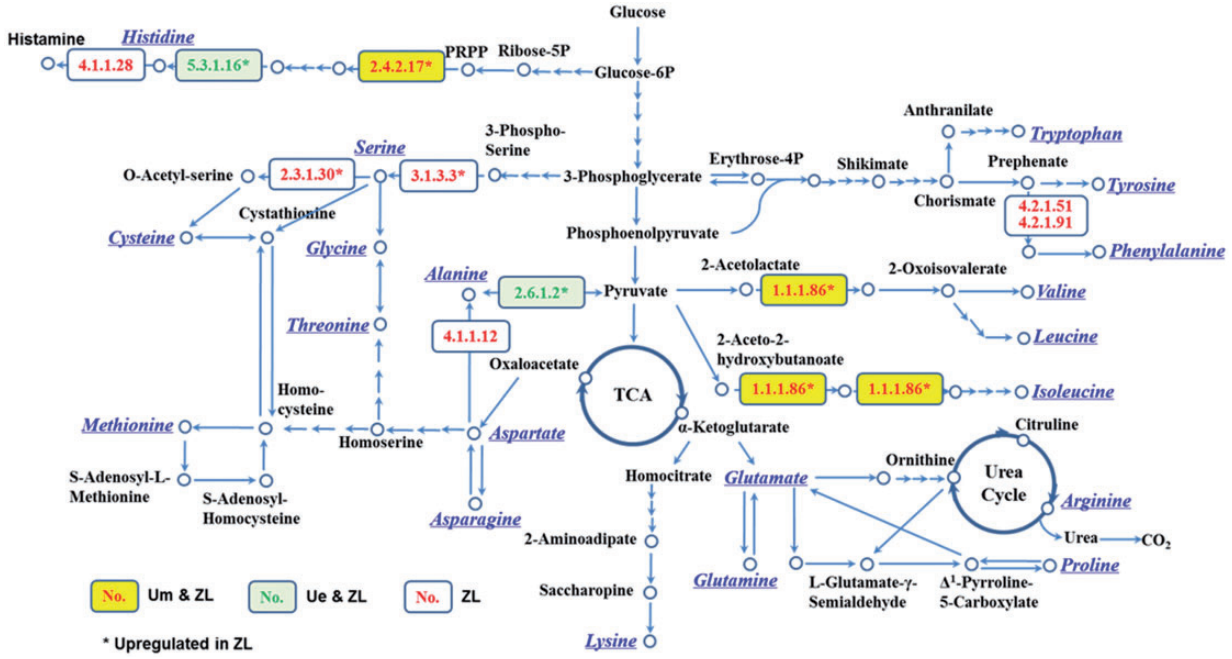

Figure 9. Amino acid biosynthesis and metabolism pathway in *U. esculenta*. Based on the amino acid biosynthesis and metabolism pathway in plants and fungi, the putative amino acid biosynthesis and metabolism pathway map of *U. esculenta* was drawn. Sets of genes encoding biosynthetic and degradative enzymes, which showed differences in the genome of *U. maydis*, *U. esculenta* and *Z. latifolia*, were marked in blue box. Therein, the red genes ID in the box filled with yellow indicated the genes cannot find in *U. esculenta*, the red genes ID in the box with white bottom colour showed the genes only can find in *Z. latifolia*, while the green genes ID indicated the genes cannot find in *U. maydis*. All the genes marked with * means the up-regulated expression during stem swollen progress based on transcriptome sequencing data basic analysis.

Table 2. Comparision of the mating response to the macro- or micro-amino acids effect *in vitro* between *U. esculenta* and *U. maydis*

No AA	*U. esculenta*				*U. maydis*			
	micro-AA		macro-AA		micro-AA		macro-AA	
		Hyphae length (mm)		Hyphae length (mm)		Hyphae length (mm)		Hyphae length (mm)
Arg	++	328.77 ± 18.40^{gh}	–	–	++	205.81 ± 56.62^{ij}	–	–
Tyr	++	375.08 ± 41.63^{defgh}	+++	348.16 ± 17.77^{fgh}	++	102.19 ± 73.26^{klm}	+++	354.32 ± 11.65^{fgh}
His	++	348.54 ± 35.81^{fgh}	+++	362.12 ± 15.82^{efgh}	+	10.86 ± 1.96^{n}	+++	454.41 ± 17.12^{bcd}
Ile	+	338.62 ± 145.33^{fgh}	++	391.33 ± 24.05^{defg}	++	133.01 ± 54.49^{jkl}	+++	206.62 ± 21.55^{ij}
Leu	+	196.78 ± 135.4^{1j}	+	194.12 ± 56.92^{j}	+++	195.51 ± 22.02^{j}	++	158.53 ± 30.17^{jk}
Val	++	524.01 ± 138.44^{ab}	++	290.63 ± 286.18^{hi}	++	163.55 ± 30.96^{jk}	+++	217.12 ± 94.81^{ij}
Met	+++	580.42 ± 36.09^{a}	+++	423.34 ± 71.74^{cdef}	+	24.38 ± 8.5^{1mn}	–	–
Phe	++	303.42 ± 15.77^{gh}	+++	593.88 ± 36.95^{a}	+	42.04 ± 11.67^{mn}	+	82.33 ± 17.05^{klmn}
Pro	++	443.20 ± 18.59^{bcde}	+++	334.24 ± 46.49^{fgh}	+	140.69 ± 39.60^{jkl}	+	52.86 ± 26.72^{lmn}
Lys	+++	487.04 ± 35.42^{bc}	++	311.14 ± 16.30^{gh}	+	59.13 ± 38.52^{lmn}	–	–

All the hyphae length data were detected 3 days after mating assays of *U. esculenta* and *U. maydis* on basic solid medium with different kind and amount amino acid added. The capital letter as a superscript showed significant differences analysis between edible and grey *jiaobai* at $P < 0.05$ level, analysed by one-way ANOVA analysis with Duncan method. All the amino acids (AA) name were abbreviation listed in the table. Also, integrated mating ability analysed of the hyphae length and amount (Supplementary Fig. S3) were estimated. Therein, '-' means no hyphae formation; '+' means less hyphae or short hyphae length; '++' means normal hyphae amount and length and '+++' means large amount hyphae and long hyphae length.

cells (Table 2, Supplementary Fig. S2), which was contrary to the promoting effects on mycelia formation in *Ceratocystis ulmi*[71] and *Candida albicans*.[72] The arginine biosynthesis pathway was up-regulated in yeast/mycelia switching,[72] which was the most essential response in the progression of infection. In *U. esculenta*, the presence of exogenous arginine may exhibit feedback inhibition on arginine biosynthesis.[73] The host plant may conveniently employ this feedback mechanism to modulate *U. esculenta* infection. It was consistent with the results of High Performance Liquid Chromatography (HPLC) detection for amino acids content in stems of grey *Jiaobai* and edible *Jiaobai*, that free arginine content is significantly higher in *Jiaobai* with disease symptom (grey *Jiaobai*) than in the edible *Jiaobai* (Supplementary Table S4). But more studies need to be done to show how the arginine regulated

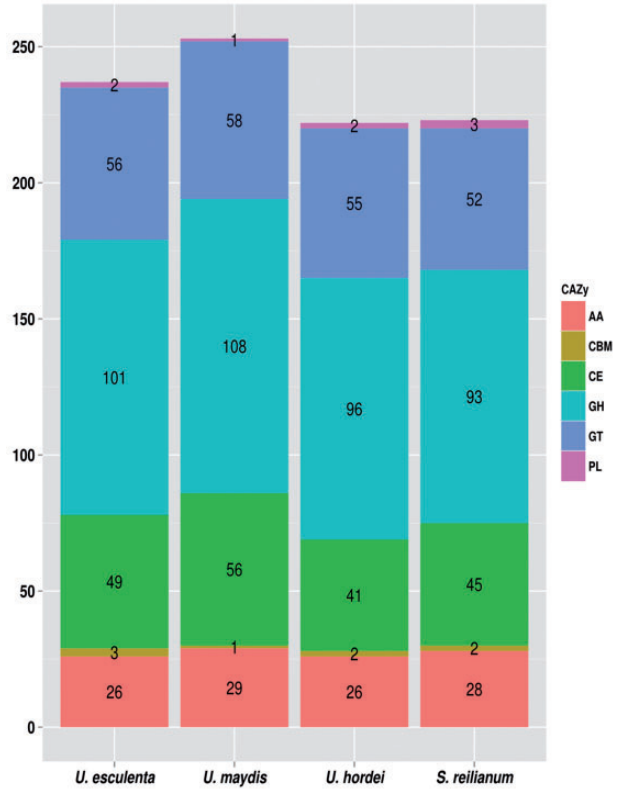

Figure 10. CAZy compositions of the four Ustilagoinaceae species. The numbers represent the number of genes encoded of each CAZy category. The CAzy categories are composed of AAs: auxiliary activities; CBMs: carbohydrate-binding modules; CEs: carbohydrate esterases; GH: glycoside hydrolase; GTs: glycosyl transferases and PLs: polysaccharide lyases.

the dimorphism and what signalled the regulation-feedback system.

Also, results of mating assays indicated some distinct regulation mechanism of amino acids in *U. esculenta*. First, in mating progress, *U. esculenta* may reduce dependence on branched chain amino acids, which were defective in *U. esculenta* (Fig. 9, Table 2, Supplementary Fig. S2 and File S7). For adding leucine, isoleucine or valine could promote mycelia formation in *U. maydis* but had little effect on *U. esculenta* (Table 2, Supplementary Fig. S2). Second, *U. esculenta* may adapt to proline, methionine, lysine, phenylalanine and their metabolism, which were closely related to plant disease resistance reaction.[74–80] In *U. maydis*, adding each of the four amino acids inhibited the mating progress, which related to fungal pathogenicity, but in *U. esculenta*, they had conspicuous promotion effect on mycelia formation (Table 2, Supplementary Fig. S2). Besides, histidine, the reactive oxygen species scavenger,[81] can also promote the yeast cells fusion in *U. esculenta*.

Overall, we speculated that *U. esculenta* developed a partial complementary mechanism to gain the host's nutrition while the hosts also utilize some nutrition metabolism pathways to defence infection. In addition, *in vitro* assays showed that His, Pro, Met, Lys and Phe, which host plant biosynthesized to respond fungus infection, were utilized by *U. esculenta* to promote its mating progress which was important for its successful infection, indicating that *U. esculenta* may change some amino acids response mechanism to regulate the balance of infection–resistance engineering for adapting the entophytic life, during 1000 yrs in *Z. latifolia*.

3.6. Fewer surface sensors and CSEPs benefit economic entophytic life of *U. esculenta*

The basic requirement for pathogens to infect host is to perceive physical and chemical stimuli when contacting with host surface and then induce to penetrate host either by high turgor pressure, like *M. oryzae* and *Colletotrichum spp.*,[82,83] or by secretion of plant cell wall-degrading enzymes (CWDEs), like *U. maydis* and *Cochliobolus carbonum*.[84,85] During the process, surface sensors, secreted proteins, CAZys and CWDEs are the indispensable factors.

Sho1 and Msb2 are conserved upstream proteins of MAP kinase cascade not only in the phytopathogenic fungi *Fusarium oxysporum*, *M. oryzae* and *U. maydis* but also in model fungus *S. cerevisiae*, and human fungal pathogen *C. albicans*, having been approved to recognize the surface signals, direct the central transcriptional network towards penetration.[22] It is intriguing that there is no *Sho1* orthologue in *U. esculenta*. We further investigated 61 genes (42 of which are secreted proteins) related with *Sho1/Msb2* and also involved in filament and appressoria formation.[22] Forty-two orthologues were found in *U. esculenta*, and more than 50% of them up-regulated in mycelium when compared with sporidial. Nineteen (17 of which are secreted protein genes) out of 61 genes were not found in *U. esculenta* (Supplementary File S8), including some important virulence factors, for example *pit2* (protease inhibitor in conjunction with Pit1 to maintain biotrophy during plant infection), *am1* (appressoria marker gene), *mig2-5*, *mig2-6*, *dkh6* (a 7TM protein, which might function in plant/pathogen interaction).[86]

The whole *mig2* gene cluster, a maize-specifically expressed cluster, is missing in *U. esculenta*. The *Ustilago scitaminea*, pathogen of sugar cane, appears to lack the *mig2* cluster too.[87] And interestingly, some *U. maydis* strains isolated from South America lost the *mig2-4* and were unable to trigger disease symptoms in regional maize variety.[87] The *mig2* cluster was related with host range,[88] which may partially explain why *U. esculenta* cannot infect maize and has a narrow host range. Lanver et al.[22] identified 139 plant surface cue-induced secreted proteins in *U. maydis*. *Ustilago esculenta* shares 112 these genes including 31 CWDEs and majority of *b*-induced genes (Supplementary File S9), loses 27 important secreted protein such as *exg1*,[89] *egl1* (cellulase gene expressed only in the filament),[90] *hum2* (the former hydrophobin gene in aerial hyphae formation)[91] and *tin2* (promotes virulence by targeting anthocyanin biosynthesis).[92]

Fungi produce a variety of CAZys to facilitate infection and/or gain nutrition. A total of 237 genes including auxiliary activities, glycoside hydrolases (GHs), polysaccharide lyases, carbohydrate esterases (CEs) and glycosyl transferases as well as carbohydrate-binding modules were identified in *U. esculenta* (Fig. 10). Those genes belonging to the fungal CAZy families are presumably involved in nutritional uptake and infection mechanisms. The *U. esculenta* genome contains 55 genes (including 9 CEs and 46 GHs) involved in plant cell wall-degradation, which was similar to *U. maydis*,[93] *U. hordei*[25] and *S. reilianum*,[26] but much fewer than hemibiotrophic and necrotrophic fungi such as *M. oryzae*[94] and *Botrytis cinerea*.[95] The compositions of CEs and GHs are similar between the four Ustilagonales species. There is also no significant change in the total number or composition of enzymes coding for the digestion of the three major plant cell wall polysaccharides classes: cellulose, hemicelluloses and pectin (Supplementary File S10).

Effector proteins, which are typically secreted by the pathogen after contacting with host, have been recognized to govern the interaction between plants and biotrophic pathogens. Totally, 633 proteins were identified as secreted proteins. Of these secreted proteins, 300 were predicted to be candidate secreted effector proteins (CSEPs) since they could not be assigned to enzymatic functions. Compared with *U. maydis* and *U. hordei*, the genome of *U. esculenta* has nearly 20% more predicted secreted proteins and a 10∼20% fewer predicted CSEPs (554 secreted proteins and 386 CSEPs for *U. maydis*. 515 secreted proteins and 333 CSEPs for *U. hordei*).[25] Ninety-two (15%) of 633 *U. esculenta* secreted proteins were species specific. Among them, 87 were CSEPs which were 29.6% of total predicted CSEPs. Four hundred and fifteen of 633 (66%) *U. esculenta* secreted proteins are shared among *U. maydis*, *U. hordei* and *S. reilianum* (Supplementary Table S5). Of them, 97 secreted proteins could be matched in pathogen–host interaction database (PHI-database)[96] with an amino acid identity >30%. Of all PHI hits, 55 out of total proteins were characterized as increased or reduced virulence. The remaining hits were either characterized as unaffected virulence or loss of pathogenicity.

There were 57 (11%) secreted proteins assigned to 14 clusters in *U. esculenta*, in comparison to 21, 7 and 22 clusters in *U. maydis*, in *U. hordei* and in *S. reilianum*. Seventy-six percent of Ue genes in clusters had orthologue Um genes, with amino acid identities ranging from 23.3 to 94.2%. Among the 14 clusters identified in this study, seven clusters were previously described Um effector gene clusters (Supplementary File S2), including all clusters whose deletion reduced virulence or non-pathogenic. Fungal secreted proteins are assumed to contribute to tumour formation,[90] and organ-specific effectors in both host and pathogen are required for host tumour formations.[24] In *U. esculenta*, the homologues with um3615 and um3616 secreted proteins were found, which were important genes of cluster 9A for tumour symptoms in adult tissue. When considering the responsible gene cluster 19A for tumour formation in *U. maydis* seedlings, only 4 out of 15 homologues of cluster 19A were present in *U. esculenta*. It may be the reason that *U. esculenta* could not trigger gall formation during seedlings. And more attention should be paid to find stem-specific gene clusters.

Besides, when we compared 92 CSEPs being important for successfully colonization of grass hosts and a core set of 248 CSEPs needed for pathogenicity (PSEPs) which were concluded from genome comparison among smut fungi *M. pennsylvanicum, U. maydis, U. hordei* and *S. reilianum*,[16] 55 grass host-related CSEPs and 182 PSEPs can find orthologues in *U. esculenta* (Supplementary Files S11 and S12). But it is remarkable that some important host-related CSEPs, including um03223 (maize-induced gene *mig1*), um12216 (related to Mig1 proteins) and 9 Ustilago-specific proteins, were not found in *U. esculenta* (Supplementary File S11). Furthermore, nearly 95% of *U. esculenta*-specific secreted proteins were CSEPs, which was one-third of total CSEPs. Missing of host-related CSEPs can further confirmed the narrow host range of *U. esculenta*, and species-specific CSEPs may give more hints in explanation. It has been proposed that effector genes, as speciation genes in fungal plant pathogens,[97] co-evolve with their host targets and reflect the host adaptation.[98]

All in all, although the *U. esculenta* shared core set of secreted proteins with some smut fungi indicating the close relation, the fungus has evolved distinguish characters of fewer surface sensors, high percentage of species-specific CSEPs and fewer CSEPs, lost some important virulence factors and host range-related effectors. All these characters may be the nature need of maintaining lower cost endophytic life and limitation of small genome size, which was the obliged evolution under long time pressure on keeping the mycelia phase and making no need to re-infect the host.

Acknowledgements

The authors thank Ying Rong for the preparation of the fungus collection, Cao Qianchao for mating assay, Gui Yijie for analysing assistant with host gene expression, Michael Feldbrügge for providing plasmids. They are grateful to Vera Göhre and Michael Feldbrügge for their critical comments on the article.

Accession number

EMBARGO: No

Funding

Our work was supported by funds from the National Natural Science Foundation of China (31470785, 31600634). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the article.

Supplementary data

Supplementary data are available at *DNARES* online.

Conflict of interest

None declared.

References

1. Chung, K.R. and Tzeng, D.D. 2004, Nutritional requirements of the edible gall-producing fungus *Ustilago esculenta*, *J. Biol. Sci.*, **4**, 246–52.
2. Xu, X., Walters, C., Antolin, M.F., et al. 2010, Phylogeny and biogeography of the eastern Asian-North American disjunct wild-rice genus (*Zizania L.*, Poaceae), *Mol. Phylogenet. Evol.*, **55**, 1008–17.
3. Guo, H.B., Li, S.M., Peng, J. and Ke, W.D. 2007, *Zizania latifolia* Turcz. cultivated in China, *Genet. Resour. Crop Evol.*, **54**, 1211–7.
4. Wang, Y., Huang, L. and Fan, L. 2013, Main agronomic traits, domestication and breeding of Gu (*Zizania latifolia*), *J. Zhejiang Univ.*, **39**, 629–35.
5. Yang, H.C. and Leu, L.S. 1978, Formation and histopathology of galls induced by *Ustilage esculenta* in *Zizania latifolia*, *Phytopathology*, **68**, 1572–6.
6. Qian, B., Luo, Y., Deng, Y., et al. 2012, Chemical composition, angiotensin-converting enzyme-inhibitory activity and antioxidant activities of few-flower wild rice (*Zizania latifolia* Turcz.), *J. Sci. Food Agric.*, **92**, 159–64.
7. Yan, N., Wang, X.Q., Xu, X.F., et al. 2013, Plant growth and photosynthetic performance of *Zizania latifolia* are altered by endophytic *Ustilago esculenta* infection, *Physiol. Mol. Plant Pathol.*, **83**, 75–83.
8. Dong, S.Z., Z.G., Yu, X.P., Fu, C.H. 2012, Highly efficient and safe production of *Jiaobai*, *Sci. Agric.*, **69**, 142–6.
9. Couch, B.C., Fudal, I., Lebrun, M.H., et al. 2005, Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice, *Genetics*, **170**, 613–30.

10. Stukenbrock, E., Banke, S., Javan-Nikkhah, M. and Mcdonald, B. 2007, Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation, *Mol. Biol. Evol*., **24**, 398–411.
11. Olsen, K.M. and Wendel, J.F. 2013, A bountiful harvest: genomic insights into crop domestication phenotypes, *Annu. Rev. Plant Biol*., **64**, 47–70.
12. Chen, Y.H., Gols, R. and Benrey, B. 2015, Crop domestication and its impact on naturally selected trophic interactions, *Annu. Rev. Entomol*., **60**, 35–58.
13. Stukenbrock, E.H., Bataillon, T., Dutheil, J.Y., et al. 2011, The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species, *Genome Res*., **21**, 2157–66.
14. Xue, M., Yang, J., Li, Z., et al. 2012, Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*, *PLoS Genet*., **8**, e1002869.
15. Stukenbrock, E.H. 2013, Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens, *New Phytol*., **199**, 895–907.
16. Sharma, R., Mishra, B., Runge, F. and Thines, M. 2014, Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the Smut Fungus *Melanopsichium pennsylvanicum*, *Genome Biol. Evol*., **6**, 2034–49.
17. Heinze, B. 2009, Comparative analysis of the Maize Smut Fungi *Ustilago maydis* and *Sporisorium reilianum*. Philipps-Universität Marburg, Germany.
18. Xu, X., Ke, W., Yu, X., Wen, J. and Ge, S. 2008, A preliminary study on population genetic structure and phylogeography of the wild and cultivated *Zizania latifolia* (Poaceae) based on Adh1a sequences, *Theor. Appl. Genet*., **116**, 835–43.
19. Lanver, D., Mendoza-Mendoza, A., Brachmann, A. and Kahmann, R. 2010, Sho1 and Msb2-related proteins regulate appressorium development in the smut fungus *Ustilago maydis*, *Plant Cell*, **22**, 2085–101.
20. Doehlemann, G., Berndt, P. and Hahn, M. 2006, Different signalling pathways involving a G alpha protein, cAMP and a MAP kinase control germination of *Botrytis cinerea* conidia, *Mol. Microbiol*., **59**, 821–35.
21. Doehlemann, G., Wahl, R., Horst, R.J., et al. 2008, Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*, *Plant J*., **56**, 181–95.
22. Lanver, D., Berndt, P., Tollot, M., et al. 2014, Plant surface cues prime *Ustilago maydis* for biotrophic development, *PLoS Pathog*., **10**, 295.
23. Brefort, T., Tanaka, S., Neidig, N., Doehlemann, G., Vincon, V. and Kahmann, R. 2014, Characterization of the largest effector gene cluster of *Ustilago maydis*, *PLoS Pathog*., **10**, 191–8.
24. Skibbe, D.S., Doehlemann, G., Fernandes, J. and Walbot, V. 2010, Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen, *Science*, **328**, 89–92.
25. Laurie, J.D., Ali, S., Linning, R., et al. 2012, Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements, *Plant Cell*, **24**, 1733–45.
26. Schirawski, J., Mannhaupt, G., Münch, K., et al. 2010, Pathogenicity determinants in smut fungi revealed by genome comparison, *Science*, **330**, 1546–8.
27. Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. and Pirovano, W. 2011, Scaffolding pre-assembled contigs using SSPACE, *Bioinformatics*, **27**, 578–9.
28. Boetzer, M. and Pirovano, W. 2012, Toward almost closed genomes with GapFiller, *Genome Biol*., **13**, 1–9.
29. Xu, Z. and Wang, H. 2007, LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons, *Nucleic Acids Res*., **35**, 265–8.
30. Stanke, M., Diekhans, M. and Baertsch, R., D. 2008, Using native and syntenically mapped cDNA alignments to improve de novo gene finding, *Bioinformatics*, **24**, 637–44.
31. Elsik, C.G., Mackey, A.J., Reese, J.T., Milshina, N.V., Roos, D.S. and Weinstock, G.M. 2007, Creating a honey bee consensus gene set, *Genome Biol*., **8**, R13.
32. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990, Basic local alignment search tool, *J. Mol. Biol*., **215**, 403–10.
33. Quevillon, E., Silventoinen, V., Pillai, S., et al. 2005, InterProScan: protein domains identifier, *Nucleic Acids Res*., **33**, 116–20.
34. Li, L., Stoeckert, C.J. and Roos, D.S. 2003, OrthoMCL: identification of ortholog groups for eukaryotic genomes, *Genome Res*., **13**, 2178–89.
35. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. 2012, dbCAN: a web resource for automated carbohydrate-active enzyme annotation, *Nucleic Acids Res*., **40**, 445–51.
36. Khaldi, N., Seifuddin, F.T., Turner, G., et al. 2010, SMURF: genomic mapping of fungal secondary metabolite clusters, *Fungal Genet. Biol*., **47**, 736–41.
37. Nielsen, H., Engelbrecht, J., Brunak, S. and Heijne, G.V. 1997, Identification of prokaryotic and eukaryotic signal peptides prediction of their cleavage sites, *Protein Eng*., **10**, 1–6.
38. Emanuelsson, O., Nielsen, H., Brunak, S. and Heijne, G.V. 2000, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, *J. Mol. Biol*., **300**, 1005–16.
39. Lowe, T.M. 1997, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res*., **25**, 955–64.
40. Nawrocki, E.P., Kolbe, D.L. and Eddy, S.R. 2009, Infernal 1.0: inference of RNA alignments, *Bioinformatics*, **25**, 1335–7.
41. Burge, S.W., Daub, J., Eberhardt, R., et al. 2013, Rfam 11.0: 10 years of RNA families, *Nucleic Acids Res*., **41**, 226–32.
42. Kurtz, S., Phillippy, A., Delcher, A.L., et al. 2004, Versatile and open software for comparing large genomes, *Genome Biol*., **5**, R12.
43. Li, R., Yu, C., Li, Y., et al. 2009, SOAP2: an improved ultrafast tool for short read alignment, *Bioinformatics*, **25**, 1966–7.
44. Mortazavi, A., Williams, B.A., Mccue, K., Schaeffer, L. and Wold, B. 2008, Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nat. Methods*, **5**, 621–8.
45. Robinson, M.D., Mccarthy, D.J. and Smyth, G.K. 2010, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, **26**, 139–40.
46. Guo, L., Qiu, J., Han, Z., et al. 2015, A host plant geneme (*Zizania latifolia*) after a century-long endophyte infection, *Plant J*., **83**, 600–9.
47. Sievers, F., Wilm, A., Dineen, D., et al. 2011, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol*., **7**, 1429–32.
48. Roy, S.W. and Gilbert, W. 2006, The evolution of spliceosomal introns: patterns, puzzles and progress, *Nat. Rev. Genet*., **7**, 211–21.
49. Laurie, J.D., Linning, R., Wong, P. and Bakkeren, G. 2013, Do TE activity and counteracting genome defenses, RNAi and methylation, shape the sex lives of smut fungi?, *Plant Signal. Behav*., **8**, e23853.
50. Selker, E.U. 1990, Premeiotic instability of repeated sequences in *Neurospora crassa*, *Annu. Rev. Genet*., **24**, 579–613.
51. Vogel, K.J. and Moran, N.A. 2013, Functional and evolutionary analysis of the genome of an obligate fungal symbiont, *Genome Biol. Evol*., **5**, 891–904.
52. Kellner, R., Vollmeister, E., Feldbrügge, M. and Begerow, D. 2011, Interspecific sex in grass smuts and the genetic diversity of their pheromone-receptor system, *PLoS Genet*., **7**, e1002436.
53. Klaus, B., Lengeler, D.S.F., Fraser, J.A., et al. 2002, Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes, *Eukaryot. Cell*, **1**, 704–18.
54. Hood, M.E., Antonovics, J. and Koskella, B. 2004, Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms, *Genetics*, **168**, 141–6.
55. Gordon, J.L., Armisén, D., Proux-Wéra, E., Ó S.S., Byrne, K.P. and Wolfe, K.H. 2011, Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents, *Proc. Natl. Acad. Sci. USA*., **108**, 20024–9.
56. Tollis, M. and Boissinot, S. 2012, The evolutionary dynamics of transposable elements in eukaryote genomes, *Genome Dyn*., **7**, 68–91.
57. Ochogavía, A.C., Seijo, J.G., González, A.M., et al. 2011, Characterization of retrotransposon sequences expressed in inflorescences

of apomictic and sexual *Paspalum notatum* plants, *Sex. Plant Reprod.*, **24**, 231–46.
58. Lokody, I. 2014, Evolution: transposons drive sex chromosome evolution, *Nat. Rev. Genet.*, **15**, 1.
59. Fedler, M., Luh, K.S., Stelter, K., Fernanda, N.J., Basse, C.W. 2009, The *a2* mating-type locus genes *lga2* and *rga2* direct uniparental mitochondrial DNA (mtDNA) inheritance and constrain mtDNA recombination during sexual development of *Ustilago maydis*, *Genetics*, **181**, 847–60.
60. Bortfeld, M., Auffarth, K., Kahmann, R. and Basse, C.W. 2004, The *Ustilago maydis a2* mating-type locus genes *lga2* and *rga2* compromise pathogenicity in the absence of the mitochondrial p32 family protein Mrb1, *Plant Cell*, **16**, 2233–48.
61. Breton, S., Stewart, D.T. and Bonen, L. 2015, Atypical mitochondrial inheritance patterns in eukaryotes, *Genome*, **58**, 423–31.
62. Yan, Z., Sun, S., Shahid, M. and Xu, J. 2007, Environment factors can influence mitochondrial inheritance in the fungus *Cryptococcus neoformans*, *Fungal Genet. Biol.*, **44**, 315–22.
63. Heimel, K., Scherer, M., Vranes, M., et al. 2010, The transcription factor Rbf1 is the master regulator for *b*-mating type controlled pathogenic development in *Ustilago maydis*, *PLoS Pathog.*, **6**, 761–4.
64. Brefort, T., Doehlemann, G., Mendoza-Mendoza, A., Reissmann, S., Djamei, A. and Kahmann, R. 2009, *Ustilago maydis* as a pathogen, *Phytopathology*, **47**, 423–45.
65. Feldbrügge, M. and Al, E. 2004, Regulation of mating and pathogenic development in *Ustilago maydis*, *Curr. Opin. Microbiol.*, **7**, 666–72.
66. Zarnack, K., Eichhorn, H., Kahmann, R. and Feldbrügge, M. 2008, Pheromone-regulated target genes respond differentially to MAPK phosphorylation of transcription factor Prf1, *Mol. Microbiol.*, **69**, 1041–53.
67. Heimel, K., Scherer, M., Schuler, D. and Kämper, J. 2010, The *Ustilago maydis* Clp1 protein orchestrates pheromone and *b*-dependent signaling pathways to coordinate the cell cycle and pathogenic development, *Plant Cell*, **22**, 2908–22.
68. Helfer, H. and Gladfelter, A.S. 2006, AgSwe1p regulates mitosis in response to morphogenesis and nutrients in multinucleated *Ashbya gossypii* cells, *Mol. Biol. Cell*, **17**, 4494–512.
69. Mccann, M.P. and Snetselaar, K.M. 2008, A genome-based analysis of amino acid metabolism in the biotrophic plant pathogen *Ustilago maydis*, *Fungal Genet. Biol.*, **45**(**Suppl 1**), S77–87.
70. Hansen, A.K. and Moran, N.A. 2011, Aphid genome expression reveals host-symbiont cooperation in the production of amino acids, *Proc. Natl. Acad. Sci. USA.*, **108**, 2849–54.
71. Hornby, J.M., Jacobitzkizzier, S.M., Mcneel, D.J., Jensen, E.C., Treves, D.S. and Nickerson, K.W. 2004, Inoculum size effect in dimorphic fungi: extracellular control of yeast-mycelium dimorphism in *Ceratocystis ulmi*, *Appl. Environ. Microbiol.*, **70**, 1356–9.
72. Jiménezlópez, C., Collette, J.R., Brothers, K.M., et al. 2013, *Candida albicans* induces arginine biosynthetic genes in response to host-derived reactive oxygen species, *Eukaryot. Cell*, **12**, 91–100.
73. Davis, R.H. 1986, Compartmental and regulatory mechanisms in the arginine pathways of *Neurospora crassa* and *Saccharomyces cerevisiae*, *Microbiol. Rev.*, **50**, 280–313.
74. Senthilkumar, M. and Mysore, K.S. 2012, Ornithine-delta-aminotransferase and proline dehydrogenase genes play a role in non-host disease resistance by regulating pyrroline-5-carboxylate metabolism-induced hypersensitive response, *Plant Cell Environ.*, **35**, 1329–43.
75. Zhu, X.L., Qi, L., Liu, X., et al. 2014, The wheat ethylene response factor transcription factor pathogen-induced ERF1 mediates host responses to both the necrotrophic pathogen *Rhizoctonia cerealis* and freezing stresses, *Plant. Physiol.*, **164**, 1499–514.
76. Boubakri, H., Wahab, M.A., Chong, J., et al. 2013, Methionine elicits H_2O_2 generation and defense gene expression in grapevine and reduces *Plasmopara viticola* infection, *J. Plant Physiol.*, **170**, 1561–8.
77. Arasimowicz-Jelonek, M., Floryszak-Wieczorek, J., Gzyl, J. and Chmielowska-Bąk, J. 2013, Homocysteine over-accumulation as the effect of potato leaves exposure to biotic stress, *Plant Physiol. Biochem.*, **63**, 177–84.
78. Yang, H. and Ludewig, U. 2014, Lysine catabolism, amino acid transport, and systemic acquired resistance: what is the link?, *Plant Signal. Behav.*, **9**, e28933.
79. Kim, D.S. and Hwang, B.K. 2014, An important role of the pepper phenylalanine ammonia-lyase gene (PAL1) in salicylic acid-dependent signalling of the defence response to microbial pathogens, *J. Exp. Bot.*, **65**, 2295–306.
80. Kurth, F., Mailänder, S., Bönn, M., et al. 2014, Streptomyces-induced resistance against oak powdery mildew involves host plant responses in defense, photosynthesis, and secondary metabolism pathways, *Mol. Plant Microbe Interact.*, **27**, 891–900.
81. Hara, M., Kondo, M. and Kato, T. 2013, A KS-type dehydrin and its related domains reduce Cu-promoted radical generation and the histidine residues contribute to the radical-reducing activities, *J. Exp. Bot.*, **64**, 1615–24.
82. Bechinger, C., Giebel, K.F., Schnell, M., Leiderer, P., Deising, H.B. and Bastmeyer, M. 1999, Optical measurements of invasive forces exerted by appressoria of a plant pathogenic fungus, *Science*, **285**, 1896–9.
83. Howard, R.J., Ferrari, M.A., Roach, D.H. and Money, N.P. 1991, Penetration of hard substrates by a fungus employing enormous turgor pressures, *Proc. Natl. Acad. Sci. USA.*, **88**, 11281–4.
84. Schirawski, J., Böhnert, H.U., Steinberg, G., Snetselaar, K., Adamikowa, L. and Kahmann, R. 2005, Endoplasmic reticulum glucosidase II is required for pathogenicity of *Ustilago maydis*, *Plant Cell*, **17**, 3532–43.
85. Tonukari, N.J., Scott-Craig, J.S. and Walton, J.D. 2000, The *Cochliobolus carbonum SNF1* gene is required for cell wall-degrading enzyme expression and virulence on maize, *Plant Cell*, **12**, 237–47.
86. Kulkarni, R.D., Thon, M.R., Pan, H. and Dean, R.A. 2005, Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*, *Genome Biol.*, **6**, R24.
87. Basse, C.W., Kolb, S. and Kahmann, R. 2002, A maize-specifically expressed gene cluster in *Ustilago maydis*, *Mol. Microbiol.*, **43**, 75–93.
88. Farfsing, J.W., Auffarth, K. and Basse, C.W. 2005, Identification of cis-active elements in *Ustilago maydis* mig2 promoters conferring high-level activity during pathogenic growth in maize, *Mol. Plant Microbe Interact.*, **18**, 75–87.
89. Schaeffer, H.J., Leykam, J. and Walton, J.D. 1994, Cloning and targeted gene disruption of EXG1, encoding exo-beta 1, 3-glucanase, in the phytopathogenic fungus *Cochliobolus carbonum*, *Appl. Environ. Microbiol.*, **60**, 594–8.
90. Schauwecker, F., Wanner, G. and Kahmann, R. 1995, Filament-specific expression of a cellulase gene in the dimorphic fungus *Ustilago maydis*, *Biol. Chem. Hoppe Seyler*, **376**, 617–25.
91. Teertstra, W.R., Deelstra, H.J., Vranes, M., et al. 2006, Repellents have functionally replaced hydrophobins in mediating attachment to a hydrophobic surface and in formation of hydrophobic aerial hyphae in *Ustilago maydis*, *Microbiology*, **152**, 3607–12.
92. Tanaka, S., Brefort, T., Neidig, N., et al. 2014, A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize, *Elife*, **3**, 125–31.
93. Kämper, J., Kahmann, R., Bölker, M., et al. 2006, Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*, *Nature*, **444**, 97–101.
94. Quoc, N.B. and Chau, N.N. 2016, The role of cell wall degrading enzymes in pathogenesis of *Magnaporthe oryzae*, *Curr. Protein Pept. Sci.*, **17**, [Epub ahead of print].
95. Blanco-Ulate, B., Morales-Cruz, A., Amrine, K.C., Labavitch, J.M., Powell, A.L. and Cantu, D. 2013, Genome-wide transcriptional profiling of *Botrytis cinerea* genes targeting plant cell walls during infections of different hosts, *Front. Plant Sci.*, **5**, 435.
96. Winnenburg, R., Urban, M., Beacham, A., et al. 2008, PHI-base update: additions to the pathogen host interaction database, *Nucleic Acids Res.*, **36**, 572–6.
97. Giraud, T., Gladieux, P. and Gavrilets, S. 2010, Linking emerging fungal plant diseases and ecological speciation, *Trends Ecol. Evol.*, **25**, 387–95.
98. Stukenbrock, E.H. and Mcdonald, B.A. 2009, Population genetics of fungal and Oomycete effectors involved in gene-for-gene interactions, *Mol. Plant Microbe Interact.*, **22**, 371–80.