

Genetic Diversity and Population Structure of Rice Pathogen *Ustilaginoidea virens* in China

Xianyun Sun¹✉, Shu Kang¹✉, Yongjie Zhang²✉, Xinqiu Tan³, Yufei Yu¹, Haiyong He⁴, Xinyu Zhang¹, Yongfeng Liu⁵, Shu Wang⁶, Wenxian Sun⁷, Lei Cai¹, Shaojie Li^{1*}

1 State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, P. R. China, **2** School of Life Sciences, Shanxi University, Taiyuan, P. R. China, **3** Institute of Plant Protection, Hunan Academy of Agricultural Science, Changsha, P. R. China, **4** Guizhou Institute of Plant Protection, Guiyang, P. R. China, **5** Institute of Plant Protection, Jiangsu Academy of Agricultural Science, Nanjing, P. R. China, **6** Institute of Plant Protection, Liaoning Academy of Agricultural Science, Shenyang, P. R. China, **7** Department of Plant Pathology, China Agricultural University, Beijing, P. R. China

Abstract

Rice false smut caused by the fungal pathogen *Ustilaginoidea virens* is becoming a destructive disease throughout major rice-growing countries. Information about its genetic diversity and population structure is essential for rice breeding and efficient control of the disease. This study compared the genome sequences of two *U. virens* isolates. Three SNP-rich genomic regions were identified as molecular markers that could be used to analyze the genetic diversity and population structure of *U. virens* in China. A total of 56 multilocus sequence types (haplotypes) were identified out of 162 representative isolates from 15 provinces covering five major rice-growing areas in China. However, the phylogeny, based on sequences at individual SNP-rich regions, strongly conflicted with each other and there were significant genetic differences between different geographical populations. Gene flow between the different geographical populations and genetic differentiation within each geographical population were also detected. In addition, genetic recombination and genetic isolation resulting from geographic separation was also found.

Citation: Sun X, Kang S, Zhang Y, Tan X, Yu Y, et al. (2013) Genetic Diversity and Population Structure of Rice Pathogen *Ustilaginoidea virens* in China. PLoS ONE 8(9): e76879. doi:10.1371/journal.pone.0076879

Editor: Vishnu Chaturvedi, California Department of Public Health, United States of America

Received: May 27, 2013; **Accepted:** August 28, 2013; **Published:** September 30, 2013

Copyright: © 2013 Sun et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This project is supported by the Special Fund for Agro-scientific Research in the Public Interest of China (No. 200903039-8) and a "100 Talent Program" grant from Chinese Academy of Sciences (to Shaojie Li). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: shaojieli@gmail.com

✉ These authors contributed equally to this work.

Introduction

False smut (green smut) is one of the most destructive diseases of rice. It is caused by the ascomycete fungus *Ustilaginoidea virens* (Cooke) Takah (Anamorph). *U. virens* was recently placed in Clavicipitaceae as the new name *Villosiclava virens* (Teleomorph) [1,2], which can reproduce both sexually and asexually with multiple propagules [3,4]. It was found that *U. virens* could invade rice coleoptiles and roots at the young seedling stage and the stamen filaments at the earlier booting stage [5-7]. However, it is not clear if *U. virens* mycelia, that colonized in young seedlings, can extend to the spikelets and form the ball-like symptom. At rice booting stage, *U. virens* invades rice through small gap at the apex of spikelet before heading [6] but does not penetrate host cell walls [7]. The characteristic symptom of rice false smut is the formation of ball-like colonies in spikelets. Sclerotia and chlamydo spores are commonly found on the surface of and inside smut balls,

respectively. In addition to yield loss, *U. virens* also contaminates rice seed and straw by producing antimetabolic cyclic peptides that are poisonous to both humans and animals [8,9]. False smut of rice was previously regarded as a minor disease that occurred sporadically in certain regions, but in recent years, the disease has become significant throughout major rice-growing countries [10-12]. The use of nitrogen fertilizers and large-scale planting of hybrid cultivars have been regarded as responsible for the increased disease severity of false smut in China [13].

Previous studies on rice false smut mainly focused on its occurrence, pathogen detection, toxin analysis, pathogen life cycle and disease control [11,14-19]. However, little is known about the genetic diversity of *U. virens* populations. Information on the genetic diversity of *U. virens* populations will help create efficient strategies to reduce the yield loss caused by this pathogen. Zhou et al. [20] analyzed the genetic diversity of *U. virens* populations from two provinces in North China using



Figure 1. Geographical distributions of *U. virens* samples in China. A total of 162 *U. virens* isolates from 15 provinces, covering the five major rice-growing areas of China, were used in this study. Each rice-growing area is marked with a unique color on the map.

doi: 10.1371/journal.pone.0076879.g001

amplified fragment length polymorphism (AFLP) markers and discovered that the genetic variation in *U. virens* populations was small in these areas. However, the samples used for the analysis were collected from only two provinces. The genetic diversity and population structure of *U. virens* in the majority of the rice-growing areas across China are unknown.

Methods based on molecular markers have been developed in order to analyze genetic diversity in fungal species. These markers include: allozyme, AFLP and microsatellites. DNA sequence-based markers have been widely used over recent years due to the reduced cost of DNA sequencing. Since an individual gene contains limited information, multi-locus sequence typing (MLST) has become a mainstream method for both molecular systematics and population genetics analyses [21,22]. MLST has been demonstrated to be highly discriminatory when analyzing the genetic diversities of many fungal species that are pathogenic to humans, such as: *Histoplasma capsulatum* [23], *Candida albicans* [24], *Aspergillus fumigatus* [25] and *Cryptococcus neoformans* [26]. Most of the DNA sequences used for MLST in these studies were limited to housekeeping genes. For plant pathogenic fungi, however, reports on genetic diversity, based on MLST, only found in *Magnaporthe grisea* [27].

Chlamydospores of *U. virens* can persist on seeds or in soil for several years [28]. The accumulation of chlamydospores in field or movement of chlamydospores to other areas can increase disease severity. Analysis of genetic diversity in *U. virens* isolates from China on a large geographical scale will improve the design of efficient management strategies for rice false smut. In this study, three SNP-rich genomic regions were identified as novel molecular markers that could be used to study the population structure of *U. virens*. The genetic diversity and population structure of *U. virens* in China were examined by comparing sequences located in SNP-rich genomic regions of 162 *U. virens* isolates collected from five major rice-growing areas in China.

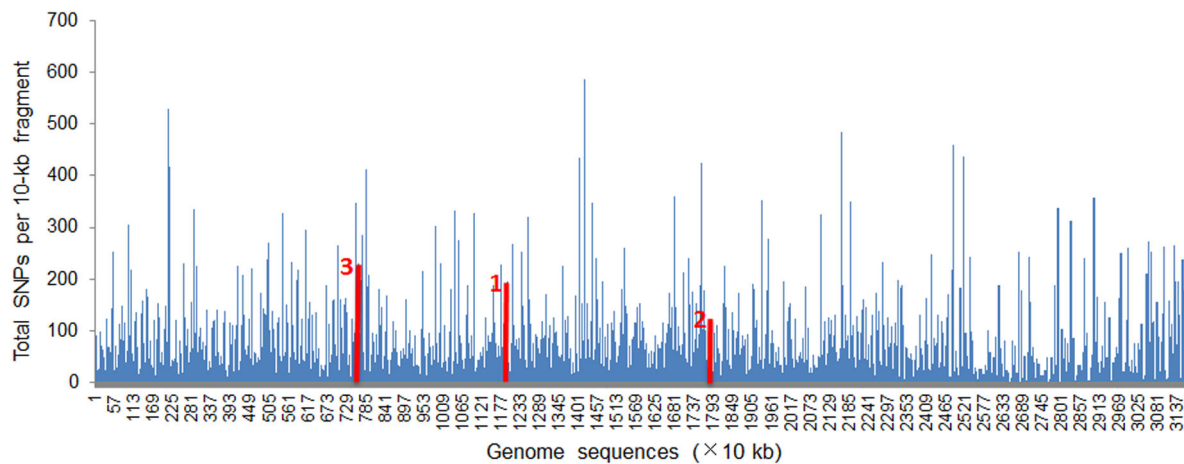
Materials and Methods

Ethics statement: No specific permissions were required for sample collection on the locations mentioned in this manuscript and the fungus used in this study is not an endangered or protected species.

Table 1. Primer pairs used for the SNP marker assays in this study and their nucleotide variations.

Fragment	Primers sequence (5'→3')	T _m (°C)	Size of PCR products	Aligned length	No. of SNP sites	No. of alleles
F1	F: GGTCGGATACTCGGTGCC	58.0	853 bp	775 bp	30	16
	R: CGCTTAGGGCATCTTTAC	55.8				
F2	F: GGTTCCGCTAGGGGCGATTG	66.3	776 bp	712 bp	25	9
	R: TGACGGGGCGTAGTAAGTTT	61.4				
F3	F: TTGGCGAGGAGATCAGGGTG	66.7	668 bp	594 bp	70	7
	R: TGCTGGTGGGAGCGTTGA	65.2				

doi: 10.1371/journal.pone.0076879.t001

**Figure 2.** SNP sites comparison between the genome sequences of two *U. virens* isolates. The SNP numbers per 10 kb DNA were compared using two genomes sequences and Biolign software version 4.0.6 (Hall 2001). Locations of the three SNP-rich markers are indicated by red lines.

doi: 10.1371/journal.pone.0076879.g002

Collection and isolation of *U. virens* isolates

All isolates were obtained from infected rice spikelets by single chlamydospore isolation. The *U. virens* isolates were recovered from five rice cropping regions (Figure 1; Table S1), based on rice cropping regionalization in China [29], including 32 isolates from Sichuan-Shaanxi basin (Sichuan and South Shaanxi) where single-cropping hybrid rice is cultivated, 28 isolates from South China (Guangdong, Guangxi and Fujian) where double-cropping hybrid rice is cultivated, 66 isolates from Central China rice-growing areas (Hubei, Hunan, South Henan, South Jiangsu, South Anhui, Jiangxi and Zhejiang) where both single-cropping and double-cropping systems are applied and the cultivars used include hybrid rice and japonica rice, 20 isolates from the Yunnan-Guizhou Plateau (Yunnan and Guizhou) which is a single- and double-cropping rice hybrid area and the main cultivar is hybrid rice but japonica rice is also cultivated on a small scale, and 16 isolates from the Northeast China rice-growing area (Liaoning) where single cropping japonica rice is cultivated.

DNA extraction

Each isolate was cultured in 20 ml liquid potato-dextrose broth in a 100 ml conical flask at 26°C and was shaken at 200 rpm. After 7 days, the culture was transferred to 50 ml liquid potato-dextrose broth, in a 250 ml conical flask, and incubated under the same conditions for 3 days. The mycelia were harvested and immediately ground into a fine powder in liquid nitrogen for DNA extraction following the protocol described by Nakada et al. [30]. The DNA was precipitated with ethanol after two phenol/chloroform (1:1) extractions and one chloroform extraction. All isolates were confirmed as *U. virens* by PCR amplification using a pair of *U. virens* specific primers: US1-5/US3-3, which target the nrDNA ITS region [14].

Molecular marker determination

In order to identify the genomic regions where SNPs are highly enriched, the genome sequences (unpublished data) of the two *U. virens* isolates: cce007 and csc086, were comparatively analyzed by Biolign software version 4.0.6 [31]. Isolate cce007 was collected from Hubei province of Central China rice-growing areas while isolate csc086 was collected from Sichuan province of Sichuan-shaanxi basin rice-growing

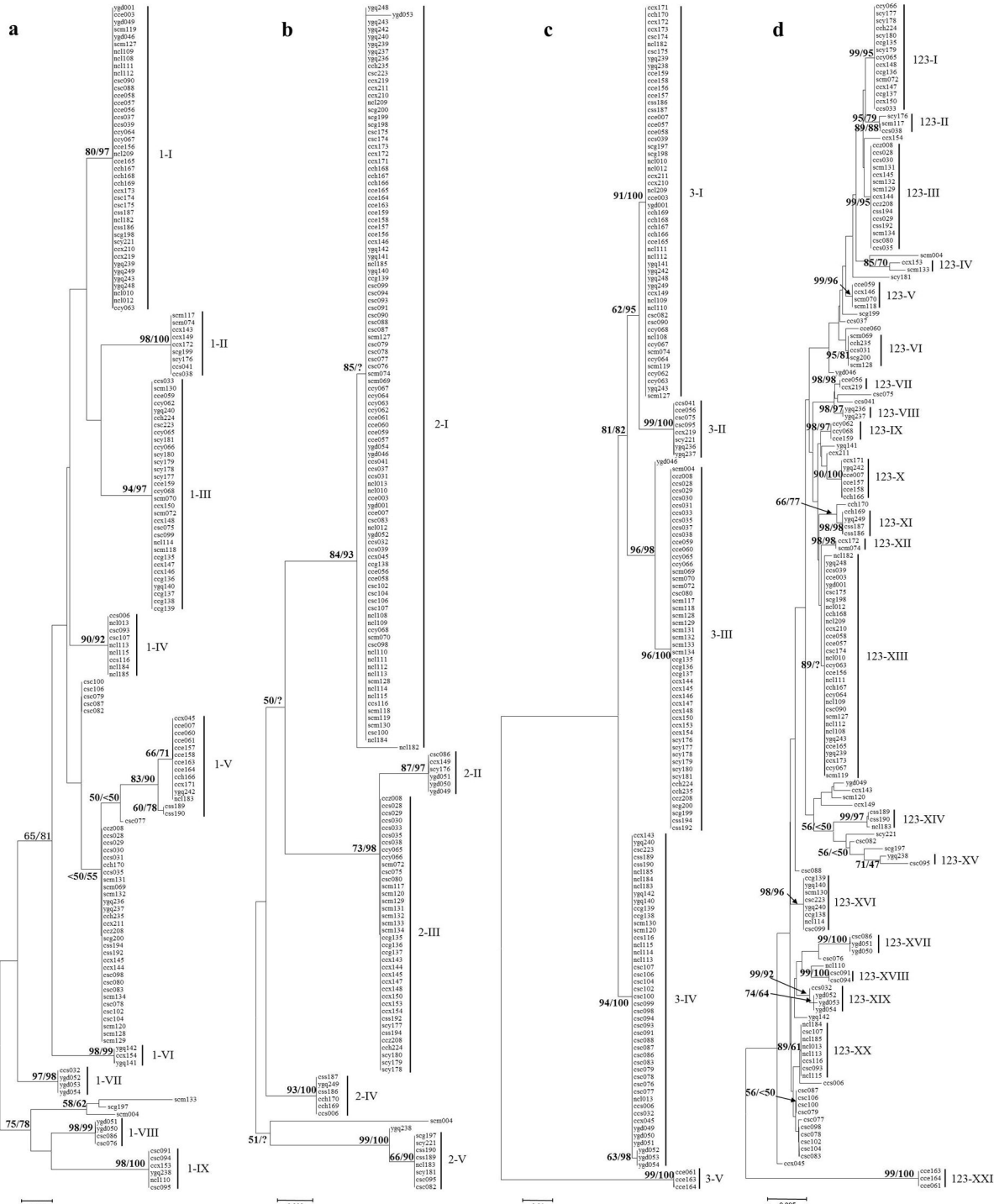


Figure 3. Phylogenetic analyses based on sequences at individual marker (a–c) or the combined sequences at three markers (d). NJ and ML phylogenetic analyses were conducted based on DNA sequences at Marker 1(a), Marker 2 (b), Marker 3 (c) and the combined sequences at the three markers (d). The two methods gave almost identical topologies. Nodal support values for NJ (before the slash) and ML analysis (after the slash) are given for branches receiving more than 70% support. Clades were named with marker names and Roman numbers.

doi: 10.1371/journal.pone.0076879.g003

Table 2. Genetic variation patterns within populations.

Population	No. of haplotypes (Frequency of the most common haplotype)		Nucleotide diversity, Pi ($\times 10^{-2}$)	Haplotype diversity/ I_A	r_d	PcP	
	No. of samples						
Sichuan-Shaanxi basin	32	16 (5)	0.75	0.95	0.44** (0.27*)	0.23** (0.15*)	0.00
Yunnan-Guizhou Plateau	20	13 (4)	0.79	0.95	0.19* (-0.11)	0.09* (-0.06)	0.66
South China	28	16 (5)	0.82	0.94	0.25** (-0.05)	0.13** (-0.03)	0.00
Central China	66	26 (13)	0.98	0.92	0.50** (-0.03)	0.26** (-0.02)	0.00
Northeast China	16	6 (7)	0.49	0.74	0.56** (-0.05)	0.28** (-0.03)	1.00
Total	162	56 (30)	0.93	0.94	0.38** (-0.02)	0.19** (-0.01)	0.00

Note: I_A , index of association; r_d , scaled index of association (I_A) by the number of loci ($m-1$) and PcP, percentage of phylogenetically compatible pairs of loci. I_A and r_d values were estimated for both uncorrected (outside parentheses) and clone-corrected (within parentheses) data sets. * $P < 0.05$; ** $P < 0.01$.

doi: 10.1371/journal.pone.0076879.t002

Table 3. Pairwise F_{ST} values between different populations.

Population	Sichuan-Shaanxi basin	Yunnan-Guizhou Plateau	South China	Central China
Yunnan-Guizhou Plateau	0.02			
South China	0.24***	0.24***		
Central China	0.12***	0.09***	0.06**	
Northeast China	0.08*	0.02	0.34***	0.13***

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

doi: 10.1371/journal.pone.0076879.t003

areas. Since these two rice-growing areas are different in topography, climate, cropping regionalization and rice cultivars, genetic difference might exist between these two isolates. PCR primers, which were used to amplify 600–1000 bp DNA fragments in selected SNP-rich loci, were then designed using the online software, Primer3 (<http://frodo.wi.mit.edu/primer3/>). The SNP sites of each selected genomic locus were identified by aligning the sequences at each selected locus in 10 isolates from 10 different provinces using the DNAMAN software (Lynnon Corporation, Quebec, Canada). The putative functions of SNP-rich DNA fragments were deduced from BLAST alignment against sequences in GenBank (<http://www.ncbi.nlm.nih.gov/>) and gene prediction using GlimmerHMM software [32].

Amplification and sequencing of DNA fragments

SNP-rich DNA fragments were amplified using specific primers (Table 1). Each reaction mixture was composed of a 1× PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 200 μ M dNTP, 0.4 μ M of each primer and 1 U of *Pfu* DNA polymerase (TaKaRa). The PCR was performed in a thermocycler (Labnet, TC9600; NJ, USA) under the following conditions: denaturation at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 20 s, annealing at 56°C for 30 s and elongation at 72°C for 1 min. The PCR products were isolated by electrophoresis and purified for bi-directional sequencing.

Phylogenetic analysis

DNA sequences were aligned using Clustal X 2.1 software [33]. The resulting alignments were manually adjusted where necessary. Neighbor-joining (NJ) phylogenetic analysis was performed with the Kimura 2-parameter model using MEGA 5 software [34]. Maximum likelihood (ML) analysis was performed with the GTR+CAT model using the version RAxML v7.2.8 [35]. Bootstrap tests were based on 500 re-samplings and midpoint rooting was used for every data set. A bootstrap portion greater than 70% was considered significant. In addition, we tested for heterogeneity between the three markers using the Incongruence Length Difference test, implemented as the partition homogeneity test in PAUP (Sinauer Associates).

Population genetics analysis

Population genetics analysis was performed within and between populations (i.e. the above mentioned five rice cropping regions) using concatenated DNA sequence data set at three genomic loci. For within-population genetic analysis, we calculated the haplotype diversity and analyzed linkage disequilibrium (1,000 randomizations) using MultiLocus 1.3 [36]. Evidence of recombination was examined by calculating the index of associations (I_A and r_d) and by analyzing the allelic combinations. The number of haplotypes and nucleotide diversity were calculated using DnaSP software (version 5.10) [37]. For cross-population analysis, F_{ST} values were calculated using Arlequin 3.5.1.2 [38], based on variable nucleotides from the concatenated data set. Sources of genetic differentiation were investigated using analysis of molecular variance

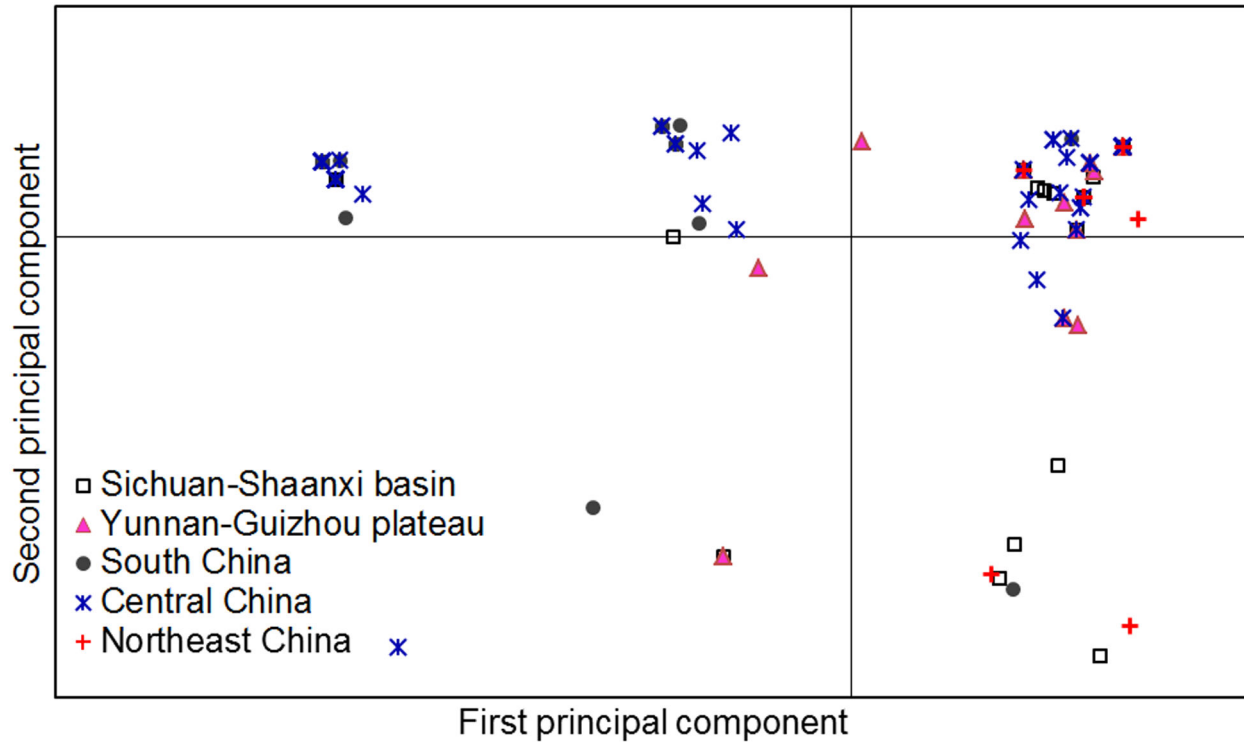


Figure 4. Principal component analysis (PCA) for 162 individual isolates from five major rice-growing areas in China. Individuals within the same population are marked using the same symbols. The first and second principal coordinates account for 41.6% and 23.0% of the variation, respectively. Note that individuals from a population may have separated and gathered with individuals from another population.

doi: 10.1371/journal.pone.0076879.g004

(AMOVA) tests, which were undertaken using GenALEx 6.5 [39]. Population structure was inferred using the Structure 2.3 program [40,41] with the default settings based on the concatenated data set and the result was visualized using Distruct 1.1 [42]. Principal coordinate analysis (PCA) was performed with individual isolates using GenALEx. The Mantel test was performed using GenALEx in order to examine the level of genetic isolation by geographical distances.

Results

Identification of SNP-rich genomic DNA markers

In order to analyze the genetic diversity and population structure of *U. virens*, we explored new DNA markers that contained multiple SNP sites. First, two *U. virens* isolates (cce007 and csc086) were chosen to do comparative analysis between their genome sequences. The assembled genome of cce007 is 38.8 Mb with a contig N50 length of 29.3 kb and the assembled genome of csc086 is 35.5 Mb with a contig N50 length of 27 kb (data not shown).

By comparing the genome sequences of two *U. virens* isolates, a total of 149669 SNP sites were found within 31.67 Mb matched genomic DNA between two isolates. As shown in Figure 2, the SNP sites were highly enriched in some regions. Ten genomic regions, each of which was approximately 700 bp

in size and contained more than 10 SNP sites, were then randomly selected. The PCR products of each selected locus, from 10 isolates collected from 10 different provinces, were then sequenced and compared by alignment. Six genomic regions, from which PCR products were successfully amplified in all 10 isolates, were selected to generate PCR products in 162 isolates. As a result, only three genomic regions, that contain dense SNP sites and could be easily amplified by PCR in all isolates, were selected as DNA markers in order to analyze the genetic diversity and population structure of *U. virens*. For the other genomic regions, their PCR products could be obtained in some isolates but not in other isolates. The presence of SNP sites at the primer regions in some isolates might affect the efficacy of PCR application.

These DNA markers were named as Marker 1 (775 bp with 30 SNPs), Marker 2 (712 bp with 25 SNPs), and Marker 3 (594 bp with 70 SNPs) (Table 1). Gene prediction indicated that Marker 1 was a coding sequence without any introns, Marker 2 was a non-coding region and Marker 3 covered a coding region (50 bp at the 3' end) and a non-coding region (544 bp at the 5' end). A BLASTp search showed that the protein encoded by Marker 1 was homologous to a hypothetical protein (MGG_13116) from *Magnaporthe oryzae*. The nucleotide sequences of the three markers from the representative isolate, cce003, were deposited in GenBank under the following

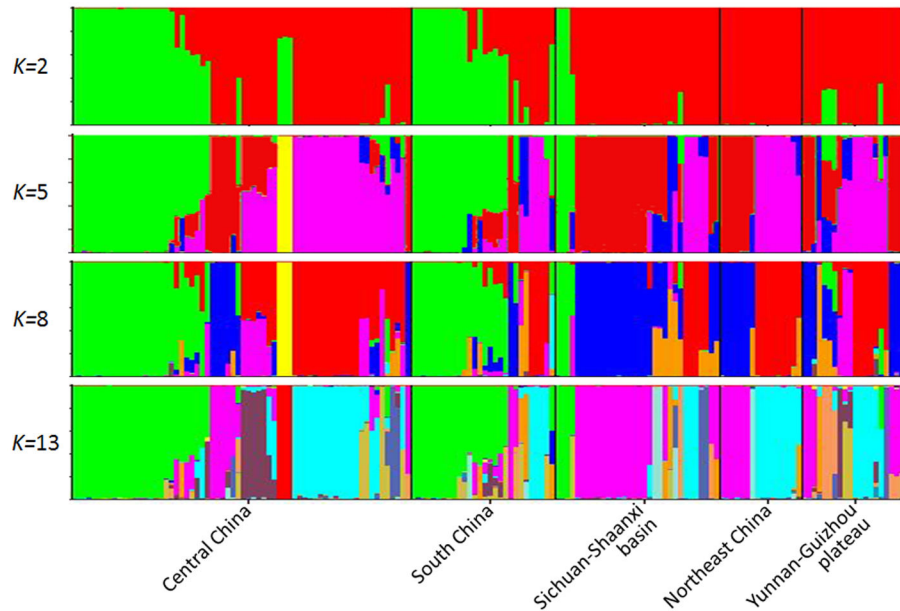


Figure 5. Population structure of *U. virens* inferred from 125 SNPs of 162 isolates using the program Structure (version 2.3.1), with different *K* values. An admixture model with correlated allele frequencies was used. Each isolate is represented by a single vertical line broken into *K* colored segments, with lengths proportional to each of the *K* inferred clusters.

doi: 10.1371/journal.pone.0076879.g005

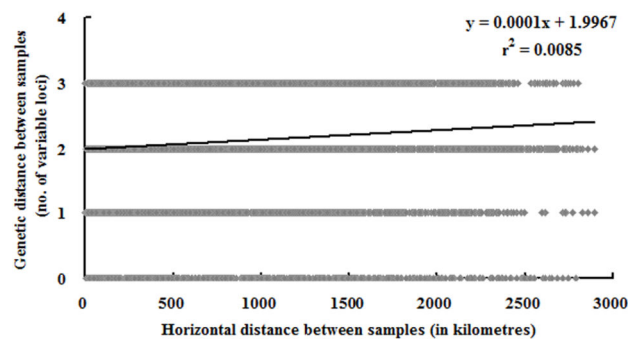


Figure 6. Mantel test between genetics and horizontal distance. Horizontal geographical distances means distances based on longitudinal–latitudinal coordinates.

doi: 10.1371/journal.pone.0076879.g006

accession numbers: JX483852 for Marker 1, JX483853 for Marker 2 and JX483854 for Marker 3. Sequence variations in these markers among the *U. virens* isolates are shown in Table S2.

Nucleotide variations of SNP-rich markers among *U. virens* isolates

A total of 162 *U. virens* isolates from 15 provinces, covering the five major rice-growing areas of China, were used to investigate the genetic diversity and population structure of the fungus (Figure 1; Table S1). The PCR products targeting the three SNP-rich DNA markers were successfully amplified from 162 isolates. Sequence analysis of the PCR products identified 30, 25 and 70 SNP sites and 16, 9 and 7 sequence types at

Markers 1, 2 and 3, respectively. These SNP sites, 125 in total, included 118 parsimony informative sites and seven singleton variable sites. The parsimony informative sites included one insertion/deletion, 95 transitional substitutions, 19 transversal substitutions and three A/C/G substitutions (Table S2).

Haplotypes within the 162 *U. virens* isolates

Analyses of the combined sequences at the SNP-rich DNA regions in the 162 isolates by software DnaSP version 5.10 [37] identified a total of 56 multilocus sequence types (haplotypes). Thirty-five haplotypes were represented by only one isolate and the remaining 21 haplotypes had at least two isolates (Table S2). Haplotypes 1, 10 and 11 were dominant

and contained 30 (18.5% of total number of isolates), 15 (9.3%), and 15 (9.3%) isolates, respectively (Table S2).

Seven haplotypes, each of which contained at least two isolates, were only found in one area. For instance, all five isolates of haplotype 42 were collected from the Sichuan-Shaanxi basin rice-growing areas, and all three isolates of haplotype 6 were collected from Central China (Table S2). However, 12 haplotypes were shared by three or more areas. Haplotype 1 and haplotype 17 were found in all five areas.

Phylogenetic analysis

Three phylogenetic trees were built using the DNA sequences of the 162 isolates found in each individual SNP-rich region. As shown in Figure 3A–C, the 162 isolates could be grouped into 9, 5 and 5 clades, based on sequences at Marker 1, Marker 2 and Marker 3, respectively. Most of the clades contained isolates from different rice cropping regions. The partition homogeneity test revealed that the phylogenetic analysis results, based on sequences at individual SNP-rich markers, strongly conflict with each other ($P = 0.001$). The 162 isolates could be grouped into more than 20 clades, based on the combined sequences at the three SNP-rich markers, and the isolate numbers for most clades were less than five (Figure 3D).

Population genetic structures

The 162 *U. virens* isolates used in this study were grouped into five geographical populations, each of which had 6–26 haplotypes. As shown in Table 2, with the exception of Northeast China, most of these populations displayed high nucleotide diversity and haplotype diversity. Central China possessed the highest genetic variation. The linkage disequilibrium analysis revealed that recombination occurred within most of the populations when all the populations were considered as a whole (Table 2). Genetic recombination occurred amongst all possible locus pairs (Figure S1). False smut pathogen has sexual stage because it produces sclerotia from which fruiting bodies form. When two compatible strains meet, their hyphae might fuse and the recombination might take place between two strains. However, the formation of fruiting bodies in fields has not previously reported in *U. virens*. To verify its genetic recombination revealed by this SNP-rich marker-based study, it needs to find the sexual stage of this fungus in fields first.

When two populations were compared, significant genetic differentiation could be found in most of the compared pairs, based on F_{ST} values. The only two exceptions were the pairs: Yunnan-Guizhou Plateau versus Sichuan-Shaanxi basin and Yunnan-Guizhou Plateau versus Northeast China (Table 3). AMOVA analysis found that the difference between these populations contributed to 11% of the total genetic variation while 89% of the genetic variation came from within individual populations. This is consistent with the fact that several haplotypes were shared by samples from different populations (Table S2), which indicated that gene flow occurred between populations. Genetic differences among isolates within populations and gene flow between different populations were also verified by principal component analysis (Figure 4) and

model-based clustering algorithm analysis by STRUCTURE (Figure 5). Furthermore, a significant and positive correlation ($P = 0.001$) between genetic and geographic distances was detected with a correlation coefficient (r) of 0.09 (Figure 6). Further investigations found that the genetic distances were significantly correlated with longitude gradients ($r = 0.08$; $P = 0.01$), but not with latitude gradients ($r = 0.01$; $P = 0.30$).

Discussion

Sequence-based markers are important in both molecular systematic and population genetic analyses because they can provide unambiguous information on defined loci and enable the direct assessment of variability [43]. In this study, three SNP-rich DNA markers, located in three different genomic regions, were identified by comparing the genomic sequences of two *U. virens* isolates and this was further confirmed by the sequence analysis of 10 isolates from 10 different provinces. Most of the DNA sequences used for MLST in other fungi are housekeeping genes [23–26]. However, none of the SNP-rich DNA markers found in this study were housekeeping genes. Therefore, this study provided novel DNA markers for analyzing the genetic diversity of fungal species, especially for species with poor polymorphism in commonly used housekeeping genes. In addition, these defined markers can be used to identify field *U. virens* isolates in the future. As far as can be ascertained, this is the first report on the DNA-sequence-based analysis of the genetic diversity and population structure of *U. virens*.

Analysis and comparison of the sequences of these SNP-rich regions revealed that *U. virens* populations in China have at least 56 multilocus sequence types (haplotypes). This means that *U. virens* isolates in the field are genetically diverse. However, the correlation between virulence and genotype is unknown. Although several haplotypes, such as haplotype 1 and 17, were widely distributed over the rice-growing areas, the genetic differentiations between isolates from geographically distant rice-growing areas were higher than isolates between adjacent rice-growing areas. In addition, the *U. virens* isolates collected from similar regions displayed considerable DNA composition stability. These results suggested that, naturally, *U. virens* in China might not be spread over long distances. The wide distribution of some *U. virens* haplotypes might be due to the transportation of rice seeds contaminated with smut balls containing chlamydospores.

This study demonstrates that the genetic diversities of *U. virens* populations are high in most of rice-growing areas. The 162 isolates could be grouped into more than 20 clades. The high genetic diversity in *U. virens* might be caused by genetic recombination through sexual cross between mating compatible strains in fields. In laboratory condition, *U. virens* has sexual stage during which fruiting bodies emerge from sclerotia and ascospores are released [1]. Our data clearly suggest that a sexual stage in fields does take place and does contribute to the genetic diversity of *U. virens* populations. The genetic diversity of *U. virens* in Northeast is relatively lower than other areas. Similar to this, an AFLP-based analysis

showed that the genetic variation *U. virens* populations of Liaoning province were low [20]. The lower genetic divergence of *U. virens* in Northeast suggests that sexual reproduction in this region might be less active than other regions. Climate conditions might affect sexual reproduction of *U. virens*.

Supporting Information

Table S1. Geographical distributions and haplotypes of *U. virens* isolates.

(XLS)

Table S2. Sequence differences among 56 haplotypes found at the three loci with enriched SNP sites.

(XLS)

Figure S1. Allelic combination showing evidence of recombination.

References

- Tanaka T, Ashizawa T, Sonoda R, Tanaka C (2008) *Villosiclava virens* gen. nov., com. nov., teleomorph of *Ustilagoidea virens*, the causal agent of rice false smut. Mycotaxon 106: 491-501.
- Kepler RM, Sung GH, Harada Y, Tanaka K, Tanaka E et al. (2012) Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and *Ustilagoidea* (Clavicipitaceae). Am J Bot 99: 552–561. doi:10.3732/ajb.1100124. PubMed: 22334447.
- Fu R, Ding L, Zhu J, Li P, Zheng AP (2012) Morphological structure of propagules and electrophoretic karyotype analysis of false smut *Villosiclava virens* in rice. J Microbiol 50: 263-269. doi:10.1007/s12275-012-1456-3. PubMed: 22538655.
- Singh RA, Dubey KS (1984) Sclerotial germination and ascospore formation of *Claviceps oryzae-sativae* in India, Indian. Phytopathology 37: 168-170.
- Schroud P, TeBeest DO (2005) Germination and infection of rice roots by spores of *Ustilagoidea virens*. AAES Research Series 540: 143-151
- Ashizawa T, Takahashi M, Arai M, Arie T (2012) Rice false smut pathogen, *Ustilagoidea virens*, invades through small gap at the apex of a rice spikelet before heading. J Gen Plant Pathol 78: 255-259. doi: 10.1007/s10327-012-0389-3.
- Tang YX, Jin J, Hu DW, Yong ML, Xu Y et al. (2013) Elucidation of the infection process of *Ustilagoidea virens* (teleomorph: *Villosiclava virens*) in rice spikelets. Plant Pathol 62: 1-8. doi:10.1111/j.1365-3059.2012.02629.x.
- Koiso Y, Li Y, Iwasaki S, Hanaoka K, Kobayashi T et al. (1994) Ustiloxin, antimitotic cyclic peptides from false smut balls on rice panicles caused by *Ustilagoidea virens*. J Antibiot 47: 765-773. doi: 10.7164/antibiotics.47.765. PubMed: 8071121.
- Ludueña RF, Roach MC, Prasad V, Banerjee M, Koiso Y et al. (1994) Interaction of ustiloxin A with bovine brain tubulin. Biochem Pharmacol 47: 1593–1599. doi:10.1016/0006-2952(94)90537-1. PubMed: 8185673.
- Yaegashi H, Fujita Y, Sonoda R (1989) Severe outbreak of false smut of rice in 1988. Plants Protect 43: 311-314 (in Japanese).
- Rush MC, Shahjahan AKM, Jones JP (2000) Outbreak of false smut of rice in Louisiana. Plant Dis 84: 100. doi:10.1094/PDIS.2000.84.1.100D.
- Singh AK, Pophaly DJ (2010) An unusual rice false smut epidemic reported in Raigarh District, Chhattisgarh, India. Int Rice Res Notes 35: 1-3.
- Deng GS (1989) Present status of research on false smut in China. Plants Protect 15: 39-40 (in Chinese).
- Zhou YL, Izumitsu K, Sonoda R, Nakazaki T et al. (2003) PCR-based specific detection of *Ustilagoidea virens* and *Ephelis japonica*. J Phytopathol 151: 513-518. doi:10.1046/j.1439-0434.2003.00761.x.
- Kim KW, Park EW (2007) Ultrastructure of spined conidia and hyphae of the rice false smut fungus *Ustilagoidea virens*. Micron 38: 626–631. doi:10.1016/j.micron.2006.09.006. PubMed: 17092730.
- Li YS, Zhu Z, Zhang YD, Zhao L, Wang CL (2008) Genetic analysis of rice false smut resistance using mixed major genes and polygenes inheritance model. Acta Agron Sinica 34: 1728-1733 (in Chinese). doi: 10.1016/S1875-2780(09)60007-9.
- Miyazaki S, Matsumoto Y, Uchihara T, Morimoto K (2009) High-performance liquid chromatographic determination of Ustiloxin A in forage rice silage. J Vet Med Sci 71: 239-241. doi:10.1292/jvms.71.239. PubMed: 19262041.
- Brooks SA, Anders MM, Yeater KM (2009) Effect of cultural management practices on the severity of false smut of rice. Plant Dis 93: 1202-1208. doi:10.1094/PDIS-93-11-1202.
- Ashizawa T, Takahashi M, Moriwaki J, Hirayae K (2010) Quantification of the rice false smut pathogen *Ustilagoidea virens* from soil in Japan using real-time PCR. Eur J Plant Pathol 128: 221-232. doi:10.1007/s10658-010-9647-4.
- Zhou YL, Pan YJ, Xie XW, Zhu LH, Xu JL et al. (2008) Genetic diversity of rice false smut fungus, *Ustilagoidea virens* and its pronounced differentiation of populations in North China. J Phytopathol 156: 559-564. doi:10.1111/j.1439-0434.2008.01387.x.
- Taylor JW, Fisher MC (2003) Fungal multi-locus sequence typing - it's not just for bacteria. Curr Opin Microbiol 6: 1–6. doi:10.1016/S1369-5274(02)00008-5.
- Xu JP (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. Mol Ecol 15: 1713-1731. doi:10.1111/j.1365-294X.2006.02882.x. PubMed: 16689892.
- Kasuga T, White TJ, Koenig G, McEwen J, Restrepo A et al. (2003) Phylogeography of the fungal pathogen *Histoplasma capsulatum*. Mol Ecol 12: 73–84. PubMed: 14629354.
- Li J, Fan SR, Liu XP, Li DM, Nie ZH et al. (2008) Biased genotype distributions of *Candida albicans* strains associated with vulvovaginal candidosis and candidal balanoposthitis in China. Clin Infect Dis 47: 1119–1125. doi:10.1086/592249. PubMed: 18808351.
- Bain JM, Tavanti A, Davidson AD, Jacobsen MD, Shaw D et al. (2007) Multilocus sequence typing of the pathogenic fungus *Aspergillus fumigatus*. J Clin Microbiol 45: 1469-1477. doi:10.1128/JCM.00064-07. PubMed: 17376886.
- Meyer W, Aanensen DM, Boekhout T, Cogliati M, Diaz MR et al. (2009) Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med Mycol 47: 561-570. doi: 10.1080/13693780902953886. PubMed: 19462334.
- Choi J, Park S-Y, Kim B-R, Roh J-H, Oh I-S et al. (2013) Comparative analysis of pathogenicity and phylogenetic relationship in *Magnaporthe grisea* species complex. PLOS ONE 8(2): e57196. doi:10.1371/journal.pone.0057196. PubMed: 23468934.
- TeBeest DO, Jecmen A, Dittmore M (2010) Infection of rice by the false smut fungus, *Ustilagoidea virens*. AAES Research Series 591: 70-80
- Mei FQ, Wu XZ, Yao CX, Li LP, Wang L et al. (1988) Rice cropping regionalization in China. Chin J Rice Science 2: 97-110.

(DOC)

Acknowledgements

The authors are grateful to Prof. Fu Huang, Prof. Yunliang PenG, Prof. Xiujuan Yang, Prof. Dongwei Hu, Dr. Maolin Hu, Dr. Yu Chen and Dr. Wei Zhang for their assistance in sample collection. We also thank Fatima Ismail for proofreading.

Author Contributions

Conceived and designed the experiments: SL XS LC WS. Performed the experiments: XS SK. Analyzed the data: XS YZ YY XZ. Contributed reagents/materials/analysis tools: XS XT HH YL SW. Wrote the manuscript: DL XS.

30. Nakada M, Tanaka C, Tsunewaki K, Tsuda M (1994) RFLP analysis for species separation in genera *Bipolaris* and *Curvularia*. *Mycoscience* 35: 271-278. doi:10.1007/BF02268449.
31. Hall T (2001) Biolign alignment and multiple contig editor. Available: <http://en.bio-soft.net/dna/BioLign.html>. Accessed 2013 May 8.
32. Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. *Bioinformatics* 20: 2878-2879. doi:10.1093/bioinformatics/bth315. PubMed: 15145805.
33. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948. doi:10.1093/bioinformatics/btm404. PubMed: 17846036.
34. Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731-2739. doi:10.1093/molbev/msr121. PubMed: 21546353.
35. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 57: 758-771. doi: 10.1080/10635150802429642. PubMed: 18853362.
36. Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* 1: 101-102. doi:10.1046/j.1471-8278.2000.00014.x.
37. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452. doi: 10.1093/bioinformatics/btp187. PubMed: 19346325.
38. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564-567. doi:10.1111/j.1755-0998.2010.02847.x. PubMed: 21565059.
39. Peakall ROD, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537-2539. doi:10.1093/bioinformatics/bts460. PubMed: 22820204.
40. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959. PubMed: 10835412.
41. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587. PubMed: 12930761.
42. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4: 137-138.
43. Xu JP (2006) Fundamentals of fungal molecular population genetic analyses. *Curr Issues Mol Biol* 8: 75-90. PubMed: 16875415.