In silico Analysis of the Potential Infection Mechanisms of *Magnaporthe grisea* from Horizontal Gene Transfer Hypothesis

Chunyang Li[1#], Ying Wang[2#], Hao Peng[1], Hejiao Bian[1], Mingwei Min[1], Longfei Chen[1], Qian Liu[1], and Jinku Bao[1*]

[1] *School of Life Sciences, Sichuan University, Chengdu 610064, China;* [2] *China National Center for Biotechnology Development, Beijing 100081, China.*

[#]Equal contribution.

*Corresponding author. E-mail: jinkubao@yahoo.com

DOI: 10.1016/S1672-0229(08)60036-4

Horizontal gene transfer (HGT) has long been considered as a principal force for an organism to gain novel genes in genome evolution. Homology search, phylogenetic analysis and nucleotide composition analysis are three major objective approaches to arguably determine the occurrence and directionality of HGT. Here, 21 genes that possess the potential to horizontal transfer were acquired from the whole genome of *Magnaporthe grisea* according to annotation, among which three candidate genes (corresponding protein accession numbers are EAA55123, EAA47200 and EAA52136) were selected for further analysis. According to BLAST homology results, we subsequently conducted phylogenetic analysis of the three candidate HGT genes. Moreover, nucleotide composition analysis was conducted to further validate these HGTs. In addition, the functions of the three candidate genes were searched in COG database. Consequently, we conclude that the gene encoding protein EAA55123 is transferred from *Clostridium perfringens*. Another HGT event is between EAA52136 and a certain metazoan's corresponding gene, but the direction remains uncertain. Yet, EAA47200 is not a transferred gene.

Key words: *Magnaporthe grisea*, infection mechanism, horizontal gene transfer, homology, phylogenetic analysis, nucleotide composition

Introduction

Magnaporthe grisea, also commonly known as rice blast fungus, rice rotten neck, rice seedling blight, blast of rice, oval leaf spot of graminea, pitting disease, ryegrass blast, and johnson spot, is one of the most destructive plant-pathogenic fungus that causes an important disease affecting rice (*1*, *2*). Conservatively estimated, the quantity of rice destroyed by *M. grisea* each year can feed at least 60 million people (*3*). Many countries would be threatened by hunger if this blast disease cannot be controlled, especially in developing countries who plant rice as their main food. Moreover, *M. grisea* is capable of infecting not only rice but also other domesticated grasses, such as wheat and barley. For instance, it was reported that the disease caused by this fungal pathogen had broken out on wheat in South America (*4*). Although it has been studied intensively, there are still myriads of mysteries on the potential infection mechanisms of *M. grisea* (*5*).

Horizontal gene transfer (HGT) has long been considered as a principal force for an organism to gain novel genes in genome evolution (*6*). Thanks to HGTs, one organism can gain novel genes from other species and along with these genes, new functions can be obtained and expressed in the recipient organism (*7*). These acquired new functions might be very important for species to adapt to the changing environment and avoid extinction (*8*). The possibility of HGT is strongly correlated with gene function. HGT event is more likely to happen in operational genes participating in housekeeping functions than in informational genes involved in the processes of transcription and translation (*9*). Furthermore, the genome sequences of bacteria and fungi have revealed the existence of genes whose similarity to loci in distantly related organisms is explained most parsimoniously by HGT events (*10*). Therefore, it is notable that the publication of the infection mechanisms of fungi

has focused on HGT events for several years (*11*). Of note, homology search, phylogenetic analysis and nucleotide composition analysis are three major objective approaches to arguably determine the occurrence and directionality of HGTs (*12*).

In the present study, we took advantage of the three main bioinformatics methods to determine the occurrence and directionality of possible HGT events between *M. grisea* and other organisms. According to the genome annotation of *M. grisea*, we found that 21 genes were possibly transferred from other species (*13*). Considering the results of BLAST homology and phylogenetic analyses of every one of the 21 genes, three candidate genes were selected for further analysis. Subsequently, nucleotide composition analysis was also utilized to verify these possible HGT events. The results provide new evidence to discover whether these HGTs would be available therapeutic targets in future investigations.

Results

Candidate gene selection

According to the annotated gene function of *M. grisea* genome, our study involved a set of 21 genes (**Table S1**) that had been proposed as instances of possible HGT between *M. grisea* and other organisms (*9*). The results of BLAST homology and phylogenetic analyses of every one of the 21 genes revealed three candidate genes with relatively more credibility as horizontal transferred genes (corresponding protein accession numbers are EAA55123, EAA47200 and EAA52136, respectively) (*12*).

BLAST homology analyses

According to BLAST results of protein EAA55123 between *M. grisea* and other organisms, the E-values of Fungi and Plantae species are obviously different (**Table S2**). For example, the E-value of *Arabidopsis thaliana* is 5e-84 and its identity value is 35%, and the E-value of *Oryza sativa* is le-79 and its identity value is 36%. These two plants have much lower E-values than fungi. Of note, *O. sativa* is the plant host of *M. grisea*. Thus, there is an opportunity for HGT to occur. Furthermore, this gene of *M. grisea* should have been similar with those of other fungi; however, the result showed that this gene is more analogous with those of plants and bacteria. Consequently, this gene was selected for subsequent phylogenetic analysis.

According to BLAST results of protein EAA47200 (**Table S3**), the E-value of *A. thaliana* is 2e-76 and its identity value is 27%, and the E-value of *O. sativa* is 3e-75 and its identity value is 27%. Compared with them, the E-value of *Yarrowia lipolytica* (fungus) is 5e-42 and its identity value is 27%. The result showed that this gene of *M. grisea* is more similar with *O. sativa* and *A. thaliana* than with other fungi. Therefore, this gene was also retained for the following phylogenetic analysis.

According to BLAST results of protein EAA52136 (**Table S4**), it was found that the most similar species with *M. grisea* is *O. sativa* with E-value 3e-57 and identity value 47%. Meanwhile, the E-value of *A. thaliana* is 2e-56 and its identity value is 47%. In comparison, there are relatively higher E-values and lower identities among fungi. Thus this gene was also selected for the next phylogenetic analysis.

Phylogenetic analyses

In several instances, BLAST searching resulted in a large number of homologous sequences, which served as input for a preliminary phylogeny designed to assess the paralogy and orthology of species. These phylogenetic trees then served as frameworks on which to exclude the more distantly related paralogies in a follow-up alignment and phylogeny. Therefore, we conducted phylogenetic analysis of the three selected proteins by neighbor joining, minimum evolution and UPGMA methods. In **Figure 1** for protein EAA55123, it was found that *M. grisea* is grouped together with *Clostridium perfringens*, which is a bacterial organism, and their bootstrap value is 53. This result is consistent with BLAST results in which the top hit is *C. perfringens* (E-value 2e-87). Moreover, while searching for a bigger group of *M. grisea*, it was discovered that all group members are bacteria. Another interesting phenomenon is that *A. thaliana* (plant) is near *Streptococcus pneumoniae* (bacterium), and *O. sativa* (plant) is near *Xylella fastidiosa* (bacterium). In **Figure 2** for protein EAA47200, it demonstrated that *M. grisea* and *Ustilago maydis* (fungus) are in the same branch. In addition, *Caenorhabditis elegans* (metazoan) and *Cryptococcus neoformans* (fungus) are grouped together, so are *Y. lipolytica* (fungus) and *Anopheles gambiae* (metazoan). In **Figure 3** for protein EAA52136, it showed that *M. grisea* is close to the group involving *Canis familiaris* (dog), *Rattus norvegicus* (rat) and *Bos taurus* (cattle). In

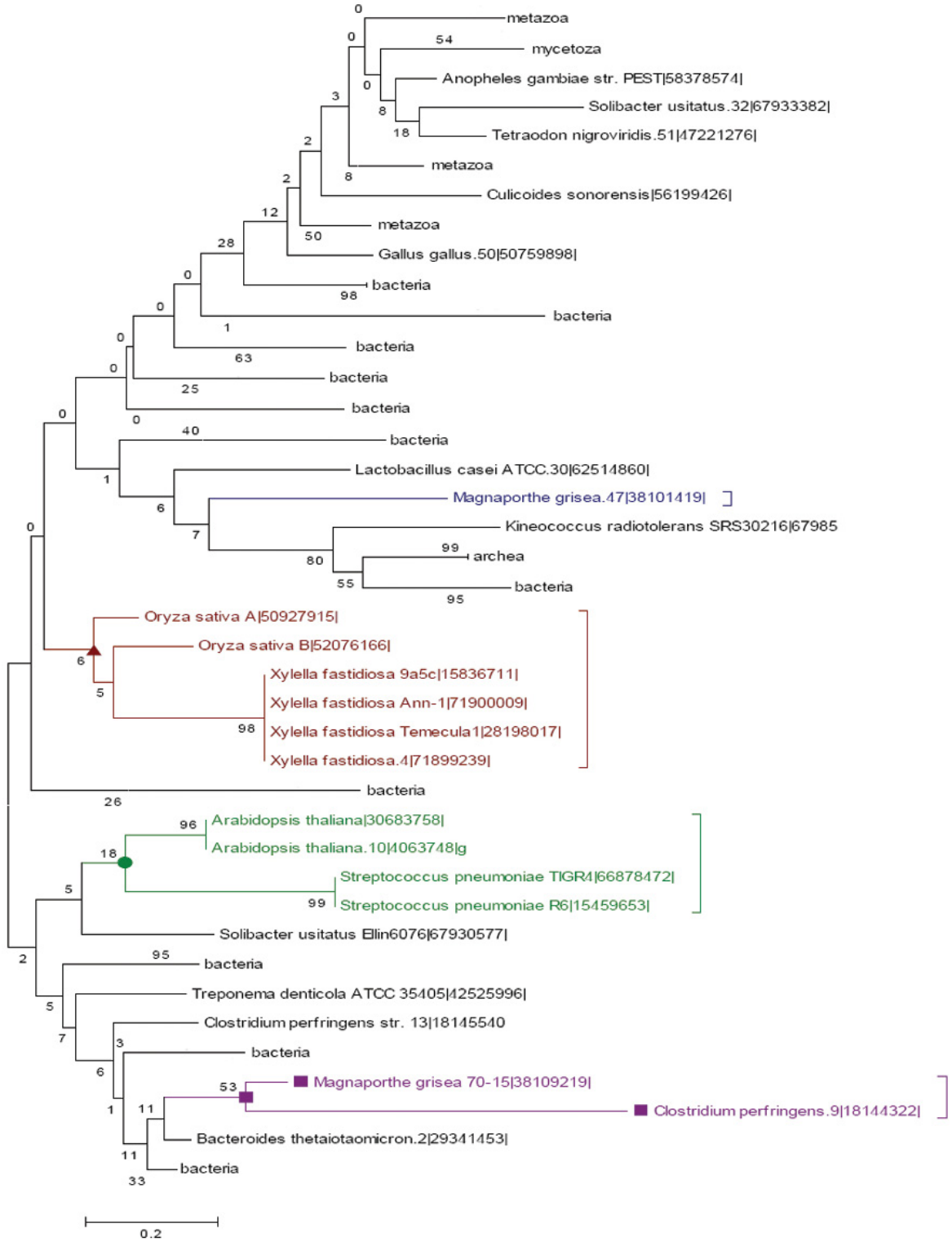

Figure 1 Phylogenetic tree of protein EAA55123 by neighbor joining method.

addition, *Erythrobacter litoralis* (bacterium) and *M. grisea* are in the same clade, so are *Arthrobacter* (bacterium) and *Strongylocentrotus purpuratus* (metazoan).

Nucleotide composition comparison

Next, we compared the nucleotide composition of the three putative transferred genes encoding proteins

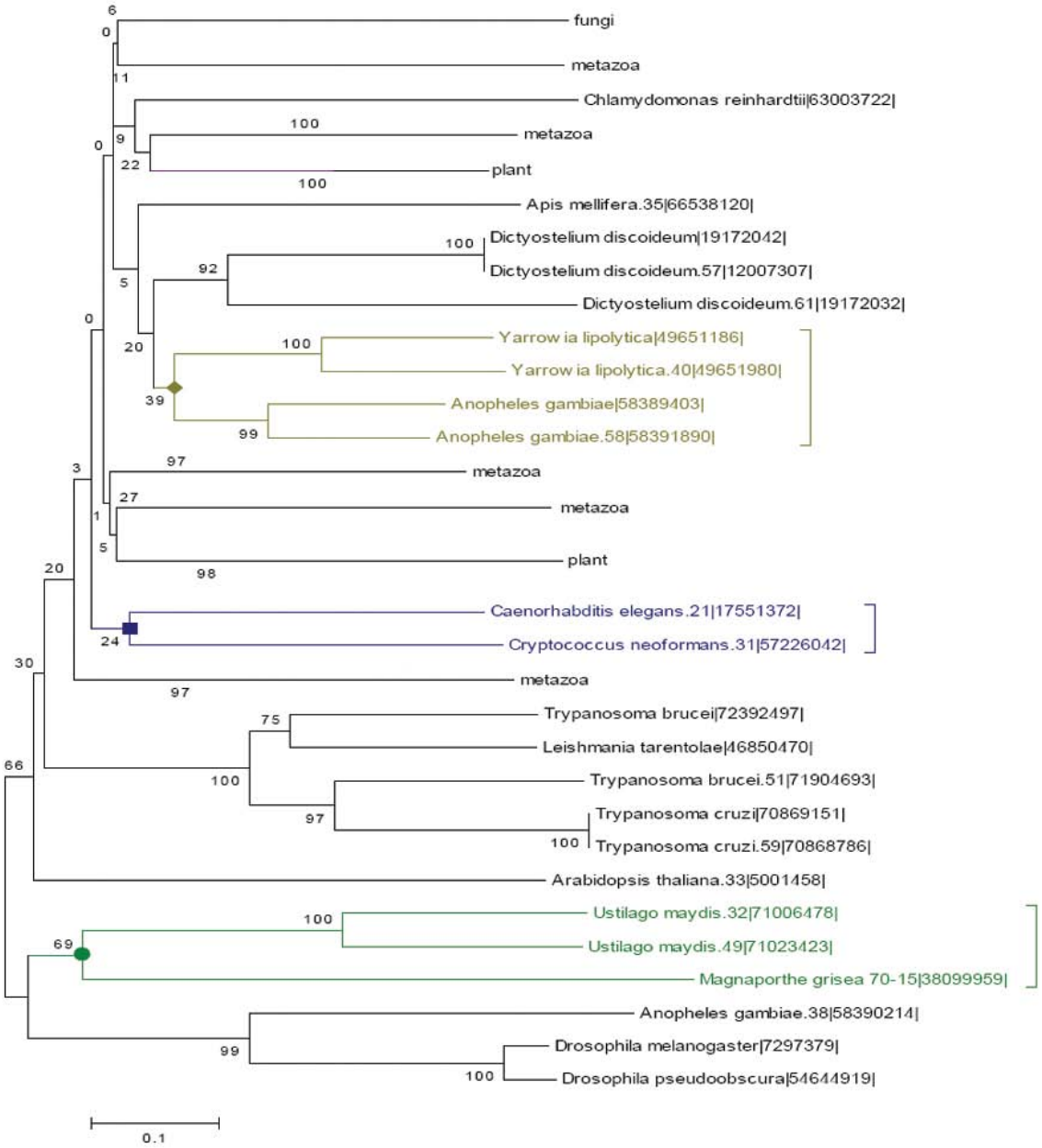

Figure 2 Phylogenetic tree of protein EAA47200 by neighbor joining method.

EAA55123, EAA47200 and EAA52136 with the average nucleotide composition of all the genes from *M. grisea* (**Figure 4A**). Results of the standard deviation (SD) values of nucleotide composition (nucleotide composition of putative transferred genes compared with that of the average of the whole genes) are shown in **Figure 4B**. For the gene encoding protein EAA55123, the SD value of T3s, A3s, G3s, GC3s and GC is greater than 1.5, while Nc's value is equal to 1.5. The highest SD value is 4.5 (G3s), which reflects that the G3s content of EAA55123 is quite different from the average G3s content of the whole genes. For the gene that encodes EAA47200, the SD value of C3s, A3s, G3s, Nc and GC3s is greater than 1.5, with the highest value of 4.35 (C3s). For the gene encoding EAA52136, the SD value of T3s, C3s and GC is greater than 1.5, with the most distinct one of 2.325 (C3s).

Function classification

Functions of the three putative transferred genes were obtained according to the function annotation provided by the COG database (*14*). The top hit E-value of protein EAA55123 is 5e-34 and its function is alpha-L-fucosidase. The top hit E-value of protein

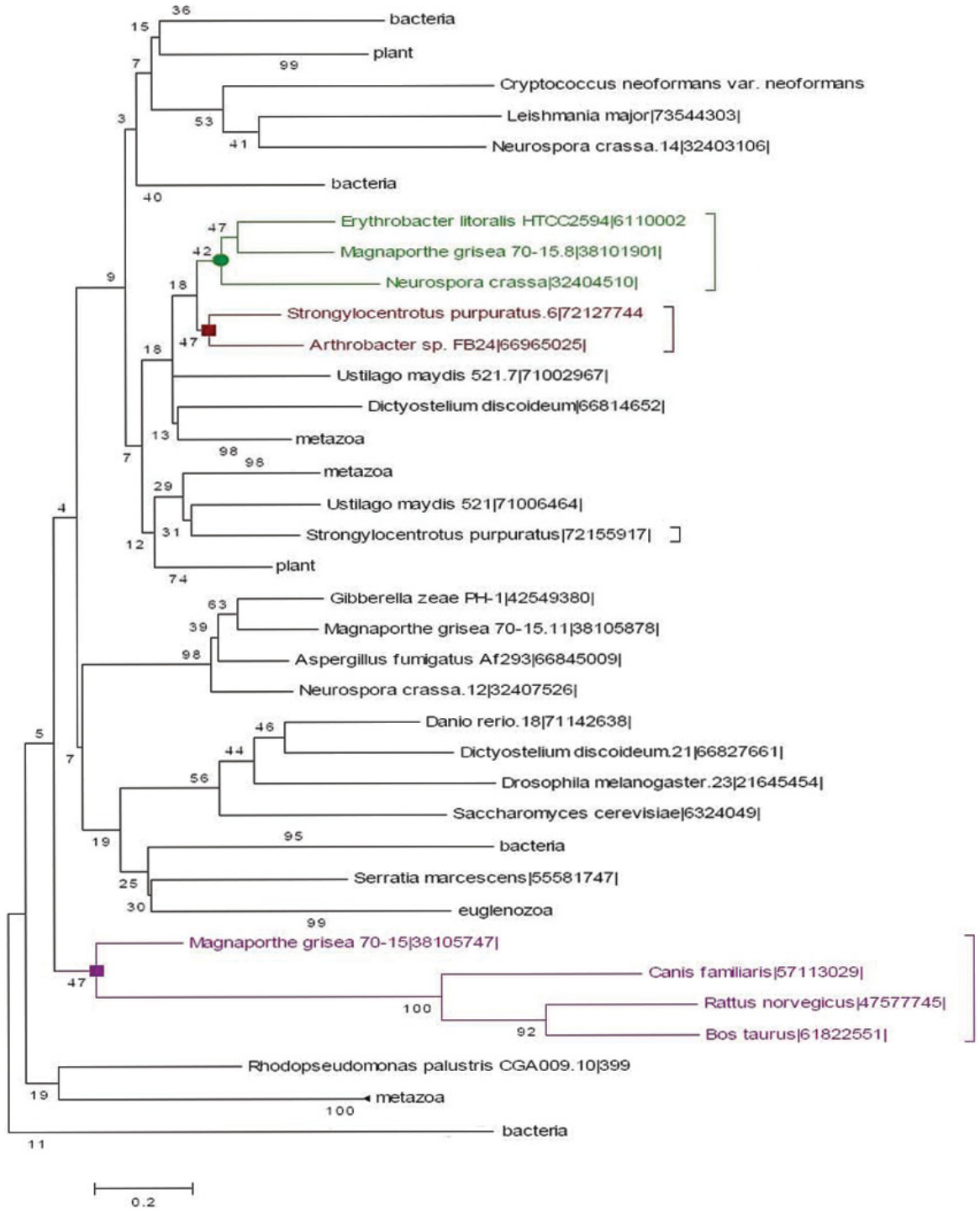

Figure 3 Phylogenetic tree of protein EAA52136 by minimum evolution method.

EAA47200 is 9e-63 and its function is ATP-binding cassette (ABC)-type multi-drug/protein/lipid transport system, ATPase component. The top hit E-value of protein EAA52136 is 3e-16 and its function is an oxidoreductase activity, acting on the CH–CH group of donors (uncharacterized membrane protein) (*15*).

Discussion

Three bioinformatics methods, homology search, phylogenetic analysis and nucleotide composition analysis, were used to indicate that HGT events might occur between *M. grisea* and other organisms. From

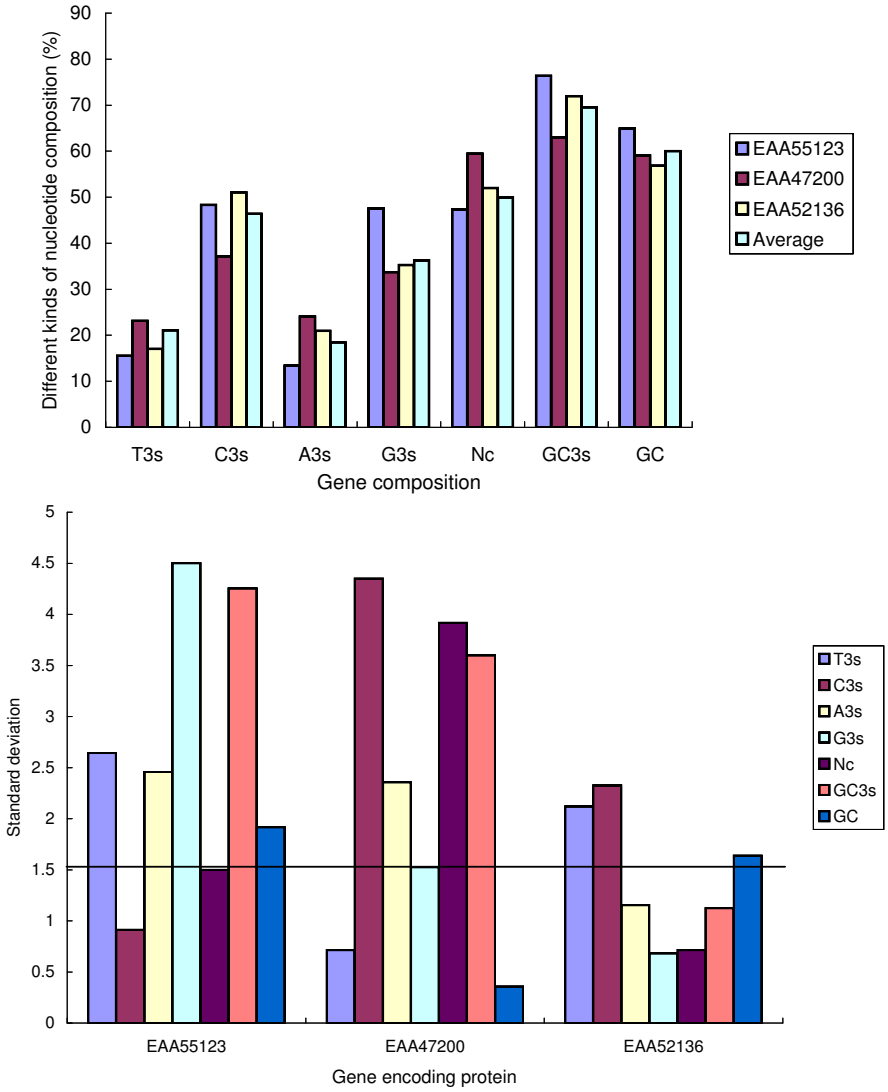

Figure 4 Nucleotide composition analysis of three putative transferred genes. GC3s: G+C content of the third position of synonymous codon; Nc: the effective number of codons; A3s, T3s, C3s and G3s: the silent base composition.

an evolutionary theoretical prospective, HGT, particularly when it occurs between a fungus and other organisms, especially its plant host, is the pivotal testimony to a remarkable clue to discover a new way to tackle diseases such as the rice blast disease caused by *M. grisea* (*16*).

In the present study, according to BLAST homology results, it was found that the protein EAA55123 of *M. grisea* has high sequence similarity with proteins of *C. pefringens*, *A. thaliana* and *O. sativa* (Table S2). Why do these proteins show such high similarity? There are two possibilities: one is that an HGT event happened between *M. grisea* and other species; the other is that this similarity might be an artifact. Because BLAST homology is based on corrected genetic distances, if the genetic distances are uncorrected, some unpredictable results would be acquired. From BLAST homology results, the protein EAA55123 is conserved among the four species (*M. grisea*, *C. pefringens*, *A. thaliana* and *O. sativa*), but *M. grisea* is far from the other species in phylogenetic relationship. It indicates that the gene encoding EAA55123 should be transferred by HGT occurring among these species. Table S3 demonstrates that the protein EAA47200 of *M. grisea* has the highest sequence similarity with *A. thaliana* (E-value 2e-76) and *O. sativa* (E-value 3e-75). Hence, it is possible that an HGT event happened between *M. grisea* and *A. thaliana* or *O. sativa*. Moreover, as *M. grisea* is a parasite of *O. sativa*, this relationship reinforces the

point that an HGT event might occur between them. Since *A. thaliana* and *O. sativa* belong to the same subclass and the E-value of *A. thaliana* is significant, *A. thaliana* might also be a host of *M. grisea*. This phenomenon was also found in Table S4 for protein EAA52136, where *O. sativa* has the lowest E-value (3e-57) and *A. thaliana* has the second lowest one (2e-56). Therefore, the gene encoding protein EAA52136 might also be transferred from *A. thaliana* and *O. sativa* to *M. grisea*, or from *M. grisea* to *A. thaliana* and *O. sativa* in bi-direction.

Based upon phylogenetic analysis, the protein EAA55123 of *M. grisea* is the nearest to a protein of *C. perfringens* (Figure 1), suggesting that an HGT event happened between *M. grisea* and *C. perfringens*. In addition, there are two homologous proteins of *M. grisea*, protein [38109219] (EAA55123) and protein [38101419] (EAA48384), which are distributed in different bacterium groups instead of being grouped together. It indicated that there was a gene duplication long time ago with accompanying gene duplication, which might have happened after an HGT event from *C. perfringens* to *M. grisea*. It also suggested that another HGT event might have happened between bacterium and *M. grisea*. On the other hand, *A. thaliana* and *O. sativa* are both plants; thereby they should be grouped together with each other or other plants. Nevertheless, they did not show a close phylogenetic relationship. It shows that *A. thaliana* is grouped together with *S. pneumoniae* whereas *O. sativa* is with *X. fastidiosa*. Thus there might be an HGT event from *S. pneumoniae* to *A. thaliana* or from *X. fastidiosa* to *O. sativa*. In the phylogenetic tree constructed by minimum evolution method, the protein [38101419] (EAA48384) of *M. grisea* is grouped together with a protein of bacterium *Lactobacillus casei*, demonstrating that the direction of HGT is from *L. casei* to *M. grisea*. Interestingly, it was not found in UPGMA tree that *M. grisea* is grouped together with *C. perfringens*, while *M. grisea* and *A. thaliana* are in the same branch, and all other species around them are bacteria. It suggested that there are indeed HGTs from bacteria to *M. grisea* or *A. thaliana* (neighbor joining tree described the exact donors of *M. grisea* and *A. thaliana*). The relationship between *O. sativa* and *X. fastidiosa* in UPGMA tree is the same as that in neighbor joining tree. The relationship between *L. casei* and *M. grisea* in UPGMA tree is the same as that in minimum evolution tree. Why could these phenomena happen? There are two possibilities. Firstly, the original gene was transferred from *C. perfringens* to *M. grisea*, and after HGT, the transferred gene duplicated into two homology genes. The first one encodes protein EAA55123, and the second one encodes protein EAA48384. Secondly, the gene encoding protein EAA55123 of *M. grisea* was transferred from *C. perfringens,* and the gene encoding protein EAA48384 of *M. grisea* was transferred from *L. casei*. Besides the above two genes of *M. grisea,* other possible HGTs were also discovered. Genes that encode proteins XP-473485 and BAD 46679 of *O. sativa* might be transferred from *X. fastidiosa*. Genes that encode proteins NP-180377 and AAC98456 of *A. thatliana* might be transferred from *S. pneumoniae*.

Although in BLAST homology results of protein EAA47200, *A. thaliana* and *O. sativa* showed the lowest E-values, in neighbor joining, minimum evolution and UPGMA trees this protein of *M. grisea* was grouped together with a protein of fungus *U. maydis* (Figure 2). It indicated that the gene encoding protein EAA47200 is not acquired from HGT event between *M. grisea* and other organisms. Nevertheless, they were separated from other fungus group, suggesting that perhaps after gene duplication, this gene was deleted from other fungi. All trees constructed by neighbor joining, minimum evolution and UPGMA methods were described as incongruent tree topologies, for example *Y. lipolytica* (fungus) and *A. gambiae* (metazoan) are in the same branch, while *C. elegans* (metazoan) and *C. neoformans* (fungus) are in the same branch. Thus, after eliminating gene duplication and deletion or insufficient sample, there might be HGTs between *Y. lipolytica* and *A. gambiae* as well as between *C. elegans* and *C. neoformans*. Extraordinarily, in neighbor joining and minimum evolution trees, *Y. lipolytica* stayed in metazoan group near *A. gambiae*. It suggested that if it is an HGT event, the direction is from *A. gambiae* to *Y. lipolytica*. Minimum evolution tree showed that *C. neoformans* stayed in the metazoan group near *C. elegans*. It pointed out that the direction of HGT is from *C. elegans* to *C. neoformans*.

In the phylogentic tree, protein EAA52136 of *M. grisea* was found to be grouped with *C. familiaris*, *R. norvegicus* and *B. taurus* (Figure 3). It is not consistent with BLAST homology results that showed significant E-values of *A. thaliana* and *O. sativa*. Thus, if ignoring gene duplication, deletion and other artifacts, there might be an HGT event between *M. grisea* and metazoan. From minimum evolution and neighbor joining trees, it was found that in the phylogenetic

group of *E. litoralis* (bacterium), *M. grisea* and *Neurospora crassa* (fungus), *M. grisea* is more close to *E. litoralis* than *N. crassa.* This result might be due to an HGT event. In addition, metazoan *S. purpuratus* and bacterium *Arthrobacter* are also in one group. Thus, there might be an HGT between them as well. Because the phylogenetic tree was constructed in accordance with BLAST results of protein EAA52136, the samples for EAA52136 are relatively sufficient in this tree. Therefore, HGTs were deduced occurring between *M. grisea* and *E. litoralis* as well as between *S. purpuratus* and *Arthrobacter*, which might be an artifact of gene duplication.

Shifts in nucleotide composition such as an increase in the frequencies of GC bases over AT bases may indicate that a gene or non-coding region from another species has been inserted into the genome (*12*). Some reports showed that this variation in species and in the same genome is caused by mutational bias (*17*). In the premise that the function remains unchanged, the way for discovering local mutational bias is to detect nucleotide in the third position of synonymous codons (*18*). Different species have different mutational bias and patterns of nucleotide composition (*14*), so if a gene's nucleotide composition is different from its genome, this gene might be transferred from other species. Therefore, it is reasonable to utilize T3s, C3s, A3s, G3s, Nc, GC3s and GC to discover HGT in this investigation. For the gene encoding protein EAA55123, the nucleotide compositions of T3s, A3s, G3s, GC3s and GC are different from the average of *M. grisea* genome except C3s. Thus, this gene might not be originated from this genome and should be transferred from other species. A long time after HGTs, the transferred gene's mutation would be subject to the local mutation bias of the acceptor genome, and this process is termed "amelioration" (*19*). Hence, its C3s is identical with the genome of *M. grisea*, which is possible that C3s is subject to local mutation bias of the *M. grisea* genome. For the gene encoding protein EAA47200, it might be transferred from other species, since its C3S, A3s, G3s, Nc and GC3s are different from the genome of *M. grisea*. However, since the detection method of phylogenetic tree remains to be the most credible method to indicate HGT events (*12*), we finally adopt the results of the phylogenetic trees of protein EAA47200, which indicated that no HGT event of this gene happened between *M. grisea* and other organisms. Yet, for the gene encoding protein EAA52136, only T3s, C3s and GC are distinct from the *M. grisea* genome. Accordingly, it might be an ancient transferred gene, with A3s, G3s, Nc and GC3s subject to local mutation of the acceptor genome.

In summary, we conclude that the gene encoding protein EAA55123 of *M. grisea* is transferred from *C. perfringens*. Another HGT event is between the gene encoding protein EAA52136 and a certain metazoan's corresponding gene, but the direction remains uncertain. Yet, the gene encoding protein EAA47200 is not a transferred gene. In most instances, homology search, phylogenetic analysis and nucleotide composition analysis can only demonstrate that the genes have probably been inserted into the given genome by horizontal transfer. Therefore, more thoughtful analysis of cellular pathways could explain the extent and significance of putative HGTs or gene-loss events more thoroughly. Accordingly, our understanding of biological implications of HGTs should also require some directly experimental studies with these genes. The consequences of these recent developments are yet to come, but they would undoubtedly help filling in many gaps in our knowledge and understanding of the potential infection mechanisms of *M. grisea* in the near future.

Materials and Methods

Data collection and BLAST homology method

The infection mechanisms of *M. grisea* are intensively associated with biosynthesis, metabolism and transport (*12*), so our study involved a set of 21 relatively credible horizontal transferred genes that participate in the process of biosynthesis, metabolism or transport according to their genome annotation. We subsequently analyzed the 21 proteins encoded by these 21 genes in order to elucidate their biological functions on proteomic level. Each protein sequence was selected using BLAST (*20*) and phylogenetic analysis (*12*). ClustalW (*21*) was used to align the resulting set of significant homologues, and the alignments were then refined manually using Genetics Computer Group (GCG) software. After analyzing the results of BLAST and phylogenetic analysis of the 21 genes, three candidate transferred genes, AACU01000750, AACU01001862 and ACCU01000751, were used for further analyses. Their protein accession numbers are EAA55123, EAA47200 and EAA52136, respectively.

Phylogenetic method

The amino acid sequence alignments were analyzed using neighbor joining, minimum evolution and UPGMA methods as implemented in the Molecular Evolutionary Genetics Analysis (MEGA) software package (*22*). Phylogenetic analysis randomized the input order 20±50 times, depending on the number of sequences in the alignments. All potentially ambiguous gaps were removed before phylogenetic analysis. We coded all remaining gaps as missing data. Gap opening penalty of multiple alignments was 10. Gap extension penalty of multiple alignments was 0.2. Clade strength was assessed with bootstrap using 1,000 replications.

Nucleotide composition analysis

The three candidate transferred genes were analyzed by nucleotide composition analysis. To further detect these possible HGTs, CodonW software (*23*) was exerted including calculation of G+C. All mRNAs were gained from the website (http://www.broad.mit.edu/annotation/) and analyzed by CodonW. The averages of G+C, GC3s, Nc, A3s, T3s, G3s and C3s of all genes of the *M. grisea* genome were computed by Perl program (*24*). Accordingly, the differences between compositions of the putative transferred genes and the average of all genes were determined. The criterion of discovering HGT events is whether the differences are over 1.5 standard deviations. T-test was taken for statistical analysis.

Acknowledgements

We thank Dr. Bo Liu for his constructive suggestion and critical review of this manuscript. This work was supported in part by grants from the National Natural Science Foundation of China (General Programs No. 30270331 and No. 30670469), Director Fund of the State Key Laboratory of Oral Diseases (Sichuan University), the Science and Technology Fund for Distinguished Young Scholars of Sichuan Province (No. 06ZQ026-035) and the Key Technologies R&D Program of Sichuan Province (2006Z08-010).

Authors' contributions

CL conducted data analyses and prepared the manuscript. YW conceived the idea of using this approach and assisted with manuscript preparation. HP and MM assisted with data analysis. HB assisted with manuscript preparation. LC and QL collected the datasets. JB supervised the project and co-wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1. Dean, R.A., *et al.* 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434: 980-986.
2. Kanhere, A. and Vingron, M. 2009. Horizontal Gene Transfers in prokaryotes show differential preferences for metabolic and translational genes. *BMC Evol. Biol.* 9: 9.
3. Zeigler, R.S., *et al.* 1994. *Rice Blast Disease*. CAB International, Wallingford, UK.
4. Valent, B. and Chumley, F.G. 1994. Avirulence genes and mechanisms of genetic instability in the rice blast fungus. In *Rice Blast Disease* (eds. Zeigler, R.S., *et al.*), pp.111-134. CAB International, Wallingford, UK.
5. Davis, C.C. and Wurdack, K.J. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malphigialis. *Science* 305: 676-678.
6. Roulin, A., *et al.* 2009. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon *Route66* in *Poaceae*. *BMC Evol. Biol.* 9: 58.
7. Jain, R., *et al.* 2003. Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* 20: 1598-1602.
8. Richards, T.A., *et al.* 2003. Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* 154: 17-32.
9. Jain, R., *et al.* 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* 96: 3801-3806.
10. Marraffini, L.A. and Sontheimer, E.J. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322: 1843-1845.
11. Lawrence, J.G. 2005. Horizontal and vertical gene transfer: the life history of pathogens. *Contrib. Microbiol.* 12: 255-271.
12. Brown, J.R. 2003. Ancient horizontal gene transfer. *Nat. Rev. Genet.* 4: 121-132.
13. Altschul, S.F., *et al.* 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

14. Tatusov, R.L., *et al.* 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
15. Ochman, H. and Moran, N.A. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292: 1096-1098.
16. Ochman, H., *et al.* 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405: 299-304.
17. Heinrich, J., *et al.* 2004. A comparison of methods for the assessment of postural load and duration of computer use. *Occup. Environ. Med.* 61: 1027-1031.
18. Tatusov, R.L., *et al.* 1997. A genomic perspective on protein families. *Science* 278: 631-637.
19. Ragan, M.A. and Charlebois, R.L. 2002. Distributional profiles of homologous open reading frames among bacterial phyla: implications for vertical and lateral transmission. *Int. J. Syst. Evol. Microbiol.* 52: 777-787.
20. Altschul, S.F., *et al.* 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
21. Thompson, J.D., *et al.* 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
22. Tamura, K., *et al.* 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
23. Jia, W. and Higgs, P.G. 2008. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Mol. Biol. Evol.* 25: 339-351.
24. Anwar, T. and Khan, A.U. 2006. MAP MUTATION: a program for analyzing mutations in protein sequences. *Bioinformation* 1: 92-93.

Supporting Online Material

Tables S1–S4

DOI: 10.1016/S1672-0229(08)60036-4