

Open Access

## Analysis of *in planta* Expressed Orphan Genes in the Rice Blast Fungus *Magnaporthe oryzae*

Md. Abu Sadat<sup>1†</sup>, Junhyun Jeon<sup>1†</sup>, Albely Afifa Mir<sup>1</sup>, Seongbeom Kim<sup>1</sup>, Jaeyoung Choi<sup>1</sup> and Yong-Hwan Lee<sup>1,2,3\*</sup>

<sup>1</sup>Department of Agricultural Biotechnology, <sup>2</sup>Center for Fungal Pathogenesis, <sup>3</sup>Center for Fungal Genetic Resources, Plant Genomics and Breeding Institute, and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea

(Received on August 1, 2014; Revised on August 25, 2014; Accepted on August 29, 2014)

Genomes contain a large number of unique genes which have not been found in other species. Although the origin of such “orphan” genes remains unclear, they are thought to be involved in species-specific adaptive processes. Here, we analyzed seven orphan genes (*MoSPC1* to *MoSPC7*) prioritized based on *in planta* expressed sequence tag data in the rice blast fungus, *Magnaporthe oryzae*. Expression analysis using qRT-PCR confirmed the expression of four genes (*MoSPC1*, *MoSPC2*, *MoSPC3* and *MoSPC7*) during plant infection. However, individual deletion mutants of these four genes did not differ from the wild-type strain for all phenotypes examined, including pathogenicity. The length, GC contents, codon adaptation index and expression during mycelial growth of the four genes suggest that these genes formed during the evolutionary history of *M. oryzae*. Synteny analyses using closely related fungal species corroborated the notion that these genes evolved *de novo* in the *M. oryzae* genome. In this report, we discuss our inability to detect phenotypic changes in the four deletion mutants. Based on these results, the four orphan genes may be products of *de novo* gene birth processes, and their adaptive potential is in the course of being tested for retention or extinction through natural selection.

**Keywords :** fungal pathogenesis, gene birth, *Magnaporthe oryzae*, orphan gene

Orphan genes are protein-coding regions that lack any recognizable orthologs in other organisms (Ekman and

Elofsson, 2010). Based on the large number of genome sequences available, orphans are a universal feature of all genomes (Khalturin et al., 2009). Two major models have been proposed for the origin of new genes; the duplication-divergence model and *de novo* evolution model (Tautz and Domazet-Loso, 2011). In the first model, new genes emerge through gene duplication followed by rapid divergence leading to loss of similarity to the gene from which it was duplicated (Domazet-Loso and Tautz, 2003). This model has been well-supported and is considered the major mechanism for creating evolutionary novelties. However, within the framework of this model, it is difficult to explain how natural selection can choose one gene for divergence while retaining the duplicate to maintain the ancestral function (Conant and Wolfe, 2008; Lynch and Katju, 2004). Furthermore, this model is based on the extensive accumulation of substitutions over the entire length of the protein during divergence to the point at which paralogous relationships cannot be detected. This assumption rarely holds due to the existence of functional protein domains in many genes. The second model postulates that a new gene can emerge directly from non-coding sequences via a random combination of sequences, giving rise to functional sites such as transcription initiation regions and polyadenylation sites (Siepel, 2009).

Although such *de novo* gene birth is considered very rare, recent reports provided evidence for this type of gene birth in a variety of organisms (Begun et al., 2007; Cai et al., 2008; Donoghue et al., 2011; Heinen et al., 2009; Levine et al., 2006; Li et al., 2010a; Yang and Huang, 2011; Zhou et al., 2008). Furthermore, comparative analyses of genome sequences revealed that orphan genes emerge at high rates and account for 10–20% of total genes within individual eukaryotic genomes (Domazet-Loso and Tautz, 2003; Khalturin et al., 2009). Characteristics of orphan genes are well-documented in diverse eukaryotes. They are relatively short (both with respect to gene and ORF length), contain

<sup>†</sup>These authors contributed equally to this study

\*Corresponding author.

Phone) +82-2-880-4674, FAX) +82-2-873-2317

E-mail) yonglee@snu.ac.kr

a low number of exons and the detectable domains and are less expressed and evolve more rapidly than non-orphan or old genes (Carvunis et al., 2012; Neme and Tautz, 2013).

Orphan genes are believed to be involved in species-specific processes of adaptation (Kaessmann, 2010). A number of studies provide examples supporting such roles of orphan genes. In human, *FLJ33706* that encodes newly organized genes is associated with brain function (Li et al., 2010a). In the yeast *Saccharomyces cerevisiae*, the newly born *BSC4* gene is suggested to be involved in the DNA damage repair pathway during the stationary phase, whereas another new gene, *MDF1*, promotes vegetative growth and decreases mating efficiency in rich media (Cai et al., 2008; Li et al., 2010b). The *EED1* gene, which is found only in *Candida albicans*, is crucial for hyphal extension and maintenance of filamentous growth both on solid surfaces and during the interaction with host cells (Martin et al., 2011).

In this report, we analyzed orphan genes and determined whether they are involved in fungal pathogenesis in a model plant pathogenic fungus, *Magnaporthe oryzae*. The fungus is a causal agent of the rice blast disease, which is one of the most devastating global fungal diseases of cultivated rice (Talbot, 2003; Valent and Chumley, 1991). Leaf infection by this fungus is initiated by landing of a conidium on the leaf surface. Melanized appressorium develops to mechanically penetrate the cuticular layer, following conidial germination (Wilson and Talbot, 2009). Rice blast is used as a model system to investigate host-pathogen interactions due to the genetic tractability and availability of genome sequences for both organisms (Dean et al., 2005; Goff, 2005; Yu et al., 2002). In this study, we report for the first time the identification and analysis of orphan genes in *M. oryzae*.

## Materials and Methods

**Computational analysis.** Orphan genes were identified using BLASTP searches against the NCBI non-redundant dataset using protein sequences of *M. oryzae* as queries. The E-value threshold was set at  $10^{-3}$ . The BLAST matrix function embedded in CFGP 2.0 was used to visualize the presence of the gene in *M. oryzae* among species with genome sequences archived in CFGP 2.0 (Choi et al., 2013) (<http://cfgp.snu.ac.kr/>). Comparison between orphan and non-orphan genes based on gene length, GC content, transcript abundance, and codon adaptation index was performed using R (<http://www.R-project.org/>).

**Fungal isolates and culture conditions.** *Magnaporthe*

*oryzae* strain KJ201 was provided by the Center for Fungal Genetic Resources (CFGR, <http://genebank.snu.ac.kr>) and was used as the wild-type strain for this study. All strains including mutants were grown on V8 agar [V8; 8% V8 juice (v/v) and 1.5% agar (w/v), adjusted to pH 6.0 using NaOH] or oatmeal agar [OMA; 5% oatmeal (w/v) and 2% agar (w/v)] at 25°C in constant light to promote conidial production (Park et al., 2010). For conidial production, strains were cultured on V8 juice agar medium for 7 days or on OMA for 10 to 15 days at 25°C under continuous light conditions. To compare mycelial growth, modified complete agar medium (MCA) or modified minimal agar medium (MMA) (Talbot et al., 1993) was used.

**Nucleic acid manipulation and expression analysis.** Two methods were used for fungal genomic DNA isolation based on two different purposes. A quick and easy genomic DNA extraction method was used for PCR-based screening of transformants (Chi et al., 2009). Genomic DNA was isolated from mycelia according to a standard protocol for southern hybridization (Sambrook and Russell, 2001). Southern DNA hybridization was performed with the selected transformants to ensure correct gene replacement events and absence of ectopic integration. Genomic DNA was digested with *Bam*HI, *Pst*I, *Xho*I and *Nhe*I, and blots were probed with 1-kb 5'-flanking or 3'-flanking sequences (Fig. S1). Southern DNA hybridization was performed using a standard method (Sambrook and Russell, 2001). To perform expression analysis using quantitative real-time PCR (qRT-PCR), cDNA synthesis was performed with 5 µg of total RNA using the oligo dT primer with the ImProm-IITM Reverse Transcription System kit (Promega, Madison, WI, USA) following the manufacturer's instructions. Primer pairs used in this study were listed in Table S1.

**Targeted gene deletion.** Knock-out constructs of individual genes were generated by double-joint PCR where ~1 kb flanking sequences of each gene was amplified and fused with the hygromycin resistance gene (HPH) cassette. The knock-out constructs were transferred to wild-type protoplasts and the resulting transformants were primarily selected by PCR-based screening using specific primer pairs *MoSPC1\_NF*, *MoSPC1\_NR*, *MoSPC2\_NF*, *MoSPC2\_NR*, *MoSPC3\_NF*, *MoSPC3\_NR*, *MoSPC7\_NF* and *MoSPC7\_NR* (Table S1). Knock-outs were confirmed by Southern blot analysis using one of the flanking sequences as a probe.

**Developmental phenotype assays.** Radial mycelial growth

was measured on modified complete agar medium (MCA) or modified minimal agar medium (MMA) 12 days after inoculation (DAI) with three replications (Kim et al., 2014). Conidia were harvested from 7-day-old mycelia grown on V8 juice agar plates and the conidial suspension was adjusted to  $2 \times 10^4$  conidia/ml. For conidial germination and appressorium formation, 40  $\mu$ l of conidial suspension was dropped onto plastic coverslips with three replicates and incubated in a moistened box at room temperature. At 2 and 4 h after incubation, the frequency of conidial germination was determined by counting at least 100 conidia per replicate under a microscope. The frequency of appressorium formation was measured from germinated conidia at 8 h after incubation. Conidiation was determined by counting the number of conidia using a hemacytometer under a microscope. These assay processes were performed with three replicates in three independent experiments.

**Pathogenicity assay.** For the pathogenicity assay by spray inoculation, conidia were collected from 7-day-old V8 juice agar medium and 10 ml of conidial suspension were adjusted to  $1 \times 10^5$  conidia/ml containing Tween 20 (250 ppm final concentration) and sprayed onto the rice seedlings (*Oryza sativa* cv. Nakdongbyeo) of three to four leaf stages. Sprayed rice seedlings were placed in a dew chamber for 24 h under dark conditions at 25°C, then transferred to a rice growth incubator at 25°C, 80% humidity and a 16-h photoperiod with fluorescent lights.

## Results

**Identification of orphan genes with transcriptional activity in *M. oryzae*.** To identify orphan genes in the *M. oryzae* genome, we performed BLASTP searches against the NCBI non-redundant database (NCBI nr) with an e-value threshold of  $10^{-3}$ . This search demonstrated that 2,740 out of 12,991 genes (~21%) encode putative proteins with no match in the NCBI nr, indicating that they are likely orphan genes. We hypothesized that there should be evidence of transcription of orphan genes with roles in the host plant interaction. Thus, we utilized a previously reported *in planta* EST library (Kim et al., 2010). Although a total of 712 genes were initially reported to have *in planta* ESTs, we found only 542 genes in *M. oryzae* genome version 8. Comparison of orphan genes and *in planta* EST data showed that seven genes were shared between the two datasets. Since these genes were specific to *M. oryzae*, we named them *MoSPCs* (*Magnaporthe oryzae* specific). These genes were typically less than 1 kb in length (excluding *MoSPC6*), and *MoSPC3* was predicted

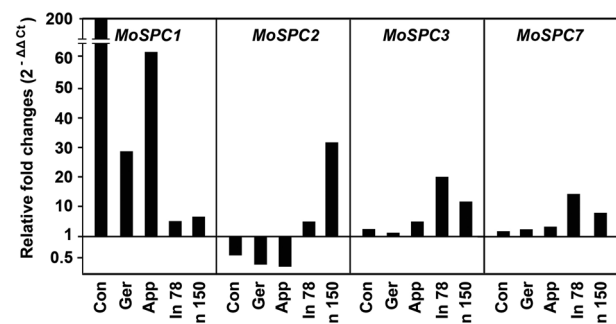
**Table 1.** Structural characteristics of *MoSPC* genes

Gene name	Gene size (bp)	No. of exons	Longer exon (bp)	Secretion possibility
<i>MoSPC1</i>	809	3	330	No
<i>MoSPC2</i>	546	2	251	No
<i>MoSPC3</i>	704	3	198	Yes
<i>MoSPC4</i>	233	2	134	No
<i>MoSPC5</i>	984	1	298	No
<i>MoSPC6</i>	1979	7	218	No
<i>MoSPC7</i>	777	2	222	No

to contain a signal peptide (Table 1).

**Expression analysis of *in planta* expressed *M. oryzae* orphan genes.** The presence of expressed sequence tag (EST)s is a good indication of transcriptional activity. However, it does not provide information on differential expression of a gene, which is associated with the gene's role. Therefore, we examined the transcriptional activity of seven genes during different developmental stages, including plant infection, using qRT-PCR. Expression profiling suggested that four genes (*MoSPC1*, *MoSPC2*, *MoSPC3* and *MoSPC7*) among the seven are differentially expressed during plant infection (appressorium formation and invasive growth) compared to mycelial growth conditions (Fig. 1). However, both *MoSPC5* and *MoSPC6* were down-regulated during plant infection, and *MoSPC4* was not differentially expressed under the same conditions (data not shown). Based on this expression analysis, we selected *MoSPC1*, *MoSPC2*, *MoSPC3* and *MoSPC7* for targeted deletion and further functional studies.

**Targeted deletion of selected genes and phenotypes analysis.** To investigate the possible roles of the four se-



**Fig. 1.** Transcript abundance of *MoSPC* genes in different development stages of *Magnaporthe oryzae*. Con, conidia; Ger, germinating conidia; App, appressoria; In 78, infection stage at 78 h post-inoculation (hpi); In 150, infection stage at 150 hpi.

**Table 2.** Phenotypes of wild-type,  $\Delta MoSPC1$ ,  $\Delta MoSPC2$ ,  $\Delta MoSPC3$  and  $\Delta MoSPC7$ 

Strain	Mycelial growth (mm) <sup>a</sup>		Conidiation ( $\times 10^4$ /ml) <sup>b</sup>	Conidial germination (%) <sup>c</sup>	Appressorium formation (%) <sup>d</sup>
	MCA	MMA			
KJ201	83.6 $\pm$ 0.6	78.6 $\pm$ 0.6	27.8 $\pm$ 4.6	90.2 $\pm$ 1.1	84.3 $\pm$ 5.4
$\Delta MoSPC1$	83.2 $\pm$ 0.7	78.0 $\pm$ 0.6	24.3 $\pm$ 9.6	90.2 $\pm$ 1.1	79.4 $\pm$ 3.9
$\Delta MoSPC2$	82.2 $\pm$ 0.3	77.3 $\pm$ .03	28.3 $\pm$ 3.3	88.9 $\pm$ 0.9	79.4 $\pm$ 3.9
$\Delta MoSPC3$	81.6 $\pm$ 1.6	77.0 $\pm$ 1.3	29.2 $\pm$ 4.2	90.3 $\pm$ 1.3	82.3 $\pm$ 3.3
$\Delta MoSPC7$	82.3 $\pm$ 0.6	76.0 $\pm$ 1.5	26.6 $\pm$ 0.6	87.1 $\pm$ 3.3	81.2 $\pm$ 3.5

<sup>a</sup>Mycelial growth was measured at 12 days after inoculation (DAI) on modified complete agar medium (MCA) and minimal agar medium (MMA).

<sup>b</sup>Conidiation was measured as the number of conidia from a culture flooded with 5 ml of sterilized distilled water.

<sup>c</sup>Percentage of conidial germination was measured on plastic coverslips under a light microscope using conidia harvested from 7-day-old V8 juice agar plates.

<sup>d</sup>Percentage of appressorium formation was measured on plastic coverslips using conidia harvested from 7-day-old V8 juice agar plates.

Data are presented as means  $\pm$  standard deviation from three independent experiments.

lected genes in *M. oryzae*, we generated deletion mutants of individual genes. Gene deletion constructs were prepared through double joint PCR (Yu et al., 2004), in which the HPH cassette was combined with ~1-kb 5'/3' flanking regions of the target gene (Fig. S1). The gene deletion constructs were used directly for transformation with wild-type protoplasts. The resulting hygromycin-resistant transformants were screened by PCR and correct gene replacement, and the resulting transformants were confirmed using Southern hybridization analysis (Fig. S1).

When examining the phenotypes of the deletion mutant in each of the four genes, we found that deletion mutants of *MoSPCs* were comparable to the wild-type with respect to mycelial growth and conidiation (Table 2). Conidia of all mutants were capable of germinating and developing appressorium on a germ tube tip. The morphology of appressoria formed by the four mutants was indistinguishable from the wild-type. Furthermore, when conidial suspensions of the mutants were spray-inoculated onto rice plants of a susceptible cultivar, Nakdongbyeo, all mutants showed virulence similar to the wild-type strain (Fig. 2A and 2B). These results suggested that *MoSPCs* are not required for the traits examined, including mycelial growth, conidiation, conidial germination, appressorium formation, and pathogenicity.

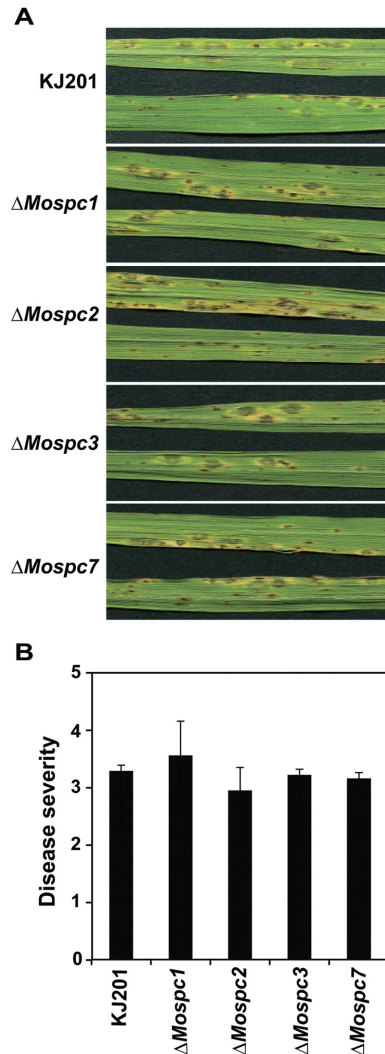
### Genome-wide analysis of orphan genes in *M. oryzae*.

In parallel to the targeted deletion of the four selected genes, we analyzed orphan genes at the genome scale by comparing the features of orphan and non-orphan genes, including gene length, GC contents, transcription and codon adaptation index. Our analysis showed that orphan genes including *MoSPCs* (excluding *MoSPC6*) were relatively short (Fig. 3A), showed a low GC content (Fig.

3B), low transcription (Fig. 3C) and less-biased codon usage (Fig. 3D) than their non-orphan counterparts. Gene length is positively associated with both GC content and gene expression (Jansen and Gerstein, 2000; Pozzoli et al., 2008). We also observed such a positive correlation among gene features. The differences between orphan and non-orphans in our analysis are in concordance with the results of previous studies on *de novo* emergence of orphan genes from model organisms (Cai et al., 2008; Carvunis et al., 2012; Li et al., 2010b; Neme and Tautz, 2013).

**Synteny analysis of orphan genes.** The most stringent criterion for involvement of *de novo* processes in explaining orphan genes requires that syntenic blocks spanning an orphan gene are present in outgroup organisms as non-coding sequences that are not transcribed (Cai et al., 2008; Knowles and McLysaght, 2009). To further support *de novo* gene birth of the four selected genes, we performed synteny analysis with phylogenetically closely related fungal species, including *Gaeumannomyces graminis* (a member of the Magnaporthaceae family) (Table S2). Our synteny analysis demonstrated that the selected genes (excluding *MoSPC2*) were present in the well-conserved synteny block (Fig. 4). We did not detect a synteny block containing *MoSPC2*, even by expanding the number of fungal species included in the analysis. Based on our analysis, *MoSPC3* was the orphan gene that best fit our stringent criterion. In synteny blocks harboring *MoSPC1* and *MoSPC7*, predicted ORFs (GGTG\_00636 and GGTG\_00730) of *G. graminis* were found in a region corresponding to the location of *MoSPC1* and *MoSPC7*. However, these two ORFs did not have any similarity to *MoSPC1* or *MoSPC7* (data not shown). BLASTP searches against the NCBI nr database showed that GGTG\_00636 and GGTG\_00730



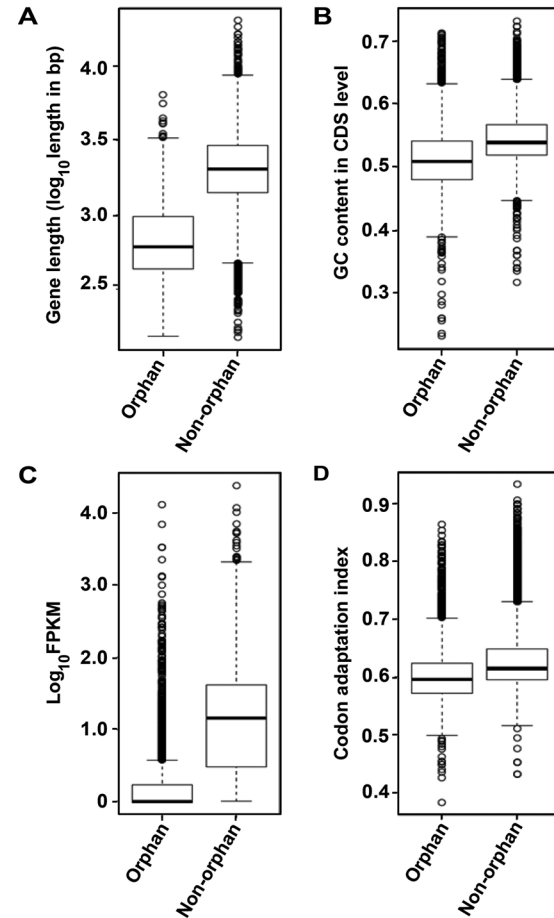


**Fig. 2.** Pathogenicity assay of wild-type,  $\Delta Mospc1$ ,  $\Delta Mospc2$ ,  $\Delta Mospc3$  and  $\Delta Mospc7$ . (A) Disease development after spraying conidial suspension onto rice leaves. Conidial suspension ( $1 \times 10^5$ /ml) was sprayed onto the leaves and diseased leaves were harvested 7 days after inoculation. (B) Disease score measurement was performed on 7-day diseased leaves of all strains, as described by Valent et al., 1991. The data are the mean  $\pm$  standard deviation of three independent experiments and all data are statistically analyzed using Tukey test ( $p < 0.05$ ).

were also likely orphan genes in the *G. graminis* genome. Based on our observations, at least one of the selected orphan genes (*MoSPC3*) originated *de novo* during the evolutionary history of *M. oryzae*.

## Discussion

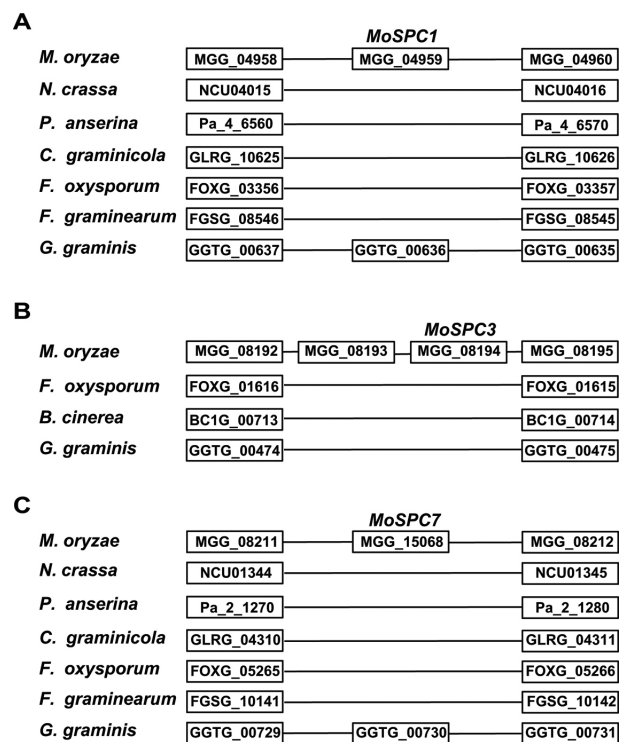
*De novo* gene birth is an important mechanism in species-specific adaptation processes (Cai et al., 2008; Kaessmann,



**Fig. 3.** Genomic features of new born genes in *M. oryzae*. (A) Gene length, (B) GC content in coding DNA sequence (CDS) level, (C) expression pattern and (D) codon adaptation between new and old genes using the Mann-Whitney Wilcoxon test ( $p < 2.2e-16$ ).

2010). Examples of newborn genes have shown that they quickly become essential and play pivotal roles during development, reproduction and survival (Ding et al., 2010; Li et al., 2010a). However, there have been no fully documented cases in plant pathogenic fungi regarding *de novo*-originated genes, although it is possible that such novel genes contribute to fungal pathogenesis, host specificity and host jumping.

To identify orphan genes in *M. oryzae*, we performed BLASTP searches against the NCBI nr database. To characterize orphan genes, it is important to understand whether a gene is absent in other species or whether the observed absence is the result of technical limitations of the methods used. This concern was addressed by a recent study showing that BLAST is sufficiently sensitive to detect the majority of remote homologues (Alba and Castresana, 2007). Furthermore, the NCBI nr database includes a variety of



**Fig. 4.** Synteny relationships of *MoSPC* genes with other fungal species. Protein sequences from the upstream and downstream flanking regions of *MoSPC* genes were used for BLASTP searches with a cutoff e-value of  $10^{-5}$  to identify homologous genes in other organisms. Gene birth of *M. oryzae* is shown in the upper region.

sequences, including those from metagenomic studies (Pruitt et al., 2007). A gene is considered to be absent in the genomes of other species if there is no match to the corresponding gene product in a BLASTP search against the NCBI nr database. Using this approach, we identified ~2,700 orphan genes that account for ~21% of genes annotated in the genome. Considering that orphan genes account for approximately 10-20% of total genes in other organisms (Khalturin et al., 2009), the observed proportion of orphans in the genome of *M. oryzae* was reasonable. Our analysis of orphan versus non-orphan genes showed that orphan genes of *M. oryzae* possess typical characteristics of genes formed via *de novo* evolution; namely, short length, low GC contents, low transcriptional activity, and less bias in codon usage.

In this report, we targeted four orphan genes of *M. oryzae* among ~2,700 orphan genes prioritized by *in planta* EST data and qRT-PCR to explore the involvement of orphan genes in fungal pathogenesis. Synteny analysis of these four genes indicated that at least one of these genes (*MoSPC3*) has a *de novo* origin. Our data suggest that the

other three genes may have originated *de novo*. It is highly likely that homologous genes of *MoSPC2* flanking genes are spread in other organisms during evolution, in contrast to *M. oryzae* in which these genes remain linked. Predicted ORFs of *G. graminis* (GGTG\_00636 and GGTG\_00730) are present in the corresponding location of *MoSPC1* and *MoSPC7* without sequence similarities, which is suggestive of two possibilities. First, the two genes are orthologous between *M. oryzae* and *G. graminis* but underwent rapid divergence after the speciation event. Second, each gene emerged *de novo* from non-coding sequences independently in both species. We argue that the second possibility is more parsimonious than the first considering the small amount of time since speciation events and that the divergence is significant enough to be undetectable using a BLAST search.

Deletion of individual genes (*MoSPC1*, *MoSPC2*, *MoSPC3* and *MoSPC7*) did not result in phenotypic changes for traits we examined in this study. This suggests that these four genes are not required for fungal development and pathogenesis. It is also possible that assays we performed to examine phenotypes were not sufficiently sensitive to detect subtle differences that exist between the mutants and wild-type. The absence of phenotypes in deletion mutants of an orphan gene was reported previously in *M. grisea*. The authors investigated a gene called *MIRI*, which is unique to *M. grisea* (Li et al., 2007). Deletion of *MIRI* did not cause changes in phenotypes examined. Based on its upregulation during plant infection, *MIRI* may be important for the fitness and virulence of *M. grisea* under field conditions, although it is dispensable for plant infection under laboratory conditions. This may also be true for the four orphan genes investigated in this study.

Alternatively, the absence of phenotypes could be related to evolutionary processes through which orphan genes are gained and retained in the genome. The *de novo* evolution model of orphan genes suggests a balance between gene emergence and gene loss over time (Palmieri et al., 2014). These genes can be retained if they confer selective advantages, otherwise they are quickly lost. Thus, we may have observed orphan genes that are being selected for retention or extinction through natural selection. In addition, orphan genes present in the majority of field isolates are more likely to have an impact on the fungal life cycle.

In summary, we investigated the role of orphan genes during pathogenesis of *M. oryzae* through gene deletions. Our data demonstrated that none of these genes are important for fungal development and pathogenicity, despite their upregulation during plant infection processes. We confirmed that at least one of the genes has a *de novo* origin

based on our analyses, including synteny comparison. These results suggest that a significant proportion of the fungal genome is comprised of orphan genes. These genes should be examined to determine their origin and increase our understanding of fungal pathogenesis and evolution.

### Acknowledgments

This work was supported by National Research Foundation of Korea grants funded by the Korean government (2008-0061897 and 2013-003196), and by the Next-Generation BioGreen 21 Program of the Rural Development Administration in Korea (PJ00821201). Sadat is grateful for a graduate fellowship through the Brain Korea 21 Program.

### References

- Alba, M. M. and Castresana, J. 2007. On homology searches by protein Blast and the characterization of the age of genes. *BMC Evol. Biol.* 7:53.
- Begun, D. J., Lindfors, H. A., Kern, A. D. and Jones, C. D. 2007. Evidence for *de novo* evolution of testis-expressed genes in the *Drosophila yakuba*/*Drosophila erecta* clade. *Genetics* 176:1131–1137.
- Cai, J., Zhao, R. P., Jiang, H. F. and Wang, W. 2008. *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* 179:487–496.
- Carvunis, A. R., Rolland, T., Wapinski, I., Calderwood, M. A., Yildirim, M. A., Simonis, N., Charlotiaux, B., Hidalgo, C. A., Barbette, J., Santhanam, B., Brar, G. A., Weissman, J. S., Regev, A., Thierry-Mieg, N., Cusick, M. E. and Vidal, M. 2012. Proto-genes and *de novo* gene birth. *Nature* 487:370–374.
- Chi, M. H., Park, S. Y. and Lee, Y. H. 2009. A quick and safe method for fungal DNA extraction. *Plant Pathol. J.* 25:108–111.
- Choi, J., Cheong, K., Jung, K., Jeon, J., Lee, G. W., Kang, S., Kim, S., Lee, Y. W. and Lee, Y. H. 2013. CFGP 2.0: a versatile web-based platform for supporting comparative and evolutionary genomics of fungi and Oomycetes. *Nucleic Acids Res.* 41:D714–D719.
- Conant, G. C. and Wolfe, K. H. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nature Rev. Genet.* 9:938–950.
- Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J. R., Pan, H. Q., Read, N. D., Lee, Y. H., Carbone, I., Brown, D., Oh, Y. Y., Donofrio, N., Jeong, J. S., Soanes, D. M., Djonovic, S., Kolomiets, E., Rehmeier, C., Li, W. X., Harding, M., Kim, S., Lebrun, M. H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L. J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J. E. and Birren, B. W. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434:980–986.
- Ding, Y., Zhao, L., Yang, S. A., Jiang, Y., Chen, Y. A., Zhao, R. P., Zhang, Y., Zhang, G. J., Dong, Y., Yu, H. J., Zhou, Q. and Wang, W. 2010. A Young *Drosophila* duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. *PLoS Genet.* 6:e1001255.
- Domazet-Loso, T. and Tautz, D. 2003. An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res.* 13:2213–2219.
- Donoghue, M. T. A., Keshavaiah, C., Swamidatta, S. H. and Spillane, C. 2011. Evolutionary origins of *Brassicaceae* specific genes in *Arabidopsis thaliana*. *BMC Evol. Biol.* 11:47.
- Ekman, D. and Elofsson, A. 2010. Identifying and quantifying orphan protein sequences in fungi. *J. Mol. Biol.* 396:396–405.
- Goff, S. A. 2005. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 309:879–879.
- Heinen, T. J. A. J., Staubach, F., Haming, D. and Tautz, D. 2009. Emergence of a new gene from an intergenic region. *Curr. Biol.* 19:1527–1531.
- Jansen, R. and Gerstein, M. 2000. Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. *Nucleic Acids Res.* 28:1481–1488.
- Kaessmann, H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20:1313–1326.
- Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. and Bosch, T. C. G. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25:404–413.
- Kim, H. J., Han, J. H., Kim, K. S. and Lee, Y. H. 2014. Comparative functional analysis of the velvet gene family reveals unique roles in fungal development and pathogenicity in *Magnaporthe oryzae*. *Fungal Genet. Biol.* 66:33–43.
- Kim, S., Park, J., Park, S. Y., Mitchell, T. K. and Lee, Y. H. 2010. Identification and analysis of *in planta* expressed genes of *Magnaporthe oryzae*. *BMC Genomics* 11:104.
- Knowles, D. G. and McLysaght, A. 2009. Recent *de novo* origin of human protein-coding genes. *Genome Res.* 19:1752–1759.
- Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A. and Begun, D. J. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. USA* 103:9935–9939.
- Li, C. Y., Zhang, Y., Wang, Z. B., Zhang, Y., Cao, C. M., Zhang, P. W., Lu, S. J., Li, X. M., Yu, Q., Zheng, X. F., Du, Q., Uhl, G. R., Liu, Q. R. and Wei, L. P. 2010a. A human-specific *de novo* protein-coding gene associated with human brain functions. *PLoS Comput. Biol.* 6:e1000734.
- Li, D., Dong, Y., Jiang, Y., Jiang, H. F., Cai, J. and Wang, W. 2010b. A *de novo* originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. *Cell Res.* 20:408–420.
- Li, L., Ding, S. L., Sharon, A., Orbach, M. and Xu, J. R. 2007. Mir1 is highly upregulated and localized to nuclei during infectious hyphal growth in the rice blast fungus. *Mol. Plant-Microbe Interact.* 20:448–458.
- Lynch, M. and Katju, V. 2004. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* 20:544–549.

- Martin, R., Moran, G. P., Jacobsen, I. D., Heyken, A., Domey, J., Sullivan, D. J., Kurzai, O. and Hube, B. 2011. The *Candida albicans*-specific gene *EED1* encodes a key regulator of hyphal extension. *PLoS ONE* 6:e18394.
- Neme, R. and Tautz, D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent *de novo* evolution. *BMC Genomics* 14:117.
- Palmieri, N., Kosiol, C. and Schlotterer, C. 2014. The life cycle of *Drosophila* orphan genes. *eLife* 3:e01311.
- Park, S. Y., Chi, M. H., Milgroom, M. G., Kim, H., Han, S. S., Kang, S. and Lee, Y. H. 2010. Genetic stability of *Magnaporthe oryzae* during successive passages through rice plants and on artificial medium. *Plant Pathol. J.* 26:313–320.
- Pozzoli, U., Menozzi, G., Fumagalli, M., Cereda, M., Comi, G. P., Cagliani, R., Bresolin, N. and Sironi, M. 2008. Both selective and neutral processes drive GC content evolution in the human genome. *BMC Evol. Biol.* 8:99.
- Pruitt, K. D., Tatusova, T. and Maglott, D. R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35:D61–D65.
- Sambrook J. and Russell D. W. (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Siepel, A. 2009. Darwinian alchemy: Human genes from noncoding DNA. *Genome Res.* 19:1693–1695.
- Talbot, N. J. 2003. On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. *Annu. Rev. Microbiol.* 57:177–202.
- Talbot, N. J., Ebbole, D. J. and Hamer, J. E. 1993. Identification and characterization of Mpg1, a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. *Plant Cell* 5:1575–1590.
- Tautz, D. and Domazet-Loso, T. 2011. The evolutionary origin of orphan genes. *Nature Rev. Genet.* 12:692–702.
- Valent, B. and Chumley, F. G. 1991. Molecular genetic analysis of the rice blast fungus, *Magnaporthe grisea*. *Annu. Rev. Phytopathol.* 29:443–467.
- Valent, B., Farrall, L. and Chumley, F. G. 1991. *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics* 127:87–101.
- Wilson, R. A. and Talbot, N. J. 2009. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Rev. Microbiol.* 7:185–195.
- Yang, Z. F. and Huang, J. L. 2011. De novo origin of new genes with introns in *Plasmodium vivax*. *FEBS Lett.* 585:641–644.
- Yu, J., Hu, S. N., Wang, J., Wong, G. K. S., Li, S. G., Liu, B., Deng, Y. J., Dai, L., Zhou, Y., Zhang, X. Q., Cao, M. L., Liu, J., Sun, J. D., Tang, J. B., Chen, Y. J., Huang, X. B., Lin, W., Ye, C., Tong, W., Cong, L. J., Geng, J. N., Han, Y. J., Li, L., Li, W., Hu, G. Q., Huang, X. G., Li, W. J., Li, J., Liu, Z. W., Li, L., Liu, J. P., Qi, Q. H., Liu, J. S., Li, L., Li, T., Wang, X. G., Lu, H., Wu, T. T., Zhu, M., Ni, P. X., Han, H., Dong, W., Ren, X. Y., Feng, X. L., Cui, P., Li, X. R., Wang, H., Xu, X., Zhai, W. X., Xu, Z., Zhang, J. S., He, S. J., Zhang, J. G., Xu, J. C., Zhang, K. L., Zheng, X. W., Dong, J. H., Zeng, W. Y., Tao, L., Ye, J., Tan, J., Ren, X. D., Chen, X. W., He, J., Liu, D. F., Tian, W., Tian, C. G., Xia, H. G., Bao, Q. Y., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W. M., Li, P., Chen, W., Wang, X. D., Zhang, Y., Hu, J. F., Wang, J., Liu, S., Yang, J., Zhang, G. Y., Xiong, Y. Q., Li, Z. J., Mao, L., Zhou, C. S., Zhu, Z., Chen, R. S., Hao, B. L., Zheng, W. M., Chen, S. Y., Guo, W., Li, G. J., Liu, S. Q., Tao, M., Wang, J., Zhu, L. H., Yuan, L. P. and Yang, H. M. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296:79–92.
- Yu, J. H., Hamari, Z., Han, K. H., Seo, J. A., Reyes-Dominguez, Y. and Scazzocchio, C. 2004. Double-joint PCR: a PCR-based molecular tool for gene manipulations in filamentous fungi. *Fungal Genet. Biol.* 41:973–981.
- Zhou, Q., Zhang, G. J., Zhang, Y., Xu, S. Y., Zhao, R. P., Zhan, Z. B., Li, X., Ding, Y., Yang, S. A. and Wang, W. 2008. On the origin of new genes in *Drosophila*. *Genome Res.* 18:1446–1455.