농학박사 학위논문

지의류 형성 곰팡이와 내생균의 비교유전체학적 분석

Comparative genomics of lichen-forming and endophytic fungi

2023년 2월

서울대학교 대학원

협동과정 농생명유전체학 전공

송 현 정

Comparative genomics of lichen-forming and endophytic fungi

A dissertation submitted in partial
fulfillment of the requirement for
the degree of

DOCTOR OF PHILOSOPHY

to the Faculty of
Interdisciplinary Program in Agricultural Genomics
at

SEOUL NATIONAL UNIVERSITY

by

Hyeunjeong Song

FEBRUARY 2023

농학박사 학위논문

지의류 형성 곰팡이와 내생균의 비교유전체학적 분석

지도교수 이 용 환

이 논문을 농학박사 학위논문으로 제출함
2022년 12월

서울대학교 대학원
협동과정 농생명유전체학 전공

송 현 정

송현정의 박사 학위논문을 인준함

2022년 12월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

A THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Comparative genomics of lichen-forming and endophytic fungi

UNDER THE DIRECTION OF DR. YONG-HWAN LEE

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

BY

HYEUNJEONG SONG

INTERDISCIPLINARY PROGRAM IN AGRICULTURAL GENOMICS

DECEMBER 2022

APPROVED AS A QUALIFIED THESIS OF HYEUNJEONG SONG
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN _____

VICE CHAIRMAN _____

MEMBER _____

MEMBER _____

MEMBER _____

ABSTRACT

Comparative genomics of lichen-forming and endophytic fungi

Hyeunjeong Song
Interdisciplinary Program in Agricultural Genomics
The Graduate School
Seoul National University

Symbiosis is one of the interactions between different organisms in the ecosystem in which one of them has benefits. Most plants are associated with symbiotic fungi to obtain water and nutrients more efficiently from the external environment. Fungal symbiosis which interact with plants is classified into mycorrhizal, lichen-forming, and endophytic fungi which interact with plant roots, algae, and entire plant tissue respectively. Although there have been limitations in understanding fungal symbiosis using molecular biology, the recent advancement of whole genome sequencing technology has enabled a comprehensive comparative genomics approach to evolutionary perspectives. For example, comparative genomics of a large number of mycorrhizal fungi genomes revealed their lifestyle has evolved with the loss of plant cell wall degrading enzymes and the gain of lineage-specific effector proteins. However, comparative genomics of lichen-

forming and endophytic fungi were poorly conducted compared with that of mycorrhizal fungi.

This study analyzed the genomic characteristics of symbiotic, lichen-forming, and endophytic fungi, in two chapters. In the first chapter, the genomes of six lichen-forming fungi were compared to find how their symbiotic relationships gained in evolutionary history. Whole genome sequences of lichen-forming fungi were not conserved, which indicates that six lichen-forming fungi were derived from different common ancestors. However, two lichen-forming fungi lineages have undergone similar changes in gene families such as gene family expansions of cytochrome P450 and gene family contractions of plant cell wall degrading enzymes, sugar transporters, and transcription factors. These results correspond with the formation of non-penetrating interfaces in fungi-algal interactions and the obtaining of photosynthetic products such as specific polyols from algal partners. In addition, time-series transcriptome profiles showed that lichen-forming fungi-specific and conserved genes were induced in the early and late stages of symbiosis establishment, respectively.

In chapter 2, two host-transited *Magnaporthe grisea* strains, JDJ2F and YHL-684, were isolated from rice showing an endophytic lifestyle. Even though two strains were isolated from rice, their genomic evidence indicated that they had originated from crabgrass. To reveal the host transition evidence of two strains, effector protein sequences were compared among *M. grisea* and *M. oryzae* sequences.

The repertories of effectors were not significantly different among *M. grisea* strains, but two *M. oryzae* effector genes showed a significant difference which could clarify the host transition evidence. One of the effector proteins, AVR-Pi9 in the *M. grisea* JDJ2F strain, has amino acid polymorphism similar to sequences of *M. oryzae* strains. Moreover, rice-infecting *M. oryzae*-specific AVR-Pik gene was observed in the *M. grisea* JDJ2F strain with a large number of transposable elements. Based on these results, we hypothesized that *M. grisea* strain JDJ2F is in an intermediate stage of the host transition process from crabgrass to rice and exhibit the endophytic lifestyle before they obtain virulence in rice.

This comparative genomic study aims to elucidate how lichen-forming fungi obtain and maintain their beneficial interaction and how endophytic fungi are evolved as an intermediate stage during host transition process. The results of this study will contribute to the understanding of fungal symbiosis and provide the basic clue for further molecular genetic studies of fungal symbiosis and their lifestyles.

Keywords: Fungal symbiosis, Comparative genomics, Fungal lifestyle, Lichen-forming fungi, Endophytic fungi, Host transition

Student number: 2014-21922

CONTENTS

LIST OF TABLES

page

CHAPTER I

CHAPTER II

LIST OF FIGURES

page

CHAPTER I

CHAPTER II

GENERAL INTRODUCTION

Most plants in the ecosystem have symbiotic relationships with microorganisms, and these associations emerged since the transition of plants from aquatic to terrestrial environments (Rai and Agarkar, 2016). Symbiosis has been widely described as mutualism, but the original definition of symbiosis is interactions between co-existed two organisms, from which at least one of the partners benefits (Das and Varma, 2009). Symbiotic associations are divided into mutualism, commensalism, and parasitism depending on their biological interactions (Lewis, 1985;Das and Varma, 2009). Mutualistic relations both organism benefits, which was successfully co-evolved (Martin et al., 2017), and commensalism is one partner benefit and the other partner have no influence, and parasitism is one partner benefit but the other partner has harmed. In fungi, mycorrhizal, endophytic, and lichen-forming fungi are colonized in plant roots, algae or cyanobacteria, and multiple tissues of plant respectively, showing beneficial relationships.

The mycorrhizal association is the most well-studied fungal symbiosis in both molecular and genomic approaches. Mycorrhizal fungi that existed in the soil establish beneficial associations by extending hyphae to the root of the host plant. Ectomycorrhizal fungi colonize the intercellular space of the root with a developing harting net, and endomycorrhizal fungi or arbuscular mycorrhizal fungi penetrate the epidermal cell and develop arbuscules inside the cells (Bonfante and Genre, 2010). Both types of mycorrhizal fungi and also endophytic fungi promote the growth of

the plant by transferring soil nutrients such as nitrogen and phosphorous to the plant and increase biotic or abiotic stress tolerance of the plant (Zamioudis and Pieterse, 2012;Trouvelot et al., 2015). Phylogenetic analysis and fossil evidence showed ancestor of mycorrhizal fungi and lichen-forming fungi appeared approximately 400 million years ago (MYA) (Taylor et al., 1995).

Another beneficial fungal association is lichen, which is composed of a fungal partner (mycobiont) and photosynthetic green algae or cyanobacteria (photobiont). Photosynthetic partners transfer the carbon sources converted to specific polyols, and fungi provide minerals, water, and stable shelter for photobionts (Nash, 2008;Behie and Bidochka, 2014). Recent studies revealed that other organisms such as bacteria, yeast, protest, and virus also be a component of lichen association (Spribille et al., 2016;Morillas et al., 2022). Unlike plant-associated mycorrhizal and endophytic fungi, lichen-forming fungi are the more dominant partners in their symbiosis. The structure of lichens is mainly composed of hyphae of fungal partners, and the name of lichens is dependent on the species name of fungal partner because one photobiont can be associated with diverse lichen fungal species (Greuter, 2000;Honegger, 2009). Lichenization is common among fungi, with approximately 21% of fungal species forming lichens, about 14,000 species were so far discovered. Most are belongs to Ascomycota (99.6%) including Lecanoromycetes, Eurotiomycetes, Ostropomycetidae, and Ostropomycetidae and only a few species belong to Basidiomycetes (0.4%) (Hawksworth and Hill, 1984).

Fungal species which associated with plants but do not shows any disease symptoms are defined as endophytic fungi (Wilson, 1995). Unlike mycorrhizal fungi

which colonize only in plant roots, endophytes could colonize within multiple tissues of the plants (Herrera et al., 2010). Not only promoting plant growth by nitrogen uptake and water supply, co-existing with endophytic fungi could prevent the invasion of pathogens (Yang et al., 2019 Gruber et al., 2012 Microbiology). Most fungal species with antifungal activity contains expanded carbohydrate-active enzymes for degrading the cell wall of pathogenic fungi (Yang et al., 2019). This biological property made endophytic fungi to attention as biocontrol agents. Many endophytic fungus also evolved from pathogenic ancestors (Xu et al., 2014), but it is known as an unstable lifestyle because they easily switched from other lifestyles such as pathogen or saprotroph (Delaye et al., 2013) depending on the genomic differences (Freeman and Rodriguez, 1993) or environmental changes (Rodriguez et al., 2005;Alvarez-Loayza et al., 2011;Kuo et al., 2014). Horizontal gene transfer (HGT) can lead to rapid lifestyle transitions through the gain and loss of virulence genes (Melnyk et al., 2019).

As genomic approaches have been commonly used through advanced next-generation sequencing methods, prior studies tried to reveal the evolutionary mechanisms of the emergence of fungal symbiotic lifestyles. *Laccaria bicolor* which has a mycorrhizal association with the root of trees is the first genome-sequenced symbiotic fungi (Martin et al., 2008). Genomic analysis of this mutualistic fungus revealed that mycorrhizal fungi have larger genome size and number of protein-coding genes than fungi with other lifestyles, and that small secreted proteins are involved in the establishment of symbiosis. Endophytic fungi (*Epichloe festucae*) and lichen-forming fungi (*Xanthoria parietina* and *Cladonia grayi*) were also first

sequenced in 2009 and 2011 respectively (Armaleo et al., 2019), but few genomes were reported compared to mycorrhizal fungi up to date. Recently, however, a large number of symbiotic fungal genomes were released to elucidate the evolutionary mechanism of the symbiotic association through comparative genomic analyses (Kohler et al., 2015;Miyauchi et al., 2020;Resl et al., 2022). Mycorrhizal symbiosis undergoes loss of repertories of plant cell wall degrading enzymes and gain of mycorrhizal-specific small secreted proteins when they evolved from non-symbiotic ancestors, and further molecular studies supported that small secreted proteins are involved in the establishment of mycorrhizal symbiosis (Kloppholz et al., 2011;Plett et al., 2011;Tsuzuki et al., 2016). However, genomic analysis of lichen-forming and endophytic fungi poorly understood compared to mycorrhizae and fungi with other lifestyles.

This study presents the genomic perspective on the evolution and establishment of fungal symbiotic associations focused on the mutualistic lichen-forming and endophytic fungi. We sequenced genomes of symbiotic fungal species and compared them with non-symbiotic fungal genomes to find how they obtained and maintained their beneficial associations. Lichen-forming fungi located in multiple origins have evolved to gain repertories of cytochrome P450 gene families and lost plant cell wall degrading enzymes, transcription factors, and sugar transporter gene families. Moreover, lichen-specific genes including small secreted proteins were induced in the early stage for the establishment of lichen symbiosis, whereas other genes were involved in the late stage. In another chapter, we identified rice-isolated *Magnaporthe grisea* strains that originally infected crabgrass, and they showed an

endophytic lifestyle in the isolated host. We hypothesized that rice-isolated *M. grisea* strains are in the preadaptation stage of the host transition from crabgrass to rice, but have not yet obtained virulence in rice despite being able to colonize in rice tissues. Comparative genomic analyses showed evidence of host transition in *M. oryzae* effector genes AVR-Pi9 and AVR-Pik. These comparative resources provide new insights into fungal symbiotic associations and help the comprehensive understanding of plant-microbe interactions along with pathogenic and saprotrophic lifestyles.

LITERATURE CITED

Alvarez-Loayza, P., White Jr, J.F., Torres, M.S., Balslev, H., Kristiansen, T., Svenning, J.-C., and Gil, N. (2011). Light converts endosymbiotic fungus to pathogen, influencing seedling survival and niche-space filling of a common tropical tree, *Iriartea deltoidea*. *PLoS One* **6**: e16386.

Armaleo, D., Müller, O., Lutzoni, F., Andrésson, S., Blanc, G., Bode, H.B., Collart, F.R., Dal Grande, F., Dietrich, F., and Grigoriev, I.V. (2019). The lichen symbiosis re-viewed through the genomes of *Cladonia grayi* and its algal partner *Asterochloris glomerata*. *BMC Genome.* **20**: 1-33.

Behie, S.W., and Bidochka, M.J. (2014). Nutrient transfer in plant–fungal symbioses. *Trends Plant Sci.* **19**: 734-740.

Bonfante, P., and Genre, A. (2010). Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat. Commun.* **1**: 48.

Das, A., and Varma, A. (2009). "Symbiosis: the art of living," in Symbiotic Fungi. Springer, 1-28.

Delaye, L., García-Guzmán, G., and Heil, M. (2013). Endophytes versus biotrophic and necrotrophic pathogens—are fungal lifestyles evolutionarily stable traits? *Fungal Divers.* **60**: 125-135.

Freeman, S., and Rodriguez, R.J. (1993). Genetic conversion of a fungal plant pathogen to a nonpathogenic, endophytic mutualist. *Science* **260**: 75-78.

Greuter, W. (2000). International code of botanical nomenclature (Saint Louis Code). *Regnum Veg.* **138**: 1-474.

Hawksworth, D.L., and Hill, D.J. (1984). The lichen-forming fungi. New York: Blackie

Honegger, R. (2009). "Lichen-forming fungi and their photobionts," in Plant relationships. Springer, 307-333.

Kloppholz, S., Kuhn, H., and Requena, N. (2011). A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr. Biol.* **21**: 1204-1209.

Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canback, B., Choi, C., Cichocki, N., Clum, A., Colpaert, J., Copeland, A., Costa, M.D., Dore, J., Floudas, D., Gay, G., Girlanda, M., Henrissat, B., Herrmann, S., Hess, J., Hogberg, N., Johansson, T., Khouja, H.R., Labutti, K., Lahrmann, U., Levasseur, A., Lindquist, E.A., Lipzen, A., Marmeisse, R., Martino, E., Murat, C., Ngan, C.Y., Nehls, U., Plett, J.M., Pringle, A., Ohm, R.A., Perotto, S., Peter, M., Riley, R., Rineau, F., Ruytinx, J., Salamov, A., Shah, F., Sun, H., Tarkka, M., Tritt, A., Veneault-Fourrey, C., Zuccaro, A., Mycorrhizal Genomics Initiative, C., Tunlid, A., Grigoriev, I.V., Hibbett, D.S., and Martin, F. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47**: 410-415.

Kuo, H.-C., Hui, S., Choi, J., Asiegbu, F.O., Valkonen, J., and Lee, Y.-H. (2014). Secret lifestyles of *Neurospora crassa*. *Sci. Rep.* **4**: 1-6.

Lewis, D. (1985). “Symbiosis and mutualism: crisp concepts and soggy semantics” in Biology of mutualism, Boucher DH eds. Oxford University Press. 400 p.

Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., and Pereda, V. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88-92.

Martin, F.M., Uroz, S., and Barker, D.G. (2017). Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.

Melnyk, R.A., Hossain, S.S., and Haney, C.H. (2019). Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated *Pseudomonas*. *ISME J.* **13**: 1575-1588.

Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sanchez-Garcia, M., Morin, E., Andreopoulos, B., Barry, K.W., Bonito, G., Buee, M., Carver, A., Chen, C., Cichocki, N., Clum, A., Culley, D., Crous, P.W., Fauchery, L., Girlanda, M., Hayes, R.D., Keri, Z., Labutti, K., Lipzen, A., Lombard, V., Magnuson, J., Maillard, F., Murat, C., Nolan, M., Ohm, R.A., Pangilinan, J., Pereira, M.F., Perotto, S., Peter, M., Pfister, S., Riley, R., Sitrit, Y., Stielow, J.B., Szollosi, G., Zifcakova, L., Stursova, M., Spatafora, J.W., Tedersoo, L., Vaario, L.M., Yamada, A., Yan, M., Wang, P., Xu, J., Bruns, T., Baldrian, P., Vilgalys, R., Dunand, C., Henrissat, B., Grigoriev, I.V., Hibbett, D., Nagy, L.G., and Martin, F.M. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat. Commun*. **11**: 5125.

Morillas, L., Roales, J., Cruz, C., and Munzi, S. (2022). Lichen as multipartner symbiotic relationships. *Encyclopedia* **2**: 1421-1431.

Nash, T.H. (2008). Lichen biology. Cambridge ; New York: Cambridge University Press.

Plett, J.M., Kemppainen, M., Kale, S.D., Kohler, A., Legue, V., Brun, A., Tyler, B.M., Pardo, A.G., and Martin, F. (2011). A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr. Biol.* **21**: 1197-1203.

Rai, M., and Agarkar, G. (2016). Plant–fungal interactions: what triggers the fungi to switch among lifestyles? *Crit. Rev. Microbiol.* **42**: 428-438.

Resl, P., Bujold, A.R., Tagirdzhanova, G., Meidl, P., Freire Rallo, S., Kono, M., Fernández-Brime, S., Guðmundsson, H., Andrésson, S., and Muggia, L. (2022). Large differences in carbohydrate degradation and transport potential among lichen fungal symbionts. *Nat. Commun.* **13**: 1-13.

Rodriguez, R.J., Redman, R.S., and Henson, J.M. (2005). Symbiotic lifestyle expression by fungal endophytes and the adaptation of plants to stress: unraveling the complexities of intimacy. *Mycology* **23**: 683.

Spribille, T., Tuovinen, V., Resl, P., Vanderpool, D., Wolinski, H., Aime, M.C., Schneider, K., Stabentheiner, E., Toome-Heller, M., Thor, G., Mayrhofer, H., Johannesson, H., and Mccutcheon, J.P. (2016). Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **353**: 488-492.

Taylor, T.N., Remy, W., Hass, H., and Kerp, H. (1995). Fossil arbuscular mycorrhizae from the early Devonian. *Mycologia* **87**: 560-573.

Trouvelot, S., Bonneau, L., Redecker, D., Van Tuinen, D., Adrian, M., and Wipf, D. (2015). Arbuscular mycorrhiza symbiosis in viticulture: a review. *Agron. Sustain. Dev.* **35**: 1449-1467.

Tsuzuki, S., Handa, Y., Takeda, N., and Kawaguchi, M. (2016). Strigolactone-induced putative secreted protein 1 is required for the establishment of symbiosis by the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Mol. Plant Microbe Interact*. **29**: 277-286.

Wilson, D. (1995). Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos*, 274-276.

Xu, X.-H., Su, Z.-Z., Wang, C., Kubicek, C.P., Feng, X.-X., Mao, L.-J., Wang, J.-Y., Chen, C., Lin, F.-C., and Zhang, C.-L. (2014). The rice endophyte *Harpophora oryzae* genome reveals evolution from a pathogen to a mutualistic endophyte. *Sci. Rep.* **4**: 1-9.

Yang, Y., Liu, X., Cai, J., Chen, Y., Li, B., Guo, Z., and Huang, G. (2019). Genomic characteristics and comparative genomics analysis of the endophytic fungus *Sarocladium brachiariae*. *BMC genom.* **20**: 1-20.

Zamioudis, C., and Pieterse, C.M. (2012). Modulation of host immunity by beneficial microbes. *Mol. Plant Microbe Interact*. **25**: 139-150.

CHAPTER I

A comparative genomic analysis of lichen-forming fungi reveals new insights into fungal lifestyles

This chapter was published in Scientific Reports.

Song, H. et al. (2022) *Sci. Rep.* 12: 1-14.

ABSTRACT

Lichen-forming fungi are mutualistic symbionts of green algae or cyanobacteria. We report the comparative analysis of six genomes of lichen-forming fungi in classes Eurotiomycetes and Lecanoromycetes to identify genomic information related to their symbiotic lifestyle. The lichen-forming fungi exhibited genome reduction via the loss of dispensable genes encoding plant-cell-wall-degrading enzymes, sugar transporters, and transcription factors. The loss of these genes reflects the symbiotic biology of lichens, such as the absence of pectin in the algal cell wall and obtaining specific sugars from photosynthetic partners. The lichens also gained many lineage- and species-specific genes, including those encoding small secreted proteins. These genes are primarily induced during the early stage of lichen symbiosis, indicating their significant roles in the establishment of lichen symbiosis. Our findings provide comprehensive genomic information for six lichen-forming fungi and novel insights into lichen biology and the evolution of symbiosis.

INTRODUCTION

Lichens exists in symbiosis, in which at least one fungus (mycobiont) lives in a mutually beneficial relationship with photosynthetic algae and/or cyanobacteria (photobiont) (Nash, 2008;Grube and Wedin, 2016). Since this dual nature was discovered by Schwendener in 1867 (Honegger, 2000), numerous studies have demonstrated that basidiomycetes yeast (Millanes et al., 2016;Spribille et al., 2016;Černajová and Škaloud, 2019;Tuovinen et al., 2019), as well as diverse microbiomes (Hawksworth and Grube, 2020), may cohabitate within lichen thalli. In lichen association, dominant fungal partners which produce basic morphological structure of lichens are determine the classification of lichens. The thallus structure composed of the fungal component retained the water for drought tolerance in extreme conditions (Honegger, 2006;Kranner et al., 2008) as well as has a role as a shelter; protecting the photobionts from the external environment (Nash, 2008). Moreover, the algal partner synthesizes carbohydrate products by photosynthesis and transfers this carbon source to the fungal partner to maintain the lichen association (Sanders and Masumoto, 2021).

Lichenization is common among fungi, with approximately 21% of fungal species forming lichens (Nash, 2008). Lichenized symbiosis is not derived from a single phylogenetic clade (Gargas et al., 1995), but found in the Ascomycota classes Lecanoromycetes, Eurotiomycetes, Leotiomycetes, Dothideomycetes, and Arthoniomycetes, as well as in a few Basidiomycota classes (Nash, 2008;Grube and Wedin, 2016). Previous phylogenetic studies have suggested that lichenization evolved

independently at least five times in distantly related lineages (Gargas et al., 1995). Such studies have also demonstrated that lichenization has been continuously maintained from the common ancestor of Lecanoromycetes, but was lost during the evolution of Lecanoromycetes. Due to this complex evolutionary history, many hypotheses have been proposed to account for the evolutionary time required for lichenization and its loss and re-evolution (Nelsen et al., 2020).

To date, many studies have been conducted to elucidate the symbiotic nature of lichens. The successful re-association of lichen symbionts under laboratory conditions has facilitated microscopic observations of the fungal-algal interface during lichen establishment (Ahmadjian et al., 1978;Athukorala et al., 2014). Thus, the early stages of lichenization, which ranges from 'pro-contact' to 'growth together', have been well investigated (Trembley et al., 2002;Joneson et al., 2011;Athukorala et al., 2014); however, its late stages remain poorly understood. Although the aim of many studies is to identify symbiosis-related genes, until recently, we lacked the genetic transformation tools required to perform gene manipulation in lichen biology (Park et al., 2013a;Liu et al., 2021). Thus, recent molecular studies have applied genetic transformation systems to elucidate lichen symbiosis.

However, the slow growth of several lichens and the difficulty of their culture in the laboratory have further required the development of genomic-level studies to gain an evolutionary understanding of lichen symbiosis. Genomics have advanced greatly since the sequencing of *Xanthoria parietina* (Armstrong et al., 2018;Dal Grande et al., 2018;Wang et al., 2018;Armaleo et al., 2019;Pizarro et al., 2019;Mead and Gueidan, 2020). Numerous studies have approached lichen symbiosis from a genomic perspective to

identify evolutionary process of lichenization and symbiosis-related genes. *Endocarpon pusillum* was the first lichen to have been subjected to genomic analysis; early studies reported its symbiosis-related genes involved in nitrogen/sugar transport and metabolism with their expression during the re-synthesis stages (Wang et al., 2014). Although continuous genomic studies investigating the key factors of lichen symbiosis (Wang et al., 2014;Armaleo et al., 2019;Pizarro et al., 2019;Pizarro et al., 2020), recent descriptions of several additional genome sequences (Resl et al., 2021), and the application of systems biology approach to lichen associations (Nazem-Bokaee et al., 2021) improve the knowledge of lichen symbiotic systems but determining how a symbiotic lifestyle evolved remains challenging. Mycorrhizal fungi, which are mutualistic symbionts associated with >90% of land plants, have been studied extensively to identify their symbiotic nature. Large-scale genomic sequencing of mycorrhizal fungi has revealed that convergent evolution occurred via the loss of plant cell wall-degrading enzymes (PCWDEs) and the enrichment of transposable elements (TEs) and mycorrhiza-induced small secreted proteins (MiSSPs) (Kohler et al., 2015;Pellegrin et al., 2015;Miyauchi et al., 2020). Several molecular studies have also reported that secreted proteins play a crucial role in mycorrhizal symbiotic associations (Kloppholz et al., 2011;Plett et al., 2011;Tsuzuki et al., 2016).

In this study, we aim to conduct a comparative genomic analysis of four Lecanoromycetes species (*Gyalolechia flavorubescens* [Park et al., 2013c], *Cladonia macilenta* [Park et al., 2013b], *Cladonia metacorallifera* [Park et al., 2014b], and *Umbilicaria muehlenbergii* [Park et al., 2014a]) and two Eurotiomycetes species (*Endocarpon pusillum* Z07020 [Wang et al., 2014] and R61883 [Park et al., 2014c]) of

lichen-forming fungi, which are evolutionary distant species. An additional 50 genomes from fungi with diverse lifestyles, including mycorrhizal fungi and close relatives of lichen-forming fungi, were also used to support the identification of the lichen-specific genomic features. We were going to perform a gene family gain/loss analyses in comparison with non-lichenized fungi to identify specific gene families of lichen-forming fungi. Finally, we re-synthesized *G. flavorubescens* and its algal partner *Trebouxia gelatinosa* and performed a time-series transcriptomic analysis of this re-synthesized lichen through RNA sequencing to reveal unique features of lichen symbiosis.

MATERIALS AND METHODS

I. Genome resources of fungal species and ortholog clustering

We sequenced genome of five lichen-forming fungi including *G. flavorubescens* KoLRI002931 (accession no. AUPK01000000) (Park et al., 2013c), *C. macilenta* KoLRI003786 (AUPP01000000) (Park et al., 2013b), *C. metacorallifera* KoLRI002260 (AXCT02000000) (Park et al., 2014b), *U. muehlenbergii* KoLRILF000956 (JFDN01000000) (Park et al., 2014a) and *E. pusillum* R61883 (JFDM01000000) (Park et al., 2014c). *E. pusillum* Z07020 (APWS00000000) (Wang et al., 2014) in previous researches and the other 50 fungal genomes used for comparative analysis were downloaded from Broad Institute (http://www.broadinstitute.org/), JGI fungal genome portal MycoCosm (http://jgi.doe.gov/fungi) and NCBI-GenBank database. The predicted protein sequences from 56 fungal genomes were clustered by OrthoFinder v2.2.7 with the default program settings (Emms and Kelly, 2015). The conserved and lichen-specific genes were annotated by GO term annotation using InterProScan version 60 (Jones et al., 2014).

II. Phylogenetic analysis and divergence time estimation

Total proteins of 56 fungal genomes were used to construct a whole genome-based phylogenomic tree using CVtree3 with k-tuple 7 (Zuo and Hao, 2015). Initial curation of the divergence time for the major fungal taxa was achieved by Timetree (Hedges et al., 2015), and the divergence times were estimated by MCMCtree in PAML package version 4.8 (Yang, 2007) using molecular markers, including actin (ACT1), translation elongation

factor EF1-α (TEF1), RNA polymerase II large subunits (RPB1 and RPB2) and β-tubulins (TUB1 and TUB2). The final phylogenomic tree with divergence times was visualized by MEGA version 7.0.26 (Kumar et al., 2016).

III. Repetitive sequence and whole genome synteny analysis

The repeat contents were analyzed using TRF and rmBlastN, parts of RepeatMasker v4.0.5 package with RepBase 21.05 fungi library (Smit et al., 2013). For pairwise genomic comparisons, MUMmer v3.23 (Kurtz et al., 2004) was used for aligning and comparing the whole genomes between lichen-forming fungi and the other fungal genomes.

IV. Gene family evolution analysis and gene family annotation

CAFE (Computational analysis of gene family evolution) v2 was used to find out the gene families with significant changes in size ($P < 0.01$) (Kurtz et al., 2004). The time-calibrated phylogenetic tree and gene families identified by ortholog clustering were used for this analysis. Functional annotations of expanded and contracted gene families were identified by domain-based InterProScan v60 (Jones et al., 2014). The cytochrome P450 genes were firstly identified with Fungal Cytochrome P450 Database (FCPD) (Park et al., 2008), and then BLAST analysis against the P450 database in David Nelson cytochrome P450 web site (Nelson, 2009) for nomenclature. The secondary metabolite biosynthesis gene clusters, including PKS, NRPS, and DMATs were identified by SMURF (Fedorova et al., 2012). Candidate MFS transporters were obtained using the Transporter Classification Database (TCDB) (Saier et al., 2016). Diverse polyol and monosaccharide

transporters of *G. flavorubescens* were predicted by BLAST search (identity > 30, query coverage > 60) with functionally characterized transporter genes listed in (Yoshino et al., 2019). The transcription factors were predicted by InterProScan v60 using the previously annotated DNA-binding domains (Shelest, 2017), and PCWDE encoding CAZyme families were annotated by HMMER search against dbCAN CAZyme domain HMM database (Yin et al., 2012). The secretomes and SSPs were predicted by the method described previously (Kim et al., 2016). The maximum-likelihood phylogenetic trees of CYP, PKS, PCWDEs, homeodomain-like and helix-turn-helix psq TF genes were constructed using RAxML version 8.2.9 with a bootstrap value of 1000 (Stamatakis, 2014). Aligning protein sequences using ClustalW 2.1 (Larkin et al., 2007) and remove poorly aligned regions by trimAl v1.2 (Capella-Gutierrez et al., 2009) were preceded before phylogenetic analysis. We reconciled the gene trees of PCWDEs resulting from this analysis with the species tree using NOTUNG 2.9 (Darby et al., 2017).

IV. RNA extraction and expression analysis

Actively growing *G. flavorubescens* KoLRI002931 mycelia were collected and macerated the mycelia into 10 mL of sterilized distilled water using a homogenizer (Ika, T10 basic, German). The macerated mycelia were dropped on malt extract agar medium (Difco), incubated at 15 ℃ for 4-6 weeks, and then covered 50 μL of 2 weeks old *Trebouxia gelatinosa* cell suspension (1×10^8/mL) which is partner alga of *G. flavorubescens*. The plates were incubated at 15 °C without light. For harvesting samples at different time points during re-synthesis between *G. flavorubescens* and *T. gelatinosa*, all samples were

collected 0 h, 12 h, 24 h, 48 h, 72 h, 4 weeks, and 6 weeks after re-synthesis, immediately frozen using liquid nitrogen, and stored at -80℃ until processed. The whole samples on the medium were collected from three replicates of three biological repeats except 4 and 6 weeks. Total RNA was extracted using an Easy-Spin Total RNA Extraction Kit (iNtRON Biotechnology, Seoul, Korea). RNA sequencing was performed at Macrogen Inc. (Seoul, Korea) using Illumina HiSeq platform. The NGS QC Toolkit ver. 2.3.3 (Patel and Jain, 2012) was used to remove adaptors, low-quality sequences and sequences containing more than 5% N to obtain clean reads. Since the sequences of *G. flavorubescens* and *T. gelatinosa*, the partner alga, were mixed in the clean reads, the algal reads were eliminated using BWA (0.7.9a-r786) (Li and Durbin, 2009). Then the paired-end clean reads were aligned to *G. flavorubescens* genome using TopHat v2.0.12 (Kim et al., 2013) and the gene expression levels were calculated as FPKM (Fragment Per Kilobase of transcript per Million mapped reads) using cufflink v.2.2.1 (Trapnell et al., 2010) and cuffdiff v.2.21 (Trapnell et al., 2013). The FPKM value of 0h, 12 h, 24 h, 48 h, and 72 h PCI were calculated with three biological repeats, and 4 weeks and 6 weeks PCI were calculated without repeats. Fold changes were calculated simply using a modified function, $\log2([FPKM_{SYMBIOSIS}+1]/[FPKM_{MYCELIA}+1])$. The hieratical clustering of protein expressions in the heatmaps was performed using the Euclidean clustering distance by Morpheus run by Broad institute (https://software.broadinstitute.org/morpheus). Significantly enriched GO terms in symbiosis induced genes were identified using R package topGO version 2.38.1 with threshold $p<0.05$ (Alexa and Rahnenfuhrer, 2016). KEGG pathway mapping analysis using differentially expressed genes was performed by KEGG Automatic Annotation Server (KAAS) web sites (Moriya et al., 2007).

RESULTS

I. Phylogenomic relationships and genomic similarity among lichen-forming fungi

The Lecanoromycetes and Eurotiomycetes lichen-forming fungi have similar genome sizes, ranging from 34.5 to 37.3 Mb, and similar numbers of genes (8,294 to 9,695) (Table 1). We conducted phylogenomic analysis of the six lichen-forming fungi including 50 fungal genome sequences (Table 2). The phylogenomic tree showed that the four Lecanoromycetes species (*G. flavorubescens*, *C. macilenta*, *C. metacorallifera*, and *U. muehlenbergii*) and the two *E. pusillum* isolates were distantly related (Figure 1A). This finding is consistent with previous reports that lichenization events evolved independently in multiple lineages (Gargas et al., 1995;Lutzoni et al., 2001;Nash, 2008). The time-calibrated phylogeny suggests that Lecanoromycetes diverged approximately 258 million years ago from an ancestral fungus that may have been lichen-forming, and that divergence between *E. pusillum* isolates and the plant pathogen *Phaeomoniella chlamydospora* occurred approximately 52 million years ago.

We analyzed the synteny of lichen-forming fungal genomes using *C. macilenta* as a reference. Dot plots revealed that both *Cladonia* species had a robust syntenic relationship, with several inverted blocks (Figure 1B); however, as the evolutionary distance of species from *C. macilenta* increased, the syntenic relationship weakened. Because *C. macilenta* and *E. pusillum* belong to different classes, their syntenic relationship is nearly random, despite their both being lichen-forming fungi. The syntenic region of the two *Cladonia*

species had 65.7–66.5% similarity, whereas the syntenic similarities between *C. macilenta* and other lichen-forming *G. flavorubescens* and *U. muehlenbergii* in Lecanoromycetes were 6.1–6.6% and 6.8–7.3%, respectively (Table 3), and the two *Endocarpon* species in Eurotiomycetes had 3.4% and 3.5% similarity compared with *C. macilenta*.

Repetitive sequence content was also analyzed in lichen-forming fungi. The simple repeat content account for approximately 1% of all lichen-forming fungal genomes, but most DNA transposons were observed only in *E. pusillum* R61883, rather than other lichen-forming fungi (Figure 1C). The portion of retroelements differ among lichen-forming fungi, Lecanoromycetes fungi (less than 1%), *E. pusillum* (1.5%), and *U. muehlenbergii* (4%). The total composition of repeat sequences in lichen-forming fungi was lower than that of other fungal species (Figure 2).

Table 1. Genome statistics of the lichen-forming fungi

Lichen species	Genome size (Mb)	Number of Scaffolds	Number of genes	GC content (%)	Reference
Endocarpon pusillum R61883	37.13	59	9,252	48.54%	Park et al., 2014c
Endocarpon pusillum Z07020	37.33	308	9,238	45.56%	Wang et al., 2014
Gyalolechia flavorubescens	34.47	36	9,695	41.79%	Park et al., 2013c
Umbilicaria muehlenbergii	34.81	7	8,294	46.82%	Park et al., 2014a
Cladonia metacorallifera	36.68	30	9,030	43.81%	Park et al., 2014b
Cladonia macilenta	37.12	240	8,773	42.85%	Park et al., 2013b

Table 2. Genome information of fungal species used in this study

Species name	Abbr.	Total length (Mb)	Scaffold number	GC contents (%)	Gene number	Phylum	Subphylum	Class	Lifestyle	Genome Reference
Microsporum canis	Mcan	23.24	16	47.26	8,765	Ascomycota	Pezizomycotina	Eurotiomycetes	Animal pathogen	Martinez *et al.*, 2012
Trichophyton rubrum	Trub	22.50	35	47.45	10,418	Ascomycota	Pezizomycotina	Eurotiomycetes	Animal pathogen	Martinez *et al.*, 2012
Coccidioides immitis	Cimm	28.95	6	46.01	9,910	Ascomycota	Pezizomycotina	Eurotiomycetes	Animal pathogen	Neafsey *et al.*, 2010
Blastomyces dermatitidis	Bder	66.57	25	36.89	11,539	Ascomycota	Pezizomycotina	Eurotiomycetes	Animal pathogen	Munoz *et al.*, 2015
Histoplasma capsulatum	Hcap	32.99	275	42.83	9,313	Ascomycota	Pezizomycotina	Eurotiomycetes	Animal pathogen	Sharpton *et al.*, 2009
Aspergillus fumigatus	Afum	29.38	8	48.82	9,630	Ascomycota	Pezizomycotina	Eurotiomycetes	Animal pathogen	Nierman *et al.*, 2005
Aspergillus nidulans	Anid	30.24	90	50.03	9,561	Ascomycota	Pezizomycotina	Eurotiomycetes	Saprotroph	Galagan *et al.*, 2005
Endocarpon pusillum R61883	EpusR	37.13	59	48.54	9,252	Ascomycota	Pezizomycotina	Eurotiomycetes	Lichen	Park *et al.*, 2014
Endocarpon pusillum Z07020	EpusZ	37.33	908	45.56	9,238	Ascomycota	Pezizomycotina	Eurotiomycetes	Lichen	Wang *et al.*, 2014
Phaeomoniella chlamydospora	Pchl	27.51	702	47.01	6,907	Ascomycota	Pezizomycotina	Eurotiomycetes	Plant pathogen	Morales-Cruz *et al.*, 2015
Caloplaca flavorubescens	Cfla	34.47	36	41.79	9,695	Ascomycota	Pezizomycotina	Lecanoromycetes	Lichen	Park *et al.*, 2013
Umbilicaria muehlenbergii	Umue	34.81	7	46.82	8,294	Ascomycota	Pezizomycotina	Lecanoromycetes	Lichen	Park *et al.*, 2014
Cladonia metacorallifera	Cmet	36.68	30	43.81	9,030	Ascomycota	Pezizomycotina	Lecanoromycetes	Lichen	Park *et al.*, 2014
Cladonia macilenta	Cmac	37.12	240	42.85	8,773	Ascomycota	Pezizomycotina	Lecanoromycetes	Lichen	Park *et al.*, 2013
Cenococcum geophilum	Cgeo	177.56	268	37.48	14,709	Ascomycota	Pezizomycotina	Dothideomycetes	Mycohrrizae	Peter *et al.*, 2016
Mycosphaerella graminicola	Mgra	39.69	21	52.13	10,963	Ascomycota	Pezizomycotina	Dothideomycetes	Plant pathogen	Goodwin *et al.*, 2011
Blumeria graminis	Bgra	118.73	6,843	32.55	6,495	Ascomycota	Pezizomycotina	Leotiomycetes	Plant pathogen	Spanu *et al.*, 2010
Botrytis cinerea	Bcin	42.66	588	39.15	16,447	Ascomycota	Pezizomycotina	Leotiomycetes	Plant pathogen	Staats *et al.*, 2012
Oidiodendron maius	Omai	46.43	100	46.93	16,702	Ascomycota	Pezizomycotina	Leotiomycetes	Mycohrrizae	Kohler *et al.*, 2015
Ascocoryne sarcoides	Asar	34.17	219	46.38	10,672	Ascomycota	Pezizomycotina	Leotiomycetes	Endophyte	Gianoulis *et al.*, 2012
Podospora anserina	Pans	35.01	7	51.79	10,588	Ascomycota	Pezizomycotina	Sordariomycetes	Saprotroph	Espagne *et al.*, 2008
Neurospora crassa	Ncra	41.04	20	48.20	10,813	Ascomycota	Pezizomycotina	Sordariomycetes	Saprotroph	Galagan *et al.*, 2003

Magnaporthe oryzae	Mory	41.03	8	51.51	12,991	Ascomycota	Pezizomycotina	Sordariomycetes	Plant pathogen	Dean *et al*., 2005
Colletotrichum graminicola	Cgra	51.60	653	48.42	12,006	Ascomycota	Pezizomycotina	Sordariomycetes	Plant pathogen	O'Connell *et al*., 2012
Beauveria bassiana	Bbas	33.69	235	51.36	10,364	Ascomycota	Pezizomycotina	Sordariomycetes	Endophyte	Xiao *et al*., 2012
Fusarium oxysporum	Foxy	61.39	114	47.28	27,347	Ascomycota	Pezizomycotina	Sordariomycetes	Plant pathogen	Ma *et al*., 2010
Fusarium graminearum	Fgra	36.46	31	48.04	13,313	Ascomycota	Pezizomycotina	Sordariomycetes	Plant pathogen	Cuomo *et al*., 2007
Tuber melanosporum	Tmel	124.95	398	44.35	7,496	Ascomycota	Pezizomycotina	Pezizomycetes	Mycohrrizae	Martin *et al*., 2010
Schizosaccharomyces pombe	Spom	12.57	3	36.06	5,132	Ascomycota	Taphrinomycotina	Schizosaccharomycetes	Saprotroph	Wood *et al*., 2002
Candida albicans	Calb	27.56	413	33.44	14,217	Ascomycota	Saccharomycotina	Saccharomycetes	Animal pathogen	Jones *et al*., 2004
Saccharomyces cerevisiae	Scer	12.07	16	38.30	5,888	Ascomycota	Saccharomycotina	Saccharomycetes	Saprotroph	Foury *et al*., 1998
Puccinia graminis	Pgra	88.64	392	39.87	15,979	Basidomycota	Pucciniomycotina	Pucciniomycetes	Plant pathogen	Duplessis *et al*., 2011
Melampsora laricis-populina	Mlar	101.13	62	39.60	16,372	Basidomycota	Pucciniomycotina	Pucciniomycetes	Plant pathogen	Duplessis *et al*., 2011
Ustilago maydis	Umay	19.66	27	53.97	6,783	Basidomycota	Ustilaginomycotina	Ustilaginomycetes	Plant pathogen	Kämper *et al*., 2006
Cryptococcus neoformans	Cneo	18.89	14	48.23	7,826	Basidomycota	Agaricomycotina	Tremellomycetes	Animal pathogen	Janbon *et al*., 2014
Tulasnella calospora	Tcal	62.39	1,335	42.90	19,554	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Sebacina vermifera	Sver	38.09	546	45.56	15,245	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Piriformospora indica	Pind	24.98	1,884	50.25	11,791	Basidomycota	Agaricomycotina	Agaricomycetes	Endophyte	Zuccaro *et al*., 2011
Laccaria amethystina	Lame	52.58	1,299	45.71	17,553	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Laccaria bicolor	Lbic	60.71	55	46.00	23,132	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Martin *et al*., 2008
Hebeloma cylindrosporum	Hcyl	38.23	176	45.82	15,376	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Amanita muscaria	Amus	40.70	1,101	41.86	18,093	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Suillus luteus	Slut	41.74	649	47.03	16,744	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Suillus brevipes	Sbre	52.03	1,550	47.44	21,458	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Branco *et al*., 2015
Pisolithus tinctorius	Ptin	71.01	610	42.43	22,653	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Pisolithus microcarpus	Pmic	53.03	1,064	43.81	20,982	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Scleroderma citrinum	Scit	56.14	938	45.67	20,995	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Paxillus involutus	Pinv	58.30	2,681	42.27	17,984	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015
Paxillus adelphus	Pade	64.46	1,671	45.47	18,999	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al*., 2015

Serpula lacrymans	Slac	42.80	46	44.90	12,925	Basidomycota	Agaricomycotina	Agaricomycetes	Saprotroph	Eastwood *et al*., 2011
Piloderma croceum	Pcro	59.33	715	40.81	21,524	Basidomycota	Agaricomycotina	Agaricomycetes	Mycohrrizae	Kohler *et al.*, 2015
Heterobasidion irregulare	Hirr	33.65	15	52.23	13,275	Basidomycota	Agaricomycotina	Agaricomycetes	Plant pathogen	Olson *et al.*, 2012
Phanerochaete chrysosporium	Pchr	35.15	232	52.55	13,602	Basidomycota	Agaricomycotina	Agaricomycetes	Saprotroph	Ohm *et al.*, 2014
Phycomyces blakesleeanus	Pbla	53.94	80	35.40	16,528		Mucoromycotina		Saprotroph	Corrochano *et al*., 2016
Rhizopus oryzae	Rory	46.09	81	34.97	17,467		Mucoromycotina		Saprotroph	Ma *et al*., 2009
Rhizophagus irregularis	Rirr	90.30	28,033	27.47	29,830	Glomeromycota		Glomeromycetes	Mycohrrizae	Tisserant *et al*., 2013

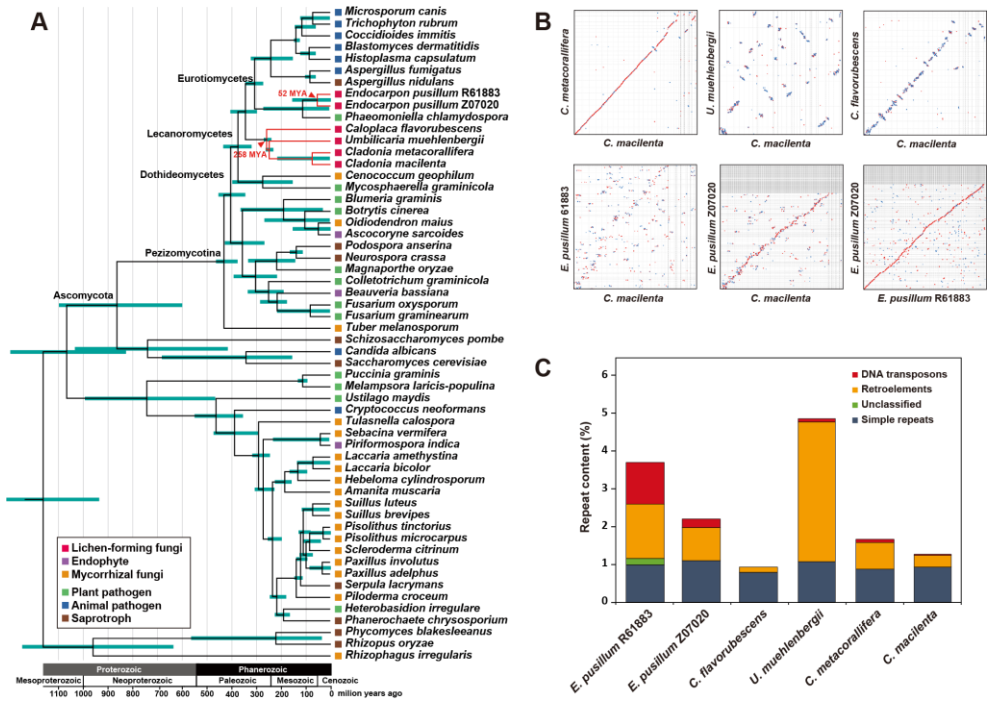


Figure 1. Phylogenomic and syntenic relationships among lichen-forming fungi and their repeat contents

(A) The phylogenomic tree shows that lichen-forming fungi are derived from many ancestors. Red branches indicate lichen-forming fungal lineages, and colored squares indicate lifestyle of a fungal species. The scale of the phylogenomic tree is millions of years, as calculated using the mcmctree function in the Phylogenetic Analysis by Maximum Likelihood software package. Blue error bars at each node indicate 95% highest posterior density (HPD) for node age. (B) Synteny dot plots of lichen-forming fungi. Red (blue) dots indicate forward (reverse) matches. (C) Repetitive sequence contents of lichen-forming fungi identified using RepeatMasker software. DNA transposons, retroelements, and unclassified repeats are classes of interspersed repeats.

Table 3. Conserved synteny regions between lichen-forming fungi

	Overlapping Length (bp)		Overlapping Ratio (%)	
	Query	Reference	Query	Reference
C. macilenta - C. metacorallifera	24,397,165	24,403,396	65.7%	66.5%
C. macilenta - U. muehlenbergii	2,537,858	2,549,399	6.8%	7.3%
C. macilenta - G. flavorubescens	2,274,687	2,276,745	6.1%	6.6%
C. macilenta - E. pusillum R61883	1,306,448	1,300,376	3.5%	3.5%
C. macilenta - E. pusillum Z07020	1,277,251	1,272,538	3.4%	3.4%
E. pusillum R61883 –E. pusillum Z07020	18,978,499	18,989,224	50.8%	51.1%

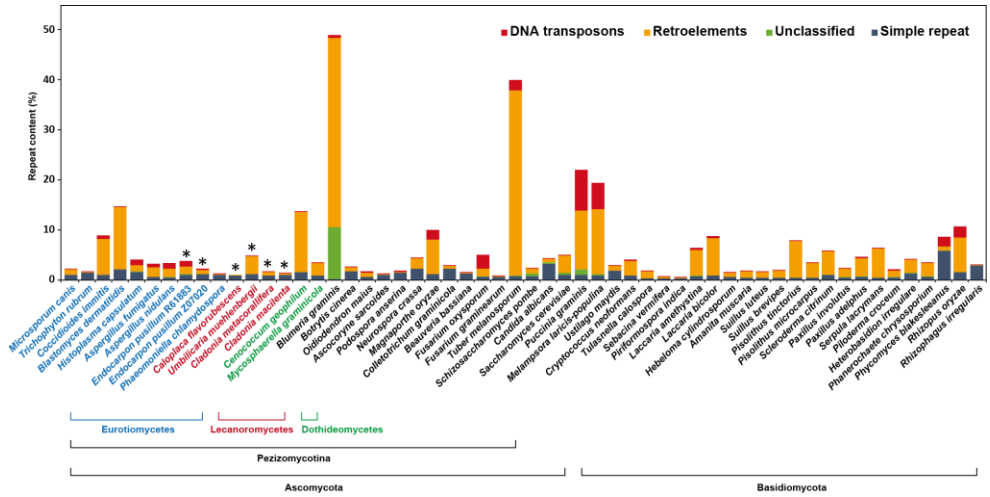


Figure 2. Repeat contents of 56 fungal species

Repetitive sequences of 56 fungal species analyzed by RepeatMasker (Smit et al., 2013). Each colored bars represent a class of repeats. The species name in blue represents Eurotiomycetes, red represents Lecanoromycetes, and green represents Dothideomycetes. The bar graphs with an asterisk indicate the lichen-forming fungi.

II. Gene family expansion and contraction during the evolution of lichen-forming fungi

Gene family expansion and contraction were analyzed based on orthologous genes across 56 fungal species including six lichen-forming fungi (Figure 3). We estimated changes in gene family size when the two lichen-forming fungal clades Lecanoromycetes and *E. pusillum* diverged from different non-lichenized common ancestors. In Lecanoromycetes, 106 families expanded and 3,049 contracted. Among the *E. pusillum* isolates, 238 families expanded and 886 contracted. Contractions were dominant in the lichen lineages, leading to a small total gene number in the lichen-forming fungi (Table 2). In both lichen-forming fungal clades, the cytochrome P450 (CYP) family expanded, whereas the glycoside hydrolase (GH), transcription factor (TF), and major facilitator superfamily (MFS) contracted (Table 4).

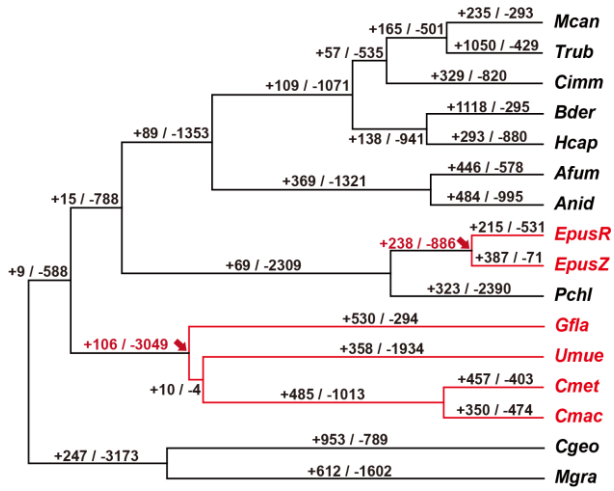


Figure 3. Gene family evolution in lichen-forming fungi and their relatives

Estimation of gene family expansion and contraction in lichen-forming fungi using the CAFE computational tool ($P > 0.01$). Red arrows indicate branch points where lichen-forming fungi diverged from non-lichenized ancestors. + and – indicate the numbers of expanded and contracted gene families, respectively. Only 16 species belonging to the Lecanoromyces, Eurotiomyces, and Dotidomyces closely related to lichen-forming fungi were shown. Species abbreviations: EpusR, *Endocarpon pusillum* R61883; EpusZ, *E. pusillum* Z07020; Gfla, *Gyalolechia flavorubescens*; Umue, *Umbilicaria muehlenbergii*; Cmet, *Cladonia metacorallifera*; Cmac, *Cladonia macilenta*. Abbreviations of other species names are provided in Table 2.

Table 4. Commonly expanded and contracted gene families in lineages of lichen-forming fungi

Interpro ID	Term	Number of changed gene families in both lichen-forming fungi clades
Expanded		
IPR001128	Cytochrome P450	3
Contracted		
IPR013781	Glycoside hydrolase, catalytic domain	16
IPR007219	Transcription factor domain, fungi	12
IPR011701	Major facilitator superfamily	10
IPR001138	Zn(2)-C6 fungal-type DNA-binding domain	10
IPR005828	Major facilitator, sugar transporter-like	7

III. Loss of plant cell wall degrading enzymes (PCWDEs) in lichen associations

In both pathogens and symbionts, PCWDEs play essential roles in plant host cell wall remodeling for fungal colonization (Lionetti and Metraux, 2014). However, gene family expansion and contraction analysis (Figure 4A) and the profiles of carbohydrate-active enzyme (CAZyme) genes (Figure 5) revealed a remarkable reduction of PCWDEs in six lichen-forming fungi. Plants and green algae have similar cell wall components, such as cellulose and hemicellulose, whereas pectin is unique to land plants and Charophycean green algae (Popper et al., 2011). Lichen-forming fungi have fewer CAZyme genes involved in PCWDEs compared with plant-associated fungi, and a similar number compared with animal pathogens (Figure 5). Almost all polysaccharide lyase (PL) family genes, which act mainly in pectin degradation, have been lost in lichen-forming fungi. Only a few genes acting on cellulose (Auxiliary activity [AA] family 9, GH5, and GH3), hemicellulose (GH5, GH27, GH31, Carbohydrate esterase [CE] family 1, GH2, GH43, and GH3), and pectin (CE1, GH2, GH43, and GH3) are conserved in lichen-forming fungi (Figure 4A). The number of PCWDE genes is dramatically decreased among lichen-forming fungi; decreasing patterns of PCWDE genes have been similarly observed in ectomycorrhizal fungi, which cannot penetrate host plant cell walls during colonization (Kohler et al., 2015).

We conducted gene tree–species tree reconciliation analysis to further infer the evolutionary relationships of PCWDE genes in lichen-forming fungi and their relatives (Figure 4B to D). Lichen-forming fungi belonging to Lecanoromycetes lost

15 cellulose-degrading enzyme genes from their ancestral gene repertory (Figure 4B). *E. pusillum* underwent two steps of gene loss: Chaetothyriomycetidae lost 6 genes, and then *Endocarpon* lost 12 cellulose-degrading enzyme genes. The plant pathogen *P. chlamydospora*, which is also a member of Chaetothyriomycetidae, regained several PCWDE genes subsequent to their loss in Chaetothyriomycetidae. The propensity of gene loss pattern in hemicellulose and pectin genes is similar to that of cellulose (Figure 4C and D). The Eurotiomycetes lineage, which comprises only animal pathogens, also underwent massive loss of PCWDE genes whereas Leotiomycetes and Sordariomycetes, which include many plant-associated fungi, gained repertoires of PCWDEs for host invation. These results indicate that most genes related to the degradation of cellulose, hemicellulose, and pectin have been lost in lichen-forming fungi, but that these event occurred independently during the evolution of lichenization in different evolutionary lineages.

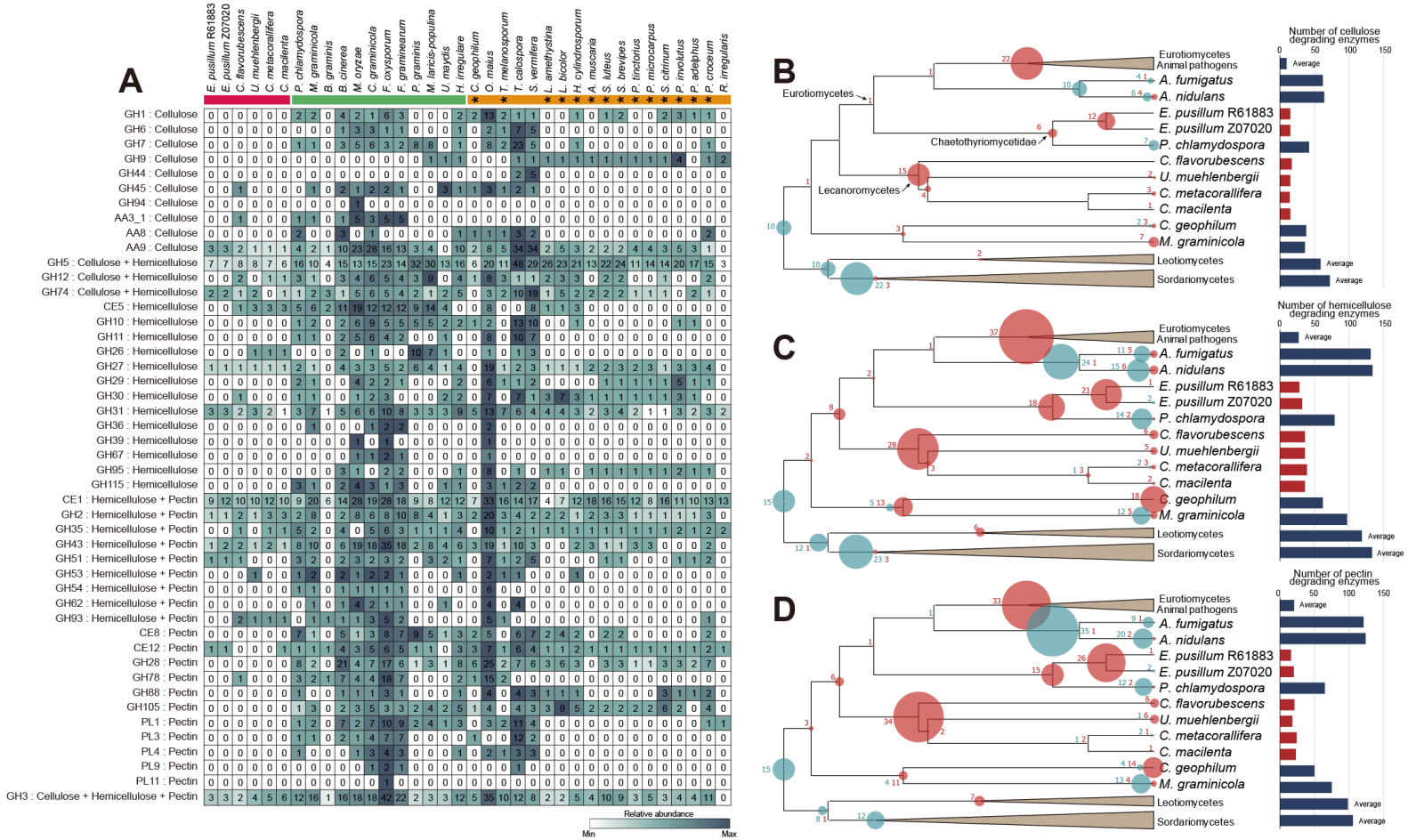

A
E. pusillum R61883
E. pusillum Z07020
C. flavorubescens
U. muehlenbergii
C. metacorallifera
C. macilenta
P. chlamydospora
M. graminicola
B. graminis
B. cinerea
M. oryzae
C. graminicola
F. oxysporum
F. graminearum
P. graminis
M. laricis-populina
U. maydis
H. irregulare
C. geophilum
O. maius
T. melanosporum
T. calcispora
S. vermifera
L. amethystina
L. bicolor
H. cylindrosporum
A. muscaria
S. luteus
S. brevipes
P. tinctorius
P. microcarpus
S. citrinum
P. involutus
P. adelphus
P. croceum
R. irregularis
GH1 : Cellulose
GH6 : Cellulose
GH7 : Cellulose
GH9 : Cellulose
GH44 : Cellulose
GH45 : Cellulose
GH94 : Cellulose
AA3_1 : Cellulose
AA8 : Cellulose
AA9 : Cellulose
GH5 : Cellulose + Hemicellulose
GH12 : Cellulose + Hemicellulose
GH74 : Cellulose + Hemicellulose
CE5 : Hemicellulose
GH10 : Hemicellulose
GH11 : Hemicellulose
GH26 : Hemicellulose
GH27 : Hemicellulose
GH29 : Hemicellulose
GH30 : Hemicellulose
GH31 : Hemicellulose
GH36 : Hemicellulose
GH39 : Hemicellulose
GH67 : Hemicellulose
GH95 : Hemicellulose
GH115 : Hemicellulose
CE1 : Hemicellulose + Pectin
GH2 : Hemicellulose + Pectin
GH35 : Hemicellulose + Pectin
GH43 : Hemicellulose + Pectin
GH51 : Hemicellulose + Pectin
GH53 : Hemicellulose + Pectin
GH54 : Hemicellulose + Pectin
GH62 : Hemicellulose + Pectin
GH93 : Hemicellulose + Pectin
CE8 : Pectin
CE12 : Pectin
GH28 : Pectin
GH78 : Pectin
GH88 : Pectin
GH105 : Pectin
PL1 : Pectin
PL3 : Pectin
PL4 : Pectin
PL9 : Pectin
PL11 : Pectin
GH3 : Cellulose + Hemicellulose + Pectin
Relative abundance
Min
Max
B
Eurotiomycetes
Chaetothyriomycetidae
Lecanoromycetes
Eurotiomycetes
Animal pathogens
A. fumigatus
A. nidulans
E. pusillum R61883
E. pusillum Z07020
P. chlamydospora
C. flavorubescens
U. muehlenbergii
C. metacorallifera
C. macilenta
C. geophilum
M. graminicola
Leotiomycetes
Sordariomycetes
Number of cellulose degrading enzymes
0
50
100
150
Average
Average
Average
C
Number of hemicellulose degrading enzymes
D
Number of pectin degrading enzymes

Figure 4. Loss of plant cell wall-degrading enzymes (PCWDEs) in lichen-forming fungi for association with algal partners

(A) Distribution of selected carbohydrate-active enzyme (CAZyme) families related to cellulose, hemicellulose, and pectin degradation among lifestyles. Red, green, and orange boxes indicate lichen-forming fungi, plant pathogens, and mycorrhizal fungi, respectively. Asterisk indicates ectomycorrhizal fungi, a class of mycorrhizal fungi. The distribution of PCWDEs in lichen-forming fungi was compared with plant pathogens known to have a large number of PCWDEs for pathogenicity and mycorrhizal fungi known to be associated with symbiosis formations. CAZyme family abbreviations: GH, glycoside hydrolase; AA, auxiliary activities; CE, carbohydrate esterase; PL, polysaccharide lyase. Gene gain and loss analysis of (B) cellulose, (C) hemicellulose, and (D) pectin-degrading CAZyme families through species tree–gene tree reconciliation. Blue (red) circles indicate the number of gene gains (losses). Bubble size indicates the number of genes gained or lost. Bar graph indicates the total number of genes encoding PCWDEs in each species.

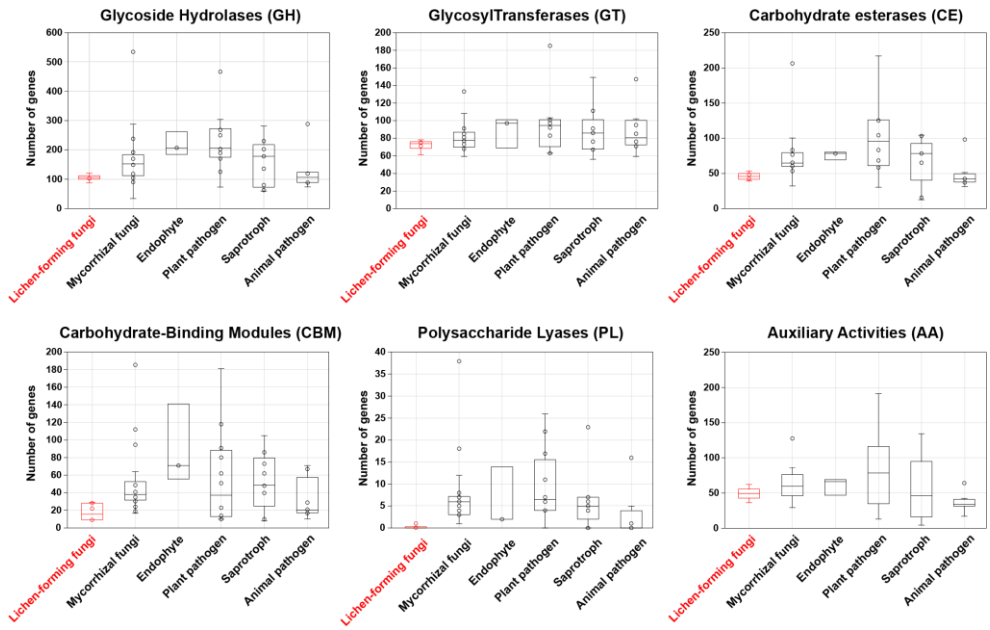


Figure 5. Distribution of CAZyme genes among diverse fungal lifestyles

Comparative analysis of CAZyme genes families among fungal lifestyles. In order from the top left glycoside hydrolases, glycosyltransferases, carbohydrate esterases, carbohydrate-binding modules, polysaccharide lyases, and auxiliary activities classes of CAZymes. The red box plot indicates lichen-forming fungi.

IV. Loss of sugar transporters during lichenization

The MFS is the largest family of secondary transporters related to the movement of diverse solutes, especially sugar uptake (Pao et al., 1998). However, MFS-type transporters underwent extensive contraction in six lichen-forming fungi, including the sugar porter (2.A.1), anion:cation symporter (2.A.1.14), aromatic acid:H+ symporter (2.A.1.15), and siderophore-iron transporter (2.A.1.16) families (Figure 6A). Because the type of sugar alcohols in symbiosis depends on the photosynthetic partners (Richardson et al., 1968), we further characterized the sugar transporters in lichen-forming fungi using dataset of *G. flavorubescens* expression during lichen resynthesis. A previous study defined 1 day post co-inoculation (PCI) as the 'pre-contact' stage of lichen fungi and algal partners, followed by 8 days PCI as the 'contact' stage, and 21 days PCI as the 'growth together' stage (Athukorala et al., 2014). We measured gene expression during the early (12, 24, 48, and 72 h) and late (4 and 6 weeks) stages after lichen resynthesis. During resynthesis, the expression levels of four ribitol transporter genes were high at 4–6 weeks PCI, whereas those of other transporter genes were low (Figure 6B). These findings are consistent with the reception of ribitol sugar alcohols by *G. flavorubescens* from its algal partner *Trebouxia* spp. (Richardson et al., 1968;Hill and Ahmadjian, 1972), and suggest that despite the extensive contraction of MFS-type transporters in lichen-forming fungi, these transporters may play important roles from 72 h to 4 weeks of lichenization.

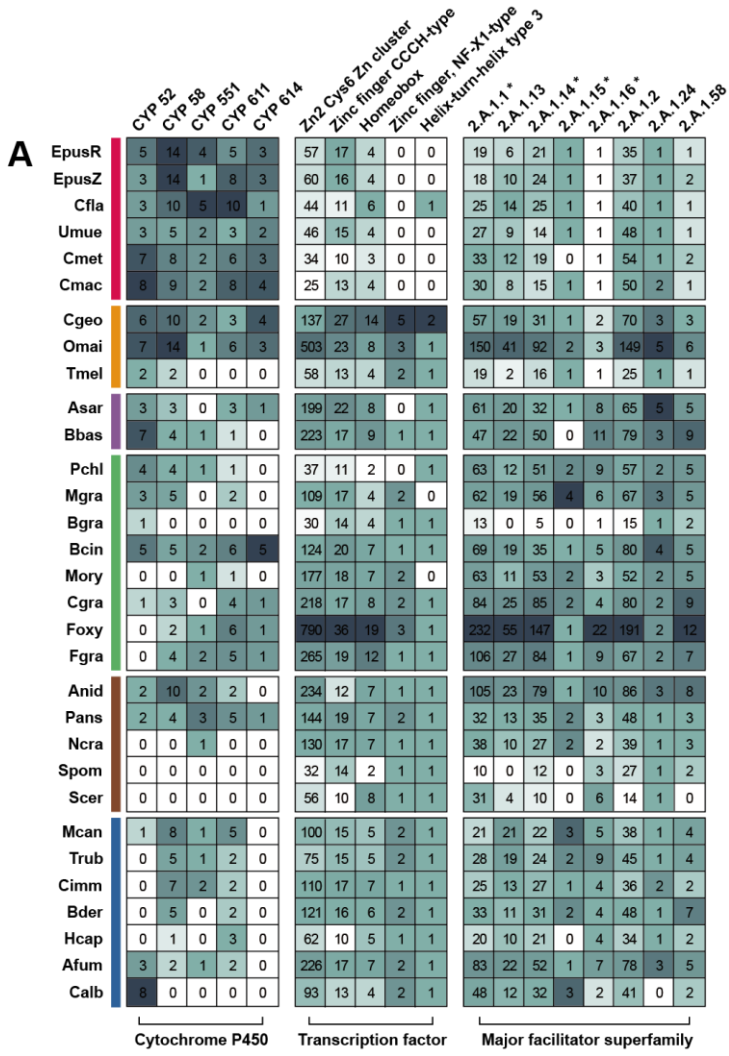

A
CYP 52
CYP 58
CYP 551
CYP 611
CYP 614
Zn2 Cys6 Zn cluster
Zinc finger CCCH-type
Homeobox
Zinc finger, NF-X1-type
Helix-turn-helix type 3
2.A.1.1*
2.A.1.13
2.A.1.14*
2.A.1.15*
2.A.1.16*
2.A.1.2
2.A.1.24
2.A.1.58
EpusR 5 14 4 5 3 57 17 4 0 0 19 6 21 1 1 35 1 1
EpusZ 3 14 1 8 3 60 16 4 0 0 18 10 24 1 1 37 1 2
Cfla 3 10 5 10 1 44 11 6 0 1 25 14 25 1 1 40 1 1
Umue 3 5 2 3 2 46 15 4 0 0 27 9 14 1 1 48 1 1
Cmet 7 8 2 6 3 34 10 3 0 0 33 12 19 0 1 54 1 2
Cmac 8 9 2 8 4 25 13 4 0 0 30 8 15 1 1 50 2 1
Cgeo 6 10 2 3 4 137 27 14 5 2 57 19 31 1 2 70 3 3
Omai 7 14 1 6 3 503 23 8 3 1 150 41 92 2 3 149 5 6
Tmel 2 2 0 0 0 58 13 4 2 1 19 2 16 1 1 25 1 1
Asar 3 3 0 3 1 199 22 8 0 1 61 20 32 1 8 65 5 5
Bbas 7 4 1 1 0 223 17 9 1 1 47 22 50 0 11 79 3 9
Pchl 4 4 1 1 0 37 11 2 0 1 63 12 51 2 9 57 2 5
Mgra 3 5 0 2 0 109 17 4 2 0 62 19 56 4 6 67 3 5
Bgra 1 0 0 0 0 30 14 4 1 1 13 0 5 0 1 15 1 2
Bcin 5 5 2 6 5 124 20 7 1 1 69 19 35 1 5 80 4 5
Mory 0 0 1 1 0 177 18 7 2 0 63 11 53 2 3 52 2 5
Cgra 1 3 0 4 1 218 17 8 2 1 84 25 85 2 4 80 2 9
Foxy 0 2 1 6 1 790 36 19 3 1 232 55 147 1 22 191 2 12
Fgra 0 4 2 5 1 265 19 12 1 1 106 27 84 1 9 67 2 7
Anid 2 10 2 2 0 234 12 7 1 1 105 23 79 1 10 86 3 8
Pans 2 4 3 5 1 144 19 7 2 1 32 13 35 2 3 48 1 3
Ncra 0 0 1 0 0 130 17 7 1 1 38 10 27 2 2 39 1 3
Spom 0 0 0 0 0 32 14 2 1 1 10 0 12 0 3 27 1 2
Scer 0 0 0 0 0 56 10 8 1 1 31 4 10 0 6 14 1 0
Mcan 1 8 1 5 0 100 15 5 2 1 21 21 22 3 5 38 1 4
Trub 0 5 1 2 0 75 15 5 2 1 28 19 24 2 9 45 1 4
Cimm 0 7 2 2 0 110 17 7 1 1 25 13 27 1 4 36 2 2
Bder 0 5 0 2 0 121 16 6 2 1 33 11 31 2 4 48 1 7
Hcap 0 1 0 3 0 62 10 5 1 1 20 10 21 0 4 34 1 2
Afum 3 2 1 2 0 226 17 7 2 1 83 22 52 1 7 78 3 5
Calb 8 0 0 0 0 93 13 4 2 1 48 12 32 3 2 41 0 2
Cytochrome P450
Transcription factor
Major facilitator superfamily
B
Log2 fold change
4
2
0
-2
-4
-6
-8
L-arabinose, ribitol, L-arabitol
sorbitol, mannitol, ribitol, arabitol
glucose, fructose
glycerol
Others
12h
24h
48h
72h
4 week
6 week

Figure. 6. Distribution of expanded and contracted gene families in lichen-forming fungi

(A) Comparative analysis of selected families among the cytochrome P450 (CYP) and transcription factor (TF) families and major facilitator superfamily (MFS), which expanded or contracted in lichen-forming fungi. Colored boxes indicate fungal species lifestyles. Asterisk indicates rapidly contracted gene families. (B) Expression patterns of diverse polyol transporters in *G. flavorubescens*.

V. Massive contraction of transcription factor (TF) genes implies streamlined lichen-forming fungal genomes

Gene family expansion and contraction analysis revealed that the Zn2 cys6 Zn cluster DNA-binding domain, which is a fungal-specific TF family, was reduced in independent lineages of lichen-forming fungi (Table 4 and Figure 6A). We found that most other TF gene families had also contracted in lichen-forming fungi (Figure 7), particularly the zinc finger CCCH-type (IPR000571) and homeobox (IPR001356) families (Figure 6A). The zinc finger, NF-X1 type (IPR000967), and helix-turn-helix type 3 (IPR001387) TF-type DNA-binding domains were not detected in the six lichen fungal genomes we analyzed. These losses in TF families led to the small number of TF genes in lichen-forming fungi compared with those of fungi with different lifestyles. Although the number of TF genes depends on the total number of proteins (Shelest, 2017), lichen-forming and mycorrhizal fungi have fewer TF genes than expected (Figure 8A). The Zn2 cys6 Zn cluster, zinc finger C2H2-type, and homeodomain-like DNA-binding domains are major contributors to the total number of TFs (Shelest, 2017); therefore we normalized these TF genes according to the total number of genes (Figure 8B to D). However, only the Zn2 cys6 Zn cluster TF genes were responsible for the low number of TF genes in lichen-forming fungi, because percentage of zinc finger C2H2-type and homeodomain-like DNA binding domains were similar in lichen-forming fungi and in fungi with other lifestyles. Mycorrhizal fungi had similar distributions, suggesting that the reduction in TFs due to contraction of the Zn2 cys6 Zn cluster occurred in lichen-forming fungi and other symbionts.

Only *E. pusillum* R61883 underwent possible duplication in specific TF families, including homeodomain-like (IPR009057) and helix-turn-helix psq (IPR007889) families (Figure 7 and Figure 9A). We hypothesize that transposons and transposition of DNA-mediated genes (GO:0006313) particularly abundant in this sample may have caused these duplications (Figure 9A). Because DNA transposons were near (3 kb) the duplicated homeodomain-like and helix-turn-helix psq families, we hypothesize that the expansion of TF families was influenced by repeat elements (Figure 9B).

Figure 7. Transcription factor (TF) gene families in 56 fungal species

Total TF gene families are classified by DNA-binding domains, which are mostly found in fungal species (Shelest, 2017). The white color indicates the minimum number of genes and the blue color indicates the maximum number of genes. The part shown in the red line is the result of lichen-forming fungi. Black asterisk means massively contracted TF families in lichen-forming fungi, and the red asterisk indicates duplicated TF families only in *E. pusillum* R61883.

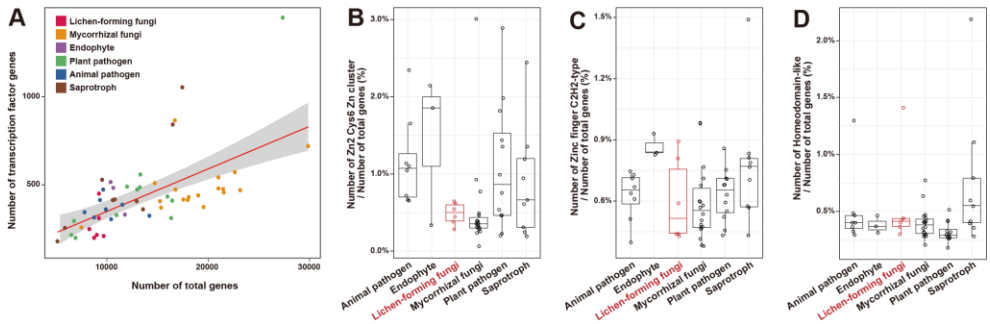


Figure 8. Major contributors of TF gene family contraction in lichen-forming fungi

(A) Correlation between the number of TF genes and the total number of genes. Red and grey lines are regression and error lines, respectively. Colored dots indicate fungal species lifestyles. (B–D) The three major TF families known to affect the total TF size. Red boxes indicate the distribution of lichen-forming fungi. (B) Zn2 Cys6 Zn cluster; (C) zinc finger C2H2-type; and (D) homeodomain-like.

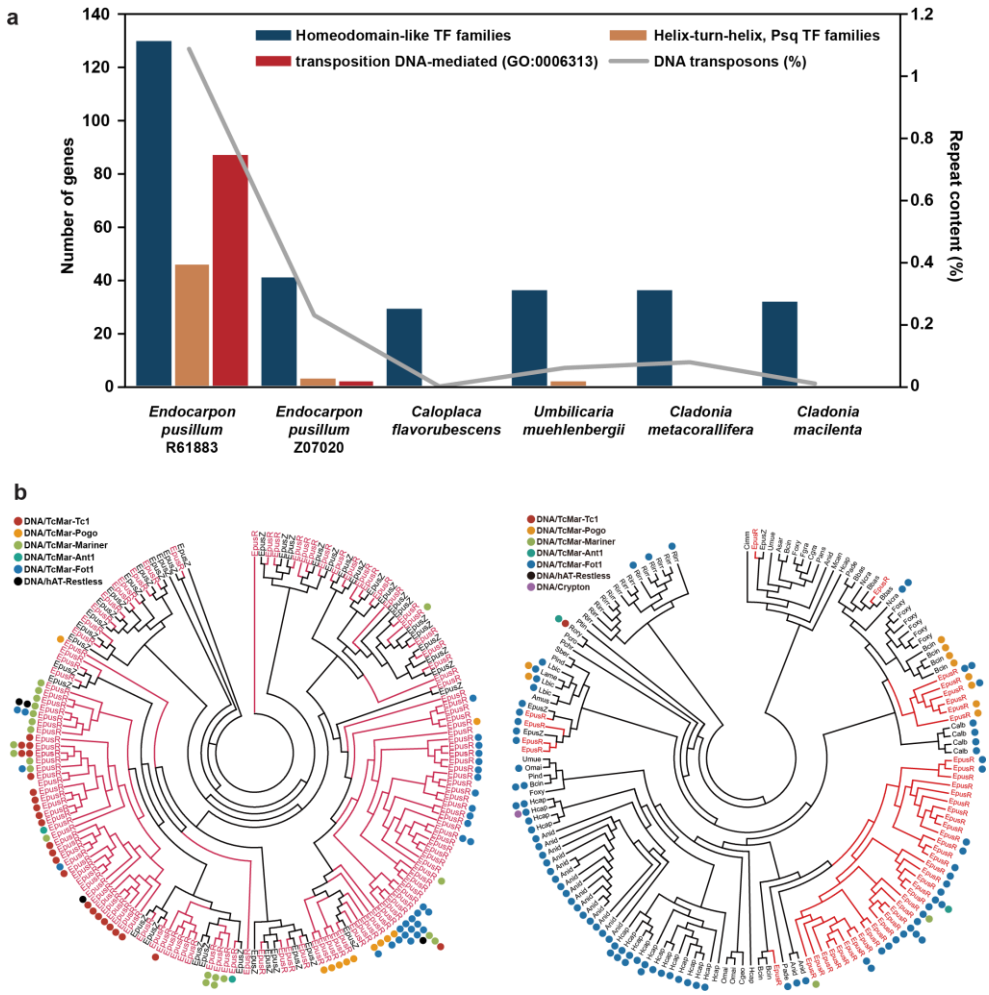


Figure 9. Specific TF duplication in *E. pusillum* R61883 driven by repeat sequences.

(A) Correlation among the expanded gene families, transposition DNA-mediated genes (GO:0006313), and DNA transposons. The bar graph represents the number of homeodomain-like, helix-turn-helix, psq, and transposition DNA-mediated genes, and the line graph represent the percentage of DNA transposons in each species. (B) Phylogenetic tree of homeodomain-like and helix-turn-helix, psq genes in *E. pusillum* isolates respectively (bootstrap value: 1000). Red color represents genes of

E. pusillum R61883, which are significantly expanded. Colored circles indicate the number of DNA transposons within 3kb of TF genes and colors are a classification of DNA transposons.

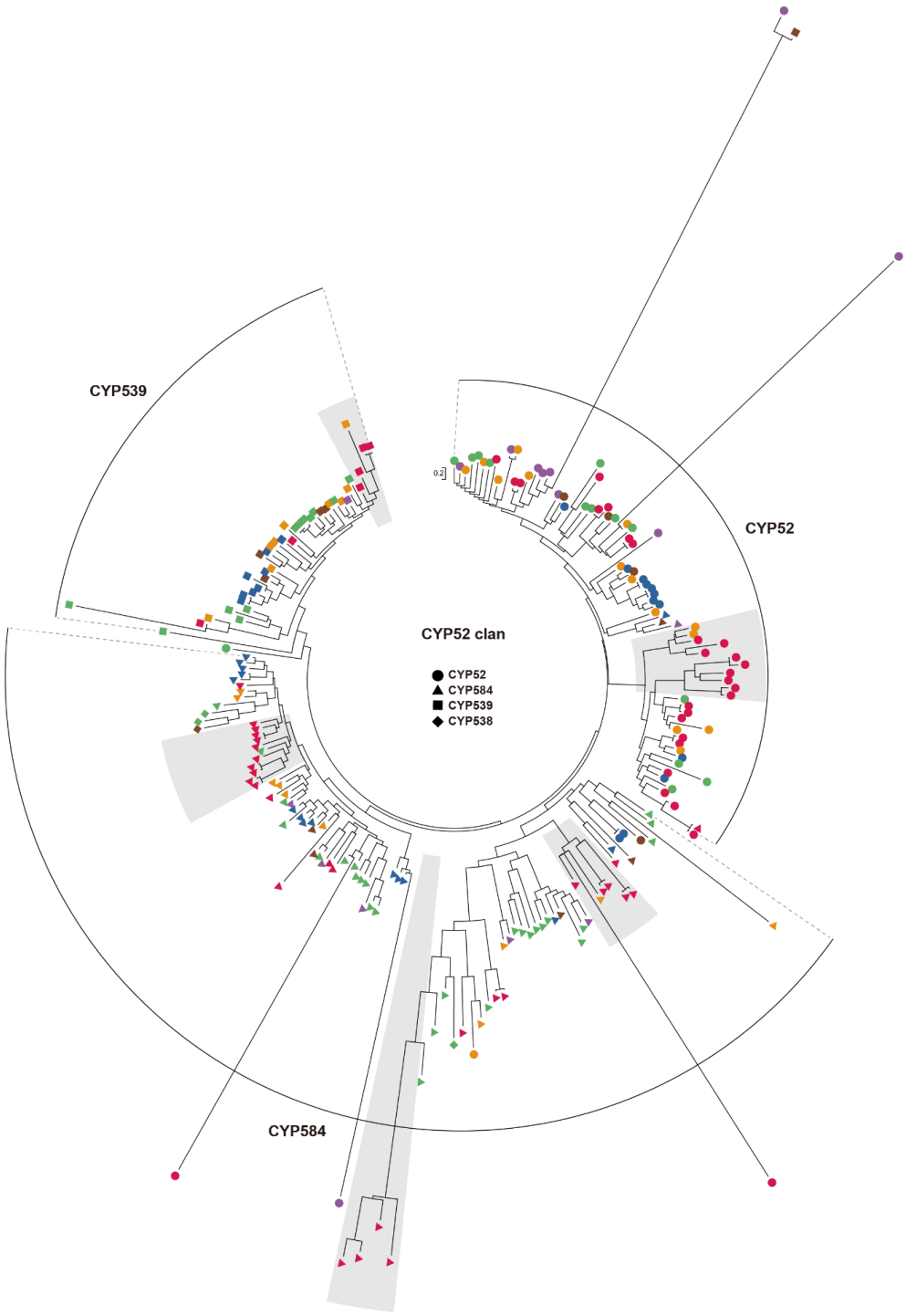
VI. Expanded cytochrome P450 (CYP) genes and secondary metabolites involved in lichen symbiosis

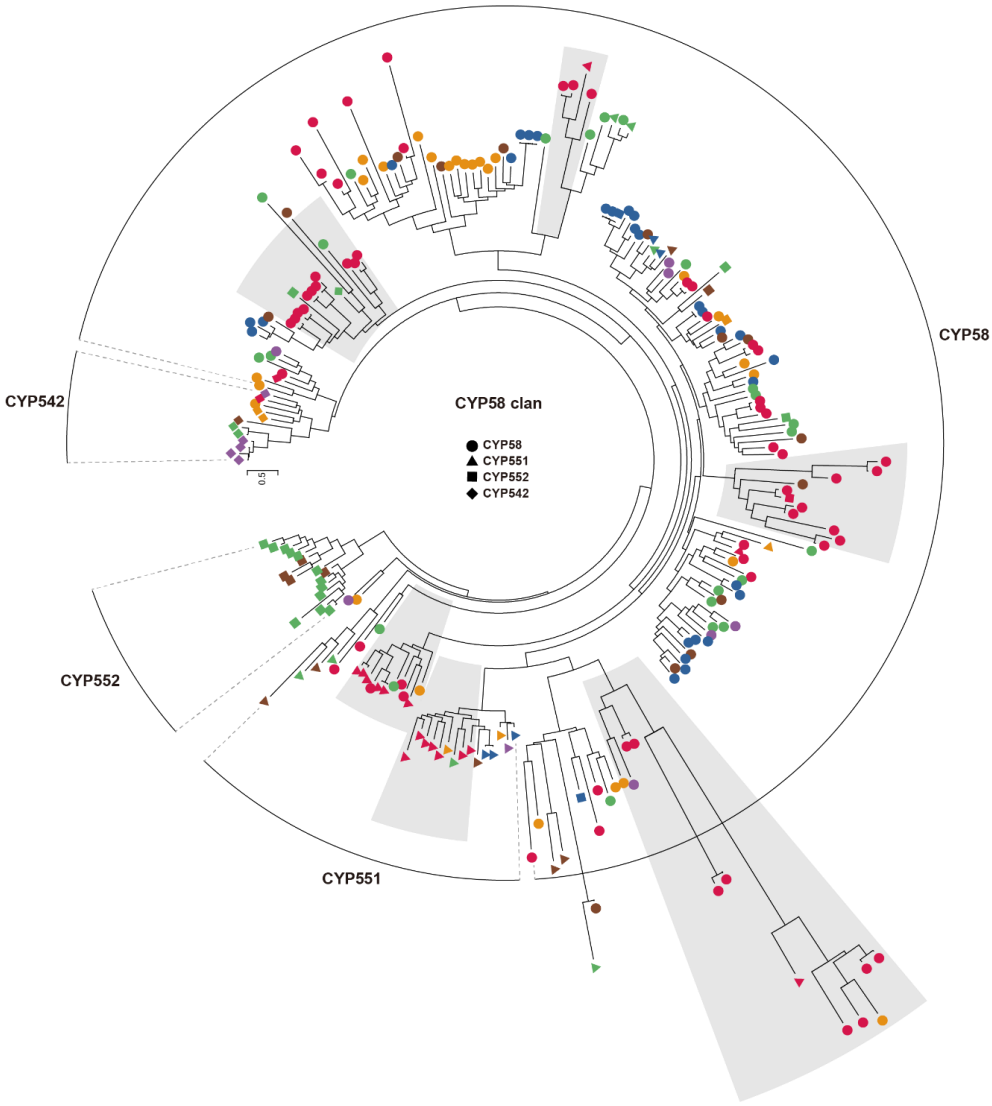
CYPs are heme-containing monooxygenases involved in a variety of metabolic processes (Park et al., 2008). Genes in the CYP52, CYP58, CYP551, CYP611, and CYP614 families are more numerous in lichen-forming fungi than in other analyzed genomes (Figure 6A). Expanded CYP genes in lichen-forming fungi are separated from those of fungi with other lifestyles, indicating that this feature evolved uniquely from other fungi. CYP52 and CYP58 are involved in n-alkane and fatty acid assimilation and trichothecene biosynthesis (Shin et al., 2018). The CYP551, CYP611, and CYP614 families have not been characterized, but may be involved in the symbiotic lifestyle because most of these CYP genes are lichen genes (Figure 6A and Figure 10).

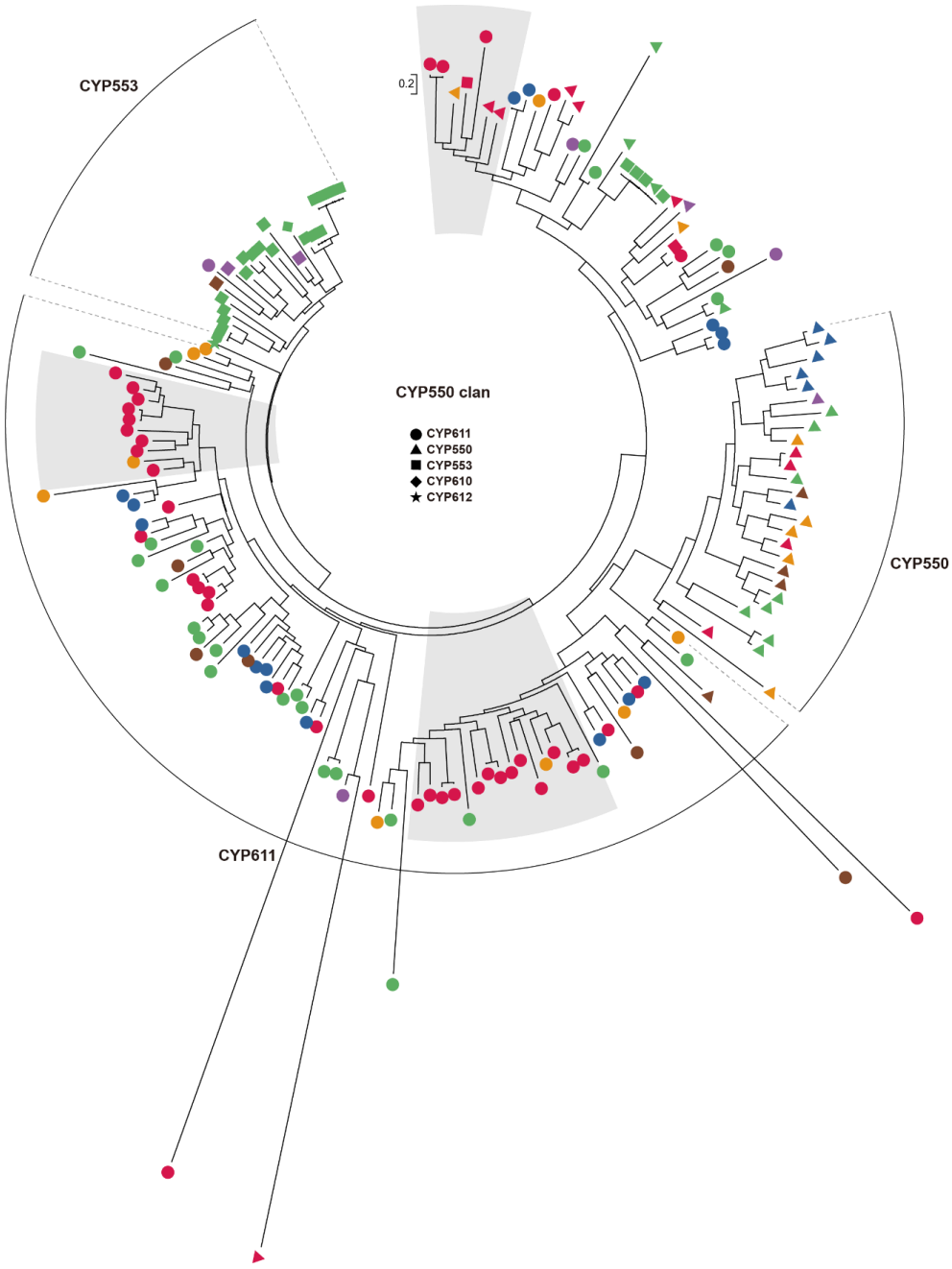
Lichen-forming fungi synthesize various unique secondary metabolites (Nash, 2008). We found more polyketide synthase (PKS) genes in lichen-forming fungi, mainly in Lecanoromycetes, than in their close relatives (Table 5). Reconciliation analysis revealed that the gain of these PKS genes occurred after lichen-forming fungi emerged from non-lichenized ancestors (Figure 11). Although the *E. pusillum* isolates are distantly related to Lecanoromycetes lichens, lichen-forming fungi shared many PKS genes in the phylogenetic analysis (Figure 12); we identified a lichen-specific PKS group that consisted entirely of lichen species, except *G. flavorubescens*. Sequence similarity analysis based on a blast search revealed that this is a lichen-specific PKS gene, with no ortholog in other fungal species (Figure 13A). Instead, *G. flavorubescens* have species-specific PKS genes (Figure 13B).

This genomic evidence is consistent with previous findings of the presence of unique secondary metabolites synthesized in lichen-forming fungi (Boustie and Grube, 2005;Ranković, 2015;Calcott et al., 2018).

Transcriptomic data for the resynthesis of *G. flavorubescens* and *T. gelatinosa* were used to investigate the relationships among two expanded gene families (PKS and CYP) and lichen symbiosis. Several PKS and CYP genes were highly expressed only during the early stage (at 12, 24, 48, and 72 h), whereas other genes were induced only during the late stage (4 and 6 weeks) (Figure 14). All lichen-specific PKS genes were induced only during the early stage of symbiosis (Figure 14B). The expanded gene families appear to be involved in producing various compounds and secondary metabolites, as previously described (Boustie and Grube, 2005;Nash, 2008;Ranković, 2015;Calcott et al., 2018). However, the expression patterns of lichen-specific PKS genes indicate that the lichen-specific PKS products are induced during the early stage of symbiosis.


CYP539
CYP52
CYP584
CYP52 clan
CYP52
CYP584
CYP539
CYP538
0.2


CYP58 clan
CYP58
CYP551
CYP552
CYP542
0.5
CYP542
CYP552
CYP551
CYP58


CYP553
0.2
CYP550 clan
CYP611
CYP550
CYP553
CYP610
CYP612
CYP550
CYP611

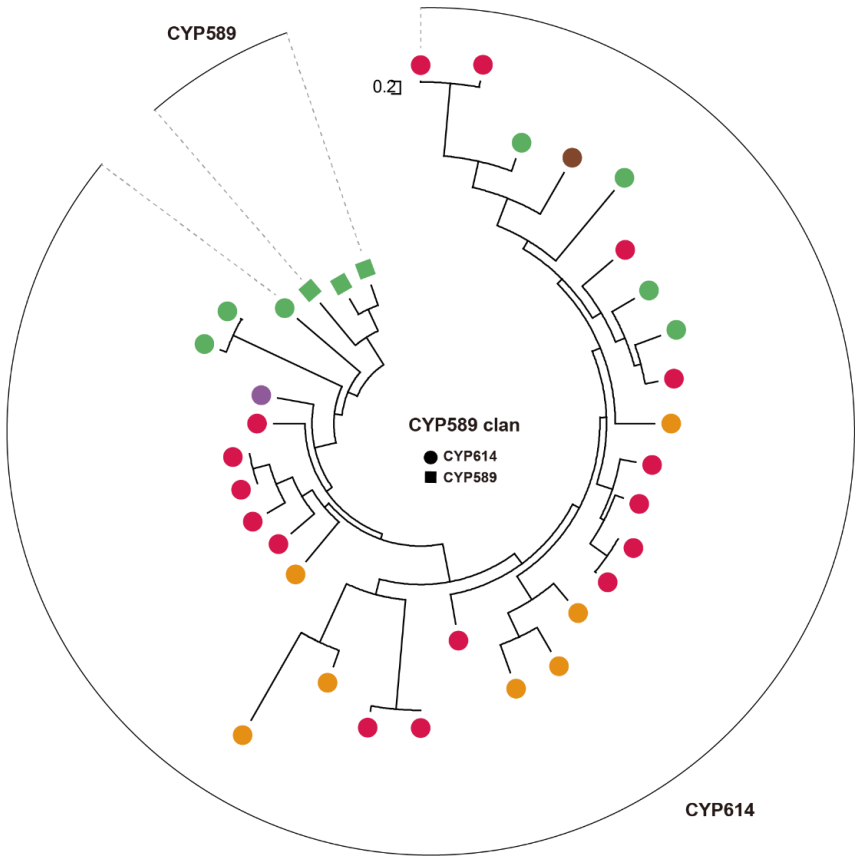


Figure 10. Maximum-likelihood phylogenetic analysis of expanded cytochrome P450 families

Phylogeny of each expanded cytochrome P450 families in ascomycetes fungi. Expanded families were shown with clans to which each family belongs. The maximum-likelihood phylogenetic tree was constructed by RAxML with 1000 bootstrap values. Families in each clan were classified with the shape of symbols. Colored marks mean each lifestyle of fungi. Red, orange, purple, green, brown, and blue indicate lichen-forming fungi, mycorrhizal fungi, endophyte, plant pathogen, saprotroph, and animal pathogen respectively. Gray shaded region shows lichen-forming fungi expanded groups or lichen-forming fungi-specific groups.

Table 5. Predicted secondary metabolite synthase genes

Species name	PKS	PKS-like	NRPS	NRPS-like	HYBRID	DMAT	Total
Microsporum canis	18	1	13	8	3	6	49
Trichophyton rubrum	9	0	12	1	1	3	26
Coccidioides immitis	9	2	5	3	1	1	21
Blastomyces dermatitidis	2	1	8	6	0	8	25
Histoplasma capsulatum	1	1	5	4	0	1	12
Aspergillus fumigatus	13	2	13	5	1	7	41
Aspergillus nidulans	23	4	11	12	1	5	56
Endocarpon pusillum R61883	13	1	2	3	0	3	22
Endocarpon pusillum Z07020	17	2	3	3	1	2	28
Phaeomoniella chlamydospora	9	1	5	2	0	0	17
Gyalolechia flavorubescens	16	3	1	7	2	2	31
Umbilicaria muehlenbergii	20	1	0	2	0	1	24
Cladonia metacorallifera	28	11	2	9	2	0	52
Cladonia macilenta	27	8	4	8	1	0	48
Cenococcum geophilum	10	2	7	3	1	2	25
Mycosphaerella graminicola	11	2	6	6	0	0	25
Blumeria graminis	1	1	1	1	0	0	4
Botrytis cinerea	16	6	6	8	0	1	37
Oidiodendron maius	44	2	9	10	2	3	70
Ascocoryne sarcoides	2	1	3	4	1	1	12
Podospora anserina	17	2	6	5	3	1	34
Neurospora crassa	8	1	3	2	0	1	15
Magnaporthe oryzae	24	2	8	6	5	3	48
Colletotrichum graminicola	37	1	5	7	4	6	60
Beauveria bassiana	12	1	13	7	3	0	36
Fusarium oxysporum	12	0	13	11	2	3	41
Fusarium graminearum	14	1	10	11	1	0	37
Tuber melanosporum	2	1	1	2	0	0	6
Schizosaccharomyces pombe	0	1	1	1	0	0	3
Candida albicans	0	2	0	0	0	0	2
Saccharomyces cerevisiae	0	1	0	0	0	0	1
Puccinia graminis	2	0	0	1	0	0	3
Melampsora laricis-populina	0	1	0	2	0	0	3
Ustilago maydis	3	2	3	6	0	2	16
Cryptococcus neoformans	0	0	0	1	0	0	1
Tulasnella calospora	1	0	0	1	0	0	2
Sebacina vermifera	0	0	0	0	0	0	0
Piriformospora indica	1	0	0	1	0	1	3
Laccaria amethystina	3	0	2	1	0	1	7
Laccaria bicolor	3	1	0	2	0	1	7
Hebeloma cylindrosporum	3	1	1	4	0	1	10
Amanita muscaria	3	0	0	1	0	1	5
Suillus luteus	3	0	1	4	0	0	8
Suillus brevipes	3	0	1	5	0	0	9
Pisolithus tinctorius	2	0	1	2	0	0	5

Pisolithus microcarpus	2	0	1	4	0	0	7
Scleroderma citrinum	1	2	0	4	1	0	8
Paxillus involutus	4	2	3	9	2	0	20
Paxillus adelphus	2	0	1	2	2	0	7
Serpula lacrymans	6	0	8	6	2	3	25
Piloderma croceum	6	1	1	8	0	3	19
Heterobasidion irregulare	3	2	0	8	0	1	14
Phanerochaete chrysosporium	2	0	1	15	0	0	18
Phycomyces blakesleeanus	0	1	0	4	0	0	5
Rhizopus oryzae	0	0	0	6	0	0	6
Rhizophagus irregularis	0	1	0	0	0	0	1

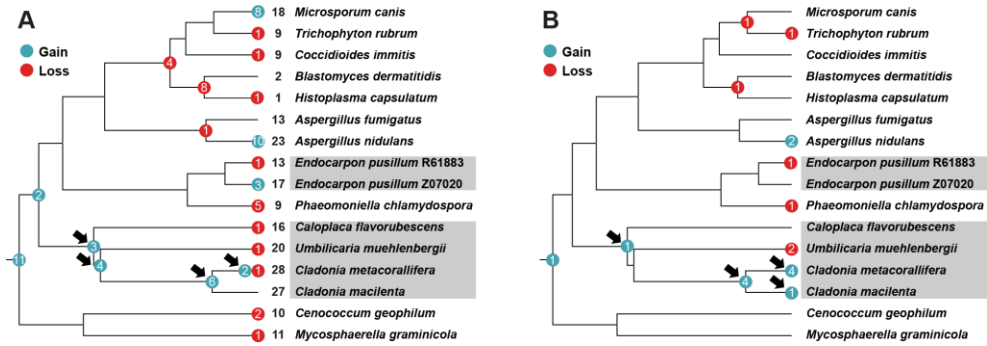


Figure 11. Gain and loss of PKS and PKS-like genes in lichen-forming fungi

Reconciliation analysis of PKS and PKS-like genes. The black arrow shows a gain of genes during the evolution of lichen-forming fungi. (A) Gain and loss of PKS genes in Eurotiomycetes, Lecanoromycetes, and Dothideomycetes. Blue circles indicate a gain of genes and a red circle means a loss of genes. (B) Changes in PKS-like genes during lichen-forming fungi evolution.

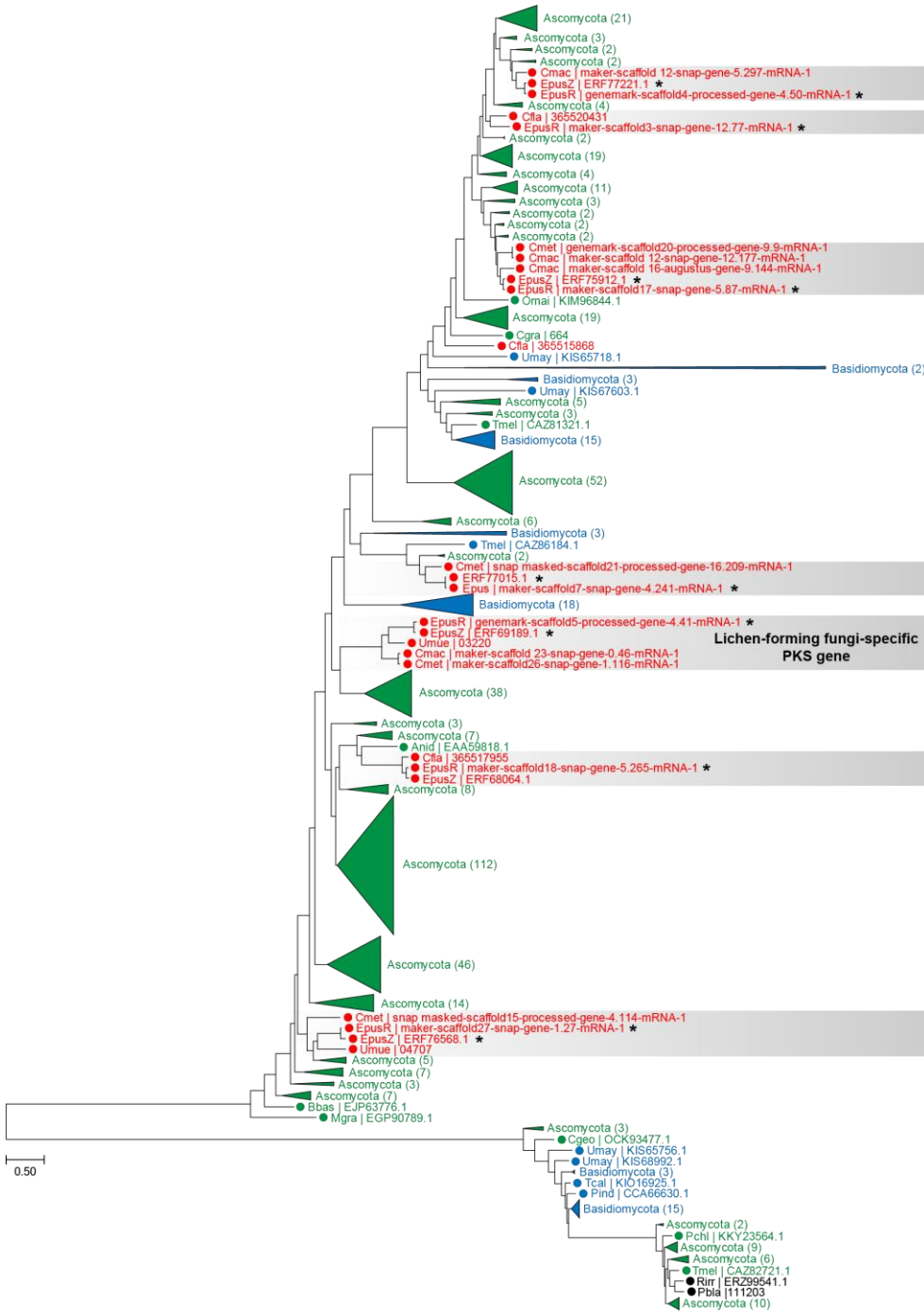

Ascomycota (21)
Ascomycota (3)
Ascomycota (2)
Ascomycota (2)
Cmac | maker-scaffold 12-snap-gene-5.297-mRNA-1
EpusZ | ERF77221.1 ★
EpusR | genemark-scaffold4-processed-gene-4.50-mRNA-1 ★
Ascomycota (4)
Cfla | 365520431
EpusR | maker-scaffold3-snap-gene-12.77-mRNA-1 ★
Ascomycota (2)
Ascomycota (19)
Ascomycota (4)
Ascomycota (11)
Ascomycota (3)
Ascomycota (2)
Ascomycota (2)
Ascomycota (2)
Cmet | genemark-scaffold20-processed-gene-9.9-mRNA-1
Cmac | maker-scaffold 12-snap-gene-12.177-mRNA-1
Cmac | maker-scaffold 16-augustus-gene-9.144-mRNA-1
EpusZ | ERF75912.1 ★
EpusR | maker-scaffold17-snap-gene-5.87-mRNA-1 ★
Omai | KIM96844.1
Ascomycota (19)
Cgra | 664
Cfla | 365515868
Umay | KIS65718.1
Basidiomycota (2)
Basidiomycota (3)
Umay | KIS67603.1
Ascomycota (5)
Ascomycota (3)
Tmel | CAZ81321.1
Basidiomycota (15)
Ascomycota (52)
Ascomycota (6)
Basidiomycota (3)
Tmel | CAZ86184.1
Ascomycota (2)
Cmet | snap masked-scaffold21-processed-gene-16.209-mRNA-1
ERF77015.1 ★
Epus | maker-scaffold7-snap-gene-4.241-mRNA-1 ★
Basidiomycota (18)
EpusR | genemark-scaffold5-processed-gene-4.41-mRNA-1 ★
EpusZ | ERF69189.1 ★
Umue | 03220
Cmac | maker-scaffold 23-snap-gene-0.46-mRNA-1
Cmet | maker-scaffold26-snap-gene-1.116-mRNA-1
Lichen-forming fungi-specific PKS gene
Ascomycota (38)
Ascomycota (3)
Ascomycota (7)
Anid | EAA59818.1
Cfla | 365517955
EpusR | maker-scaffold18-snap-gene-5.265-mRNA-1 ★
EpusZ | ERF68064.1
Ascomycota (8)
Ascomycota (112)
Ascomycota (46)
Ascomycota (14)
Cmet | snap masked-scaffold15-processed-gene-4.114-mRNA-1
EpusR | maker-scaffold27-snap-gene-1.27-mRNA-1 ★
EpusZ | ERF76568.1 ★
Umue | 04707
Ascomycota (5)
Ascomycota (7)
Ascomycota (3)
Ascomycota (7)
Bbas | EJP63776.1
Mgra | EGP90789.1
0.50
Ascomycota (3)
Cgeo | OCK93477.1
Umay | KIS65756.1
Umay | KIS68992.1
Basidiomycota (3)
Tcal | KIO16925.1
Pind | CCA66630.1
Basidiomycota (15)
Ascomycota (2)
Pchl | KKY23564.1
Ascomycota (9)
Ascomycota (6)
Tmel | CAZ82721.1
Rirr | ERZ99541.1
Pbla |111203
Ascomycota (10)

Figure 12. Lichen-forming fungi unique polyketide synthase genes

Polyketide synthase (PKS) genes in 56 fungal species. PKS gene of lichen-forming fungi represented by red color. Green color means PKS genes in Ascomycota, blue color is Basidiomycota and other PKS genes are black color. The numbers in the brackets are the number of PKS genes in each group. Gray shading indicates the PKS genes, which are shared with only *E. pusillum* isolates and Lecanoromycetes lichen-forming fungi. Genes of *E. pusillum* isolates were marked with an asterisk.

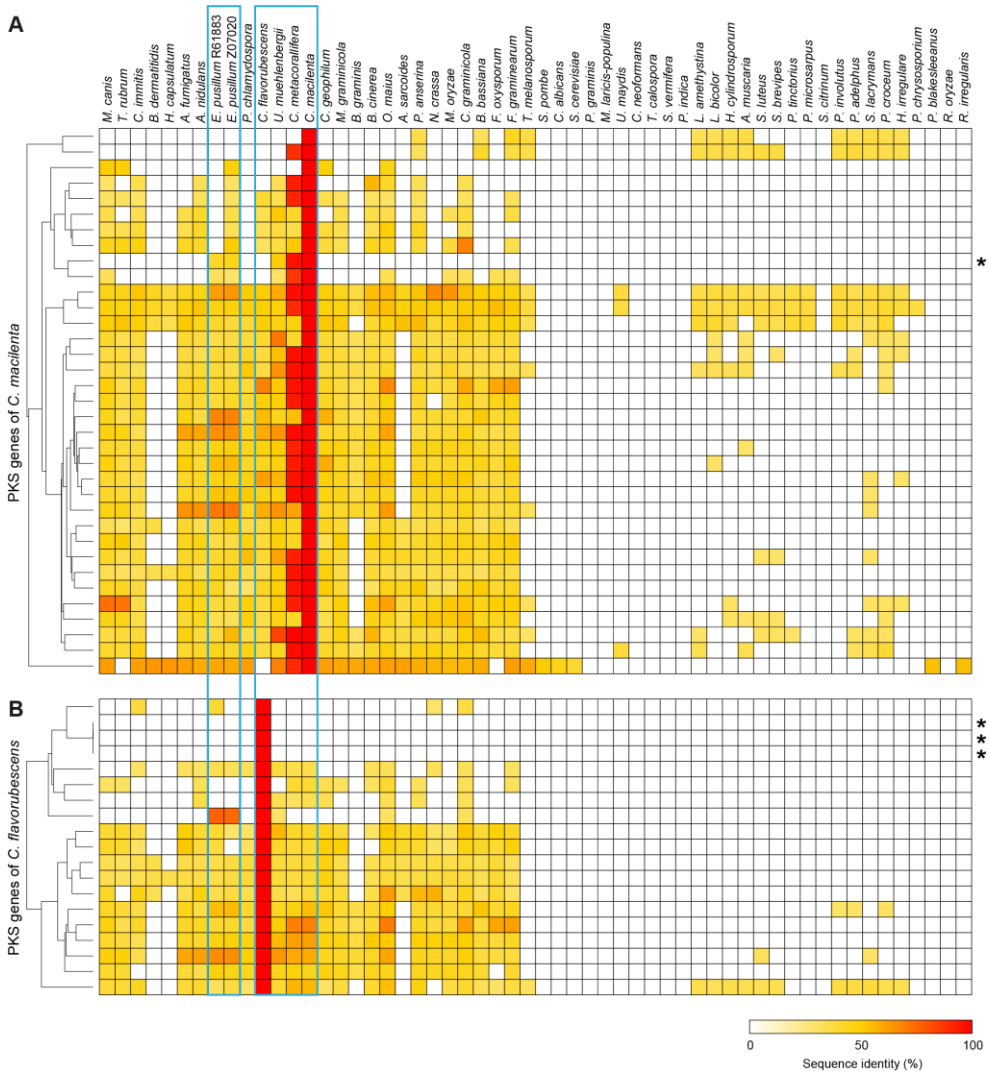


Figure 13. Presence and absence of lichen-forming fungi PKS genes in 56 fungal species

(A) BLAST search using PKS genes of *C. macilenta* as a reference. Red color means high sequence identity. The asterisk indicates lichen-specific PKS genes or species-specific PKS genes. (B) Sequence identity analysis of SSPs in *G. flavorubescens*.

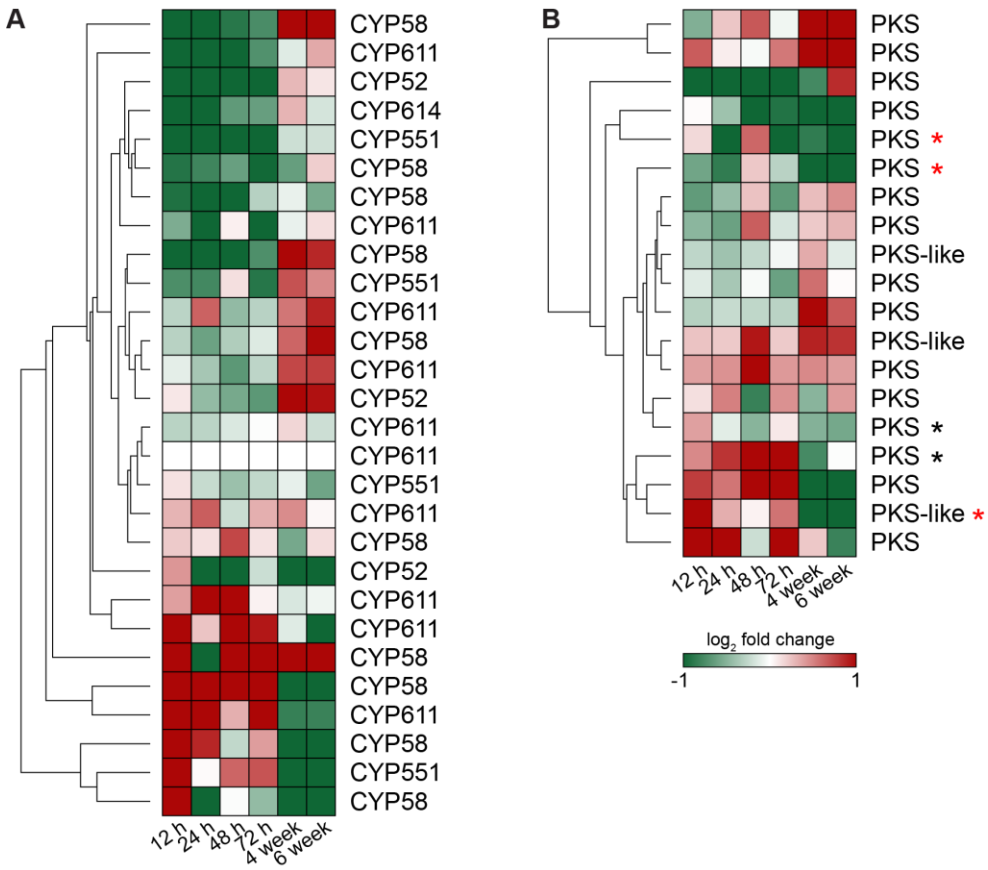


Figure 14. Expression profile of cytochrome P450 and PKS genes in *G. flavorubescens*

Expression of (A) expanded CYP families and (B) PKS genes in symbiosis of *G. flavorubescens*. Each gene family was clustered using a hierarchical clustering method. (B) The red asterisk indicates the *G. flavorubescens*-specific PKS genes and the black asterisk means the *G. flavorubescens* PKS genes that shared with *Endocarpon* spp.

VI. Lichen-specific genes of six lichen-forming fungi

In addition to the loss of unnecessary genes in lichen-forming fungi, we attempted to identify newly gained genes that may contribute to their unique symbiotic lifestyle. Ortholog clustering analysis of the six lichen-forming fungi identified 3,051 core groups, whereas clustering with only four Lecanoromycetes identified 3,468 core groups (Figure 15A and B; Table 6). Thus, the number of core gene clusters in lichen-forming fungi remained consistent regardless of the lichen-forming fungi in different classes.

We identified 5,498 lichen-specific orthogroups, including species-specific genes, after clustering with an additional 50 fungal genomes. The number of core groups was substantially reduced among lichen-forming fungi, leaving only one lichen-specific core group (Figure 15C and D). This finding suggests that no universal lichen-forming fungal gene sets are involved in their symbiosis and that, rather than core genes, they have many genus- or species-specific genes (Table 7), which are also important in mycorrhizal symbiosis (Kohler et al., 2015).

Lichen-specific genes were functionally annotated through gene ontology (GO) analysis using biological process terms. GO terms revealed no association with approximately 90% of the genes in *E. pusillum* R61883, 97% in *E. pusillum* Z07020, 99% in *G. flavorubescens*, 99% in *U. muehlenbergii*, 88% in *C. metacorallifera*, and 89% in *C. macilenta* (Figure 16A); therefore, these genes were likely newly gained during the evolution of lichen symbiosis. Common function of the functionally annotated genes in the six lichen-forming fungi included oxidation–reduction

processes, protein phosphorylation, transmembrane transport, carbohydrate metabolic processes, and transcription regulation (Figure 16B). However, genes involved in DNA-mediated transposition were found only in *Endocarpon* species, primarily in *E. pusillum* R61883. This difference appears to be related to the abundance of DNA transposons and the expansion of specific TF families *in E. pusillum* R61883, as mentioned above (Figure 9).

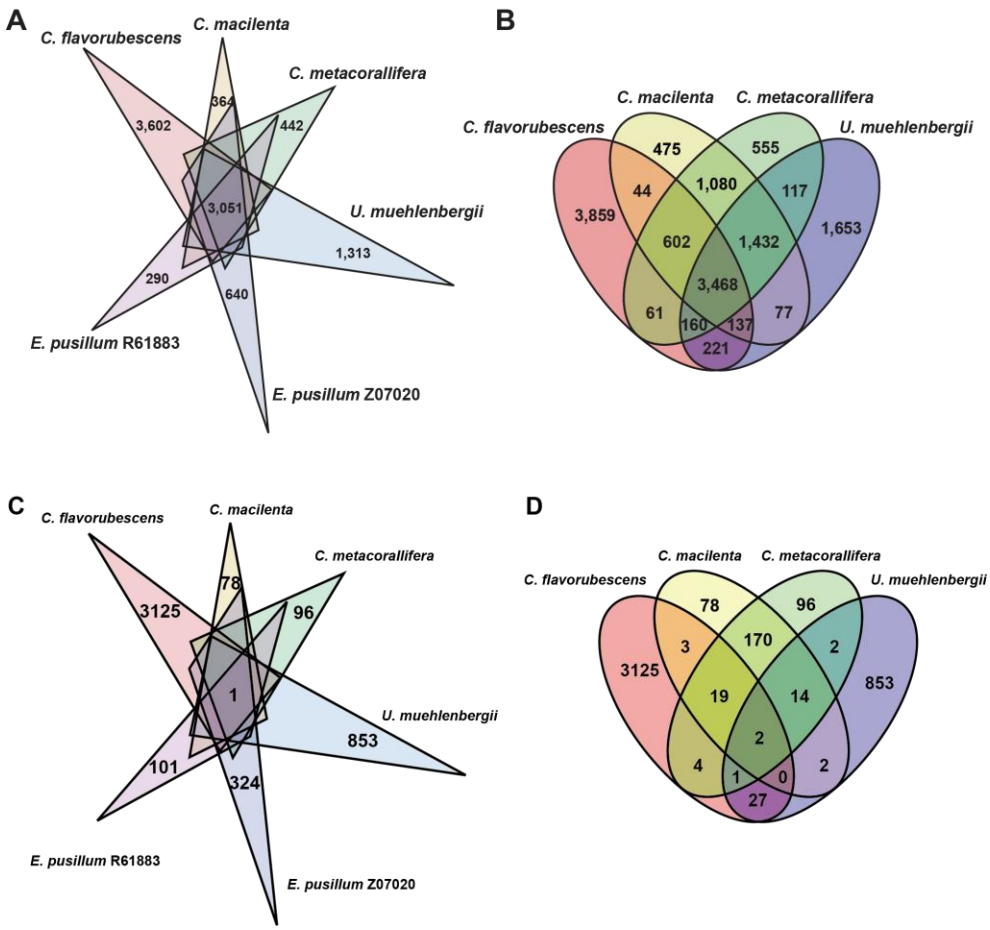


Figure 15. Core and specific genes among lichen-forming fungi

(A) The number of species-specific clusters and a core-cluster for the six lichen-forming fungi. The species-specific clusters are in each apex and the core-cluster is in the middle. (B) Ortholog clusters shared among the four Lecanoromycetes lichen-forming fungi. (C) and (D) Lichen-specific ortholog groups which are the result of ortholog clustering with 56 fungal species. Lichen-specific gene means that it is not present in the other non-lichen fungi.

Table 6. Summary of ortholog clustering with lichen-forming fungi

Fungal species	Orthogroups	Categories	Number of clusters	Number of genes
Lichen-forming fungi (6 species)	Pan clusters		15,712	54,282
	Core clusters	Total	3,051	23,234
		E. pusillum R61883 core gene	-	4,130
		E. pusillum Z07020 core gene	-	3,761
		G. flavorubescens core gene	-	3,735
		U. muehlenbergii core gene	-	3,626
		C. metacorallifera core gene	-	4,025
		C. macilenta core gene	-	3,957
	Lineage-specific clusters	*Endocarpon* lineage-specific clusters	1,220	2,970
		Cladonia lineage-specific clusters	739	1,586
	Orphan genes	*E. pusillum* R61883 orphan gene	290	324
		E. pusillum Z07020 orphan gene	640	650
		G. flavorubescens orphan gene	3,602	3,624
		U. muehlenbergii orphan gene	1,313	1,340
		C. metacorallifera orphan gene	442	447
		C. macilenta orphan gene	364	364
Lecanoromycetes (4 species)	Pan clusters		13,941	35,792
	Core clusters	Total	3,468	17,365
		G. flavorubescens core gene	-	4,226
		U. muehlenbergii core gene	-	4,138
		C. metacorallifera core gene	-	4,540
		C. macilenta core gene	-	4,461
	Lineage-specific clusters	*Cladonia* lineage-specific clusters	1,080	2,358
	Orphan genes	*G. flavorubescens* orphan gene	3,859	3,887
		U. muehlenbergii orphan gene	1,653	1,667
		C. metacorallifera orphan gene	555	562
		C. macilenta orphan gene	475	475

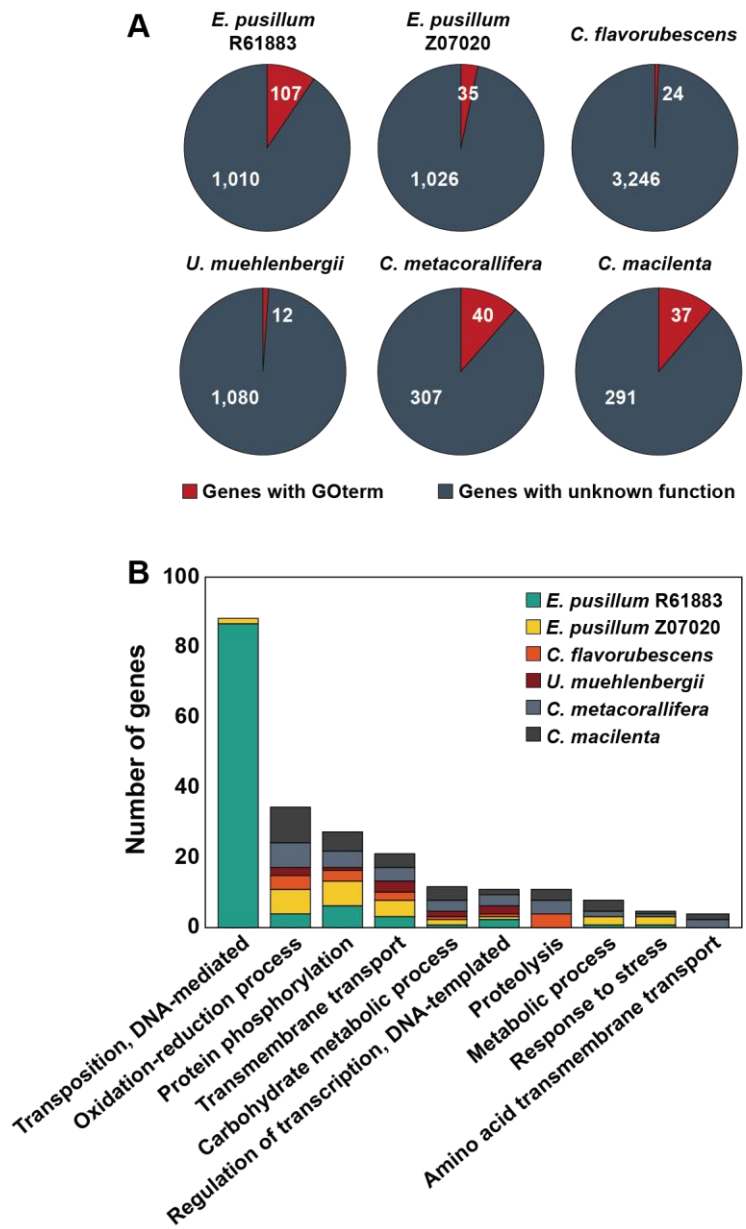


Figure 16. Major functions of the lichen-specific genes

The functions of lichen-specific genes were predicted by Gene Ontology (GO) term analysis. (A) The proportion of genes with GO terms in the lichen-specific genes. (B) The top 10 popular GO terms in the lichen-specific genes.

VI. Symbiosis-induced genes in *G. flavorubescens*

Genome-wide expression profiling was performed using *G. flavorubescens* and its algal partner *T. gelatinosa* to determine the possible roles of lichen-specific genes and conserved genes during lichenization. Of the conserved genes with orthologs in non-lichenized fungi, 17–20% were upregulated 4–6 weeks PCI, whereas 8–9% were upregulated at 12–72 h PCI (log2 fold-change > 1) (Figure 17A). Most genes that were upregulated during the late-stage were not affected during the early stage (Figure 17B), indicating that different conserved genes are associated with the early and late stages. Lichen-specific genes exhibited a different pattern from conserved genes, with more genes upregulated during the early stage (17–18%) than the late stage (6%) (Figure 17A). Similar to conserved genes, different genes were upregulated during the early and late stages, suggesting that lichen-specific genes that are highly induced during the early stage are no longer necessary during the late stage. Different lichen-specific upregulated genes were involved at each time point, even within each stage (Figure 17B).

GO enrichment analysis was performed on functionally annotated conserved genes that were differentially expressed during the early and late stages (Table 7). Genes that were up- or downregulated only during the early or late stage were defined as differentially expressed. Genes that were differentially upregulated during the early stage were significantly enriched in terms of epigenetic mechanisms, including chromosome organization (GO:0051276), DNA repair (GO:0006281), peptidyl-amino acid modification (GO:0018193), protein acylation (GO:0043543), and histone modification (GO:0016570). In contrast, terms related to glucose

(GO:0006096 and GO:0009070) and lipid (GO:0042157, GO:0042158, and GO:1903509) metabolism were significantly enriched during the late stage, suggesting that the early and late stages play different roles in lichen symbiosis.

Glucose metabolism is important in lichen-forming fungi, which absorb photosynthetic products from their algal partners and convert them into glucose or fructose for fungal metabolism (Wang et al., 2014). Glycolysis/gluconeogenesis pathway analysis of the differentially expressed genes mapped only genes induced during the late stage to the pathways (Figure 18), indicating that conversion of the obtained monosaccharides into energy sources occurs actively during the late stage.

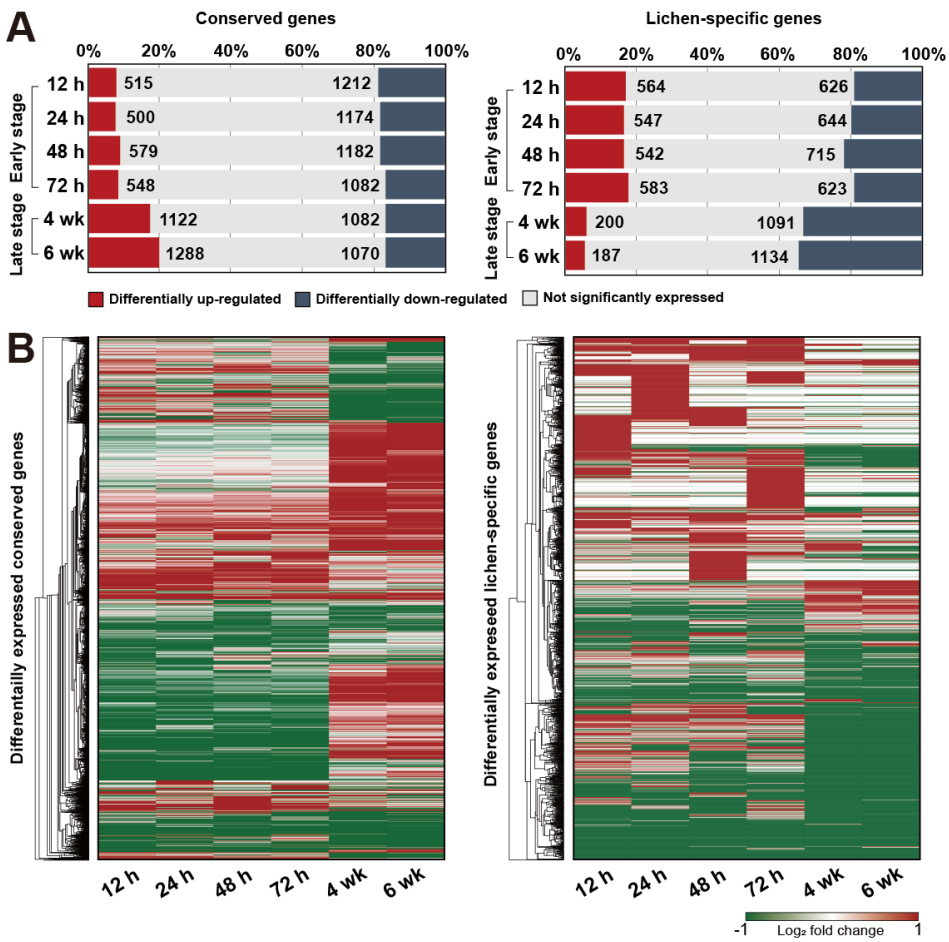


Figure 17. Symbiosis-induced genes in *G. flavorubescens*

Gene expression profiles of conserved and lichen-specific genes in *G. flavorubescens*. (A) Differentially up- ($\log_2$ fold change > 1) and downregulated ($\log_2$ fold change ≤ −1) genes at each time stage. Gray bars indicate genes that were not significantly expressed. (B) Expression patterns of lichen-specific and conserved genes.

Table 7. Go term enrichment test with differentially expressed conserved genes in *G. flavorubescens*

GO IDs	Terms	Annotated	Observed	Expected	Significance
<Up-regulated in early stage>					
GO:0051276	chromosome organization	40	10	4.01	0.0048
GO:0006281	DNA repair	68	14	6.83	0.0062
GO:0009147	pyrimidine nucleoside triphosphate metabolic process	2	2	0.2	0.01
GO:0018193	peptidyl-amino acid modification	28	7	2.81	0.0176
GO:0006631	fatty acid metabolic process	12	4	1.2	0.0256
GO:0043543	protein acylation	3	2	0.3	0.0281
GO:0006457	protein folding	38	8	3.81	0.0313
GO:0006464	cellular protein modification process	215	30	21.58	0.0338
GO:0016570	histone modification	13	4	1.3	0.0342
<Down-regulated in early stage>					
GO:0006355	regulation of transcription, DNA-templated	126	39	21.67	0.00007
GO:0055085	transmembrane transport	272	62	46.79	0.0073
GO:0042157	lipoprotein metabolic process	10	5	1.72	0.0174
GO:0042158	lipoprotein biosynthetic process	10	5	1.72	0.0174
GO:1903509	liposaccharide metabolic process	10	5	1.72	0.0174
GO:0055114	oxidation-reduction process	440	90	75.69	0.0283
GO:0044550	secondary metabolite biosynthetic process	2	2	0.34	0.0295
GO:0044036	cell wall macromolecule metabolic process	2	2	0.34	0.0295
GO:0007034	vacuolar transport	8	4	1.38	0.0339
GO:0001932	regulation of protein phosphorylation	5	3	0.86	0.0385
GO:0051338	regulation of transferase activity	5	3	0.86	0.0385
GO:0071554	cell wall organization or biogenesis	5	3	0.86	0.0385
<Up-regulated in late stage>					

GO:0006096	glycolytic process	8	6	1.79	0.0023
GO:0009070	serine family amino acid biosynthetic process	4	4	0.89	0.0025
GO:0042157	lipoprotein metabolic process	10	6	2.24	0.0111
GO:0042158	lipoprotein biosynthetic process	10	6	2.24	0.0111
GO:1903509	liposaccharide metabolic process	10	6	2.24	0.0111
GO:0070646	protein modification by small protein removal	13	7	2.91	0.013
GO:0006351	transcription, DNA-templated	181	53	40.48	0.0146
GO:0051186	cofactor metabolic process	65	26	14.54	0.0168
GO:0065009	regulation of molecular function	14	7	3.13	0.021
GO:0051336	regulation of hydrolase activity	6	4	1.34	0.0252
GO:0019438	aromatic compound biosynthetic process	274	78	61.28	0.0272
GO:0018130	heterocycle biosynthetic process	287	81	64.19	0.0285
GO:1901362	organic cyclic compound biosynthetic process	298	83	66.65	0.0352
GO:0042401	cellular biogenic amine biosynthetic process	4	3	0.89	0.0371
GO:0006796	phosphate-containing compound metabolic process	233	67	52.11	0.0394
GO:0070647	protein modification by small protein conjugation or removal	19	8	4.25	0.0427
GO:0006793	phosphorus metabolic process	234	67	52.34	0.0429
GO:0035556	intracellular signal transduction	58	19	12.97	0.043
GO:0019219	regulation of nucleobase-containing compound metabolic process	128	37	28.63	0.0459
<Down-regulated in late stage>					
GO:0006189	'de novo' IMP biosynthetic process	2	2	0.16	0.0065
GO:0015703	chromate transport	2	2	0.16	0.0065
GO:0061024	membrane organization	6	3	0.48	0.0087
GO:0071166	ribonucleoprotein complex localization	3	2	0.24	0.0185
GO:0015931	nucleobase-containing compound transport	3	2	0.24	0.0185

GO:0051236	establishment of RNA localization	3	2	0.24	0.0185
GO:0006611	protein export from nucleus	3	2	0.24	0.0185
GO:0006403	RNA localization	3	2	0.24	0.0185
GO:0051189	prosthetic group metabolic process	4	2	0.32	0.0349
GO:0006694	steroid biosynthetic process	4	2	0.32	0.0349
GO:0006414	translational elongation	10	3	0.81	0.0408

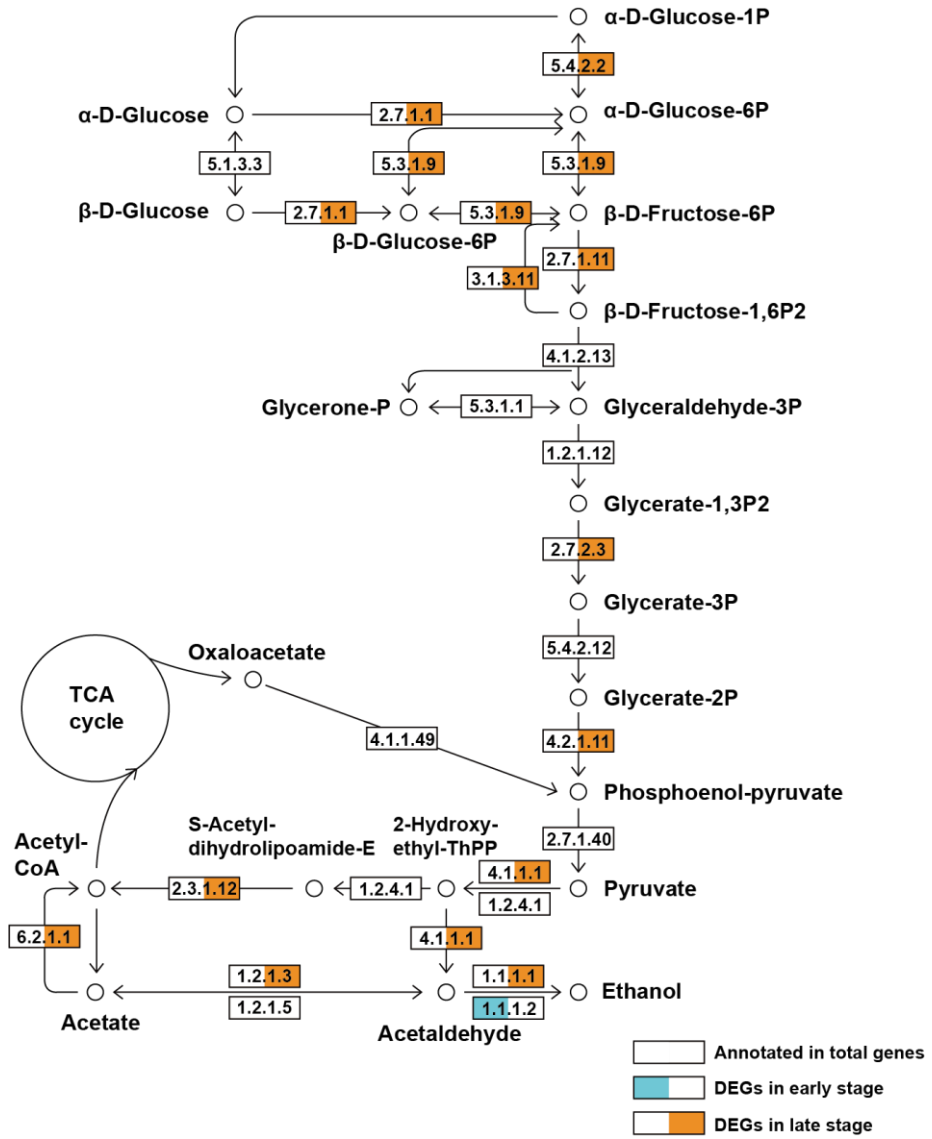


Figure 18. Pathway analysis of glycolysis/gluconeogenesis in *G. flavorubescens*

The simplified schematic diagram of the glycolysis/gluconeogenesis pathway in the KEGG pathway. Only nodes with mapped genes of *G. flavorubescens* are displayed as white boxes. Nodes marked in blue are associated with differentially expressed genes (DEGs) in the early stage of *G. flavorubescens* resynthesis and nodes marked in orange are associated with DEGs in the late stage.

VI. Small secreted proteins (SSPs) in lichen-forming fungi are involved in establishment and maintain the symbiosis

Although small secreted proteins (SSPs) are virulence factors in pathogenic fungi and are important for symbiosis in mycorrhizal fungi (Kim et al., 2016), their roles in lichen-forming fungi remain unclear. We found that lichen-forming fungi had 286–482 secreted proteins and 107–207 SSPs (Figure 19); these numbers were smaller than those for other fungi with different lifestyles, especially plant-associated symbionts (Figure 20A). Most SSPs found in lichen-forming fungi were genus- or species-specific according to the blast results (Figure 20B and Figrue 21). Sequence identity analysis of SSPs in *C. macilenta* revealed that 32% of the SSPs were *Cladonia*-specific and 39% were species-specific, whereas 43% of SSPs in *E. pusillum* R61883 were *Endocarpon*-specific and 31% were species-specific (Figure. 20B). In contrast, 76% of SSPs in *G. flavorubescens* and 75% in *U. muehlenbergii* were species-specific, as both species lacked closely related species, in our dataset (Figure 21). Although both of these lichen-forming fungal species belong to Lecanoromycetes, and both have *Trebouxia* spp. as algal partners, their SSPs differed markedly based on sequence similarity, suggesting that SSPs of lichen-forming fungi may not just dependent on their photobiont, but have been independently gained during speciation.

Like all genes, SSPs in *G. flavorubescens* had different expression profiles for conserved or specific genes. Most conserved SSPs were highly upregulated during the late stage of resynthesis, although some were also constitutively expressed during the early stages (Figure. 20C). However, most genus- and species-specific lichen

SSPs were upregulated at different time points during the early stage.

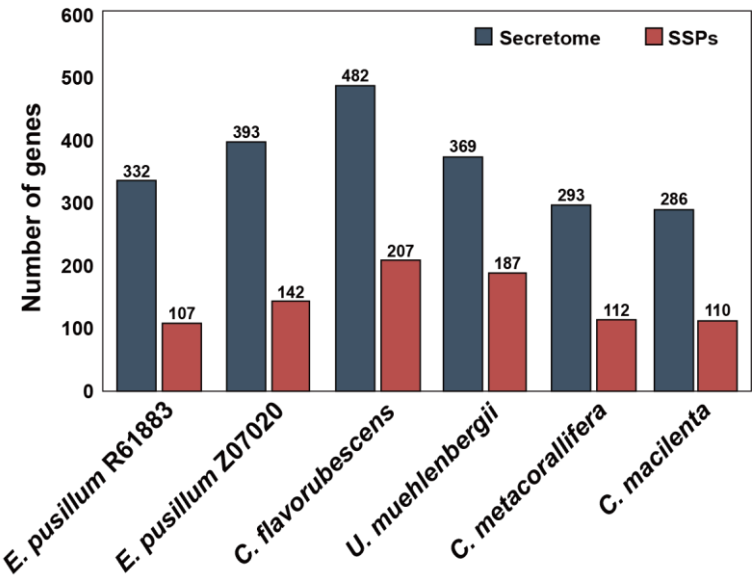


Figure 19. Number of secreted proteins and SSPs in lichen-forming fungi

Predicted secreted proteins and small secreted proteins (SSPs) in six lichen-forming fungi. The blue bar and red bar represent the number of secreted protein and SSPs respectively.

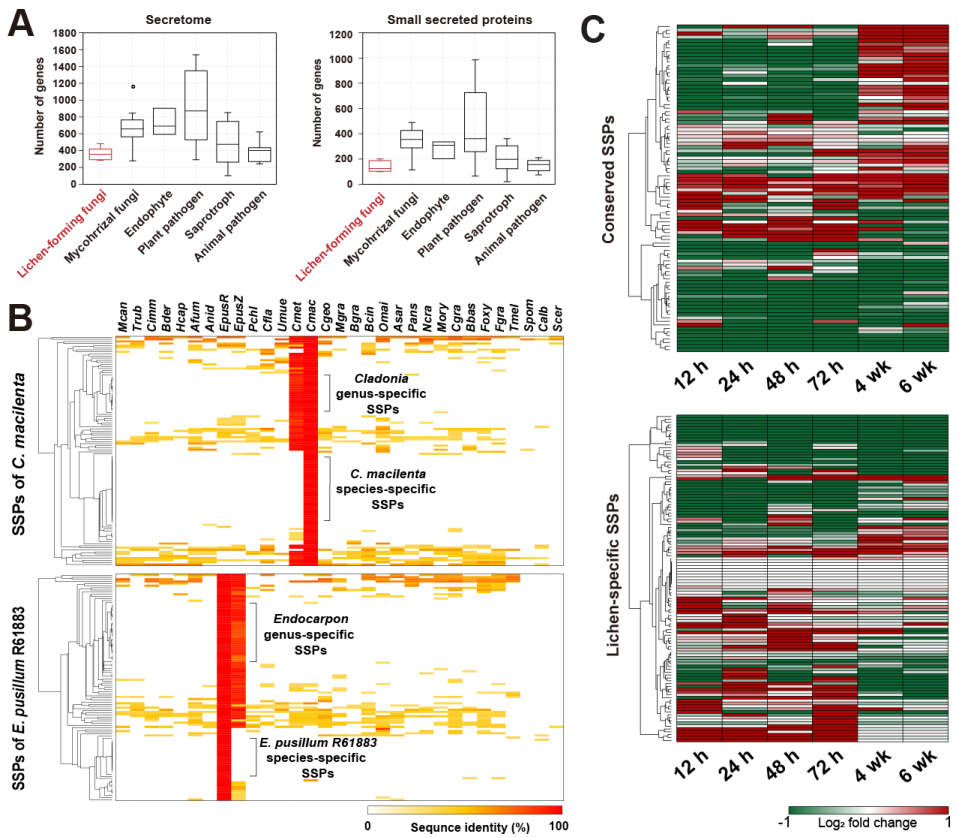


Figure 20. Lichen-specific small secreted proteins (SSPs) from *G. flavorubescens* induced in early lichen symbiosis

(A) Box plot of the number of secretomes and SSP distributions in lichen-forming fungi compared with other lifestyles. (B) Lichen-specific SSPs in *C. macilenta* and *E. pusillum* R61883. Ortholog SSPs of 56 fungal species were identified using blast and SSPs *of C. macilenta* and *E. pusillum* R61883 as references ($E = 1 \times 10^{-5}$). Abbreviations for fungal species are provided in Supplementary Dataset 1. (C) SSP expression in *G. flavorubescens* was classified as conserved (top) or lichen-specific (bottom).

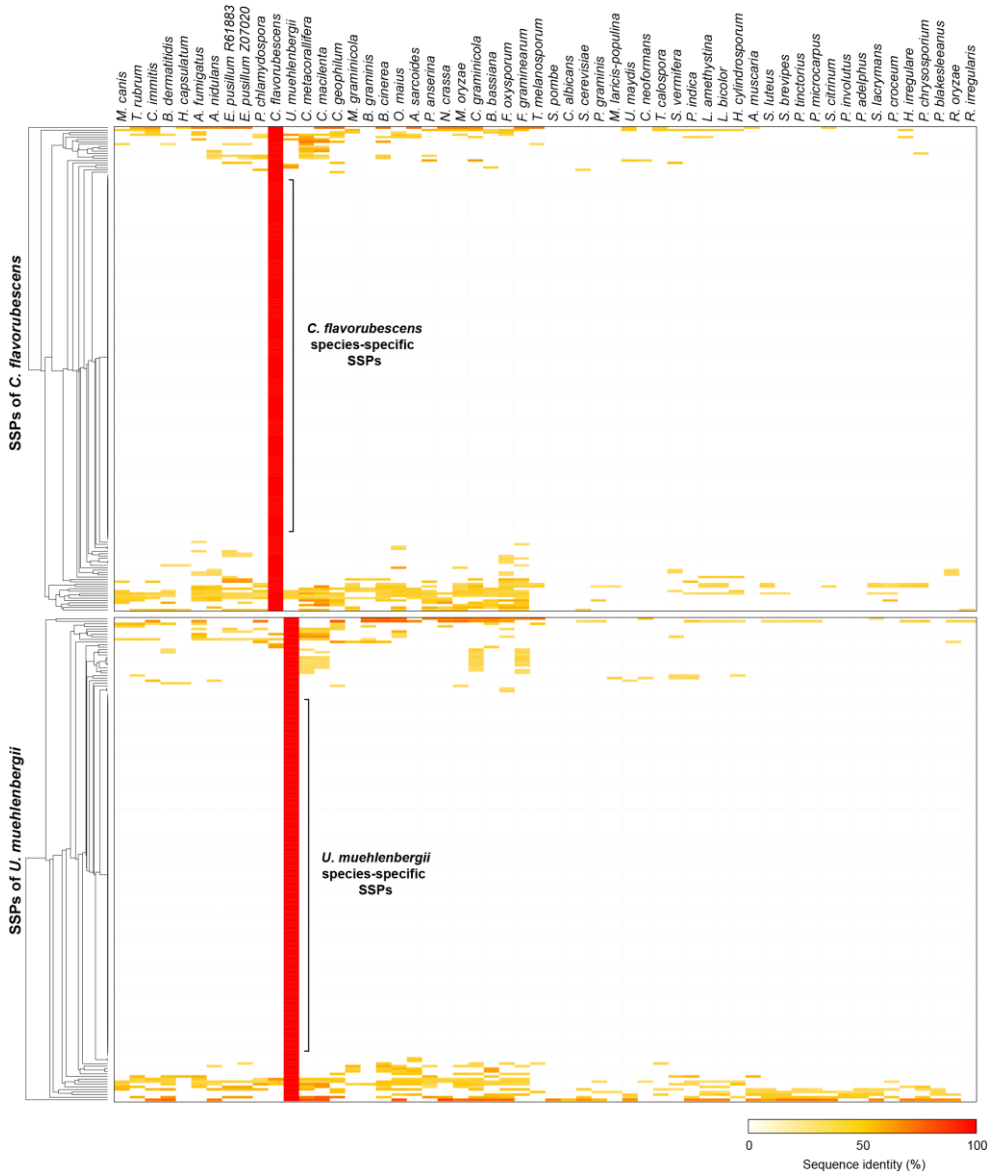


Figure 21. Species-specific SSPs in *G. flavorubescens* and *U. muehlenbergii*

Sequence identity analysis with SSPs of 56 fungal species (BLAST E-value 1×10^{-5}). SSPs of *G. flavorubescens* and *U. muehlenbergii* were used as a reference respectively, and almost all of the genes were species-specific.

VII. Delineation of the two *E. pusillum* strains as different species by comparative genomics

E. pusillum strains Z07020 and R61883 have been described as the same species in previous studies because they are morphologically indistinguishable (Park et al., 2014c;Wang et al., 2014). However, morphological classification is insufficient for species delineation, and molecular genetics has been used to solve this problem (Lumbsch and Leavitt, 2011;Boluda et al., 2019). To investigate this possibility, we compared the conserved proteins (ACT1, TEF1, TUB1, and TUB2) and the whole genomes of the two *E. pusillum* strains. Genomes of other species in the same genus and other strains in the same species were used for comparison. The sequences of conserved proteins among strains in the same species were identical, whereas the proteins of different species in the same genus had differences (Table 8). Comparison between the two *E. pusillum* strains showed that TEF1 and TUB2 genes contained mismatches, similar to the results for different species in the same genus. Whole-genome synteny analysis also showed that the *E. pusillum* strains had a relationship similar to fungi of different species in the same genus (Figure 22 and Table 9). Our results show that their genomic synteny and repeat contents differ significantly. This difference provides decisive evidence that their genomes have evolved into different species. Therefore, we determined that the two *E. pusillum* strains belong to different species within the same genus. Moreover, some gene families such as TFs expanded exclusively in *E. pusillum* R61883. This finding suggests that repetitive sequences, including transposable elements, can induce changes in genome structure that lead to speciation (Rose and Doolittle, 1983;Feschotte and Pritham, 2007) and may have

driven the rapid evolution of *E. pusillum* R61883.

Table 8. Alignment with conserved genes between *E. pusillum* isolates

	Query	Reference	ACT1	TEF1	TUB1	TUB2
E. pusillum	*E. pusillum* Z07020	*E. pusillum* R61883	100%	99.6%	100%	99.5%
Same genus, different species	*F. graminearum*	*F. oxysporum*	100%	94.1%	97.7%	99.5%
	C. macilenta	*C. metacorallifera*	100%	100%	100%	100%
	L. amethystina	*L. bicolor*	99.7%	98.5%	100%	99.5%
	C. graminicola	*C. higginsianum*	100%	97.6%	97.5%	99.5%
Same species, different strain	*M. oryzae* 70-15	*M. oryzae* KJ201	100%	100%	100%	100%
	A. fumigatus A1163	*A. fumigatus* Af293	100%	100%	100%	100%
	S.cerevisiae S288C	*S. cerevisiae* YJM993	100%	100%	100%	99.7%

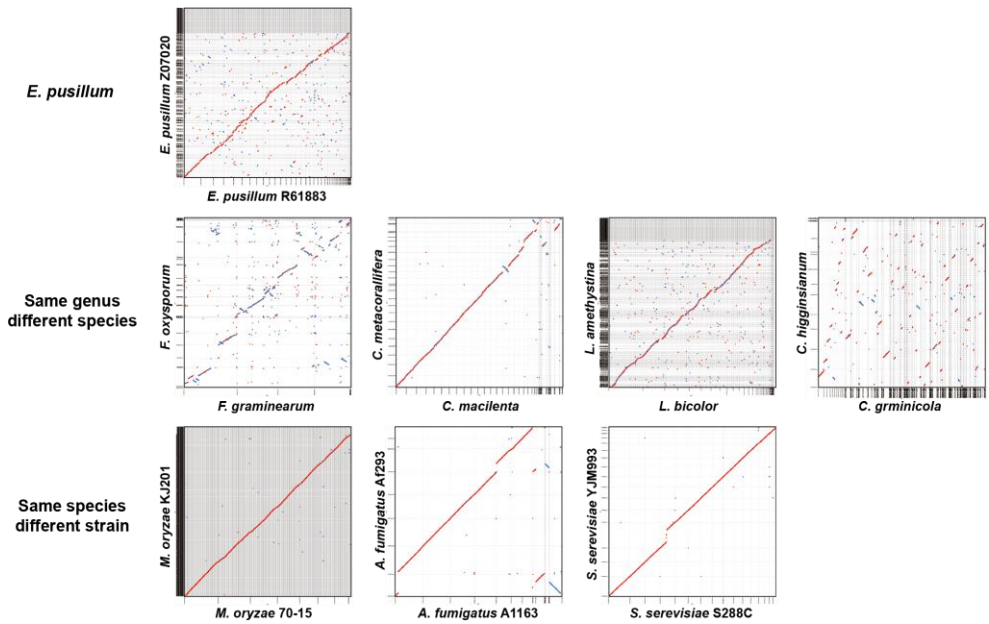


Figure 22. Synteny analysis between *E. pusillum* isolates

To delineate the two *E. pusillum* isolates Z07020 and R61883, dot plots were constructed using different species belong to the same genus and different strains belong to the same species of the selected fungal species. The *E. pusillum* dot plot is more similar to the dot plots of different species in the same genus.

Table 9. Synteny analysis of *E. pusillum* isolates

	Reference species - Query species	Length (bp)		Ratio (%)	
		Query	Reference	Query	Reference
E. pusillum	*E. pusillum R61883 - E. pusillum Z07020*	18,978,499	18,989,224	51%	51%
Same genus different species	*F. graminearum - F. oxysporum*	11,510,880	11,502,471	31%	19%
	C. macilenta - C. metacorallifera	24,397,165	24,403,396	66%	67%
	L. bicolor - L. amethystina	15,635,185	15,640,216	26%	30%
	C. graminicola - C. higginsianum	12,770,527	12,834,010	25%	25%
Same species different strain	*M. oryzae 70-15 - M. oryzae KJ201*	41,114,602	41,109,952	100%	115%
	A. fumigatus A1163 - A. fumigatus Af293	28,533,389	28,532,000	98%	97%
	S. cerevisiae S288C - S. cerevisiae YJM993	11,838,757	11,840,737	98%	95%

DISCUSSION

The main goal of symbiosis research is to determine how the beneficial associations evolved and to identify genes involved in the establishment and functioning of symbioses. However, to date, these features have been only partially investigated in lichen-forming fungi. All that is known about the evolution of lichen symbionts independently from non-lichenized ancestors was acquired from studies of small subunit and large subunit rDNA markers (Gargas et al., 1995;Lutzoni et al., 2001). We used whole-genome sequences to determine phylogenetic relationships more accurately. Our comparative analysis revealed that the lichen-forming fungi experienced massive reductions in unnecessary genes during symbiosis with their algal partners. Newly acquired lineage- and species-specific genes are involved in establishing lichen symbiosis, whereas conserved genes maintain the relationship.

PCWDEs are involved in host cell wall remodeling in mycorrhizal symbiosis (Balestrini and Bonfante, 2014) and confer virulence to fungal plant pathogens (Kubicek et al., 2014). However, we found that lichen-forming fungi experienced large contractions in PCWDE genes compared with their non-symbiotic ancestors. Pectin-degrading enzymes may no longer be necessary for algal host association in lichen symbiosis, because this cell wall component is unique to Charophyceae algae and land plants, but is not present in Chlorophyte green algae, which can form lichen (Popper et al., 2011). Consequently, most pectin-degrading enzyme genes have been lost in lichen-forming fungi derived from non-lichenized fungi. Nevertheless,

cellulose and hemicellulose enzyme genes underwent contractions similar to pectin, although they are common cell wall components in both plants and green algae (Popper et al., 2011). Lichen-forming fungi have simple wall-to-wall apposition or develop highly differentiated non-breaking symbiotic structures called intraparietal haustoria, as well as intracellular haustoria, which penetrate algal cell walls (Honegger, 1986). Although the haustoria of the lichen species used in this study were not observed directly, we suggest that they do not penetrate the host algal cell walls for colonization, as proposed in previous studies (Honegger, 1984;1986;Valladares et al., 1993). These symbiotic forms may lead to the non-functionalization of PCWDEs, which leads to gene loss (Albalat and Canestro, 2016). However, a recent study using a vast number of lichen-forming fungal genomes revealed that not all lichen-forming fungi lost large numbers of PCWDE genes (Resl et al., 2021). Taking this findings into consideration, the overall trend of our data suggests the loss of PCWDE genes in lichen, with several exceptional cases. Ectomycorrhizal fungi have similar symbiotic relationships with their hosts and have lost many PCWDE genes, unlike endomycorrhizal fungi and plant pathogens (Kuo et al., 2014;Kohler et al., 2015;Miyauchi et al., 2020;Resl et al., 2021). Their symbiotic structure also does not penetrate the plant cell walls (Bonfante and Genre, 2010;Balestrini and Bonfante, 2014); which suggests that the loss of the ability to degrade cell walls in lichen-forming and mycorrhizal fungi is a consequence of their symbiotic fungus–host interface.

Many MFS-type transporters were also lost, although carbohydrate movement from the algal partner to the lichen-forming fungus is important in lichen symbiosis (Smith, 1968). Although previous lichen genomic studies have also reported

reductions in sugar transporters (Wang et al., 2014;Armaleo et al., 2019), we found that these losses were common in lichen lineages, not only in each species. Lichen-forming fungi likely use specific transporters, as they receive different mobile carbohydrates, such as ribitol, sorbitol, and glucose, from their algal partners (Richardson et al., 1968). We propose that the extensive loss of sugar transporter genes is a result of the dispensability of common sugar transporters. The upregulation of ribitol transporter genes during the late stage of *G. flavorubescens* symbiosis supports this hypothesis, whereas other sugar transporters exhibited no significant changes. Algal partners export carbohydrates only during symbiosis (Smith, 1968), such that the completion of lichen symbiosis and the initiation of nutrient exchange occur between 72 h and 4 weeks PCI.

As a result of contractions in diverse gene families (e.g., PCWDEs, sugar transporters, and TFs), the genome size and total number of genes of lichen-forming fungi are lower than those of other fungal species, especially plant-associated fungi. This is unsurprising because gene losses are widespread among all organisms (Albalat and Canestro, 2016), and genome reduction is a dominant evolutionary process resulting in the loss of non-functionalized genes (Wolf and Koonin, 2013). Symbionts have significantly reduced genomes due to their dependence on photosynthetic partners (McCutcheon and Moran, 2011;Pogoda et al., 2018). The loss of energy-production genes in the mitochondrial genomes of lichen-forming fungi is an example of this reductive evolution (Pogoda et al., 2018). Although to date genome streamlining has been evaluated only in bacterial and mitochondrial genomes, we suggest that it can also occur in the nuclear genome. The loss of TF genes is another consequence of this evolutionary mechanism, and the loss of many

genes and dependency on the host may influence the size of TF families.

Both massive gene losses and independent gene gains have occurred in lichen-forming fungi. Each species has many unique genes. In this study, we attempted to identify lichen-specific core genes, but found only one orthogroup. Similarly, mycorrhizal fungi lack universal symbiosis genes (Kuo et al., 2014); instead, newly gained lichen-specific genes, including lineage- and species-specific genes, appear to be more related to their lifestyles, with high expression during early symbiosis, when they influence their partners (Ahmadjian et al., 1978). Most SSPs were also genus- or species-specific and had similar expression patterns. The transient expression of each set of specific genes suggests that they are activated differently during each period of the early stage. Based on the sequential expression of effector proteins in the plant pathogen *Colletotrichum higginsianum*, different SSPs may be involved during each stage of host-pathogen interaction (Kleemann et al., 2012). Because the functions of most lichen-specific genes are unknown, and the SSPs of symbionts play essential roles in maintaining mycorrhizal symbiotic relationships (Kloppholz et al., 2011;Plett et al., 2011), lineage- and species-specific genes may play significant roles in the establishment of lichen symbiosis. The functionally annotated conserved genes that may be involved in maintaining their symbiotic relationships were induced mainly during the late stage of lichen symbiosis, when recognition and contact with the partner are completed, growth is continued (Athukorala et al., 2014), and metabolic processes such as nutrient exchange are activated. For example, genes involved in glycolysis are expressed differentially during the late stage of lichen symbiosis, allowing the use of sugars obtained from photosynthetic partners. These findings indicate that lichen-specific genes and

conserved genes play roles in different stages of lichen resynthesis.

The evolutionary pattern of gene loss of lichen-forming fungi is similar to that of ectomycorrhizal fungi. Both symbionts lost the ability to degrade cell walls and gained lineage-specific genes that may be involved in symbiosis; this evolutionary process is well known in mycorrhizal fungi (Kohler et al., 2015;Miyauchi et al., 2020). However, ectomycorrhizal fungi still retain PCWDEs, including GH28, GH88, CE8, and GH30, which are induced in mycorrhizal symbiosis for host cell wall modification (Kohler et al., 2015), whereas lichen-forming fungi have lost most of these genes. The number of effector proteins remains unknown (Harris et al., 2020). However, because lichen-forming fungi have fewer SSPs than mycorrhizal fungi or other plant-associated fungi, we suggest that the number of SSPs depends on the complexity of their host. Because green algae are the ancestors of land plants and have evolved to become more complex in terms of cellular organization (Turmel et al., 2007), lichen-forming fungi may not require many SSPs to interact with the defense mechanism of their living host. Most gene family expansion occurred during the speciation of each mycorrhizal fungus, which suggests that lichen-forming fungi and mycorrhizal fungi underwent different unknown evolutionary processes to develop their lifestyles.

This study is the first comparative analysis of diverse lichen-forming fungi using whole genomes to clarify elements of lichen symbiosis. We found that the loss of non-essential genes, such as specific families of PCWDEs, sugar transporters, and TFs, streamlined the genomes of lichen-forming fungi, providing new insights on lichen symbiosis. Lineage- and species-specific genes, including SSPs, play a role during the early stage of lichen symbiosis, and may be involved in recognition

between lichen-forming fungi and their partners. These findings advance understating of the evolution of symbiotic lifestyles and the determinants contributing to lichen symbiosis. Genomic resources may contribute to future molecular functional studies of the unrevealed biological functions of significant factors in lichen symbiosis.

LITERATURE CITED

Ahmadjian, V., Jacobs, J.B., and Russell, L.A. (1978). Scanning electron microscope study of early lichen synthesis. *Science* **200**: 1062-1064.

Albalat, R., and Canestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* **17**: 379-391.

Alexa, A., and Rahnenfuhrer, J. (2016). topGO: Enrichment analysis for Gene Ontology. R package version 2.28. 0. *BioConductor. Published online*.

Armaleo, D., Müller, O., Lutzoni, F., Andrésson, S., Blanc, G., Bode, H.B., Collart, F.R., Dal Grande, F., Dietrich, F., and Grigoriev, I.V. (2019). The lichen symbiosis re-viewed through the genomes of *Cladonia grayi* and its algal partner *Asterochloris glomerata*. *BMC Genome.* **20**: 1-33.

Armstrong, E.E., Prost, S., Ertz, D., Westberg, M., Frisch, A., and Bendiksby, M. (2018). Draft genome sequence and annotation of the lichen-forming fungus *Arthonia radiata*. *Genome Announc.* **6**: e00281-00218.

Athukorala, S.N., Huebner, E., and Piercey-Normore, M.D. (2014). Identification and comparison of the 3 early stages of resynthesis for the lichen *Cladonia rangiferina*. *Can. J. Microbiol.* **60**: 41-52.

Balestrini, R., and Bonfante, P. (2014). Cell wall remodeling in mycorrhizal symbiosis: a way towards biotrophism. *Front. Plant Sci.* **5**: 237.

Boluda, C., Rico, V., Divakar, P., Nadyeina, O., Myllys, L., Mcmullin, R., Zamora, J., Scheidegger, C., and Hawksworth, D. (2019). Evaluating methodologies for species delimitation: the mismatch between phenotypes and genotypes

in lichenized fungi (Bryoria sect. Implexae, Parmeliaceae). *Pers.: Mol. Phylogeny Evol. Fungi* **42**: 75-100.

Bonfante, P., and Genre, A. (2010). Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat. Commun.* **1**: 48.

Boustie, J., and Grube, M. (2005). Lichens—a promising source of bioactive secondary metabolites. *Plant Genet. Resour.* **3**: 273-287.

Branco, S., Gladieux, P., Ellison, C.E., Kuo, A., Labutti, K., Lipzen, A., Grigoriev, I.V., Liao, H.L., Vilgalys, R., Peay, K.G., Taylor, J.W., and Bruns, T.D. (2015). Genetic isolation between two recently diverged populations of a symbiotic fungus. *Mol. Ecol.* **24**: 2747-2758.

Calcott, M.J., Ackerley, D.F., Knight, A., Keyzers, R.A., and Owen, J.G. (2018). Secondary metabolism in the lichen symbiosis. *Chem. Soc. Rev.* **47**: 1730-1760.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.

Černajová, I., and Škaloud, P. (2019). The first survey of Cystobasidiomycete yeasts in the lichen genus *Cladonia*; with the description of *Lichenozyma pisutiana* gen. nov., sp. nov. *Fungal biol.* **123**: 625-637.

Corrochano, L.M., Kuo, A., Marcet-Houben, M., Polaino, S., Salamov, A., Villalobos-Escobedo, J.M., Grimwood, J., Alvarez, M.I., Avalos, J., Bauer, D., Benito, E.P., Benoit, I., Burger, G., Camino, L.P., Canovas, D., Cerda-Olmedo, E., Cheng, J.F., Dominguez, A., Elias, M., Eslava, A.P., Glaser, F., Gutierrez, G., Heitman, J., Henrissat, B., Iturriaga, E.A., Lang, B.F., Lavin,

J.L., Lee, S.C., Li, W., Lindquist, E., Lopez-Garcia, S., Luque, E.M., Marcos, A.T., Martin, J., Mccluskey, K., Medina, H.R., Miralles-Duran, A., Miyazaki, A., Munoz-Torres, E., Oguiza, J.A., Ohm, R.A., Olmedo, M., Orejas, M., Ortiz-Castellanos, L., Pisabarro, A.G., Rodriguez-Romero, J., Ruiz-Herrera, J., Ruiz-Vazquez, R., Sanz, C., Schackwitz, W., Shahriari, M., Shelest, E., Silva-Franco, F., Soanes, D., Syed, K., Tagua, V.G., Talbot, N.J., Thon, M.R., Tice, H., De Vries, R.P., Wiebenga, A., Yadav, J.S., Braun, E.L., Baker, S.E., Garre, V., Schmutz, J., Horwitz, B.A., Torres-Martinez, S., Idnurm, A., Herrera-Estrella, A., Gabaldon, T., and Grigoriev, I.V. (2016). Expansion of signal transduction pathways in fungi by extensive genome duplication. *Curr. Biol.* **26**: 1577-1584.

Cuomo, C.A., Guldener, U., Xu, J.R., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L.J., Baker, S.E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.L., Decaprio, D., Gale, L.R., Gnerre, S., Goswami, R.S., Hammond-Kosack, K., Harris, L.J., Hilburn, K., Kennell, J.C., Kroken, S., Magnuson, J.K., Mannhaupt, G., Mauceli, E., Mewes, H.W., Mitterbauer, R., Muehlbauer, G., Munsterkotter, M., Nelson, D., O'donnell, K., Ouellet, T., Qi, W., Quesneville, H., Roncero, M.I., Seong, K.Y., Tetko, I.V., Urban, M., Waalwijk, C., Ward, T.J., Yao, J., Birren, B.W., and Kistler, H.C. (2007). The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**: 1400-1402.

Dal Grande, F., Meiser, A., Tzovaras, B.G., Jürgen, O., Ebersberger, I., and Schmitt, I. (2018). The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo. *Lichenologist* **50**: 329-340.

Darby, C.A., Stolzer, M., Ropp, P.J., Barker, D., and Durand, D. (2017). Xenolog classification. *Bioinformatics* **33**: 640-649.

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., 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., Rehmeyer, C., Li, W., 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.

Duplessis, S., Cuomo, C.A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L., Chiu, R., Coutinho, P.M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M.G., Kodira, C.D., Kohler, A., Kues, U., Lindquist, E.A., Lucas, S.M., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J.L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A.A., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G.A., Henrissat, B., Van De Peer, Y., Rouze, P., Ellis, J.G., Dodds, P.N., Schein, J.E., Zhong, S., Hamelin, R.C., Grigoriev, I.V., Szabo, L.J., and Martin, F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 9166-9171.

Eastwood, D.C., Floudas, D., Binder, M., Majcherczyk, A., Schneider, P., Aerts, A., Asiegbu, F.O., Baker, S.E., Barry, K., Bendiksby, M., Blumentritt, M., Coutinho, P.M., Cullen, D., De Vries, R.P., Gathman, A., Goodell, B., Henrissat, B., Ihrmark, K., Kauserud, H., Kohler, A., Labutti, K., Lapidus,

A., Lavin, J.L., Lee, Y.H., Lindquist, E., Lilly, W., Lucas, S., Morin, E., Murat, C., Oguiza, J.A., Park, J., Pisabarro, A.G., Riley, R., Rosling, A., Salamov, A., Schmidt, O., Schmutz, J., Skrede, I., Stenlid, J., Wiebenga, A., Xie, X., Kues, U., Hibbett, D.S., Hoffmeister, D., Hogberg, N., Martin, F., Grigoriev, I.V., and Watkinson, S.C. (2011). The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* **333**: 762-765.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**: 157.

Espagne, E., Lespinet, O., Malagnac, F., Da Silva, C., Jaillon, O., Porcel, B.M., Couloux, A., Aury, J.M., Segurens, B., Poulain, J., Anthouard, V., Grossetete, S., Khalili, H., Coppin, E., Dequard-Chablat, M., Picard, M., Contamine, V., Arnaise, S., Bourdais, A., Berteaux-Lecellier, V., Gautheret, D., De Vries, R.P., Battaglia, E., Coutinho, P.M., Danchin, E.G., Henrissat, B., Khoury, R.E., Sainsard-Chanet, A., Boivin, A., Pinan-Lucarre, B., Sellem, C.H., Debuchy, R., Wincker, P., Weissenbach, J., and Silar, P. (2008). The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biol.* **9**: R77.

Fedorova, N.D., Moktali, V., and Medema, M.H. (2012). Bioinformatics approaches and software for detection of secondary metabolic gene clusters. *Methods Mol. Biol.* **944**: 23-45.

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331-368.

Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B. (1998). The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* **440**: 325-331.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., Fitzhugh, W., Ma, L.J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C.B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M.A., Werner-Washburne, M., Selitrennikoff, C.P., Kinsey, J.A., Braun, E.L., Zelter, A., Schulte, U., Kothe, G.O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R.L., Perkins, D.D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R.J., Osmani, S.A., Desouza, C.P., Glass, L., Orbach, M.J., Berglund, J.A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D.O., Alex, L.A., Mannhaupt, G., Ebbole, D.J., Freitag, M., Paulsen, I., Sachs, M.S., Lander, E.S., Nusbaum, C., and Birren, B. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859-868.

Galagan, J.E., Calvo, S.E., Cuomo, C., Ma, L.J., Wortman, J.R., Batzoglou, S., Lee, S.I., Basturkmen, M., Spevak, C.C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G.H., Draht, O., Busch, S., D'enfert, C., Bouchier, C., Goldman, G.H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J.H., Yu, J., Vienken, K., Pain, A.,

Freitag, M., Selker, E.U., Archer, D.B., Penalva, M.A., Oakley, B.R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W.C., Denning, D.W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M.S., Osmani, S.A., and Birren, B.W. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105-1115.

Gargas, A., Depriest, P.T., Grube, M., and Tehler, A. (1995). Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. *Science* **268**: 1492-1495.

Gianoulis, T.A., Griffin, M.A., Spakowicz, D.J., Dunican, B.F., Alpha, C.J., Sboner, A., Sismour, A.M., Kodira, C., Egholm, M., Church, G.M., Gerstein, M.B., and Strobel, S.A. (2012). Genomic analysis of the hydrocarbon-producing, cellulolytic, endophytic fungus *Ascocoryne sarcoides*. *PLoS Genet.* **8**: e1002558.

Goodwin, S.B., M'barek S, B., Dhillon, B., Wittenberg, A.H., Crane, C.F., Hane, J.K., Foster, A.J., Van Der Lee, T.A., Grimwood, J., Aerts, A., Antoniw, J., Bailey, A., Bluhm, B., Bowler, J., Bristow, J., Van Der Burgt, A., Canto-Canche, B., Churchill, A.C., Conde-Ferraez, L., Cools, H.J., Coutinho, P.M., Csukai, M., Dehal, P., De Wit, P., Donzelli, B., Van De Geest, H.C., Van Ham, R.C., Hammond-Kosack, K.E., Henrissat, B., Kilian, A., Kobayashi, A.K., Koopmann, E., Kourmpetis, Y., Kuzniar, A., Lindquist, E., Lombard, V., Maliepaard, C., Martins, N., Mehrabi, R., Nap, J.P., Ponomarenko, A., Rudd, J.J., Salamov, A., Schmutz, J., Schouten, H.J., Shapiro, H., Stergiopoulos, I., Torriani, S.F., Tu, H., De Vries, R.P., Waalwijk, C., Ware, S.B., Wiebenga,

A., Zwiers, L.H., Oliver, R.P., Grigoriev, I.V., and Kema, G.H. (2011). Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet.* **7**: e1002070.

Grube, M., and Wedin, M. (2016). Lichenized fungi and the evolution of symbiotic organization. *Microbiol. Spectr.* **4**:6.

Harris, J.M., Balint-Kurti, P., Bede, J.C., Day, B., Gold, S., Goss, E.M., Grenville-Briggs, L.J., Jones, K.M., Wang, A.M., Wang, Y.C., Mitra, R.M., Sohn, K.H., and Alvarez, M.E. (2020). What are the top 10 unanswered questions in molecular plant-microbe interactions? *Mol. Plant Microbe Interact.* **33**: 1354-1365.

Hawksworth, D.L., and Grube, M. (2020). Lichens redefined as complex ecosystems. *New Phytol.* **227**: 1281.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., and Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**: 835-845.

Hill, D.J., and Ahmadjian, V. (1972). Relationship between carbohydrate movement and symbiosis in lichens with green-algae. *Planta* **103**: 267–277.

Honegger, R. (1984). Cytological aspects of the mycobiont–phycobiont relationship in lichens: haustorial types, phycobiont cell wall types, and the ultrastructure of the cell surface layers in some cultured and symbiotic myco-and phycobionts. *Lichenologist* **16**: 111-127.

Honegger, R. (1986). Ultrastructural studies in lichens. *New phytol.* **103**: 797-808.

Honegger, R. (2000). Simon Schwendener (1829-1919) and the dual hypothesis of lichens. *Bryologist* **103**: 307-313.

Honegger, R. (2006). "Water relations in lichens." in *Fungi in the environment*, GM Gadd, SC Watkinson, P Dyer, eds. (Cambridge, UK: Cambridge University Press), 185-200.

Janbon, G., Ormerod, K.L., Paulet, D., Byrnes, E.J., 3rd, Yadav, V., Chatterjee, G., Mullapudi, N., Hon, C.C., Billmyre, R.B., Brunel, F., Bahn, Y.S., Chen, W., Chen, Y., Chow, E.W., Coppee, J.Y., Floyd-Averette, A., Gaillardin, C., Gerik, K.J., Goldberg, J., Gonzalez-Hilarion, S., Gujja, S., Hamlin, J.L., Hsueh, Y.P., Ianiri, G., Jones, S., Kodira, C.D., Kozubowski, L., Lam, W., Marra, M., Mesner, L.D., Mieczkowski, P.A., Moyrand, F., Nielsen, K., Proux, C., Rossignol, T., Schein, J.E., Sun, S., Wollschlaeger, C., Wood, I.A., Zeng, Q., Neuveglise, C., Newlon, C.S., Perfect, J.R., Lodge, J.K., Idnurm, A., Stajich, J.E., Kronstad, J.W., Sanyal, K., Heitman, J., Fraser, J.A., Cuomo, C.A., and Dietrich, F.S. (2014). Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. grubii reveals complex RNA expression and microevolution leading to virulence attenuation. *PLoS Genet.* **10**: e1004261.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., Mcanulla, C., Mcwilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Jones, T., Federspiel, N.A., Chibana, H., Dungan, J., Kalman, S., Magee, B.B., Newport, G., Thorstenson, Y.R., Agabian, N., Magee, P.T., Davis, R.W., and Scherer, S. (2004). The diploid genome sequence of *Candida albicans*. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 7329-7334.

Joneson, S., Armaleo, D., and Lutzoni, F. (2011). Fungal and algal gene expression in early developmental stages of lichen-symbiosis. *Mycologia* **103**: 291-306.

Kamper, J., Kahmann, R., Bolker, M., Ma, L.J., Brefort, T., Saville, B.J., Banuett, F., Kronstad, J.W., Gold, S.E., Muller, O., Perlin, M.H., Wosten, H.A., De Vries, R., Ruiz-Herrera, J., Reynaga-Pena, C.G., Snetselaar, K., Mccann, M., Perez-Martin, J., Feldbrugge, M., Basse, C.W., Steinberg, G., Ibeas, J.I., Holloman, W., Guzman, P., Farman, M., Stajich, J.E., Sentandreu, R., Gonzalez-Prieto, J.M., Kennell, J.C., Molina, L., Schirawski, J., Mendoza-Mendoza, A., Greilinger, D., Munch, K., Rossel, N., Scherer, M., Vranes, M., Ladendorf, O., Vincon, V., Fuchs, U., Sandrock, B., Meng, S., Ho, E.C., Cahill, M.J., Boyce, K.J., Klose, J., Klosterman, S.J., Deelstra, H.J., Ortiz-Castellanos, L., Li, W., Sanchez-Alonso, P., Schreier, P.H., Hauser-Hahn, I., Vaupel, M., Koopmann, E., Friedrich, G., Voss, H., Schluter, T., Margolis, J., Platt, D., Swimmer, C., Gnirke, A., Chen, F., Vysotskaia, V., Mannhaupt, G., Guldener, U., Munsterkotter, M., Haase, D., Oesterheld, M., Mewes, H.W., Mauceli, E.W., Decaprio, D., Wade, C.M., Butler, J., Young, S., Jaffe, D.B., Calvo, S., Nusbaum, C., Galagan, J., and Birren, B.W. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**: 97-101.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**: R36.

Kim, K.T., Jeon, J., Choi, J., Cheong, K., Song, H., Choi, G., Kang, S., and Lee, Y.H. (2016). Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Front. Plant Sci.* **7**: 186.

Kleemann, J., Rincon-Rivera, L.J., Takahara, H., Neumann, U., Ver Loren Van Themaat, E., Van Der Does, H.C., Hacquard, S., Stuber, K., Will, I., Schmalenbach, W., Schmelzer, E., and O'connell, R.J. (2012). Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* **8**: e1002643.

Kloppholz, S., Kuhn, H., and Requena, N. (2011). A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr. Biol.* **21**: 1204-1209.

Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canback, B., Choi, C., Cichocki, N., Clum, A., Colpaert, J., Copeland, A., Costa, M.D., Dore, J., Floudas, D., Gay, G., Girlanda, M., Henrissat, B., Herrmann, S., Hess, J., Hogberg, N., Johansson, T., Khouja, H.R., Labutti, K., Lahrmann, U., Levasseur, A., Lindquist, E.A., Lipzen, A., Marmeisse, R., Martino, E., Murat, C., Ngan, C.Y., Nehls, U., Plett, J.M., Pringle, A., Ohm, R.A., Perotto, S., Peter, M., Riley, R., Rineau, F., Ruytinx, J., Salamov, A., Shah, F., Sun, H., Tarkka, M., Tritt, A., Veneault-Fourrey, C., Zuccaro, A., Mycorrhizal Genomics Initiative, C., Tunlid, A., Grigoriev, I.V., Hibbett, D.S., and Martin,

F. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat. Genet. **47**: 410-415.

Kranner, I., Beckett, R., Hochman, A., and Nash, T.H. (2008). Desiccation-tolerance in lichens: a review. *Bryologist* **111**: 576-593.

Kubicek, C.P., Starr, T.L., and Glass, N.L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annu. Rev. Phytopathol.* **52**: 427-451.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**: 1870-1874.

Kuo, A., Kohler, A., Martin, F.M., and Grigoriev, I.V. (2014). Expanding genomics of mycorrhizal symbiosis. *Front. Microbiol.* **5**: 582.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* **5**: R12.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., Mcwilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Lionetti, V., and Metraux, J.P. (2014). Plant cell wall in pathogenesis, parasitism and symbiosis. *Front. Plant. Sci.* **5**: 612.

Liu, R., Kim, W., Paguirigan, J.a.A., Jeong, M.-H., and Hur, J.-S. (2021). Establishment of *Agrobacterium tumefaciens*-mediated transformation of *Cladonia macilenta*, a model lichen-forming fungus. *J. Fungi* **7**: 252.

Lumbsch, H.T., and Leavitt, S.D. (2011). Goodbye morphology? A paradigm shift in the delimitation of species in lichenized fungi. *Fungal Divers.* **50**: 59-72.

Lutzoni, F., Pagel, M., and Reeb, V. (2001). Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* **411**: 937-940.

Ma, L.J., Ibrahim, A.S., Skory, C., Grabherr, M.G., Burger, G., Butler, M., Elias, M., Idnurm, A., Lang, B.F., Sone, T., Abe, A., Calvo, S.E., Corrochano, L.M., Engels, R., Fu, J., Hansberg, W., Kim, J.M., Kodira, C.D., Koehrsen, M.J., Liu, B., Miranda-Saavedra, D., O'leary, S., Ortiz-Castellanos, L., Poulter, R., Rodriguez-Romero, J., Ruiz-Herrera, J., Shen, Y.Q., Zeng, Q., Galagan, J., Birren, B.W., Cuomo, C.A., and Wickes, B.L. (2009). Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* **5**: e1000549.

Ma, L.J., Van Der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G., Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-

Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C., and Rep, M. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367-373.

Martin, F., Aerts, A., Ahren, D., Brun, A., Danchin, E.G., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., Salamov, A., Shapiro, H.J., Wuyts, J., Blaudez, D., Buee, M., Brokstein, P., Canback, B., Cohen, D., Courty, P.E., Coutinho, P.M., Delaruelle, C., Detter, J.C., Deveau, A., Difazio, S., Duplessis, S., Fraissinet-Tachet, L., Lucic, E., Frey-Klett, P., Fourrey, C., Feussner, I., Gay, G., Grimwood, J., Hoegger, P.J., Jain, P., Kilaru, S., Labbe, J., Lin, Y.C., Legue, V., Le Tacon, F., Marmeisse, R., Melayah, D., Montanini, B., Muratet, M., Nehls, U., Niculita-Hirzel, H., Oudot-Le Secq, M.P., Peter, M., Quesneville, H., Rajashekar, B., Reich, M., Rouhier, N., Schmutz, J., Yin, T., Chalot, M., Henrissat, B., Kues, U., Lucas, S., Van De Peer, Y., Podila, G.K., Polle, A., Pukkila, P.J., Richardson, P.M., Rouze, P., Sanders, I.R., Stajich, J.E., Tunlid, A., Tuskan, G., and Grigoriev, I.V. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88-92.

Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., Montanini, B., Morin, E., Noel, B., Percudani, R., Porcel, B., Rubini, A., Amicucci, A., Amselem, J., Anthouard, V., Arcioni, S., Artiguenave, F., Aury, J.M., Ballario, P., Bolchi, A., Brenna, A., Brun, A., Buee, M., Cantarel, B., Chevalier, G., Couloux, A., Da Silva, C., Denoeud, F., Duplessis, S.,

Ghignone, S., Hilselberger, B., Iotti, M., Marcais, B., Mello, A., Miranda, M., Pacioni, G., Quesneville, H., Riccioni, C., Ruotolo, R., Splivallo, R., Stocchi, V., Tisserant, E., Viscomi, A.R., Zambonelli, A., Zampieri, E., Henrissat, B., Lebrun, M.H., Paolocci, F., Bonfante, P., Ottonello, S., and Wincker, P. (2010). Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033-1038.

Martinez, D.A., Oliver, B.G., Graser, Y., Goldberg, J.M., Li, W., Martinez-Rossi, N.M., Monod, M., Shelest, E., Barton, R.C., Birch, E., Brakhage, A.A., Chen, Z., Gurr, S.J., Heiman, D., Heitman, J., Kosti, I., Rossi, A., Saif, S., Samalova, M., Saunders, C.W., Shea, T., Summerbell, R.C., Xu, J., Young, S., Zeng, Q., Birren, B.W., Cuomo, C.A., and White, T.C. (2012). Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. *MBio* **3**: e00259-00212.

Mccutcheon, J.P., and Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**: 13-26.

Mead, O.L., and Gueidan, C. (2020). Complete genome sequence of an Australian strain of the lichen-forming fungus *Endocarpon pusillum* (Hedwig). *Microbiol. Resour. Announc.* **9**: e01079-01020.

Millanes, A.M., Diederich, P., and Wedin, M. (2016). Cyphobasidium gen. nov., a new lichen-inhabiting lineage in the Cystobasidiomycetes (Pucciniomycotina, Basidiomycota, Fungi). *Fungal Biol.* **120**: 1468-1477.

Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sanchez-Garcia, M., Morin, E., Andreopoulos, B., Barry, K.W., Bonito, G., Buee, M., Carver, A., Chen,

C., Cichocki, N., Clum, A., Culley, D., Crous, P.W., Fauchery, L., Girlanda, M., Hayes, R.D., Keri, Z., Labutti, K., Lipzen, A., Lombard, V., Magnuson, J., Maillard, F., Murat, C., Nolan, M., Ohm, R.A., Pangilinan, J., Pereira, M.F., Perotto, S., Peter, M., Pfister, S., Riley, R., Sitrit, Y., Stielow, J.B., Szollosi, G., Zifcakova, L., Stursova, M., Spatafora, J.W., Tedersoo, L., Vaario, L.M., Yamada, A., Yan, M., Wang, P., Xu, J., Bruns, T., Baldrian, P., Vilgalys, R., Dunand, C., Henrissat, B., Grigoriev, I.V., Hibbett, D., Nagy, L.G., and Martin, F.M. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat. Commun.* **11**: 5125.

Morales-Cruz, A., Amrine, K.C., Blanco-Ulate, B., Lawrence, D.P., Travadon, R., Rolshausen, P.E., Baumgartner, K., and Cantu, D. (2015). Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genome.* **16**: 469.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**: W182-185.

Munoz, J.F., Gauthier, G.M., Desjardins, C.A., Gallo, J.E., Holder, J., Sullivan, T.D., Marty, A.J., Carmen, J.C., Chen, Z., Ding, L., Gujja, S., Magrini, V., Misas, E., Mitreva, M., Priest, M., Saif, S., Whiston, E.A., Young, S., Zeng, Q., Goldman, W.E., Mardis, E.R., Taylor, J.W., Mcewen, J.G., Clay, O.K., Klein, B.S., and Cuomo, C.A. (2015). The dynamic genome and transcriptome of

the human fungal pathogen *Blastomyces* and close relative *Emmonsia*. *PLoS Genet.* **11**: e1005493.

Nash, T.H. (2008). *Lichen biology*. New York: Cambridge University Press.

Nazem-Bokaee, H., Hom, E.F., Warden, A.C., Mathews, S., and Gueidan, C. (2021). Towards a systems biology approach to understanding the lichen symbiosis: opportunities and challenges of implementing network modelling. *Front. Microbiol.* **12**: 1028.

Neafsey, D.E., Barker, B.M., Sharpton, T.J., Stajich, J.E., Park, D.J., Whiston, E., Hung, C.Y., Mcmahan, C., White, J., Sykes, S., Heiman, D., Young, S., Zeng, Q., Abouelleil, A., Aftuck, L., Bessette, D., Brown, A., Fitzgerald, M., Lui, A., Macdonald, J.P., Priest, M., Orbach, M.J., Galgiani, J.N., Kirkland, T.N., Cole, G.T., Birren, B.W., Henn, M.R., Taylor, J.W., and Rounsley, S.D. (2010). Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. *Genome Res.* **20**: 938-946.

Nelsen, M.P., Lucking, R., Boyce, C.K., Lumbsch, H.T., and Ree, R.H. (2020). The macroevolutionary dynamics of symbiotic and phenotypic diversification in lichens. Proc. Natl. Acad. Sci. U. S. A. **117**: 21495-21503.

Nelson, D.R. (2009). The cytochrome p450 homepage. *Hum. Genomics* **4**: 59-65.

Nierman, W.C., Pain, A., Anderson, M.J., Wortman, J.R., Kim, H.S., Arroyo, J., Berriman, M., Abe, K., Archer, D.B., Bermejo, C., Bennett, J., Bowyer, P., Chen, D., Collins, M., Coulsen, R., Davies, R., Dyer, P.S., Farman, M., Fedorova, N., Fedorova, N., Feldblyum, T.V., Fischer, R., Fosker, N., Fraser, A., Garcia, J.L., Garcia, M.J., Goble, A., Goldman, G.H., Gomi, K., Griffith-Jones, S., Gwilliam, R., Haas, B., Haas, H., Harris, D., Horiuchi, H., Huang,

J., Humphray, S., Jimenez, J., Keller, N., Khouri, H., Kitamoto, K., Kobayashi, T., Konzack, S., Kulkarni, R., Kumagai, T., Lafon, A., Latge, J.P., Li, W., Lord, A., Lu, C., Majoros, W.H., May, G.S., Miller, B.L., Mohamoud, Y., Molina, M., Monod, M., Mouyna, I., Mulligan, S., Murphy, L., O'neil, S., Paulsen, I., Penalva, M.A., Pertea, M., Price, C., Pritchard, B.L., Quail, M.A., Rabbinowitsch, E., Rawlins, N., Rajandream, M.A., Reichard, U., Renauld, H., Robson, G.D., Rodriguez De Cordoba, S., Rodriguez-Pena, J.M., Ronning, C.M., Rutter, S., Salzberg, S.L., Sanchez, M., Sanchez-Ferrero, J.C., Saunders, D., Seeger, K., Squares, R., Squares, S., Takeuchi, M., Tekaia, F., Turner, G., Vazquez De Aldana, C.R., Weidman, J., White, O., Woodward, J., Yu, J.H., Fraser, C., Galagan, J.E., Asai, K., Machida, M., Hall, N., Barrell, B., and Denning, D.W. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**: 1151-1156.

O'connell, R.J., Thon, M.R., Hacquard, S., Amyotte, S.G., Kleemann, J., Torres, M.F., Damm, U., Buiate, E.A., Epstein, L., Alkan, N., Altmuller, J., Alvarado-Balderrama, L., Bauser, C.A., Becker, C., Birren, B.W., Chen, Z., Choi, J., Crouch, J.A., Duvick, J.P., Farman, M.A., Gan, P., Heiman, D., Henrissat, B., Howard, R.J., Kabbage, M., Koch, C., Kracher, B., Kubo, Y., Law, A.D., Lebrun, M.H., Lee, Y.H., Miyara, I., Moore, N., Neumann, U., Nordstrom, K., Panaccione, D.G., Panstruga, R., Place, M., Proctor, R.H., Prusky, D., Rech, G., Reinhardt, R., Rollins, J.A., Rounsley, S., Schardl, C.L., Schwartz, D.C., Shenoy, N., Shirasu, K., Sikhakolli, U.R., Stuber, K., Sukno, S.A., Sweigard, J.A., Takano, Y., Takahara, H., Trail, F., Van Der Does, H.C., Voll, L.M., Will, I., Young, S., Zeng, Q., Zhang, J., Zhou, S., Dickman, M.B.,

Schulze-Lefert, P., Ver Loren Van Themaat, E., Ma, L.J., and Vaillancourt, L.J. (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* **44**: 1060-1065.

Ohm, R.A., Riley, R., Salamov, A., Min, B., Choi, I.G., and Grigoriev, I.V. (2014). Genomics of wood-degrading fungi. *Fungal Genet. Biol.* **72**: 82-90.

Olson, A., Aerts, A., Asiegbu, F., Belbahri, L., Bouzid, O., Broberg, A., Canback, B., Coutinho, P.M., Cullen, D., Dalman, K., Deflorio, G., Van Diepen, L.T., Dunand, C., Duplessis, S., Durling, M., Gonthier, P., Grimwood, J., Fossdal, C.G., Hansson, D., Henrissat, B., Hietala, A., Himmelstrand, K., Hoffmeister, D., Hogberg, N., James, T.Y., Karlsson, M., Kohler, A., Kues, U., Lee, Y.H., Lin, Y.C., Lind, M., Lindquist, E., Lombard, V., Lucas, S., Lunden, K., Morin, E., Murat, C., Park, J., Raffaello, T., Rouze, P., Salamov, A., Schmutz, J., Solheim, H., Stahlberg, J., Velez, H., De Vries, R.P., Wiebenga, A., Woodward, S., Yakovlev, I., Garbelotto, M., Martin, F., Grigoriev, I.V., and Stenlid, J. (2012). Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytol.* **194**: 1001-1013.

Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. (1998). Major facilit.ator superfamily. *Microbiol. Mol. Biol. Rev.* **62**: 1-34.

Park, J., Lee, S., Choi, J., Ahn, K., Park, B., Park, J., Kang, S., and Lee, Y.H. (2008). Fungal cytochrome P450 database. *BMC Genome.* **9**: 402.

Park, S.-Y., Jeong, M.-H., Wang, H.-Y., Kim, J.A., Yu, N.-H., Kim, S., Cheong, Y.H., Kang, S., Lee, Y.-H., and Hur, J.-S. (2013a). *Agrobacterium tumefaciens*-

mediated transformation of the lichen fungus, *Umbilicaria muehlenbergii*. *PLoS One* **8**: e83896.

Park, S.Y., Choi, J., Kim, J.A., Jeong, M.H., Kim, S., Lee, Y.H., and Hur, J.S. (2013b). Draft genome sequence of *Cladonia macilenta* KoLRI003786, a lichen-forming fungus producing biruloquinone. *Genome Announc.* **1**: e00695-00613..

Park, S.Y., Choi, J., Kim, J.A., Yu, N.H., Kim, S., Kondratyuk, S.Y., Lee, Y.H., and Hur, J.S. (2013c). Draft genome sequence of lichen-forming fungus *Caloplaca flavorubescens* strain KoLRI002931. *Genome Announc.* 1: e00678-13.

Park, S.Y., Choi, J., Lee, G.W., Jeong, M.H., Kim, J.A., Oh, S.O., Lee, Y.H., and Hur, J.S. (2014a). Draft genome sequence of *Umbilicaria muehlenbergii* KoLRILF000956, a lichen-forming fungus amenable to genetic manipulation. *Genome Announc.* **2**: e00357-14.

Park, S.Y., Choi, J., Lee, G.W., Kim, J.A., Oh, S.O., Jeong, M.H., Yu, N.H., Kim, S., Lee, Y.H., and Hur, J.S. (2014b). Draft genome sequence of lichen-forming fungus *Cladonia metacorallifera* strain KoLRI002260. *Genome Announc.* **2**: e01065-13.

Park, S.Y., Choi, J., Lee, G.W., Park, C.H., Kim, J.A., Oh, S.O., Lee, Y.H., and Hur, J.S. (2014c). Draft genome sequence of *Endocarpon pusillum* strain KoLRILF000583. *Genome Announc.* **2**: e00452-14.

Patel, R.K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7**: e30619.

Pellegrin, C., Morin, E., Martin, F.M., and Veneault-Fourrey, C. (2015). Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Front. Microbiol.* **6**: 1278.

Peter, M., Kohler, A., Ohm, R.A., Kuo, A., Krutzmann, J., Morin, E., Arend, M., Barry, K.W., Binder, M., Choi, C., Clum, A., Copeland, A., Grisel, N., Haridas, S., Kipfer, T., Labutti, K., Lindquist, E., Lipzen, A., Maire, R., Meier, B., Mihaltcheva, S., Molinier, V., Murat, C., Poggeler, S., Quandt, C.A., Sperisen, C., Tritt, A., Tisserant, E., Crous, P.W., Henrissat, B., Nehls, U., Egli, S., Spatafora, J.W., Grigoriev, I.V., and Martin, F.M. (2016). Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nat. Commun*. **7**: 12662.

Pizarro, D., Dal Grande, F., Leavitt, S.D., Dyer, P.S., Schmitt, I., Crespo, A., Thorsten Lumbsch, H., and Divakar, P.K. (2019). Whole-genome sequence data uncover widespread heterothallism in the largest group of lichen-forming fungi. *Genome Biol. Evol.* **11**: 721-730.

Pizarro, D., Divakar, P.K., Grewe, F., Crespo, A., Dal Grande, F., and Lumbsch, H.T. (2020). Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with absence of usnic acid in lichen-forming fungi. *Genome Biol. Evol.* **12**: 1858-1868.

Plett, J.M., Kemppainen, M., Kale, S.D., Kohler, A., Legue, V., Brun, A., Tyler, B.M., Pardo, A.G., and Martin, F. (2011). A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr. Biol.* **21**: 1197-1203.

Pogoda, C.S., Keepers, K.G., Lendemer, J.C., Kane, N.C., and Tripp, E.A. (2018). Reductions in complexity of mitochondrial genomes in lichen-forming fungi

shed light on genome architecture of obligate symbioses. Mol. Ecol. **27**: 1155-1169.

Popper, Z.A., Michel, G., Herve, C., Domozych, D.S., Willats, W.G., Tuohy, M.G., Kloareg, B., and Stengel, D.B. (2011). Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.* **62**: 567-590.

Ranković, B. (2015). *Lichen secondary metabolites*. Cham: Springer International Publishing.

Resl, P., Bujold, A.R., Tagirdzhanova, G., Meidl, P., Rallo, S.F., Kono, M., Fernández-Brime, S., Guðmundsson, H., Andrésson, S., and Muggia, L. (2021). Large differences in carbohydrate degradation and transport potential in the genomes of lichen fungal symbionts. *bioRxiv*.

Richardson, D.H., Hill, D.J., and Smith, D.C. (1968). Lichen physiology .11. Role of alga in determining pattern of carbohydrate movement between lichen symbionts. *New Phytol.* **67**: 469.

Rose, M.R., and Doolittle, W.F. (1983). Molecular biological mechanisms of speciation. *Science* **220**: 157-162.

Saier, M.H., Jr., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., and Moreno-Hagelsieb, G. (2016). The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.* **44**: D372-379.

Sanders, W.B., and Masumoto, H. (2021). Lichen algae: the photosynthetic partners in lichen symbioses. *Lichenologist* **53**: 347-393.

Sharpton, T.J., Stajich, J.E., Rounsley, S.D., Gardner, M.J., Wortman, J.R., Jordar, V.S., Maiti, R., Kodira, C.D., Neafsey, D.E., and Zeng, Q. (2009).

Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Res.* **19**: 1722-1731.

Shelest, E. (2017). Transcription factors in fungi: TFome dynamics, three major families, and dual-specificity TFs. *Front. Genet.* **8**: 53.

Shin, J., Kim, J.E., Lee, Y.W., and Son, H. (2018). Fungal Cytochrome P450s and the P450 Complement (CYPome) of *Fusarium graminearum*. *Toxins* **10**: 3.

Smit, A., Hubley, R., and Green, P. (2013). "2013–2015. RepeatMasker Open-4.0".

Smith, D. (1968). The movement of carbohydrate from alga to fungus in lichens. *Bull. Soc. Bot. France* **115**: 129-133.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stuber, K., Ver Loren Van Themaat, E., Brown, J.K., Butcher, S.A., Gurr, S.J., Lebrun, M.H., Ridout, C.J., Schulze-Lefert, P., Talbot, N.J., Ahmadinejad, N., Ametz, C., Barton, G.R., Benjdia, M., Bidzinski, P., Bindschedler, L.V., Both, M., Brewer, M.T., Cadle-Davidson, L., Cadle-Davidson, M.M., Collemare, J., Cramer, R., Frenkel, O., Godfrey, D., Harriman, J., Hoede, C., King, B.C., Klages, S., Kleemann, J., Knoll, D., Koti, P.S., Kreplak, J., Lopez-Ruiz, F.J., Lu, X., Maekawa, T., Mahanil, S., Micali, C., Milgroom, M.G., Montana, G., Noir, S., O'connell, R.J., Oberhaensli, S., Parlange, F., Pedersen, C., Quesneville, H., Reinhardt, R., Rott, M., Sacristan, S., Schmidt, S.M., Schon, M., Skamnioti, P., Sommer, H., Stephens, A., Takahara, H., Thordal-Christensen, H., Vigouroux, M., Wessling, R., Wicker, T., and Panstruga, R. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543-1546.

Spribille, T., Tuovinen, V., Resl, P., Vanderpool, D., Wolinski, H., Aime, M.C., Schneider, K., Stabentheiner, E., Toome-Heller, M., Thor, G., Mayrhofer, H., Johannesson, H., and Mccutcheon, J.P. (2016). Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **353**: 488-492.

Staats, M., and Van Kan, J.A. (2012). Genome update of *Botrytis cinerea* strains B05.10 and T4. *Eukaryot. Cell* **11**: 1413-1414.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei Dit Frey, N., Gianinazzi-Pearson, V., Gilbert, L.B., Handa, Y., Herr, J.R., Hijri, M., Koul, R., Kawaguchi, M., Krajinski, F., Lammers, P.J., Masclaux, F.G., Murat, C., Morin, E., Ndikumana, S., Pagni, M., Petitpierre, D., Requena, N., Rosikiewicz, P., Riley, R., Saito, K., San Clemente, H., Shapiro, H., Van Tuinen, D., Becard, G., Bonfante, P., Paszkowski, U., Shachar-Hill, Y.Y., Tuskan, G.A., Young, J.P., Sanders, I.R., Henrissat, B., Rensing, S.A., Grigoriev, I.V., Corradi, N., Roux, C., and Martin, F. (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 20117-20122.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**: 46-53.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and

quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**: 511-515.

Trembley, M.L., Ringli, C., and Honegger, R. (2002). Morphological and molecular analysis of early stages in the resynthesis of the lichen Baeomyces rufus. *Mycol. Res.* **106**: 768-776.

Tsuzuki, S., Handa, Y., Takeda, N., and Kawaguchi, M. (2016). Strigolactone-induced putative secreted protein 1 is required for the establishment of symbiosis by the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. Mol. Plant Microbe Interact. **29**: 277-286.

Tuovinen, V., Ekman, S., Thor, G., Vanderpool, D., Spribille, T., and Johannesson, H. (2019). Two basidiomycete fungi in the cortex of wolf lichens. *Curr. Biol.* **29**: 476-483. e475.

Turmel, M., Pombert, J.F., Charlebois, P., Otis, C., and Lemieux, C. (2007). The green algal ancestry of land plants as revealed by the chloroplast genome. *Int. J. Plant Sci.* **168**: 679-689.

Valladares, F., Wierzchos, J., and Ascaso, C. (1993). Porosimetric study of the lichen family Umbilicariaceae: anatomical interpretation and implications for water storage capacity of the thallus. *Am. J. Bot.* **80**: 263-272.

Wang, Y., Yuan, X., Chen, L., Wang, X., and Li, C. (2018). Draft genome sequence of the lichen-forming fungus *Ramalina intermedia* strain YAF0013. *Genome Announc.* **6**: e00478-00418.

Wang, Y.Y., Liu, B., Zhang, X.Y., Zhou, Q.M., Zhang, T., Li, H., Yu, Y.F., Zhang, X.L., Hao, X.Y., Wang, M., Wang, L., and Wei, J.C. (2014). Genome characteristics reveal the impact of lichenization on lichen-forming fungus

Endocarpon pusillum Hedwig (Verrucariales, Ascomycota). *BMC Genome.* **15**: 34.

Wolf, Y.I., and Koonin, E.V. (2013). Genome reduction as the dominant mode of evolution. *Bioessays* **35**: 829-837.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., Mcdonald, S., Mclean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'neil, S., Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871-880.

Xiao, G., Ying, S.H., Zheng, P., Wang, Z.L., Zhang, S., Xie, X.Q., Shang, Y., St Leger, R.J., Zhao, G.P., Wang, C., and Feng, M.G. (2012). Genomic perspectives

on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci. Rep.* **2**: 483.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586-1591.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**: W445-451.

Yoshino, K., Yamamoto, K., Hara, K., Sonoda, M., Yamamoto, Y., and Sakamoto, K. (2019). The conservation of polyol transporter proteins and their involvement in lichenized Ascomycota. *Fungal Biol.* **123**: 318-329.

Zuccaro, A., Lahrmann, U., Guldener, U., Langen, G., Pfiffi, S., Biedenkopf, D., Wong, P., Samans, B., Grimm, C., Basiewicz, M., Murat, C., Martin, F., and Kogel, K.H. (2011). Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. *PLoS Pathog.* **7**: e1002290.

Zuo, G., and Hao, B. (2015). CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. *Genom. Proteom. Bioinform.* **13**: 321-331.

CHAPTER II

Genomic evidence of host and lifestyle transition in *Magnaporthe grisea*

ABSTRACT

Some plant pathogens have evolved mechanisms of host jumping to expand their host range. It is important to understand host specificity to prevent devastating crop yield loss and protect global food security. We identified host-transited *Magnaporthe grisea* JDJ2F and YHL-684, strains which were originally known to infect crabgrass (*Digitaria* spp.) but were isolated from rice (*Oryzae sativa*). Genotype and pathogenicity tests showed that the original host of two rice-isolated *M. grisea* strains was crabgrass, and their nonvirulence on rice suggested an endophytic lifestyle in the isolated host. The endophytic association of *M. grisea* JDJ2F was examined by artificial colonization in rice. Genome-wide comparisons revealed that the genomes of the *M. grisea* JDJ2F and YHL-684 strains were highly similar to *M. grisea* isolated from crabgrass. Effector repertoires, which are affected by host specificity, were also not significantly different from crabgrass-infecting *M. grisea* strains. However, genomic evidence of host transition was identified in *M. oryzae* effector proteins AVR-Pi9 and AVR-Pik. Several polymorphisms of AVR-Pi9 in *M. grisea* JDJ2F showed amino acid similarity to *M. oryzae*, and the unexpected presence of AVR-Pik was identified in *M. grisea* JDJ2F. Numerous transposable elements (TEs) accumulated close to the AVR-Pik locus, suggesting TE-mediated gain of this gene. These results suggested that *M. grisea* JDJ2F and YHL-684 are in the stage preceding the host jump from crabgrass to rice, and exhibit an endophytic lifestyle before obtaining virulence in rice. This genomic evidence improves our understanding of host jump events in fungal pathogens.

INTRODUCTION

Most plant pathogenic fungi are host specific and have a limited host range in which they cause disease (Li et al., 2020). In many cases, pathogenic fungi infect new plant hosts as a consequence of host jump or host range expansion, which are common evolutionary mechanisms in fungal pathogens. Host jump events are associated with the evolution of effector proteins, which are secreted proteins that modulate the host plant immune defense system and enable pathogen infection (Schulze-Lefert and Panstruga, 2011;Sonah et al., 2016). Effector proteins evolve rapidly under strong selection pressure for adaption to the new host environment, and most genes encoding these molecules are located in repeat-rich and gene-poor regions of the genome characterized by a high rate of change in accordance with the "two-speed genome" model (Dong et al., 2015). Effector variation following a host jump derived by transposable elements (TEs), chromosomal rearrangements (Seidl and Thomma, 2014;Fouché et al., 2018), and nucleotide substitutions (Yoshida et al., 2009;Huang et al., 2014). Horizontal gene transfer (HGT) and horizontal chromosome transfer (HCT) are also mechanisms underlying host jumping associated with the transfer of effector genes from pathogens to nonpathogenic fungal species (Fouché et al., 2018;Li et al., 2020).

Magnaporthe oryzae, a destructive fungal pathogen that infects a wide range of cereals and grasses, recently caused severe economic losses due to wheat blast disease in Bangladesh (Islam et al., 2019). The outbreak suggested that *M. oryzae* has the potential to evolve rapidly and infect new hosts. Therefore, understanding

host shifts has become an important aim for the protection of global food security. Several effector genes including avirulence genes were identified in *M. oryzae*, and each pathotype has a different effector repertoire to respond to the cognate host immune system (Kim et al., 2019;Chung et al., 2020). AVR1-CO39 is absent in rice-infecting *M. oryzae* isolates, whereas other pathotypes carry this effector gene (Farman et al., 2002;Tosa et al., 2005). AVR-Pik is also restricted to rice capable of infecting *M. oryzae* strains. PWL2, which was identified from rice isolates, prevents host expansion to weeping lovegrass (*Eragrostis curvula*) and is considered to be a host-determinant in *M. oryzae* (Kang et al., 1995;Sweigard et al., 1995). Although presence of these effector genes are conserved in most *M. oryzae* strains, polymorphisms depending on their associated host have been reported (Li et al., 2009; Yoshida et al., 2009). APikL2, a homologue of AVR-Pik, shows amino acid polymorphism among blast fungi with different host species, and is considered to be associated with host range expansion (Bentham et al., 2021). In addition to effector genes, gain and loss of noneffector genes are also involved in host jump adaptation (Sharma et al., 2014), such as PWT3 and PWT4, the loss of function of which induced host jump to wheat (Inoue et al., 2017). However, the mechanism underlying host jumping is not clearly understood.

Magnaporthe grisea, which belongs to the *Magnaporthe grisea* species complex that includes *M. oryzae*, mainly infects *Digitaria* spp. and is taxonomically distinct from *M. oryzae* (Couch and Kohn, 2002). The effector repertoires of *M. grisea* and *M. oryzae* are also highly dissimilar to pathotypes of *M. oryzae*, suggesting roles in infection of crabgrass. However, cross-infection exists between

Oryza sativa and *Digitaria* spp., which implies potential host jump or host range expansion between *M. oryzae* and *M. grisea* (Tosa et al., 2004;Choi et al., 2013;Chung et al., 2020). Pathogenicity of fungi isolated from rice toward crabgrass and other Poaceae species have been reported, but several studies showed that isolates from crabgrass could not infect rice or other hosts. In comparison to host jump and host specificity among pathotypes of *M. oryzae*, there have been few studies of these phenomena in *M. oryzae* and *M. grisea*.

We identified host transition in two rice-isolated *M. grisea* strains, JDJ2F (Kim et al., 2020) and YHL-684 (Park et al., 2003), by sequence comparison of housekeeping genes and pathogenicity testing on original and isolated hosts. *M. grisea* JDJ2F was obtained from surface-sterilized rice seeds, suggesting an endophytic lifestyle in rice. The endophytic lifestyle of *M. grisea* JDJ2F was examined by monitoring survival and growth in rice seedlings. Based on previous studies showing that host jump affects the pathogen genome structure, we sequenced the genomes of rice-isolated *M. grisea* JDJ2F and YHL-694 strains, and performed comparative genomic analyses with crabgrass-isolated *M. grisea* and rice-isolated *M. oryzae* strains. Synteny, phylogenetic, single-nucleotide polymorphism (SNP), and orthologue clustering analyses showed that the general genomic features of the two rice-isolated *M. grisea* strains were highly similar to crabgrass-isolated *M. grisea* strains. However, polymorphism in AVR-Pi9 and gain of AVR-Pik effector genes coincided with the distribution of large numbers of TEs, suggesting host transition from crabgrass to rice. Based on these results, we hypothesized that rice-isolated *M. grise*a strains are in an intermediate stage of host transition, and have an endophytic

lifestyle with successful colonization in the host plant before obtaining complete virulence.

MATERIALS AND METHODS

I. Fungal strains and culture conditions

Magnaporthe grisea JDJ2F strain was isolated from surface-sterilized rice seed (Kim et al., 2020) and *M. grisea* YHL-684 was isolated during a collection of rice field isolates (Park et al., 2003). These two strains were obtained from the Center for Fungal Genetic Resources (CFGR) at Seoul National University, Seoul, Korea (http://genebank.snu.ac.kr). Freeze-dried mycelium of JDJ2F and YHL-684 strains were regenerated on potato dextrose agar (PDA) medium at 25°C with continuous fluorescent light, and the JDJ2F strain was purified by single spore isolation.

II. PCR and DNA sequencing of genes for multilocus sequence typing (MLST)

M. grisea JDJ2F and YHL-684 isolates were grown in a complete medium (CM; 10 g sucrose, 6g yeast extract, 6g casamino acid) in dark conditions for 5 days to extract genomic DNA. Freeze-dried mycelium of strains was used The PCR reactions for amplifying actin, beta-tubulin, and calmodulin sequences were performed using ACT-512F, ACT-783R, Bt1a, Bt1b, CAL-228F and CAL-737R primers which were used in the previous study (Choi et al., 2013). DNA sequencing was performed at the National Instrumentation Center for Environmental Management (NICEM) at Seoul National University.

III. Pathogenicity test and sporulation on the lesion of crabgrass

5-week-old seedlings of the rice cultivar Nakddong and crabgrass *D. sanguinalis* were used for *M. oryzae* and *M. grisea* infection assays. Collected rice and crabgrass leaves were wounded with a 1ml syringe 10 times and placed the mycelium plugs from PDA cultures onto the wounded leaf surface. Inoculated crabgrass and rice leaves were incubated for 3 and 5 days respectively in the plastic humidity box in a 25°C growth chamber. These experiments were performed in triplicate three times. 5-days-old seedlings of rice were used for more efficient infection. Peeled rice seeds were surface sterilized with 4% NaClO and incubated on the MS medium (2% Murashige & Skoog powder, 1% sucrose, and 0.8% agar) for 5 days in the dark. Cut a slit on the rice seedlings and put it on the mycelial agar plug of *M. oryzae* and *M. grisea* strains. Disease symptoms were identified after 7 days in the dark and 7 days in the light incubation. To observe sporulation in artificial media, fungal strains were grown on the oatmeal agar media (50 g of oatmeal and 25g of agar per liter) at 25°C under constant fluorescent light and scraped the aerial mycelia for inducing spores.

IV. Detection of mycelia growth of *M. grisea* JDJ2F strain mycelia growth in rice

Plasmids containing EF1α::eGFP::TrpC were used to observe colonization of M. grisea JDJ2F strain in rice. EF1α promoter and TrpC terminator induced

continuous high expression of eGFP, and we could consistently detect the mycelia of M. grisea JDJ2F strain in rice. Fusion construct containing EF1α promoter (originated from *Fusarium verticillioides*) and eGFP was amplified from YL1320 and TrpC terminator was amplified from pBCATPH. EF1α promoter:eGFP (SalI-HindIII) and TrpC terminator (HindIII-EcoRI) constructs were generated using double-joint PCR and TA-cloned into pGEMT-easy. The cloned vectors were digested with the corresponding restriction enzymes and fused into pCB1004. *M. grisea* JDJ2F transformant expressing this construct was generated by PEG-mediated transformation. For detecting the mycelia growth of *M. grisea* JDJ2F in rice, we artificially inoculated mycelium on the 5 days rice seedlings using modified protocol from previously described (Becker et al., 2018).

V. Genome sequencing and de novo assembly

For genomic DNA extraction, mycelia of M. grisea JDJ2F and YHL-684 strains were cultured in a liquid complete medium (6g of yeast extract, 6g of casamino acids, and 10g of sucrose per liter) in the dark for 5 days at 25°C on a shaker. The total DNA of M. grisea JDJ2F and YHL684 strains was extracted using the modified cetyltrimethylammonium bromide (CTAB) method. DNA extraction, sequencing library preparation, whole genome sequencing, and de novo assembly were performed at the National Instrumentation Center for Environmental Management (NICEM) at Seoul National University. Briefly, adapter sequences of raw data were demultiplexed and index sequences of raw data were removed using SMRT Link (ver. 10.1.0.119588). Then, de novo assembly and contig polishing were performed

using SMRT Link (ver. 10.1.0.119588) and NextPolish v1.3.1 (Hu et al., 2020).

VI. Genome annotation

Repetitive sequences were annotated and masked by a custom repeat sequence library of 39 previously sequenced *M. oryzae* and *M. grisea* genomes using RepeatModeler v2.0.2 (Flynn et al., 2020) and RepeatMasker v4.0.5 package with RepBase 21.05 fungi library (Smit et al., 2013). Structural gene annotation was performed using two genome annotation tools, BRAKER v2 (Brůna et al., 2021) and MAKER v2 (Holt and Yandell, 2011). Peptide sequences of *M. oryzae* 70-15 and *M. grisea* NI907 were used as protein homology evidence during structural gene annotation. Functional gene annotation was implemented using InterProScan v5.20-59.0 (Jones et al., 2014). Gene density and transposable elements distribution plots were generated using python scripts from a previous study (Wyka et al., 2021).

VII. Whole-genome synteny analysis

Genome sequences of JDJ2F and YHL-684 were compared with four additional genomes, two *M. oryzae* (70-15 and KJ201) and *M. grisea* strains (DS0505 and NI907) using the dna-diff module in MUMMER v3 (Kurtz et al., 2004) program with the default option. Principal component analysis (PCA) was performed using SNPRelate R package with SNPs from genome comparison results. Genome synteny between JDJ2F and YHL-684 was visualized by circos v0.69.9 (Krzywinski et al., 2009).

VIII. Gene orthology and phylogenetic analysis

To obtain the ortholog sets of *M. grisea* and *M. oryzae* strains, homology-based protein clustering was performed using OrthoFinder v2.3.3 (Emms and Kelly, 2019). Subsequently, ortholog sets were visualized using TBTools v2 (Chen et al., 2020). In addition, to infer the phylogenetic relationship of JDJ2F and YHL-684, a total of 22 genomes of M. oryzae and M. grisea strains were used for ortholog clustering with two outgroup sets, *Magnaporthiopsis poae* and *Gaeumannomycetes graminis*. Single-copy orthologs from protein clustering results were used to construct the phylogenetic tree using IQ-TREE v2.1.3 (Minh et al., 2020) with 1,000 bootstrap replicates. Expression patterns of species-specific gene families were analyzed using previously published transcriptome datasets of *M. oryzae* 70-15 (Jeon et al., 2020).

IX. Prediction and distribution of effector candidate genes

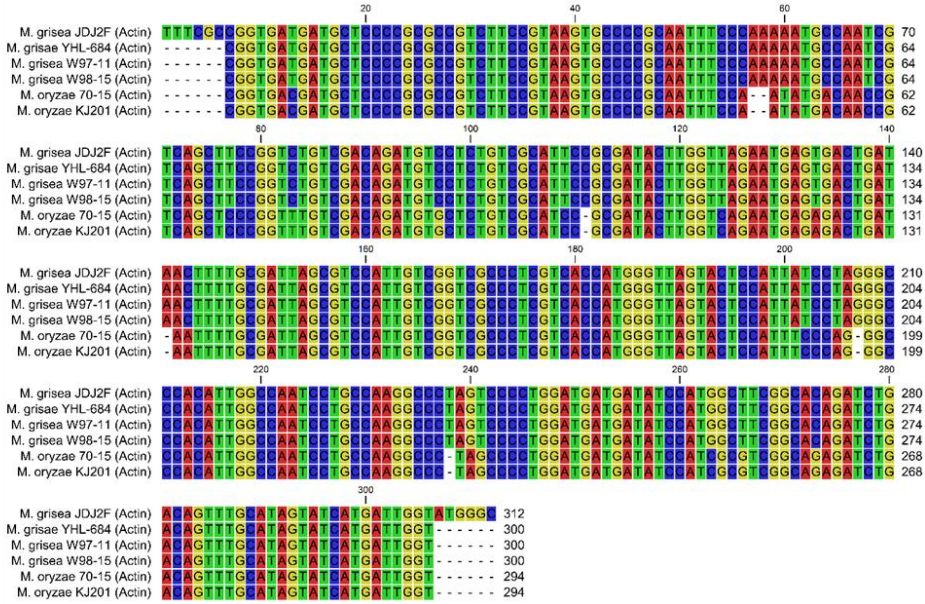
We detected secretion signal and transmembrane domain in proteins by SignalP v5.0 (Almagro Armenteros et al., 2019) and TMHMM v2.0 (Krogh et al., 2001) respectively. Proteins with a signal peptide and no transmembrane domain were analyzed as candidate effector genes by EffectorP v3.0 (Sperschneider and Dodds, 2022). Previously reported *M. oryzae* effector genes were aligned to *M. oryzae* and *M. grisea* peptide sequences using BLASTP. Homolog protein sequences were aligned using clustal-omega 1.2.4 (Sievers et al., 2011).

RESULTS

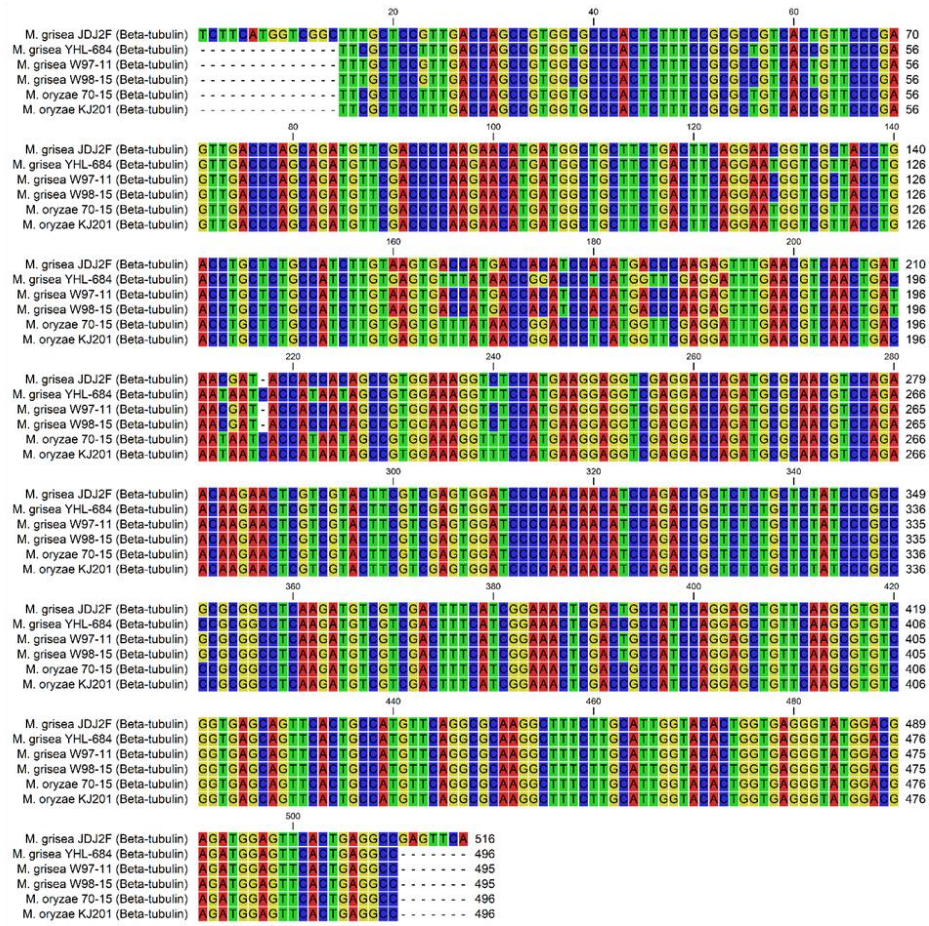
I. Species delineation and haplotype of two rice-isolated *Magnaporthe* strains

Based on the multilocus sequence typing (MLST) method used previously to determine the haplotype of the *M. grisea* species complex (Choi et al., 2013), we amplified and sequenced the actin, beta-tubulin, and calmodulin gene sequences of JDJ2F and YHL-684 strains isolated from rice. Sequence alignment with previously sequenced *M. grisea* strains (W97-11 and W98-15) and *M. oryzae* strains (70-15 and KJ201) (Choi et al., 2013) showed that three housekeeping genes of the JDJ2F and YHL-684 strains were identical to *M. grisea* strains, with the exception of the beta-tubulin gene of YHL-684 (Figure 1). The sequence of the calmodulin gene at ~30 nt in the two strains indicated the crabgrass-infecting *M. grisea* haplotype, suggesting that they originated from crabgrass (Figure 2A). Based on these results, we defined the JDJ2F and YHL-684 strains as *M. grisea* species. Interestingly, the beta-tubulin sequence of *M. grisea* YHL-684 was the same as *M. oryzae*, although beta-tubulin is widely used as a molecular marker for species delineation (Raja et al., 2017).

A


M. grisea JDJ2F (Actin)
M. grisae YHL-684 (Actin)
M. grisea W97-11 (Actin)
M. grisea W98-15 (Actin)
M. oryzae 70-15 (Actin)
M. oryzae KJ201 (Actin)

B


M. grisea JDJ2F (Beta-tubulin)
M. grisea YHL-684 (Beta-tubulin)
M. grisea W97-11 (Beta-tubulin)
M. grisea W98-15 (Beta-tubulin)
M. oryzae 70-15 (Beta-tubulin)
M. oryzae KJ201 (Beta-tubulin)

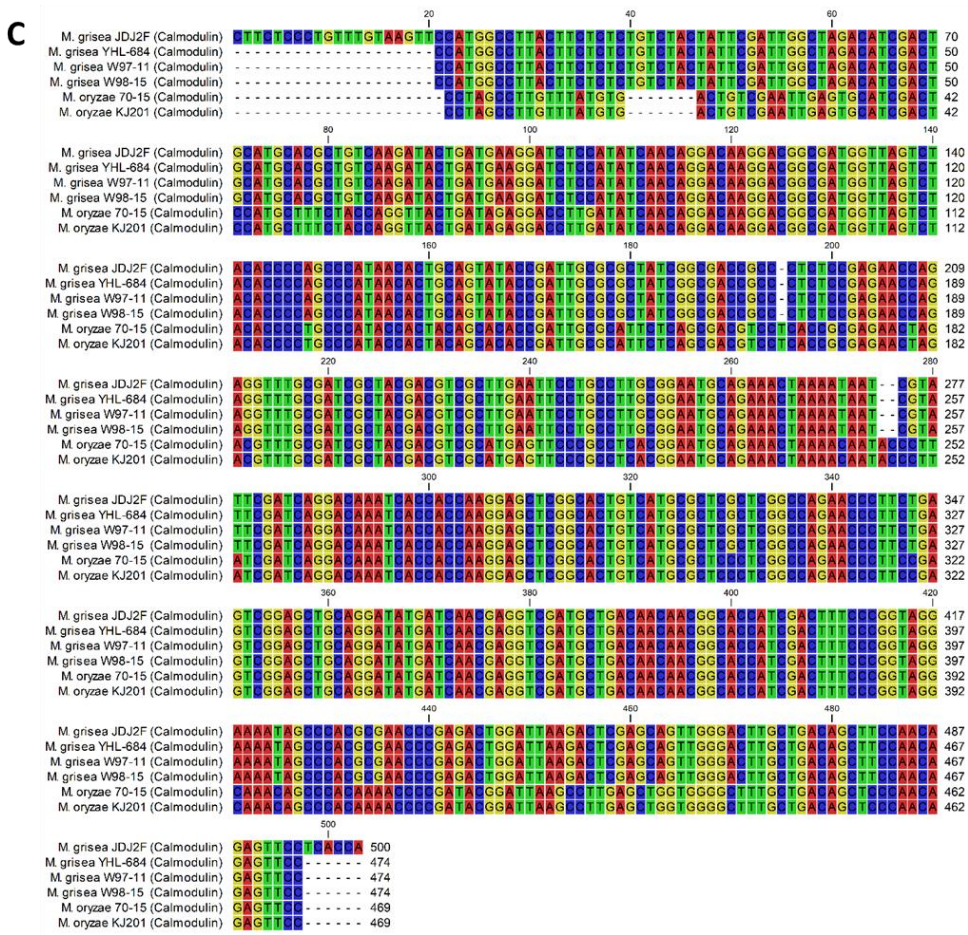


Figure 1. Haplotype identification of two rice-isolated *Magnaporthe* strains based on the multilocus sequence typing (MLST) method

The aligned nucleotide sequences of (A) actin, (B) beta-tubulin, and (C) calmodulin genes in *M. grisea* and *M. oryzae* strains for identification of haplotype and species delineation. The asterisk indicates rice-isolated *Magnaporthe* strains identified in this study.

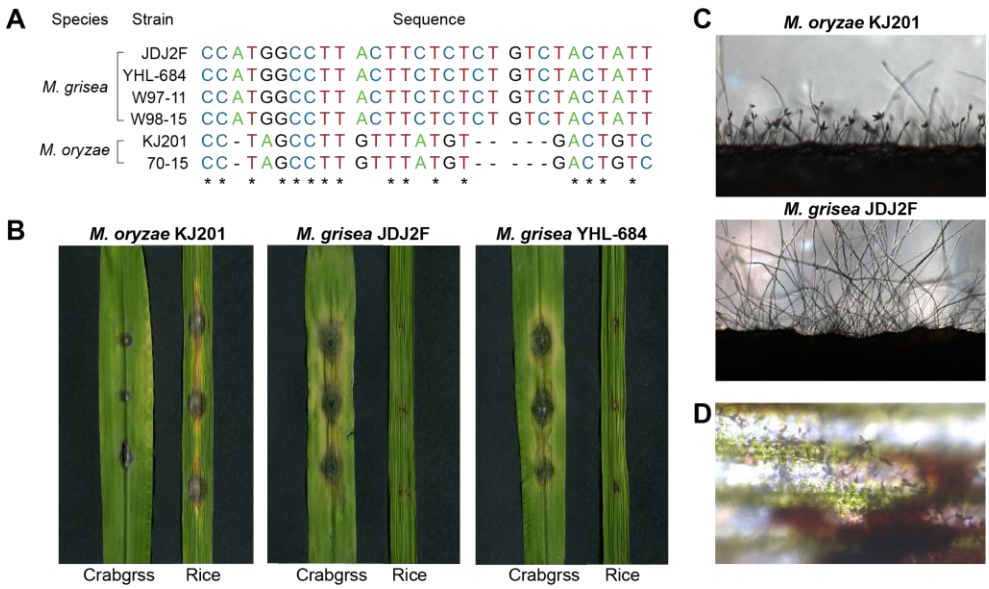


Figure 2. Characterization of two *Magnaporthe* strains, *M. grisea* JDJ2F and YHL-684.

(A) Aligned nucleotide sequences of the calmodulin genes in *M. grisea* and *M. oryzae* strains. Asterisks indicate conserved nucleotides among six *Magnaporthe* strains. (B) Pathogenicity tests of crabgrass (*Digitaria sanguinalis*) and rice (*Oryzae sativa*) by wound inoculation. Mycelial agar blocks were placed on the wounded site. Disease symptoms were assessed 3 days postinoculation in crabgrass and 5 days postinoculation in rice. (C) Conidiogenesis of *M. oryzae* KJ201 and *M. grisea* JDJ2F was observed after 24-h incubation on the surface of scraped oatmeal agar medium. (D) Conidia production on the lesions of crabgrass inoculated with *M. grisea* JDJ2F.

II. Original host of *M. grisea* JDJ2F and YHL-684 strains

Analyses of pathogenicity in crabgrass and rice were performed to confirm their host origins and nonpathogenic lifestyle in the isolated host. As the *M. grisea* JDJ2F and YHL-684 strains did not produce any spores on artificial medium (Figure 2C), we used the wound inoculation method, which employs a mycelial agar block as an inoculum. The crabgrass leaves infected by *M. grisea* strains exhibited disease lesions, whereas the *M. oryzae* KJ201 strain did not induce any symptoms in the crabgrass (Figure 2B). Moreover, conidiophores of the *M. grisea* JDJ2F strain were induced at the site of infection on crabgrass leaves, indicating that *M. grisea* JDJ2F can complete its life cycle in crabgrass (Figure 2D). Rice leaves and seedlings were also inoculated for virulence assessment (Figure 2B and Figure 3). Despite being isolated from rice, *M. grisea* JDJ2F and YHL-684 strains had no virulence on rice, indicating the potential to exist as endophytic fungi. To examine the endophytic lifestyle of *M. grisea* JDJ2F in rice, we examined mycelia growth on rice seedlings by a PCR-based method (Figure 4). Two weeks after artificially inoculating rice seedlings with *M. grisea* JDJ2F mycelia carrying GFP, we collected stems containing the inoculation site, stems distant from the inoculated site, and leaves. PCR amplification of the GFP construct in the *M. grisea* JDJ2F strain showed colonization of mycelia in all parts of the seedlings (Figure 4B). The GFP of *M. grisea* JDJ2F was detected not only at the site of inoculation, but also in distant parts of the stem, indicating growth of mycelia in rice seedlings. Moreover, we examined the linear hyphal growth of *M. grisea* JDJ2F in rice (Figure 5), which suggested the ability to colonize as an endophytic fungus (Abdellatif et al., 2009). Primers used in these

analyses are listed in Table 1.

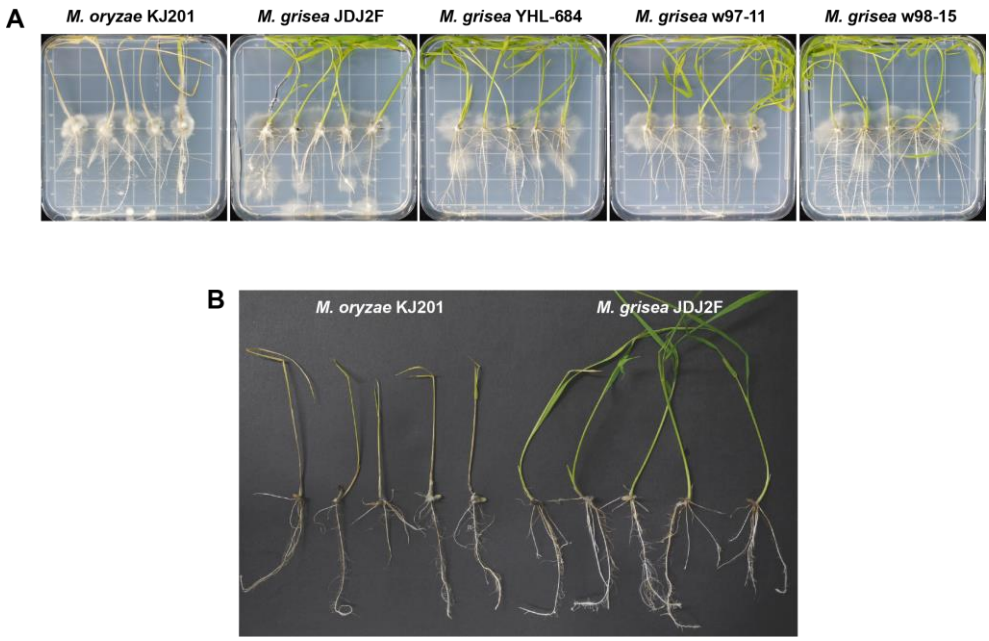


Figure 3. Pathogenicity test in rice seedlings

(A) Wound inoculation in rice seedlings. Mycelial blocks of *M. grisea* JDJ2F were placed on the slightly wounded rice stem and incubated for 2 weeks. (B) Virulence phenotypes of infected rice seedlings.

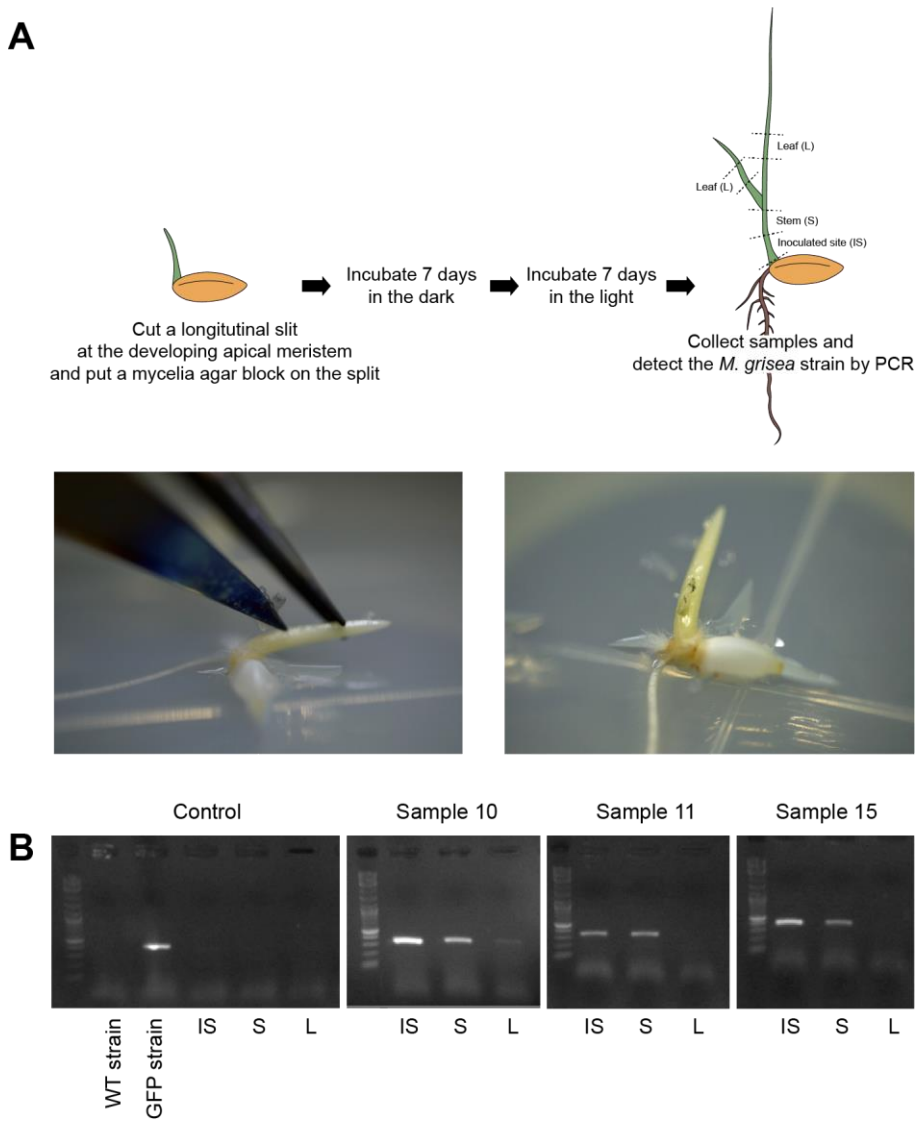


Figure 4. Detecting growth of *M. grisea* JDJ2F strain in rice seedlings

(A) Scheme for artificial infection of *M. grisea* JDJ2F into rice seedlings. (B) PCR amplification of GFP construct exists in GFP-labeled *M. grisea* JDJ2F strain collected from each part of rice seedlings. WT: Wild type; GFP: GFP-labeled strain; IS: Collected from inoculated site; S: Collected from stem; L: Collected from leaves.

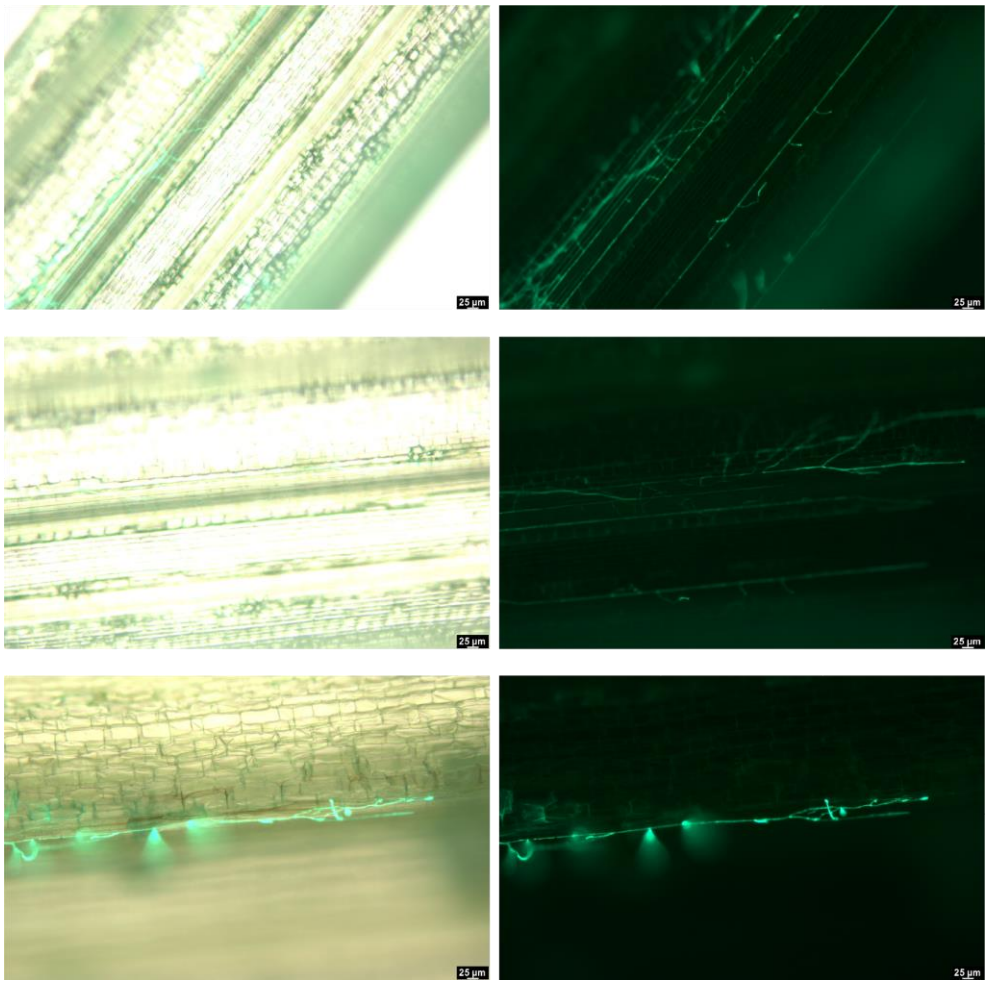


Figure 5. Endophytic colonization of *M. grisea* JDJ2F in rice

Hyphal growth of GFP-labeled *M. grisea* JDJ2F in rice sheath. Hyphae parallel with the longitudinal to rice stem observed by fluorescence microscopy. Mycelia was inoculated on the rice sheath and examined 2 days after inoculation. Bar = 100µm.

Table 1. Primers used in this study

EF1α+GFP+TrpC cloning	
EF1α_eGFP_F	GTCGACCGGTACCTATAGGGCGAATTG
EF1α_eGFP_R	AAGCTTTTACTTGTACAGCTCGTCCATG
trpC_terminator_F	AAGCTTGATCCACTTAACGTTACTGAAATCATCA
trpC_terminator_R	GAATTCAAGAAGGATTACCTCTAAACAAGTGT
GFP construct detection	
EF1α_F	CGGTACCTATAGGGCGAATTG
Inter_GFP_R	GCCGGTGGTGCAGATGAACT
Multilocus sequence typing	
actin_F	ATGTGCAAGGCCGGTTTCGC
actin_R	TACGAGTCCTTCTGGCCCAT
beta-tubulin_F	TTCCCCCGTCTCCACTTCTTCATG
beta-tubulin_R	GACGAGATCGTTCATGTTGAACTC
calmoduline_F	GAGTTCAAGGAGGCCTTCTCCC
calmoduline_R	CATCTTTCTGGCCATCATGG

III. Genome similarity among *M. grisea* strains

To obtain evidence of host transition and lifestyle switching, we sequenced the whole genomes of *M. grisea* JDJ2F and YHL-684 isolated from rice. We also included previously sequenced *M. grisea* NI907 (Gómez Luciano et al., 2019), *M. grisea* DS0505 (Zhong et al., 2016), *M. oryzae* 70-15 (Dean et al., 2005), and *M. oryzae* KJ201 genomes for comparative analysis, to optimize the bioinformatics parameters. The total assembly size (44.9–45.3 Mbp), GC content (47.8%–48.0%), and number of predicted genes (12,149–12,185) in *M. grisea* JDJ2F and YHL-684 strains were similar to previously sequenced *M. grisea* genomes (Table 2). However, these two strains had higher contents of repeat sequences (16.4%–16.5%) than other genomes (10.3%–15.9%) because of the large proportions of long terminal repeats (LTRs) (7.1%–7.0%) and unclassified interspersed repeats (5.4%–5.7%) (Table 3). The overall structure of chromosomes was largely conserved between the two *M grisea* strains (JDJ2F and YHL-684), except for several inverted regions (Figure 6). Genomes of these two stains showed greater similarity to *M. grisea* than *M. oryzae* strains (Figure 7A), and the phylogenetic tree also showed that the JDJ2F and YHL-684 strains were closely related with *M. grisea* strains despite being isolated from rice (Figure 7B). The presence of large numbers of SNPs confirms low genomic similarity between the *M. grisea* JDJ2F and *M. oryzae* strains relative to *M. grisea* strains (Figure 7C). However, in comparisons between *M. grisea* strains, there were higher rates of SNPs than in comparisons between *M. oryzae* strains, suggesting greater nucleotide variation among *M. grisea* strains. The SNPs with *M. grisea* JDJ2F as the reference genome were used for principal component analysis (PCA)

of SNP correlations (Figure 7C). *M. grisea* strains, including *M. grisea* YHL-684, were clustered together but distantly related to *M. oryzae* strains.

Table 2. Genome statistics of *M. grisea* and *M. oryzae* strains used in this study

Species	Strain	Isolated host	Genome size (bp)	Number of contigs	N50 (bp)	GC contents (%)	Repetitive sequences (%)	Number of genes
Magnaporthe grisea	JDJ2F	*Oryza sativa*	45,279,436	15	6,063,667	47.76	16.46	12,149
	YHL684	*Oryza sativa*	44,918,660	14	5,540,751	47.99	16.36	12,185
	NI907	*Digitaria sanguinalis*	44,557,582	43	5,912,490	47.80	15.87	12,452
	DS0505	*Digitaria sanguinalis*	42,710,978	1,980	73,656	48.58	12.93	12,078
Magnaporthe oryzae	70-15	*Oryza sativa*	40,979,121	53	6,606,598	51.57	12.55	12,989
	KJ201	*Oryza sativa*	45,096,509	123	2,318,557	47.09	10.29	12,780

Table 3. Proportion of repetitive sequences in genomes

Class		*M. grisea* JDJ2F	*M. grisea* YHL684	*M. grisea* NI907	*M. grisea* DS0505	*M. oryzae* 70-15	*M. oryzae* KJ201
Interspersed repeats	Short interspersed nuclear element (SINE)	0%	0%	0%	0%	0%	0%
	Long interspersed nuclear element (LINE)	0.81%	0.84%	0.72%	0.35%	2.01%	1.43%
	Long terminal repeat (LTR)	7.06%	7.13%	7.41%	5.46%	6.53%	4.84%
	DNA transposons	1.32%	1.32%	1.23%	0.88%	2.14%	1.56%
	Unclassified	5.68%	5.41%	5.06%	4.82%	1.73%	3.56%
Small RNA		0.12%	0.16%	0.02%	0.04%	0.02%	0.02%
Satellites		0.01%	0.01%	0.01%	0.01%	0.01%	0.01%
Simple repeats		1.26%	1.28%	1.24%	1.16%	1.06%	1%
Low complexity		0.21%	0.20%	0.20%	0.21%	0.16%	0.15%

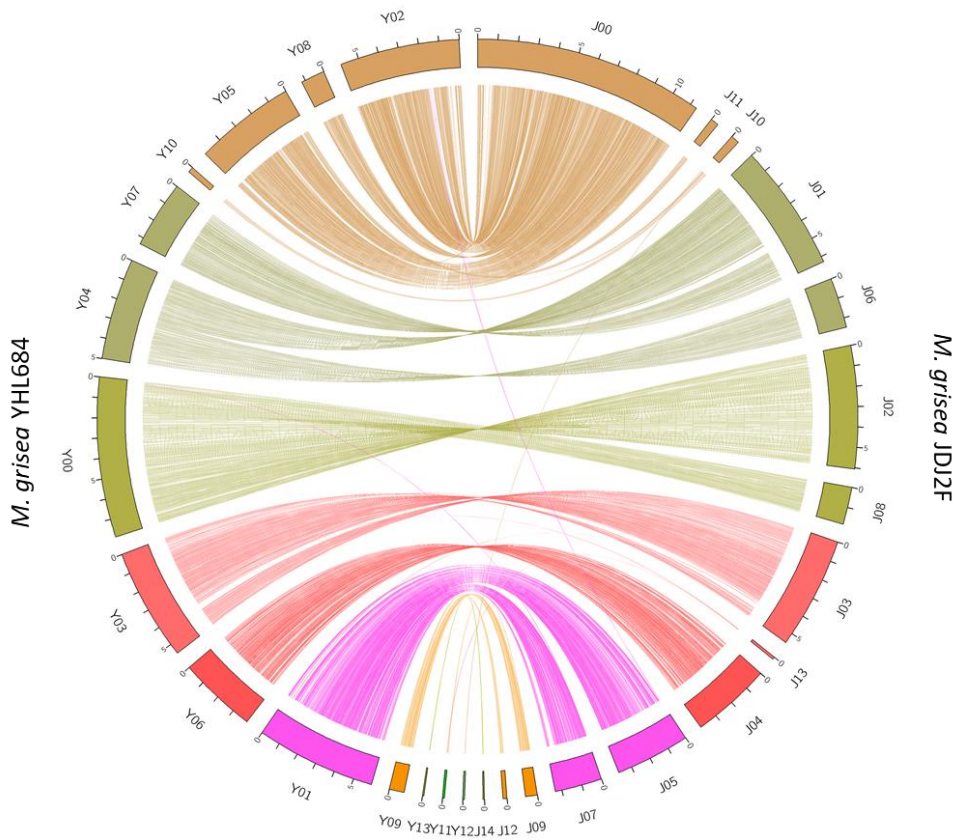


Figure 6. Genome similarity between two rice-isolated *M. grisea* strains

Circos plot demonstrating the sequence similarity between the *M. grisea* YHL-684 (left) and *M. grisea* JDJ2F (right). Links represent >98% sequence similarity with >500bp of the minimum length.

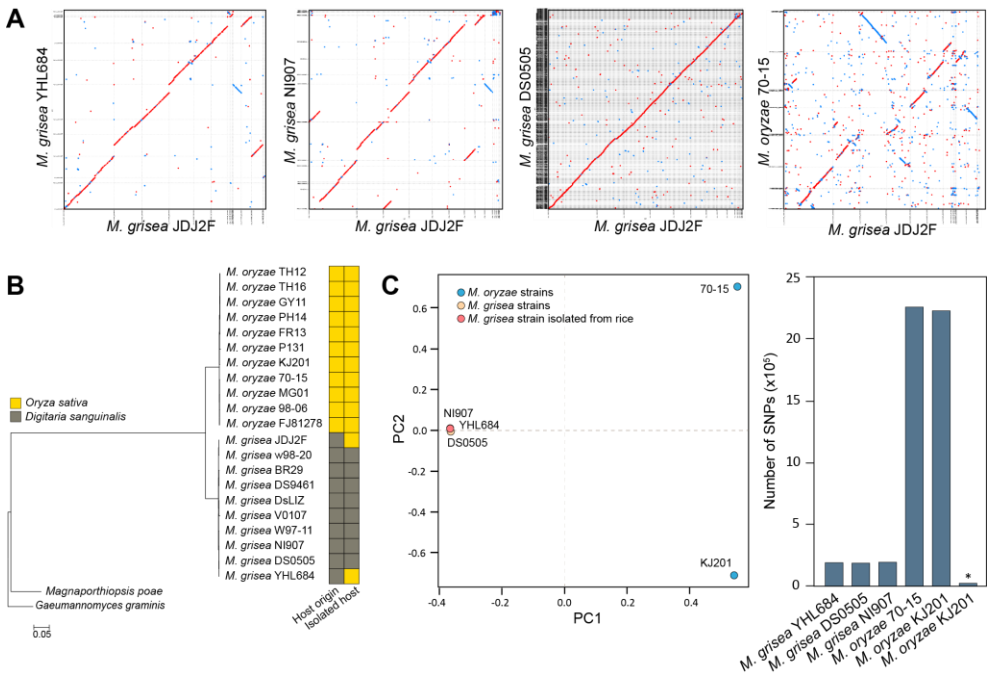


Figure 7. Whole-genome comparison among *M. grisea* and *M. oryzae* strains

(A) Analysis of synteny between *M. grisea* JDJ2F and other strains. Red lines and dots indicate matched parts with the same orientation and blue lines and dots indicate inverted matches between two strains. (B) Maximum-likelihood phylogenetic tree of *M. oryzae*, *M. grisea*, and two outgroup species, *Magnaporthiopsis poae* and *Gaeumannomycetes graminis*. Single-copy orthologues were used to construct the phylogenetic tree. Yellow and brown boxes indicate the host origin and isolated host of each strain, respectively. (C) PCA of SNPs (right) and total number of SNPs (left) between *M. grisea* JDJ2F and other strains. Asterisks in the bar graph indicate SNPs between two *M. oryzae* strains, and blue, yellow, and pink circles indicate *M. oryzae*, *M. grisea*, and rice-isolated *M. grisea* strains, respectively.

IV. Gain of *M. oryzae*-specific genes involved in the biotrophic lifestyle

Although *M. grisea* JDJ2F and YHL-684 strains showed high genomic similarity with *M. grisea* strains, we examined genomic features similar to rice-infecting *M. oryzae* strains. In a total of 13,620 orthologous gene clusters among six *M. grisea* and *M. oryzae* strains, we identified 11 gene clusters shared by *M. oryzae* 70-15, *M. oryzae* KJ201, *M. grisea* JDJ2F, and *M. grisea* YHL-684 (Figure 8A). The expression of *M. oryzae* 70-15 genes included in these clusters showed that most of these genes were induced at 18–36 hours postinfection (hpi), corresponding to the biotrophic infection stage (Figure 8B). These observations suggested that *M. grisea* JDJ2F and YHL-684 strains gained the *M. oryzae*-specific genes necessary for infection and life in rice hosts. Genes shared with *M. oryzae* were involved in DNA replication (Myb domain, transcription initiation factor TFIID, endonuclease, CMS–1) and plant cell wall degradation (glycoside hydrolase, family 43) (Table 4). *M. grisea* JDJ2F and YHL-684 had lost three genes shared among all other *M. grisea* and *M. oryzae* strains, and gene expression analysis of *M. oryzae* 70-15 orthologues showed that they were associated with the necrotrophic stage (36–45 hpi) (Figure 8A–B). *M. grisea* JDJ2F and YHL-684 had also gained 26 specific genes and lost 23 *M. grisea*-specific genes, but most of these genes were not functionally annotated (Figure 8A and Table 4).

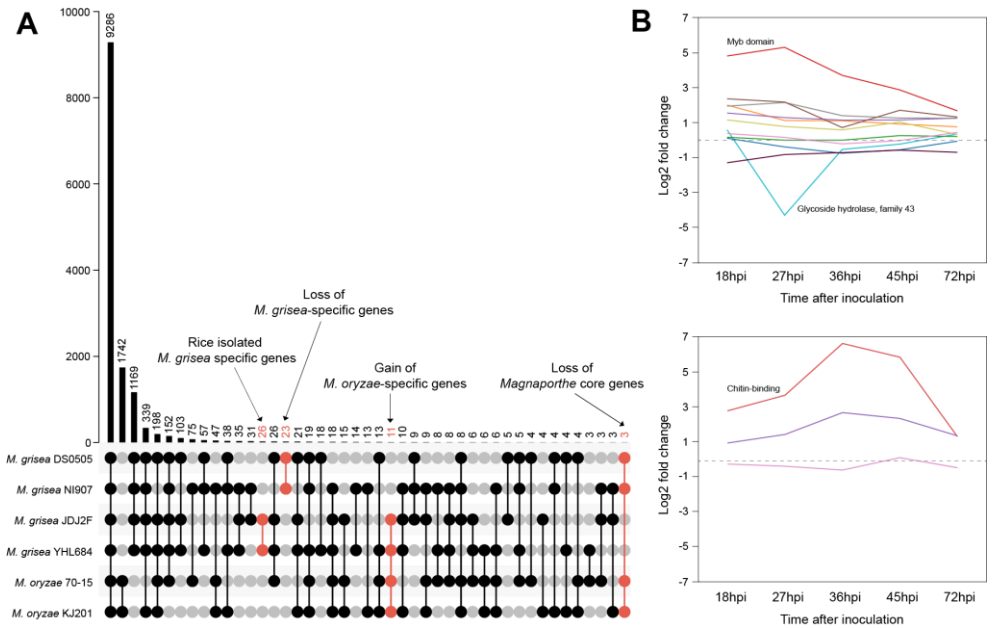


Figure 8. Identification of *M. oryzae*-specific genes in *M. grisea* JDJ2F and YHL-684 genomes

(A) UpSet plot of orthologues among *M. grisea* and *M. oryzae* strains. The bar graph at the top shows the number of shared orthogroups, depicted by dots and lines at the bottom of the figure. (B) Expression patterns of *M. oryzae* 70-15 genes in the selected orthogroups. Genes shared with *M. grisea* JDJ2F, YHL-684, and *M. oryzae* strains are shown at the top of the figure, and genes shared between strains except *M. grisea* JDJ2F and YHL-684 are shown at the bottom.

Table 4. Gene contents and functions of selected orthogroups

Orthogroup	Interpro function	Number of genes in orthogroup					
		M. grisea JDJ2F	*M. grisea* YHL684	*M. grisea* DS0505	*M. grisea* NI907	*M. oryzae* 70-15	*M. oryzae* KJ201
Gain of *M. oryzae*-specific genes							
OG0000082	Unknown	1	3	0	0	1	8
OG0011376	Unknown	1	1	0	0	1	1
OG0011366	Glycoside hydrolase, family 43	1	1	0	0	1	1
OG0011385	Unknown	1	1	0	0	1	1
OG0010153	Heterokaryon incompatibility	1	2	0	0	1	1
OG0011386	Protein Cms1	1	1	0	0	1	1
OG0011364	Endonuclease/exonuclease/phosphatase	1	1	0	0	1	1
OG0011383	Unknown	1	1	0	0	1	1
OG0011362	Myb domain	1	1	0	0	1	1
OG0011381	Transcription initiation factor TFIID, 23-30kDa subunit	1	1	0	0	1	1
OG0011382	Unknown	1	1	0	0	1	1
Loss of *Magnaporthe* conserved genes							
OG0010791	Chitin-binding, type 1	0	0	1	1	1	1
OG0010814	Unknown	0	0	1	1	1	1
OG0011360	Unknown	0	0	1	1	1	1
Loss of *M. grisea*-specific genes							
OG0011726	Unknown	0	0	1	1	0	0
OG0011727	Unknown	0	0	1	1	0	0
OG0011731	Unknown	0	0	1	1	0	0
OG0011734	Unknown	0	0	1	1	0	0

OG0011736	Unknown	0	0	1	1	0	0
OG0011737	Unknown	0	0	1	1	0	0
OG0011739	Major facilitator superfamily	0	0	1	1	0	0
OG0011743	NAD-dependent epimerase/dehydratase	0	0	1	1	0	0
OG0011745	Unknown	0	0	1	1	0	0
OG0011746	Protein kinase domain	0	0	1	1	0	0
OG0011747	Unknown	0	0	1	1	0	0
OG0011749	Unknown	0	0	1	1	0	0
OG0011751	Unknown	0	0	1	1	0	0
OG0011752	Unknown	0	0	1	1	0	0
OG0011756	Unknown	0	0	1	1	0	0
OG0011757	Unknown	0	0	1	1	0	0
OG0011758	AMP-dependent synthetase/ligase	0	0	1	1	0	0
OG0011759	Unknown	0	0	1	1	0	0
OG0011761	Unknown	0	0	1	1	0	0
OG0011765	Unknown	0	0	1	1	0	0
OG0011767	Cytochrome P450	0	0	1	1	0	0
OG0011768	Unknown	0	0	1	1	0	0
OG0011769	Unknown	0	0	1	1	0	0
M. grisea JDJ2F and YHL684 specific genes							
OG0011778	Unknown	1	1	0	0	0	0
OG0011782	BTB/POZ domain	1	1	0	0	0	0
OG0011783	Unknown	1	1	0	0	0	0
OG0011784	Unknown	1	1	0	0	0	0
OG0011785	tRNA-splicing endonuclease subunit Sen15	1	1	0	0	0	0
OG0011788	Unknown	1	1	0	0	0	0
OG0011790	Unknown	1	1	0	0	0	0

OG0011791	Unknown	1	1	0	0	0	0
OG0011794	Unknown	1	1	0	0	0	0
OG0011795	Unknown	1	1	0	0	0	0
OG0011796	Unknown	1	1	0	0	0	0
OG0011797	Unknown	1	1	0	0	0	0
OG0011799	Unknown	1	1	0	0	0	0
OG0011800	Unknown	1	1	0	0	0	0
OG0011805	Unknown	1	1	0	0	0	0
OG0011810	BTB/POZ domain	1	1	0	0	0	0
OG0011811	Unknown	1	1	0	0	0	0
OG0011812	Unknown	1	1	0	0	0	0
OG0011813	Unknown	1	1	0	0	0	0
OG0011819	Unknown	1	1	0	0	0	0
OG0011821	Major facilitator, sugar transporter-like	1	1	0	0	0	0
OG0011826	Unknown	1	1	0	0	0	0
OG0011827	Major facilitator, sugar transporter-like	1	1	0	0	0	0
OG0011828	Unknown	1	1	0	0	0	0
OG0011830	Condensation domain	1	1	0	0	0	0
OG0011833	Unknown	1	1	0	0	0	0

V. Effector repertoires and distribution in rapidly evolving genomic compartments

We predicted the effector candidate genes in *M. grisea* and *M. oryzae* genomes, but the number of effector genes was not significantly different among *M. grisea* strains including JDJ2F and YHL-684 (Table 5). In addition, sequence similarity and the presence of effector candidate genes in *Magnaporthe* strains compared to *M. grisea* JDJ2F showed that the effector repertoires of rice-isolated *M. grisea* JDJ2F and YHL-684 were highly similar to *M. grisea* strains compared to *M. oryzae* strains (Figure 9). We hypothesized that effector genes in *M. grisea* JDJ2F and YHL-684 strains may be located in regions with frequent genomic changes based on the "two-speed genome" model (Dong et al., 2015). We calculated the distance between each effector and noneffector gene to the closest repetitive elements to identify their positional relations with TEs. The effector genes of six *Magnaporthe* strains were located closer to repeat sequences compared to noneffector genes (Figure 10B). However, there were no significant differences among the six strains examined, although *M. grisea* JDJ2F and YHL-684 had higher repeat coverage than the other strains (Figure 10A). Moreover, we analyzed the lengths of 5'- and 3'- flanking intergenic regions and displayed them on two-dimensional density plots (Figure 10C). Effector genes had larger intergenic regions than noneffector genes in the six fungal genomes, indicating that the effector genes were located in gene-sparse regions compared to noneffector genes. In *M. grisea* JDJ2F and YHL-684, several noneffector genes were also distributed in gene-sparse regions compared to other

Magnaporthe strains, suggesting that the noneffector genes of two rice-isolated *M. grisea* strains were also located in regions that undergo frequent changes.

Table 5. Prediction of effector candidate genes

Species	Strain	Secreted protein	Effector candidate	Total protein
M. grisea	Mg_DS0505	1,638	485	12,078
	Mg_NI907	1,657	475	12,452
	Mg_JDJ2F	1,672	491	12,149
	Mg_YHL684	1,674	499	12,185
M. oryzae	Mo_70-15	1,739	523	12,989
	Mo_KJ201	1,746	525	12,780

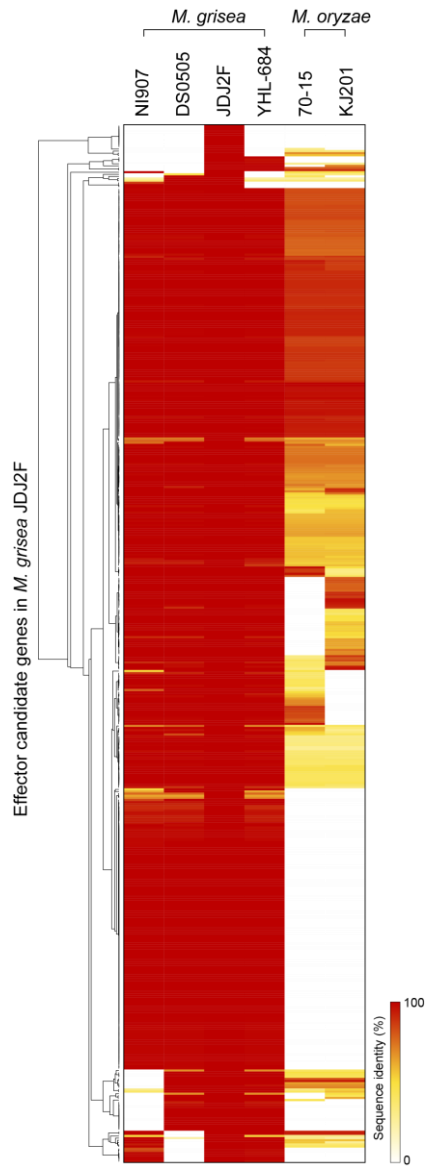


Figure 9. Sequence similarity of effector candidate genes from *M. grisea* JDJ2F in *M. grisea* and *M. oryzae* strains.

Hierarchical clustered *M. grisea* JDJ2F effector candidate genes and sequence similarity to *M. grisea* and *M. oryzae* strains were shown by heatmap. The color scale is located at the bottom of the right, and red and white color means high and low percentage sequence identity respectively.

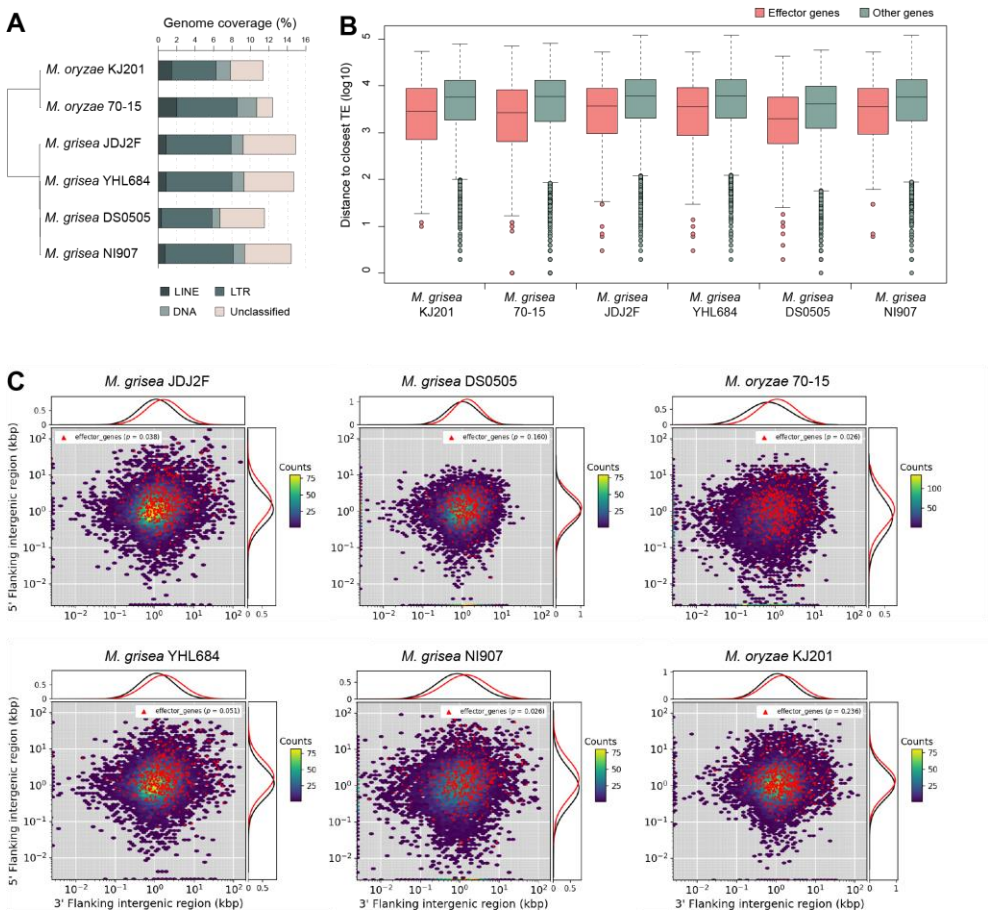


Figure 10. Genomic locations of effector and noneffector genes according to the distribution of transposable elements (TEs)

(A) Comparison of repeat contents in *Magnaporthe* strains. The genome proportions of each type of repetitive sequence are shown by cumulative bar graphs. (B) Distances of effector and noneffector genes to the closest TE fragment. Red and green boxes and dots indicate effector and noneffector genes, respectively. (C) Density plots of the 5′- and 3′-flanking intergenic regions of each genome. Intergenic region length and frequency distribution of effector genes are represented by red dots and lines (line graph located at the top and right of the density plot), respectively.

VI. Evidence of host shift from crabgrass to rice in *M. oryzae* effector genes

We analyzed the presence/absence of *M. oryzae*-specific AVR effector genes to determine the evolution of *M. grisea* JDJ2F and YHL-684 in terms of the use rice as their host. Most of the AVR effector gene repertoires of rice-isolated *M. grisea* strains were similar to crabgrass-isolated *M. grisea* strains (Figure 11A). All *Magnaporthe* strains carried AVR-Pi54, AVR-Pi9, AVR-Pii, AVR-Pita3, AVR-Piz-t, and PWL2, but there were differences in sequence identity between *M. grisea* and *M. oryzae* strains. AVR-Co39, AVR-Pita2, and PWL4 were only present in *M. grisea* strains, while AVR-Pib, PWL3, MoHTR1, and MoHTR2 were only present in *M. oryzae* strains. However, only AVR-Pi9 in *M. grisea* JDJ2F exhibited higher sequence similarity (60%) than the other *M. grisea* strains (51%), and we identified several polymorphisms at the amino acid sequence level (Figure 11A, B). Interestingly, the sequence of the AVR-Pi9 gene in *M. grisea* JDJ2F encoded *M. grisea* JDJ2F-specific amino acids or amino acids identical to *M. grisea* or *M. oryzae* (Figure 11B). The genomic context around AVR-Pi9 homologues were conserved with other *M. grisea* strains, with only *M. grisea* JDJ2F showing a difference in intergenic region length and repeat distribution (Figure 11C). These results suggest that polymorphisms in AVR-Pi9 may be affected by repeat sequences, such as repeat-induced point mutations (RIPs). The AVR-Pik gene, which is known as the host specificity gene in rice-infecting *M. oryzae*, was only present in *M. grisea* JDJ2F with high sequence similarity (73%); it was absent in other *M. grisea* strains (Figure 11A). The AVR-Pik gene of *M. grisea* JDJ2F had long upstream (19,427 bp) and downstream (12,966 bp)

intergenic lengths, and we examined the locations of repetitive sequences in these intergenic regions (Figure 12). Both intergenic regions of AVR-Pik consisted of numerous repeat units, including LTRs, long interspersed nuclear elements (LINEs), DNA transposons, and unclassified interspersed repeats. Large numbers of TE sequences distributed closed to the gene sequence may have been involved in the acquisition and sequence variation of the AVR-Pik gene. Moreover, the genomic context of AVR-Pik included *M. grisea* JDJ2F-specific genes, genes duplicated only in *M. grisea* JDJ2F, and *M. oryzae*-specific genes shared with *M. grisea* JDJ2F and YHL-684. All genes ware short (encoding products of 29–569 aa) with the exception of one gene that was duplicated only in *M. grisea* JDJ2F (encoding a product of 1,058 aa). These neighboring genes were also accompanied by numerous TE sequences (Table 6). Taken together, these observations suggest that this genomic region close to AVR-Pik was newly obtained via TE elements, especially LTRs.

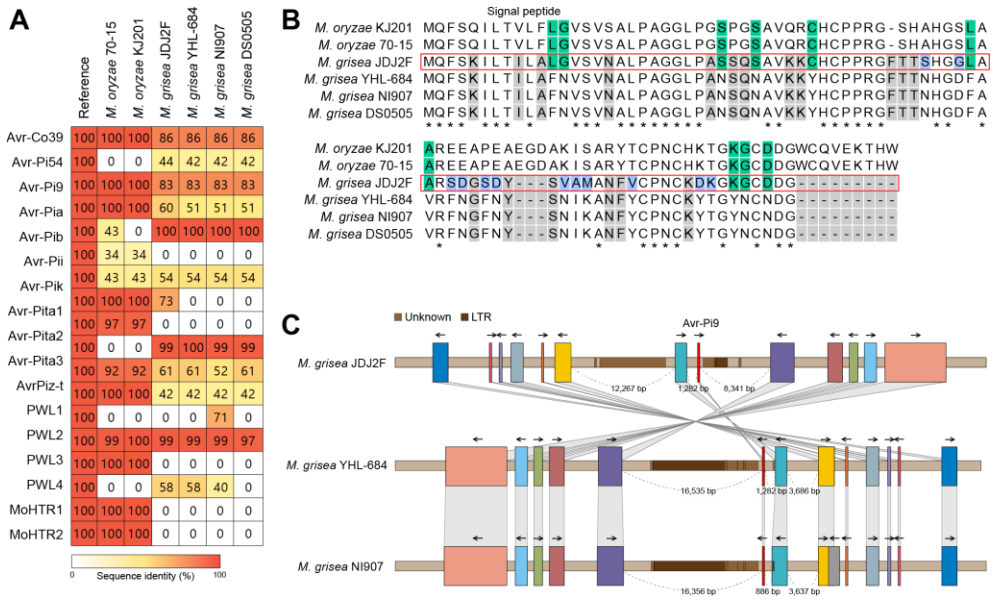


Figure 11. Distribution of *M. oryzae* effector genes in *Magnaporthe* strains and polymorphism of the AVR-Pi9 sequence in *M. grisea* JDJ2F

(A) Presence and absence of effector genes known as host-determinant factors. Sequence identity in each *M. grisea* and *M. oryzae* genome is shown as a color scale bar (bottom of heat map). (B) Sequence alignment of the AVR-Pi9 genes. Red boxes indicate the sequence of *M. grisea* JDJ2F. Gray, green, and blue shading indicate amino acid sequences specific to *M. grisea* JDJ2F, identical to *M. oryzae*, and identical to *M. grisea*, respectively. Asterisks indicate conserved sites in the six strains. (C) Genomic contexts of AVR-Pi9 in *M. grisea* strains. Colored squares indicate gene contents and orthologue between two strains are connected by gray lines. Arrows indicate the orientations of genes and brown bands represent distributed TEs.

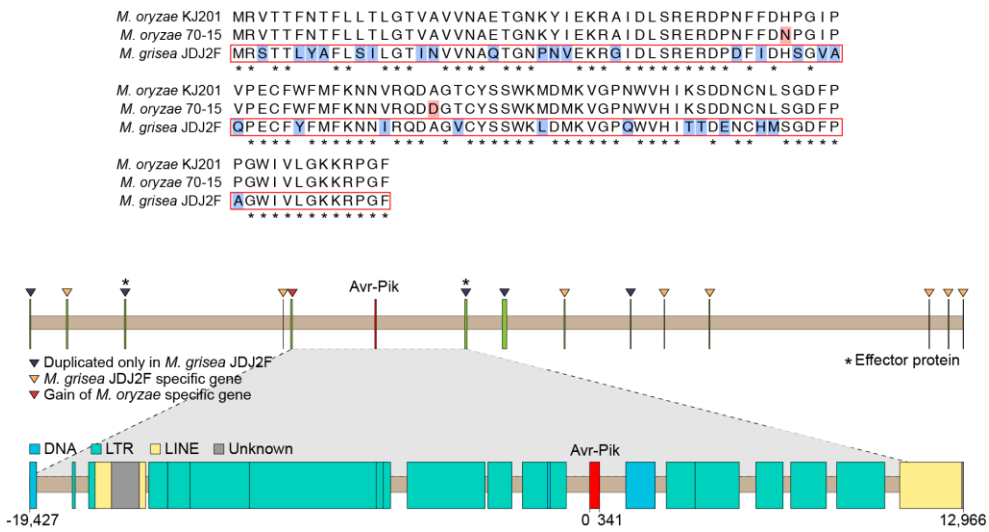


Figure 12. Genomic context of the AVR-Pik effector gene in *M. grisea* JDJ2F

(A) Comparison of AVR-Pik gene sequences among *M. grisea* JDJ2F and *M. oryzae* strains. Asterisks indicate conserved amino acids in three genomes. (B) Chromosomal region of AVR-Pik in *M. grisea* JDJ2F. Colored triangles indicate genes gained specifically in *M. grisea* JDJ2F located upstream or downstream of AVR-Pik. The intergenic region of AVR-Pik is expanded at the bottom of the figure. The red colored box indicates the AVR-Pik gene, and blue, green, yellow, and gray boxes represent distributed TE sequences close to AVR-Pik.

Table 6. Dsitribution of TEs in AVR-Pik genomic context of *M. grisea* JDJ2F

Description	Gene id/Repeat class	Length	Location in *M. grisea* JDJ2F genome	
			Start	End
	Unknown	59	4,180,195	4,180,254
	Unknown	5,965	4,180,369	4,186,334
	Unknown	858	4,185,830	4,186,688
	DNA/TcMar-Pogo	1,860	4,186,689	4,188,549
	LINE/Tad1	5,983	4,197,020	4,203,003
	Unknown	672	4,219,904	4,220,576
	LTR/Copia	261	4,220,577	4,220,838
	DNA/TcMar-Fot1	56	4,220,840	4,220,896
	LTR/Copia	258	4,220,897	4,221,155
	LTR/Gypsy	31	4,221,157	4,221,188
	LTR/Gypsy	146	4,221,189	4,221,335
	LTR/Gypsy	3,652	4,221,336	4,224,988
	LTR/Gypsy	124	4,224,963	4,225,087
	LTR/Gypsy	223	4,239,948	4,240,171
	LTR/Gypsy	124	4,240,469	4,240,593
	Unknown	602	4,247,589	4,248,191
	DNA/CMC-EnSpm	1,171	4,248,192	4,249,363
	Unknown	170	4,249,364	4,249,534
	LTR/Copia	169	4,249,364	4,249,533
	LTR/Copia	354	4,249,739	4,250,093
	LTR/Copia	1,798	4,250,129	4,251,927
	LTR/Copia	3,547	4,251,928	4,255,475
	DNA/TcMar-Fot1	1,864	4,255,476	4,257,340
	LTR/Copia	312	4,257,341	4,257,653
	LTR/Copia	63	4,257,606	4,257,669
	LTR/Copia	2,705	4,257,653	4,260,358
	LTR/Copia	339	4,260,359	4,260,698
	LTR/Copia	1,005	4,260,870	4,261,875
	LTR/Copia	4,505	4,261,876	4,266,381
Duplicated only in *M. grisea* JDJ2F	Mg_JDJ2F_00004436	248	4,267,234	4,267,482
	LTR/Copia	469	4,269,410	4,269,879
	LTR/Copia	548	4,270,385	4,270,933
	DNA/TcMar-Fot1	628	4,270,941	4,271,569
M. grisea JDJ2F-specific	Mg_JDJ2F_00004437	347	4,276,157	4,276,504
	LINE/Tad1	175	4,282,868	4,283,043
	LINE/Tad1	5,964	4,283,076	4,289,040
Duplicated only in *M. grisea* JDJ2F	Mg_JDJ2F_00004438	389	4,290,273	4,290,662
	DNA/TcMar-Ant1	686	4,291,482	4,292,168
	DNA/TcMar-Ant1	252	4,292,364	4,292,616
	LTR/Copia	309	4,292,618	4,292,927
	LTR/Copia	70	4,292,928	4,292,998
	LTR/Copia	404	4,293,017	4,293,421
	LTR/Copia	3,408	4,293,858	4,297,266

	LTR/Copia	107	4,297,779	4,297,886
	LTR/Copia	409	4,297,887	4,298,296
	LTR/Gypsy	2,492	4,298,297	4,300,789
	Unknown	985	4,300,731	4,301,716
	Unknown	5,414	4,301,759	4,307,173
	Unknown	1,380	4,307,502	4,308,882
	LTR/Gypsy	2,076	4,308,807	4,310,883
	LTR/Gypsy	7,239	4,310,884	4,318,123
	Unknown	1,523	4,318,124	4,319,647
	LTR/Copia	340	4,319,648	4,319,988
	DNA/TcMar-Ant1	1,623	4,319,990	4,321,613
	DNA/TcMar-Fot1	269	4,321,614	4,321,883
	LTR/Gypsy	3,498	4,322,428	4,325,926
	LINE/Tad1	820	4,325,927	4,326,747
	LINE/Tad1	235	4,328,578	4,328,813
M. grisea JDJ2F-specific	Mg_JDJ2F_00004439	29	4,328,852	4,328,881
	LTR/Gypsy	55	4,330,262	4,330,317
	LTR/Gypsy	108	4,330,510	4,330,618
	LTR/Gypsy	54	4,330,606	4,330,660
M. oryzae-specific	Mg_JDJ2F_00004440	434	4,330,672	4,331,106
	LINE/Tad1	240	4,331,730	4,331,970
	LTR/Gypsy	101	4,333,211	4,333,312
	LTR/Gypsy	227	4,333,774	4,334,001
	DNA/TcMar-Fot1	572	4,333,997	4,334,569
	Unknown	966	4,334,570	4,335,536
	DNA/TcMar-Fot1	219	4,335,537	4,335,756
	LTR/Gypsy	651	4,335,862	4,336,513
	LTR/Gypsy	771	4,336,515	4,337,286
	LTR/Gypsy	2,075	4,337,288	4,339,363
	LTR/Gypsy	4,403	4,339,356	4,343,759
	LTR/Gypsy	287	4,343,751	4,344,038
	LTR/Gypsy	257	4,343,993	4,344,250
	LTR/Gypsy	2,700	4,344,826	4,347,526
	LTR/Gypsy	830	4,347,625	4,348,455
	LTR/Gypsy	962	4,348,822	4,349,784
	LINE/Tad1	103	4,349,693	4,349,796
	LTR/Gypsy	564	4,349,784	4,350,348
AVR-Pik	Mg_JDJ2F_00004441	341	4,351,157	4,351,498
	LINE/Tad1	1,026	4,352,405	4,353,431
	LTR/Gypsy	998	4,353,813	4,354,811
	LTR/Gypsy	1,528	4,354,812	4,356,340
	LTR/Gypsy	950	4,356,916	4,357,866
	LTR/Gypsy	1,237	4,358,126	4,359,363
	LTR/Gypsy	1,667	4,359,722	4,361,389
	DNA/TcMar-Ant1	2,153	4,361,909	4,364,062
	Unknown	60	4,364,063	4,364,123
Duplicated only in *M. grisea* JDJ2F	Mg_JDJ2F_00004442	569	4,372,996	4,373,565
	DNA/TcMar-Fot1	604	4,374,599	4,375,203
	LTR/Gypsy	469	4,375,206	4,375,675

	Unknown	933	4,375,657	4,376,590
	DNA	124	4,376,591	4,376,715
	DNA/TcMar-Pogo	737	4,378,311	4,379,048
	DNA/TcMar-Fot1	1,136	4,379,265	4,380,401
Duplicated only in *M. grisea* JDJ2F	Mg_JDJ2F_00004443	1,058	4,382,136	4,383,194
	LTR/Gypsy	162	4,383,456	4,383,618
	Unknown	115	4,383,978	4,384,093
	DNA/Academ	132	4,384,079	4,384,211
	Unknown	5,030	4,384,095	4,389,125
	LTR/Copia	503	4,389,126	4,389,629
	LTR/Gypsy	124	4,389,648	4,389,772
	Unknown	2,624	4,389,769	4,392,393
	Unknown	866	4,392,376	4,393,242
	LTR/Copia	2,993	4,393,235	4,396,228
M. grisea JDJ2F-specific	Mg_JDJ2F_00004444	257	4,397,152	4,397,409
	Unknown	368	4,399,704	4,400,072
	Unknown	66	4,400,103	4,400,169
	LTR/Gypsy	83	4,400,175	4,400,258
	Unknown	917	4,400,386	4,401,303
	LTR/Gypsy	1,058	4,401,087	4,402,145
	LINE/CRE-Cnl1	865	4,412,170	4,413,035
Duplicated only in *M. grisea* JDJ2F	Mg_JDJ2F_00004445	200	4,413,225	4,413,425
	LTR/Gypsy	1,487	4,415,407	4,416,894
	LTR/Gypsy	85	4,417,601	4,417,686
	LTR/Gypsy	134	4,418,448	4,418,582
	LINE/Tad1	235	4,420,713	4,420,948
M. grisea JDJ2F-specific	Mg_JDJ2F_00004446	113	4,421,469	4,421,582
	LINE/Tad1	108	4,425,579	4,425,687
	LTR/Copia	104	4,425,914	4,426,018
	LTR/Gypsy	69	4,426,510	4,426,579
	LINE/Tad1	1,107	4,426,610	4,427,717
	LINE/CRE-Cnl1	280	4,427,549	4,427,829
	Unknown	332	4,427,557	4,427,889
	LINE/Tad1	73	4,432,214	4,432,287
M. grisea JDJ2F-specific	Mg_JDJ2F_00004447	206	4,432,380	4,432,586
	Unknown	1,141	4,436,020	4,437,161
	Unknown	756	4,437,160	4,437,916
	Unknown	124	4,443,349	4,443,473
	Unknown	1,173	4,445,292	4,446,465
	DNA	59	4,446,445	4,446,504
	Unknown	2,468	4,446,502	4,448,970
	DNA/TcMar-Fot1	1,164	4,449,853	4,451,017
	LTR/Gypsy	2,571	4,450,886	4,453,457
	Unknown	1,673	4,453,458	4,455,131
	Unknown	6,872	4,454,386	4,461,258
	Unknown	7,160	4,455,732	4,462,892

	Unknown	173	4,463,383	4,463,556
	Unknown	6,129	4,463,710	4,469,839
	Unknown	448	4,470,119	4,470,567
	Unknown	530	4,470,820	4,471,350
	Unknown	6,264	4,470,914	4,477,178
	Unknown	2,924	4,477,194	4,480,118
	Unknown	1,328	4,480,144	4,481,472
	Unknown	758	4,481,473	4,482,231
	DNA/TcMar-Fot1	113	4,482,434	4,482,547
	Unknown	146	4,482,543	4,482,689
	Unknown	2,839	4,482,695	4,485,534
M. grisea JDJ2F-specific	Mg_JDJ2F_00004448	122	4,485,848	4,485,970
	Unknown	196	4,490,156	4,490,352
M. grisea JDJ2F-specific	Mg_JDJ2F_00004449	174	4,490,537	4,490,711
	Unknown	162	4,491,664	4,491,826
	LTR/Gypsy	1,117	4,491,952	4,493,069
	DNA/TcMar-Fot1	465	4,493,381	4,493,846
	LTR/Gypsy	67	4,493,846	4,493,913
M. grisea JDJ2F-specific	Mg_JDJ2F_00004450	119	4,494,139	4,494,258
	Unknown	527	4,497,030	4,497,557
	Unknown	545	4,498,345	4,498,890
	Unknown	1,181	4,498,890	4,500,071
	LINE/Tad1	99	4,503,692	4,503,791
	LTR/Gypsy	124	4,504,007	4,504,131
	LTR/Gypsy	75	4,504,617	4,504,692
	LINE/Tad1	1,107	4,504,723	4,505,830
	LINE/Tad1	347	4,507,257	4,507,604
	LINE/Tad1	5,455	4,507,589	4,513,044
	Unknown	213	4,513,685	4,513,898

DISCUSSION

Host jumping in plant pathogenic fungi occurs frequently, resulting in expansion of their host range, and numerous studies have attempted to elucidate the underlying mechanisms. We showed that the *M. grisea* JDJ2F and YHL-684 strains have undergone host transition from crabgrass to rice; their presence did not simply represent contamination because mycelial growth of *M. grisea* JDJ2F was observed in rice. In previous studies, host jump events between pathotypes in *M. oryzae* were identified, along with cross-infection between isolates from rice and crabgrass (Choi et al., 2013;Yoshida et al., 2016;Inoue et al., 2017;Chung et al., 2020;Bentham et al., 2021), but this is first case of host transition between rice and crabgrass. These observations suggest that *M. grisea* has evolved to infect rice as a host, and has the potential to cause rice yield loss. Therefore, understanding the host jump process between crabgrass and rice is important to prevent such issues. Before complete host jump, preadaptation is necessary to suppress the defense mechanisms of the new host plant and allow colonization (Thines, 2019). *M. grisea* JDJ2F may aid this preadaptation stage of host jumping, because it is able to colonize rice but does not yet show virulence toward the new host. Host jumping is defined as the infection of a new host by pathogenic fungi, which must evolve to enhance infection and transmission (Thines, 2019;Zess et al., 2021). However, as the two strains examined here have not yet obtained the ability of virulecen to rice and could not identified complete lifecycle in new host, the term “host transition” is more suitable than “host jump” in this study. We sequenced the genome of host-transited *M. grisea* strains to

obtain genomic evidence of host transition from crabgrass to rice. However, the genomic features, including effector repertoires, were highly similar between rice-isolated *M. grisea* and crabgrass-isolated *M. grisea*. We could identified evidences in *M. oryzae* effector genes, which sequences of *M. grisea* JDJ2F were similar to those of rice-infecting *M. oryzae* strains. The AVR-Pi9 and AVR-Pik effector genes are *M. oryzae*-specific, with AVR-Pik being restricted to rice-infecting *M. oryzae* (Kim et al., 2019;Li et al., 2019;Wu et al., 2021). Similar to *M. grisea* strains NI907 and DS0505, there were also no AVR-Pik genes in *M. grisea* strains Dig41, BR29, DS0505, and DS9461 (Yoshida et al., 2016;Kim et al., 2019), indicating an extraordinary gain of AVR-Pik in *M. grisea* JDJ2F. We suggest that AVR-Pik may have been derived from rice-infecting *M. oryzae* by HGT or HCT. Horizontal transfer in plant pathogens facilitates adaptation to new host plants, and TEs may promote this process (Mehrabi et al., 2011;McDonald et al., 2019). The presence of a large number of TEs in AVR-Pik of *M. grisea* JDJ2F supports this transposon-mediated horizontal transfer. Moreover, the ~18 kb surrounding AVR-Pik contained genes only duplicated in *M. grisea* JDJ2F, shared with *M. oryzae* strains, and specific to *M. grisea* JDJ2F with numerous repeat sequences. The wheat-specific virulence gene ToxA was also horizontally transferred among wheat blast pathogens with a surrounding ~14 kb of genomic context (McDonald et al., 2019). Therefore, this genomic region may be a consequence of transposon-driven horizontal transfer and represent evidence of host adaptation to rice. Genetic diversity existed in AVR-Pik of *M. oryzae* according to five polymorphic sites (AVR-PikA, AVR-PikB, AVR-PikC, AVR-PikD, AVR-PikE, AVR-PikF) (Li et al., 2019;Longya et al., 2019). AVR-Pik

gene sequence in *M. oryzae* 70-15 and KJ201 correspond to AVR-PikC and AVR-PikD or AVR-PikE repectively, but AVR-Pik in *M. grisea* JDJ2F, which showed sequence differences in neumorus residue, did not belongs to this variation. Therefore we suggested to classify AVR-Pik of *M. oryzae* as AVR-Pik1 and *M. grisea* as AVR-Pik2. In contrast to AVR-Pik, AVR-Pi9 is conserved in diverse pathotypes of *M. oryzae*, and is also found in *M. grisea* with low similarity. *M. grisea* JDJ2F exhibited specific amino acid differences from other *M. grisea* strains, and these substituted sequences were identical to *M. oryzae* sequences. Inversion of the genomic context, including AVR-Pi9, and differences in the length of intergenic regions and transposon distribution, suggest that the genomic region surrounding AVR-Pi9 has also undergone evolutionary changes. This polymorphism in AVR-Pi9 of *M. girsea* JDJ2F was considered a significant finding because the sequences of *M. oryzae* and *M. grisea* AVR-Pi9 are well-conserved. However, the mechanisms responsible for these substitutions were not identified. Several fungal strains show cross infection between rice and crabgrass (Choi et al., 2013;Chung et al., 2020), but crabgrass-isolated *M. grisea* strains identified in this study unusally show an endophytic lifestyle in an isolated host, although it is pathogenic in the original host. Transition between pathogen and endophyte lifestyle has occurred multiple times in evolution, indicating that the endophytic lifestyle is not an evolutionarily stable trait (Delaye et al., 2013). Many studies have shown that the endophytic lifestyle can be changed by mutation in a single locus (Freeman and Rodriguez, 1993;Redman et al., 2001;Rodriguez et al., 2004;Rai and Agarkar, 2016), thereby breaking the balance between host and pathogen (Kuo et al., 2014), and by environmental changes

(Alvarez-Loayza et al., 2011).We hypothesized that the endophytic fungal strain *M. grisea* JDJ2F may exist transiently in rice as an intermediate step of the host jump to rice. Phylogenetic analysis revealed that biotrophic and saprotrophic fungi can be derived from endophytic fungi and vice versa (Delaye et al., 2013). This evolutionary pattern showed that an endophytic lifestyle could exist in the intermediate stage of the evolution of a pathogen to a different host, thus supporting the hypothesis that *M. grisea* JDJ2F is in an intermediate stage in the process of host shifting from crabgrass to a rice pathogen. We showed that *M. grisea* JDJ2F and YHL-684 shared genes with rice-isolated *M. oryzae* strains not found in crabgrass-isolated *M. grisea* strains. Most of these genes are assciated with DNA replication and cell wall degradation, and are required for penetration of rice cell walls in biotrophic infection (Quoc and Bao Chau, 2017;Fernandez and Orth, 2018). RNA sequencing analysis of *M. grisea* JDJ2F was not performed in this study, but we estimated the expression of gained and lost genes using previously reported time-series expression data for *M. oryzae* 70-15 (Jeon et al., 2020). The genes shared with rice-isolated *M. oryzae* strains, and those conserved in all *Magnaporthe* species but lost in rice-isolated *M. grisea*, were highly induced in the biotrophic and necrotrophic stages, respectively, suggesting that the gain and loss of genes may contribute to the fungal transition to an endophytic lifestyle.

LITERATURE CITED

Abdellatif, L., Bouzid, S., Kaminskyj, S., and Vujanovic, V. (2009). Endophytic hyphal compartmentalization is required for successful symbiotic Ascomycota association with root cells. *Mycol. Res.* **113**: 782-791.

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., Von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**: 420-423.

Alvarez-Loayza, P., White Jr, J.F., Torres, M.S., Balslev, H., Kristiansen, T., Svenning, J.-C., and Gil, N. (2011). Light converts endosymbiotic fungus to pathogen, influencing seedling survival and niche-space filling of a common tropical tree, *Iriartea deltoidea*. *PLoS One* **6**: e16386.

Becker, Y., Green, K.A., Scott, B., and Becker, M. (2018). Artificial inoculation of *Epichloë festucae* into *Lolium perenne*, and visualisation of endophytic and epiphyllous fungal growth. *Bio-protoc.* **8**: e2990-e2990.

Bentham, A.R., Petit-Houdenot, Y., Win, J., Chuma, I., Terauchi, R., Banfield, M.J., Kamoun, S., and Langner, T. (2021). A single amino acid polymorphism in a conserved effector of the multihost blast fungus pathogen expands host-target binding spectrum. *PLoS Pathog.* **17**: e1009957.

Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+

and AUGUSTUS supported by a protein database. *NAR genom. bioinform.* **3**: lqaa108.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**: 1194-1202.

Choi, J., Park, S.-Y., Kim, B.-R., Roh, J.-H., Oh, I.-S., Han, S.-S., and Lee, Y.-H. (2013). Comparative analysis of pathogenicity and phylogenetic relationship in *Magnaporthe grisea* species complex. *PLoS One* **8**: e57196.

Chung, H., Goh, J., Han, S.S., Roh, J.H., Kim, Y., Heu, S., Shim, H.K., Jeong, D.G., Kang, I.J., and Yang, J.W. (2020). Comparative pathogenicity and host ranges of *Magnaporthe oryzae* and related species. *Plant Pathol. J.* **36**: 305-313.

Couch, B.C., and Kohn, L.M. (2002). A multilocus gene genealogy concordant with host preference indicates segregation of a new species, *Magnaporthe oryzae*, from *M. grisea*. *Mycologia* **94**: 683-693.

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., and Pan, H. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980-986.

Delaye, L., García-Guzmán, G., and Heil, M. (2013). Endophytes versus biotrophic and necrotrophic pathogens—are fungal lifestyles evolutionarily stable traits? *Fungal Divers.* **60**: 125-135.

Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of

filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.* **35**: 57-65.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**: 1-14.

Farman, M.L., Eto, Y., Nakao, T., Tosa, Y., Nakayashiki, H., Mayama, S., and Leong, S.A. (2002). Analysis of the structure of the AVR1-CO39 avirulence locus in virulent rice-infecting isolates of *Magnaporthe grisea*. *Mol. Plant Microbe Interact*. **15**: 6-16.

Fernandez, J., and Orth, K. (2018). Rise of a cereal killer: the biology of *Magnaporthe oryzae* biotrophic growth. *Trends Microbiol.* **26**: 582-597.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* **117**: 9451-9457.

Fouché, S., Plissonneau, C., and Croll, D. (2018). The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr. Opin. Microbiol.* **46**: 34-42.

Freeman, S., and Rodriguez, R.J. (1993). Genetic conversion of a fungal plant pathogen to a nonpathogenic, endophytic mutualist. *Science* **260**: 75-78.

Gómez Luciano, L.B., Tsai, I.J., Chuma, I., Tosa, Y., Chen, Y.-H., Li, J.-Y., Li, M.-Y., Lu, M.-Y.J., Nakayashiki, H., and Li, W.-H. (2019). Blast fungal genomes show frequent chromosomal changes, gene gains and losses, and effector gene turnover. *Mol. Biol. Evol.* **36**: 1148-1161.

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**: 1-14.

Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**: 7.

Huang, J., Si, W., Deng, Q., Li, P., and Yang, S. (2014). Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genet.* **15**: 1-10.

Inoue, Y., Vy, T.T., Yoshida, K., Asano, H., Mitsuoka, C., Asuke, S., Anh, V.L., Cumagun, C.J., Chuma, I., and Terauchi, R. (2017). Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science* **357**: 80-83.

Islam, M.T., Kim, K.H., and Choi, J. (2019). Wheat blast in bangladesh: the current situation and future impacts. *Plant Pathol. J.* **35**: 1-10.

Jeon, J., Lee, G.-W., Kim, K.-T., Park, S.-Y., Kim, S., Kwon, S., Huh, A., Chung, H., Lee, D.-Y., and Kim, C.-Y. (2020). Transcriptome profiling of the rice blast fungus *Magnaporthe oryzae* and its host *Oryza sativa* during infection. *Mol. Plant Microbe Interact.* **33**: 141-144.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., Mcanulla, C., Mcwilliam, H., Maslen, J., Mitchell, A., and Nuka, G. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Kang, S., Sweigard, J.A., and Valent, B. (1995). The PWL host specificity gene family in the blast fungus *Magnaporthe grisea*. *Mol. Plant Microbe Interact.*

8: 939-948.

Kim, H., Lee, K.K., Jeon, J., Harris, W.A., and Lee, Y.-H. (2020). Domestication of *Oryza* species eco-evolutionarily shapes bacterial and fungal communities in rice seed. *Microbiome* **8**: 1-17.

Kim, K.-T., Ko, J., Song, H., Choi, G., Kim, H., Jeon, J., Cheong, K., Kang, S., and Lee, Y.-H. (2019). Evolution of the genes encoding effector candidates within multiple pathotypes of *Magnaporthe oryzae*. *Front. Microbiol.* **10**: 2575.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**: 567-580.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**: 1639-1645.

Kuo, H.-C., Hui, S., Choi, J., Asiegbu, F.O., Valkonen, J., and Lee, Y.-H. (2014). Secret lifestyles of *Neurospora crassa*. *Sci. Rep.* **4**: 1-6.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* **5**: 1-9.

Li, J., Cornelissen, B., and Rep, M. (2020). Host-specificity factors in plant pathogenic fungi. *Fungal Genet. Biol.* **144**: 103447.

Li, J., Wang, Q., Li, C., Bi, Y., Fu, X., and Wang, R. (2019). Novel haplotypes and networks of AVR-Pik alleles in *Magnaporthe oryzae*. *BMC Plant Biol.* **19**: 1-12.

Longya, A., Chaipanya, C., Franceschetti, M., Maidment, J.H., Banfield, M.J., and Jantasuriyarat, C. (2019). Gene duplication and mutation in the emergence of a novel aggressive allele of the AVR-Pik effector in the rice blast fungus. *Mol. Plant Microbe Interact.* **32**: 740-749.

Mcdonald, M.C., Taranto, A.P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., Milgate, A., and Solomon, P.S. (2019). Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *MBio* **10**: e01515-01519.

Mehrabi, R., Bahkali, A.H., Abd-Elsalam, K.A., Moslem, M., Ben M'barek, S., Gohari, A.M., Jashni, M.K., Stergiopoulos, I., Kema, G.H., and De Wit, P.J. (2011). Horizontal gene and chromosome transfer in plant pathogenic fungi affecting host range. *FEMS Microbiol. Rev.* **35**: 542-554.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**: 1530-1534.

Park, S.-Y., Milgroom, M.G., Han, S.-S., Kang, S., and Lee, Y.-H. (2003). Diversity of pathotypes and DNA fingerprint haplotypes in populations of *Magnaporthe grisea* in Korea over two decades. *Phytopathology* **93**: 1378-

1385.

Quoc, N.B., and Bao Chau, N.N. (2017). The role of cell wall degrading enzymes in pathogenesis of *Magnaporthe oryzae*. *Curr. Protein Pept. Sci.* **18**: 1019-1034.

Rai, M., and Agarkar, G. (2016). Plant–fungal interactions: what triggers the fungi to switch among lifestyles? *Crit. Rev. Microbiol.* **42**: 428-438.

Raja, H.A., Miller, A.N., Pearce, C.J., and Oberlies, N.H. (2017). Fungal identification using molecular tools: a primer for the natural products research community. *J. Nat. Prod.* **80**: 756-770.

Redman, R.S., Dunigan, D.D., and Rodriguez, R.J. (2001). Fungal symbiosis from mutualism to parasitism: who controls the outcome, host or invader? *New Phytol.* **151**: 705-716.

Rodriguez, R.J., Redman, R.S., and Henson, J.M. (2004). The role of fungal symbioses in the adaptation of plants to high stress environments. *Mitig. Adapt. Strateg. Glob. Chang.* **9**: 261-272.

Schulze-Lefert, P., and Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends Plant Sci.* **16**: 117-125.

Seidl, M.F., and Thomma, B.P. (2014). Sex or no sex: evolutionary adaptation occurs regardless. *Bioessays* **36**: 335-345.

Sharma, R., Mishra, B., Runge, F., and Thines, M. (2014). Gene loss rather than gene

gain is associated with a host jump from monocots to dicots in the smut fungus *Melanopsichium pennsylvanicum*. *Genome Biol. Evol.* **6**: 2034-2049.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., Mcwilliam, H., Remmert, M., and Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**: 539.

Smit, A., Hubley, R., and Green, P. (2013). "2013–2015. RepeatMasker Open-4.0".

Sonah, H., Deshmukh, R.K., and Bélanger, R.R. (2016). Computational prediction of effector proteins in fungi: opportunities and challenges. *Front. Plant Sci.* **7**: 126.

Sperschneider, J., and Dodds, P.N. (2022). EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant Microbe Interact.* **35**: 146-156.

Sweigard, J.A., Carroll, A.M., Kang, S., Farrall, L., Chumley, F.G., and Valent, B. (1995). Identification, cloning, and characterization of PWL2, a gene for host species specificity in the rice blast fungus. *Plant Cell* **7**: 1221-1233.

Thines, M. (2019). An evolutionary framework for host shifts–jumping ships for survival. *New Phytol.* **224**: 605-617.

Tosa, Y., Hirata, K., Tamba, H., Nakagawa, S., Chuma, I., Isobe, C., Osue, J., Urashima, A., Don, L., and Kusaba, M. (2004). Genetic constitution and pathogenicity of Lolium isolates of *Magnaporthe oryzae* in comparison with host species-specific pathotypes of the blast fungus. *Phytopathology* **94**:

454-462.

Tosa, Y., Osue, J., Eto, Y., Oh, H.-S., Nakayashiki, H., Mayama, S., and Leong, S.A. (2005). Evolution of an avirulence gene, AVR1-CO39, concomitant with the evolution and differentiation of *Magnaporthe oryzae*. *Mol. Plant Microbe Interact.* **18**: 1148-1160.

Wu, Q., Wang, Y., Liu, L.-N., Shi, K., and Li, C.-Y. (2021). Comparative genomics and gene pool analysis reveal the decrease of genome diversity and gene number in rice blast fungi by stable adaption with rice. *J. Fungi.* **8**: 5.

Wyka, S.A., Mondo, S.J., Liu, M., Dettman, J., Nalam, V., and Broders, K.D. (2021). Whole-genome comparisons of ergot fungi reveals the divergence and evolution of species within the genus *Claviceps* are the result of varying mechanisms driving genome evolution and host range expansion. *Genome Biol. Evol.* **13**: evaa267.

Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., and Win, J. (2009). Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* **21**: 1573-1591.

Yoshida, K., Saunders, D.G., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue, Y., Chuma, I., Tosa, Y., and Cano, L.M. (2016). Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genom.* **17**: 1-18.

Zess, E.K., Dagdas, Y.F., Peers, E., Maqbool, A., Banfield, M.J., Bozkurt, T.O., and

Kamoun, S. (2022). Regressive evolution of an effector following a host jump in the Irish potato famine pathogen lineage. *PLoS Pathog.* **18**: e1010918.

Zhong, Z., Norvienyeku, J., Chen, M., Bao, J., Lin, L., Chen, L., Lin, Y., Wu, X., Cai, Z., and Zhang, Q. (2016). Directional selection from host plants is a major force driving host specificity in *Magnaporthe* species. *Sci. Rep.* **6**: 1-13.

지의류 형성 곰팡이와 내생균의 비교유전체학적 분석

송 현 정

초 록

공생(symbiosis)은 생태계에 존재하는 관계 중 하나로, 두 유기체 중 적어도 한쪽이 이득을 얻으며 서로 영향을 주고받는 형태를 말한다. 대부분의 식물들은 외부환경으로부터 물과 양분을 좀 더 효율적으로 얻기 위해 공생균과 상호작용을 하고 있다. 분자유전학적 연구만으로는 공생 생활사를 이해하는데 한계가 있었지만, 최근 염기서열 분석 기술의 발달로 인해 포괄적인 비교 유전체학적 연구와 진화적 관점에 대한 접근이 가능해졌다. 공생균은 각각 식물의 뿌리, 조류, 그리고 식물의 전체 조직과 상호작용하는 균근(mycorrhizal fungi), 지의류 형성 곰팡이(lichen-forming fungi), 내생균(endophytic fungi)으로 나뉜다. 최근에는 많은 수의 균근 유전체 서열이 밝혀졌고, 비교 유전체학 분석을 통해 식물 세포벽 분해효소(plant cell wall degrading enzyme) 유전자의 손실과 계통 특이적인 이펙터(effector) 유전자들의 획득으로 인해 균근의 공생생활사가 나타났음이 밝혀졌다. 하지만 이에 비해 지의류와 내생균의 유전체 연구는 부족한 것으로 보인다.

본 연구에서는 공생균인 지의류 형성 곰팡이와 내생균의 유전체적 특성을 두 장에 걸쳐 분석하였다. 첫 번째 장에서는 여섯 종의 지의류 형성 곰팡이 유전체를 비교분석하여 조류와의 공생관계 형성에 관여하는 요소를 유전체 서열에서 찾아보고자 하였다. 지의류 형성 곰팡이들의 유전체 서열은 서로 잘 유지 되어있지 않았으며, 각각 다른

공통조상으로부터 유래된 것을 확인할 수 있었다. 하지만 지의류 계통에서 공통적으로 사이토크롬 P450(cytochrome P450) 유전자군의 확장과 식물 세포벽 분해효소, 당 수송체(sugar transporter), 전사 인자(transcription factor) 유전자군의 수축이 일어났음을 밝혔다. 이는 지의류 형성 곰팡이가 조류와의 상호작용에서 비 침투 표면(non-penetrating interface)을 형성하고, 조류로부터 광합성 산물을 특정 폴리올(polyol) 형태로 제공받는 것과 상응하는 결과이다. 또한 시간대별 전사체 정보를 통해 지의류 형성 곰팡이 특이적 유전자와 다른 곰팡이에서도 보존 되어있는 유전자들이 각각 공생 형성과정의 초기와 후기 단계에 관여한다는 것을 밝혀내었다.

두 번째 장에서는 바랭이에서 벼로 기주 전이(host transition)된 *Magnaporthe grisea* JDJ2F와 YHL-684 균주들을 찾아내었으며, 이들은 벼에서 내생균으로 존재함을 확인하였다. 두 균주는 벼에서 분리되었지만 유전형과 전체 유전자 서열 비교를 통해 원래 기주 (origin host)는 바랭이임이 밝혀졌다. 기주 결정 요인으로 알려진 이펙터 유전자들은 *M. grisea* 균주들 사이에서 큰 차이를 보이지 않았지만, *M. oryzae* 이펙터 유전자들에서 *M. grisea* JDJ2F 균주의 기주 전이의 증거를 확보할 수 있었다. AVR-Pi9은 일부 아미노산 서열에서 다형성(polymorphism)을 보이는 것을 확인하였고, 이는 *M. oryzae* 서열과 동일한 것으로 나타났다. 또한, 벼를 기주로 가지는 *M. oryzae*에 특이적인 AVR-Pik 유전자가 *M. grisea* JDJ2F에 존재하며 주변에 많은 수의 전이 인자(transposable element)와 *M. grisea* JDJ2F 특이적으로 획득된 유전자들이 분포하는 것을 확인하였다. 이러한 증거들을 토대로, *M. grisea* JDJ2F 균주는 바랭이에서 벼로 기주가 전이되는 과정에 있으며 벼에 대한 병원성을 얻기 전에 내생균으로 존재할 수 있다는 가설을 설정할 수 있었다.

본 비교유전체 연구는 지의류 형성 곰팡이가 어떻게 공생관계를 형성하고 유지할 수 있었는지, 그리고 기주 전이의 중간단계로 보이는

내생균에서의 기주 전이 증거는 무엇인지를 밝혀내는 것을 목표로 하였다. 이 연구결과들은 공생관계를 형성하는 곰팡이들의 유전체적 특성을 보여줄 뿐만 아니라 추후 분자생물학적 연구의 기반을 제공할 것이다. 궁극적으로는 곰팡이의 생활양식을 이해하는데 새로운 식견을 제공할 것이다.

주요어: 공생균, 비교유전체학, 지의류 형성 곰팡이, 내생균, 기주 전이

학번: 2014-21922