Corrections

CELL BIOLOGY


The authors note that Figure 1 appeared incorrectly. The corrected figure and its legend appear below.

Fig. 1 Dual inhibition of MEK and PI3K induces phosphorylation of DDR substrates. (A) A2058 or 888MEL melanoma cells were treated for 6 h with DMSO or GDC-0973+GDC-0941 (MEKi/PI3Ki combo; 4× EC50) and were subjected to KinomeView Profiling. A2058 and 888MEL cells were treated with 10 μM GDC-0973 + 10 μM GDC-0941 or 0.2 μM GDC-0973 + 10 μM GDC-0941, respectively. Blots were probed using an antibody mixture recognizing pRSK, pAKT, pERK, and pS6 or phosphomotif antibodies (e.g., DDR substrates with the [s/t]Q motif and AKT substrates with the RXX[s/t] and RXRXX[s/t] motifs). (B) A2058 lysates probed with the [s/t]Q and RXX[s/t] antibodies after treatment with DMSO, 10 μM GDC-0973 (MEKi), 10 μM GDC-0941 (PI3Ki), or the combination. (C) Dose response of A2058 cells to increasing concentrations of MEKi and PI3Ki alone or in combination. Blots were performed against DDR (p53 pSer15, histone 2AX pSer139), cell survival/cell death (AKT pThr308, cleaved PARP), and cell signaling (ERK1/2 pThr202/Tyr204) markers and controls. Actin and GAPDH served as loading controls.

www.pnas.org/cgi/doi/10.1073/pnas.1322630111
ENVIRONMENTAL SCIENCES

The authors note that the author name Peter W. Young should instead appear as J. Peter W. Young. The corrected author line appears below. The online version has been corrected.


GENETICS

The authors note that the following statement should be added to the Acknowledgments: “Y.S. is supported by the Israel Science Foundation (Grants 1604/13 and 877/13) and the European Research Council (Grant StG-335377).”

MEDICAL SCIENCES

The authors note that on page 10295, right column, 2nd full paragraph, line 8 “10 mg/kg xylene” should instead appear as “10 mg/kg xylazine.”
Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis

Emilie Tissierant1, Mathilde Malbreil1, Alan Kuo1, Annett Gret Koller1, Aikaterini Symeoni1,2, Raffaella Balestrini2, Philippe Charron9, Nina Duensing9, Nicolas Frei dit Frey9, Vivienne Gianinazzi-Pearson1, Luz B. Gilbert1, Yoshihiro Handa1, Joshua R. Herr3, Mohamed Hijri1, Raman Koul1, Masayoshi Kawaguchi1, Franziska Krajinski1, Peter J. Lambers2, Frederic G. Masclaux2,3, Claude Murat4, Emmanuelle Morin4, Steve Ndiikumana5,6, Marco Pagni7, Denis Petitpierre8, Natalia Requena8, Pavel Rosikiewicz8, Rohan Riley8, Katsuhiro Saigo7, Hélène San Clemente8,9, Harris Shapiro10, Diederik van Tuinen11, Guillaume Bècard12, Paola Bonfante13, Uta Paszkowski14, Yair Y. Shachar-Hill15, Gerald A. Tuskani16, J. Peter W. Young16, Ian R. Sanders17, Bernard Hennessat18,19,20, Stefan A. Rensing20,21, Igor V. Grigoriev22, Nicolas Corradi9, Christoph Roux20, and Francis Martin23,24

1Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1136, Interactions Arbres/Microorganismes, Centre de Nancy, Université de Lorraine, 54280 Champenoux, France; 2Centre National de la Recherche Scientifique, Université Paul Sabatier, Unité Mixte de Recherche 5546, Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, F-31326 Castanet-Tolosan, France; 3US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598; 4BioSS Centre for Biological Signalling Studies and Faculty of Biology, University of Freiburg, D-79104 Freiburg, Germany; 5Faculty of Biology, University of Marburg, D-35043 Marburg, Germany; 6Institut für Biologie des Lebens, Zentrum für Integrative Biologie, Universität Freiburg, 79104 Freiburg, Germany; 7Department of Sciences della Vita e Biologia dei Sistemi, Istituto per la Protezione delle Piante del Consiglio Nazionale delle Ricerche, 10125 Torino, Italy; 8Canadian Institute for Advanced Research, Department of Biology, University of Ottawa, Ottawa, ON, Canada K1N 6N5; 9Max Planck Institut für Molekulare Pflanzenphysiologie, 14476 Potsdam, Germany; 10Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1347, Agroécologie, Pôle Interaction Plantes–Microorganismes, Université de Bourgogne, 21065 Dijon, France; 11Department of Evolutionary, Biology and Biodiversity, Division of Symbiotic Systems, National Institute for Basic Biology, Aichi 444-8585, Japan; 12Institut de la Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, Montréal, QC, Canada H3X 2B2; 13Department of Chemistry and Biochemistry, New Mexico State University, Las Cruces, NM 88003-8801; 14Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland; 15Vital-IT Group, Swiss Institute for Bioinformatics, 1015 Lausanne, Switzerland; 16Botanical Institute, Plant-Microbial Interaction, Karlsruhe Institute of Technology, D-76187 Karlsruhe, Germany; 17Faculty of Agriculture, Shinshu University, Nagano 399-4588, Japan; 18Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom; 19Department of Plant Biology, Michigan State University, East Lansing, MI 48824; 20Bioscience Department, Oak Ridge National Laboratory, Oak Ridge, TN 37831; 21Department of Biology, University of York YO10 5DD, United Kingdom; 22Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, 13288 Marseille Cedex 9, France; 23Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7257, Aix-Marseille Université, 13288 Marseille Cedex 9, France; and 24Department of Cellular and Molecular Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark

Edited by Paul Schulze-Lefert, Max Planck Institute for Plant Breeding Research, Cologne, Germany, and approved October 28, 2013 (received for review July 18, 2013)

The arbuscular mycorrhizal symbiosis involving Glomeromycota, a distinctive phylum of early diverging fungi, is widely hypothesized to have promoted the evolution of land plants during the middle Paleozoic. These arbuscular mycorrhizal fungi (AMF) perform vital functions in the phosphorus cycle that are fundamental to sustainable crop plant productivity. The unusual biological features of AMF have long fascinated evolutionary biologists. The coenocytic hyphae host a community of hundreds of nuclei and reproduce clonally through large multinucleated spores. It has been suggested that the AMF maintain a stable assemblage of several different genomes during the life cycle, but this genomic organization has been questioned. Here we introduce the 153-Mb haploid genome of *Rhizophagus irregularis* and its repertoire of 28,232 genes. The observed low level of genome polymorphism (0.43 SNP per kb) is consistent with the occurrence of multiple, highly diverged genomes. The expansion of mating-related genes suggests the existence of cryptic sex-related processes. A comparison of gene categories confirms that *R. irregularis* is close to the Mucoromycotina. The AMF obligate biotrophy is not explained by genome erosion or any related loss of metabolic complexity in central metabolism, but is marked by a lack of genes encoding plant cell wall-degrading enzymes and of genes involved in toxin and thiamine synthesis. A battery of mycorrhiza-induced secreted proteins is expressed in symbiotic tissues. The present comprehensive repertoire of *R. irregularis* genes provides a basis for future research on symbiosis-related mechanisms in Glomeromycota.

carbohydrate-active enzymes | effector | fungal evolution | glomales | mutualism

The arbuscular mycorrhizal symbiosis between fungi of the Glomeromycota phylum and plants involves more than two-thirds of all known plant species, including important crop species. This mutualistic symbiosis, involving one of the oldest fungal lineages, is arguably the most ecologically and agriculturally important symbiosis in terrestrial ecosystems. The Glomeromycota are unique in that their spores and coenocytic hyphae contain hundreds of nuclei in a common cytoplasm, which raises important questions about the natural selection, population genetics, and gene expression of these highly unusual organisms. Study of the genome of *Rhizophagus irregularis* provides insight into genes involved in obligate biotrophy and mycorrhizal symbioses and the evolution of an ancient asexual organism, and thus is of fundamental importance to the field of genome evolution.

Significance

The arbuscular mycorrhizal symbiosis between fungi of the Glomeromycota phylum and plants involves more than two-thirds of all known plant species, including important crop species. This mutualistic symbiosis, involving one of the oldest fungal lineages, is arguably the most ecologically and agriculturally important symbiosis in terrestrial ecosystems. The Glomeromycota are unique in that their spores and coenocytic hyphae contain hundreds of nuclei in a common cytoplasm, which raises important questions about the natural selection, population genetics, and gene expression of these highly unusual organisms. Study of the genome of *Rhizophagus irregularis* provides insight into genes involved in obligate biotrophy and mycorrhizal symbioses and the evolution of an ancient asexual organism, and thus is of fundamental importance to the field of genome evolution.


This article is a PNAS Direct Submission.

Data deposition: Sequences have been deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank (accession no. AUPC00000000). RNA-Seq reads have been submitted at the National Center for Biotechnology Information’s Sequence Read Archive (accession nos. SRX1072885, SRX1329862 and SRX132214).

*E.T. and M.M. contributed equally to this work.

1To whom correspondence should be addressed. E-mail: fmart@nancy.inra.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1313452110/DCSupplemental.
These arbuscular mycorrhizal fungi (AMF) perform vital functions in the phosphorus cycle (3) that are fundamental to sustainable crop plant productivity (4). They also drive plant diversity (5). The extraradical mycelium of the symbiont acts as an extension of the root system and increases the uptake of key nutrients, particularly phosphorus (3). Furthermore, because the colonization of plants by AMF also can result in a 20% net increase in photosynthesis, these fungi make a very large, poorly understood contribution to the global carbon cycling budget of terrestrial ecosystems.

The Glomeromycota are unique in that their spores and coenocytic hyphae contain multiple nuclei in a common cytoplasm. With no known sexual cycle (6), AMF reproduce clonally through large and multinucleated spores (Fig. 1), although a conserved meiotic machinery (7, 8) might allow individuals to shuffle their genetic material (9, 10) and reduce their mutational load. Genetic variation has been observed within AMF in ribosomal DNA and in other regions of the genome (11). It has been hypothesized that AMF maintain an assemblage of genetically different nuclei (heterokaryosis) and transmit them from generation to generation (12, 13). Another study failed to find evidence for heterokaryosis, however, and the authors suggested that the genetic variation that they observed potentially could be due to polyploidy, although no studies of ploidy were conducted (14). A recent study of Rhizophagus irregularis (but not isolate DAOM-197198) provided further evidence in favor of heterokaryosis (15), so this remains an open question. Another hypothesis, specific for multicopy ribosomal DNA, is that variation among copies could exist within nuclei of a homokaryotic AM fungus and this is supported by some studies (14, 16).

Here we introduce the assembly and annotation of the genome of R. irregularis DAOM-197198 (formerly Glomus intraradices) in association with transcriptome data, and show that it is remarkably different from other sequenced fungal genomes in content and organization. We focus on the genome polymorphism, annotation, and transcript profiling of gene families likely to be involved in symbiosis, to reveal adaptation processes for growth in planta.

**Results and Discussion**

**Genome Assembly.** To investigate the gene repertoire and genome polymorphism of R. irregularis, we sequenced the genomic DNA from multinucleated hyphae of the strain DAOM-197198 grown in root culture of carrot (Daucus carota) (Fig. 1). The size of the genome assembly is 101 megabases (Mb), and the coding space is 98% complete on the basis of conserved core eukaryotic single-copy genes (17) (SI Appendix, sections 1.1–1.3, Figs. S1–S5). As expected, this genome is rich in A and T bases (A + T content, 72%) (Fig. 2). Based on flow cytometry assays and Feulgen densitometry measurements, the size of the DAOM-197198 genome has been estimated as 154.8 ± 6.2 Mb (18). Using the frequency distribution of 17-base oligomers in the usable sequencing reads to determine sequencing depth (19), we obtained an estimated genome size of 153 Mb (SI Appendix, Fig. S6). Thus, the genome size of R. irregularis is among the largest fungal genome sequenced to date, along with the obligate biotrophic powdery mildews (20) and the ectomycorrhizal symbiont Tuber melanosporum (21). Interestingly, no contigs corresponding to multiple haplotypes were identified. Transposable elements (TEs) (Fig. 2; 11% of the assembly) (SI Appendix, section 1.5, Table S1 and Figs. S3–S5), but fosmid Sanger sequencing showed a much higher TE abundance (36%), including several retrotransposons. The long (9–25 kb), highly repetitive, and nested nature of the transposons (SI Appendix, Fig. S4) is the main explanation for the observed fragmentation of this assembly. Assuming that the TE abundance of fosmids (36%) holds for the whole genome, this would contain ~55 Mb of repeated sequences, possibly reflecting a lack of efficient control mechanisms to prevent the expansion of repetitive elements and their subsequent elimination.

**Genome Polymorphism.** Evidence exists from this and other Glomeromycota species that these fungi are heterokaryotic, that is, harbor genetically different nuclei (11, 12). To investigate whether large differences could exist among nuclei, we analyzed the assembled contigs for local similarities to other contigs within the assembly. A BLAST search of all contigs against all contigs was carried out. The proportion of contigs sharing significant similarity (>90% identity) with at least one other contig of >1,000 bp was low (5.6%) and involved mostly repetitive sequences. Neither segmental duplication nor distinct haplotype contigs were detected, suggesting that the assembled data are not composed of multiple genomes.

We also investigated the possibility that alleles had been collapsed during the assembly through genomic and RNA-Seq read...
mapping and single nucleotide polymorphism (SNP) calling (SI Appendix, section 1.4). The polymorphism in the assembled genome was estimated as 0.43 SNP per kb (Table 1). The SNP rate for expressed genes was identical (0.40 SNP per kb) as measured by mapping DAOM-197198 RNA-Seq reads to the assembly (Table 1), and only a few genes were found to have more than three SNPs (SI Appendix, Table S2). Using identical coverage and quality thresholds, we estimated the SNP rate of the assembled genomes as 0.06 SNP per kb in the homokaryotic ascomycete T. melanosporum and 0.78 SNP per kb in the dikaryotic basidiomycete Laccaria bicolor. These findings allow us to reject the hypothesis that large sequence polymorphism occurs among coexisting nuclei in the DAOM-197198 isolate; however, the density of SNPs suggests some allelic variation among nuclei, likely reflecting the limited opportunities for recombination and related sequence homogenization during the life cycle.

In contrast, the SNP/substitution density was 6.5- and 20-fold higher when RNA-Seq reads from the strains R. irregularis C2 (2.6 SNPs per kb) and R. diaphanum (8.0 substitutions per kb), respectively, were mapped to the DAOM-197198 genome (Table 1). We also assessed the intrastrain and interstrain/species SNPs/substitutions by mapping RNA-Seq reads to assembled transcripts. The intrastrain SNP density of the transcripts was 0.40, 0.25, and 0.22 per kb for the DAOM-197198, C2, and MUCL43196 isolates, respectively. The interstrain SNP/substitution density of the transcripts ranged from 2.3 for DAOM-197198 vs. C2 to 9.3 for DAOM-197198 vs. R. diaphanum MUCL43196 (Table 1), confirming significant intraspecific and interspecific genetic variability. Only 291 SNPs/substitutions were shared among all strains analyzed (SI Appendix, Fig. S7).

We calculated the number of synonymous substitutions per synonymous site of paralogs, but found no evidence for any recent whole genome duplication. The distribution of paralog duplication age (SI Appendix, section 1.8, Fig. S8) did not follow the typical steep exponential decay pattern (22) that is the hallmark of gene birth and death by constantly occurring small-scale duplication events. Gene loss in DAOM-197198 may occur at a much lower rate than typically observed; however, eight hidden components contributing to the distribution were detected (SI Appendix, Fig. S8), potentially representing remnants of large-scale duplication events under an alternative scenario. Strikingly, all but the model component representing the youngest paralogs were enriched for genes annotated to contribute to biological processes related to phosphorus metabolism and

![Fig. 2. Circos circular visualization of the genome assembly. (A) The 30 largest scaffolds of the genome assembly. (B) Locations of gene models (blue), repeated elements (red) and sequence gaps (gray). (C) Genomic SNP density. (D) Read coverage on genome (scale, 0–100). (E) Expressed gene SNP density (outer to inner: RNA-Seq tags from DAOM-197198 spores, R. irregularis C2 spores, and R. diaphanum spores). (F) Transcript coverage, RNA-Seq reads from spores (scale, 0–10,000). (G) Transcript coverage, RNA-Seq reads from symbiotic roots (scale, 0–200). (H) Guanosine and cytosine (GC) content based on a sliding window of 100 bp (red, >40%; green, <20%; midline, 33%).](image)

### Table 1. SNP and substitution density in the R. irregularis genome and transcriptome

<table>
<thead>
<tr>
<th>SNP or substitution per kb</th>
<th>DAOM-197198 genome</th>
<th>DAOM-197198 exons</th>
<th>C2 transcriptome</th>
<th>R. diaphanum transcriptome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genomic reads</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrastrain SNPs per kb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genomic, raw</td>
<td>1.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genomic, filtered*</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genomic, exons, filtered*</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RNA-Seq reads</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of intrastrain SNPs per kb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAOM-197198</td>
<td>0.40</td>
<td></td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. diaphanum</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of interstrain SNPs/substitutions per kb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAOM-197198–C2</td>
<td>2.6</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAOM-197198–R. diaphanum</td>
<td>8.0</td>
<td>9.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2–R. diaphanum</td>
<td>7.7</td>
<td>9.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Illumina genomic reads from R. irregularis DAOM-197198 were mapped to the R. irregularis Gloin1 assembly. RNA-Seq reads from R. irregularis DAOM-197198, R. irregularis C2, and R. diaphanum were mapped to the R. irregularis DAOM-197198 exons, R. irregularis C2 transcriptome, and R. diaphanum transcriptome, respectively. Fixed differences between species are termed substitutions, and variable positions within species are SNPs.

* Filtered values correspond to values without positions with coverage <5 and coverage in the top 5%, to remove potential artifactual SNPs caused by repetitive and paralogous sequences.
signaling via phosphorylation (SI Appendix, Fig. S9), the vast majority of which were annotated as kinases. We suggest that the expansion of the kinase-based signaling network observed in the R. irregularis genome (see below) may be derived from frequent retention of duplicated kinase genes over a long period.

The Rhizophagus Gene Repertoire. Of the 28,232 protein-coding genes predicted (Fig. 2 and SI Appendix, Figs. S1 and S10), 23,561 high-confidence genes had transcriptomic support (RNA-Seq) and/or showed sequence similarity to documented proteins and/or domains (SI Appendix, sections 1.6 and 1.8 and Tables S3 and S4). Only 62% of the high-confidence genes showed sequence similarity to documented proteins and/or domains (SI Appendix, Figs. S11 and S12). Compared with representative sequenced fungi, including the taxonomically related Mucoromycotina, the percentage of proteins encoded by species-specific genes (SI Appendix, Fig. S12) and multigene families (SI Appendix, Fig. S13) in R. irregularis was among the highest observed. Expansion of protein family sizes was prominent in the lineage-specific multigene families (SI Appendix, Fig. S13 and Table S3), but marked gene family expansions were also seen in genes predicted to have roles in signal transduction mechanisms [e.g., tyrosine kinase-like genes (TKLs)], in protein–protein interactions (e.g., Sel1-domain-containing proteins), and RNA interference-related mechanisms (e.g., the Argonaute proteins) (SI Appendix, Tables S6 and S7). Notably, several TKL-containing proteins are associated with Sel1 repeats and were highly expressed in germinating spores and intraradical mycelium (SI Appendix, Fig. S14).

Together with RNA-dependent RNA polymerases, DICER (IPR011545), and C-5 cytosine-specific DNA methylases, the unusually high number of Argonaute genes (SI Appendix, Table S7) are likely involved in silencing the abundant TEs. We identified gene families with a smaller size in R. irregularis compared with other fungi (SI Appendix, Table S8). The inability of R. irregularis to grow in vitro suggests that the obligate biotrophic genome may lack genes typically present in autotrophic fungi. Thus, we systematically searched for genes absent in R. irregularis, the obligate biotrophic pathogen Blumeria graminis (20), and early diverging Mucoromycotina and Chytridiomycota genomes (23) but present in the well-annotated yeast (Saccharomyces cerevisiae) genome (SI Appendix, Table S9). Genes encoding enzymes of primary metabolism are retained in R. irregularis [(Kyoto Encyclopedia of Genes and Genomes; KEGG) Metabolic Pathways database; jgi.doe.gov/Rhizophagus], but several key genes are missing in the genome assemblies of both the obligate biotrophs R. irregularis and B. graminis (see below), suggesting that the lack of these genes is an evolutionary adaptation to the obligate biotrophy.

Like obligate biotrophic pathogens (20) and ectomycorrhizal symbionts (21, 24), R. irregularis has a decreased repertoire of genes involved in the degradation of plant cell wall (PCW) polysaccharides and in the biosynthesis of secondary metabolite toxins. None of the glycoside hydrolases (GHs) identified in R. irregularis are involved in degrading PCW polysaccharides. No gene encoding cellulohydrolases (GHs and GH7), polysaccharide lyases (PL1, PL3, PL4, and PL9), or proteins with cellulose-binding motif 1 (CBM1) were identified (Fig. 3 and SI Appendix, Table S10 and Fig. S15). Lytic polysaccharide mono-oxygenases (AA9,
formerly GH61) that are abundant in the ectomycorrhizal L. bicolor and T. melanosporum are missing from R. irregularis. Similarly, no genes involved in lignin decomposition, such as class II peroxidases (SI Appendix, Fig. S15), were found. No orthologs of bacterial genes coding for enzymes involved in symbiotic lipochitooligosaccharide factors (25) were identified.

Key enzymes that catalyze the biosynthesis of fungal toxins, such as polyketide synthases, modular nonribosomal peptide synthetases, terpene cyclases, and dimethylallyl diphosphate tryptophan synthases, are also lacking in R. irregularis (Fig. 3). Thus, it appears that biotrophy is associated with a convergent loss of secondary metabolic enzymes and PCW-degrading enzymes (20). No previously sequenced plant-interacting biotrophic fungus has such a minimal set of degrading enzymes, however. This finding strongly suggests an evolutionary adaptation to minimize the release of effector molecules that could trigger the plant immune system. The lack of PCW-degrading enzymes in the mycobiont also implies that penetration of the PCW, a prerequisite to the development of the intracellular symbiotic arbuscules, relies on plant enzymes. In addition, R. irregularis has no gene coding for the thiamine biosynthetic pathway (SI Appendix, Table S9). Interestingly, haustorial oomycetes (Albugo laibachii, Hyaloperonospora arabidopsidis, and Phytophthora spp.) also lack the gene for thiamine biosynthesis (26).

No secreted invertebrate or sucrose transporter was identified, implying that this fungus likely relies on the host plant to provide monosaccharides as a carbon source (27). It is unlikely that the aforementioned gene sets were missed because of incomplete genome coverage, given that the R. irregularis assembly encompasses 98% of conserved core eukaryotic single-copy genes (17). In contrast, genes coding for nitrate and nitrite reductases, nitrate transporter, and sulfite reductase that are missing from B. graminis (20) were found in R. irregularis (SI Appendix, Table S9). Thus, the obligate mycobiont R. irregularis retains the ability to take up and assimilate nutrients from its soil environment, a key issue for a soil-borne fungus providing its host plant with mineral nutrients (28).

Additional genes are missing from R. irregularis and are also lacking in the genomes of Mucoromycotina and Chytridiomycota (SI Appendix, Table S9), suggesting that this is a genomic feature of early diverging fungi. A comparison of the metabolic and cellular (KEGG) gene categories confirmed that R. irregularis is closer to the early diverging fungi, such as the Mucoromycotina, than to Dikarya (SI Appendix, Figs. S16 and S17).

We identified hundreds of fungal symbiosis-related genes by mapping RNA-Seq reads from germinating spores and Medicago-colonized roots (SI Appendix, section 1.7, Table S4). Of the 22,647 expressed genes, 1,068 (4.7%) were up-regulated in planta by at least twofold (false discovery rate-adjusted P value < 0.05). Most of the highly up-regulated genes code for proteins with no known function which are specific to R. irregularis (SI Appendix, Table S11), but several enzymes and membrane transporters are also induced during the interaction, as shown by eukaryotic orthologous groups (KOG) and InterPro (IPR) analyses of induced transcripts (SI Appendix, Table S12 and Fig. S18). Twenty-nine of the 50 most highly up-regulated fungal transcripts in mycorrhizal roots code for small secreted proteins with a predicted size of <150 aa (SI Appendix, Tables S11, S13, and S14), several of which were detected in laser microdissected arbusculated cells (SI Appendix, Table S11). These mycorrhiza-induced small secreted proteins (MiSSPs) belong to R. irregularis-specific orphan gene families (SI Appendix, Table S15) and may act as effector proteins to manipulate host cell signaling or to suppress defense pathways during infection, as has been shown for the R. irregularis SP7 effector, which interacts with a plant nuclear ethylene-responsive transcriptional factor (29). Although the repertoire of carbohydrate-active enzymes (CAZyme) of R. irregularis is limited, a few genes coding for CBM-containing proteins (LysM), glycosyl transferases (e.g., glycogen synthase, chitin synthase, UDP-glucosyltransferases), and GHSs (e.g., α-amylase, lysozyme, glucosaminidase) were dramatically up-regulated (SI Appendix, Table S16), suggesting a role in mycorrhiza metabolism.

Presence of Sex-Related Genes in R. irregularis. AMF have long been considered ancient clonal organisms, but recent studies have revealed the presence of many AMF homologs of genes known to be involved in sexual reproduction in other fungi, including meiosis-specific genes (7, 30) and an expanded gene family harboring a mating-type-related high mobility group domain (MAT-HMGs) (31). Here we confirm the previous findings of sex-related genes, and also show that the number of MATA-HMGs present in the genome of R. irregularis that harbor a MATA domain, or that share similarities with mating-type (MAT)-like HMGs from other fungi, is much larger than previously identified by mining transcriptome sequences (SI Appendix, section 1.6, Fig. S19 and Table S17). Specifically, 146 AMF genes were found to share similarities with homologs present in the mating type locus of various fungal lineages, including SexM/P from Phycomyces blakesleeanus (Zygomycota) (32). We confirm that none of the MATA-HMG genes are located in close proximity to orthologs of genes known to compose the MAT locus of other fungi, and that six scaffolds harbor MATA-HMGs that are repeated in tandem (31). We investigated the presence of potential idiomorphs in these MATA-HMGs by isolating their respective alleles from three genetically different strains of R. irregularis (strains A4, B3, and C2). Only 3 of the 146 homologs failed to amplify from one of three isolates using nonstringent procedures, raising the intriguing possibility that these represent idiomorphs of a heterothallic AMF mating-type locus.

In conclusion, the occurrence of multiple highly diverged genomes in the multinucleated R. irregularis is not supported by the present study. Ancient whole-genome duplication and TE proliferation likely have promoted massive gene duplications. The R. irregularis genome shares many features with fungi belonging to Mucoromycotina (e.g., homokaryotic organization in coenocytic hyphae, similar core metabolic pathways), suggesting that Glomeromycota have strong phylogenetic relationships with these early diverging fungi. On the other hand, the expression of effector-like MiSSPs and the lack of PCW-degrading and toxin-synthesizing enzymes are hallmarks of R. irregularis. These features suggest a functional converging evolution with phylogenetically unrelated biotrophic pathogens (20, 26) and ectomycorrhizal symbionts (21, 24). In contrast, the obligate biotrophic lifestyle of R. irregularis is not associated with a significant reduction in genes involved in nitrogen and sulfur assimilation, as observed in many obligate biotrophic leaf pathogens (20, 26), but is associated with the high expression of genes involved in nutrient uptake (8). Thus, R. irregularis has the dual ability to interact with the soil environment with respect to mineral nutrient uptake and to integrate the complex cues imposed by its in planta life. The present comprehensive repertoire of R. irregularis genes provides a basis for future research on symbiosis-related mechanisms and the ecological genomics of Glomeromycota.

Methods
Detailed descriptions of materials and methods are provided in SI Appendix, Methods. In brief, the genome of the multinucleated mycelium of R. irregularis DAOM-197198 was sequenced using Sanger, 454, Illumina, and PacBio platforms and assembled using the CLC Genomic Workbench assembler. Gene models were predicted and validated using computational tools, and RNA-Seq transcriptomics were annotated using the Joint Genome Institute annotation pipeline. Gene expression of R. irregularis and Medicago truncatula was assessed using RNA-Seq.

ACKNOWLEDGMENTS. We acknowledge Y. C. Li, H. Niculita-Herzel, and A. Brachman (from the former Joint Genome Institute Glomus consortium)
for their genome analyses that were not included in this study. We also thank the Lausanne University Genomic Technologies Facility, especially K. Harshman and E. Beauvoino, for PacBio sequencing support, and the Génome et Transcriptome-Plateforme Génomique (GeT-PlaGE) Facility of Toulouse, especially N. Marsaud and N. Ladouce, for Illumina sequencing support. The computations were performed at the Institut National de la Recherche Agronomique Nycogenomics facilities and in part at the Vital-IT Center for high-performance computing of the Swiss Institute of Bioinformatics. E.T. is supported by a postdoctoral fellowship from the European Commission (project EcoINDERS FP7-264465). This work was supported by the French National Research Agency through the Clusters of Excellence ARBRE (Advanced Research on the Biology of Tree and Forest Ecosystems) (ANR-11-LABX-0002-01) and TULIP (Toward a Uniform Theory of Biotic Interactions: Role of Environmental Perturbations) (ANR-10-LABX-41). This work was also funded by grants from the US Department of Energy’s Oak Ridge National Laboratory Scientific Focus Area for Genomics Foundation-Tional Sciences (to F.M. and G.A.T.); the Conseil Régional Midi-Pyrénées (to C.R.); the Natural Sciences and Engineering Research Council of Canada (to N.C.); the German Federal Ministry of Education and Research (to S.A.R.); the Swiss National Science Foundation (to I.R.S.); the Italian Regional Project Con-verging Technologies-BIOBIT (to P.B.); the Ministry of Education, Culture, Sports, Science, and Technology of Japan (to M.K.); and the Programme for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (to K.S.). The work conducted by the US Department of Energy’s Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231.