

The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont

E. Tisserant^{1*}, A. Kohler^{1*}, P. Dozolme-Seddas², R. Balestrini³, K. Benabdellah⁴, A. Colard^{5,6}, D. Croll^{5,6}, C. Da Silva⁷, S. K. Gomez⁸, R. Koul⁹, N. Ferrol⁴, V. Fiorilli³, D. Formey¹⁰, Ph. Franken¹¹, N. Helber¹², M. Hijri¹³, L. Lanfranco³, E. Lindquist¹⁴, Y. Liu², M. Malbreil¹⁰, E. Morin¹, J. Poulain⁷, H. Shapiro¹⁴, D. van Tuinen², A. Waschke¹¹, C. Azcón-Aguilar⁴, G. Bécard¹⁰, P. Bonfante³, M. J. Harrison⁸, H. Küster¹⁵, P. Lammers⁹, U. Paszkowski¹⁶, N. Requena¹², S. A. Rensing¹⁷, C. Roux¹⁰, I. R. Sanders⁵, Y. Shachar-Hill¹⁸, G. Tuskan¹⁹, J. P. W. Young²⁰, V. Gianinazzi-Pearson² and F. Martin¹

¹Institut National de la Recherche Agronomique (INRA), UMR 1136 INRA/University Henri Poincaré, Interactions Arbres/Micro-organismes, Centre de Nancy, 54280 Champenoux, France; ²UMR 1088 INRA/5184 CNRS/Burgundy University Plante-Microbe-Environnement, INRA-CMSE, BP 86510, 21065 Dijon, France; ³Istituto per la Protezione delle Pianta del CNR, sez. di Torino and Dipartimento di Biologia Vegetale, Università degli Studi di Torino, Viale Mattioli, 25, 10125 Torino, Italy; ⁴Departamento de Microbiología del Suelo y Sistemas Simbióticos, Estación Experimental del Zaidín, CSIC, C. Profesor Albareda, 1, 18008 Granada, Spain; ⁵Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland; ⁶ETH Zürich, Plant Pathology, Universitätsstrasse 3, CH-8092 Zürich, Switzerland; ⁷CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France; ⁸Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14853-1801, USA; ⁹Department of Chemistry and Biochemistry, New Mexico State University, Department 3MLS, PO Box 3001, Las Cruces, NM 88003-8001, USA; ¹⁰Université de Toulouse & CNRS, UPS, UMR 5546, Laboratoire de Recherche en Sciences Végétales, BP 42617, F-31326, Castanet-Tolosan, France; ¹¹Leibniz-Institute of Vegetable and Ornamental Crops, Department of Plant Nutrition, Theodor-Echtermeyer-Weg 1, D-14979 Grossbeeren, Germany; ¹²Karlsruhe Institute of Technology, Botanical Institute, Plant-Microbial Interaction, Hertzstrasse 16, D-76187 Karlsruhe, Germany; ¹³Institut de la Recherche en Biologie Végétale, Département de sciences biologiques, Université de Montréal, 4101 Rue Sherbrooke est, Montréal, Que., Canada H1X 2B2; ¹⁴Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA; ¹⁵Institut für Pflanzengenetik, Naturwissenschaftliche Fakultät, Leibniz Universität Hannover, D-30419 Hannover, Germany; ¹⁶Department de Biologie Moléculaire Végétale, Université de Lausanne, Biophore, 4419, CH-1015 Lausanne, Switzerland; ¹⁷BIOSS Centre for Biological Signalling Studies, Freiburg Initiative for Systems Biology and Faculty of Biology, University of Freiburg, Hauptstr. 1, D-79104 Freiburg, Germany; ¹⁸Department of Plant Biology, Michigan State University, East Lansing, MI 48824-1312, USA; ¹⁹Oak Ridge National Laboratory, BioSciences, PO Box 2008, Oak Ridge, TN 37831, USA; ²⁰Department of Biology, University of York, York YO10 5DD, UK

Summary

Author for correspondence:

Francis Martin

Tel: +33 383 39 40 80

Email: fmartin@nancy.inra.fr

Received: 31 August 2011

Accepted: 26 September 2011

New Phytologist (2011)

doi: 10.1111/j.1469-8137.2011.03948.x

Key words: Glomeromycota, *Glomus*, meiosis and recombination genes, mycorrhiza, small secreted proteins, symbiosis, transcript profiling.

- The arbuscular mycorrhizal symbiosis is arguably the most ecologically important eukaryotic symbiosis, yet it is poorly understood at the molecular level. To provide novel insights into the molecular basis of symbiosis-associated traits, we report the first genome-wide analysis of the transcriptome from *Glomus intraradices* DAOM 197198.
- We generated a set of 25 906 nonredundant virtual transcripts (NRVTs) transcribed in germinated spores, extraradical mycelium and symbiotic roots using Sanger and 454 sequencing. NRVTs were used to construct an oligoarray for investigating gene expression.
- We identified transcripts coding for the meiotic recombination machinery, as well as meiosis-specific proteins, suggesting that the lack of a known sexual cycle in *G. intraradices* is not a result of major deletions of genes essential for sexual reproduction and meiosis. Induced expression of genes encoding membrane transporters and small secreted proteins in intraradical mycelium, together with the lack of expression of hydrolytic enzymes acting on plant cell wall polysaccharides, are all features of *G. intraradices* that are shared with ectomycorrhizal symbionts and obligate biotrophic pathogens.
- Our results illuminate the genetic basis of symbiosis-related traits of the most ancient lineage of plant biotrophs, advancing future research on these agriculturally and ecologically important symbionts.

*These authors contributed equally to this work.

Introduction

The arbuscular mycorrhizal (AM) symbiosis between fungi in the Glomeromycota (Schüssler *et al.*, 2001) and plants involves over two-thirds of all known plant species, including important crop species, such as wheat (*Triticum aestivum*), rice (*Oryza sativa*), maize (*Zea mays*), soybean (*Glycine max*) and poplar (*Populus* spp.). This mutualistic symbiosis, involving one of the oldest fungal lineages, is arguably the most ecologically and agriculturally important symbiosis in terrestrial ecosystems (Fitter *et al.*, 2011). The extraradical mycelium (ERM) of the symbiont acts as an extension of the root system and increases the uptake of key nutrients, particularly phosphorus (P) and zinc (Zn) and possibly also nitrogen (N) (Smith & Smith, 2011). These fungi are therefore crucial to plant growth (Smith & Read, 2008) and also define the diversity of plants in ecosystems (van der Heijden *et al.*, 1998). Furthermore, because the colonization of plants by AM fungi can also result in a 20% net increase in photosynthesis (Smith & Read, 2008), these universal mycosymbionts make a very large, poorly understood contribution to the global carbon cycling budget of ecosystems.

The Glomeromycota are unique in that their spores and coenocytic hyphae contain multiple nuclei in a common cytoplasm. No sexual cycle is known, although anastomosis and nuclear movement between hyphae of the same species have been described (Giovannetti *et al.*, 2001), as well as genetic exchange between AM fungal individuals (Croll *et al.*, 2008, 2009). Furthermore, the concept of an individual is unclear as nuclei within a single AM fungus appear to be genetically different in some species, which raises substantial questions about the natural selection, population genetics and gene expression of these highly unusual organisms (Jany & Pawlowska, 2010; Sanders & Croll, 2010). AM fungi are thought to grow clonally, but controversy clouds their ploidy, their genome size, and whether or not they can reproduce sexually (for a review, see Sanders & Croll, 2010). Evidence for recombination has been provided, but whether mating and meiosis are involved is unknown (Croll & Sanders, 2009). There is therefore a need to identify genes whose products are required for proper completion of meiotic recombination.

Root colonization by AM fungi follows a series of distinct steps (for a review, see Bonfante & Genre, 2010) starting with a pre-symbiotic molecular dialogue that involves root-released strigolactones (Akiyama *et al.*, 2005) and AM fungal signaling molecules, such as lipochitoooligosaccharides (Maillet *et al.*, 2011). AM fungi are obligate mutualistic symbionts that can only grow for a limited time without colonizing a susceptible host root (Bécard *et al.*, 2004). This has been suggested to be a consequence of some nutritional deficiencies and loss of metabolic pathways, such as an absence of *de novo* fatty acid synthesis (Bago *et al.*, 2000), during the asymbiotic and presymbiotic phases related to putative genome erosion (Ercolin & Reinhardt, 2011). After spore germination, the first contact at the root surface is marked by the differentiation of the fungal hyphae into an appressorium from which a penetration hypha invades the root epidermis (Genre *et al.*, 2005). After the fungus has traversed the outer cell layers it spreads in the inner cortex, that is, intraradical

mycelium (IRM), and forms highly branched structures inside cortical cells, so-called arbuscules. Arbuscules are the main site of nutrient transfer from the fungus to the plant (Javot *et al.*, 2007; Bonfante & Genre, 2010; Smith & Smith, 2011). Arising from the colonized roots, the ERM proliferates in the growth medium, where it absorbs and assimilates the available nutrients before their transfer to the host plant. Major transcriptome shifts are thus expected during these different key developmental stages, but little is known of the repertoire of effector-like proteins, membrane transporters and assimilative enzymes involved in these different steps of symbiosis development and functioning.

The ability to establish a sophisticated zone of interaction, such as the haustorium in pathogenic oomycetes and fungi interacting with plants, requires sophisticated host defense suppression (Dodds & Rathjen, 2010), which is predominantly achieved via secreted proteins delivered into the host cell (Kamoun, 2007). The protein SP7 of *Glomus intraradices* is secreted and transferred to the plant cell nucleus in colonized roots of *Medicago truncatula*, where it binds to the host ethylene-responsive transcriptional factor regulating the expression of several defense-related genes (Kloppholz *et al.*, 2011). It remains to be determined whether *G. intraradices* expresses genes coding for additional secreted effector-like proteins during its interaction with the host plant.

The mycorrhizal colonization leads to quantitative and qualitative changes in the host transcriptome. Plant genes that are specifically regulated during the establishment and/or functioning of the AM symbiosis have been identified by both targeted and high-throughput expression profiling in several sequenced model and crop plants (Güimil *et al.*, 2005; Balestrini & Lanfranco, 2006; Gomez *et al.*, 2009; Grunwald *et al.*, 2009; Guether *et al.*, 2009a). In contrast, regulation of gene expression in the AM fungal symbionts caused by interactions with roots has so far been poorly explored, and studies have mainly been based on targeted approaches providing only narrow insights into fungal adaptation to the symbiotic mode (Gomez *et al.*, 2009; Seddas *et al.*, 2009; Tani *et al.*, 2009; Kuznetsova *et al.*, 2010). Most studies have focused on P transport and metabolism (Viereck *et al.*, 2004; Benedetto *et al.*, 2005; Balestrini *et al.*, 2007; Javot *et al.*, 2007; Gómez-Ariza *et al.*, 2009; Grunwald *et al.*, 2009) and primary carbon metabolism (Bago *et al.*, 2000, 2003; Schaarschmidt *et al.*, 2006). Further studies adopting a microarray or high-throughput sequencing approach are definitely warranted, because they could provide a wider understanding of how genetic changes in the plant affect overall patterns of fungal gene expression, as well as the impact of the fungal genotypes on symbiosis fitness and plant growth (Angelard *et al.*, 2010).

Biotrophs are widely accepted to intimately interact and co-evolve with their hosts. The genetic changes that brought about the evolution of obligate biotrophy in the Glomeromycota are, however, unknown. Recent research on obligate biotrophic pathogens, such as the powdery mildew fungus *Blumeria graminis* (Spanu *et al.*, 2010), the downy mildew oomycete *Hyaloperonospora arabidopsidis* (Baxter *et al.*, 2010), the oomycetous white rust *Albugo laibachii* (Kemen *et al.*, 2011),

and the poplar and stem rusts *Melampsora larici-populina* and *Puccinia graminis*, respectively (Duplessis *et al.*, 2011), reveals a close correlation between the biotrophic life style and gene losses in primary and secondary metabolism. In the oomycetes, all haustorium-forming species have lost the thiamine biosynthetic pathway, suggesting that haustorial oomycetes obtain thiamine from the host (Kemen *et al.*, 2011). It appears that biotrophy in these obligate plant pathogens is also associated with a convergent loss of secondary metabolic enzymes and an extremely reduced set of carbohydrate-active enzymes devoted to plant cell wall depolymerization. Genes for nitrate and nitrite reductases, a nitrate transporter, and sulfite reductase are also often missing. This streamlining of metabolism presumably reflects adaptation to life within host cells, as it is not observed in nonobligate biotrophic pathogens. Whether similar events have also characterized AM fungal evolution remains to be determined.

As mentioned above, transcriptome studies have been pursued on AM roots in multiple plant species under a variety of experimental conditions, but they have mainly focused on the host plant transcriptome. No single study has yet brought together genome-wide transcriptomic data for the fungal component of this important symbiosis. The aim of this study was to establish a comprehensive, genome-wide inventory of gene expression in *G. intraradices* DAOM197198 by sequencing cDNA libraries from different fungal structures (germinated spores, extra- and intraradical mycelium, and arbuscules). Through transcriptomic analyses, we also wished to identify which of these genes were induced upon symbiosis and, further, to assign the differentially expressed transcripts to key symbiotic mechanisms, such as nutrient transport and assimilation, host colonization, and signaling pathways. Several factors have led to the choice of *G. intraradices* for the first large-scale transcriptome sequencing of an AM fungus. As a symbiont, *G. intraradices* is highly effective in mobilizing, taking up and transferring mineral nutrients, such as inorganic orthophosphate ions (Pi), N and sulfur (S), from soils to plants (Govindarajulu *et al.*, 2005; Allen & Shachar-Hill, 2009; Tian *et al.*, 2010; Smith & Smith, 2011) and it readily colonizes many plants, including agriculturally important crop species (e.g. alfalfa (*Medicago truncatula*), poplar, rice and wheat) as well as model plants such as *M. truncatula* and *Lotus japonicus*. *Glomus intraradices* is one of the most studied AM fungi as it rapidly colonizes its host plants, and it is a model species for dissecting the molecular and cellular biology of the Glomeromycota (Seddas *et al.*, 2009; Sanders & Croll, 2010). It is readily amenable to *in vitro* culture on transformed host roots (Chabot *et al.*, 1992) and is the only species for which spores are available commercially in pure form in large quantities.

The comparative analyses of gene repertoires in *G. intraradices*, pathogenic obligate biotrophs and ectomycorrhizal fungi offer insights into genes that may be involved in obligate biotrophy and mycorrhizal symbioses. In the absence of a whole-genome sequence for any member of the phylum Glomeromycota (Martin *et al.*, 2008a), the availability of large-scale expressed sequenced tag (EST) collections represents the core foundation for understanding genome functionality in the Glomeromycota.

Materials and Methods

Biological material

Glomus intraradices Schenck & Smith DAOM 197198 (recently reassigned to *G. irregulare* and then *Rhizophagus irregularis* (Błaszk., Wubet, Renker & Buscot) C. Walker & A. Schüßler comb. nov.; see Stockinger *et al.*, 2009) was produced in monoxenic cultures maintained on *Agrobacterium rhizogenes*-transformed carrot (*Daucus carota*; clone DC2) roots (Bécard & Fortin, 1988). Germinated spores were produced as described by Chabot *et al.* (1992). Biological materials used for cDNA library construction and microarray transcript profiling are summarized in Table 1. Protocols for producing the biological materials, RNA purification and cDNAs are described in the online Supporting Information Methods S1. For array profiling, *G. intraradices* ERM was grown on liquid M medium without sucrose for 3 wk before harvesting. *Glomus intraradices*- and mock-inoculated roots of *M. truncatula* were prepared as described for the cDNA library AKNA (Methods S1). Germinated spores were sampled as described for the cDNA library AKND (Methods S1) (Balestrini *et al.*, 2007). Samples were snap-frozen in liquid N₂.

Table 1 *Glomus intraradices* tissues used in the extraction of transcripts for cDNA library construction and oligoarray profiling

Samples	Sanger sequencing	454 sequencing	Array profiling
Germinated spores	CCHU		Three replicates
Germinated spores mix = spores + spores in GR24 + spores in API		EXTA	
Germinated spores in GR24		AKNC	
Germinated spores in API		AKNB	
<i>Daucus carota</i> ERM	CACE	EXTB	Three replicates
<i>Medicago truncatula</i> ERM			
<i>M. truncatula</i> IRM		AKNA*	Three replicates
LMD-microdissected arbuscule-containing cells		AKND	

GR24, strigolactone GR24; API, apigenin; ERM, extraradical mycelium; IRM, intraradical mycelium; AKNA, AKNB, AKNC, AKND, CACE, CCHU, EXTA and EXTB stand for cDNA library IDs. AKNA, *M. truncatula* IRM; AKNB, germinated spores incubated in apigenin for 24 h; AKNC, germinated spores incubated in the strigolactone GR24 for 24 h; AKND, laser-microdissected *M. truncatula* arbuscled cells; CACE, *M. truncatula* ERM; CCHU, germinated spores; EXTA, germinated spore mix; EXTB, *M. truncatula* ERM; LMD, laser microdissection system.

*The cDNA library AKNA has been constructed using mRNA from a mix of the three replicates also used for microarray transcript profiling.

EST sequencing, filtering and assembly

Sanger sequencing was performed on ABI3730xl analyzers (Applied Biosystems, Nutley, NJ, USA) and pyrosequencing on the Genome Sequencer FLX System (454 Life Sciences/Roche, Nutley, NJ, USA). Reads (797 394) were filtered and trimmed for low quality, low complexity, and adaptor sequences using SeqClean (TIGR (The Institute for Genomic Research); <http://sourceforge.net/projects/seqclean/>) (Supporting Information Fig. S1). Sequences ≤ 100 bp were discarded. The resulting high-quality sequences were screened to detect potential bacterial and fungal contaminant sequences, and plant sequences present in cDNAs from symbiotic tissues. Sequences having a GC% higher than that of *G. intraradices* (i.e. $\geq 45\%$) were removed (Fig. S2). Sequences with a high nucleotide sequence similarity (BLASTN e -value $\leq 1e-5$, score ≥ 150 , % identity ≥ 95) with identified contaminants, *Aspergillus* spp. and *Chromobacterium violaceum*, as well as the *M. truncatula* genome and transcripts, and other plant DNA (e.g. carrot) were also removed (Fig. S1). Finally, to avoid the loss of *G. intraradices* outlier sequences, removed sequences were aligned to the *G. intraradices* draft genome using BLASTN (e -value $\leq 1e-5$, score ≥ 150) at the INRA GlomusDB website (<http://mycor.nancy.inra.fr/IMG/ClomusGenome/index3.html>) and retained when sequence similarity was $\geq 95\%$. Filtered reads were then assembled using the MIRA assembler (Chevreux *et al.*, 2004). To choose the best assembly, several assemblies were constructed using different parameters and several indexes were measured, including the number of nonredundant virtual transcripts (NRVTs), the percentage of assembled reads and the percentage of NRVTs with hits against gene models from the taxonomically related, although distant, Mucoromycotina *Rhizopus oryzae* (<http://genome.jgi-psf.org/Rhior3/Rhior3.home.html>) (Supporting Information Table S1). The assembly generated by MIRASEARCHESTSNPs was selected. The number of detected NRVTs increased with the number of reads and the rarefaction curve did not reach a plateau (Fig. S3a). However, many of these sequences may be the result of DNA pyrosequencing errors creating false transcripts. This is an acute problem for the AT-rich sequences of *G. intraradices*. In contrast, the number of protein sequences documented in the Swiss-Prot DNA database (<http://www.uniprot.org/>) and *R. oryzae* gene repertoire showing a similarity to *G. intraradices* NRVTs plateaued at approx. 5500 homologs (Fig. S3b), suggesting that most of the transcriptome of *G. intraradices* was covered. This was confirmed by a TBLASTN search (cut-off e -value of $1e-5$) using the protein sequences of the core eukaryotic genes (CEGs) (Parra *et al.*, 2009); 245 (98.7%) of the 248 CEG proteins were found in the current *G. intraradices* transcriptome. The missing protein sequences were a metal-binding protein, a spindle assembly checkpoint protein and a monooxygenase involved in coenzyme Q (ubiquinone) biosynthesis. Eighty-six per cent of NRVTs matched the current *G. intraradices* 52.5 Mb-genome assembly (Martin *et al.*, 2008a), confirming that most filtered, high-quality ESTs were from *G. intraradices*.

Single-nucleotide polymorphisms (SNPs) in NRVTs were identified with MIRASEARCHESTSNPs using the following

parameters: minimum reads, 2 (Sanger) or 4 (454); minimum quality value, 30 (Sanger) or 25 (454); minimum neighbor quality value, 20.

For the digital northern gene expression, the number of reads for each NRVT in each library was counted and the relative frequency (reads of a given NRVT divided by the total number of reads) was obtained. Significant differences in gene expression between libraries for each NRVT were calculated (Audic & Claverie, 1997).

EST sequences are available at the National Center for Biotechnology Information (NCBI) (accession numbers GW091323–GW125581 and GW086621–GW090678). A MySQL/PHP database compiling NRVT sequences and annotations is available at the INRA GlomusDB website. This server can also be used to query the ESTs and the draft genome assembly using BLAST programs.

Functional annotation

NRVTs were compared against Swiss-Prot using BLASTX with a significance threshold (e -value $< 1e-5$). Gene annotations were assigned to each NRVT based on the best BLAST hits. Sequences were searched against the *G. intraradices* draft genome contigs (version 'test14') at the INRA GlomusDB using BLASTN (e -value $\leq 1e-5$). NRVTs were also compared with proteins of Basidiomycota (*Cryptococcus neoformans*, *Laccaria bicolor*, *M. larici-populina*, *Phanerochaete chrysosporium* and *Ustilago maydis*), Ascomycota (*Aspergillus nidulans*, *B. graminis*, *Botrytis cinerea*, *Magnaporthe grisea*, *Neurospora crassa* and *Tuber melanosporum*) and Mucoromycotina (*R. oryzae* and *Mucor circinelloides*) using BLASTX (e -value $\leq 1e-5$). To identify conserved protein domains in NRVTs, predicted protein sequences were compared with the Eukaryotic Orthologous Groups (KOG) database (Tatusov *et al.*, 2003). The counts of each KOG domain by species were transformed into a z -scores matrix to center the data, and visualization was performed using MeV (MultiExperiment Viewer) (Saeed *et al.*, 2006). The gene ontology (GO) terms (Ashburner *et al.*, 2000) were assigned to each sequence using BLAST2GO (Conesa *et al.*, 2005). Enrichment analysis of GO annotations was carried out with BINGO (<http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html>) using the hypergeometric test and Benjamini and Hochberg's false discovery rate (FDR) multiple testing correction ($P \leq 0.05$) (Maere *et al.*, 2005).

To predict *G. intraradices* metabolic pathways, sequences were queried for Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) assignments (Kanehisa & Goto, 2000) using the KEGG Automatic Annotation Server (KAAS) (Moriya *et al.*, 2007) and results were then plotted into the KEGG Global metabolic atlas (map01100) using iPATH (Letunic *et al.*, 2008). A high-resolution version of this metabolic pathway map can be downloaded at: http://mycor.nancy.inra.fr/IMG/ClomusGenome/download/Gi_metabolic_map.pdf.

To identify sequences encoding secreted proteins, NRVTs were translated into protein using FrameFinder (<http://www.ebi.ac.uk/~guy/estate/>). *In silico* predictions of secreted proteins

were then carried out using SIGNALP 3.0 (Bendtsen *et al.*, 2004), TARGETP 1.1 (Emanuelsson *et al.*, 2000) and TMHMM 2.0 (prediction of transmembrane helices in proteins) (Krogh *et al.*, 2001) as previously reported in Duplessis *et al.* (2011). Small secreted proteins (SSPs) were selected based on an arbitrary cut-off of 300 amino acids (Duplessis *et al.*, 2011).

Oligoarray design

For the design of custom oligoarrays, a MIRA assembly (v. 1.0) comprising 22 254 NRVTs was constructed using cDNA libraries CACE, CCHU, EXTA and EXTB. At the time, sequences from the AKNA, AKNB, AKNC and AKND libraries were not available. To ensure that all sequences were sufficiently long to allow the design of high-quality probes, we selected the subset of sequences ≥ 200 bp in length ($n = 19\,465$). To remove redundancy among sequences that would produce cross-hybridization, the assembled NRVTs were clustered based on nucleotide sequence similarity using BLASTClust (<http://www.ncbi.nlm.nih.gov/Web/News/str/Spring04/blastlab.html>). Within each cluster, sequences were ranked by length, and the longer member of each cluster was selected ($n = 15\,932$). We also selected singleton sequences with matches in NCBI databases and with length ≥ 200 bp ($n = 2046$). We then added 2327 NRVTs generated by the Paracel Transcript Assembler (Striking Development; <http://www.paracel.com/>) (S. Rensing, unpublished results) which are not found in the MIRA NRVT set. We also included 12 *G. intraradices* reference sequences downloaded from the NCBI, 23 *G. intraradices* mitochondrial sequences (C. Roux, unpublished results) and 70 *G. intraradices* sequences generated from the manual annotation of the *G. intraradices* genome and not found in the NRVTs. Finally, 4471 remaining singleton Sanger sequences without matches in public databases were added. For 2477 sequences of this set, no probe could be designed, probably because of their relative high A + T content, which resulted in a higher-than-usual prevalence of homopolymer runs and other forms of low-complexity sequences. This procedure produced a total set of 22 404 sequences containing 14 828 MIRA NRVTs, 1976 PARACEL NRVTs, 5496 singletons and 104 additional sequences. This set did not include NRVTs from symbiotic tissues not available when the custom array was constructed. However, *c.* 80% of the NRVTs used to construct the oligoarray were found by BLASTN in the 25 906 MIRASEARCHESTSNPs NRVTs.

The *G. intraradices* expression array ($4 \times 72\text{K}$) manufactured by Roche NimbleGen Systems Limited (Madison, WI, USA) (<http://www.nimblegen.com/products/exp/index.html>) contained three independent, nonidentical, 60-mer probes per sequence. Included in the oligoarray were *G. intraradices* sequences, and 5785 random 60-mer control probes.

Transcript profiling

Whereas biological samples used for cDNA sequencing were produced by several different collaborators, biological materials used for microarray transcript profiling were generated in triplicate

in a small number of laboratories: spores at the Laboratoire de Recherche en Sciences Végétales (Université de Toulouse, France), *M. truncatula* ERM at the Departamento de Microbiología del Suelo y Sistemas Simbióticos (Estación Experimental del Zaidín, CSIC, Granada, Spain) and *M. truncatula* IRM at the Boyce Thompson Institute for Plant Research (Tower Road, Ithaca, NY, USA). cDNA synthesis was carried out at INRA-Nancy.

RNA was extracted from mycorrhizal roots with Trizol reagent (Invitrogen Corporation, Carlsbad, CA, USA) with additional phenol : chloroform (1 : 1, v/v) purification steps, from ERM and spores using the RNeasy Plant Mini Kit (Qiagen) with RLT lysis buffer. RNA from mycorrhizal roots was treated with Turbo DNase I (Ambion Inc., Austin, TX, USA) and purified with the RNeasy MinElute CleanUp kit (Qiagen). DNase treatment of RNA from ERM was performed using RNase-free DNase (Qiagen). RNA quality was checked before cDNA synthesis using the Bio-Rad Experion analyzer. RNA preparations (three biological replicates) were amplified using the SMART PCR cDNA Synthesis Kit (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France) according to the manufacturer's instructions. Single dye labeling of cDNA samples, hybridization procedures and data acquisition were performed at the NimbleGen facilities (NimbleGen Systems, Reykjavik, Iceland) following their standard protocol. As ERM and IRM materials were not produced in strictly similar conditions, transcript profiles of ERM and IRM should be compared with caution.

Average expression levels were calculated for each gene from the independent probes on the array and were used for further analysis. Raw array data were normalized by the robust multi-array average (RMA) routine using the ARRAYSTAR software (Dnastar Inc., Madison, WI, USA). A transcript was deemed expressed when its signal intensity was three-fold higher than the mean signal-to-noise threshold (cut-off value) of the random oligonucleotide probes present on the array (50–100 arbitrary units). The maximum signal intensity values for the most abundant transcripts were *c.* 65 000 arbitrary units. A control hybridization was performed with cDNA from uninfected *M. truncatula* roots to evaluate any plant RNA cross-hybridization with *G. intraradices* sequences and to detect any plant-derived probes on the array. The later probe signals were discarded. A Student's *t*-test with FDR correction (Benjamini–Hochberg) was applied to the data using ARRAYSTAR. Transcripts with a significant *P*-value (< 0.05) and \geq five-fold change in transcript level were considered to be differentially expressed. The complete expression data set is available as series (accession number GSE29866) at the Gene Expression Omnibus at NCBI (<http://www.ncbi.nlm.nih.gov/geo/>).

Validation of oligoarray data

To obtain the biological material for the PCR validation of oligoarray data, 6000 arbuscule-containing cells from *M. truncatula* roots (2000 cells for each of three biological replicates) were microdissected using a laser microdissection system (LMD) according to Balestrini *et al.* (2007). ERM from monoxenic *M. truncatula* cultures and *c.* 6000 germinated spores were

sampled, snap-frozen in liquid N₂ and stored at -80°C until used. Total RNA was extracted using the RNeasy Microarray Tissue Mini Kit (Qiagen, Courtaboeuf, France) following the manufacturer's instructions. All reverse transcriptase (RT)-PCR assays were carried out using the One Step RT-PCR kit (Qiagen). DNA contamination in RNA samples was evaluated using the *G. intraradices* elongation factor *GintEF1 α* gene specific primers for *G. intraradices*. To determine the amount of extracted RNA, a semiquantitative RT-PCR using *GintEF1 α* specific primers was performed. Reactions were carried out in a final volume of 25 μ l. Amplification reactions were run for 40 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s, and an aliquot of the PCR reaction was taken after 35, 37 and 40 cycles.

Results and Discussion

An obligate biotroph with a large gene repertoire

The cDNA libraries of *G. intraradices* listed in Table 1 were sequenced to assess the size and diversity of the fungus gene repertoire and explore the responses of the AM fungus at key developmental stages, germinated spores, ERM and IRM. After sequencing, the *de novo* hybrid clustering of the 437 411 filtered reads produced 25 906 NRVTs (Table S1); 87% of the reads were in the assembly corresponding to 20.3 Mbp of nonredundant sequences. Only 10 823 NRVTs (41.8%) matched database sequences (cut-off *e*-value of 1e-5) in BLAST searches against genes from fungi other than Glomeromycota (Fig. S4). Homologous NRVTs corresponded to 8773 unique known genes in the DNA databases. Among the predicted peptide sequences, 9035 and 9265 showed significant sequence similarity to proteins from *R. oryzae* and *M. circinelloides*, respectively, distantly related fungi in the order Mucorales (Fig. S4). NRVTs that did not match any of the known genes still gave highly relevant hits against the *G. intraradices* draft genome, suggesting that these represent novel orphan genes. As many *G. intraradices* transcripts are represented by two NRVTs (i.e. aligning to the 5' or 3' end of the transcript) with an NRVT:gene ratio of 1.4, the number of nonredundant expressed transcripts in *G. intraradices* was *c.* 18 500 (25 906/1.4).

Based on BLASTClust analysis, the number of NRVTs found in multigene families was low, with 120 (90% sequence identity; 80% sequence coverage) to 500 (90% sequence identity; 50% sequence coverage) NRVTs, most of the gene families having only two members.

Although the vast majority of genes encoding enzymes of primary metabolism are retained in obligate biotrophic pathogens (downy and powdery mildews, white rust, and poplar and stem rusts), notable exceptions include anaerobic fermentation, biosynthesis of glycerol from glycolytic intermediates, biosynthesis of thiamine, and nitrate and sulfate assimilation (Baxter *et al.*, 2010; Spanu *et al.*, 2010; Duplessis *et al.*, 2011; Kemen *et al.*, 2011). Out of the 25 906 *G. intraradices* NRVTs, 5296 (20.4%) had significant matches in the KEGG pathway database and were assigned to 175 KEGG pathways. Most KEGG enzymes were mapped to NRVTs (see the metabolic pathway map online at

<http://mycor.nancy.inra.fr/IMGC/GlomusGenome/download3.php?select=anno>), indicating the occurrence of these active metabolic processes in *G. intraradices* mycelium. In contrast to oomycete and fungal biotrophic pathogens, transcripts coding for nitrate and nitrite reductases, nitrate transporter, and sulfite reductase were highly expressed in *G. intraradices*. However, *G. intraradices* has lost the thiamine biosynthetic pathway (Table S2), as have most haustorium-forming species (Kemen *et al.*, 2011). We infer that *G. intraradices* obtains thiamine from the host. There is no invertase transcript in the *G. intraradices* transcriptome (Table S2), implying that this fungus depends on the plant to both provide and hydrolyse sucrose, with the glucose moiety then preferentially transferred to the mycobiont. This is consistent with earlier observations (Schaarschmidt *et al.*, 2006). Additional genes missing from biotrophic pathogens (Spanu *et al.*, 2010), such as those coding for the allantoin permease DAL4p, the uracil permease FUR4p, and several enzymes involved in aromatic amino acid metabolism and detoxification, were also missing from *G. intraradices* (Table S2). Although some metabolic pathways are missing in *G. intraradices*, we hypothesize that evolution to biotrophy in *G. intraradices* was not initiated by massive loss of metabolic complexity as observed in obligate biotrophic pathogens (Spanu *et al.*, 2010). The mycobiont retains the ability to take up and assimilate nutrients from its soil environment.

Protein domain distributions

About half (5056) of the 10 823 peptide sequences from *G. intraradices* with sequence similarity to documented proteins in databases were shared with all other fungal species and can thus be considered ubiquitous among fungi. The comparison of protein sets of *G. intraradices* with those of *R. oryzae* and *M. circinelloides* showed that 1076 proteins were unique to these basal lineages of aseptate, coenocytic fungi. Interestingly, a small set of *G. intraradices* protein sequences were only shared with one or other of the sequenced mycorrhizal symbionts, the ascomycete *T. melanosporum* (49 NRVTs) or the basidiomycete *L. bicolor* (107 NRVTs). They mainly code for hypothetical proteins with no known function, although a few predicted proteins belong to signaling pathways. A total of 9321 NRVTs contained at least a part of a conserved protein domain in the KOG database. Compared with other fungal gene repertoires, *G. intraradices* showed an overrepresentation of proteins involved in signaling pathways and ubiquitin-related metabolism (Fig. 1). The expansion of the tyrosine kinase-encoding gene family involved in signaling pathways is also a feature of the ectomycorrhizal *L. bicolor* genome (Fig. 1; Martin *et al.*, 2008b).

Sequence polymorphism of transcripts

There were a total of 43 872 SNPs in 3963 NRVTs (15.3% of the total number of NRVTs; 2.1 SNP/kb); 1102 NRVTs contained < 5 SNPs (Fig. 2a). When this SNP analysis was conducted on NRVT regions having a similarity with known protein coding sequences, 846 polymorphic NRVTs (8%; 1.3

○ ■ ● ○ ■ ● ■ ■ ● ■ ■ ● ○ ■
Gi Mc Ro Lb Cn Pc Um Ml An Bc Mg Nc Tm Bg



Fig. 1 Most abundant Eukaryotic Orthologous Groups (KOG) domains in *Glomus intraradices* compared with representative saprotrophic (black circle), pathogenic (black square) and mutualistic fungi (open circle). The heatmap is based on the relative z-score of KOG conserved domains. The top 55 KOG domains found in *G. intraradices* were selected. Colors indicate abundance: from dark green (highly abundant) to white (weakly abundant). An, *Aspergillus nidulans*; Bc, *Botrytis cinerea*; Bg, *Blumeria graminis*; Cn, *Cryptococcus neoformans*; Lb, *Laccaria bicolor*; Ml, *Melampsora larici-populina*; Mg, *Magnaporthe grisea*; Mc, *Mucor circinelloides*; Nc, *Neurospora crassa*; Pc, *Phanerochaete chrysosporium*; Ro, *Rhizopus oryzae*; Tm, *Tuber melanosporum*; and Um, *Ustilago maydis*.

SNP/kb) were identified among 10 823 homologs. Most polymorphic NRVTs contained < 5 SNPs (Fig. 2b). The NRVTs with known function having the highest SNPs were a 60S ribosomal protein L17 (15 SNPs) and a Ras-related Rab (11 SNPs).

The presence of multiple SNPs in hundreds of NRVTs confirmed the within-isolate DNA sequence polymorphism that has repeatedly been reported for a number of genes in *G. intraradices* (Croll *et al.*, 2008; Sanders & Croll, 2010). The transcriptome

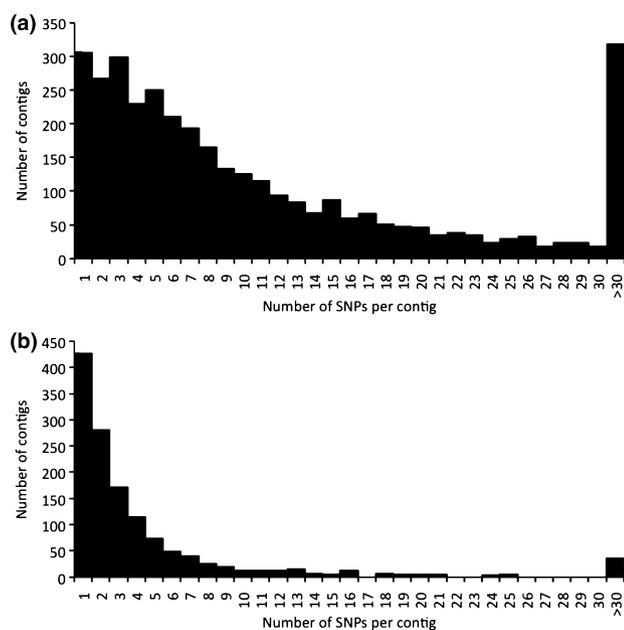


Fig. 2 Single nucleotide polymorphisms (SNPs) in nonredundant virtual transcript (NRVT) sequences from *Glomus intraradices*. (a) SNPs were identified by the MIRASEARCHESTSNPs program by aligning the complete nucleotide sequences from NRVTs, including 5' and 3' untranslated regions (UTRs). (b) SNPs were calculated in the coding regions of NRVTs having a homolog documented in the Swiss-Prot database.

data show that this polymorphism is widespread in the genome, and that > 1 variant of each polymorphic gene is transcriptionally active. This implies that the sequence variation may be of functional importance.

Symbiosis induces alterations in the *G. intraradices* transcriptome

Little is known about the impact of symbiosis on the mycobiont transcriptome and molecular factors driving developmental pathways (Güimil *et al.*, 2005; Gomez *et al.*, 2009; Seddas *et al.*, 2009; Kuznetsova *et al.*, 2010; Sanders & Croll, 2010). Our transcript profiling thus provides the first large-scale discovery of fungal symbiosis-related genes. Of the 18 751 coding sequences detected by oligoarrays, 395 (2.1%) and 569 (3.0%) were up- and down-regulated (\geq five-fold; P -value \leq 0.05), respectively, in IRM in comparison to germinated spores, whereas 202 (1.1%) and 74 (0.4%) were up- and down-regulated, respectively, in ERM in comparison to germinated spores. Highly up-regulated genes in IRM and ERM are shown in Tables 2, 3, S3 and S4. In this study, mRNA concentrations were used as a proxy of protein concentrations. Gene expression is, however, a multistep process that involves the transcription, translation and turnover of messenger RNAs and proteins, and transcript levels are not always related to protein levels and enzyme activities (Schwanhäusser *et al.*, 2011). It remains to be determined whether the observed changes in transcripts lead to alteration of the proteome, enzyme activities and/or metabolic fluxes. In addition, regulatory enzymes exhibiting allosteric properties are likely to be effective agents for fine tuning of metabolic fluxes in symbiotic tissues.

Table 2 Most highly up-regulated *Glomus intraradices* transcripts (nonredundant virtual transcripts (NRVTs)) in the intraradical mycelium (IRM) compared with germinated spores identified using expression oligoarrays and ranked by decreasing fold changes

NRVT ID	Ratio IRM:spore	IRM expression level	Spore expression level	Putative function
EXTB106676.b0	36 516	38 488	1	No hit
Glomus_c7119	10 427	10 427	1	Cytochrome P450
Glomus_c13017	4057	7579	2	No hit
Gi.12932_C1	3919	3919	1	No hit
EXTB106330.b0	3428	3428	1	No hit
Glomus_lrc986	2759	6911	3	No hit
EXTB28921.b0	2581	2581	1	No hit
Gi.304_C1	1996	30 567	15	No hit
CCHU1751.g1	1481	8199	6	Centromere-related protein
Glomus_c14959	1395	9891	7	Hypothetical protein
Gi.167_C1	1339	44 618	33	No hit
CEXTB101194.b0	1206	1307	1	No hit
Glomus_c19831	878	6131	7	Hypothetical protein
Glomus_c5958	761	46 426	61	No hit
Glomus_c3804	667	667	1	No hit
Glomus_c8796	638	638	1	No hit
EXTB112691.b0	576	576	1	No hit
Glomus_c8747	561	561	1	No hit
Glomus_c17585	426	426	1	Hypothetical protein
Gi.7237_C1	425	13 107	31	No hit
Glomus_c10269	378	3150	8	No hit
EXTB121165.b0	339	3765	11	No hit
EXTA118370.b0	333	1353	4	No hit
Glomus_c8579	329	16 088	49	No hit
Glomus_c2327	309	2838	9	No hit
Glomus_c12964	246	25 666	105	Hypothetical protein
Gi.8222_C1	236	18 627	79	Zinc-regulated transporter

A value of 1 was given to transcripts not detected in germinated spores.

The majority of IRM- and ERM-induced genes were lineage-specific genes, as 80% of the most highly up-regulated transcripts coded for orphan proteins (Tables 2, S3). Differentially expressed genes coding for proteins of known function were categorized into functional classes (Tables 3, S4). Genes overrepresented in IRM vs germinated spores showed a GO enrichment (FDR-corrected $P \leq$ 0.05) for several categories of biological processes associated with ion transport, lipid/steroid metabolism and DNA replication (data not shown). Out of 25 transcripts that appeared to be highly up-regulated in IRM, 17 were validated by RT-PCR (data not shown). While 13 out of 17 transcripts were detected in at least one other stage of the life cycle besides arbuscules, four were arbuscule-specific transcripts (Table S5). Three were orphan genes, while the fourth showed similarity to a multidrug resistance protein (ATP Binding Cassette (ABC) transporter).

Table 3 All transcripts (nonredundant virtual transcripts (NRVTs)) with functional annotation that were up-regulated by at least 10-fold in *Glomus intraradices*

NRVT ID	Ratio IRM:spores	IRM expression level	Putative function	Functional categories
EXTB25580.b0	17.6	8653	MFS monocarboxylic acid transporter	Transport
Glomus_c20826	13.7	25 099	Amino acid transporter	Transport
Glomus_c3620	27.4	21 270	MFS nitrate transporter	Transport
Glomus_c10810	34.5	22 106	MFS nitrate transporter	Transport
EXTA118370.b0	333.4	1353	Iron permease	Transport
Gi.8222_C1	236.2	18 627	ZIP (Zrt- and Irt-like) Zn transporter	Transport
Glomus_c13887	15.8	8222	P-type ATPase	Transport
Gi.5720_C1	12.5	3337	Cation transport-related protein	Transport
EXTA115701.b0	24.7	7920	ABC transporter	Transport
Glomus_c14346	16.1	1921	P-type ATPase	Transport
Glomus_c9272	77.1	3766	Fatty acyl-CoA elongase	Lipid metabolism
EXTB126208.b0	10.6	9156	Peroxisomal 3-ketoacyl-CoA-thiolase	Lipid metabolism
Glomus_c8371	40.2	9435	Phosphatidylglycerol/phosphatidylinositol transfer protein	Lipid metabolism
Gi.6806_C1	48.2	23 206	Lipase	Lipid metabolism
Glomus_c11625	15.4	26 531	Acetyl-CoA acetyltransferase	Lipid metabolism
Glomus_c15563	10.0	10 357	Adenine phosphoribosyl transferase	Nucleotide metabolism
Gi.1390_C2	32.1	15 107	Ribonucleoside-diphosphate reductase	Nucleotide metabolism
Glomus_c2075	23.9	12 720	Thymidylate synthase	Nucleotide metabolism
Glomus_c12729	26.4	10 113	Endopolyphosphatase	Phosphate metabolism
EXTB107807.b0	18.8	17 134	Endopolyphosphatase	Phosphate metabolism
Glomus_c4491	27.0	26 571	Carbonic anhydrase	Carbon metabolism
Gi.8293_C1	43.4	7083	Cytochrome P450	Secondary metabolism
Glomus_c7119	10426.8	10 427	Cytochrome P450	Secondary metabolism
EXTB18295.b0	27.7	20 682	Multicopper oxidase	Secondary metabolism
Glomus_c1906	10.0	11 236	Glutathione S-transferase	Secondary metabolism
Glomus_c1140	10.9	24 021	Iron sulfur cluster assembly protein	Protein modification
Glomus_c14063	21.0	7074	Ubiquitin-conjugating enzyme	Protein modification
Glomus_c22386	58.6	13 312	Protease, Ulp1 family	Protein modification
Glomus_c3253	13.6	3141	Ubiquitin/ribosomal protein S27a	Protein modification
Glomus_c8960	10.0	8892	Menaquinone biosynthesis methyltransferase	Coenzyme metabolism
Glomus_c10425	13.6	22 927	Steroidogenic acute regulatory protein	Signal transduction
CACE6804.g1	11.3	2292	Mitogen-activated protein kinase kinase kinase	Signal transduction
EXTB27526.b0	13.5	2738	Tyrosine-like kinase	Signal transduction
Glomus_lrc4409	50.5	37 996	Calcium-binding protein	Signal transduction
Glomus_lrc7412	33.1	50 044	Calcium-binding protein	Signal transduction
Glomus_c9394	14.4	5290	Chitin synthase	Cell wall biogenesis
Glomus_c1601	40.8	50 339	Cell wall anchor domain protein	Cell wall biogenesis
CCHU1751.g1	1480.7	8199	Centromere protein-related protein	Cell cycle control
Glomus_c10811	27.2	9906	Kinetochore protein	Cell cycle control
Glomus_c3780	17.9	20 774	Targeting protein for Xklp2	Cell cycle control
Glomus_c5080	50.3	16 455	Kinesin-like protein	Cytoskeleton
Glomus_c7425	35.1	4759	DNA topoisomerase type II	Chromatin structure
Glomus_lrc174	1214.8	22 989	Chromatin assembly factor I	Chromatin structure
Glomus_c505	26.3	37 147	Histone H3.1	Chromatin structure
CCHU9624.1b	12.5	450	DNA repair protein RAD51	Replication, repair
Glomus_c9197	12.7	7425	DNA replication licensing factor	Replication, repair
Glomus_c16484	73.5	2903	Uracil-DNA glycosylase	Replication, repair
Glomus_c13628	31.5	3433	BTB/POZ domain-containing protein	Transcription
Glomus_c18585	13.2	4751	Zn finger-containing protein	Transcription
Glomus_c10966	80.6	9500	Zn finger-containing protein	Transcription
Glomus_c1198	11.2	25 711	Yabby-like transcription factor	Transcription
Glomus_lrc739	11.3	11 898	Translation initiation factor 3	Translation

IRM compared with germinated spores based on oligoarray profiling (ratio ≥ 10 ; P -value < 0.05).

ABC, ATP-binding cassette; BTB (for BR-C, ttk and bab) or POZ (for Pox virus and Zinc finger); IRM, intraradical mycelium; MFS, major facilitator family.

As the oligoarray was constructed before the 454 ESTs from IRM and LMD arbuscule-containing cells were obtained, we also carried out a digital northern analysis based on the differential frequency of 454 reads (Audic & Claverie, 1997). This transcript profiling confirmed the above oligoarray expression patterns (data not shown). However, we detected a significant set of IRM transcripts coding for SSPs not detected in ERM and spores. These SSPs showed a striking up-regulation in IRM and arbuscule-containing cells (Table 4). The most highly up-regulated IRM orphan transcript 'step3_c3163' coded for an SSP of 280 amino acids of unknown function with similarity to proteins of several biotrophic basidiomycetes (e.g. *L. bicolor* and *M. larici-populina*). This transcript was only detected in mycorrhizal roots and arbuscules.

Recent studies of plant–microbe interactions involving pathogenic and ectomycorrhizal fungi have shown that many lineage-specific orphan genes code for effector proteins playing key roles in host colonization and *in planta* accommodation by controlling the plant immune system (De Wit *et al.*, 2009; Plett *et al.*, 2011). The set of differentially expressed *G. intraradices* orphan genes included the effector protein SP7, which interacts with the pathogenesis-related transcription factor Ethylene Response Factor (ERF19) in the host nucleus (Kloppholz *et al.*, 2011) (Table S4). Given the key results obtained for SP7, the role played by other mycorrhiza-up-regulated SSPs of *G. intraradices* should be elucidated, as well as the identity of plant-based signals that may control their expression within the root space.

Stealth colonization to evade host defenses

Several obligate biotrophic pathogens and ectomycorrhizal symbionts have a decreased repertoire of carbohydrate-acting enzymes

Table 4 The transcripts coding for mycorrhizal-induced small secreted proteins (MiSSPs) showing up-regulation in *Glomus intraradices* intraradical mycelium (IRM) compared with germinated spores based on 454 read frequency

NVRT ID	IRM (reads)	Arbuscules (reads)	Spores (reads)	Putative function	Size (aa)	# Cys
step3_c3163	62	4	0	Hypothetical protein	280	12
remain_c2375	24	16	0	No hit	156	3
step3_c3587	15	2	0	No hit	149	0
remain_c4548	12	4	0	MD-2-related lipid recognition domain	173	4
step3_c2834	11	19	0	No hit	61	9
remain_c2379	11	7	0	No hit	120	0
step3_c3282	10	1	0	No hit	86	2
remain_c8131	9	0	0	Hypothetical protein	150	3
remain_c2779	7	20	0	No hit	67	0
remain_c8897	7	1	0	No hit	71	7
remain_c12045	5	3	0	No hit	122	4

These sequences have no probe on oligoarrays.

aa, amino acids; # Cys, number of cysteine residues; NVRT, nonredundant virtual transcript.

involved in the degradation of plant cell wall (PCW) polysaccharides (Martin *et al.*, 2008b, 2010; Baxter *et al.*, 2010; Spanu *et al.*, 2010; Duplessis *et al.*, 2011). We identified 139 NVRTs encoding carbohydrate-active enzymes in the *G. intraradices* transcriptome (Table S6). They included enzymes that are involved in acquiring and converting carbohydrates and in maintaining fungal cell wall plasticity. Genes encoding glycosyl hydrolases (GHs) involved in degrading PCW lignocellulosic polymers were not found in the IRM predicted proteome. No genes encoding GH from families GH6 and GH7 (which include cellobiohydrolases and are involved in the attack of crystalline cellulose), polysaccharide lyases or proteins with cellulose-binding motif 1 (CBM1) were identified (Table S6). Enzymes that attack amorphous cellulose, hemicellulose, and pectin (e.g. β -glucosidases and cellobiose dehydrogenases of family GH3, GH10 xylanases, and GH28 pectinases) were absent. Genes encoding carbohydrate esterases and secreted feruloyl esterase which could hydrolyze cross-links in PCW were also not expressed. However, genes encoding GH5 endoglucanases (which act on mannan and/or cellulose) and a GH9 cellobiohydrolase were identified (Table S6). As observed in ectomycorrhizal symbionts (Martin *et al.*, 2008b, 2010; Nagendran *et al.*, 2009), the minimal set of PCW-degrading enzymes in the transcriptome of *G. intraradices* might be an evolutionary adaptation to avoid the release of polysaccharide fragments and their detection by the host immune system during the obligate biotrophic phase of the fungus. By contrast, a large set of genes encoding enzymes involved in chitin metabolism were expressed, including chitin synthases, chitin deacetylases, and chitinases (Table S6). They probably play a role in remodeling the fungal cell wall during growth and symbiosis. The abundance of transcripts coding for dolichyl-related mannosylation enzymes suggests an intense biosynthesis of cell wall mannans and manno-proteins.

Nutrient assimilation

We identified numerous transcripts coding for different predicted proteins associated with the uptake and assimilation of major soil nutrients, such as nitrate and Pi, and conversion of metabolites shuttled between partners (amino acids and carbohydrates) (Table S7). Although a few genes involved in nutrient transport and assimilation have previously been identified in *G. intraradices* and *Glomus mosseae* (Benedetto *et al.*, 2005; Govindarajulu *et al.*, 2005; Javot *et al.*, 2007; Tian *et al.*, 2010), there is no comprehensive molecular description of N and P assimilation pathways, despite the crucial importance of metabolite exchange in this symbiosis (Kiers *et al.*, 2011; Smith & Smith, 2011). The striking induction of several genes coding for metal transporters suggests that metals, such as Zn, are important for plant colonization, and the expression pattern of genes coding for assimilative enzymes fully supports rapid uptake, translocation and transfer of metabolites, as discussed in the following sections.

Nitrogen metabolism In the *G. intraradices*-colonized *M. truncatula* roots, transcripts for transporters and enzymes involved in N uptake and assimilation showed high constitutive expression, supporting the

contention that a high N turnover is taking place in the symbiosis (Gomez *et al.*, 2009; Guether *et al.*, 2009b). The up-regulation of transcripts for nitrate permease and nitrate reductase (Table S8), together with the higher expression of the gene coding for the ammonium transporter gene *GintAMT2* (Table S8; Pérez-Tienda *et al.*, 2011), might favor the accumulation of NH_4^+ for incorporation into amino acids and further translocation of N to the host cells. The IRM transcript profile is consistent with assimilation of NH_4^+ via the glutamine synthetase/glutamate synthase (GS/GOGAT) cycle. Transcripts encoding enzymes of arginine biosynthesis and degradation were highly expressed in spores, ERM and IRM (Table S8), indicating an intense cycling of N through the arginine pathway (Tian *et al.*, 2010). The higher expression of genes coding for arginase, urease and urease accessory protein (*UreG*) in IRM might lead to a greater release of NH_4^+ able to feed the GS/GOGAT cycle and/or the passive efflux of ammonia to the interfacial apoplast and the uptake of NH_4^+ by the host ammonium transporter(s) (Govindarajulu *et al.*, 2005; Gomez *et al.*, 2009; Tian *et al.*, 2010). High expression of amino acid transporter genes in ERM and IRM (Table S7) suggests that the mycobiont can taken up amino acids present in both soil and the interfacial apoplast.

Phosphate metabolism *Glomus intraradices* expressed a wide spectrum of secreted phosphatase transcripts, including those coding for the *p*-nitrophenylphosphatase Pho13p, acid phosphatase Pho3p, repressible alkaline phosphatase Pho8p, and magnesium-dependent phosphatase, able to act on various phosphate esters (Table S9). No phytase transcript was detected. *Glomus intraradices* has a combination of low-affinity and high-affinity transporter genes, including the high-affinity Pi transporter gene *GintPT* (equivalent to *Pho84p* from *Saccharomyces cerevisiae*), and genes encoding the low-affinity Pi transporter Pho91p and the Na^+ /Pi symporter Pho89p (Table S9). The transcripts coding for the interactor protein Pho88p which promotes maturation and trafficking of Pho84p in yeast have also been identified. Pi transporter genes are expressed in spores, ERM and IRM. The expression pattern of *GintPT* is in agreement with the results obtained on *GmosPT* in *G. mosseae* (Benedetto *et al.*, 2005) and the observed expression of *GmosPT* in arbuscules (Balestrini *et al.*, 2007; Gómez-Ariza *et al.*, 2009). Although high net transfer of P from soil to the plant takes place via the mycobiont (Kiers *et al.*, 2011; Smith & Smith, 2011), our findings suggest that the IRM may re-absorb some of the Pi released in the symbiotic apoplastic space. As shown by Kiers *et al.* (2011), the fungus might exert some control on the delivery of nutrients to its host plant, and reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis.

Following Pi uptake, PolyP accumulates in *G. intraradices* hyphae and is stored in the vacuolar compartments, where it buffers cytoplasmic Pi concentrations. PolyP is translocated along hyphae toward the interfacial apoplast (Viereck *et al.*, 2004; Javot *et al.*, 2007; Hijikata *et al.*, 2010). The rapid synthesis of PolyP is thus crucial for the maintenance of effective hyphal Pi uptake (Tani *et al.*, 2009). Transcripts coding for the PolyP polymerase/Vacuolar Transporter Chaperone Complex Vtc4p protein involved in the synthesis and transfer of PolyP to the vacuole

(Hothorn *et al.*, 2009), and for its accessory proteins, such as Vtc1p, were expressed in *G. intraradices* (Table S9). Transcripts for vacuolar endopolyphosphatase Ppn1p and the exopolyphosphatase Ppx1p were also identified. The endopolyphosphatase transcript was strikingly up-regulated in IRM. Several components of the complex phosphate regulatory system (PHO regulon) found in *S. cerevisiae* and *N. crassa* were expressed in *G. intraradices*, including the ankyrin-repeat-containing NUC2-related protein (Pho81p), the cyclin-dependent protein kinase regulator Pho80p and the mitogen-activated serine/threonine Cdc2 (Cell Cycle Control 2) cyclin-dependent protein kinase Pho85p (Table S9), suggesting putative regulatory similarities.

Lipid metabolism The profuse development of fungal membranes associated with arbuscule development in host cells dramatically increases the need for plasma membrane fatty acids and sterols. The oligoarray analysis identified transcripts in IRM coding for the subunits α and β of the fatty acid synthase complex, together with the acetyl-CoA carboxylase transcript (Table S10), confirming that the mycobiont does possess fatty acid synthetic capacities (Table S10). Similarly, mycorrhiza-induced genes predicted to be involved in fatty acid metabolism are up-regulated in *M. truncatula* (Gomez *et al.*, 2009), suggesting a complementary regulation of fungal and plant lipid metabolism. In this host plant, the gene *MtMSBPI*, encoding a membrane steroid (progesterone)-binding protein, is also induced early by a diffusible AM fungal signal produced by *G. intraradices* branched hyphae (Kuhn *et al.*, 2010). The high activity of enzymes involved in sterol and steroid metabolism in IRM, together with the AM fungal induction of *MtMSBPI*, might be related to the need to alter sterol metabolism (e.g. lanosterol) to allow plasma membrane invagination and intracellular accommodation of the fungal symbiont in the cortical cells.

An expanded inventory of conserved meiotic genes provides evidence for cryptic sex

The lack of an observed sexual stage in any member of the Glomeromycota led to the suggestion, which has been debated, that AM fungi were ancient asexuals. Results from Croll & Sanders (2009) strongly suggest that recombination occurred among some *G. intraradices* genotypes in the field, although the majority of populations examined so far have been clonal (Rosendahl, 2008). As Glomeromycota could represent one of the earliest diverging fungal lineages, their meiotic processes could represent an ancestral state. We surveyed the transcriptome of *G. intraradices* for a set of sex and meiotic genes conserved among eukaryotes (Malik *et al.*, 2008) (Table S11) and identified several 'meiosis-specific' genes (*HOP2* (*Homologous-pairing protein 2*) and *MND1* (*Meiotic nuclear division protein 1*)) which are only known to function in meiosis in other eukaryotes (Malik *et al.*, 2008). These genes are hypothesized to be present in organisms with sexual ancestry (Table S11). The homolog of a transcript coding for the key meiotic recombinase SPO11 (REC12) was not found, but has been identified in the genomic sequences of *Glomus diaphanum* (MUCL 43196), *G. intraradices* (DAOM

197198), *Glomus clarum* (DAOM 234281) and *Glomus cerebri-forme* (DAOM 227022) (Halary *et al.*, 2011). Transcripts coding for high-mobility group (HMG) domain-containing transcriptional factors with a significant similarity (55%) to the *sexP* and *sexM* genes from *Phycomyces blakesleeanus* (Idnurm *et al.*, 2008) were also retrieved. These genes are master switches controlling mating type in fungi. Mechanisms controlling the *G. intraradices* sexual cycle need to be further examined within AM fungal populations as they are likely to allow the mixing of nuclei and subsequent recombination among different individuals of this fungus interacting in soil.

In summary, induced expression of genes coding for membrane transporters and SSPs during the symbiotic interaction and the lack of expression of hydrolytic enzymes acting on PCW polysaccharides are hallmarks of *G. intraradices*. These results extend conserved patterns of gene expression profiles observed in obligate biotrophic pathogens (Spanu *et al.*, 2010; Kemen *et al.*, 2011) and ectomycorrhizal symbionts (Plett & Martin, 2011) to the Glomeromycota lineage. By contrast, obligate biotrophy in *G. intraradices* is not associated with a striking reduction of metabolic complexity (e.g. lack of N and S assimilation pathways), as observed in many obligate biotrophic pathogens, so that the ability to interact with the soil environment with respect to nutrient uptake is maintained in the symbiotic fungus. Finally, we can hypothesize that biotrophy in AM fungi has evolved through a series of steps requiring effectors, such as the secreted SP7 (Kloppholz *et al.*, 2011), coupled with a reduced inventory of PCW-hydrolyzing enzymes to suppress or attenuate host defense reactions, and weak selection forces to maintain certain biosynthetic pathways if products (e.g. thiamine and sucrose) can be directly obtained from the host. The present comprehensive repertoire of *G. intraradices* genes, the first for Glomeromycota, provides a basis for future research in environmental genomics and for accessing symbiosis-related functional features in other members of this unique phylum.

Acknowledgements

E.T. is supported by a scholarship from the Région Lorraine and the European Commission ENERGYPOPLAR project (FP7-211917), and Y.L. by a scholarship from the Education Ministry of China. This work was supported by the Network of Excellence EVOLTREE (FP6-016322) and the US Department of Energy (DOE) – Oak Ridge National Laboratory Scientific Focus Area for Genomics Foundational Sciences (grant to F.M.). Ph.F. was supported by the German Research Foundation and the Ministries of Food, Agriculture, and Consumer Protection of the FRG and the Brandenburg and Thüringen States. P.B. received funding from the Regional Project Converging Technologies-BI-OBIT, and N.F. and C.A.A. from the project AGL2009-08868 from the Spanish Ministry of Science and Education. Support for development of mycorrhizal *M. truncatula* root materials was provided by the US National Science Foundation (grant IOS-0842720 to M.J.H.). A portion of the writing of this manuscript was sponsored by the DOE Genomic Science Program (DE-AC05-00OR22725). EST sequencing was funded by

INRA, Genoscope, New Mexico University, and Michigan State University. The genomic and Sanger EST sequencing was conducted by JGI (DE-AC02-05CH11231). We would like to thank Nicolas Corradi for helpful discussions on the meiotic genes, and three anonymous referees for their insightful comments and suggestions.

References

- Akiyama K, Matsuzaki KI, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435: 824–827.
- Allen JW, Shachar-Hill Y. 2009. Sulfur transfer through an arbuscular mycorrhiza. *Plant Physiology* 149: 549–560.
- Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. 2010. Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology* 20: 1216–1221.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25: 25–29.
- Audic S, Claverie JM. 1997. The significance of digital gene expression profiles. *Genome Research* 7: 986–995.
- Bago B, Pfeffer PE, Abubaker J, Jun J, Allen JW, Brouillette J, Douds DD, Lammers PJ, Shachar-Hill Y. 2003. Carbon export from arbuscular mycorrhizal roots involves the translocation of carbohydrate as well as lipid. *Plant Physiology* 131: 1496–1507.
- Bago B, Pfeffer PE, Shachar-Hill Y. 2000. Carbon metabolism and transport in arbuscular mycorrhizas. *Plant Physiology* 124: 949–957.
- Balestrini R, Gómez-Ariza J, Lanfranco L, Bonfante P. 2007. Laser microdissection reveals that transcripts for five plant and one fungal phosphate transporter genes are contemporaneously present in arbusculated cells. *Molecular Plant-Microbe Interactions* 20: 1055–1062.
- Balestrini R, Lanfranco L. 2006. Fungal and plant gene expression in arbuscular mycorrhizal symbiosis. *Mycorrhiza* 16: 509–524.
- Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W *et al.* 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330: 1549–1551.
- Bécard G, Fortin JA. 1988. Early events of vesicular arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytologist* 108: 211–218.
- Bécard G, Kosuta S, Tamasloukht M, Séjalon-Delmas N, Roux C. 2004. Partner communication in the arbuscular mycorrhizal interaction. *Canadian Journal of Botany* 8: 1186–1197.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340: 783–795.
- Benedetto A, Magurno F, Bonfante P, Lanfranco L. 2005. Expression profiles of a phosphate transporter gene (*GmosPT*) from the endomycorrhizal fungus *Glomus mosseae*. *Mycorrhiza* 15: 620–627.
- Bonfante P, Genre A. 2010. Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nature Communications* 1: 48.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* 37: D233–D238.
- Chabot S, Bécard G, Piché Y. 1992. The life cycle of *Glomus intraradices* in root organ culture. *Mycologia* 84: 315–321.
- Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* 14: 1147–1159.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.

- Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. 2009. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* 181: 924–937.
- Croll D, Sanders IR. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evolutionary Biology* 9: 13.
- Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders IR. 2008. Genetic diversity and host plant preferences revealed by simple sequence repeat and mitochondrial markers in a population of the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* 178: 672–687.
- De Wit PJGM, Mehrabi R, Van Den Burg HA, Stergiopoulos I. 2009. Fungal effector proteins: past, present and future. *Molecular Plant Pathology* 10: 735–747.
- Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics* 8: 539–548.
- Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL *et al.* 2011. Obligate biotrophy features unraveled by the genomic analysis of the rust fungi, *Melampsora larici-populina* and *Puccinia graminis* f. sp. *tritici*. *Proceedings of the National Academy of Sciences, USA* 108: 9166–9171.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300: 1005–1016.
- Ercolin F, Reinhardt D. 2011. Successful joint ventures of plants: arbuscular mycorrhiza and beyond. *Trends in Plant Science* 16: 356–362.
- Fitter AH, Helgason T, Hodge A. 2011. Nutritional exchanges in the arbuscular mycorrhizal symbiosis: implications for sustainable agriculture. *Fungal Biology Reviews* 25: 68–72.
- Genre A, Chabaud M, Timmers T, Bonfante P, Barker DG. 2005. Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. *Plant Cell* 17: 3489–3499.
- Giovannetti M, Fortuna P, Citernesi AS, Morini S, Nuti MP. 2001. The occurrence of anastomosis formation and nuclear exchange in intact arbuscular mycorrhizal networks. *New Phytologist* 151: 717–724.
- Gomez SK, Javot H, Deewatthanawong P, Torres-Jerez I, Tang Y, Blancafort EB, Udvardi MK, Harrison MJ. 2009. *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. *BMC Plant Biology* 22: 10.
- Gómez-Ariza J, Balestrini R, Novero M, Bonfante P. 2009. Cell-specific gene expression of phosphate transporters in mycorrhizal tomato roots. *Biology and Fertility of Soils* 45: 845–853.
- Govindarajulu M, Pfeffer PE, Jin HR, Abubaker J, Douds DD, Allen JW, Bucking H, Lammers PJ, Shachar-Hill Y. 2005. Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature* 435: 819–823.
- Grunwald U, Guo W, Fischer K, Isayenkov S, Ludwig-Müller J, Hause B, Yan X, Küster H, Franken P. 2009. Overlapping expression patterns and differential transcript levels of phosphate transporter genes in arbuscular mycorrhizal, Pi-fertilised and phytohormone-treated *Medicago truncatula* roots. *Planta* 229: 1023–1034.
- Guether M, Balestrini R, Hannah MA, Udvardi MK, Bonfante P. 2009a. Genome-wide reprogramming of regulatory networks, transport, cell wall and membrane biogenesis during arbuscular mycorrhizal symbiosis in *Lotus japonicus*. *New Phytologist* 182: 200–212.
- Guether M, Neuhäuser B, Balestrini R, Dynowski M, Ludewig U, Bonfante P. 2009b. A mycorrhizal-specific ammonium transporter from *Lotus japonicus* acquires nitrogen released by arbuscular mycorrhizal fungi. *Plant Physiology* 150: 73–83.
- Güimil S, Chang H-S, Zhu T, Sesma A, Osbourn A, Roux C, Ioannidis V, Oakeley EJ, Docquier M, Descombes P *et al.* 2005. Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proceedings of the National Academy of Sciences, USA* 102: 8066–8070.
- Halary S, Malik SB, Lildhar L, Slamovits CH, Hijri M, Corradi N. 2011. Conserved meiotic machinery in *Glomus* spp., a putatively ancient asexual fungal lineage. *Genome Biology and Evolution* 3: 950–958.
- van der Heijden MGA, Klironomos JN, Ursic M, Moutoglou P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. 1998. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396: 69–72.
- Hijikata N, Murase M, Tani C, Ohtomo R, Osaki M, Ezawa T. 2010. Polyphosphate has a central role in the rapid and massive accumulation of phosphorus in extraradical mycelium of an arbuscular mycorrhizal fungus. *New Phytologist* 186: 285–289.
- Hothorn M, Neumann H, Lenherr ED, Wehner M, Rybin V, Hassa PO, Uttenweiler A, Reinhardt M, Schmidt A, Seiler J *et al.* 2009. Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science* 324: 513–516.
- Idnurm A, Walton FJ, Floyd A, Heitman J. 2008. Identification of the sex genes in an early diverged fungus. *Nature* 451: 193–197.
- Jany JL, Pawlowska TE. 2010. Multinucleate spores contribute to evolutionary longevity of asexual Glomeromycota. *The American Naturalist* 175: 424–435.
- Javot H, Pumplin N, Harrison MJ. 2007. Phosphate in the arbuscular mycorrhizal symbiosis: transport properties and regulatory roles. *Plant, Cell & Environment* 30: 310–322.
- Kamoun S. 2007. Groovy times: filamentous pathogen effectors revealed. *Current Opinion in Plant Biology* 10: 358–365.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28: 27–30.
- Kemen A, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, Maclean D *et al.* 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biology* 9: e1001094.
- Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A *et al.* 2011. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* 333: 880–882.
- Kloppholz S, Kuhn H, Requena N. 2011. A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Current Biology* 21: 1204–1209.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* 305: 567–580.
- Kuhn H, Küster H, Requena N. 2010. Membrane steroid-binding protein 1 induced by a diffusible fungal signal is critical for mycorrhization in *Medicago truncatula*. *New Phytologist* 185: 716–733.
- Kuznetsova E, Seddas-Dozolme PMA, Arnould C, Tollot M, van Tuinen D, Borisov A, Gianinazzi S, Gianinazzi-Pearson V. 2010. Symbiosis-related pea genes modulate fungal and plant gene expression during the arbuscule stage of mycorrhiza with *Glomus intraradices*. *Mycorrhiza* 20: 427–443.
- Letunic I, Yamada T, Kanehisa M, Bork P. 2008. iPath: interactive exploration of biochemical pathways and networks. *Trends in Biochemical Sciences* 33: 101–103.
- Maere S, Heymans K, Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448–3449.
- Maillet F, Poinso V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey D, Niebel A *et al.* 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* 469: 58–63.
- Malik SB, Pightling AW, Stefaniak LM, Schurko AM, Logsdon JM. 2008. An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. *PLoS ONE* 3: e2879.
- Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al.* 2008b. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452: 88–92.
- Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders IR, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JPW. 2008a. The long hard road to a completed *Glomus intraradices* genome. *New Phytologist* 180: 747–750.
- Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R *et al.* 2010. Périgord Black Truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464: 1033–1038.

- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35: W182–W185.
- Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. 2009. Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. *Fungal Genetics and Biology* 46: 427–435.
- Parra G, Bradnam K, Ning Z, Keane T, Korfi I. 2009. Assessing the gene space in draft genomes. *Nucleic Acids Research* 37: 289–297.
- Pérez-Tienda J, Testillano PS, Balestrini R, Fiorilli V, Azcón-Aguilar C, Ferrol N. 2011. GintAMT2, a new member of the ammonium transporter family in the arbuscular mycorrhizal fungus *Glomus intraradices*. *Fungal Genetics and Biology* 48: 1044–1055.
- Plett JM, Kempainen M, Kale SD, Kohler A, Legué V, Brun A, Tyler B, Pardo A, Martin F. 2011. A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Current Biology* 21: 1197–1203.
- Plett JM, Martin F. 2011. Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. *Trends in Genetics* 27: 14–22.
- Rosendahl S. 2008. Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytologist* 178: 253–266.
- Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 microarray software suite. *Methods in Enzymology* 411: 134–193.
- Sanders IR, Croll D. 2010. Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annual Review of Genetics* 44: 271–292.
- Schaarschmidt S, Roitsch T, Hause B. 2006. Arbuscular mycorrhiza induces gene expression of the apoplastic invertase LIN6 in tomato (*Lycopersicon esculentum*) roots. *Journal of Experimental Botany* 57: 4015–4023.
- Schüssler A, Schwarzott D, Walker C. 2001. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* 105: 1413–1421.
- Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* 473: 337–342.
- Seddas PM, Arias C, Arnould C, van Tuinen D, Godfroy O, Benhassou H, Gouzy J, Morandi D, Dessaint F, Gianinazzi-Pearson V. 2009. Symbiosis-related plant genes modulate molecular responses in an arbuscular mycorrhizal fungus during early root interactions. *Molecular Plant-Microbe Interactions* 22: 341–351.
- Smith SE, Read DJ. 2008. *Mycorrhizal symbiosis*. London, UK: Academic press.
- Smith SE, Smith FA. 2011. Roles of arbuscular mycorrhizas in plant nutrition and growth: new paradigms from cellular to ecosystem scales. *Annual Review of Plant Biology* 62: 227–250.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM, Butcher SA, Gurr SJ *et al.* 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* 330: 1543–1546.
- Stockinger H, Walker C, Schüssler A. 2009. ‘*Glomus intraradices* DAOM197198’, a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytologist* 183: 1176–1187.
- Tani C, Ohtomo R, Osaki M, Kuga Y, Ezawa T. 2009. ATP-dependent but proton gradient-independent polyphosphate-synthesizing activity in extraradical hyphae of an arbuscular mycorrhizal fungus. *Applied and Environmental Microbiology* 75: 7044–7050.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN *et al.* 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
- Tian C, Kasiborski B, Koul R, Lammers PJ, Bucking H, Shachar-Hill Y. 2010. Regulation of the nitrogen transfer pathway in the arbuscular mycorrhizal symbiosis: gene characterization and the coordination of expression with nitrogen flux. *Plant Physiology* 153: 1175–1187.
- Viereck N, Hansen PE, Jakobsen I. 2004. Phosphate pool dynamics in the arbuscular mycorrhizal fungus *Glomus intraradices* studied by *in vivo* ³¹P NMR spectroscopy. *New Phytologist* 162: 783–794.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Flow chart describing the *in silico* filtering process of *Glomus intraradices* expressed sequence tags (ESTs).

Fig. S2 GC% distribution of *Glomus intraradices* raw and filtered expressed sequence tag (EST) sequences for each cDNA library.

Fig. S3 Rarefaction curves depicting the effect of *Glomus intraradices* read number (a) on the number of assembled nonredundant virtual transcripts (NRVTs) and (b) on the number of detected proteins in the Swiss-Prot database and *Rhizopus oryzae* gene repertoire.

Fig. S4 Similarity of *Glomus intraradices* nonredundant virtual transcripts (NRVTs) to gene coding sequences from other selected fungi.

Table S1 The number of assembled nonredundant virtual transcripts (NRVTs) was affected by the parameters used for assembling the *Glomus intraradices* expressed sequence tags (ESTs) using the MIRA assembler

Table S2 Genes missing in the *Glomus intraradices* transcriptome, and gene repertoires of *Blumeria graminis* (Spanu *et al.*, 2010) and other obligate biotrophic pathogens (Kemen *et al.*, 2011)

Table S3 Most highly up-regulated *Glomus intraradices* transcripts in the extraradical mycelium (ERM) vs germinated spores identified using expression oligoarrays and ranked by decreasing fold changes

Table S4 All transcripts with functional annotations that were up-regulated by at least 10-fold in *Glomus intraradices* extraradical mycelium (ERM) compared with germinated spores based on oligoarray profiling (ratio ≥ 10 ; *P*-value < 0.05)

Table S5 Expression profiles, as obtained by semiquantitative RT-PCR, of genes showing up-regulation in *Glomus intraradices* intraradical mycelium (IRM) compared with extraradical mycelium (ERM) in oligoarray profiling

Table S6 Expression of nonredundant virtual transcripts (NRVTs) coding for carbohydrate-active enzymes (Cantarel *et al.*, 2009) identified in the *Glomus intraradices* transcriptome in germinated spores, extraradical mycelium (ERM) and intraradical mycelium (IRM)

Table S7 Expression of nonredundant virtual transcripts (NRVTs) coding for membrane transporters identified in the *Glomus intraradices* transcriptome in germinated spores, extraradical mycelium (ERM) and intraradical mycelium (IRM)

Table S8 Expression of nonredundant virtual transcripts (NRVTs) coding for transporters and enzymes involved in nitrogen metabolism in the *Glomus intraradices* transcriptome

Table S9 Expression of nonredundant virtual transcripts (NRVTs) coding for transporters and enzymes involved in phosphate metabolism in the *Glomus intraradices* transcriptome

Table S10 Expression of nonredundant virtual transcripts (NRVTs) coding for lipid metabolism enzymes in the *Glomus intraradices* transcriptome

Table S11 Expression levels and regulation ratio of transcripts related to meiosis in the *Glomus intraradices* transcriptome

Methods S1 Construction of the cDNA libraries.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.