Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal Amanita

Maryam Chaib De Mares1,2*, Jaqueline Hess2,3*, Dimitrios Floudas4, Anna Lipzen5, Cindy Choi5, Megan Kennedy5, Igor V. Grigoriev5 and Anne Pringle6

1Department of Microbial Ecology, University of Groningen, 9747 AG Groningen, the Netherlands; 2Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA; 3Department of Biosciences, University of Oslo, 0371 Oslo, Norway; 4Department of Biology, Clark University, Worcester, MA 01610, USA; 5US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA; 6Harvard Forest, 324 North Main Street, Petersham, MA 01366, USA

Summary

- The genus Amanita encompasses both symbiotic, ectomycorrhizal fungi and asymbiotic litter decomposers; all species are derived from asymbiotic ancestors. Symbiotic species are no longer able to degrade plant cell walls. The carbohydrate esterases family 1 (CE1s) is a diverse group of enzymes involved in carbon metabolism, including decomposition and carbon storage. CE1 genes of the ectomycorrhizal A. muscaria appear diverged from all other fungal homologues, and more similar to CE1s of bacteria, suggesting a horizontal gene transfer (HGT) event.
- In order to test whether Amanita CE1s were acquired horizontally, we built a phylogeny of CE1s collected from across the tree of life, and describe the evolution of CE1 genes among Amanita and relevant lineages of bacteria.
- CE1s of symbiotic Amanita were very different from CE1s of asymbiotic Amanita, and are more similar to bacterial CE1s. The protein structure of one CE1 gene of A. muscaria matched a depolymerase that degrades the carbon storage molecule poly((R)-3-hydroxybutyrate) (PHB). Asymbiotic Amanita do not carry sequence or structural homologues of these genes.
- The CE1s acquired through HGT may enable novel metabolisms, or play roles in signaling or defense. This is the first evidence for the horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal fungi.

Introduction

Horizontal gene transfer (HGT) is the mobilization and stable integration of genetic material between distinct, reproductively isolated genomes (Richards et al., 2011). HGT is ubiquitous among bacteria (Brown & Doolittle, 1997; Lawrence & Ochman, 1997; Nelson et al., 1999; Koonin et al., 2001), and is also a major force in evolution among eukaryotes, enabling diversification and the adaptation of organisms to new environments. Horizontally transferred genes have facilitated changes in the host ranges of rumin and pathogenic fungi (Garcia-Vallvé et al., 2000; Juhas et al., 2009; Mehrabi et al., 2011; Sun et al., 2013), the spread of antibiotic resistance (Weldhagen, 2004; Roberts, 2005; Hanssen & Ericson-Söllid, 2006), and the evolution of novel metabolic capabilities (Lawrence & Ochman, 1998; Kanhere & Vingron, 2009; Marchetti et al., 2009; Ma et al., 2010; Christin et al., 2012).

Although HGT appears to be less frequent among eukaryotes, as compared with bacteria, there is ample evidence for the movement of genes from bacteria to fungi, as well as among different species of fungi and between plants and fungal pathogens (Richards et al., 2009, 2011; Fitzpatrick, 2011; Armijos-Jaramillo et al., 2013; Sun et al., 2013; Bruto et al., 2014). Ectomycorrhizal (EM) fungi form intimate associations with the roots of plants, but also extend into the surrounding soil, an environment teeming with bacteria (Gans et al., 2005). Nevertheless, to date there is scant evidence for HGT into EM lineages. Research targeting the EM fungus Amanita muscaria and transgenic poplar trees found no evidence for HGT between fungus and plant in laboratory settings (Zhang et al., 2005; Nehls et al., 2006). But agrobacteria have been used to transform other EM fungi, including Hebeloma cylindrosporum (Combier et al., 2003), Tuber borchii (Grimaldi et al., 2005) and Laccaria bicolor (Kemppainen & Pardo, 2011); demonstrating that lateral acquisition of genes from bacteria is theoretically possible. Whether HGT enables the movement of genes into EM fungi in nature remains an open question.

The movement of carbohydrate metabolism genes from bacteria to fungi or between fungi may enable fungi to establish in novel habitats or niches. For example, the glycosyl hydrolases

*These authors contributed equally to this work.
(GH) of rumen fungi are bacterial in origin, and allow the fungi to degrade cellulose and hemicellulose in the rumens of herbivorous mammals (Garcia-Vallvé et al., 2000). The transfer of a high affinity nitrate assimilation gene cluster from a basidiomycete to an ancestor of the ascomycetous mold Trichoderma reesei may have facilitated a change in the mold’s nutritional mode, allowing it to become an efficient wood degrader (Slot & Hibbett, 2007). Moreover, an earlier HGT of nitrate assimilation genes into Dikarya may have facilitated exploitation of nitrate in aerobic soils (Slot & Hibbett, 2007).

The genus Amanita encompasses a diversity of EM and closely related saprotrophic (SAP) fungi. A recent phylogeny documents a single origin of symbiosis within the Amanita; asymbiotic Amanita form a strongly supported clade basal to a monophyletic clade of symbiotic species (Wolfe et al., 2012b). Comparative genomics of EM and SAP Amanita reveal large-scale losses of carbohydrate-active enzymes from symbiotic genomes (Nagendran et al., 2009; Wolfe et al., 2012b; Chaib De Mares et al., 2013; Hess & Pringle, 2014). The result appears to be a general one; plant cell wall degrading enzymes are frequently lost after fungi become obligately dependent on plants for carbon (Martin et al., 2008, 2010).

The carbohydrate esterases family 1 (CE1s; Cantarel et al., 2009) are a diverse group of enzymes, encompassing at least seven classes within the CAZy database (http://www.cazy.org/). The enzymes are heterogeneous with respect to both substrate specificity and structure. Some CE1 enzymes target esters or amides, deacetylating the side group components of hemicellulose (Towler et al., 1988; Biely, 2012). These side groups covalently link and physically mask potentially fermentable substrates in plant cell walls, perhaps protecting them from degradation (Akin, 2008). CE1s of this group are hemicellulose accessory enzymes (McDonald et al., 2009) and enable microorganisms to attack and partially degrade plant tissues, working with xylanases and pectinases to break apart plant cell walls (Kubicek et al., 2010). Other CE1 enzymes, structurally related to lipases and proteases, consist of depolymerases that degrade a bacterial polymer, poly-((R)-3-hydroxybutyrate) (PHB; Dawes, 1988). PHB is built from glucose, and is used by microorganisms to store energy; it is metabolized when other carbon sources are unavailable.

Phylogenetic analyses by the Mycorrhizal Genomics Initiative (MGI, http://mycor.nancy.inra.fr/IMGC/MycoGenomes/) identify three phylogenetically distinct clades of CE1 genes in fungal genomes (genes annotated as CE1s identified by B. Henrissat, Cantarel et al., 2009). Two clades show sequence homology to hemicellulose accessory enzymes. But within the analyses, a third group of CE1 genes from the EM species A. muscaria appears diverged from other clades, and similar to CE1s of bacteria. Complex evolutionary relationships among CE1s of fungi, bacteria and plants are common (Udatha et al., 2011), and the patterns suggest a potential HGT event between A. muscaria and bacteria.

In order to explore and test for potential HGT events in A. muscaria and the genus Amanita, we identified four key questions: are the CE1 genes described from A. muscaria structurally and functionally integrated into the genome? What kind of CE1 genes are found in other Amanita species, and are the CE1s of ectomycorrhizal Amanita different from CE1s of saprotrophs? Do phylogenies built from a comprehensive dataset of CE1s suggest HGT? What do phylogenies tell us about the history of CE1s within the lineage of ectomycorrhizal Amanita? We took a variety of genetic and bioinformatic approaches to answer these questions, and then more fully characterized the EM Amanita CE1 genes, as well as the function of one predicted protein.

**Materials and Methods**

**Identification of CE1s in Amanita genomes and a transcriptome**

In order to investigate the origins of Amanita CE1 genes, we first identified the complete set of CE1 genes in A. muscaria, homologues present in available genomes of other Amanita species, and homologues in the outgroup Volvariella volvacea. The genome of A. muscaria var. guessowii (Koide BX008, Pennsylvania, USA; Hess & Pringle, 2014) has been sequenced twice; one genome is deposited at Joint Genome Institute (JGI, genome.jgi.doe.gov/Amusc;aryia), and the other was sequenced by the Pringle laboratory (Cambridge, MA, USA). The Pringle laboratory has also sequenced genomes for the EM fungi A. brunneescens and A. polyfarramis, and SAP fungi A. inapinata and V. volvacea (Hess et al., 2014). Genome sequences are available at NCBI under the accession nos. PRJNA236753, PRJNA236755, PRJNA236758, PRJNA236757 and PRJNA236756. The genome of A. thiissii is also available through JGI (genome.jgi.doe.gov/Athiissii, Wolfe et al., 2012a).

In addition to genomic data, a transcriptome of A. crenulata was sequenced at JGI in the context of a different experiment and Illumina RNA-Seq data are available at NCBI SRA under the accessions SRX141954 and SRX141955. Cultures were maintained as in Wolfe et al. (2012b). Mycelia were collected and immediately stored in RNALater (Qiagen). RNA was isolated with the RNeasy Maxi Kit (Qiagen). Poly A RNA was isolated from 10 ug total RNA using the Absolutely mRNA purification kit (Stratagene, Santa Clara, CA, USA). This procedure was repeated twice, to ensure that the sample was free from rRNA contamination. Detailed protocols for RNA isolation, sequencing libraries preparation, sequencing and assembly are available in Supporting Information Methods S1.

In order to generate a catalogue of candidate CE1 loci, two previously annotated CE1 genes – one from A. muscaria (JGI protein ID 166350) and a second from A. thiissii (JGI protein ID 1897), both annotated by B. Henrissat (pers. comm.) – were used as probes to screen all available genomes and the transcriptome with TBLASTN (Altschul et al., 1990), using an E-value cutoff of 10^-5 (Table S1).

**Naming conventions**

We adopt the following naming conventions: CE1 genes in A. muscaria are labelled as CE1_AmX, where X is a number used to distinguish among individual genes. The CE1 genes of other species are named as CE1_Ab (A. brunneescens, where only one
gene was identified), CE1_AcrX (A. crenulata) and CE1_AcoX (A. constricta). Carbohydrate-binding module 1 (CBM1) is an additional domain found in CE1 genes of A. thiersii and V. volvacea, and so we named the CE1s of these saprotrophic species as CE1-CBM1_Ath (A. thiersii) and CE1-CBM1_VvX (X is a number). The fungal genes flanking CE1 genes of A. muscaria’s scaffold 57 (Fig. 1) are labelled as FX (X is a number).

Confirming physical integration of A. muscaria CE1 genes

In order to confirm a subset of candidate A. muscaria genes as physically linked and integrated within the genome, and not the result of contaminant DNA, we used a single, long-range PCR to amplify a 24 435 bp region of A. muscaria JGI genome scaffold 57, where eight CE1 and five fungal genes are found (Fig. 1). We used primers spanning genes CE1_Am1 and CE1_Am3 (Table 1; see also the ‘Primer Design’ section below). To ensure high fidelity amplification, we used LongAmpTM Taq DNA polymerase (New England BioLabs Inc., Ipswich, MA, USA) and followed the manufacturer’s protocol implementing 16-min extension cycles. The resulting long-range construct was used as a template for subsequent PCR reactions, and was therefore diluted 1 : 1000. CE1 genes (CE1_Am1, CE1_Am5, CE1_Am9, CE1_Am10, CE1_Am2 and CE1_Am3; Table S2) and interspersed fungal genes (Table S3) were successfully amplified from the long-range construct using specific primer pairs (Table 1), and confirmed by resequencing (Genewiz Inc., Cambridge, MA, USA).

Expression of A. muscaria CE1 genes

The expression of A. muscaria CE1 genes was confirmed using RNAseq data available at JGI (http://genome.jgi-psf.org/cgi-bin/browserLoad/?db=Amamu1). Assembled transcripts were downloaded and aligned to the genome using GMAP with default settings (Wu & Watanabe, 2005) and visualized using IGV (Thorvaldsdóttir et al., 2013).

Identification of CE1s in Amanita without sequenced genomes


Genomic DNA was also newly isolated from the original, sequenced strain of A. muscaria var. guessowii (Koide BX008), two additional strains (FP01, collected in Cambridge, MA, USA, and PS #283, from the Penn State Spawn Collection, originally collected in PA, USA), and A. brunnescens (BW HF10C), using a modified version of an extraction protocol developed for Phytophthora infestans (http://my.jgi.doe.gov/general/protocols/). We used a 1 : 1 phenol chloroform ratio in our extractions.

Primer design

CE1 genes found in sequenced genomes of A. muscaria and A. brunnescens, and the transcriptome of A. crenulata, are variable enough so that individual genes can be distinguished. To confirm the presence of multiple individual CE1s in genomic DNA, we designed highly specific primers for each CE1 gene, based on either A. muscaria or A. brunnescens sequences (Table 1). To increase the likelihood that each primer pair designed from A. muscaria would amplify only one CE1 gene, primers were simultaneously designed for all of the different CE1 loci identified on scaffold 57 of the A. muscaria JGI assembly (Table S2) using Geneious v1.6 (http://www.geneious.com/, Biomatters, Newark, NJ, USA; Table 1). Our design strategy was successful, but because we designed these primers to target specific, variable regions, in some cases the primers do not amplify the full length of a gene. Moreover, successfully amplified fragments were more easily sequenced from species closely related to A. muscaria.

Identification of CE1s from across the tree of life

We next collected CE1 homologues from across the tree of life, using the predicted protein sequences of CE1s from A. muscaria, A. brunnescens, A. thiersii and V. volvacea as probes (Table S4). Before probing, we removed the carbohydrate binding modules.
Table 1 Primers used to amplify CE1 genes across the Amanita

<table>
<thead>
<tr>
<th>Gene name*</th>
<th>Primer pair</th>
<th>Primer name</th>
<th>Sequence (5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE1_Am1</td>
<td>1</td>
<td>CE1.1.i.fw</td>
<td>CCATGCGGTGACTCTCCTGAAGAC</td>
</tr>
<tr>
<td>CE1_Am2</td>
<td>2</td>
<td>CE1.1.i.rv</td>
<td>CAGCGCGTGACTCTCAAGC</td>
</tr>
<tr>
<td>CE1_Am3</td>
<td>3</td>
<td>CE1.2.i.fw</td>
<td>TGTTTTCCGTGGCATCTTGTG</td>
</tr>
<tr>
<td>CE1_Am4</td>
<td>4</td>
<td>CE1.2.i.rv</td>
<td>CCAGGAGCCGCTATTAGTAT</td>
</tr>
<tr>
<td>CE1_Am5</td>
<td>5</td>
<td>CE1.3.i.fw</td>
<td>TCTTTGCATCTCAGCTGCT</td>
</tr>
<tr>
<td>CE1_Am6</td>
<td>6</td>
<td>CE1.3.i.rv</td>
<td>ATCACCTGGGCTACCTGCT</td>
</tr>
<tr>
<td>CE1_Am7</td>
<td>7</td>
<td>CE1.5.i.fw</td>
<td>TGGGTACGCAATGTCGGTAA</td>
</tr>
<tr>
<td>CE1_Am8</td>
<td>8</td>
<td>CE1.5.i.rv</td>
<td>GACAAGCGCTGCTGTTGGTA</td>
</tr>
<tr>
<td>CE1_Am9</td>
<td>9</td>
<td>CE1.7.i.fw</td>
<td>CCTCCCGGATATACGGGCC</td>
</tr>
<tr>
<td>CE1_Am10</td>
<td>10</td>
<td>CE1.7.i.rv</td>
<td>ATGGTATCAGGGCTACCT</td>
</tr>
<tr>
<td>CE1_Ab</td>
<td>1b</td>
<td>A_brun_CE1.fw</td>
<td>ATACAGGCCTTCCCGCTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A_brun_CE1.rv</td>
<td>ATTTCTGATGCGACCAGGTG</td>
</tr>
</tbody>
</table>

*Names as described in the ‘Materials and Methods’ section.

(CBM1s) found in the *A. thiersii* and *V. volvacea* homologues, to prevent nonspecific matching. We conducted a BLASTP search in the NR database at NCBI with default parameters and an E-value cutoff of $10^{-5}$ on 26 June 2013. Because JGI is a central repository for fungal genome sequences, we also ran BLASTP searches in Mycocosm (http://genome.jgi.doe.gov/pages/blast.jsf?db=fungi), using the same parameters. These searches involved the predicted proteomes of all available fungal genomes; 134 as of 26 September 2013. To ensure we did not miss any homologues due to gene annotation artifacts, we also screened unmasked genome sequences using TBLASTN (Altschul et al., 1990).

Phylogenetic analyses

Searches and screens provided a collection of CE1 sequence data from *Amanita* and outgroup genomes, from *Amanita* without sequenced genomes, and from all species found in databases. We used these sequences to build a comprehensive phylogeny of the full diversity of CE1 genes. Sequences were aligned using PAGAN (Löytynoja et al., 2012) and then trimmed using trimAl v1.2 with –gt 0.1 (Capella-Gutierrez & Silla-Martinez, 2009). The best model of protein evolution was determined to be WAG + Γ, using ProtTest 3 (Darriba et al., 2011). A maximum-likelihood (ML) phylogeny was built with RAxML v7.7.5 (Stamatakis, 2006), using WAG + Γ with four rate categories and the ‘–autoMRE’ automatic stopping criterion for rapid bootstraps, which converged after 250 replicates. Attempts to implement a complementary, Bayesian approach failed; amino acid models are computationally intensive and after running analyses for 6 wk or more it became clear they would not converge on a meaningful result (see also Bruto et al., 2014; and Nikolaidis et al., 2014, for comparison).

Evolutionary history of CE1s in EM *Amanita*

In order to explore the evolutionary dynamics of CE1 genes within symbiotic *Amanita*, we reconstructed the CE1 gene tree of the four species for which we have full length CE1 sequence data; *A. brunnescens*, *A. muscaria*, *A. cremulata* and *A. constricata*. Nucleotide sequences were aligned twice using PRANK (Löytynoja & Goldman, 2008, 2010). The first alignment produced by PRANK was used to build a ML phylogeny with the phylogenetic software RAxML v7.4.9 (Stamatakis, 2006) and the GTR + Γ substitution model. The tree produced by this analysis was used to guide the second alignment. We iterated the alignment twice because using an improved guide tree (from the first alignment) often results in a more accurate second alignment; the evolutionary algorithms implemented by PRANK are strongly dependent on the guide tree (results not shown). We trimmed the resulting alignment with trimAl v1.2 (Capella-Gutierrez & Silla-Martinez, 2009), using a gap threshold of 0.5 and keeping at least 70% of the aligned positions. An ML gene tree for 16 genes was generated with RAxML v7.4.9 (Stamatakis, 2006) using the GTR + Γ model and 100 bootstrap replicates. The resulting tree was reconciled with the species tree generated by Wolfe et al. (2012b) using the software TreeFix (Wu et al., 2013). We estimated branch lengths for the ML reconciled tree and projected bootstrap values from the ML gene tree onto the ML reconciled gene tree using RAxML v7.4.9 (Stamatakis, 2006). The numbers of duplications and losses were mapped manually on the resulting reconciled tree.

HGT gene characterization: nucleotide composition and codon usage bias

Horizontally transferred genes often reflect the nucleotide composition of the donor genome at the time of transfer. Following acquisition, sequences will be subject to the same genome pressures as native genes and over time those genes will resemble genes in the recipient genome; a process named amelioration (Lawrence & Ochman, 1997). To test whether HGT genes are ameliorated to host genomes, we characterized nucleotide composition and codon usage. Analyses were based on the G + C content and codon usage of CE1 genes of the EM species *A. muscaria* and *A. brunnescens*, because we had access to both, the full-length CE1 sequences and genomic background data for these species. To visualize codon bias, we used the base composition of silent sites at the third position of synonymous codons with either a G or C (GC3s) as a proxy (Roth et al., 2012). We used the effective number of codons (Nc) to measure the deviation from uniform codon usage (Wright, 1990); values of Nc range from 20 (when only one codon is used per amino acid) to 61 (the standard genetic code, where all possible synonymous codons are used with equal frequency). These three measures (Nc, GC3 and GC) were also calculated for CE1 genes of the saprotrophs *A. thiersii* (Stamatakis, 2006).
and *V. volvacea*, the CE1 gene of the bacteria *Ktedonobacter racemifer*, the closest homologue to a potential donor lineage. Measures were calculated using CodonW (Peden, 2000; http://codonw.sourceforge.net/).

### HGT gene characterization: evolution of gene structure

We reconstructed patterns of exon and intron structure for CE1 gene sequences found in the EM *A. muscaria, A. brunnescens*, and the saprotrophs *A. thiersii* and *V. volvacea* using WebScipio (Odronitz *et al.*, 2013; Hatje *et al.*, 2011; http://www.webscipio.org). These are the same species and sequences used in analyses of nucleotide composition and codon usage bias. We then used GenePainter (Hammesfahr *et al.*, 2013) to align gene structures.

The analyses focused on comparing position and phase: position is defined by where an intron is inserted, and phase is defined by what the intron interrupts. For example, phase 0 introns interrupt an ORF between two consecutive codons, phase 1 introns are found between the first and second nucleotide of one codon, and phase 2 introns between second and third nucleotide of a codon.

### HGT gene characterization: putative cellular location of proteins

In order to explore whether the proteins coded for by CE1 genes in the fungi *A. muscaria, A. brunnescens, A. thiersii* and *V. volvacea* are secreted, we searched for signal peptide cleavage sites in...
amino acid sequences using Signal P v4.1 (Petersen et al., 2011; http://www.cbs.dtu.dk/services/SignalP/), and predicted the putative cellular location of the proteins using Target P v1.1 (Emanuelsson et al., 2007) and WoLF PSORT (Horton et al., 2000; http://psort.hgc.jp/). We also scanned for transmembrane \( \alpha \)-helices using TMHMM v2.0 (Krogh et al., 2001; http://www.cbs.dtu.dk/services/TMHMM/). We labelled a protein as likely to be secreted if it possessed a signal peptide (Signal P); was predicted to be extracellular (Target P or WoLF PSORT); and had no transmembrane helices.

Structural analyses of a predicted protein

The horizontal acquisition of genes suggests functional relevance for the recipient species. Function may be better predicted by tertiary structure, as opposed to primary sequence (Bajaj & Blundell, 1984; Chothia & Lesk, 1986), and so we reconstructed the tertiary structure of CE1_Am1 (see later Fig. 5) using Phyre2 (Kelley & Sternberg, 2009; http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). We did not target the CE1_Am1 protein for any particular reason, and chose it simply because we named this gene first. Next, we generated a list of proteins with structural similarities to the CE1_Am1 protein model using the Dali server (Holm & Rosenstrom, 2010; http://ekhidna.biocenter.helsinki.fi/dali_server/). A 3D model of the best match, protein PDB 2D80 (Hisano et al., 2006), was retrieved from the Protein Data Bank (PDB; Berman et al., 2000; http://www.rcsb.org). Manipulations, structural alignments and comparisons between the 3D models of CE1_Am1 and PDB 2D80 used PyMOL (Schrödinger, 2010; http://www.pymol.org).

Based on results found for CE1_Am1, and to explore whether other CE1s of EM Amanita may be active esterases, we identified the esterase domains of all EM Amanita CE1s through a search of the Pfam database (http://pfam.sanger.ac.uk) (Punta et al., 2012). We then extracted the coordinates of the domain in each of the genes, and aligned the domains to investigate the degree of conservation of the active site and substrate interacting residues.

Results

Identification of CE1 genes in A. muscaria

We identified a total of 10 CE1 genes within the A. muscaria genome. Eight of the 10 genes are located on scaffold 57 of the JGI assembly, between positions 64 720 and 103 161 (Fig. 1, Table S2). One gene is found on scaffold 120, and the other on scaffold 547. Within the 35-kb region housing the majority of CE1_Am genes, CE1 genes are interspersed with fungal genes (Fig. 1), and there are also five sequencing gaps (not shown). Iterated attempts to sequence these gaps failed, perhaps because the regions are rich in repeats (Hoskins et al., 2007; Cole et al., 2008).

CE1 genes are integrated into the A. muscaria genome

A long-range PCR approach confirmed the physical integration of CE1 genes within the A. muscaria genome. We amplified a section of scaffold 57 housing six CE1_Am, five fungal genes and all gaps (‘Long PCR’, Fig. 1). The PCR successfully generated a fragment spanning from CE1_Am1 to the 3’ end of CE1_Am3; subsequent PCR and sequencing confirmed the presence of all other annotated CE1 and fungal genes expected from the fragment (Table 1; Fig. 1). BLAST searches based on fungal gene sequences confirmed that in all cases the closest match in GenBank is to another fungus (Table S3). The length of genes CE1_Am8, CE1_Am9 and CE1_Am10 was too short to allow confirmation by sequencing (Table S2); however, the sizes of PCR fragments were visualized by agarose gel electrophoresis, and sizes matched expectations (data not shown).

In addition to being structurally integrated into the genome of A. muscaria, CE1 genes appear to be actively expressed. EST data are evidence for transcription of these genes (Fig. 1). CE1_Am1, CE1_Am3, CE1_Am6 and CE1_Am4 are contained within unambiguously mapped assembled transcripts. Transcripts mapped across the region containing CE1_Am5, CE1_Am9, CE1_Am10 and CE1_Am2 often span multiple genes. Those are likely to be mapping artifacts due to the presence of sequencing gaps in this region and high sequence similarity between genes. Further support for expression of CE1 genes in this region is evident from aligned RNAseq data in the JGI genome browser (http://genome.jgi-psf.org/cgi-bin/browserLoad?db=Amamu1&position=scaffold_57:75706-85676).

CE1 genes are common among other Amanita species

In order to investigate the diversity of CE1 genes among other species in the genus, we sought to identify additional CE1 genes from multiple, other Amanita species, and compare the CE1 repertoire of symbiotic and asymbiotic Amanita. Within species with sequenced genomes, one CE1 homologue was found in the genome of A. brunnescens, and four CE1 gene copies were identified in the transcriptome data of A. crenulata (Fig. 2a). By contrast, no homologues of CE1 genes were found in the genome of the sequenced EM fungus A. polyphoramus. Homology BLAST searches using A. muscaria CE1 genes did not retrieve genes in the closely related asymbiotic species A. thiersii or V. volvacea. Instead, highly divergent homologues in these asymbiotic species were found from gene annotations of sequenced genomes; these homologues include CBM1-type carbohydrate binding modules (Fig. 2a). However, the divergence between CE1 genes of symbiotic and asymbiotic Amanita species is apparent even when truncated sequences without CBM1 modules are used as the basis for comparisons.

We also verified the presence of CE1 genes in additional Amanita lacking sequenced genomes, from newly extracted DNA of the sequenced A. muscaria, and from two additional strains of A. muscaria var. guessowii collected from Massachusetts and Pennsylvania, USA (Fig. 2). CE1 genes appear ubiquitous among symbiotic species of subgenus Amanita, although confirmation by Sanger sequencing was not possible for each successful PCR amplification (Fig. 2b). Because our aim was to use sequencing to confirm at least one copy of a CE1 in every species where PCR was successful, we did not attempt cloning and sequencing, or other approaches which would have enabled us to generate
sequences for every PCR fragment. In fact, sequences may be greatly diverged in a subset of species. For this reason we do not interpret unsuccessful amplifications as evidence for the absence of genes. We used CE1_Ab primers to amplify DNA fragments from every species of Amanita subgenus Lepidella, the subgenus housing A. brunnescens. However, we were not able to confirm any of these products by sequencing and therefore did not include this subgenus or these species in Fig. 2; genes may be highly diverged CE1 homologues, or the result may reflect non-specific amplifications.

Phylogenetic analyses suggest CE1s of EM Amanita as HGT

The CE1 sequences of EM Amanita are most similar to sequences of soil bacteria of divergent phyla (Fig. 3, Table S4). The highest sequence identities are to PHB depolymerases in both Alicyclobacillus pohliae (51%; gi|516850840 and gi|516850844, http://www.ncbi.nlm.nih.gov/protein/018130341.1) and K. racemifer DSM 44963 (39–47%; gi|298250533 and gi|297548537) (J. A. Eisen, 2010, unpublished; http://genome.jgi.doe.gov/ktera/ktera.info.html). By contrast, the best-matching CE1 sequences of EM Amanita are only a 32% match to sequences in A. thiersii and 33% match to sequences in V. volvacea.

In order to test the hypothesis that EM CE1 genes were horizontally transferred from bacteria, we conducted a phylogenetic analysis of over 1600 CE1 homologues identified through BLAST searches in NCBI and JGI databases (Table S4). We included CE1 gene products from bacteria, archaea, nonfungal eukaryotes, as well as basidiomycetes and ascomycetes in the analysis. The fungal genes were distributed over four distinct clades (Fig. 3, Notes S1). The first we identify as ‘Fungal Clade I’, the largest fungal clade, which includes the CE1 genes of A. thiersii and V. volvacea, and may represent a clade of genes unique to fungi. Three other clades are interspersed within bacterial lineages. One clade includes both ascomycetes and basidiomycetes (‘Fungal Clade II’), the second groups a set of diverse ascomycetes (‘Fungal Clade III’) and is near Fungal Clade I. The third is the symbiotic Amanita clade (‘EM Amanita’, Fig. 3) and it includes all of the CE1 genes identified from EM Amanita. They form a strongly supported monophyletic clade, embedded within bacterial lineages, with 99% bootstrap support. The CE1 genes present in asymbiotic Amanita may have been lost from symbiotic lineages of the genus. Copies of CE1 genes in symbiotic and asymbiotic lineages are clearly highly divergent. A few other CE1 genes from various other lineages (for example, other eukaryotes) are found scattered in apparently unusual places.

Fig. 3 Phylogenetic evidence of horizontal gene transfer of carbohydrate esterase family 1 (CE1) genes. Lower left, complete CE1 phylogeny (available as a larger format in Supporting Information Notes S1). Inset, Clade containing CE1s in ectomycorrhizal (EM) Amanita. Numbers are bootstrap values above 70; not all bootstrap values shown for larger phylogeny. Black branches, lineages leading to clades where species belong to different groups (Bacteria, Ascomycota, Basidiomycota and Other). Fungal Clade 1 houses saprotrophic (SAP) Amanita and Volvariella sequences, as well as sequences of other fungi.
across the phylogeny (‘Others’, Fig. 3). These are described in detail in Table S5.

CE1 genes are dynamic elements of EM Amanita genomes

A maximum-likelihood reconciliation analysis of the Amanita species tree with all CE1 genes of the EM species A. muscaria, A. crenulata and A. constricta (subgenus Amanita) and A. brunnescens (subgenus Lepidella) suggests a single HGT event followed by a dynamic history of duplications and losses (Figs 2, 3, S1). A conservative inference using only highly supported nodes reveals at least four duplication events, but up to 10 duplications and six losses are possible (Fig. S1). The oldest duplication occurred outside subgenus Amanita. Therefore, the HGT event must have occurred before the split of subgenera Amanita and Lepidella. The CE1 genes of A. muscaria are highly dynamic; four duplications have occurred within this genome alone.

Transferred CE1 genes have been ameliorated in their host genomes

Patterns of CE1_Am and CE1_Ab nucleotide composition and codon usage are highly similar to patterns found in recipient genomes (Table S6). For example, the average GC3 content of fungal genes of A. muscaria scaffold 57 is 0.45, similar to CE1 genes on scaffold 57 (0.46); by contrast, the average GC3 content of CE1s of K. racemifer is high (0.65). Similar trends are observed for codon usage (Table S6).

The numbers and placement of introns in A. muscaria and A. brunnescens are well conserved and distinct from introns of the CE1-CBM1 genes of A. thiersii and V. volvacea (Fig. 4). Numbers of introns in A. muscaria and A. brunnescens CE1 genes range from one to four. The density of introns per gene falls at the lower limit of values for basidiomycete genomes (3.8–5.7 introns per gene; Da Lage et al., 2013), but corresponds well with the median (3) and average (4.5) number of introns per gene in the A. muscaria genome (http://genome.jgi.doe.gov/Amamu1/Amamu1.info.html).

CE1_Am1 shows structural similarity to a PHB depolymerase

In order to explore the potential function of the CE1 genes found in EM Amanita, we predicted and analysed the structure of the inferred protein sequence of CE1_Am1. The predicted structure of CE1_Am1 shows close similarity to the crystal structure of a PHB depolymerase isolated from Penicillium funiculosum (fig. 6, PDB 2D80, Hisano et al., 2006). Despite the strong structural similarities, the sequence of PDB 2D80 is sufficiently diverged from the sequence of CE1_Am1, and the other Amanita CE1s, that it was excluded by our BLAST cutoff in initial analyses. The P. funiculosum protein is therefore not present in our phylogenetic tree (Fig. 3).

The CE1_Am1 protein possesses the typical structural features of extracellular PHB depolymerases: catalytic (320–400 aa), linker (50–100 aa) and substrate-binding (40–60 aa) domains. It was also identified as a putatively secreted protein. The catalytic domain houses a lipase-like catalytic triad (serine, aspartic acid and histidine residues; Fig. 5a). The substrate binding domain in the PHB depolymerase of P. funiculosum possesses 14 binding...
residues and these combine to provide a hydrophobic environment inside a pocket formed on the surface of the enzyme (Hisano et al., 2006); seven of these residues remain conserved in CE1_Am1, and five of the residues differ between the enzymes but remain hydrophobic in CE1_Am1 (Table S7). The binding residues are essential for interaction with PHB chains and define the substrate specificity of the enzyme (Fig. 5b; Hisano et al., 2006). This is likely to be conserved in CE1_Am1, because the same catalytic residues are located within a pocket formed by the binding domain (red spheres in Fig. 5b).

The Dali search also returned good matches to several carboxylesterases, lipases and peptidases (Table S8), reflecting the diversity of enzymatic functions within the CE1 class of carbohydrate esterases. However, PHB depolymerases are structurally very different from carboxylesterases; for example, carboxylesterases lack regions of helices and coils (Fig. S2) typical in PHB depolymerases, and CE1_Am1 shows closer structural conservation to PHB depolymerases (Fig. 5).

Catalytic residues of CE1_Am1 are conserved within the esterase domains of other CE1 genes found in other EM Amanita (Fig. S2). This suggests that all of the CE1s are potentially active, and have the same function: degrading PHB or PHB-like carbon storage molecules.

Discussion

Carbohydrate metabolism genes of bacterial origin in EM Amanita

Phylogenetic analyses identify CE1 genes of bacterial origin within symbiotic Amanita; the same genes are not found in asymbiotic species of the genus. Multiple, independent lines of evidence confirm the A. muscaria genes as integrated within the genome, and expressed. CE1 homologues are found throughout subgenus Amanita, and are also found in A. brunnescens (subgenus Lepidella). The distribution suggests an HGT event around the time of the evolution of the EM niche in Amanita, and before the split of the subgenera. CE1 genes are ameliorated (Lawrence & Ochman, 1997) within host genomes, with nucleotide contents and exon/intron structures typical of basidiomycetes.

We considered and eliminated alternative explanations to HGT. For example, rapid evolution can blur homology relationships among orthologous genes. Although horizontally transferred CE1 genes resemble basidiomycete genes, amelioration does not always denote rapid evolution; amelioration may also mark an ancient transfer event (Lawrence & Ochman, 1997). Moreover, sequence homology and gene structure analyses prove that CE1 genes in symbiotic Amanita are very different from those found in the asymbiotic species A. thiersii and V. volvacea (Notes S2). A scenario in which the CE1 copy in an ancient symbiotic Amanita lineage diverged radically from its homologue in A. thiersii, and convergently evolved to be most similar to bacterial CE1s, is unlikely.

The selective loss of genes in specific lineages may also create incongruent gene and species trees (Aravind et al., 2000). However, a gene loss hypothesis would require an ancient origin of PHB depolymerase-type CE1 genes during the early diversification of eukaryotes, or fungi at least, followed by multiple losses in every lineage except the lineage housing EM Amanita. The scenario is implausible, and not the most parsimonious explanation for observed patterns. Moreover, phylogenies do not suggest a vertical origin from within the fungi, and we are confident that we retrieved a comprehensive set of homologous CE1 genes from available fungal genomes with our BLAST searches.

Although HGT is the most consistent and parsimonious explanation for observed patterns (Fig. 3), available data do not allow unequivocal determination of a donor lineage. The phylogenetic analysis clusters symbiotic Amanita CE1 genes with CE1 genes from both spore-forming soil bacteria from the phylum Firmicutes (A. pobliae and Bacillus megaterium) and with a filamentous soil bacterium from the phylum Chloroflexi (K. racemifer); bootstrap support is strong (99). However, these sequences share at most 51% identity to CE1 genes in symbiotic Amanita, and indicate that either the true donor is absent from genome databases, that gene sequences have changed considerably over time, hindering identification, or that perhaps the donor species is extinct. To identify the direction of HGT we rely on the widely accepted assumption that the taxon of the broadest representation of the gene family is the most likely source (Koonin et al., 2002). In our case, CE1 genes in symbiotic Amanita are embedded within a large bacterial clade, suggesting a bacteria to symbiotic Amanita HGT event.

HGT events among bacteria, preceding the HGT to EM Amanita, may also have obscured the origin of symbiotic Amanita CE1 genes; this idea is supported by the variety of dissimilar taxa found in phylogenetic proximity to the EM Amanita CE1 clade, and also by previous research showing CE1 genes as prone to horizontal transfers (Marcet-Houben & Gabaldon, 2010; Udata et al., 2011). The pattern observed for Fungal
Clade II may reflect an independent HGT event from bacteria to fungi. Marcet-Houben & Gabaldon (2010) searched for prokaryotic-derived HGT in 60 fully sequenced fungal genomes and reported nine putative PHB depolymerases, which they identify as originating from three independent HGT events. Eight of the events reported by Marcet-Houben & Gabaldon (2010) correspond to potential HGT genes in Fungal Clade II (Table S9).

Dynamics of CE1 genes within Amanita

Eight out of the 10 A. muscaria CE1 genes are located on the same scaffold and our phylogenetic analyses suggest that the group is derived from a single, ancestral horizontally transferred gene that was subsequently amplified (Fig. S1). In theory, once there is more than one copy of a gene, the genes can spread more easily, because redundant copies shelter replication errors, enabling duplications (Hurles, 2004). Furthermore, homologues located in close proximity to each other on the chromosome may also promote the formation of unequal crossing over events and result in accelerated gene gain and loss (Li, 1997).

Dynamic, expanded gene families often mark functionally important genes; for example, enzymes involved in the detoxification of insecticides are heavily amplified in exposed species of mosquitoes (Hemingway et al., 1998). Similarly, gene family expansions are a common theme of carbohydrate metabolism genes associated with different fungal niches, including pathogenesis (Soanes et al., 2008; Abramyan & Stajich, 2012) and decomposition (Eastwood et al., 2011). The observed expansion in CE1 genes among symbiotic Amanita lineages at least suggests a critical function; other research also suggests that HGT events are strongly associated with functional genes (Rivera et al., 1998).

Functions of transferred genes

CE1 genes of bacterial origin are only found within EM Amanita, and the CE1 genes are likely to provide some function associated with the symbiotic niche. A test of putative function based on a focal gene, the CE1_Am1, reveals structural conservation between the gene’s protein and an extracellularly secreted PHB depolymerase. Moreover, the catalytic residues of CE1_Am1 are conserved across the EM Amanita CE1s, suggesting that function is conserved across the genes.

In the absence of an exogenous carbon supply, extracellular PHB depolymerases degrade PHB, a microbial carbon and energy storage compound (Dawes, 1988; Jendrossek & Handrick, 2002). Extracellular PHB depolymerases are found in filamentous fungi (McLellan & Halling, 1988; Matavulj & Molitoris, 1992; Lee et al., 2005), but the ecological role of PHB degradation remains largely unexplored. Soils are the habitats with the largest numbers of PHB degrading fungi (Jendrossek & Handrick, 2002).

Based on limited available knowledge, we suggest and briefly discuss three hypotheses for the function of HGT CE1s: CE1s may play a role in carbon metabolism, communication, or defence. First, symbiotic Amanita lack plant cell wall degrading enzymes, and cannot decompose organic substrates. The ability to use extracellular PHB as an alternative carbon source may represent an important adaptation. CE1 genes may enable Amanita species to grow when a symbiosis is not yet established, or when a plant is not providing enough carbon. By contrast, CE1s in V. volvacea match xylanases, plant cell wall degrading enzymes from the CE1 family, both by sequence and structural homology (Ding et al., 2007). Their role in plant cell wall degradation is further supported by the presence of a CBM1, which binds to cellulose and may target the enzyme towards the plant cell wall. CE1s in A. thiersii share close homology to CE1s in V. volvacea and probably perform a similar function.

Second, mycorrhizal symbioses grow in habitats teeming with other organisms, including ‘mycorrhiza helper bacteria’ (MHB; Garbaye, 1994), and CE1s may play a role in signalling. Available evidence suggests that symbiotic Amanita actively communicate with surrounding bacteria: A. muscaria secretes either organic acids or protons capable of modulating the spectrum of antibiotics produced by MHB (Frey-Klett et al., 2007), and a compound produced by Streptomyces sp. AcH505 seems to stimulate the presymbiotic growth of A. muscaria, and simultaneously inhibit the growth of pathogenic fungi (Keller et al., 2007). Third, PHB is able to be depolymerized into water-soluble short-chain fatty acid monomers, and these monomers can act as microbial control agents (Najdegerami et al., 2012). CE1s may play a role in defense. Whatever the function of CE1s, the genes may not be essential for the mycorrhizal niche; the mycorrhizal A. polypyramis does not appear to house PHB depolymerase-type CE1s.

The frequency of HGT and potential for HGT to provide novel metabolic tools (Garcia-Vallvé et al., 2000; Intra et al., 2008; Udatha et al., 2011; this study) may influence thinking on transitions between saprotrophic and mycorrhizal niches. Whether mycorrhizal species can evolve saprotrophy has been debated since at least Hibbett et al. (2000). The most recent evidence points to a history of independent origins of the mycorrhizal habit, with no reversals to saprotrophy (Bruns & Shefferson, 2004; Matheny et al., 2006; Wolfe et al., 2012b). Although the large-scale losses of CAZymes found within the symbiotic Amanita likely preclude a total reversal to an asymbiotic niche (Wolfe et al., 2012b), HGT may endow symbiotic species with novel functions, perhaps including access to alternative carbon sources.

Acknowledgements

We gratefully acknowledge K. Zimmerman for the collection of A. muscaria strain FP01. We thank the National Science Foundation for funding, and a Talent Grant from the Faculty of Mathematics and Natural Sciences of the University of Groningen for enabling M. Chaib De Mares to travel and work in the Pringle laboratory. The work conducted by the US Department of Energy Joint Genome Institute was supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. We are grateful to F. Martin and the Mycorrhizal Genomics Initiative consortium for access to
unpublished genome data. We also thank three anonymous referees for thoughtful, useful reviews that greatly improved our manuscript.

References


FEMS Immunology and Medical Microbiology 46: 8–20.


Supporting Information

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

Fig. S1 Maximum-likelihood reconciliation tree of ectomycorrhizal (EM) Amanita carbohydrate esterases family 1 (CE1) nucleotide sequences.

Fig. S2 Representative portion of a multiple sequence alignment of the esterase domain of carbohydrate esterases family 1 (CE1) genes in ectomycorrhizal (EM) Amanita and Ktedonobacter racemifer.

Methods S1 Detailed protocols for RNA isolation, sequencing libraries preparation, sequencing and assembly of Amanita crenulata’s transcriptome.

Table S1 Carbohydrate esterase family 1 (CE1) gene coordinates in genomes sequenced locally at the Pringle Lab identified through TBLASTN search

Table S2 List of location and Joint Genome Institute (JGI) identifiers of relevant carbohydrate esterase family 1 (CE1) genes used throughout this study

Table S3 Fungal genes flanking carbohydrate esterase family 1 (CE1) genes in scaffold 57

Table S4 Carbohydrate esterase family 1 (CE1) proteins across the tree of life identified through BLAST search, used in phylogenetic analysis

Table S5 Detailed lineages with carbohydrate esterase family 1 (CE1) genes distributed scattered across large phylogeny

Table S6 Nucleotide composition and codon usage bias in carbohydrate esterase family 1 (CE1) genes

Table S7 Conservation of structural features between CE1_Am1 and 2D80

Table S8 Structural similarity search using Dali

Table S9 CE1 genes in Fungal Clade II previously identified as potential horizontal gene transfer (HGT) events

Notes S1 Complete carbohydrate esterase family 1 (CE1) gene phylogeny.

Notes S2 Full alignment used to build the complete carbohydrate esterase family 1 (CE1) gene phylogeny.

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.