New Phytologist

Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal *Amanita*

Maryam Chaib De Mares[1,2]*, Jaqueline Hess[2,3]*, Dimitrios Floudas[4], Anna Lipzen[5], Cindy Choi[5], Megan Kennedy[5], Igor V. Grigoriev[5] and Anne Pringle[6]

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

Author for correspondence:
Maryam Chaib De Mares
Tel: +31 0 503632169
Email: m.chaib.de.mares@rug.nl

Received: *1 July 2014*
Accepted: *24 September 2014*

New Phytologist (2015) **205:** 1552–1564
doi: 10.1111/nph.13140

Key words: *Amanita*, carbohydrate metabolism, comparative genomics, evolution of symbiosis, evolutionary novelty, horizontal gene transfer (HGT), saprotrophy.

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 rumen 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-Sollid, 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 targetting 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.

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

(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. *guessowi* (Koide BX008, Pennsylvania, USA; Hess & Pringle, 2014) has been sequenced twice; one genome is deposited at Joint Genome Institute (JGI, genome.jgi.doe.gov/Amuscaria), and the other was sequenced by the Pringle laboratory (Cambridge, MA, USA). The Pringle laboratory has also sequenced genomes for the EM fungi *A. brunnescens* and *A. polypyramis*, and SAP fungi *A. inopinata* 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. thiersii* is also available through JGI (genome.jgi.doe.gov/Athiersii, 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. thiersii* (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. brunnescens,* where only one

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

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

We screened genomic DNA originally extracted by Wolfe *et al.* (2012b) from 11 species of *Amanita* subgenus *Amanita* (*A. multisquamosa*, *A. subglobosa*, *A. parvipantherina*, *A. praecox*, *A. xylinivolva*, *A. altipes*, *A. crenulata*, *A. wellsii*, *A. murinoflammeum*, *A. umbrinella* and *A. constricta*). We also screened 13 species of *Amanita* subgenus *Lepidella* (*A. flavorubens*, *A. novinupta*, *A. luteolovolvata*, *A. citrina*, *A. rubescens* var. *congolensis*, *A. marmorata*, *A. virosa*, *A. smithiana*, *A. pelioma*, *A. clelandii*, *A. conicobulbosa*, *A. chlorinosoma* and *A. cokeri*).

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 *Phytophtora infestans* (http://my.jgi.doe.gov/general/protocols/). We used a 1 : 1 phenol chloroform ratio in our extractions.

Primers (see the 'Primer Design' section below) were used to test for CE1 genes in each species and strain (Fig. 2). Amplification was performed using a touchdown PCR protocol, to increase sensitivity and specificity, as previously described (Don *et al.*, 1991). The protocol resulted in unique, clear bands. Amplicons were verified as CE1 genes using Sanger sequencing (Genewiz Inc.).

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

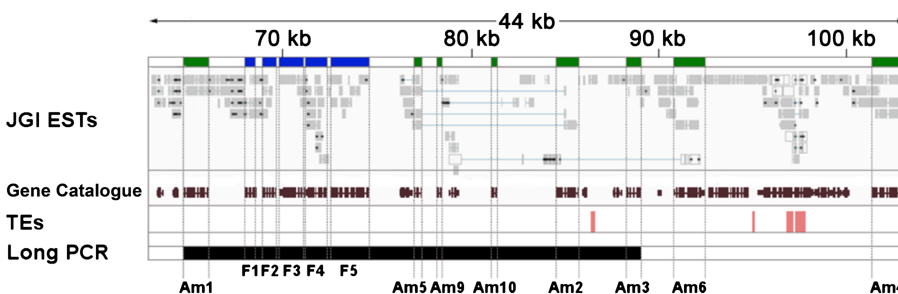

Fig. 1 Physical map of *Amanita muscaria* scaffold 57. Carbohydrate esterase family 1 (CE1) genes with close homology to bacterial CE1 genes shown in green. Blue, fungal genes; brown, gene models; pink, transposable elements (TEs); grey, evidence of gene expression from JGI EST data; black, a long-range PCR confirmed CE1 and fungal genes are linked.

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

Table 1 Primers used to amplify CE1 genes across the *Amanita*

Gene name*	Primer pair	Primer name	Sequence (5′–3′)
CE1_Am1	1	CE1.1.i.fw	CCATGGGTGACTCCTGGAAC
		CE1.1.i.rv	CAGCGCTGTACGTATAGCCA
CE1_Am2	2	CE1.2.i.fw	TGTTTTCGCTGCCATTGGTG
		CE1.2.i.rv	CCAGGAGGCAGCACTATACG
CE1_Am3	3	CE1.3.i.fw	TCTTTGCATCCTGACGTGCT
		CE1.3.i.rv	GGAGGCAGCGCTGTATGTAT
CE1_Am4	4	CE1.4.i.fw	CTTTGCCAAACCAGCTGACC
		CE1.4.i.rv	ATCACTCGGGCTACCTGTCT
CE1_Am5	5	CE1.5.i.fw	TGGAACAACGGCAGAGTTCA
		CE1.5.i.rv	GACAACGACTGGCTTGGGTA
CE1_Am6	6	CE1.6.i.fw	CGCAGCTATGACCGTCATCT
		CE1.6.i.rv	TCCATTAACCAGTGGCGGAC
CE1_Am7	7	CE1.7.i.fw	CCTTCCGGATAATCAGGCCC
		CE1.7.i.rv	ATGTCATTCCAGGAGGCAGC
CE1_Am8	8	CE1.8.i.fw	CGCGGATCTGGTAAGGGAAT
		CE1.8.i.rv	ATGACGGTCATAGCTGCACC
CE1_Am9	9	CE1.9.i.fw	TGGGAAGCCTGTAGTCCAAT
		CE1.9.i.rv	TGTGAAACTTGGACCACGGG
CE1_Am10	10	CE1.10.i.fw	CATGAATCACGCTTGGTCCG
		CE1.10.i.rv	TGGTACGCAATGTCGGTGAA
CE1_Ab	1b	A_brun_CE1.fw	ATACAGGTCCTTCCGGCTCT
		A_brun_CE1.rv1	ATTCCATGATGCCACCGTGT

*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. crenulata* and *A. constricta*. 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 produced 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*

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

Fig. 2 Presence of carbohydrate esterase family 1 (CE1) genes across the phylogeny of *Amanita* and outgroup *Volvariella volvacea* (Wolfe *et al.*, 2012a) based on (a) bioinformatics of sequenced genomes and (b) PCR amplification and Sanger sequencing. (a, b) Phylogenetic trees: green branches, ectomycorrhizal (symbiotic) fungi; grey branches, saprotrophic (free-living) fungi. (a) Empty boxes are definitive proof of the absence of genes from genomes. (b) Dark grey squares, presence of the gene as demonstrated by PCR amplification and Sanger sequencing; medium grey squares, presence of the gene as confirmed by PCR amplification only; light grey squares, negative results, but are not definitive proof of the absence of genes, because in this approach a negative result simply reflects an unsuccessful PCR.

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

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

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 α-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_Am 3; 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. polypyramis*. 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

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

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/WP_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

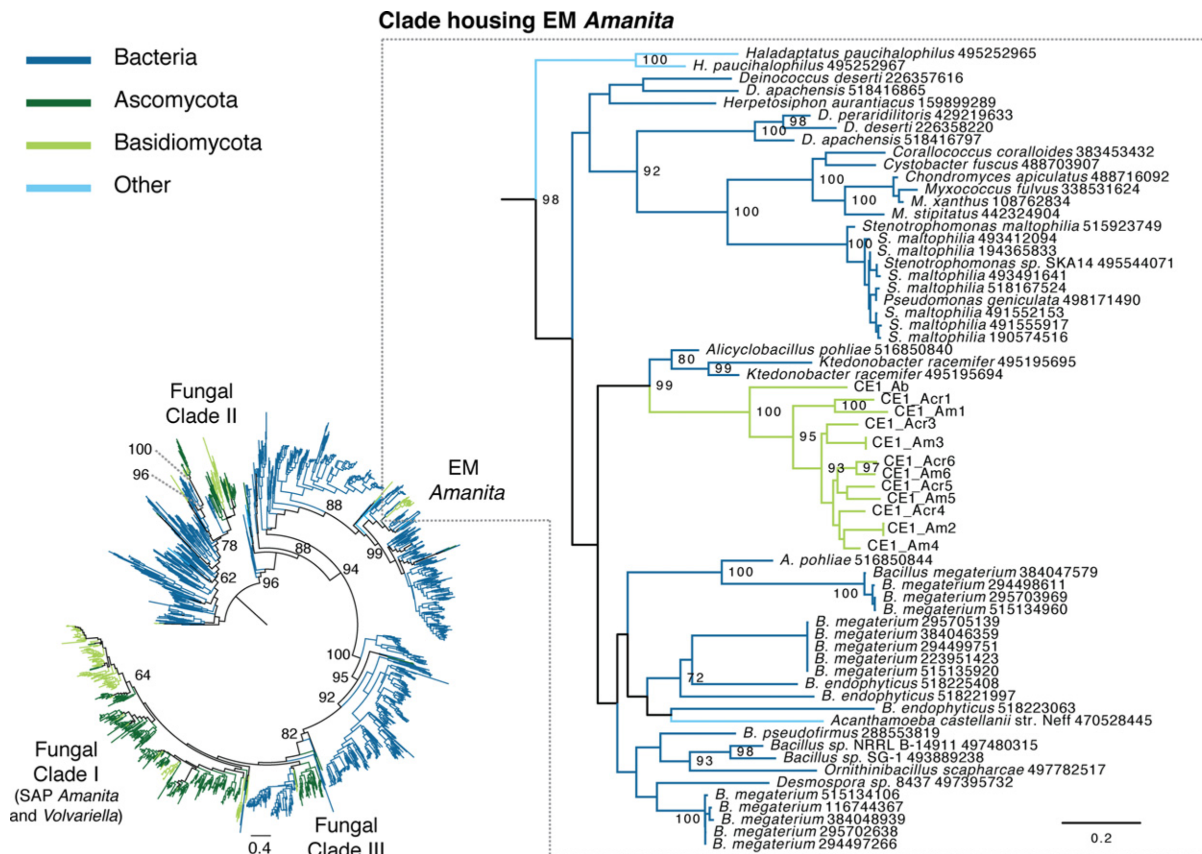

Fig. 3 Phylogenetic evidence of horizontal gene transfer of carbohydrate esterase family 1 (CE1) genes. Lower left, complete CE1 phylogeny (available as a larger format in Supporting Information Notes S1). Inset, Clade containing CE1s in ectomycorrhizal (EM) *Amanita.* Numbers are bootstrap values above 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.

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

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

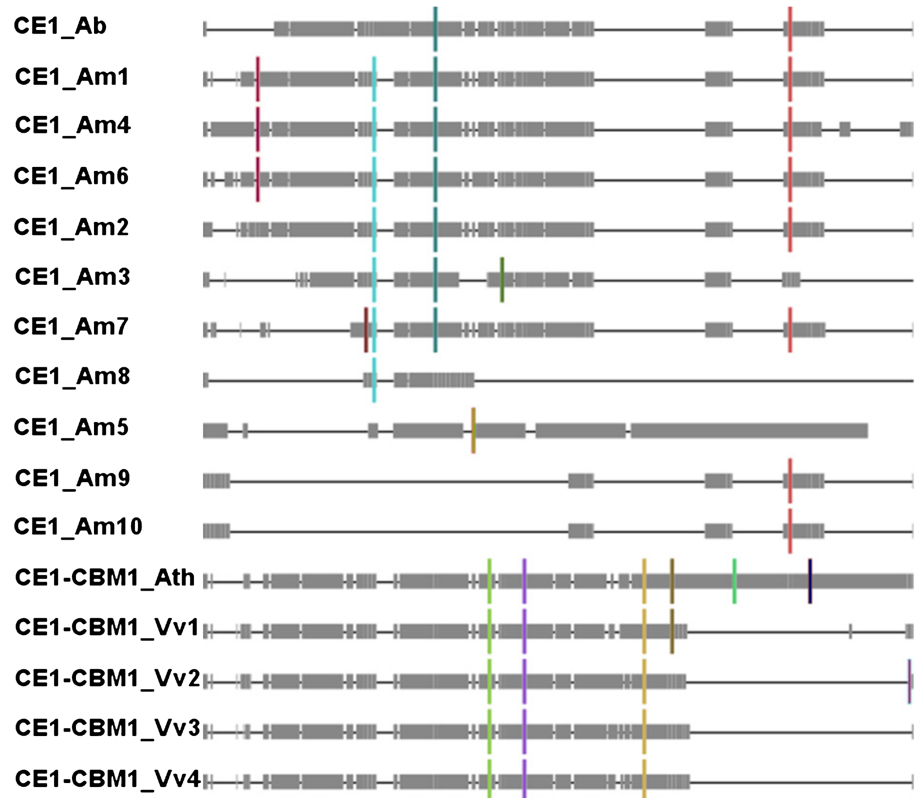

Fig. 4 Gene structure of horizontal gene transfer (HGT) carbohydrate esterase family 1 (CE1) genes in ectomycorrhizal (EM) *Amanita* compared to CE1-CBM1 genes from *A. thiersii* and *Volvariella volvacea*. Grey blocks, aligned exons; coloured bars, different introns. Introns sharing position and phase are shown in the same colour. Ab, *A. brunnescens*; Am, *A. muscaria*; Ath, *A. thiersii*; Vv, *V. volvacea*.

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

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.

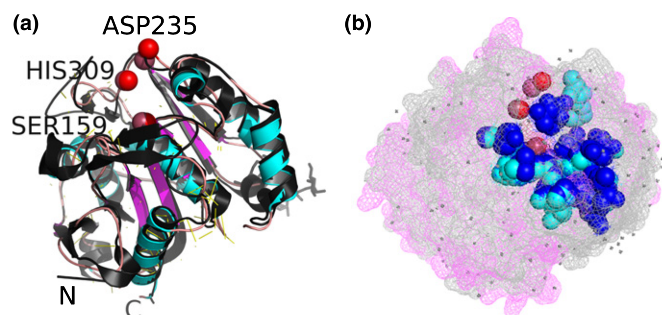

Fig. 5 (a) Alignment between mature forms of poly((*R*)-3-hydroxybutyrate) (PHB) depolymerases of *Penicillium funiculosum* (PDB 2D80) (black), and CE1_Am1 (colour). The match between the proteins is strong, with a root-mean-square-deviation of 1.2 Å and *Z*-score of 43.2. Colours in CE1_Am1 are as follows: cyan, α-helices; magenta, β-sheets; pink, coils. The catalytic triad residues Ser159, Asp235 and His309 of CE1_Am1 are marked with red spheres and overlap with the dark red spheres of residues Ser39, Asp121 and His155 in PDB 2D80. Only names of CE1_Am1 are shown. Unaligned sequence regions are shown in grey. (b) Representation of binding and catalytic sites in CE1_Am1 and PDB 2D80. Protein structures of CE1_Am1 and 2D80 are represented as violet and grey mesh, respectively. Binding sites forming a pocket on the surface of the enzymes are represented: cyan spheres, CE1_Am1; blue spheres, 2D80. Catalytic residues coloured as in (a).

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. pohliae* 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; Udatha *et al.*, 2011). The pattern observed for Fungal

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

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 mosquitos (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

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

unpublished genome data. We also thank three anonymous referees for thoughtful, useful reviews that greatly improved our manuscript.

References

Abramyan J, Stajich JE. 2012. Species-specific chitin-binding module 18 expansion in the amphibian pathogen *Batrachochrytium dendrobatidis*. *MBio* **3**: e00150–12.

Akin DE. 2008. Plant cell wall aromatics: influence on degradation of biomass. *Biofuels, Bioproducts and Biorefining* **2**: 288–303.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Aravind L, Watanabe H, Lipman DJ, Koonin EV. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proceedings of the National Academy of Sciences, USA* **97**: 11319–11324.

Armijos-Jaramillo VD, Vargas WA, Sukno SA, Thon MR. 2013. Horizontal transfer of a subtilisin gene from plants into an ancestor of the plant pathogenic fungal genus *Colletotrichum*. *PLoS ONE* **8**: e59078.

Bajaj M, Blundell T. 1984. Evolution and the tertiary structure of proteins. *Annual Review of Biophysics and Bioengineering* **13**: 453–492.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. *Nucleic Acids Research* **28**: 235–242.

Biely P. 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances* **30**: 1575–1588.

Brown JR, Doolittle WF. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiology and Molecular Biology Reviews* **61**: 456–502.

Bruns TD, Shefferson RP. 2004. Evolutionary studies of ectomycorrhizal fungi: recent advances and future directions. *Canadian Journal of Botany* **82**: 1122–1132.

Bruto M, Prigent-Combaret C, Luis P, Moënne-Loccoz Y, Muller D. 2014. Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes. *Proceedings of the Royal Society B* **281**: 20140848.

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.

Capella-Gutierrez JM, Silla-Martinez TG. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Chaib De Mares M, Hess J, Falcao Salles J, Pringle A. 2013. *A genomic perspective on a transition to symbiosis in ectomycorrhizal* Amanita. MSc thesis, University of Groningen, Groningen, the Netherlands.

Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO Journal* **5**: 823–826.

Christin P-A, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP. 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Current Biology* **22**: 445–449.

Cole CG, McCann OT, Collins JE, Oliver K, Willey D, Gribble SM, Yang F, McLaren K, Rogers J, Ning Z *et al.* 2008. Finishing the human chromosome 22 sequence. *Genome Biology* **9**: R78.

Combier J-P, Melayah D, Raffier C, Gay G, Marmeisse R. 2003. *Agrobacterium tumefaciens*-mediated transformation as a tool for insertional mutagenesis in the symbiotic ectomycorrhizal fungus *Hebeloma cylindrosporum*. *FEMS Microbiology Letters* **220**: 141–148.

Da Lage JL, Binder M, Hua-Van A, Janeček S, Casane D. 2013. Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial α-amylase gene to Basidiomycetes. *BMC Evolutionary Biology* **13**: 40–53.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**: 1164–1165.

Dawes EA. 1988. Polydroxybutyrate: an intriguing biopolymer. *Bioscience Reports* **8**: 537–547.

Ding S, Cao J, Zhou R, Zheng F. 2007. Molecular cloning, and characterization of a modular acetyl xylan esterase from the edible straw mushroom *Volvariella volvacea*. *FEMS Microbiology Letters* **274**: 304–310.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **14**: 4008.

Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M *et al.* 2011. The plant cell-wall decomposing machinery underlies the functional diversity of forest fungi. *Science* **333**: 762–765.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953–971.

Fitzpatrick DA. 2011. Horizontal gene transfer in fungi. *FEMS Microbiology Letters* **329**: 1–8.

Frey-Klett P, Garbaye J, Tarkka M. 2007. The mycorrhiza helper bacteria revisited. *New Phytologist* **176**: 22–36.

Gans J, Wolinsky M, Dunbar J. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**: 1387–1390.

Garbaye J. 1994. Helper bacteria: a new dimension to the mycorrhizal symbiosis. *New Phytologist* **128**: 197–210.

Garcia-Vallvé S, Romeu A, Palau J. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Molecular Biology and Evolution* **17**: 352–361.

Grimaldi B, de Raaf MA, Filetici P, Ottonello S, Ballario P. 2005. *Agrobacterium*-mediated gene transfer and enhanced green fluorescent protein visualization in the mycorrhizal ascomycete *Tuber borchii*: a first step towards truffle genetics. *Current Genetics* **48**: 69–74.

Hammesfahr B, Odronitz F, Mühlhausen S, Waack S, Kollmar M. 2013. GenePainter: a fast tool for aligning gene structures of eukaryotic protein families, visualizing the alignments and mapping gene structures onto protein structures. *BMC Bioinformatics* **14**: 77–88.

Hanssen AM, Ericson-Sollid JU. 2006. SCCmec in Staphylococci: genes on the move. *FEMS Immunology and Medical Microbiology* **46**: 8–20.

Hatje K, Keller O, Hammesfahr B, Pillmann H, Waack S, Kollmar M. 2011. Cross-species protein sequence and gene structure prediction with fine-tuned Webscipio 2.0 and Scipio. *BMC Research Notes* **4**: 265.

Hemingway J, Hawkes N, Prapanthadara L, Jayawardenal KG, Ranson H. 1998. The role of gene splicing, gene amplification and regulation in mosquito insect resistance. *Philosophical Transactions of the Royal Society B* **353**: 1695–1699.

Hess J, Pringle A. 2014. The natural histories of species and their genomes: asymbiotic and ectomycorrhizal *Amanita*. In: Martin F, ed. *Fungi (advances in botanical research)*. San Diego, CA, USA: Academic Press Inc., 235–257.

Hess J, Skrede I, Wolfe EB, LaButti K, Ohm RA, Grigoriev IV, Pringle A. 2014. Transposable element dynamics among asymbiotic and ectomycorrhizal *Amanita* fungi. *Genome Biology and Evolution* **6**: 1564–1578.

Hibbett DS, Gilbert LB, Donoghue MJ. 2000. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature* **407**: 506–510.

Hisano T, Kasuya KI, Tezuka Y, Ishii N, Kobayashi T, Shiraki M, Oroudjev E, Hansma H, Iwata T, Doi Y *et al.* 2006. The crystal structure of Polyhydrohybutyrate depolymerase from *Penicillium funiculosum* provides insights into the recognition and degradation of biopolyesters. *Journal of Molecular Biology* **356**: 993–1004.

Holm L, Rosenstrom P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Research* **38**: W545–W549.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2000. WoLF PSORT: protein localization predictor. *Nucleic Acids Research* **35**: W585–W587.

Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F *et al.* 2007. Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **316**: 1625–1628.

Hurles M. 2004. Gene duplication: the genomic trade in spare parts. *PLoS Biology* **2**: e206.

Intra J, Pavesi G, Horner DS. 2008. Phylogenetic analyses suggest multiple changes of substrate specificity within the Glycosyl hydrolase 20 family. *BMC Evolutionary Biology* **8**: 214–231.

Jendrossek D, Handrick R. 2002. Microbial degradation of polyhydroxyalkanoates. *Annual Review of Microbiology* **56**: 403–432.

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews* **33**: 376–393.

Kanhere A, Vingron M. 2009. Horizontal gene transfers in prokaryotes show differential preferences for metabolic and translational genes. *BMC Evolutionary Biology* **9**: 9–21.

Keller S, Schneider K, Sussmuth RD. 2007. Structure elucidation of auxofuran, a metabolite involved in stimulating growth of fly agaric, produced by the mycorrhiza helper bacterium *Streptomyces* sp. AcH505. *Journal of Antibiotics (Tokyo)* **59**: 801–803.

Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* **4**: 363–371.

Kemppainen MJ, Pardo AG. 2011. Transformation of the mycorrhizal fungus *Laccaria bicolor* by using *Agrobacterium tumefaciens*. *Bioengineered Bugs* **2**: 38–44.

Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology* **55**: 709–742.

Koonin EV, Makarova KS, Wolf YI, Aravind L. 2002. Horizontal gene transfer and its role in the evolution of prokaryotes. In: Syvanen M, Kado CI, eds. *Horizontal gene transfer*. London, UK: Academic Press, 277–304.

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.

Kubicek CP, Seidl V, Seiboth B. 2010. Plant cell wall and chitin degradation. In: Borkovich K, Ebbole DJ, eds. *Cellular and molecular biology of filamentous fungi*. Herndon, VA, USA: ASM Press, 396–413.

Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. *Journal of Molecular Evolution.* **44**: 383–397.

Lawrence JG, Ochman H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences, USA* **95**: 9413–9417.

Lee KM, Gimore DF, Huss MJ. 2005. Fungal degradation of the bioplastic PHB (poly-3-hydroxy-butyric acid. *Journal of Polymers and the Environment* **13**: 213–219.

Li W-H. 1997. *Molecular evolution*. Sunderland, MA, USA: Sinauer Associates.

Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**: 1632–1635.

Löytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**: 579.

Löytynoja A, Vilella AJ, Goldman N. 2012. Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics* **28**: 1684–1691.

Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B *et al.* 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.

Marcet-Houben M, Gabaldon T. 2010. Acquisition of prokaryotic genes by fungal genomes. *Trends in Genetics* **26**: 5–8.

Marchetti A, Parker MS, Moccia LP, Lin EO, Arrieta AL, Ribalet F, Murphy MEP, Maldonado MT, Armbrust EV. 2009. Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature* **457**: 467–470.

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

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.

Matavulj M, Molitoris HP. 1992. Fungal degradation of polyhydroxyalkanoates and a semiquantitative assay for screening their degradation by terrestrial fungi. *FEMS Microbiology Reviews* **103**: 323–331.

Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo JM, Ge Z-W, Yang Z-L, Slot JC, Ammirati JF, Baroni TJ *et al.* 2006. Major clades of Agaricales: a multilocus phylogeny overview. *Mycologia* **98**: 982–995.

McDonald AG, Boyce S, Tipton KF. 2009. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Research* **37**: D593–D597.

McLellan DW, Halling PJ. 1988. Acid-tolerant poly(3-hydroxybutyrate) hydrolases from moulds. *FEMS Microbiology Letters* **52**: 215–218.

Mehrabi R, Bahkali AH, Abd-Elsalam KA, Moslem M, Ben M'Barek S, Gohari AM, Jashni MK, Stergiopoulos I, Kema GH, de Wit PJ. 2011. Horizontal gene and chromosome transfer in plant pathogenic fungi affecting host range. *FEMS Microbiology Reviews* **35**: 542–554.

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.

Najdegerami EH, Tran TN, Defoirdt T, Marzorati M, Sorgeloos P, Boon N, Bossier P. 2012. Effects of poly-β-hydroxybutyrate (PHB) on Siberian sturgeon (*Acipenser baerii*) fingerlings performance and its gastrointestinal tract microbial community. *FEMS Microbiology Ecology* **79**: 25–33.

Nehls U, Zhang C, Tarkka M, Hampp R, Fladung M. 2006. Investigation of horizontal gene transfer from transgenic Aspen to ectomycorrhizal fungi. In: Fladung M, Ewald D, eds. *Tree transgenesis: recent developments*. Berlin, Germany: Springer, 323–333.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA *et al.* 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**: 323–329.

Nikolaidis N, Doran N, Cosgrove DJ. 2014. Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. *Molecular Biology and Evolution* **31**: 376–386.

Odronitz F, Pillmann H, Keller O, Waack S, Kollmar M. 2013. WebScipio: an online tool for the determination of gene structures using protein sequences. *BMC Genomics* **9**: 422.

Peden JF. 2000. *CodonW*. [WWW document] URL http://codonw.sourceforge.net/. [accessed 1 November 2013].

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785–786.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al.* 2012. The Pfam protein families database. *Nucleic Acids Research* **40**: D290–D301.

Richards TA, Leonard G, Soanes DM, Talbot NJ. 2011. Gene transfer into the fungi. *Fungal Biology Reviews* **25**: 98–110.

Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. 2009. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* **21**: 1897–1911.

Rivera MC, Jain R, Moore JE, Lake JA. 1998. Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences, USA* **95**: 6239–6244.

Roberts MC. 2005. Update on acquired tetracycline resistance genes. *FEMS Microbiology Letters* **245**: 195–203.

Roth A, Anisimova M, Cannarozzi GM. 2012. Measuring codon usage bias. In: Cannarozzi G, Schneider A, eds. *Codon evolution: mechanisms and models*. Oxford, UK: Oxford University Press, 198–217.

Schrödinger L. 2010. *The PyMOL molecular graphics system, version 1.5.0.4*. New York, NY, USA: Schrödinger, LLC. [WWW document] URL http://www.pymol.org [accessed 1 November 2014].

Slot JC, Hibbett DS. 2007. Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLoS ONE* **2**: e1097.

Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, Rattray M, Hubbard SJ, Oliver SG, Talbot NJ. 2008. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS ONE* **3**: e2300.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Sun BF, Xiao JH, He S, Liu L, Murphy RW, Huang DW. 2013. Multiple interkingdom horizontal transfers in *Pyrenophora* and closely related species and their contributions to phytopathogenic lifestyles. *PLoS ONE* **8**: e60029.

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14: 178–192.

Towler DA, Gordon JI, Adams SP, Glaser L. 1988. The biology and enzymology of eukaryotic protein acylation. *Annual Review of Biochemistry* 57: 69–97.

Udatha DBRK, Kouskoumvekaki I, Olsson L, Panagiotou G. 2011. The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. *Biotechnology Advances* 29: 94–110.

Weldhagen GF. 2004. Integrons and beta-lactamases – a novel perspective on resistance. *International Journal of Antimicrobial Agents* 23: 556–562.

Wolfe BE, Kuo M, Pringle A. 2012a. *Amanita thiersii* is a saprotrophic fungus expanding its range in the United States. *Mycologia* 104: 22–33.

Wolfe BE, Tulloss RE, Pringle A. 2012b. The irreversible loss of a decomposition pathway marks the single origin of an ectomycorrhizal symbiosis. *PLoS ONE* 7: e39597.

Wright F. 1990. The 'effective number of codons' used in a gene. *Gene* 87: 23–29.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21: 1859–1875.

Wu YC, Rasmussen MD, Bansal MS, Kellis M. 2013. TreeFix: statistically informed gene tree error correction using species trees. *Systematic Biology* 62: 110–120.

Zhang C, Hampp R, Nehls U. 2005. Investigation of horizontal gene transfer in poplar/*Amanita muscaria* ectomycorrhizas. *Environmental Biosafety Research* 4: 235–242.

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) identificators 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.

© 2014 The Authors
New Phytologist © 2014 New Phytologist Trust