Title
Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus Cenococcum geophilum

Permalink
https://escholarship.org/uc/item/8nd1q9w4

Authors
Peter, M
Kohler, A
Ohm, RA
et al.

Publication Date
2016-09-07

DOI
10.1038/ncomms12662

Peer reviewed
Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*

Martina Peter1,*, Annegret Kohler2,*, Robin A. Ohm3,4, Alan Kuo3, Jennifer Krützmann5, Emmanuelle Morin2, Matthias Arend1, Kerrie W. Barry3, Manfred Binder6, Cindy Choi3, Alicia Clum3, Alex Copeland3, Nadine Grisel1, Sajeet Haridas3, Tabea Kipfer1, Kurt LaButti3, Erika Lindquist3, Anna Lipzen3, Renaud Maire1, Barbara Meier1, Sirma Mihaltcheva3, Virginie Molinier1, Claude Murat2, Stefanie Pöggeler7,8, C. Alisha Quandt9, Christoph Sperisen1, Andrew Tritt3, Emilie Tisserant2, Pedro W. Crous6, Bernard Henrissat10,11,12,13, Uwe Nehls5, Simon Egli1, Joseph W. Spatafora14, Igor V. Grigoriev3 & Francis M. Martin2

The most frequently encountered symbiont on tree roots is the ascomycete *Cenococcum geophilum*, the only mycorrhizal species within the largest fungal class Dothideomycetes, a class known for devastating plant pathogens. Here we show that the symbiotic genomic idiosyncrasies of ectomycorrhizal basidiomycetes are also present in *C. geophilum* with symbiosis-induced, taxon-specific genes of unknown function and reduced numbers of plant cell wall-degrading enzymes. *C. geophilum* still holds a significant set of genes in categories known to be involved in pathogenesis and shows an increased genome size due to transposable elements proliferation. Transcript profiling revealed a striking upregulation of membrane transporters, including aquaporin water channels and sugar transporters, and mycorrhiza-induced small secreted proteins (MiSSPs) in ectomycorrhiza compared with free-living mycelium. The frequency with which this symbiont is found on tree roots and its possible role in water and nutrient transport in symbiosis calls for further studies on mechanisms of host and environmental adaptation.

---

1 Swiss Federal Research Institute WSL, Forest Dynamics, Zuercherstrasse 111, 8903 Birmensdorf, Switzerland. 2 INRA, UMR INRA-Université de Lorraine ‘Interactions Arbres/Microorganismes’, Laboratoire d’Excellence ARBRE, INRA-Nancy, 54280 Champenoux, France. 3 US Department of Energy Joint Genome Institute (JGI), Walnut Creek, California 94598, USA. 4 Microbiology, Department of Biology, Utrecht University, 3508 TB Utrecht, The Netherlands. 5 University of Bremen, Botany, Leobenerstr. 2, 28359 Bremen, Germany. 6 CBS-KNAW Fungal Biodiversity Centre, Uppsala Alan 8, 3584 CT Utrecht, The Netherlands. 7 Institute of Microbiology and Genetics, Department of Genetics of Eukaryotic Microorganisms, Georg-August-University Göttingen, 37077 Göttingen, Germany. 8 Göttingen Center for Molecular Biosciences (GZMB), Georg-August-University Göttingen, 37077 Göttingen, Germany. 9 Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109 USA. 10 Centre National de la Recherche Scientifique, UMR 7257, F-13288 Marseille, France. 11 Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille University, F-13288 Marseille, France. 12 INRA, USC 1408 AFMB, F-13288 Marseille, France. 13 Department of Biological Sciences, King Abdullah University, Jeddah 21589, Saudi Arabia. 14 Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331 USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.P. (email: martina.peter@wsl.ch) or to F.M.M. (email: fmartin@nancy.inra.fr).
The ‘symbiosis molecular toolbox’ of a dozen of ectomycorrhizal (ECM) basidiomycetes has recently been characterized by large-scale comparative genomics. These analyses have shown that through contraction and loss of major gene families such as plant cell wall-degrading enzymes (PCWDE), secreted invertases, and toxin synthesis genes, ECM symbionts have become highly reliant upon the availability of a continuous flux of photoassimilates from their host plant, while preserving plant cell integrity by avoiding the release of PCWDE and toxins. Transcript profiling of several mycorrhizal interactions revealed that a significant proportion of the symbiosis-upregulated genes are taxon-specific, restricted to a single mycorrhizal species. Within these taxon-specific orphan genes, symbiotic effectors such as mycorrhiza-induced, small secreted protein (MiSSP), suppress root defense responses and alter root metabolism to accommodate the intruding hyphae and feed them with nutrients as observed for plant pathogens. In contrast to basidiomycetes, only two genomes of ascomycetous ECM fungi, the Périgord black truffle (Tuber melanosporum) and the Hart’s truffle (Elaphomyces granulatus) have been published so far. The ascomycete Cenococcum geophilum Fr. is the most common and globally abundant ECM fungus and often dominates the root systems of trees in the arctic, temperate and subtropical zones and particularly in extreme environments. It forms a characteristic melanized mantle around root tips with darkly pigmented, stringy hyphae emanating in the surrounding soil. The ECM tips are highly resistant to desiccation and are strikingly abundant during soil drought conditions when other ECM species decline, suggesting an important role in drought resistance and resilience of host plant systems.

Figure 1 | The ectomycorrhizal ascomycete Cenococcum geophilum. (a) Phylogenomic placement within Dothideomycetes. The tree including all selected species is shown in Supplementary Fig. 1. (b) An ectomycorrhizal (ECM) root tip of C. geophilum-Pinus sylvestris showing darkly melanized hyphae emanating from the black mantle, and (c,d) longitudinal sections of a C. geophilum-P. sylvestris ECM root tip showing the mantle (m), the Hartig net hyphae (hn) between epidermal (ec) and cortical (cc) root cells as well as the central stele (cs). Scale bars, 1 mm in b; 0.2 mm in c; 50 μm in d.
trees\textsuperscript{5,8}. Although being ubiquitous, the biology of \textit{C. geophilum} is poorly understood. It forms sclerotia as resistant propagules, but no spore-producing structures are known. Studies of fine-scale genetic diversity in \textit{C. geophilum} populations revealed a high level of sequence polymorphism consistent with the occurrence of recombination mechanisms, suggesting that the fungus is reproducing sexually in nature\textsuperscript{9-11}.

The ECM association has arisen repeatedly from ancestral saprotrophs in most clades during fungal evolution leading to a surprisingly similar symbiotic organ and mutualistic relationship with host plants\textsuperscript{1}. \textit{C. geophilum} (Mytilinidiales, Glioniales) is the only ECM symbiont that evolved within the clade of Dothideomycetes\textsuperscript{12}, the largest class of Ascomycota in which sequenced pathogens and saprotophs show a large complement of genes coding for secondary metabolism (SM) and plant-cell-wall degrading enzymes (PCWDE)\textsuperscript{13}. Therefore possibly more of these genes are present in \textit{C. geophilum} than in other ECM fungi, which might have consequences on the fungal/host interaction.

Here we aimed to decipher which features of the mycorrhizal lifestyle and function are imprinted in the genome and symbiotic transcriptome of \textit{C. geophilum} and to compare them with those of the ECM basidiomycetes. We sequenced the genomes of \textit{C. geophilum} and two closely related saprotrophic Dothideomycetes in the Mytilinidiales, \textit{Glionium stellatum} and \textit{Lepidoperterella palustris}, and compared them with other fungi having saprotrophic, pathogenic or mycorrhizal lifestyles. As most ECM fungi, \textit{C. geophilum} has a large genome of 178 megabases due to the proliferation of transposable elements. This genome encodes a lower set of genes coding for PCWDEs compared with saprotrophic and pathogenic Dothideomycetes, although its arsenal of PCWDEs and toxin synthesis genes is larger than those of ECM basidiomycetes. Transcript profiling of ECM root tips indicate a striking upregulation of water channels, sugar transporters and MiSSPs. These findings call for further studies on the mechanisms of host and environmental adaptation during symbiosis.

Results
Genome assemblies and phylogenomic placement. The genome of \textit{C. geophilum} was assembled to 178 Mb, one of the largest fungal genomes ever sequenced, and is predicted to contain 14,748 gene models (Supplementary Table 1 and JGI MycoCosm portal\textsuperscript{14}). The 4-fold smaller 41 Mb and 46 Mb genomes of \textit{Glionium stellatum} and \textit{Lepidoperterella palustris} revealed similar gene counts with 14,362 and 13,870 predicted gene models, respectively. Phylogenomic analyses confirmed the placement of \textit{C. geophilum} in Dothideomycetes, order Mytilinidiales, and the close relationship with the saprotrophic \textit{G. stellatum} (Fig. 1, Supplementary Fig. 1). Intriguingly, \textit{L. palustris}, an aquatic saprotroph, was resolved as closely related to these two species, and thus was included in comparative analyses of Dothideomycetes and other selected species from Ascomycota and Basidiomycota (Supplementary Data 1).

Repeated elements and gene content. The larger genome size of \textit{C. geophilum} compared with its close relatives is explained by its high content (81\%) in repeated sequences, mainly consisting of transposable elements (TE; Fig. 2a, Supplementary Data 2). Many plant pathogenic fungi show an increased content in TE, particularly (hemi)biotrophic ones\textsuperscript{15} and this trend is even more striking for symbiotic mycorrhizal fungi (Supplementary Fig. 2).

There is considerably less variation in gene count than genome size, the former ranging from 9,739 to 14,748 in Dothideomycetes with Mytilinidiales at the upper margin (Fig. 2b, Supplementary Data 3). \textit{C. geophilum} has 2,176 unique genes within 1,422 gene families; the largest families, which hold 5 to 30 members (63 families), contain small (<300 amino acids), mostly functionally unknown and often truncated proteins that show no match to the Pfam domain database\textsuperscript{16} (76\% of families). Three of them are transposon-related (transposase, helicase), among which the third most-abundant family holding 22 proteins, but they are only weakly expressed, as are most of these large families, with 50\% of them showing no expression under any of our experimental conditions (Supplementary Data 4). Similar numbers of unique genes are present in the taxonomically closely related saprotrophic \textit{G. stellatum} and \textit{L. palustris}, the numbers of which range from 48 to 3,152 in Dothideomycetes with fewer unique genes for species with a sequenced close relative\textsuperscript{13} (Fig. 2b, Supplementary Data 3). Expanded gene families in \textit{C. geophilum} are mostly transposon-related or involved in protein–protein interactions and show domains typically found in self/non-self recognition proteins, referred to as somatic incompatibility and defense mechanisms\textsuperscript{17} (for example, HET, NACHT, WD40; Supplementary Data 5). The expression of most of them is not regulated in functioning mycorrhizas.

Although \textit{C. geophilum} is taxonomically more closely related to \textit{G. stellatum} than \textit{L. palustris}, the most recent common ancestors roughly estimated to be 26 and 71 million years ago, respectively (Supplementary Fig. 3), \textit{L. palustris} is more similar to \textit{C. geophilum} with respect to the content in gene categories (Fig. 2c). Although \textit{C. geophilum} and the aquatic \textit{L. palustris} are more similar to the hemiobiotrophic Capnodiales which generally have lower numbers of genes in categories that have been implicated in plant pathogenesis\textsuperscript{13} such as secreted hydrolytic enzymes, the saprotroph \textit{G. stellatum} is more similar to the saprotrophic Hysteriales and necrotrophic Pleosporales (Fig. 2c, Supplementary Data 6). For example, the number of secreted enzymes of \textit{C. geophilum}, \textit{L. palustris} and \textit{G. stellatum} are 595, 588 and 823, respectively, whereas sequenced Capnodiales, Hysteriales and Pleosporales have on average 629, 830 and 894 of these genes in their gene repertoire.

Plant cell wall degrading enzymes and secondary metabolism. The observed division of Mytilinidiales based on the species’ gene contents is most notable for PCWDEs where double hierarchical clustering largely follows the phylogeny of Dothideomycetes but clearly separates Mytilinidiales according to the reduced numbers of these genes in \textit{C. geophilum} and \textit{L. palustris} (Fig. 2d). This reduction in gene content can also be noted for enzymes acting on cellulose, hemicellulose and pectin (Supplementary Data 7). The lytic polysaccharide monoxygenases (LPMO; AA9) for example acting on cellulose and hemicellulose are significantly (Fisher’s exact test \(P=0.001\)) reduced to 5, 2 and on average two members in \textit{C. geophilum}, \textit{L. palustris} and Capnodiales compared with 17 in \textit{G. stellatum} and on average 25 in the Hysteriales/Pleosporales, respectively (Supplementary Data 7). Likewise, as noted previously for the white-rot versus brown-rot modes of wood decay\textsuperscript{18} and for ECM basidiomycetes\textsuperscript{1}, enzymes degrading crystalline cellulose (for example, cellobiohydrolases GH6 and GH7) or modules binding (for example, CBM1) are significantly reduced from 3, 5 and 10 members on average for \textit{G. stellatum}/Hysteriales/Pleosporales to 0, 1 and 2 members for \textit{C. geophilum}/\textit{L. palustris}/Capnodiales, respectively. Xylanases (GH10, GH11), mannanases (GH26) and glucuronidases (GH115) acting on hemicellulose as well as pectin attacking enzymes (PL1, PL3, PL4, CE12) are also reduced from two to five members to no or only one member.

Reduced decay capabilities of plant cell walls are a hallmark of the ECM symbionts in Agaricomycotina\textsuperscript{1} and Tuberae\textsuperscript{2}, a lifestyle feature, which also convergently evolved in...
Dothideomycetes (Fig. 2d, Supplementary Data 7 and 8). Generally, ECM fungi show significantly smaller numbers and less variation of total PCWDEs than saprotrophic and pathogenic fungi (Supplementary Fig. 4). Total numbers range between 6 and 43 with a mean of 19 PCWDEs in ECM fungi, whereas necrotrophic plant pathogen have on average 132 PCWDEs (Supplementary Data 7). Although the total numbers of PCWDE is lower within ECM fungi with even missing families (for example, cellobiohydrolases GH6 and GH7), ECM fungi retained a limited and unique set of these enzymes including several polysaccharide lyases (GH6, GH7, AA9, CBM1) are present in the ECM root tips (for example, GH12, GH5_5, GH5_7, Fig. 3). However, most retained PCWDEs are only weakly expressed in ECM root tips (for example, GH12, GH5_5, GH5_7, GH51, CE8; Supplementary Fig. 5) and are potentially involved in the colonization of the root apoplastic space during Hartig net formation. The expression of most genes related to SM are downregulated in ECM, except for two nonribosomal peptide synthases (NRPS, Cenge3:701851 and Cenge3:471114) and a PKS (Cenge3:706069; Supplementary Data 9). Interestingly, the former NRPS shows a high protein sequence identity (42%) to Pes1 of Aspergillus fumigatus that confers protection against oxidative stress21, which is an inevitable consequence of drought but also imposed by plants as a defense mechanism against biotic stresses22.

**Figure 2 | Genome characteristics of Conococcum geophilum as compared with other Dothideomycetes fungi with diverse lifestyles.** (a) Genome size and repeat element content, (b) number of predicted genes broken down by level of conservation, (c) number of genes in different selected functional categories. (a-c) Are grouped according to phylogeny; (d) represents a double hierarchical clustering of plant cell wall degrading (PCWDE) enzymes, which largely follows the phylogeny of Dothideomycetes but clearly separates the Mytilinidiales according to the reduced numbers of these genes in C. geophilum and L. palustris. Blue-intensity corresponds to the number of gene counts. PCWDEs colour background: Families targeting cellulose (light red), cellulose/hemicellulose (dark red), hemicellulose (light blue), hemicellulose/pectin (dark blue), pectin (grey). Lifestyles: circle colour blue, (hemi)biotrophic plant pathogens; orange, necrotrophic plant pathogens; black, saprotrophs; green, mycorrhiza. For full names of species, see Supplementary Data 1.
### Secondary metabolism enzymes

<table>
<thead>
<tr>
<th>Fungus</th>
<th>PKS</th>
<th>PKS-like</th>
<th>NRPS</th>
<th>PKS-NRPS hybrid</th>
<th>DMAT</th>
<th>Terpene synthase</th>
<th>Total Sec. Met.</th>
<th>Total PCWDE</th>
<th>CBM1</th>
<th>GH3</th>
<th>GH4</th>
<th>FMT (A9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F. fulva</td>
<td>9</td>
<td>3</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>40</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D. septosporum</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>50</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>S. musiva</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>26</td>
<td>64</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>S. populicola</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>32</td>
<td>43</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Z. tritici</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>38</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B. compliaciens</td>
<td>11</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>R. rufum</td>
<td>20</td>
<td>4</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>64</td>
<td>139</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>H. pulicar</td>
<td>23</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>42</td>
<td>113</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>L. palustris</td>
<td>19</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>33</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C. geophilum</td>
<td>9</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>36</td>
<td>43</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>G. stellatum</td>
<td>19</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>42</td>
<td>113</td>
<td>12</td>
<td>5</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>P. nodorum</td>
<td>15</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>38</td>
<td>130</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>P. lingam</td>
<td>12</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>23</td>
<td>111</td>
<td>12</td>
<td>3</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>P. grisea</td>
<td>24</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>53</td>
<td>123</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>B. sorokiniana</td>
<td>16</td>
<td>4</td>
<td>14</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>128</td>
<td>13</td>
<td>1</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>B. onyae</td>
<td>18</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>39</td>
<td>131</td>
<td>11</td>
<td>3</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>B. victoriae</td>
<td>18</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>134</td>
<td>11</td>
<td>3</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>B. zeicola</td>
<td>22</td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>56</td>
<td>132</td>
<td>12</td>
<td>3</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>B. maydiC4</td>
<td>22</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>134</td>
<td>13</td>
<td>3</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>A. brassicicola</td>
<td>22</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>134</td>
<td>13</td>
<td>3</td>
<td>6</td>
<td>24</td>
</tr>
</tbody>
</table>

**Figure 3** | Genes coding for secondary metabolism and selected plant cell wall degrading enzymes in fungi with diverse lifestyles. Circle colour blue, (hemi)biotrophic plant pathogens; orange, necrotrophic plant pathogens; black, saprotrophs; green, ectomycorrhizal and ericoid mycorrhizal (*O. maius*) fungi. PCWD enzymes, plant cell wall-degrading enzymes; PKS, polyketide synthase; DMATS, prenyltransferase; NRPS, nonribosomal peptide synthase. The heat map depicts gene numbers shaded in bright to dark blue according to their number. For full names of species, see Supplementary Data 1.
Figure 4 | Sequence conservation and functional analysis of symbiosis-induced transcripts. The heatmap depicts a double hierarchical clustering of 487 symbiosis-upregulated Cenococcum geophilum genes (rows, fold change >5, FDR-corrected P<0.05). Symbiosis-induced genes were blasted (BLASTP) against the genomes of Glonium stellatum (Glost), Lepidopterella palustris (Leppa), Hysterium pulicare (Hyspu) and Rhytidhysteron rufulum (Rhyru) to find homologues (see Supplementary Fig. 6a and b for a comparison with 50 fungal genomes and Supplementary Fig. 6c and d for downregulated genes). Homologues are coloured from yellow to red depending on the percentage of similarity. Left of heatmap, clusters are highlighted by branches of the same colour. Right of heatmap, the percentages of putative functional categories are given for each cluster as bargrams and the number and percentage of genes in each cluster are shown. Clusters significantly enriched in small secreted proteins (SSPs) are marked with an asterisk (Fisher’s exact test P<0.01). CAZyme, carbohydrate active enzymes; FOLyme, fungal oxidative lignin-degrading enzymes; no KOG, nonsecreted orphan genes.

Table 1 | The most highly upregulated genes in ectomycorrhizal roots of Cenococcum geophilum and Pinus sylvestris compared with free-living mycelium.

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>FLM—means</th>
<th>ECM—means</th>
<th>Fold change</th>
<th>Size (AA)</th>
<th>Definition</th>
<th>Predicted function</th>
<th>Conservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>690706</td>
<td>0.6</td>
<td>3,849.1</td>
<td>6,133</td>
<td>283</td>
<td>Aquaporin (major intrinsic protein)</td>
<td>Transport mechanisms</td>
<td>Conserved</td>
</tr>
<tr>
<td>600722</td>
<td>0.4</td>
<td>550.1</td>
<td>1,453</td>
<td>468</td>
<td>Sugar transporter</td>
<td>Transport mechanisms</td>
<td>Conserved</td>
</tr>
<tr>
<td>647346</td>
<td>5.3</td>
<td>5,666.3</td>
<td>1,073</td>
<td>316</td>
<td>Aquaporin (major intrinsic protein)</td>
<td>Transport mechanisms</td>
<td>Conserved</td>
</tr>
<tr>
<td>697512</td>
<td>0.1</td>
<td>112.5</td>
<td>1,018</td>
<td>287</td>
<td>Unknown</td>
<td>No KOG</td>
<td>Cenge/Glost/Leppa/Hyspu</td>
</tr>
<tr>
<td>654895</td>
<td>1.0</td>
<td>476.0</td>
<td>484</td>
<td>209</td>
<td>Peptidase</td>
<td>Metabolism</td>
<td>Conserved</td>
</tr>
<tr>
<td>655146</td>
<td>3.7</td>
<td>1,560.8</td>
<td>420</td>
<td>195</td>
<td>Unknown</td>
<td>No KOG</td>
<td>Conserved</td>
</tr>
<tr>
<td>730139</td>
<td>0.5</td>
<td>185.9</td>
<td>410</td>
<td>523</td>
<td>MFS transporter</td>
<td>Transport mechanisms</td>
<td>Conserved</td>
</tr>
<tr>
<td>612206</td>
<td>0.3</td>
<td>91.6</td>
<td>344</td>
<td>496</td>
<td>MFS transporter</td>
<td>Transport mechanisms</td>
<td>Conserved</td>
</tr>
<tr>
<td>333290</td>
<td>0.2</td>
<td>51.8</td>
<td>341</td>
<td>165</td>
<td>Unknown</td>
<td>No KOG</td>
<td>Unique</td>
</tr>
<tr>
<td>600293</td>
<td>0.3</td>
<td>97.3</td>
<td>313</td>
<td>121</td>
<td>Unknown</td>
<td>No KOG</td>
<td>Conserved</td>
</tr>
<tr>
<td>698167</td>
<td>2.1</td>
<td>636.3</td>
<td>297</td>
<td>259</td>
<td>Unknown</td>
<td>Small secreted protein (SSP)</td>
<td>Unique</td>
</tr>
<tr>
<td>616643</td>
<td>0.1</td>
<td>36.0</td>
<td>279</td>
<td>93</td>
<td>Unknown</td>
<td>No KOG</td>
<td>Conserved</td>
</tr>
<tr>
<td>660401</td>
<td>0.1</td>
<td>15.6</td>
<td>270</td>
<td>58</td>
<td>Unknown</td>
<td>SSP</td>
<td>Unique</td>
</tr>
<tr>
<td>680403</td>
<td>0.1</td>
<td>21.0</td>
<td>236</td>
<td>135</td>
<td>Unknown</td>
<td>SSP</td>
<td>Unique</td>
</tr>
<tr>
<td>605087</td>
<td>8.9</td>
<td>2,016.8</td>
<td>227</td>
<td>539</td>
<td>MFS transporter</td>
<td>Transport mechanisms</td>
<td>Conserved</td>
</tr>
<tr>
<td>610797</td>
<td>0.2</td>
<td>39.1</td>
<td>225</td>
<td>567</td>
<td>Amino acid permease</td>
<td>Transport mechanisms</td>
<td>Conserved</td>
</tr>
<tr>
<td>608762</td>
<td>0.9</td>
<td>204.9</td>
<td>222</td>
<td>269</td>
<td>Glucose/ribitol dehydrogenase</td>
<td>Poorly characterized</td>
<td>Conserved</td>
</tr>
<tr>
<td>613185</td>
<td>1.6</td>
<td>363.0</td>
<td>222</td>
<td>359</td>
<td>G-protein</td>
<td>No KOG</td>
<td>Overrep Mytil versus Dothis</td>
</tr>
<tr>
<td>676136</td>
<td>2.0</td>
<td>409.9</td>
<td>207</td>
<td>151</td>
<td>Ricin B lectin</td>
<td>CAZyme (CBM13, plant cell wall)</td>
<td>Cenge/Copci</td>
</tr>
<tr>
<td>649427</td>
<td>1.6</td>
<td>325.8</td>
<td>207</td>
<td>248</td>
<td>Unknown</td>
<td>No KOG</td>
<td>Conserved</td>
</tr>
</tbody>
</table>

ECM, ectomycorrhizal roots; FLM, free-living mycelium.

*Values are means of three replicates; given in RPKM (reads per kilobase of transcript per million reads mapped).

Conservation of proteins: protein orthologues only present in mentioned fungal species; Cenge, Cenococcum geophilum; Copci, Coprinus cinereus; Glost, Glonium stellatum; Hyspu, Hysterium pulicare; Leppa, Lepidopterella palustris; Unique, protein family unique to C. geophilum; Overrep Mytil versus Dothis, over-represented gene family in Mytilinidiales compared with other Dothideomycetes; conserved, protein orthologues present in several orders without particular pattern. KOG: EuKaryotic Orthologous Groups (http://genome.jgi-psf.org/help/kogbrowser.jsf).
Aquaporins. Intriguingly, two of the three most highly upregulated and expressed genes in ECM are coding for aquaporins (AQPs) that channel water and/or small solutes across membranes23 (Table 1). This high upregulation seems to be a feature of the C. geophilum symbiotic interaction when compared with other mycorrhizal fungi (Supplementary Fig. 7), although important functions of fungal AQPs in the ECM symbiosis have been shown for other ECM fungi24-26. C. geophilum harbours six putative functional aquaporins in its genome, a similar number to closely related Dothideomycetes relatives (Supplementary Data 11). Four are classified as classical AQPs, which facilitate mainly water transport, one aquaglyceroporin that may also channel small solutes such as urea, glycerol and ammonia, and one X intrinsic protein of which the function is less clear23 (Supplementary Fig. 8). To assess the role of the two symbiosis-induced AQPs (the aquaglyceroporin Cenge3:647346 and a classical AQP Cenge3:690706) as well as an additional classical AQP (Cenge3:604158), we studied both their water transport capabilities and expression patterns in ECM subjected to severe drought and rewatering and in the free-living mycelium under drought. The two symbiosis-upregulated AQPs enabled similar high water permeability of Xenopus laevis oocyte plasma membranes, whereas permeability of Cenge3:604158 was low but still significantly increased compared with buffer-injected controls (Fig. 5a). Interestingly, the two symbiosis-induced AQPs (Cenge3:647346 and Cenge3:690706) were downregulated, whereas the additional classical AQP (Cenge3:604158) was highly upregulated upon high drought stress in ECM of Scots pine (Pinus sylvestris) seedlings (Fig. 5b,c) and in free-living mycelia (Supplementary Fig. 9, statistics in Supplementary Table 2). We followed plant physiological parameters over the drought/rewatering experiment to assess the intensity of the drought treatment (Fig. 5c). After 9 days without watering, the seedlings hardly experienced drought yet (mean shoot water potential, −1 MPa) neither significantly affecting the plants gas exchange nor fungal AQP expression. At high drought stress (shoot water potential of −3.5 MPa), gas exchange ceased in most plants and they started to recover 2 days after rewatering. Similarly, a strong downregulation of the two symbiosis-induced fungal AQPs was observed at high drought stress, whereas the expression of the classical AQP Cenge3:604158 increased upon drought, reaching the highest transcript level of all C. geophilum AQPs (Fig. 5b,c). Transcript levels of the three AQPs returned completely (Cenge3:647346, Cenge3:604158) or almost (Cenge3:690706) to the controlled state 2 days after rewatering. Interestingly, under well-watered control conditions, the transcript levels of the drought-induced, classical AQP (Cenge3:604158) correlated best with the shoot water potential of their host plant (Pearson correlation 0.70), whereas the symbiosis-induced classical AQP significantly positively correlated with the nitrogen content of the needles (0.62) and the plant’s water use efficiency (0.60; Supplementary Fig. 10).

In the drought/rewatering experiment, we assessed the plant conditions of mycorrhizal versus non-mycorrhizal (NM) plants. We observed significantly increased needle nitrogen content, net photosynthesis and water use efficiencies in ECM compared with NM plants, confirming the mutualistic fungus–plant relation. On the other hand, no significant effect of mycorrhizal inoculation was observed in the drought treatment (Supplementary Fig. 11, Supplementary Table 3).

Symbiosis-regulated plant genes. To identify specific gene networks induced in host roots by C. geophilum, we inoculated seedlings of Scots pine simultaneously with C. geophilum and either Suillus granulatus or Rhizopogon roseolus, two common ECM fungi associated to pines27. We then measured the differential expression of Scots pine genes in the respective mycorrhizas using oligoarrays. Although 8% of these genes were differentially regulated in mycorrhizal compared with NM roots (P < 0.05; posterior probability of differential expression28 PPDE > 0.97, Supplementary Data 12), the expression of 1% significantly differed in ECM root tips challenged by C. geophilum compared with the other two symbionts (Fig. 6 and Supplementary Data 13). Several interesting patterns emerged from these data: (1) Genes for membrane transporters, such as a the SWEET sugar exporters (11.1 × fold change in C. geophilum interacting ECM tips compared with NM tips versus 4.2 × and 1.7 × in S. granulatus and R. roseolus ECM tips compared with NM tips, respectively; 3.7 × fold change C. geophilum versus other symbionts, PPDE 0.99), ammonium and phosphate transporters (1.6–5.2 × fold change in C. geophilum ECM versus other symbionts) and a NOD26 AQP (5.6 × fold change in C. geophilum ECM versus other symbionts, PPDE 0.99) were more highly upregulated in C. geophilum ECM than in ECM root tips of the other two fungi possibly indicating a higher nutrient and water transfer. (2) A K+ efflux transporter SKOR, which is induced by stressors and reactive oxygen species mediating K+ export in xylem vessels and which is thought to be involved in stress signalling29, was expressed 18.4-fold higher (PPDE 0.98) in ECM of C. geophilum than in the other ECM fungi. (3) Many oxidative stress- and pathogen resistance-related proteins such as peroxidases, anthocyanin biosynthesis30 and SOS interacting proteins31 were upregulated in C. geophilum ECM compared with NM root tips and downregulated in roots tips interacting with other ECM fungi (3.5–5.6 × fold change in C. geophilum versus other symbionts). Most remarkably, the gene showing the second highest differential expression (20.4 × fold change, PPDE 1.00) codes for an activated disease resistance (ADR1)-like protein, which is usually induced by plant pathogens and conveys broad-spectrum disease resistance32. Enhanced expression of this gene leads to drought tolerance in Arabidopsis, which suggests that interlinked networks of abiotic and biotic stress signalling exist and may have significant functional overlap32.

Presence of sex-related genes in C. geophilum. A puzzling finding of ECM field surveys is the observed global distribution and high genetic diversity of C. geophilum, a presumably asexual soil fungus relying on vegetative mycelial propagation9,10. We therefore assessed whether C. geophilum has the prerequisites to have sex and to produce fruiting bodies. The genome of C. geophilum strain 1.58 contains a MAT locus, which is conserved with its close relative G. stellatum (Supplementary Fig. 12). It includes one intact mating-type gene (MAT1-1-1) of the two typically present in the bipolar mating system of ascomycetes (MAT1-1-1, and MAT1-2-1), indicating that C. geophilum is self-sterile33. Genes encoding proteins of the pheromone response pathway and involved in fruiting body
production are equally present in *C. geophilum* and *G. stellatum* (Supplementary Data 14), indicating that *C. geophilum* is possibly able to reproduce sexually by outcrossing and to differentiate inconspicuous fruiting bodies, maybe similar to those produced by *G. stellatum*\(^12\).

**Discussion**

The *C. geophilum* genome contains many features of the symbiosis molecular toolbox found in the ECM fungi sequenced so far, that is, loss of genes involved in PCW degradation and presence of a large set of taxon-specific, symbiosis-upregulated orphan genes, including MiSSPs\(^1\). These findings support the independent convergent evolution of this ecology in Dothideomycetes, which diverged from other fungal classes approximately 280 million years ago\(^12,13\) and which is the largest class of ascomycetes. *C. geophilum* is the only known ECM member in this class comprising over 19,000 species with an unparalleled diversity of lifestyles\(^13\) and including many devastating plant pathogens. This prompted the genome sequencing of >100 Dothideomycetes providing an excellent opportunity for comparative genomics to identify the peculiarities of the ECM lifestyle.

One of the most obvious and outstanding features of the *C. geophilum* genome is its very large size resulting from ancient TE bursts (Fig. 2). Mycorrhizal fungi, such as the ECM *Tuber melanosporum* and the arbuscular mycorrhizal *Rhizophagus irregularis*, generally show an increased proportion of TE, as do many biotrophic pathogens\(^34\). However, several ECM fungi, such as *Hebeloma cylindrosporum* and *Amanita muscaria* (Supplementary Data 2) did not experience massive TE proliferations, suggesting this is not a prerequisite of the biotrophic lifestyle. Genome invasion by TEs might explain the success of *C. geophilum* to colonize a wide range of hosts and habitats\(^9\) since TE are known to contribute to the adaptability of fungi to their environment\(^15,35,36\). The high content in TE might partly explain the high genetic diversity and cryptic speciation observed in *C. geophilum* populations\(^9\)–\(^11\). On-going genome re-sequencing of geographic isolates of

---

**Figure 5** | Functional characterization of three selected aquaporins of *Cenococcum geophilum*. Two aquaporins (AQP) were highly symbiosis-induced (*Cenge3*:647346, *Cenge3*:690706), whereas *Cenge3*:604158 was not regulated by the symbiosis. (**a**) Water permeability of *Xenopus laevis* oocytes expressing *C. geophilum* AQPs as compared with ND96-buffer injected oocytes (negative control) or oocytes expressing a Laccaria bicolor AQP (positive control; mean \(P \) values, \(n = 20 \pm s.e.m\.). Compared with the ND96-buffer injected control, all three *C. geophilum* AQPs significantly increased the plasma membrane water permeability of oocytes (\(P \) values of Student’s \(t\)-test: 1.5E – 51, 4.7E – 55, 1.6E – 5, respectively). (**b, c**) Impact of drought and rewatering on the fungal AQP expression in mycorrhizal root tips of Scots pine (*Pinus sylvestris*)/*C. geophilum* seedlings grown in pots in a climate chamber experiment. (**a**) Gene expression of the three AQPs at three time points (t1, 10 days of drought; t2, 14 days of drought; t3, 3 days after rewatering) in ECM root tips of Scots pine seedlings grown in pots in a climate chamber experiment. (**b**) Stomatal conductance (mol H\(_2\)O m\(^{-2}\) × s\(^{-1}\) ) and net photosynthesis (μmol CO\(_2\) m\(^{-2}\) × s\(^{-1}\) ) of Scots pine seedlings measured at the three time points (means of four replicates (pots) \(\pm s.e.m\.)). Asterisks indicate significance of the drought treatment (\(P < 0.05\), analysis of variance). (**c**) Shoot water potential, stomatal conductance and net photosynthesis of the Scots pine seedlings measured at the three time points (means of four replicates (pots) \(\pm s.e.m\.)).
C. geophilum should provide additional information about the role of TE in genetic structuration of the symbiont’s populations.

There is considerably less variation in gene count than genome size in Dothideomycetes as previously noted\textsuperscript{13} (Fig. 2). However, gene content varies widely, in particular in categories found to be involved in pathogenesis such as those coding for plant-cell-wall degrading and secondary metabolism enzymes\textsuperscript{37}. Most remarkably, this can be noted within Mytilinidiales, where a reduced number of PCWDE separates C. geophilum and the aquatic saprotroph L. palustris from the saprotroph G. stellatum (Fig. 2), the latter likely showing the ancestral state of this family placed within Pleosporomycetidae\textsuperscript{39}. A reduced complement of such genes is meaningful for (hemi)biotrophic pathogens to avoid triggering plant defense mechanisms, one of the suggested evolutionary adaptations to the biotrophic lifestyle\textsuperscript{13,34} in both, pathogenic and mutualistic interactions\textsuperscript{39}. Although ECM fungi show consistently reduced numbers of PCWDEs, (hemi)biotrophic pathogens are more variable (Supplementary Fig. 4), which might depend on their strategy to enter plant tissues and cells, evade plant defense mechanisms and switch between biotrophic and necrotrophic phases. Saprotrophs benefit from a large repertoire of PCWDEs to degrade their strategy to enter plant tissues and cells, evade plant defense mechanisms and switch between biotrophic and necrotrophic phases. Saprotrophs benefit from a large repertoire of PCWDEs to degrade their strategy to enter plant tissues and cells, evade plant defense mechanisms and switch between biotrophic and necrotrophic phases. Saprotrophs benefit from a large repertoire of PCWDEs to degrade their strategy to enter plant tissues and cells, evade plant defense mechanisms and switch between biotrophic and necrotrophic phases.

### Figure 6 | Scots pine differentially interacts with Conococcum geophilum and other simultaneously inoculated mycorrhizal fungi. (a) The 20 most highly up- or downregulated Scots pine (Pinus sylvestris) genes expressed in ectomycorrhizas (ECM) of double-inoculated seedlings (CR, C. geophilum and Rhizopogon roseolus; CS, C. geophilum and Sullus granulatus; three replicate plants each) when interacting with C. geophilum (Cg) compared with S. granulatus (S) or R. roseolus (R). Expression of genes in non-mycorrhizal plants (NM; two replicate plants) is also indicated. The heat map depicts a double hierarchical clustering in which normalized gene expression values are coloured from red (up-) over black (mean expression over all conditions) to blue (downregulated). (b) Principal component analysis (PCA) based on the expression of 7,000 P. sylvestris genes shows that C. geophilum interacting ECM root tips (filled circles) are separated on the first axis from the other two ECM fungi (R. roseolus, open triangles and S. granulatus, open circles). Each symbol per fungal species corresponds to a plant replicate. Note the sketch showing a pine seedling colonized by C. geophilum (left root) and one of the other fungi (right root).
Having arisen from functionally diverse wood- and litter-decomposer saprotrophs, ECM fungi retained distinct arrays of PCWDEs1,2,4,38 and since residing within Pleosporomycotina39, the rather high fraction of such genes in *C. geophilum* does not come as a surprise. Those expressed in ECM roots (Supplementary Fig. 5) are likely involved in the reconstruction of the middle lamella accompanying the formation of the Hartig net. However, most of the retained PCWDEs are not expressed in ECM, neither are the genes related to secondary (toxin) metabolism. They therefore likely play a role in developmental stages other than those involved in the mutualistic fungus-plant interaction in vital and functioning ECM. For example, they might participate at later ECM stages in root turnover after senescence. Still, the increased expression of stress- and pathogen resistance-related proteins (for example, receptor-like protein kinase, NBS-LRR disease proteins) seen in the plant hosts transcriptome when associated with *C. geophilum* compared with the ECM basidiomycetes *S. granulatus* and *R. rooseli* (Fig. 6, Supplementary Data 13) indicate that these pathogenic-related features might affect the mutualistic interaction to some extent. The most highly regulated plant genes with known functions, the activated defense response gene (*ADR1*), increases drought resistance in *Arabidopsis thaliana* possibly by linking signalling networks of biotic and abiotic stress32. We therefore speculate that *C. geophilum* colonization primes pathways conferring drought tolerance as suggested previously for other organisms32.

The striking upregulation of genes coding for taxon-specific SSPs in symbiosis confirms previous observations1,41 and highlights the importance of these proteins in mutualistic interactions. The role of such candidate effector proteins has been extensively reported in pathogenic systems34,35 and they have recently been characterized in mycorrhizal mutualists3,34,34. Besides the MiSSP effector secreted by the ECM *Laccaria bicolor*, the key role of secreted effectors of the arbuscular mycorrhizal fungus *Rhizopogon irregularis* and the beneficial endophyte *Piriformospora indica* in subverting plant immunity has been confirmed34, but many MiSSPs await further characterization. Analysis of the role of *C. geophilum* MiSSPs in the symbiotic interaction is currently underway. Because of the low protein sequence conservation of secreted effector-like proteins, their function cannot be inferred by comparisons with known genes of taxonomically related pathogenic relatives. There are indications that in spite of sequence divergence, functional redundancy of taxonomically related pathogenic relatives is already available13,37. For the *C. geophilum* genome sequencing, two Illumina fragment and three long mate pair (LMP) libraries have been sequenced, assembled with AllPath-LG49, and further improved with Pacific Biosciences sequencing as follows: For fragment libraries, 500 ng of genomic DNA was sheared using the Covaris E210 (Covaris) and size-selected using Agencourt Ampure Beads (Beckman Coulter). The DNA fragments were treated with end repair, A-tailing and adapter ligation using the TruSeq DNA Sample Prep Kit (Illumina) and purified using Agencourt AMPure Beads. Two types of LMP libraries were prepared. LFPE (ligation-free paired end) mate pair fragments were generated using the 5500 SOLiD Mate-Paired Library Construction Kit (SOLiD). Fifteen micrograms of genomic DNA was sheared using the Covaris g-TUBETM (Covaris) and gel size selected for 4–4.5 kb. The sheared DNA was end repaired, A-tailed and ligated with biotinylated Illumina compatible adapters containing IDT (IDT). The adapter ligated DNA fragments were circularized via recombinaction by aCre excision reaction (NEB). The circularized DNA was digested using four base cutter restriction enzymes followed by self-ligation (NEB). The self-ligated products were immobilized on beads and released PCR was used to enrich for the final library. The libraries were quantified using KAPA Biosystem’s next-generation sequencing library quantitative PCR (qPCR) kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v3, and then generate a clustered flowcell. Sequencing of the flowcell was performed on the Illumina HiSeq2000 sequencer using a TruSeq SBS sequencing kit 200 cycles, v3, following a 2 x 100 and 2 x 150 indexed run recipe for LFPE and fragments, respectively. Genomic data from five libraries were filtered and assembled with AllPath-LG. To fill gaps in Illumina genome assembly of *C. geophilum*, an unamplified library was generated using Pacific Biosciences standard template preparation protocol, where 5 μg of gDNA was used to generate the library and the DNA was sheared using a Covaris LE220 focused-ultrasonicator with their Blue 3.0 kb miniTUBES to generate sheared fragments of 3 kb in length. The sheared DNA fragments were then prepared according to the Pacific Biosciences protocol and using the SMRTbell template preparation kit, where the fragments were treated with DNA damage repair, had their ends repaired so that they were blunt-ended, and 5’ phosphorylated. Pacific Biosciences hairpin adapters were then ligated to the fragments to create the SMRTbell template for sequencing. The SMRTbell templates were then purified using exonuclease treatments and size-selected using AMPure PB beads. Sequencing primer was then annealed to the SMRTbell templates and Version C2 sequencing polymerase was bound to them. The prepared SMRTbell template...
libraries were then sequenced on a Pacific Biosciences RS sequencer using Version C2 chemistry and running 2 * 45 min movies per SMRT cell.

Leptotyphlassa lemmata was sequenced at the AFTOL project using one fragment and one LMP libraries. For the fragment library, 100 ng of DNA was sheared to 270 bp using the Covaris LE220 (Covaris) and size selected using SPRI beads (Beckman Coulter).

The fragments were treated with end-repair, A-tailing and ligation of Illumina compatible adaptors (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA). A high-quality resulting 554 fragments were used for library construction using the 5500 SOLiD Mate-Paired Library Construction Kit (SOLID) as described for C. geophilum. The purified fragments were ligated with Illumina adaptors and amplified using eight cycles of PCR with Illumina primers (Illumina) to generate the final library. The library was quantitated using an Agilent 2100 Bioanalyzer (Agilent) and prepared for sequencing on the SOLiD platform using a SOLiD sequencing kit 100 cycles, v3, at the Core Facilities of the Center for Genome Research and Biocomputing at Oregon State University. De novo genome assembly of filtered sequences was performed using Velvet0.8. The resulting contigs were manually inspected by motive analysis and all models of contigme-scale phylogenetic analyses were conducted with the HAL pipeline54. Briefly, orthologous clusters (OC) were identified using MCL across inflation parameters 1, 2, 3 and 5. OCs were filtered to retain non-redundant, single-copy clusters (one protein per genome) with a minimum of 75% representation of the sampled genomes. The protein sequences resulting from the filtered clusters were aligned individually in MUSCLE with default settings, after which poorly aligned regions identified using Gblocks (using liberal setting) and excluded. Aligned proteins were concatenated into a superalignment that included 119,236 amino acid positions and analysed with RAxML v 7.2.6 using the PROTGAMMAWAG model of evolution with 100 bootstrap replicates.

**Estimated ages of the most recent common ancestors.** A chronogram for the 51 taxa RAxML phylogeny was constructed using R8S V1.8 (ref. 55). Estimated mean ages for the most recent common ancestors of Ascomycota (518 MYA) and Basidiomycota (521 MYA) from Floudas et al.56 were used as calibration points. Calibration points were cross-validated as follows: method = PL (penalized likelihood) cross-validation; y = yes for retention; d = 6. These estimation methods were used for 400 bp through agarose gel isolation. The fragments were treated with end-repair, A-tailing and ligation of Illumina compatible adaptors (IDT, Inc) using the NEBNext kits (New England BioLabs). The purified fragments were ligated with Illumina adaptors and amplified using 16 cycles of PCR with Illumina primers (Illumina) to generate the final library. The library was quantitated using an Agilent 2100 Bioanalyzer (Agilent) and prepared for sequencing on the Illumina platform using a TruSeq sequencing platform utilizing a TruSeq paired-end cluster kit v3. The library was sequenced one lane of a flowcell on an Illumina HiSeq2000 sequencer using a TruSeq SBS sequencing kit 100 cycles, v3, at the Core Facilities of the Center for Genome Research and Biocomputing at Oregon State University. De novo genome assembly of filtered sequences was performed using Velvet0.8, with a minimum contig length of 300 bp. For genome annotations, RNA of mycelia grown in liquid culture with different growth conditions as well as from different tissues for C. geophilum (sclerotia and mycorrhizas) were sequenced after total RNA extraction (RNasey Plant Mini Kit of Qagen with integrated DNAase step) using the Illumina platform. The resulting cDNA libraries (C. geophilum and L. palustris) were sequenced on the Illumina HiSeq2000 genome sequencer at the Beckman Coulter Genomics facilities (g. stellatum). C. geophilum messenger (mRNA) was purified from total RNA using Dynabeads mRNA Purification Kit (Invitrogen) and chemically fragmented to 200–250 bp (Ambion). mRNA was reverse transcribed with SuperScript II using random hexamers. Second-strand complementary DNA (cDNA) was synthesized using dNTPs/DTTP mix (Thermo Scientific), Escherichia coli DNA Ligase, E. coli DNA polymerase I, and E. coli RnaSEh (Invitrogen). The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation using the TruSeq Sample Preparation Kit (Illumina). Second-strand cDNA was removed by AmpErase UNG (Amplysystems) to generate stranded cDNA libraries for L. palustris were generated using the Illumina TruSeq Stranded RNA LT kit. mRNA was purified from 1 μg of total RNA using magnetic beads containing poly-T oligos. The mRNA was fragmented using divalent cations at high temperatures, and the resulting RNA was reverse transcribed using random hexamers and SSII (Invitrogen) followed by second-strand synthesis. The fragmented mRNA was treated with end-pair, A-tailing, adaptor ligation and 10 cycles of PCR. Both the libraries were quantitated using KAPA BioSystem’s next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. These were prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit v1, v3 and v4 for L. palustris and g. stellatum, respectively, and Illumina’s cBOT instrument to generate a clustered flowcell for sequencing. Sequencing of the flowcell was performed on the Illumina HiSeq2000 sequencer using a TruSeq SBS sequencing kit 200 cycles, v1 following a 2 x 100 indexed run recipe for L. palustris and v3 with 2 x 130 reads for C. geophilum. For G. stellatum, library construction and 2 x 100 bp Illumina HiSeq sequencing was performed by Beckman Coulter Genomics (Danvers, MA, USA) using their standard protocols.

**Phylogenomic placement of Conococcum geophilum.** For phylogenomic placement, we selected 48 additional species (Supplementary Data 1) of which the genomes were available on the basis of the following criteria: all published Dothideomycetes, one to five members of other ascomycetous orders and eight Basidiomycota from five orders selected to cover major orders as well as based on the diversity of lifestyles (biotrophic/hemibiotrophic/necrotrophic pathogens, white rot/brown rot/litter/aquatic saprotrophs and ericoid/ECM species). Genome-scale phylogenetic analyses were conducted with the HAL pipeline54. Briefly, orthologous clusters (OC) were identified using MCL across inflation parameters 1, 2, 3 and 5. OCs were filtered to retain non-redundant, single-copy clusters (one protein per genome) with a minimum of 75% representation of the sampled genomes. The protein sequences resulting from the filtered clusters were aligned individually in MUSCLE with default settings, after which poorly aligned regions identified using Gblocks (using liberal setting) and excluded. Aligned proteins were concatenated into a superalignment that included 119,236 amino acid positions and analysed with RAxML v 7.2.6 using the PROTGAMMAWAG model of evolution with 100 bootstrap replicates.

**Comparative analyses and annotation of functional categories.** Multigene families were predicted from 567,845 gene models of the studied 51 species using the MCL algorithm implemented in the JGI clustering pipeline as described in Ohm et al.13. Multigene families were analysed for evolutionary changes in protein family size using the CAFE program57. To allow comparing specific functional categories among the selected 51 genomes, the predicted gene models were functionally annotated using the same pipeline. These gene categories included: repeated elements, CAZymes, secreted proteins and genes from the secondary metabolism. RepeatScout58 was used to identify de novo repetitive DNA in genome drafts. The default parameters (with l = 15) were used. RepeatScout generated libraries of consensus repeated sequences that were then filtered as follows: (1) the sequences less than 100 bp were discarded; (2) repeats having less than 10 copies in the genome were removed (as they may correspond to protein-coding gene families) and (3) repeats having significant hits to known proteins in Uniprot (http://www.uniprot.org/) other than proteins known as belonging to TE were removed. The consensus sequences remaining were manually annotated by a TRASTX search against Repbase version 19.05 (http://www.girinst.org/rebase/index.html). Using the final libraries of repeated elements, the number of repeat occurrences and the percentage of genome coverage were assessed by masking the genome assemblies using RepeatMasker (http://www.repeatmasker.org). To identify full-length long terminal repeats (LTR) retrotransposons, a second de novo search was performed with LTR_STRUc59. The putative full-length LTR retrotransposons were annotated annotated by a TRASTX search against RepBase (http://www.girinst.org/rebase/index.html) to identify those belonging to copia LTR and to gypsy LTR. MISA (http://grc.ipk-gatersleben.de/misa/download/misa.pl) with default parameters was used to identify mono- to hexanucleotide simple sequence repeat (SSR) occurrences and the default parameters for the basic repeaterfinder60 was used to identify minisatellites (motif between 7 to 100bp) and satellites (motif above 100bp) with the following parameters: 2, 7, 8, 10, 15, 30, 500 (Supplementary Data 2). Annotation of carbohydrate active enzyme (CAZyme)) genes was performed using the CAZY database (www.cazy.org) annotation pipeline61. To assess differences in the numbers of genes present using CAZY families among phylogenetic groups or lifestyles, we applied non-parametric median tests in SPSS (version 23, IBM) with false discovery rate (FDR) correction. The secretion was identified using a custom pipeline as outlined in Kohler et al.3.

Proteins predicted to contain a signal-peptide, but not predicted to localize within the endoplasmic reticulum or to targeting a cytoplasmic domain were considered as secreted proteins. They were classified into four functional categories: CAZymes, lipases, proteases and small-secreted proteins (SSP; proteins less than 300 amino-acids long with no CAZyme, protease or lipase domains). The secreted proteases were identified in the genome using the MEROPS peptide database (http://merops.sanger.ac.uk) and the secreted lipases using the Lipase engineering database (www.lipase.org). Genes and gene clusters involved in secondary metabolism were predicted using a custom pipeline based on the SMURF method62. Conserved protein domains were predicted using Pfam version 27 (ref. 16). SMURF parameter d (maximum intergenic distance in base pairs) was set at 3,000 bp, and SMURF parameter y (the maximum number of non-secondary metabolism gene upstream or downstream of the backbone gene) was set at 6.

Gene models of major intrinsic proteins (aquaporins) were identified using the automatic annotation pipeline of JGI. The primary structure of the deduced proteins was manually inspected and a conserved signature for the endoplasmic reticulum or mitochondria, and displaying less than one
sequences were initially analyzed using ClustalW and then applied to a Bayesian approach on the basis of Markov chain Monte Carlo (MCMC) as implemented in the computer program MrBayes 3.2.2 (ref. 62) to estimate phylogenetic relationships. Rather than specifying an amino acid substitution model, we allowed the Markov processes to sample randomly from the substitution models implemented in MrBayes. Trees were sampled every 200 generations resulting in an overall sampling of 20,000 trees per run, from which the first 4,000 trees of each run were discarded (burn-in). The remaining 12,000 trees sampled in each run were pooled and used to compute a majority rule consensus tree to obtain estimates of the posterior probabilities. Branch lengths were averaged over the sampled trees.

To identify sex-related genes and genes involved in fruiting body formation, we used a whole-genome shotgun library of STP and TRLB with known names of ascomycete species for mating-type, pheromones and pheromone receptors as well as genes involved in fruiting body development within the predicted proteomes and genomes of C. geophilum and its closest relative G. stellatum.

Transcriptome analyses of Conococcum geophilum. Total RNA was extracted from free-living mycelium grown on agar Petri dishes containing Cg-Medium and covered by a cellophane for 2 months and from ECM root tips of C. geophilum/Pinus sylvestris (growth conditions see below) using the RNAeasy Plant Mini Kit of Qiagen with DNase step and addition of 20 mg/ml -polymethylene glycol to the RLC extraction buffer. RNA sequencing was performed using 2 to 100 pg Illumina HiSeq sequencing after mRNA library construction by IGA Technology services (Udine, Italy). Raw reads were trimmed and aligned to the C. geophilum genome (https://snp.sanger.ac.uk) using the clc-genomics Workbench 6 and 7 (http://www.clcsupport.com/clcgenomicsworkbench/900/index.php#manual=RNA_Seq_analysis.html) and as described before 1. To identify differentially regulated transcripts in mycorrhizal tissues compared with free-living mycelium, the test of Baggett et al. 2 was implemented in CLC Genomic Workbench and the Benjamini-Hochberg multiple testing algorithm was used. FDR was applied to the data. If not otherwise indicated, transcripts with ≥ 5 fold change and a FDR-corrected P value < 0.05 were kept for further analysis. To assess whether symbiosis-regulated transcripts were either conserved among fungal species or lineage-specific, their protein sequences were queried against the protein sequences of other fungi using BlastP (Supplementary Data 1) using BLASTP as indicated in Kohler et al. 1. The distributions of the symbiosis-regulated transcripts within each cluster were quantified according to their putative function as CAZymes, SPSs, with KOG classification or without KOG. The enrichment of secreted proteins and SPSs in symbiosis-regulated transcript clusters was assessed by comparing their percentages in symbiosis-regulated transcript clusters to their relative percentage in the total gene repertoire and applying a Fisher’s exact test.

Functional analyses of aquaporins in Conococcum geophilum. For heterologous expression of selected aquaporins in Xenopus laevis oocytes, total RNA was isolated from C. geophilum/P. sylvestris ECM root tips using the NucleoSpin RNA Plant kit (NAGEL GmbH, Düren, Germany) with a DNase step according to the manufacturer’s instructions. First-strand cDNA was synthesized in a total volume of 20 µl containing 30 pmol oligo-d(T)18-primer (Eurofins-MWG-Operon, Ebersberg, Germany) and 100 U Ribolock RNAase inhibitor (40 µl -1) using the First Strand cDNA Synthesis Kit (NEB) and 1 µl of the PCR products was used as a positive control. The quality of the grids was determined by counting ECM oocytes with the aid of the FLM. ECM plants were produced as follows: Mycelium was grown in LI-6400XT Portable Photosynthesis System (LI-COR) at five time points (8, 9 and 13 days of drought stress and 3 days after rewatering. For each time point, 16 pots were harvested. 4 of each combination (non-mycorrhizal/mycorrhizal, drought-stressed/control). Following plant parameters were assessed during the time course of the experiment: Net photosynthesis, stomatal conductance and transpiration using a LI-6400XT Portable Photosynthesis System (LI-COR) at five time points (8,9 and 13 days of drought, 1 and 2 days after rewatering). These non-destructive measurements were performed the day before the harvesting time points as well as at two additional days to follow the time course. Plant water use efficiency was calculated as net photosynthesis/transpiration. The pre-dawn water potential of shoots was measured using a Scholander Pressure Chamber (MWTEch support) at pre-dawn of the three harvesting days. Shoot and root dry weight was measured using a M 600 Scholander Pressure Chamber (MMMtech) or the Power Plant RNA isolation kit (MOBI) as indicated by the NC 2500 elemental analyzer of CE instruments. At harvesting days, each pot was checked for the mycorrhizal infection number by calculating as the colonization rate. shoots and mycorrhizal potential contaminants of other ECM fungi. No contaminating ECM fungi were present in NM and ECM pots. A subsample of roots was assessed for the degree of colonization by C. geophilum by counting ECM and NM root-tips. It ranged between 10 and 50% with a mean colonization of 23% in inoculated pots and 0% in NM-pots. ECM root-tips were harvested at 3 days after rewatering. These were cultured as described above for RNA-extraction using the NucleoSpin RNA Plant kit (MACHEREY-NAGEL) or the Power Plant RNA isolation kit (MOBI) as indicated by the NC 2500 elemental analyzer of CE instruments. At harvesting days, each pot was checked for the mycorrhizal infection number by calculating as the colonization rate. shoots and mycorrhizal potential contaminants of other ECM fungi.
chamber (25 °C) after which the cellophanes carrying the mycelia were transferred on new agar dishes. They were grown for another 10 days before starting the drought stress for half of the plates by opening the Petri dishes in a running sterile bench for 24 h, whereas control plates were kept closed and sealed with parafilm. Four plates of each treatment were then harvested by peeling the mycelia from the membranes and immediately shock freezing them in liquid N. The rest of the dishes that were opened for 24 h were closed again but without sealing with parafilm, kept for additional 3 days in the running sterile bench. After this time period, the agar was completely dried out (Supplementary Fig. 9) and four plates each of control and drought-stressed mycelia were harvested and kept at −80 °C before RNA extraction. RNA was extracted using the NucleoSpin RNA Plant Kit (MACHEREY-NAGEL) or the Power Plant RNA isolation kit (MOBIO) as indicated by the providers but with adding 20 mg of polyethylene glycol to the extraction buffer.

Gene expression analyses using qPCR. To quantify AQP expression, we applied the absolute quantification method with standard curves using an ABI 7500 Fast real-time cycler as indicated by the provider (Applied Biosystems). RT was performed with 20–250 ng of total RNA (mean of 140 ng and 30 ng for mycelia and mycorrhizas, respectively) using the QuantiTect Reverse Transcription Kit (Qiagen) under conditions recommended by the manufacturer. The primers were designed in the 3′ untranslated region if possible and span an exon–exon junction if possible (sequences of the primers and fragments in Supplementary Data 15). Specificity was tested by virtual PCR (FastPCR; primerdigital.com/fastpcr.html) against the genome and by sequencing the amplified fragment. PCR was performed with the SYBR Green PCR master mix (Applied Biosystems) in a volume of 15 μL containing 4–7 μg of cDNA. Cycling was carried out with the following parameters: 10 min activation of AmpliTaq Gold Polymerase at 95 °C, 40 cycles of 15 s denaturation at 95 °C, 60 s annealing and extension at 60 °C, followed by a dissociation step of 15 s at 95 °C, 60 sec at 60 °C, and 30 s at 95 °C to detect primer dimers and nonspecific amplification products. For each primer pair, we determined the PCR efficiency and the dynamic range of PCR by plotting the threshold cycle (Ct) values generated over a range of dilutions against the log input of cDNA amount from diverse samples. Primer efficiencies were also calculated in each qPCR reaction for standard curves. To obtain accurate results, only primer pairs yielding PCR efficiencies of 90–110% (slope regression between Ct and log (dilution) vs. −3.5) were considered. For several samples, reproducibility was checked by diluting cDNA and performing PCR in triplicate within the same plate and in different qPCR reactions and was confirmed to be high. To quantify the transcripts, cDNAs of at least three independent replicates (for example, mycorrhizas of plants grown in different pots, mycelia from different Petri dishes) were diluted and used in qPCR reactions. Copy numbers of the transcripts were calculated from standard curves that were obtained using synthesized single-stranded sense oligonucleotides (Sigma-Aldrich, Switzerland; Supplementary Data 15) specifying amplification of the selected genes and using the information from the manufacturer to calculate the copy number of the oligonucleotides present in respective vegetative tissues. Serial dilutions of the stocks were carried out in duplicate, and dilutions in the range from 10^1 to 10^8 copies, depending on the expression profiles of the AQPs, were used in duplicate PCR to generate standard curves. The standard curve was obtained by plotting the logs of the calculated copy number against Ct. The copy number of unknown samples were calculated from the regression line. The transcript copy numbers were normalized to 10^6 actin (IGI ID Cenge6: 653800) transcripts, which was used as internal control.

Statistics of aquaporin expression and plant parameters. To assess the effect of treatments on AQP expression, we applied linear ANOVA in SPSS statistics (IBM). If necessary, expression values were log10 transformed to meet the assumption of normality. The effect of the drought/rewatering treatment on AQP expression at single time points was tested by the Kruskal–Wallis H statistic in SPSS (IBM). We applied correlation analyses between AQP expression and diverse plant parameters such as chlorophyll content and needle nitrogen content as well as shoot and root dry weights in the drought/rewatering treatment on AQP expression, we applied linear ANOVA in SPSS statistics (IBM). We applied correlation analyses between AQP expression and diverse plant parameters such as chlorophyll content and needle nitrogen content as well as shoot and root dry weights in the drought/rewatering treatment. We applied one-way ANOVA to test the effect of mycorrhizal treatment on relative water content, needle nitrogen content and chlorophyll content and needle nitrogen content as well as shoot and root dry weights in the drought/rewatering treatment. We applied one-way ANOVA to test the effect of mycorrhizal treatment on relative water content, needle nitrogen content and chlorophyll content and needle nitrogen content as well as shoot and root dry weights in the drought/rewatering treatment.

Data availability. Genome assemblies and annotations for the organisms used in this study are available via the JGI fungal genome portal MycoCosm14 (http://jgi doe.gov/fungi). In addition, the newly sequenced genome assemblies and annotations have been deposited to GenBank under the following accession codes:BiosProjects: Cenococcum geophilum: LKKR00000000/PRJNA196023; Glomum stellatum: LKA00000000/PRJNA295956; Lepidoptera palustre: ECM fungi: N2016CA19401. Amino acid alignment and phylogenetic tree files have been deposited in the TreeBase repository under accession number S15956. The complete transcriptome data sets are available at the Gene Expression Omnibus at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/). The microarray derived data sets as series GSE80549, the RNA-Seq derived data as GSE83909. The authors declare that all other data supporting the findings of this study are available within the article and its Supplementary Information files or are available from the corresponding authors upon request.

References

Oligoarray construction and hybridization. A 12-plex oligoarray was manufactured by NimbleGen Systems using 43,926 454-sequencing-derived contigs and selected singletons of expressed sequence tags corresponding to ~7,000 and 5,500 unigenes from Pina sylvestris and Cenococcum geophilum. The expressed sequence tags used for microarray construction were produced using the SMART PCR cDNA Synthesis Kit (Clontech) according to the manufacturer’s instructions.

Mycorrhiza-induced plant genes. Double-mycorrhizal P. sylvestris seedlings were produced as described in Kipfer et al.15 using C. geophilum 1.58 and the abundant

pine symbiotic ECM species Rhizopogon roseolus (Corda) Th. Fr. or Suillus granulatus (L.) Roussel. Four months after inoculation, three and two replicate plants were harvested for each fungal combination and for non-mycorrhizal plants, respectively. Single fine root tips were collected for each ECM species per plant and for non-mycorrhizal plants in separate tubes and were shock-frozen. Total RNAs were extracted as described above and cDNA was amplified using the SMART PCR cDNA Synthesis Kit (Clontech) according to the manufacturer’s instructions.

Statistics of aquaporin expression and plant parameters. To assess the effect of treatments on AQP expression, we applied linear ANOVA in SPSS statistics (IBM). If necessary, expression values were log10 transformed to meet the assumption of normality. The effect of the drought/rewatering treatment on AQP expression at single time points was tested by the Kruskal–Wallis H statistic in SPSS (IBM). We applied correlation analyses between AQP expression and diverse plant parameters such as chlorophyll content and needle nitrogen content as well as shoot and root dry weights in the drought/rewatering treatment. We applied one-way ANOVA to test the effect of mycorrhizal treatment on relative water content, needle nitrogen content and chlorophyll content and needle nitrogen content as well as shoot and root dry weights in the drought/rewatering treatment.

Data availability. Genome assemblies and annotations for the organisms used in this study are available via the JGI fungal genome portal MycoCosm14 (http://jgi doe.gov/fungi). In addition, the newly sequenced genome assemblies and annotations have been deposited to GenBank under the following accession codes:BiosProjects: Cenococcum geophilum: LKKR00000000/PRJNA196023; Glomum stellatum: LKA00000000/PRJNA295956; Lepidoptera palustre: ECM fungi: N2016CA19401. Amino acid alignment and phylogenetic tree files have been deposited in the TreeBase repository under accession number S15956. The complete transcriptome data sets are available at the Gene Expression Omnibus at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/). The microarray derived data sets as series GSE80549, the RNA-Seq derived data as GSE83909. The authors declare that all other data supporting the findings of this study are available within the article and its Supplementary Information files or are available from the corresponding authors upon request.

References


5. LoBuglio, K. F. in Ectomycorrhizal Fungi. Key Genera in Profile


Acknowledgements
We thank Dr David Hibbett and our colleagues from the Mycorrhizal Genomics Initiative consortium for helpful discussions. We thank WSL-staff Raphael Appenzeller, Rosmarie Eppenberger and Sylvia Hutter for technical assistance and Terence Menard for microscopic cuttings. We are indebted to Dr Annette Peter and Annette Hintelmann (University of Bremen) for providing Xenopus oocytes and technical support, Nicolas Cichocki and Maìra de Freitas Pereira (INRA-Nancy) for support in RNA extractions and microscopic pictures. This material is based on work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, under contract no. DE-AC02-05CH11231. It was supported by the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01), the ARBRE-WSL joint project BLACK-SECRET, the Genomic Science Program (Plant-Microbe Interactions project) funded by the US Department of Energy, Office of Science, Biological and Environmental Research (contract DE-AC05-00OR22725), and the Lorraine Region Council (to F.M.M.), the US National Science Foundation (DEB-0732993 to J.W.S.) and the EC-supported Network of Excellence EevlTree (GOCE-016322 to M.P.).

Author contributions

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Peter, M. et al. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus Cenococcum geophilum. Nat. Commun. 7:12662 doi: 10.1038/ncomms12662 (2016).