The ISME Journal (2011) 5, 1871–1880

www.nature.com/ismej

ORIGINAL ARTICLE

A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes

Coralie Damon[1], Laurent Vallon[1], Sabine Zimmermann[2], Muhammad Z Haider[2], Virginie Galeote[3], Sylvie Dequin[3], Patricia Luis[1], Laurence Fraissinet-Tachet[1] and Roland Marmeisse[1]

[1]*Université de Lyon, Université Lyon 1, CNRS, UMR5557 d'Ecologie Microbienne, Villeurbanne, France;* [2]*CNRS, UMR5004 Biochimie et Physiologie Moléculaire des Plantes, Montpellier, France and* [3]*INRA, UMR1083 Sciences pour l'Oenologie, Montpellier, France*

Functional environmental genomics has the potential to identify novel biological functions that the systematic sequencing of microbial genomes or environmental DNA may fail to uncover. We targeted the functions expressed by soil eukaryotes using a metatranscriptomic approach based on the use of soil-extracted polyadenylated messenger RNA to construct environmental complementary DNA expression libraries. Functional complementation of a yeast mutant defective in di/tripeptide uptake identified a novel family of oligopeptide transporters expressed by fungi. This family has a patchy distribution in the Basidiomycota and Ascomycota and is present in the genome of a *Saccharomyces cerevisiae* wine strain. High throughput phenotyping of yeast mutants expressing two environmental transporters showed that they both displayed broad substrate specificity and could transport more than 60–80 dipeptides. When expressed in *Xenopus* oocytes one environmental transporter induced currents upon dipeptide addition, suggesting proton-coupled co-transport of dipeptides. This transporter was also able to transport specifically cysteine. Deletion of the two copies of the corresponding gene family members in the genome of the wine yeast strain severely reduced the number of dipeptides that it could assimilate. These results demonstrate that these genes are functional and can be used by fungi to efficiently scavenge the numerous, low concentration, oligopeptides continuously generated in soils by proteolysis.
The ISME Journal (2011) **5,** 1871–1880; doi:10.1038/ismej.2011.67; published online 9 June 2011
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: environmental genomics; metatranscriptomics; oligopeptide transport; *Saccharomyces cerevisiae*

Introduction

Microorganisms in their natural environments are confronted with a huge, uncountable number and variety of organic and inorganic molecules that need to cross the cell plasma membrane to be used as nutrient sources. Metabolic versatility of microorganisms, or their ability to use or not specific compounds as nutrient sources, is therefore partly determined by the existence of transporters at the plasma membrane, which allow assimilation of these molecules.

In eukaryotic microorganisms, transport of monomeric compounds (that is, monosaccharides, amino acids) is well-documented. For example, in the yeast *S. cerevisiae*, at least 16 different plasma membrane amino-acid transporters belonging to the yeast amino acid transporter family have been identified (Transporter Classification Database No. T.C. 2.A.3.10). They differ with respect to substrate specificity, kinetic properties and regulation patterns (De Hertogh *et al.*, 2002; Saier *et al.*, 2009). In complex environments, these monomeric molecules are often the end products of the degradation of large polymers (lignocellulosic compounds, proteins or nucleic acids) by extracellular hydrolytic enzymes. Decomposition of these polymers also produces numerous oligomeric molecules (for example, oligosaccharides, oligopeptides), which can directly enter the cells to be used as carbon and/or nitrogen (N) sources. Nevertheless, although the chemical diversity of these oligomers is necessarily higher than the diversity of the monomers, the number and diversity of the corresponding identified transporters is far lower. Indeed, concerning N sources assimilation, only two di/tripeptide transporter coding genes have been identified in the yeast

Correspondence: R Marmeisse, Université de Lyon, Université Lyon 1, CNRS, UMR5557 d'Ecologie Microbienne, Bâtiment Lwoff, 43 Boulevard du 11 Novembre, F-69622 Villeurbanne Cedex, France.
E-mail roland.marmeisse@univ-lyon1.fr
Received 23 November 2010; revised 26 April 2011; accepted 29 April 2011; published online 9 June 2011

S. cerevisiae, *PTR2* and *DAL5* (Homann *et al.*, 2005; Cai *et al.*, 2007), representing two of the three known families of eukaryotic dipeptide transporters (Frølund *et al.*, 2010). In *S. cerevisiae*, each of these two transporters display a broad substrate specificity. For example, when overexpressed in yeast cells, Ptr2p alone could transport more than 100 distinct dipeptides. However, translational and post-translational regulation mechanisms limit Ptr2p levels in wild-type yeasts and as a consequence also limit the actual diversity of dipeptides that these strains can ultimately use as N sources (Homann *et al.*, 2005).

Because of the chemical diversity of oligopeptides, it seems likely that additional dipeptide transport systems could exist in eukaryotic organisms, which are in strong competition for the use of N sources, often present in limiting amounts in natural ecosystems as exemplified by forest soils (Read and Perez Moreno, 2002). Therefore, in order to identify novel transporters, we implemented an environmental genomics approach that allows the simultaneous screening of genes present in the genomes of the numerous eukaryotic organisms cohabiting in an environmental sample. This approach, referred to as functional metatranscriptomics, is based on the extraction of polyadenylated messenger RNA from environmental samples (Grant *et al.*, 2006; Bailly *et al.*, 2007) and their conversion into complementary DNAs (cDNAs) that are cloned in a yeast expression vector (Bailly *et al.*, 2007; Kellner *et al.*, 2010). These environmental cDNA libraries are, therefore, representative of the diversity of the genes expressed by the different eukaryotic organisms present in the studied ecosystem. These libraries can be explored for novel genes by expression in a model microorganism. In our case, we performed the functional complementation of a *S. cerevisiae* mutant defective in dipeptide assimilation as a result of the deletion of the *PTR2* and *DAL5* genes (Homann *et al.*, 2005).

This strategy identified members of a novel family of fungal oligopeptide transporters (fungal oligopeptide transporter, FOT). To demonstrate, as hypothesized, that these genes contribute to the metabolic versatility of the microorganisms that carry them, we analyzed the di/tripeptide utilization profile of a mutant yeast strain deleted of the corresponding genes.

Materials and methods

Soil sampling, RNA extraction and cDNA library construction

A composite, organic matter-rich (13% total C, pH_{H_2O} 3.9) soil sample was collected in July 2007 in a 30-year-old spruce (*Picea abies*) forest stand located in central France (Breuil–Chenue forest, 47°18′10″N, 4°4′44″E, 638 m above sea level). For details about the forest, soil characteristics and sampling strategy, see Ranger *et al.* (2004) and Damon *et al.* (2010). After sieving (2 mm mesh size), the sample was frozen and stored at −70 °C.

Total, pro- and eukaryotic, RNA was extracted from ca 90 g of soil according to Bailly *et al.* (2007) and Damon *et al.* (2010). Polyadenylated messenger RNAs were isolated by affinity capture on poly-dT-coated paramagnetic beads (Dynal) and converted into double-stranded cDNAs (SMART cDNA library construction kit, Clontech, Saint-Germain-en-Laye, France). cDNAs were size-fractionated by 2-dimensional agarose gel electrophoresis as presented by Wellenreuther *et al.* (2004). The recovered 1–3 kb cDNAs were PCR-amplified (26 cycles, SMART cDNA library construction kit) and cloned downstream of the *S. cerevisiae PGK1* promoter in the pFL61 (Minet *et al.*, 1992) yeast–*E. coli* shuttle vector into which *SfiI*A&B cloning sites had been introduced.

Selection of dipeptide transporters encoding cDNAs

Environmental cDNAs encoding dipeptide transporter genes were selected by the functional complementation of the W303 (*MAT**a** ura3 can1–100 ho ptr2*Δ *dal5*Δ) *S. cerevisiae* mutant strain unable to use most dipeptides as N sources (Homann *et al.*, 2005). Yeast cells transformed by the cDNA library were plated on a yeast nitrogen base minimal medium lacking uracil and containing the three dipeptides Tyr–Ala, Ala–Leu and Ala–Tyr as sole N sources (0.25 mM each). Yeast manipulation, including transformation, total and plasmid DNA extraction, and plasmid rescue followed standard protocols (Rose *et al.*, 1990).

Phenotypic characterization of yeast strains

The di/tripeptide utilization profile of the different yeast strains and of their transformants was performed using the Biolog (Hayward, CA, USA) Phenotype MicroArrays (PM) system combined with the OmniLog reader (Biolog; see Supplementary Material). Cultures were incubated at 30 °C in duplicate over more than 48 h. We considered that a strain utilized a di/tripeptide when its growth value at 48 h on this N source represented 20% or more of its growth value measured on L-glutamine taken as a positive control for growth (= 100% growth). Below 20%, the growth signal could be confounded with spontaneous reduction of the tetrazolium dye or reduction by the resting yeast inoculum.

Functional expression in Xenopus *oocytes and electrophysiological analysis*

Procedures used to characterize by voltage-clamp recordings the EnvFOFp protein expressed in *Xenopus* oocytes are detailed in Supplementary Material.

Sequence analyses
Protein sequences homologous to the new oligopeptide (FOT) and Ptr2p transporters were identified by BLASTp and BLASTx searches against the GenBank EST and nr protein databases and genome databases at the Joint Genome Institute (http://genome.jgi-psf.org/), Broad Institute (http://www.broadinstitute.org/scientific-community/data?) and Genolevure (http://www.genolevures.org/). Prediction of transmembrane domains was performed using the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). Amino-acid sequences were aligned using MUSCLE (Edgar, 2004); phylogenetically informative positions were selected using Gblocks with relaxed parameters (Castresana, 2000) and phylogenetic reconstructions were performed with PhyML (Guindon and Gascuel, 2003) using the WAG substitution model and other default parameters. The whole process was performed using Seaview v. 4 (http://pbil.univ-lyon1.fr/software/seaview.html) (Gouy *et al.*, 2010) and Phylogeny.fr (http://www.phylogeny.fr/; Dereeper *et al.*, 2008).

Results

A novel family of fungal dipeptide transporters
Using polyadenylated messenger RNA directly extracted from a spruce (*P. abies*) forest soil, we constructed a sized cDNA expression library (1.75×10^6 clones), which was transferred into the W303 yeast mutant deleted of its two known dipeptide transporter genes, *PTR2* and *DAL5*. Transformants were plated onto a selective medium, containing three dipeptides as N sources, Tyr–Ala was used by Ptr2p but not by Dal5p, Ala–Leu was used by Dal5p but not by Ptr2p, and Ala–Tyr was neither used by Ptr2p nor by Dal5p in the W303 background (Homann *et al.*, 2005). Among 7.7×10^5 screened transformants, 25 grew on the selective medium. Following plasmid rescue and sequencing, we identified six different sequences each containing long open-reading frames beginning with ATG start codons and encoding polypeptides ranging in size between 502 and 607 amino acids (GenBank/EMBL/DDJB accession No. FR690804-808 and FR714879). All these six polypeptides were homologous among them, but were not homologous to either Ptr2p or Dal5p. In line with this result, we did not identify any DAL5 (T.C. 2.A.1.14) homologs among ca. 9000 randomly sequenced *P. abies* soil cDNAs (accession Nos. FR697056—FR706058), whereas we could identify two members of the PTR gene family (T.C 2.A.17) and two of the new dipeptide transporters in this sequence dataset.

BLAST searches against the GenBank/EMBL/DDJB database and specialized genome databases identified homologous sequences in Ascomycota and Basidiomycota fungal species (25–49% identities) and also those in a limited number of non-fungal species (21–27% identities) belonging to the Plantae (for example, *Sorghum bicolor*, *Vicia faba*), Metazoas (for example, *Trichoplax adhaerens, Caenorhabditis elegans*), Choanoflagelates (*Monosiga brevicollis*) and Alveolatas (*Perkinsus marinus*). A phylogenetic analysis, using 73 homologous proteins identified in the genome sequences of 24 fungal strains representing 23 different species (two strains of *S. cerevisiae*, the haploid laboratory S288c and the diploid EC1118 one were inspected, Supplementary Table S1), showed that the 6 environmental sequences grouped with 15 others to form a distinct and statistically well-supported, fungal-specific, clade (clade A, Figure 1). None of these 15 fungal proteins had been studied experimentally but several of them had been automatically annotated as 'putative amino-acid transporters'. Indeed, these proteins are related (ca. 23–24% identities) to fungal members of the Amino Acid/Auxin Permease family (TC 2.A.18.4.), represented by two proteins from *Neurospora crassa* and *Penicillium chrysogenum* that have been shown experimentally to transport aromatic and/or neutral amino acids (Koo and Stuart, 1991; Trip *et al.*, 2004). These two latter proteins are located in clade C of the phylogenetic tree (Figure 1). This observation suggested that the environmental transporters could also potentially transport amino acids.

If we hypothesize that all the members of clade A in the gene tree code for oligopeptide transporters, referred to as the FOT family, members of this novel class of peptide transporters were present in only 10 out of the 24 inspected fungal genomes and absent from the others (Supplementary Table S1). Distribution of the genes did not show a clear phylogenetic pattern. Species with and without these genes were found at different taxonomic levels, suggesting frequent and recurrent loss, and possibly also acquisition during fungal evolution. For example, species without FOT coding genes have been identified both within the Basidiomycota (for example, *Laccaria bicolor* or *Coprinopsis cinerea*) and the Ascomycota (for example, *N. crassa*, *Tuber melanosporum* or *Magnaporthe grisea*; Supplementary Table S1). In the same genus, the gene can also be present (*Aspergillus fumigatus*) or absent (*Asp. nidulans*). Finally, even at the species level, this gene can show an uneven distribution as illustrated by *S. cerevisiae* where the gene is absent from the S288c laboratory strain and also from the W303 one, but present in several other strains used in wine making as exemplified by the EC1118 wine strain that possesses two closely linked copies on its chromosome XV (Novo *et al.*, 2009).

A specific phylogenetic analysis of the FOT clade using 33 reference protein sequences from Ascomycota and Basidiomycota taxa showed that the EnvFOT-A sequence was nested within a statistically well-supported basidiomycete sequence clade, while EnvFOT-B to -F originated from potentially phylogenetically distinct pezizomycotina species, although a single-species origin could not be ruled

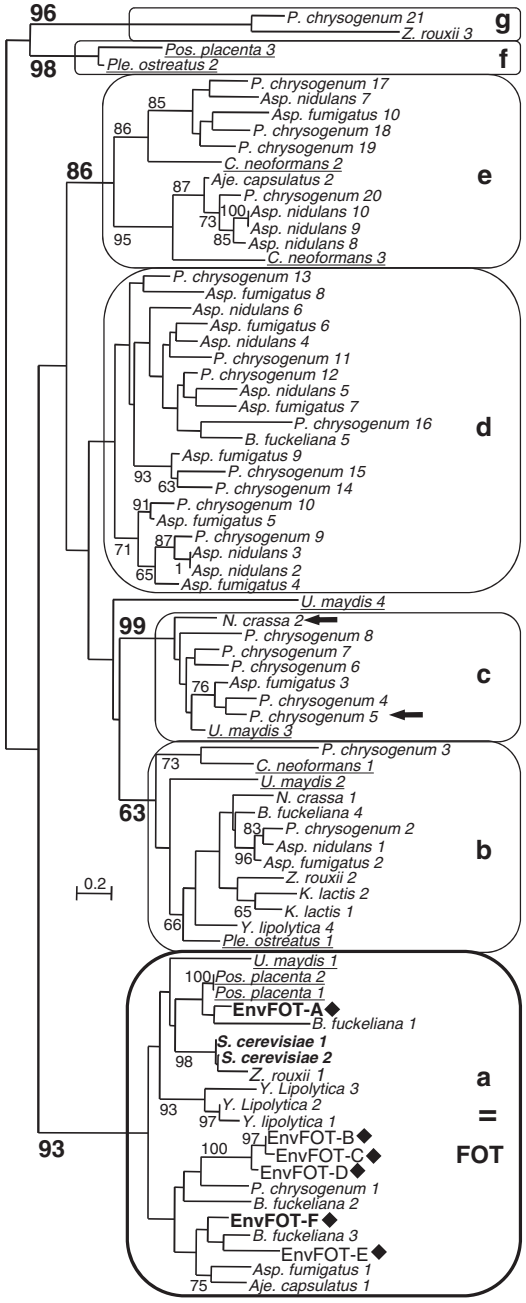

Figure 1 Phylogenetic relationships (PhyML) between 6 environmental oligopeptide transporter protein sequences (black diamonds) and the 73 protein sequences belonging to the Amino Acid/Auxin Permease family identified in 14 fungal genomes. Underlined names are from Basidiomycota, the others from Ascomycota. All six environmental transporter sequences as well as the two homologous sequences from the *S. cerevisiae* EC1118 wine strain belong to clade A, which defines the new FOT family. The two experimentally characterized amino-acid transporters from *N. crassa* and *P. chrysogenum* cluster together in clade C (arrows). Correspondence between protein names and accession number. is given in Supplementary Table S2. Bootstrap supports (200 replicates) above 60 are indicated. For this analysis, 85 phylogenetic informative positions were considered.

out as illustrated by the three highly divergent FOT sequences of *Botryotinia fuckeliana* (Supplementary Figure S1).

Patchy distribution of the FOT family and also of the entire Amino Acid/Auxin Permease gene family among the Dikarya (Basidiomycota + Ascomycota) contrasts with the systematic presence of 1–7 members of the PTR di/tripeptide transporter family in each of the 24 inspected genomes (Supplementary Table S1). However, analysis of the two different *PTR2* alleles present in the diploid EC1118 *S. cerevisiae* strain showed that they were each characterized by a specific nonsense mutation that resulted in premature stop codons likely giving non-functional Ptr2p proteins (Supplementary Figure S2). Each of these two mutations were confirmed by PCR amplification and resequencing of the corresponding regions.

FOT predicted protein structure

Number and position of transmembrane helices were predicted for all 21 protein sequences in clade A of the FOT family (Figure 1) and 12 protein sequences representative of clades B to G (Supplementary Table S3). Eleven putative transmembrane domains characterized all these proteins (Figure 2a) and almost all FOT members shared some specific features that separated them from members of clades B to G. With the exception of EnvFOT-Ap, FOT proteins were on average 116 amino acids longer than their counterparts in clades B to G (Supplementary Table S3). This resulted from a significantly longer N-terminal cytosolic tail (on average of 135 amino acids in FOT proteins *versus* 70 for clades B to G) and longer spacers between several predicted transmembrane helices (illustrated for spacer domain 10–11 in Supplementary Table S3). EnvFOT-Ap lacked a long N-terminal tail as a result of cloning of a truncated cDNA. Indeed, the most likely ATG start codon of *EnvFOT-A* is an artifactual in-frame ATG triplet found in the plasmid *Sfi*IA cloning site.

The FOTs show a broad substrate specificity

The diversity of di/tripeptides transported by the FOT proteins that were overexpressed in the W303 ptr2Δ dal5Δ mutant was evaluated using the Biolog Phenotype MicroArrays (PM) system. This system allowed evaluation of 190 dipeptides and 9 tripeptides as potential substrates for the FOT proteins (see Supplementary Material).

Substrates transported by EnvFOT-Ap and EnvFOT-Fp were determined, as these two transporters were the only two out of the six tested to allow the W303 *ptr2*Δ *dal5*Δ mutant to grow to a high cell density in the PM system. These proteins could efficiently transport 77 and 59 dipeptides, respectively, as well as 6 tripeptides (Figures 2b and c, Supplementary Table S4). Furthermore, these two

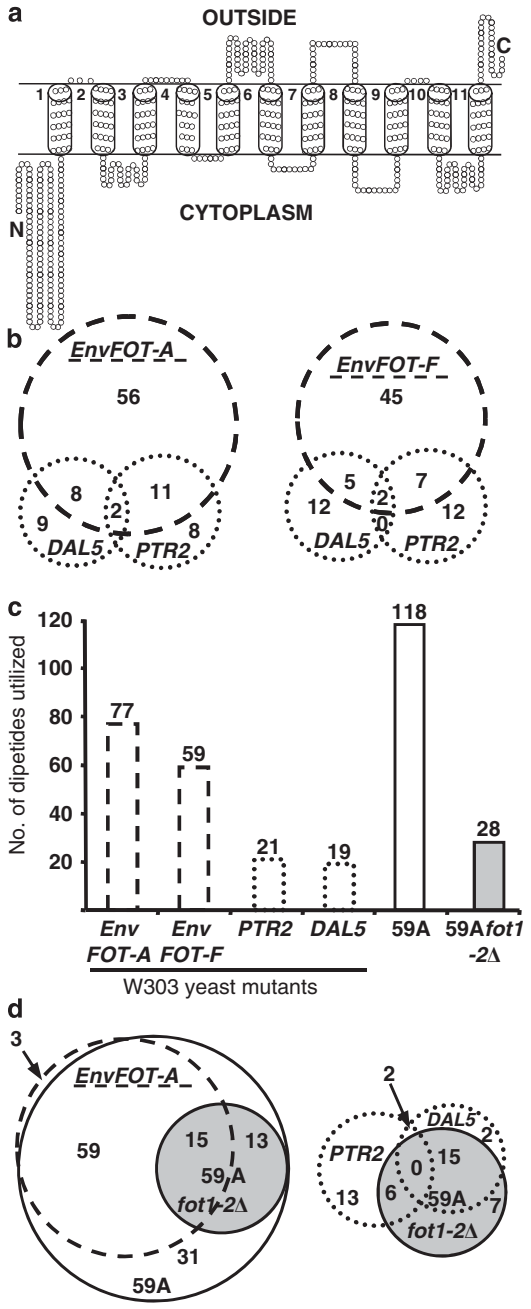

Figure 2 Predicted structure of environmental oligopeptide transporter EnvFOT-Fp (**a**) and spectra of dipeptides transported and used as N sources by yeast strains expressing different dipeptide transporters (**b–d**). (**b**, **d**), Venn diagrams illustrating the fractions of dipeptides used as substrates by either one, two or three yeast strains. (**c**) Total numbers of dipeptides transported by the different strains out of the 190 tested. Strains compared in (**b**) are the W303 *dal5*Δ or *ptr2*Δ mutants expressing either their endogenous *PTR2* or *DAL5* genes and the W303 *ptr2*Δ *dal5*Δ mutant expressing the cloned *EnvFOT-A* or *EnvFOT-F* genes. Strains compared in (**c**) are the wt 59A and corresponding mutant 59A *fot1-2*Δ wine yeast strains, the W303 *dal5*Δ or *ptr2*Δ mutants expressing either their endogenous *PTR2* or *DAL5* genes, respectively, and the W303 *ptr2*Δ *dal5*Δ mutant expressing the cloned *EnvFOT-A*. Venn diagrams were drawn to scale using the BioVenn software (http://www.cmbi.ru.nl/cdd/biovenn/) (Hulsen *et al*., 2008).

proteins had very similar substrate spectra; only 2 of the 59 dipeptides transported by EnvFOT-Fp did not support growth of the W303 mutant expressing EnvFOT-Ap and conversely 20 of the 77 dipeptides transported by EnvFOT-Ap were not transported by EnvFOT-Fp (Supplementary Table S4). A majority of the dipeptides specific to EnvFOT-Ap were characterized by an isoleucine or serine residue in N-terminal position or an isoleucine or leucine in the C-terminal position (Supplementary Table S4).

Transported dipeptides were chemically diverse. After exclusion of the cysteine, histidine or lysine-containing peptides (see Supplementary Material), 15 of the remaining 17 protein-forming L-amino acids were present in at least one of the dipeptides transported by either EnvFOT-Ap or EnvFOT-Fp (Table 1). The only two acidic amino acids, glutamate and aspartate, which can otherwise be used as N sources by *S. cerevisiae*, were not represented although they were present in 43 of the dipeptides analyzed. Dipeptides containing either a glutamate or aspartate were also not efficiently transported by either Ptr2p or Dal5p (Table 1).

Parallel study of di/tripeptide assimilation by the two single mutants, W303 *ptr2*Δ and W303 *dal5*Δ, transformed with an empty plasmid vector showed that more than half of the dipeptides taken up by either Ptr2p or Dal5p were also taken up by Env FOT-Ap and EnvFOT-Fp (Figures 2b and c, Supplementary Table S4).

EnvFOT-Fp, member of the FOT family, functions as proton-coupled dipeptide and amino-acid transporter in Xenopus *oocytes*

EnvFOT-F was chosen for the expression in *Xenopus* oocytes because of its known substrate specificity in yeast (Figures 2b and c) and because it encodes, as opposed to *EnvFOT-A*, a protein with a full-length N-terminal cytosolic domain (Figure 2a). Oocytes expressing *EnvFOT-F* showed a clear current induction at hyperpolarized potentials upon addition of the dipeptide Tyr–Ala (500 μM, $n=39$ oocytes) in the bath solution. This current induction was reversible ($n=35$), perfusion without the dipeptide added to the bath solution decreased the current. The amplitude of the observed currents was dependent on the membrane potential and was typically in the order of around 100 nA at −140 mV with current maxima of around −200 nA (Figures 3a and d). This current induction was observed with a standard physiological bath solution at pH 5.5 but not at pH 7.5 ($n=7$ oocytes; Figure 3a), suggesting a proton-coupled co-transport. Water-injected control oocytes tested with 500 μM ($n=7$) or 1 mM ($n=7$) Tyr–Ala did not show any current induction ($n=7$). Testing different concentrations of Tyr–Ala revealed a concentration-dependent effect of the current induction saturating at concentrations between 500–1000 μM with a mean K_m of 23 μM ($n=10$,

Table 1 Influence of amino-acid at position 1 (N-terminal) or 2 (C-terminal) within the dipeptides on the capacity of the latter to be used as N sources by the W303 yeast strain expressing either the *EnvFOT-A, EnvFOT-F, PTR2* or *DAL5* genes

	Total no. in PM plates		*Transported by EnvFOT-Ap/EnvFOT-Fp*		*Transported by Ptr2p*		*Transported by Dal5p*	
	Pos. 1	*Pos.2*	*Pos.1*	*Pos.2*	*Pos.1*	*Pos.2*	*Pos.1*	*Pos.2*
Ala	15	14	7/4	6/6	1	3	6	3
Arg	13	10	1/1	9/9	2	4	0	0
Asn	2	1	1/0	1/1	0	0	0	1
Asp	9	13	0/0	0/0	0	0	0	0
Glu	8	15	0/0	0/0	1	1	0	0
Gln	3	11	1/1	6/6	0	0	0	3
Gly	14	15	1/1	0/1	0	0	5	0
Ile	14	9	13/8	7/3	3	2	1	1
Leu	14	13	10/10	8/2	1	0	1	3
Met	14	9	8/7	7/5	1	0	2	3
Phe	12	14	7/7	9/8	1	3	0	0
Pro	14	10	0/0	1/1	1	1	0	1
Ser	12	12	4/1	4/5	0	1	2	2
Thr	11	3	7/7	0/0	1	0	0	2
Trp	11	12	1/1	1/0	4	2	0	0
Tyr	10	13	6/4	7/5	3	3	0	0
Val	15	15	10/7	7/3	2	1	0	0

Abbreviations: N, nitrogen; Pos., position; PM, phenotypic microarray.
The first two columns give the total no. of dipeptides in the PM microtiterplates containing a particular amino acid in position 1 or 2. Dipeptides with either a Cys, His or Lys were excluded from the analysis.

Figures 3b and c). Phe–Ala, another dipeptide that supported growth of yeast cells expressing *EnvFOT-F*, induced similar currents at 500 μM ($n=4$), than Tyr–Ala (Figure 3d). In contrast, the dipeptide Ala–Leu, which failed to support the growth of yeast cells (Supplementary Table S4), did not induce any current in *Xenopus* oocytes (500 μM, $n=8$; Figure 3d).

The multiplicity and overlapping specificities of amino-acid transporters in *S. cerevisiae* make it difficult to test the capacity of a membrane transporter to mediate permeability for a specific amino acid in this yeast. We therefore screened all 20 protein-forming L-amino acids (1 mM) for a possible current induction within *Xenopus* oocytes (see Supplementary Material). Only cysteine ($n=14$ oocytes) could activate reversible currents (Figure 3d). This current activation was also pH-dependent, at pH 7.5 no current induction was observed ($n=3$). The current amplitudes increased when testing different concentrations ($n=10$), but we could not observe saturation up to 1 mM, and higher concentrations seemed to damage the oocytes as tested with water-injected control oocytes at 5 mM cysteine compared with 1 mM ($n=5$).

FOT transporters are important for peptide uptake by a yeast strain

The completely sequenced *S. cerevisiae* EC1118 wine strain differs from the laboratory S288c strain by the presence of long blocks of genes in its genome that could have been acquired by lateral transfer from different yeast species (Novo *et al.*, 2009). In block C on chromosome XV two linked genes are present, separated by 2819 bp, and code for FOT transporters (*S. cerevisiae FOT1* and *FOT2* in Figure 1). We used this specific yeast strain as a model fungus to verify that FOT transporters do participate in the use of peptides as N sources *in vivo*. These two genes were simultaneously deleted and replaced by a kanamycin resistance cassette (KanMX) in the 59A haploid progeny of the diploid EC1118 strain, thus, giving a mutant strain called 59A *fot1-2*Δ. This mutant yeast strain has also a functional *DAL5* and non-functional *PTR2* genes in its genome (see above and Supplementary Figure S2).

Comparison of peptide uptake between the wt 59A strain and the 59A *fot1-2*Δ mutant strain was performed using the PM system. While the 59A strain efficiently utilized 118 dipeptides and 8 tripeptides, the mutant one only used 28 dipeptides (that is, 24%; Figure 2c) and 4 tripeptides, thus, demonstrating that proteins encoded by members of the new FOT family do participate in the use of di/tripeptides as N sources in this yeast. Two thirds (66%) of the 90 dipeptides that were specifically used by the 59A strain, and not by the 59A *fot1-2*Δ mutant one, were also transported by the EnvFOT-Ap protein (Figure 2d). Among the remaining 31 dipeptides specifically used by the 59A strain and not by the W303 strain expressing *EnvFOT-A*, 5 contained a glutamate and 2 an aspartate, which are the two amino acids absent from the dipeptides transported by the W303 strain expressing either EnvFOT-Ap, EnvFOT-Fp, Ptr2p or Dal5p (Table 1). Of the 28 dipeptides utilized by the 59A *fot1-2*Δ mutant strain, 15 were among the 19 dipeptides transported by Dal5p and only 6 were among the 21 dipeptides transported by Ptr2p when expressed in

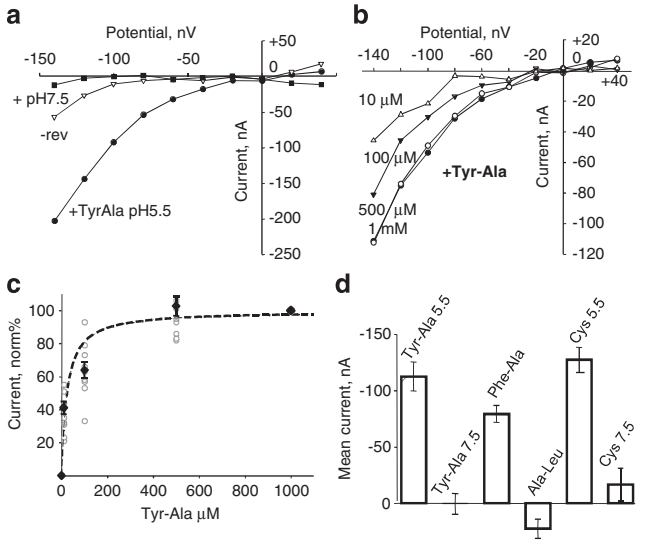

Figure 3 Functional expression and characterization of EnvFOT-Fp in *Xenopus* oocytes as a transporter of dipeptides and amino acids. (**a**), Current-voltage relation of EnvFOT-Fp mediated currents in presence of 500 μM Tyr-Ala at pH 5.5 (filled circles). The current induction is reversible (external solution without dipeptide, empty triangles) and dependent on the external pH, as no current is induced by the dipeptide at pH 7.5 (filled squares). The example of Tyr–Ala induced current is representative for $n=39$ oocytes at pH 5.5 and $n=7$ oocytes at pH 7.5. (**b**), Concentration-dependence of dipeptide-induced currents in EnvFOT-Fp expressing oocytes. The induced current increased with increasing dipeptide concentrations until saturation. Representative example for $n=10$ oocytes. (**c**) Dose-response curve for the dipeptide Tyr–Ala for EnvFOT-Fp expressing oocytes ($n=10$, $N=2$). The current values, presented as empty circles, are normalized (% of the current value with 1 mM Tyr–Ala for each oocyte). Mean values are presented as filled diamonds and fitted by a Michaelis–Menten equation (SigmaPlot), represented by the dotted line. (**d**) Mean current amplitudes in EnvFOT-Fp expressing oocytes induced by dipeptides Tyr–Ala at pH 5.5 ($n=39$, $N=6$) and pH 7.5 ($n=7$, $N=3$), Phe–Ala ($n=4$, $N=1$) and Ala–Leu ($n=8$, $N=2$) both at pH 5.5, as well as by the amino acid cysteine at pH 5.5 ($n=14$, $N=2$) and 7.5 ($n=3$, $N=1$).

the W303 strain (Figure 2d). This observation is consistent with the absence of a functional *PTR2* gene in the 59A wine yeast strain.

Discussion

In this paper, we demonstrate for the first time that functional environmental genomics, in combination with high throughput phenotyping (PM system) and cellular biology approaches (electrophysiology on *Xenopus* oocytes) can be used to discover and describe novel biological functions in the eukaryotes. Functional environmental genomics, in the form of metagenomics, had thus far only led to the characterization of novel bacterial/archaeal genes as exemplified by the proteorhodopsins (Béjà *et al.*, 2000; O'Malley, 2007; Fuhrman *et al.* 2008) and several genes coding for enzymes of biotechnological interest (recently reviewed by Ferrer *et al.*, 2009; Uchiyama and Miyazaki, 2009; Chistoserdova, 2010). As demonstrated here by the discovery of the FOT di/tripeptide transporters, functional metatranscriptomics based on the construction of expression cDNA libraries using polyadenylated environmental RNA bears similar promises. The two initial uses of this metatranscriptomic approach had thus far led to the selection of genes belonging to known families of enzymes, including imidazoleglycerol-phosphate dehydratases and an excreted phosphatase (Bailly *et al.*, 2007; Kellner *et al.*, 2010).

As opposed to the systematic sequencing of environmental nucleic acids, the functional screening of environmental libraries represents a novel way to ascribe functions to the numerous genes that have either no functionally characterized homologs or that are distantly related to proteins of known functions. In this context membrane transporters represent a special case study, as it is notoriously difficult to predict on the basis of their sole amino-acid sequences the nature of substrates transported. This makes homology-based functional annotation of these proteins hazardous as illustrated in this study where the most similar proteins to the new FOT transporters were all annotated as amino-acid transporters.

The study we initiated for di/tripeptide utilization could also be extended to other small molecules present in the natural ecosystems, including monomeric compounds such as diverse monosaccharides or amino acids. Indeed, many phylogenetic analyses of specific eukaryotic transporter families (this study, Figure 1 and also De Hertogh *et al.*, 2002, 2006; Hellborg *et al.*, 2008; López *et al.*, 2008; Lucic *et al.*, 2008), show that numerous transporter families and sub-families (as exemplified by clades B and clades D to G in Figure 1) have not been studied experimentally with respect to the substrates they transport. We therefore largely ignore how many nutrients present in the natural environment are absorbed by most microorganisms. Knowledge of the substrates transported by the different transporters is yet of prime importance in different fields of biology, including environmental sciences. In this latter area, Poretsky *et al.* (2010) suggested that the abundance of specific transporter gene families within metatranscriptomes could be interpreted in term of availability of the corresponding substrates in the environment and that the taxonomic origin of these transporters could reflect the taxonomic specialization within the microbial community for the assimilation of specific classes of molecules.

High throughput phenotyping of yeast cells expressing *FOT* genes showed that the corresponding transporters are rather unspecific and the wide spectrum of di/tripeptide they transport partially overlaps with the spectra of other di/tripeptide transporters (*PTR2*, *DAL5*). Furthermore, heterologous expression in *Xenopus* oocytes allowed us to demonstrate that at least *EnvFOT-F* could transport L-cysteine. The transport of additional nitrogenous

compounds had already been documented for other eukaryotic di/tripeptide transporters such as the yeast dal5p known to transport allantoate and ureidosuccinate (Cai *et al*., 2007), the *Arabidopsis PTR2-B* and *PTR1* shown to transport histidine (Frommer *et al*., 1994; Komarova *et al*., 2008) and the human hPAT1 that transports both amino acids and a few dipeptides (Frølund *et al*., 2010). It is therefore possible that the true range of N compounds transported by FOT transporters extends beyond di/tripeptides and cysteine.

Altogether, these results suggest that *FOT* genes, when present in a fungal genome, should increase the metabolic versatility of the corresponding fungal strain. It is however intriguing that this gene family, as several other transporter families (De Hertogh *et al*., 2006), has a patchy distribution in the fungal kingdom although proteins represent a widespread source of N in the environment and specifically in soils (Jones *et al*., 2009; Näsholm *et al*., 2009). Furthermore, presence of the FOT family does not seem to represent a marker of a specific lifestyle in the fungi as the species that harbor it are either saprotrophic (*Yarrowia lipolytica*, *Zygosaccharomyces rouxii*, *Postia placenta* or *S. lacrymans*), facultative animal pathogens (*Ajellomyces capsulatus*, *Asp. fumigatus* or *C. posadasii*), or plant pathogens (*B. fuckeliana, Ustilago maydis* or *C. parasitica*). According to the phylogenetic analysis of the cloned environmental sequences (Figure 1 and Supplementary Figure S1), they most likely come from both Basidiomycota and Pezizomycotina known to be diverse in the studied spruce forest stand (Buée *et al*., 2009; Damon *et al*., 2010). However, the actual number of reference FOT sequences in databases is too limited to identify, not only the corresponding species, but also the genera or families to which they belong. Although the two fully sequenced symbiotic ectomycorrhizal species, *L. bicolor* and *T. melanosporum,* do not have any *FOT* genes in their genomes, because of the patchy distribution of this gene family, we cannot exclude that this gene family is present in the genomes of other mycorrhizal species. In this case, the *FOT* genes could represent one of the component participating in the functioning of the ectomycorrhizal symbiosis known to allow forest trees to indirectly use soil organic N sources (Read and Perez Moreno, 2002; Näsholm *et al*. 2009).

Disruption of the cognate genes in the *S. cerevisiae* EC1118 wine strain established experimentally that *FOT* genes can make a significant contribution to oligopeptide assimilation in a fungus. Furthermore, the calculated K_m of EnvFOT-Fp for Tyr–Ala (23 μM) is lower than affinity constants reported for plant peptide transporters of the peptide transport family as estimated using *Xenopus* oocytes (for example, 130–160 μM for the root-expressed AtPTR5; Komarova *et al*., 2008, and references therein). As for the concentration of di/tripeptides in the soil solution, no quantitative data are available to our knowledge. We can however hypothesize that it is of a similar order of magnitude as the soil amino-acid concentration, estimated to be on average of 23 μM on a global scale (Jones *et al*., 2009). Similarly, probably because of their large number, no report exists on the chemical nature and abundance of individual di/tripeptides in soil. As they originate from proteolysis, the occurrence of amino acids in soil di/tripeptides could however be estimated from the average amino-acid composition of soil-extracted peptides/proteins (Isnor and Warman 1990; Yu *et al*., 2002) and also from the one of proteins from major forest organisms such as in our case the tree *P. abies* or the symbiotic basidiomycete *L. bicolor* (Supplementary Table S5). In both cases, we observe a greater abundance of the nonpolar Ala, Gly, Leu and Val and of the polar Ser, which are not necessarily the most prominent amino acids among the free soil amino acids (Yu *et al*., 2002; Werdin-Pfisterer *et al*., 2009; Supplementary Table S5). With the notable exception of Gly, these different amino acids are well-represented among the dipeptides that can be transported by EnvFOT-Fp and EnvFOT-Ap (Table 1) and by the yeast Fot1-2p (Supplementary Table S4). Altogether, these different data and observations strongly suggest that FOT transporters function as high affinity transporters, which should efficiently participate in di/tripeptide uptake in acidic soils, as it is the case for the studied site (pH 3.9).

Acknowledgements

We would like to thank the Lindquist's laboratory for the gift of the yeast strains deleted of either *PTR2* or *DAL5*; Laurence Loiseau for the use of the Omnilog reader on the PARMIC platform; Jacques Ranger for access to the Breuil forest, and Daniel Wipf for helpful discussions on oligopeptide transporters. This work was funded by the ECCO programme AO 2005 (project Microger) and the Agence Nationale de la Recherche Biodiversity program (project ANR-06-443 BDIV-006 Fundiv). MZ Haider was financially supported by a scholarship from the Higher Education Commission of Pakistan.

References

Bailly J, Fraissinet-Tachet L, Verner MC, Debaud JC, Lemaire M, Wésolowski-Louvel M *et al.* (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISMEJ* **1**: 632–642.

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al.* (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S *et al.* (2009). 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytol* **184**: 449–456.

Cai H, Hauser M, Naider F, Becker JM. (2007). Differential regulation and substrate preferences in two peptide transporters of *Saccharomyces cerevisiae*. *Euk Cell* **6**: 1805–1813.

Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.

Chistoserdova L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol Lett* **32**: 1351–1359.

Damon C, Barroso G, Férandon C, Ranger J, Fraissinet-Tachet L, Marmeisse R. (2010). Performance of COX1 gene as a marker for the study of metabolically active Pezizomycotina and Agaricomycetes fungal communities from the analysis of soil RNA. *FEMS Microbiol Ecol* **74**: 693–705.

De Hertogh B, Carvajal E, Talla E, Dujon B, Baret P, Goffeau A. (2002). Phylogenetic classification of transporters and other membrane proteins from *Saccharomyces cerevisiae*. *Funct Integr Genomics* **2**: 154–170.

De Hertogh B, Hancy F, Goffeau A, Baret PV. (2006). Emergence of species-specific transporters during evolution of the hemiascomycete phylum. *Genetics* **172**: 771–781.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F *et al.* (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465–W469.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Ferrer M, Beloqui A, Timmis KN, Golyshin PN. (2009). Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol* **16**: 109–123.

Frølund S, Holm R, Brodin B, Nielsen CU. (2010). The proton-coupled amino acid transporter, SLC36A1 (hPAT1), transport Gly-Gly, Gly-Sar and other Gly-Gly mimetics. *Br J Pharmacol* **161**: 589–600.

Frommer WB, Hummel S, Rentsch D. (1994). Cloning of an *Arabidopsis* histidine transporting protein related to nitrate and peptide transporters. *FEBS Lett* **347**: 185–189.

Fuhrman JA, Schwalbach MS, Stingl U. (2008). Proteorhodopsins: an array of physiological roles? *Nature Rev Microbiol* **6**: 488–494.

Gouy M, Guindon S, Gascuel O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221–224.

Grant S, Grant WD, Cowan DA, Jones BE, Ma Y, Ventosa A *et al.* (2006). Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. *Appl Env Microbiol* **72**: 135–143.

Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.

Hellborg L, Woolfit M, Arthursson-Hellborg M, Piskur J. (2008). Complex evolution of the *DAL5* transporter family. *BMC Genomics* **9**: 164.

Homann OR, Cai H, Becker JM, Lindquist SL. (2005). Harnessing natural diversity to probe metabolic pathways. *PLoS Genetics* **6**: e80.

Hulsen T, de Vlieg J, Alkema W. (2008). BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**: 488.

Isnor RA, Warman PR. (1990). Amino acid composition of soil peptides chromatographed by high performance liquid chromatography on C18 and C8 columns. *Biol Fertil Soils* **10**: 213–217.

Jones DL, Kielland K, Sinclair FL, Dahlgren A, Newsham KK, Farrar JF *et al.* (2009). Soil organic nitrogen mineralization across a global latitudinal gradient. *Global Biogeochem Cycles* **23**: GB1016.

Kellner H, Luis P, Portetelle D, Vandenbol M. (2010). Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. *Microbiol Res*. doi:10.1016/j.micres.2010.07.006.

Komarova NY, Thor K, Gubler A, Meier S, Dietrich D, Weichert A *et al.* (2008). AtPTR1 and AtPTR5 transport dipeptides in planta. *Plant Physiol* **148**: 856–869.

Koo K, Stuart WD. (1991). Sequence and structure of mtr, an amino acid transport gene of *Neurospora crassa*. *Genome* **34**: 644–651.

López MF, Dietz S, Grunze N, Bloschies J, Weiss M, Nehls U. (2008). The sugar porter family of *Laccaria bicolor*: function in ectomycorrhizal symbiosis and soil-growing hyphae. *New Phytol* **180**: 365–378.

Lucic E, Fourrey C, Kohler A, Martin F, Chalot M, Brun-Jacob A. (2008). A gene repertoire for nitrogen transporters in *Laccaria bicolor*. *New Phytol* **180**: 343–364.

Minet M, Dufour ME, Lacroute F. (1992). Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. *Curr Genet* **2**: 417–422.

Näsholm T, Kielland K, Ganeteg U. (2009). Uptake of organic nitrogen by plants. *New Phytol* **182**: 31–48.

Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S *et al.* (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc Natl Acad Sci USA* **106**: 16333–16338.

O'Malley M. (2007). Exploratory experimentation and scientific practice: metagenomics and the proteorhodopsin case. *Hist Philos Life Sci* **29**: 337–360.

Poretsky RS, Sun S, Mou X, Moran MA. (2010). Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ Microbiol* **12**: 616–627.

Ranger J, Andreux F, Bienaimé S, Berthelin F, Bonnaud P. (2004). Effet des substitutions d'essences sur le fonctionnement organo-minéral de l'écosystème forestier, sur les communautés microbiennes et sur la diversité des communautés fongiques mycorhiziennes et saprophytes (cas du dispositif expérimental de Breuil—Morvan). Rapport final du contrat INRA-GIP Ecofor2001-24, N° INRA 1502A. Document à distribution limitée, INRA BEF Nancy-54280 Champenoux.

Read DJ, Perez Moreno J. (2002). Mycorrhizae and nutrient cycling in ecosystems—a journey towards relevance? *New Phytol* **157**: 475–492.

Rose MD, Winston F, Hieter P. (1990). *Methods in Yeast Genetics: a Laboratory Course Manual*. Cold Spring Harbor Press: Cold Spring Harbor, NY.

Saier Jr MH, Yen MR, Noto K, Tamang DG, Elkan C. (2009). The transporter classification database: recent advances. *Nucleic Acids Res* **37**: D274–D278.

Trip H, Evers ME, Driessen AJ. (2004). PcMtr, an aromatic and neutral aliphatic amino acid permease of *Penicillium chrysogenum*. *Biochim Biophys Acta* **15**: 167–173.

Uchiyama T, Miyazaki K. (2009). Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**: 616–622.

Wellenreuther R, Schupp I, Poustka A, Wiemann S. (2004). SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. *BMC Genomics* **5**: 36.

Werdin-Pfisterer NR, Kielland K, Boone RD. (2009). Soil amino acid composition across a boreal forest successional sequence. *Soil Biol Biochem* **41**: 1210–1220.

Yu Z, Zhang Q, Kraus TEC, Dahlgren RA, Anastasio C, Zasoski RJ. (2002). Contribution of amino compounds to dissolved organic nitrogen in forest soils. *Biogeochem* **61**: 173–198.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)