High Genetic Potential for Proteolytic Decomposition in Northern Peatland Ecosystems

Emily B. Graham,[a] Fan Yang,[b] Sheryl Bell,[a] Kirsten S. Hofmockel[a,c]

aPacific Northwest National Laboratory, Richland, Washington, USA
bDepartment of Agricultural & Biosystems Engineering, Iowa State University, Ames, Iowa, USA
cDepartment of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa, USA

ABSTRACT Nitrogen (N) is a scarce nutrient commonly limiting primary productivity. Microbial decomposition of complex carbon (C) into small organic molecules (e.g., free amino acids) has been suggested to supplement biologically fixed N in northern peatlands. We evaluated the microbial (fungal, bacterial, and archaeal) genetic potential for organic N depolymerization in peatlands at Marcell Experimental Forest (MEF) in northern Minnesota. We used guided gene assembly to examine the abundance and diversity of protease genes and further compared them to those of N fixation (*nifH*) genes in shotgun metagenomic data collected across depths and in two distinct peatland environments (bogs and fens). Microbial protease genes greatly outnumbered *nifH* genes, with the most abundant genes (archaeal M1 and bacterial trypsin [S01]) each containing more sequences than all sequences attributed to *nifH*. Bacterial protease gene assemblies were diverse and abundant across depth profiles, indicating a role for bacteria in releasing free amino acids from peptides through depolymerization of older organic material and contrasting with the paradigm of fungal dominance in depolymerization in forest soils. Although protease gene assemblies for fungi were much less abundant overall than those for bacteria, fungi were prevalent in surface samples and therefore may be vital in degrading large soil polymers from fresh plant inputs during the early stage of depolymerization. In total, we demonstrate that depolymerization enzymes from a diverse suite of microorganisms, including understudied bacterial and archaeal lineages, are prevalent within northern peatlands and likely to influence C and N cycling.

IMPORTANCE Nitrogen (N) is a common limitation on primary productivity, and its source remains unresolved in northern peatlands that are vulnerable to environmental change. Decomposition of complex organic matter into free amino acids has been proposed as an important N source, but the genetic potential of microorganisms mediating this process has not been examined. Such information can inform possible responses of northern peatlands to environmental change. We show high genetic potential for microbial production of free amino acids across a range of microbial guilds in northern peatlands. In particular, the abundance and diversity of bacterial genes encoding proteolytic activity suggest a predominant role for bacteria in regulating productivity and contrasts with a paradigm of fungal dominance of organic N decomposition. Our results expand our current understanding of coupled carbon and nitrogen cycles in northern peatlands and indicate that understudied bacterial and archaeal lineages may be central in this ecosystem's response to environmental change.

KEYWORDS Marcell Experimental Forest, SPRUCE, carbon cycling, metagenomics, microbial community, microbiome, nitrogen cycling, organic nitrogen, soil carbon, soil organic matter

Understanding the processes that govern coupled carbon (C) and nutrient dynamics in northern peatlands is critical to predicting future biogeochemical cycles. These ecosystems account for 15 to 30% of global soil carbon storage (1–3), occurring primarily within layers of partially decomposed plant materials where the nitrogen (N) content is low (4, 5). Nitrogen is a critical nutrient regulating primary productivity in many terrestrial ecosystems (6) and can dictate below-ground carbon storage through impacts on soil organic matter decomposition (7, 8). Ombrotrophic peatlands are characterized by *Sphagnum* moss that has a comparatively large N requirement (approximately 40 to 50 kg ha^{-1} year^{-1} of N [9–12]). Nitrogen fixation historically has been considered to be the primary N source in peatlands (13–19). However, previous work has shown that N fixation alone cannot meet peatland N requirements (5, 20), and many studies have demonstrated the importance of organic molecules in fulfilling N demand (21–26). Symbiotic fungi have been associated with organic N acquisition (23, 26), but there is an increasing appreciation for the role of bacteria in this process. Despite these advances, our understanding of the genetic mechanisms mediating N availability remains nascent. We address this knowledge gap by exploring the genetic potential of peatland microbiomes to decompose polymeric organic N and subsequently influence peatland C and N cycles.

Depolymerization of proteinaceous organic material is an important pathway for generating bioavailable N in wide range of systems, including boreal forests, and is often considered a fungal trait (5, 7, 21, 23, 27–29). Depolymerization decomposes polymeric organic material into monomers and amino acids that can be used as C and N sources by soil microorganisms and plants (20, 21, 23, 30, 31). Several studies from terrestrial ecosystems under strong inorganic N limitation have shown that organic N, and free amino acids in particular, can be used directly by plants (20, 21, 23, 30, 31). Additionally, microorganisms (defined here as bacteria, archaea, and fungi) secrete extracellular proteases into soils to carry out organic matter depolymerization. Proteases are highly diverse and ubiquitous in soil and provide a large proportion of bioavailable N (32, 33). These enzymes catalyze the initial hydrolysis of proteins into smaller organic molecules such as oligopeptides and amino acids that can be subsequently acquired by plants (32).

In peatlands, fungi are considered more important than bacteria or archaea in proteolytic activity and decomposition more generally, particularly within the surface layer (34–36). Symbiotic ectomycorrhizal and ericoid fungi (EEM), which are supplied with C by a host plant, are especially relevant to organic N depolymerization in peatlands through N mining (37–39). EEM have been suggested to acquire N from soil organic matter (20, 40–42) and enable plants to directly compete with free-living microorganisms for N (43–45) to such an extent that Orwin et al. (46) posited a critical role for EEM in generating microbial N limitation of decomposition by enhancing plant N uptake. This fungally mediated plant organic N uptake may be particularly important in N-poor boreal ecosystems (20, 21). In these systems, free-living microorganisms should retain amino acids for growth instead of mineralizing organic N (23, 47). However, empirical evidence supporting the notion that fungi dominate proteolytic activity is sparse and derived primarily from correlative studies.

Little is known about the roles that various other microorganisms may play in peatland organic N depolymerization, and the genes that encode microbial proteases may provide valuable insight into the coupling of C and N cycles in these systems. Recent work has suggested substantial involvement of bacteria in protein depolymerization. For example, Lin et al. (35, 48) indicated that bacteria may outcompete fungal communities for plant-derived substrates, including large polymeric molecules. Consistent with this work, Bragina et al. (49) demonstrated that peatland *Sphagnum* moss microbiomes contain a high abundance of genes involved in N cycling and recalcitrant organic matter decomposition. The involvement of archaeal proteases in peatland organic N decomposition remains largely unexplored.

Here, we evaluated microbial proteolytic potential relative to N fixation within the Marcell Experimental Forest (MEF) by examining the genes encoding a suite of micro-

TABLE 1 Distribution of reads mapped to contigs identified as housekeeping genes, N fixation genes, and protease genes and the distributions of all mapped reads among archaea, bacteria, and fungi

Sample	Peatland depth (cm)	No. of reads			Total		
		Housekeeping genes	N fixation genes	Protease genes	Archaea	Bacteria	Fungi
Fen1_−10	Fen −10	3,822	559	10,296	1,851	12,754	72
Fen2_−10	Fen −10	2,506	469	5,757	894	7,701	137
T3M_−10	Bog −10	2,680	339	7,704	1,418	9,201	104
T3F_−10	Bog −10	2,785	736	6,194	663	9,003	49
T3F_−50	Bog −50	8,003	1,061	22,256	4,733	26,538	49
T3F_−100	Bog −100	4,174	266	11,420	4,933	10,888	39

bial proteases versus the *nifH* gene, with is well attributed to N fixation. Because the peatlands are highly acidic (pH 3 to 4) and organic polymers have bulky structures, we focused on microbial genes encoding extracellular proteases previously identified as active in acidic environments (50–52). This led to the identification of 19 extracellular protease gene families according to the MEROPS database (53). We also included one intracellular protease gene (encoding aminopeptidase, M1 family) in the study due to the encoded protein's role in producing single amino acids during depolymerization (50–52). We evaluated the distributions of these protease genes, as well as three house-keeping genes and the *nifH* gene, in six metagenomes across two peatland environments and three depths. Though our statistical power was limited, we know of no other work investigating the genetic potential of peatlands for organic N degradation across microbial domains, and our results contrast with the existing paradigm of fungal dominance in proteolytic decomposition from upland ecosystems. We leverage work demonstrating differences in organic matter cycling between hydrologically defined environments within peatlands (e.g., bogs versus fens) as a basis for potential differences in proteolytic activity across environments (35, 54–56). Previous research has also shown declines in fungal biomass depth as oxygen and root exudates become depleted in peatland, so we examined differential vertical stratification patterns in proteases between microbial domains (34, 57, 58). Our objectives were to provide a foundation for characterizing the mechanisms driving N availability and C decomposition patterns in peatlands through investigating shifts in proteolytic potential across (i) fungal, bacterial, and archaeal domains, (ii) peatland environment types, and (iii) depth profiles.

RESULTS AND DISCUSSION

Nitrogen fixation has long been considered the primary N source for peatlands (13–18), but N fixation alone cannot meet ecosystem N requirements (5, 20). Similarly, N assimilation has been shown to exceed gross mineralization in northern ecosystems (20, 21), and intact amino acid assimilation has been recognized as a potentially important source to meet N demand (23, 31, 37, 59). Previous work suggested that microbial proteases may be a missing link in northern peatland C and N cycling (31, 60). We investigated the proteolytic potential of peatland microbiomes across depth and environment type. Our work contrasts with the paradigm of fungal dominance in depolymerization processes and suggests that niche complementarity among diverse microorganisms is likely to play a substantial role in C and N cycling within northern peatlands.

Out of 24 genes investigated, we successfully assembled sequences aligned to 13 genes (see Tables S1 and S2 in the supplemental material). The assembled groups include three housekeeping genes (*rplB*, *rpb2_4*, and *rpb2_7*), nine protease genes (eight of which encode extracellular proteases), and the *nifH* gene. Approximately 34% of the fully covered contigs were annotated as housekeeping genes, 5% were anno-tated as *nifH* genes, and 61% were annotated as protease genes. Among all meta-genomic reads mapped to the annotated fully covered contigs, approximately 83.5% were bacterial, 16% were archaeal, and 0.5% were fungal (Table 1).

We reveal unique niches for fungal, bacterial, and archaeal proteolytic potential, as housekeeping and protease genes from each kingdom showed distinct stratification pat-
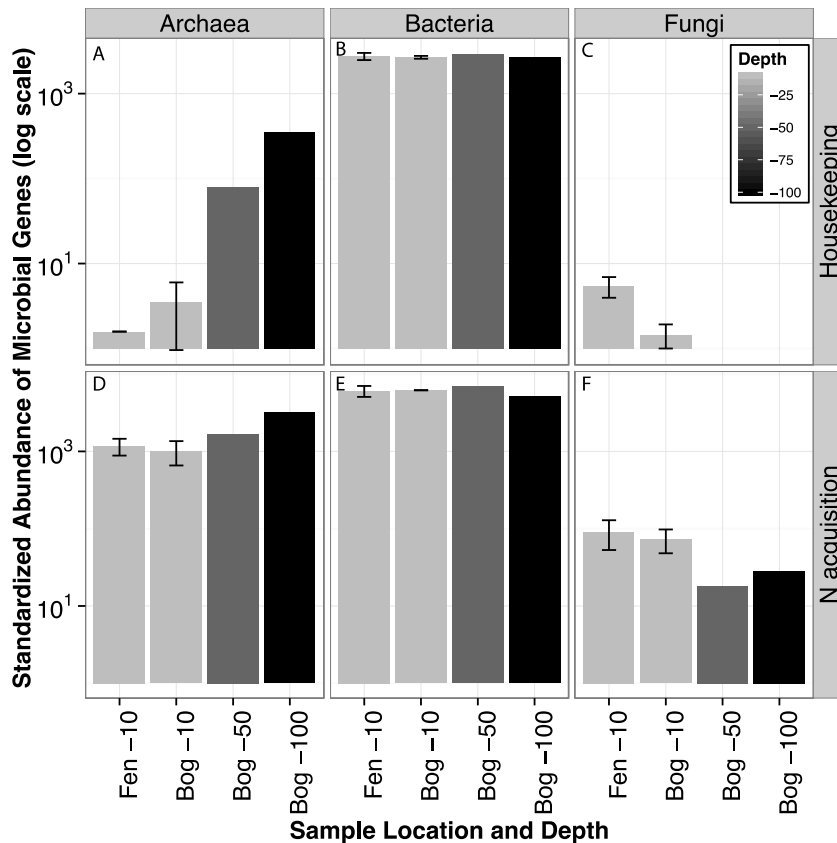
FIG 1 Standardized abundance of identified microbial genes in MEF peatlands through sampling depth. (A to C) Distribution of housekeeping genes. (D to F) Distribution of N acquisition genes. For location-depth combinations with two samples (Fen −10 and Bog −10), the bar height represents the average standardized gene abundance of two samples and the hatched lines represent the maximum and minimum values.

terns across depth. Within housekeeping genes, the standardized abundance of bacterial genes was similar across depth profiles, whereas the standardized fungal gene abundance decreased and archaeal gene abundance increased along the sampling depth (Fig. 1). Protease-encoding genes differed among archaea, bacteria, and fungi (Fig. 2). Among nine protease genes, eight were identified in bacteria, six were identified in fungi, and only three were identified in archaea. With some exceptions, archaeal protease genes increased with sampling depth, fungal protease genes decreased with depth, and bacterial protease genes varied throughout depth profiles.

Fungi were mostly found in the acrotelm, which constitutes the peat surface and is more oxygenated than deeper peat layers (Fig. 2). The acrotelm contains higher concentrations of C inputs from newly derived plant material, such as lignin-, cellulose-, and protein-based polymers (61). Lignin in particular requires oxygen for decomposition due to its comparatively high chemical complexity, such that some bonds cannot be readily cleaved by hydrolases or reductases, and other work has indicated an association between fungal proteases and lignin decomposition under N-limited conditions (62, 63). Previous work in this system has also shown that carbohydrates are enriched in the surface layer, while amino sugars and saccharides increase with depth (35). We therefore suggest that fungi are particularly relevant players in early stages of decomposition, in which fresh plant material is degraded into small oligopeptides, consistent with other evidence showing that fungi are unlikely to metabolize complex organic molecules found at depth in peatlands (58, 64).

Bacteria were the most abundant sequences detected regardless of depth or environment, in line with previous work by Lin et al. (55), and the presence of large numbers of
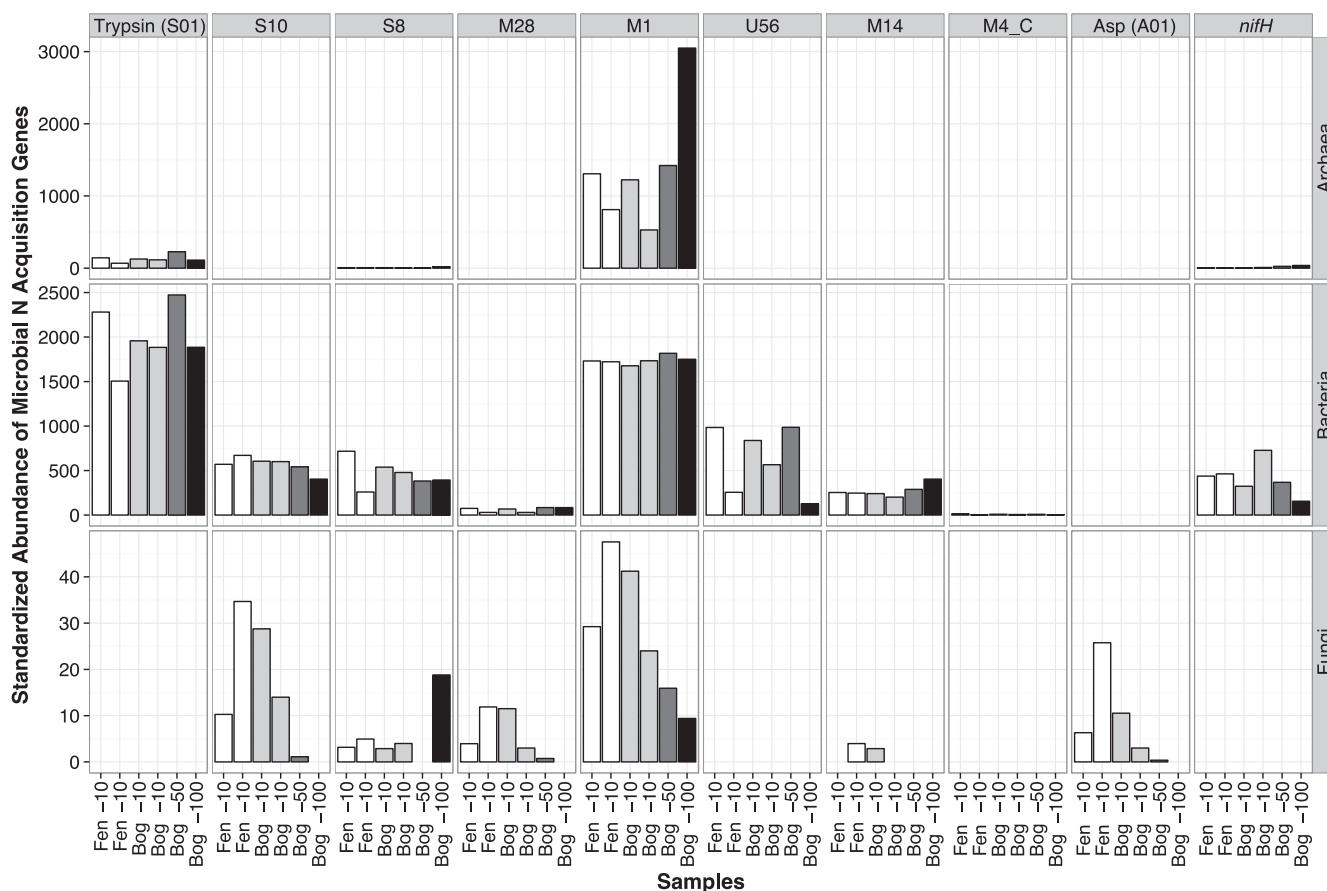
FIG 2 Abundance of identified nitrogen acquisition gene assemblies and their taxonomic distribution. Microbial proteases greatly outnumbered *nifH* genes (last column), with the most abundant genes (bacterial trypsin [S01] [column 1] and archaeal M1 [column 4]) each containing more sequences than all sequences attributed to *nifH*. Additionally, the relative abundance of most bacterial protease genes did not differ across depth profiles. Samples from the same depth and environment are colored identically, as denoted on the *x* axis. The sampling depth increases from left to right. A description of samples is given in Table 1.

bacterial proteases relative to those of fungi and archaea signifies a possible dominance of bacterial depolymerization in northern peatlands. Bacterial housekeeping genes and protease genes were approximately equally abundant throughout depth, and their sheer abundance suggests that at the community scale, bacteria may outcompete fungi and archaea in proteolytic decomposition. Additionally, bacterial diversity throughout the depth profile, despite relatively constant abundance, indicates plasticity in bacterial resource use across a variety of organic matter degradation states (Fig. 2).

Complementary to fungal and bacterial niche space, archaea had clear advantage in deep peat. Archaeal protease genes were more abundant in the mesotelm (25 to 50 cm) and catotelm (>50 cm) than in the acrotelm (0 to 10 cm). Lin et al. (55) noted the presence of archaea more generally at depth in peat, reaching up to 60% of total small-subunit rRNA gene sequences below 75 cm. Archaea are found in a variety of anaerobic and extreme environments, and specifically, methanogenic archaea are tolerant of low-oxygen conditions that persist in deep peat. Indeed, many of the protease genes we observed in deep peat were associated with methanogenic lineages (see Fig. S1 in the supplemental material). The presence of archaeal proteases at depth suggests that archaea may be vital to the decomposition of the oldest and most humified organic materials stored within peatlands. In total, the consistent differences in abundance and diversity of fungi, bacteria, and archaea across the three peat depths of the bog and fen suggest that niche partitioning across redox profiles may substantially influence the mechanisms of microbial decomposition.

With respect to differences in microbiomes across environments, we found few

TABLE 2 Standardized abundance of N acquisition genes in acrotelm peat samples[a]

Gene Categories	pfam	Average of two acrotelm replicates		Percent difference
		Bog	**Fen**	
	trypsin (S01)	2041.33	1998.47	2%
	S10	624.41	643.052	3%
	S8	516.76	497.76	4%
	M28	55.79	59.42	6%
Protease	M1	2614.43	2823.43	8%
	U56	701.53	619.60	13%
	M14	222.28	251.74	13%
	M4_C	7.31	9.49	30%
	asp (A01)	6.77	16.04	137%
Nitrogen Fixation	*nifH*	530.53	453.15	17%

[a]The values represent the average for two replicates of acrotelm samples from each geological location. The lighter color indicates a higher value. The percent difference was calculated as $100 \times (H/L - 1)$, where H is the high value and L is the low value.

differences in standardized gene abundance between bog ($n = 3$) and fen ($n = 2$) acrotelm samples, except that fungal genes were more abundant in fen acrotelm than in bog acrotelm samples (Fig. 1). However, three proteases (M14, M4_C, and asp [family A1]) were at least 12% more abundant in samples from the fen acrotelm than in those from bog acrotelm, and assemblies resembling U56 were less abundant in the fen than in the bog (Fig. 2). Remaining protease genes were less than 10% different across environments (Table 2).

Regardless of environment type or depth, bacterial protease abundance and diversity as a whole indicate a wide variety of possible niches for C- and N-cycling bacteria within peatlands. Aminopeptidase N (M1), which cleaves peptides and produces N-terminal amino acid residues (65, 66), was the most prevalent microbial protease. Work in other systems has shown that bacterial aminopeptidase N proteases can account for 99% of alanine released from substrate hydrolysis (67) and that they are critical in generating bioavailable organic N via microbial biomass turnover (68). Thus, we highlight the M1 gene family as encoding key enzymes for understanding peatland N cycles.

Beyond the protease M1 family, extracellular protease genes were highly diverse, and we propose that specific extracellular proteases we identified may fill unique steps in decomposition of plant material (Fig. 3) (69–71). In particular, asp (A01) genes were uniquely detected in fungi, while M4_C and U56 genes were found only in bacteria (Fig. 2). Each of these proteases contained a wide variety of genera with differences across both depth and environment type, and the abundance of asp (A01) genes in the acrotelm appeared to be inversely related to that of bacterial M4_C and U56 genes (Fig. 3). Previous work has shown that the fens at Marcell Experimental Forest have roughly 10% more dissolved organic C than bogs and that this difference in geochemistry explains most of the variation in microbiome composition between fen and bog samples (48). We note that fungal protease genes in particular are more abundant in the fen acrotelm than in the bog acrotelm (the layer in which most fungal biomass was found), supporting a role for fungal proteases in the decomposition of fresh carbon-rich plant material. asp (A01) genes are commonly associated with fungal wood decompo-
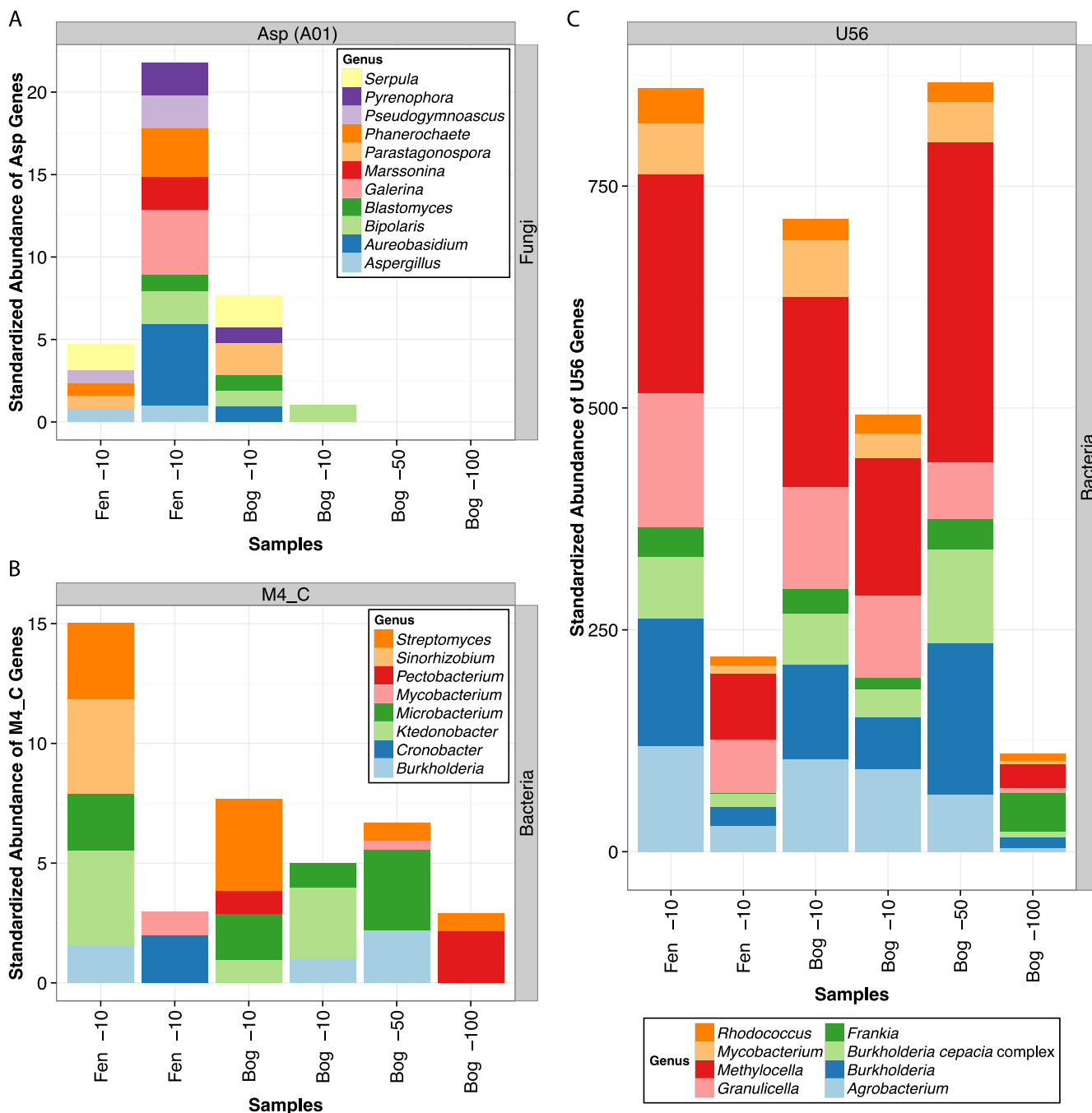
FIG 3 Phylogenetic distribution of the most abundant fungal asp (A01) (A), bacterial M4_C (B), and bacterial U56 (C) genes with variation across environments. Genus-level data are presented.

sition (62, 63). asp (A01) proteases may therefore play an important role in the early stage of peatland depolymerization, in which large polymeric molecules are degraded. asp (A01) genes consistent with the genera *Phanerochaete*, *Pseudogymnoascus*, and *Aspergillus* were detected in fen samples only (Fig. 3A), and their presence in the more C-rich fen environment may denote a distinct ecological niche for these organisms in proteolytic decomposition within fen-dominated peatlands.

Bacterial M4_C genes were particularly abundant in one fen acrotelm sample (Fen1_−10) and were taxonomically diverse within acrotelm samples more generally (Fig. 3B). M4_C genes are involved in the hydrolysis of proteins with large hydrophobic groups at P1′ (Table S2). Although bacterial and archaeal protease gene

abundances were mostly similar across environments, protease genes classified as U56 were 12% more abundant in the bog than in the fen and had diversity similar to that of the M4_C gene, with *Methylocella* and *Burkholderia* as the most abundant genera (Table 2 and Fig. 3C). We therefore propose that U56 (formerly linocin M18 [72]) may be an essential protease in peatlands. U56 is thought to hydrolyze chymotrypsin and trypsin into its component amino acids (Table S2). Though we did not explore the niches of these organisms beyond proteolytic activity, *Methylocella* spp. are commonly associated with methanotrophy (73), and *Burkholderia* spp. are functionally diverse but often are considered to be plant-associated nitrogen fixers (74, 75). The abundance of proteases associated with *Methylocella* and *Burkholderia* merits future investigation into their role in peatland biogeochemistry.

While we support previous work showing a role for N fixation in peatlands, particularly in surface peat (35), we suggest that microbial depolymerization complements N fixation in highly N-limited ecosystems. Our work is consistent with the recently proposed "LAH N acquisition strategy" framework, in which organic N decomposition supplements N fixation (76). The concept uses a cost-benefit framework to propose that proteolytic activity can be a preferred N acquisition strategy for microorganisms due to the high enzymatic cost of N fixation. We were able to assemble >21,000 archaeal and bacterial *nifH* genes, but *nifH* genes were less prevalent than the most abundant protease genes in each microbial domain (archaeal M1 and bacterial trypsin [S1 family]) (Fig. 2), in line with the observation that N accumulation in boreal ecosystems exceeded N deposition from atmosphere by 12- to 25-fold (10). *nifH* genes were detected in both archaea and bacteria, and approximately 97% of the *nifH* genes detected were bacterial. Nitrogen fixation is not known to be mediated by fungal communities (77). *Sphagnum* is known to harbor a diversity of N-fixing symbionts (49), including *Cyanobacteria* observed here (78), and is likely vital to atmospheric N fixation in peatlands.

Conclusion. We explored the genetic potential for organic matter depolymerization in a northern peatland based on previous work indicating that microbial proteases may be critical in understanding in C and N availability within these ecosystems (31, 60). We found that fungal protease genes were abundant in the acrotelm (surface layer), but bacterial proteolytic potential was orders of magnitude greater and distributed through depth profiles. Bacterial protease gene abundance was consistent across environments, and fungal protease genes were more prevalent in the higher-C fen environment. In contrast to the case for fungi and bacteria, the prevalence of archaeal protease genes at depth suggests an importance of these organisms in C and N availability below the rooting zone in peatlands. We also showed a diversity of protease genes that suggests strong niche complementarity among microorganisms with different physiologies. We identify proteases M1, U56, and asp (A01) as proteases that may be particularly important within northern peatlands. In total, protease genes greatly outnumbered *nifH* genes attributed to N fixation, emphasizing their role in peatland C and N cycles. We contrast the paradigm of fungal dominance in depolymerization processes and suggest that bacteria are imperative in releasing free amino acids from peptides through depolymerization of older organic material. Our work demonstrates a high genetic potential for depolymerization from a diverse suite of microorganisms beyond those typically considered, and we urge a broader perspective on the organisms mediating C and N cycles in northern peatlands.

MATERIALS AND METHODS

Sample description. A large-scale field manipulation experiment known as Spruce and Peatland Responses Under Changing Environments (SPRUCE) was initiated at the Marcell Experimental Forest (MEF), MN, USA, by the U.S. Department of Energy, U.S. Department of Agriculture (USDA) Forest Service, and Oak Ridge National Laboratory (http://mnspruce.ornl.gov/). The MEF itself is a 8.1-hectare acidic, forested bog (47°30′31.132′′N, 93°27′15.146′′W). Sites within the MEF are classified based on their trophic status and water source as ombrotrophic bogs (receiving precipitation only) or minerotrophic fens (fed by both groundwater and precipitation [48, 79]). Although fens are frequently considered more nutrient rich than bogs, both types of peatlands are highly limited in inorganic N. A full characterization

of the field site, including peatland hydrology and vegetation, is described by Sebestyen et al. (80). Further information on samples is available from Lin et al. (35, 55).

Six metagenomic libraries were obtained from MEF in February 2012 as described by Lin et al. (35). Briefly, peat cores were collected from hollows in bogs and fens and sectioned from 0 to 10 cm (acrotelm), 25 to 50 cm (mesotelm), and 75 to 100 cm (catotelm). While the water table was near the surface of the *Sphagnum* layer at the time of sampling, MEF experiences large seasonal variation in water table height, with the acrotelm and mesotelm layers frequently exposed to oxygen (81). For example, in the year of sampling, the water table depth fell to -20 cm below the *Sphagnum* layer (81), and microbiomes in the acrotelm contained many aerobic microorganisms (55). Each core section was homogenized. Two acrotelm samples (0 to -10 cm) were collected from bog lake fen (i.e., samples Fen1_-10 and Fen2_-10), two acrotelm samples (0 to -10 cm) from SPRUCE bog (i.e., samples T3M_-10 and T3F_-10), one mesotelm sample (-25 to -50 cm) from SPRUCE bog (i.e., sample T3F_-50), and one catotelm sample (-75 to -100 cm) from SPRUCE bog (i.e., sample T3F_-100). Sample collection is extremely limited at MEF to preserve the integrity of the SPRUCE experiment. As such, our sample numbers are not conducive to statistical analysis, but the metabolic insight generated by our in-depth metagenomic analysis merits deeper investigation into the respective roles of microbial domains in coupled C-N cycles in northern peatlands.

As described by Lin et al. (35, 55), samples differed in physicochemical properties across the depth layers. Sequencing coverage for each metagenome increased with depth and ranged from 42% to 86% (Table 1). Metagenomes were generally consistent between samples from the same depth and site, suggesting that the results reported here are likely to be robust (further details were provided by Lin et al. [35]). DNA-based approaches reflect microbial potential to catalyze a given process rather than gross activity rates; however, metagenomic sequences are often used as a proxy for identifying biogeochemical mechanisms of interest. The phylogenetic distribution of microorganisms in each sample at the community level was presented by Lin et al. (55), and that at the protease level is shown in Fig. S2 in the supplemental material. Microorganisms belonging to *Acidobacteria*, *Proteobacteria*, *Actinobacteria*, and *Verrucomicrobia* were common at both sites, with methanogenic archaea present at depth. Whole-genome shotgun metagenome sequences are available in MG-RAST (35, 55).

HMM construction. We used 24 hidden Markov models (HMMs) constructed based on protein sequences (82) to investigate the microbial genetic potential in N acquisition in MEF peatlands, 20 of which were for microbial protease genes, one for the nitrogenase gene (*nifH*), and three for microbial single-copy housekeeping genes (see Table S2 in the supplemental material). Nitrogenase enzymes are encoded by three genes (*nifH*, *nifD*, and *nifK* [reviewed in reference 83]); however *nifH* is the most commonly used marker gene for nitrogenase potential (77, 84). Evidence suggests that *nifD*- and *nifK*-based assays are consistent with *nifH*-based results (85). We use "N acquisition genes" and related terms to represent all protease genes and *nifH*. The abundance of N acquisition genes is therefore the sum of the abundance of all protease genes plus the abundance of *nifH*. The housekeeping genes used as bacterial, fungal, and archaeal markers are genes encoding ribosomal protein L2 (*rplB*), RNA polymerase second-largest subunit domain 4 (RPB2_4), and domain 7 (RPB2_7), respectively. We use the term "housekeeping genes" to represent these domain-specific single-copy housekeeping genes. The abundance of housekeeping genes is calculated as the sum of the abundance all housekeeping genes. The HMMs for *nifH* and *rplB* genes are available in Ribosomal Database Project (RDP) Fungene repository (86). Models for RPB2_4 and RPB2_7 were obtained from Pfam database (http://pfam.xfam.org). The protease genes (one intracellular and 19 extracellular) were selected based on literature characterizations (50–52). Though there are hundreds of protease genes in the MEROPS database (53), most encode intracellular enzymes not thought to be involved in decomposition or unsuitable for the acidic environments. We therefore targeted protease genes both known to encode functions relevant to peatland decomposition and known to be expressed and active in previous lab assays (50–52). To construct the most informative HMM models for targeted genes, well-studied gene sequences were first selected from existing literature. These seed sequences were cross-checked against the reviewed protein database Swiss-Prot (http://www.uniprot.org/). Genes encoding proteases were also searched against the MEROPS peptidase database (http://merops.sanger.ac.uk/) to confirm their protease identity. Protein families and existing protein HMM models were queried from Pfam database. The retrieved Pfam HMMs were then used to extrapolate archaeal, bacterial, and fungal reference protein sequences from the UniProt database. Pfam HMMs were used to search against Swiss-Prot at different cutoffs (E value or Bit score) to ensure model accuracy. If existing Pfam HMMs could not accurately query sequences, a set of well-annotated sequences would be used to construct new models (Table S2). Reference protein sequences were retrieved from UniProt and aligned using finalized HMMs.

Guided metagenomic assembly. All metagenomic reads were filtered by using RDP SeqFilters (87) to a minimal average read quality of $Q = 25$. Genes with HMM models (Table S2) were assembled from combined filtered reads by using a modified RDP Xander skeleton analysis pipeline (https://github.com/fishjord/xander_analysis_skel). Briefly, a De Bruijn graph was built for the combined shotgun metagenome data set. Potential gene start points (*K*-mer starts, $k = 30$ nucleotides) were identified from each gene reference sequence. Local assembling was carried out by searching constructed De Bruijn graphs at the given gene start points. These local assemblages were then merged to form the longest contigs possible. The final merged nucleotide sequences were dereplicated using CD-Hit 4.6.1 (-c 1.0) (88) to identify the longest unique contigs.

Data processing. All quality filtered reads were mapped (Bowtie2.2.5) (89) against the dereplicated merged contigs. Only contigs that were 100% covered (median base coverage = 1) were considered. The biological information of these fully covered contigs was identified using Basic Alignment Searching Tool

(blastx) 2.2.30+; the best-matching sequences with E values of $\leq 1 \times 10^{-5}$ were kept (90). UniProt (UniProtKB release 10, 2014) was used as the annotation database.

The final gene abundance per peat sample was determined by mapping reads from each metagenome to fully covered contigs. The mapping results indicated that ~9% of reads were mapped onto the contigs once only and the majority of the fungal housekeeping contigs were mapped once. Hence, for downstream analyses, we included all reads mapped to final fully covered contigs at least once (see Table S3 in the supplemental material). The mapped read abundances were standardized by sequencing depth for comparisons among samples. Gene abundance was used to infer the abundance of mapped reads to fully covered contigs. Final data analyses and visualization were done in R 3.1.0 (91) with packages plyr (92) and ggplot2 (93).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM .02851-18.

SUPPLEMENTAL FILE 1, PDF file, 0.3 MB.

REFERENCES

1. Limpens J, Berendse F, Blodau C, Canadell J, Freeman C, Holden J, Roulet N, Rydin H, Schaepman-Strub G. 2008. Peatlands and the carbon cycle: from local processes to global implications—a synthesis. Biogeosciences 5:1475–1491. https://doi.org/10.5194/bg-5-1475-2008.
2. Kivinen E, Pakarinen P. 1981. Geographical distribution of peat resources and major peatland complex types in the world. Suomalainen Tiedeakatemia, Helsinki, Finland.
3. Gorham E. 1991. Northern peatlands: role in the carbon cycle and probable responses to climatic warming. Ecol Appl 1:182–195. https://doi.org/10.2307/1941811.
4. Tamm CO. 1991. Nitrogen in terrestrial ecosystems, p 1–6. Springer, Basel, Switzerland.
5. Jones DL, Kielland K. 2002. Soil amino acid turnover dominates the nitrogen flux in permafrost-dominated taiga forest soils. Soil Biol Biochem 34:209–219. https://doi.org/10.1016/S0038-0717(01)00175-4.
6. LeBauer DS, Treseder KK. 2008. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. Ecology 89:371–379. https://doi.org/10.1890/06-2057.1.
7. Janssens IA, Dieleman W, Luyssaert S, Subke J-A, Reichstein M, Ceulemans R, Ciais P, Dolman AJ, Grace J, Matteucci G, Papale D, Piao SL, Schulze E-D, Tang J, Law BE. 2010. Reduction of forest soil respiration in response to nitrogen deposition. Nat Geosci 3:315. https://doi.org/10.1038/ngeo844.
8. Knicker H. 2011. Soil organic N—an under-rated player for C sequestration in soils? Soil Biol Biochem 43:1118–1129. https://doi.org/10.1016/j.soilbio.2011.02.020.
9. Wieder KR, Vitt DH, Burke-Scoll M, Scott KD, House M, Vile MA. 2010. Nitrogen and sulphur deposition and the growth of Sphagnum fuscum in bogs of the Athabasca Oil Sands Region, Alberta. J Limnol 69:161–170. https://doi.org/10.4081/jlimnol.2010.s1.161.
10. Vile MA, Kelman Wieder R, Živković T, Scott KD, Vitt DH, Hartsock JA, Iosue CL, Quinn JC, Petix M, Fillingim HM, Popma JMA, Dynarski KA, Jackman TR, Albright CM, Wykoff DD. 2014. N$_2$-fixation by methanotrophs sustains carbon and nitrogen accumulation in pristine peatlands. Biogeochemistry 121:317–328. https://doi.org/10.1007/s10533-014-0019-6.
11. Turetsky MR, Bond-Lamberty B, Euskirchen E, Talbot J, Frolking S, McGuire AD, Tuittila ES. 2012. The resilience and functional role of moss in boreal and arctic ecosystems. New Phytol 196:49–67. https://doi.org/10.1111/j.1469-8137.2012.04254.x.
12. Van Breemen N. 1995. How Sphagnum bogs down other plants. Trends Ecol Evol 10:270–275. https://doi.org/10.1016/0169-5347(95)90007-1.
13. Gundale MJ, Nilsson M, Bansal S, Jäderlund A. 2012. The interactive effects of temperature and light on biological nitrogen fixation in boreal forests. New Phytol 194:453–463. https://doi.org/10.1111/j.1469-8137.2012.04071.x.
14. Berg A, Danielsson Å, Svensson BH. 2013. Transfer of fixed-N from N$_2$-fixing cyanobacteria associated with the moss Sphagnum riparium results in enhanced growth of the moss. Plant Soil 362:271–278. https://doi.org/10.1007/s11104-012-1278-4.
15. Aerts R, Wallen B, Malmer N. 1992. Growth-limiting nutrients in Sphagnum-dominated bogs subject to low and high atmospheric nitrogen supply. J Ecol 80:131–140. https://doi.org/10.2307/2261070.
16. Schwintzer CR. 1983. Nonsymbiotic and symbiotic nitrogen fixation in a weakly minerotrophic peatland. Am J Bot 70:1071–1078. https://doi.org/10.2307/2442817.
17. Waughman G, Bellamy D. 1980. Nitrogen fixation and the nitrogen balance in peatland ecosystems. Ecology 61:1185–1198. https://doi.org/10.2307/1936837.
18. Ininbergs K, Bay G, Rasmussen U, Wardle DA, Nilsson MC. 2011. Composition and diversity of nifH genes of nitrogen-fixing cyanobacteria associated with boreal forest feather mosses. New Phytol 192:507–517. https://doi.org/10.1111/j.1469-8137.2011.03809.x.
19. Larmola T, Leppänen SM, Tuittila E-S, Aarva M, Merilä P, Fritze H, Tiirola M. 2014. Methanotrophy induces nitrogen fixation during peatland development. Proc Natl Acad Sci U S A 111:734–739. https://doi.org/10.1073/pnas.1314284111.
20. Näsholm T, Ekblad A, Nordin A, Giesler R, Högberg M, Högberg P. 1998. Boreal forest plants take up organic nitrogen. Nature 392:914. https://doi.org/10.1038/31921.
21. Näsholm T, Kielland K, Ganeteg U. 2009. Uptake of organic nitrogen by plants. New Phytol 182:31–48. https://doi.org/10.1111/j.1469-8137.2008.02751.x.
22. Lipson D, Näsholm T. 2001. The unexpected versatility of plants: organic nitrogen use and availability in terrestrial ecosystems. Oecologia 128:305–316. https://doi.org/10.1007/s004420100693.
23. Schimel JP, Bennett J. 2004. Nitrogen mineralization: challenges of a changing paradigm. Ecology 85:591–602. https://doi.org/10.1890/03-8002.
24. Gärdenäs AI, Ågren GI, Bird JA, Clarholm M, Hallin S, Ineson P, Kätterer T, Knicker H, Nilsson SI, Näsholm T, Ogle S, Paustian K, Persson T, Stendahl J. 2011. Knowledge gaps in soil carbon and nitrogen interactions—from molecular to global scale. Soil Biol Biochem 43:702–717. https://doi.org/10.1016/j.soilbio.2010.04.006.
25. Czaban W, Jämtgård S, Näsholm T, Rasmussen J, Nicolaisen M, Fomsgaard IS. 2016. Direct acquisition of organic N by white clover even in the presence of inorganic N. Plant Soil 407:91–107. https://doi.org/10.1007/s11104-016-2896-z.

26. Averill C. 2014. Divergence in plant and microbial allocation strategies explains continental patterns in microbial allocation and biogeochemical fluxes. Ecol Lett 17:1202–1210. https://doi.org/10.1111/ele.12324.

27. Jones DL, Kielland K. 2012. Amino acid, peptide and protein mineralization dynamics in a taiga forest soil. Soil Biol Biochem 55:60–69. https://doi.org/10.1016/j.soilbio.2012.06.005.

28. Kieloaho A-J, Pihlatie M, Carrasco MD, Kanerva S, Parshintsev J, Riekkola M-L, Pumpanen J, Heinonsalo J. 2016. Stimulation of soil organic nitrogen pool: the effect of plant and soil organic matter degrading enzymes. Soil Biol Biochem 96:97–106. https://doi.org/10.1016/j.soilbio.2016.01.013.

29. Andresen L, Bode S, Tietema A, Boeckx P, Rütting T. 2015. Amino acid and N mineralization dynamics in heathland soil after long-term warming and repetitive drought. Soil 1:341–349. https://doi.org/10.5194/soil-1-341-2015.

30. Kielland K. 1997. Role of free amino acids in the nitrogen economy of arctic cryptogams. Ecoscience 4:75–79. https://doi.org/10.1080/11956860.1997.11682379.

31. Hofmockel KS, Fierer N, Colman BP, Jackson RB. 2010. Amino acid abundance and proteolytic potential in North American soils. Oecologia 163:1069–1078. https://doi.org/10.1007/s00442-010-1601-9.

32. Ladd J, Butler J. 1972. Short-term assays of soil proteolytic enzyme activities using proteins and dipeptide derivatives as substrates. Soil Biol Biochem 4:19–30. https://doi.org/10.1016/0038-0717(72)90038-7.

33. Geisseler D, Horwath WR. 2008. Regulation of extracellular protease activity in soil in response to different sources and concentrations of nitrogen and carbon. Soil Biol Biochem 40:3040–3048. https://doi.org/10.1016/j.soilbio.2008.09.001.

34. Andersen R, Chapman S, Artz R. 2013. Microbial communities in natural and disturbed peatlands: a review. Soil Biol Biochem 57:979–994. https://doi.org/10.1016/j.soilbio.2012.10.003.

35. Lin X, Tfaily MM, Green SJ, Steinweg JM, Chanton P, Imvittaya A, Chanton JP, Cooper W, Schadt C, Kostka JE. 2014. Microbial metabolic potential for carbon degradation and nutrient (nitrogen and phosphorus) acquisition in an ombrotrophic peatland. Appl Environ Microbiol 80:3531–3540. https://doi.org/10.1128/AEM.00206-14.

36. Bajwa R, Abuarghub S, Read D. 1985. The biology of mycorrhiza in the Ericaceae. New Phytol 101:469–486. https://doi.org/10.1111/j.1469-8137.1985.tb02853.x.

37. Hobbie EA, Ouimette AP, Schuur EA, Kierstead D, Trappe JM, Bendiksen K, Ohenoja E. 2013. Radiocarbon evidence for the mining of organic nitrogen from soil by mycorrhizal fungi. Biogeochemistry 114:381–389.

38. Lindahl BD, Tunlid A. 2015. Ectomycorrhizal fungi—potential organic matter decomposers, yet not saprotrophs. New Phytol 205:1443–1447. https://doi.org/10.1111/nph.13201.

39. Treseder KK, Torn MS, Masiello CA. 2006. An ecosystem-scale radiocarbon tracer to test use of litter carbon by ectomycorrhizal fungi. Soil Biol Biochem 38:1077–1082. https://doi.org/10.1016/j.soilbio.2005.09.006.

40. Hodge A, Campbell CD, Fitter AH. 2001. An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material. Nature 413:297. https://doi.org/10.1038/35095041.

41. Read D, Perez-Moreno J. 2003. Mycorrhizas and nutrient cycling in ecosystems—a journey towards relevance? New Phytol 157:475–492. https://doi.org/10.1046/j.1469-8137.2003.00704.x.

42. Averill C, Finzi A. 2011. Increasing plant use of organic nitrogen with elevation is reflected in nitrogen uptake rates and ecosystem δ15N. Ecology 92:883–891. https://doi.org/10.1890/10-0746.1.

43. Averill C, Turner BL, Finzi AC. 2014. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. Nature 505:543. https://doi.org/10.1038/nature12901.

44. Lindahl BD, De Boer W, Finlay RD. 2010. Disruption of root carbon transport into forest humus stimulates fungal opportunists at the expense of mycorrhizal fungi. ISME J 4:872. https://doi.org/10.1038/ismej.2010.19.

45. Gadgil RL, Gadgil P. 1971. Mycorrhiza and litter decomposition. Nature 233:133. https://doi.org/10.1038/233133a0.

46. Orwin KH, Kirschbaum MU, St John MG, Dickie IA. 2011. Organic nutrient uptake by mycorrhizal fungi enhances ecosystem carbon storage: a model-based assessment. Ecol Lett 14:493–502. https://doi.org/10.1111/j.1461-0248.2011.01611.x.

47. Schimel JP, Chapin FS. 1996. Tundra plant uptake of amino acid and NH4+ nitrogen in situ: plants complete well for amino acid N. Ecology 77:2142–2147. https://doi.org/10.2307/2265708.

48. Lin X, Green S, Tfaily M, Prakash O, Konstantinidis K, Corbett J, Chanton J, Cooper W, Kostka J. 2012. Microbial community structure and activity linked to contrasting biogeochemical gradients in bog and fen environments of the Glacial Lake Agassiz Peatland. Appl Environ Microbiol 78:7023–7031. https://doi.org/10.1128/AEM.01750-12.

49. Bragina A, Oberauner-Wappis L, Zachow C, Halwachs B, Thallinger GG, Müller H, Berg G. 2014. The Sphagnum microbiome supports bog ecosystem functioning under extreme conditions. Mol Ecol 23:4498–4510. https://doi.org/10.1111/mec.12885.

50. Wildeboer D, Jeganathan F, Price RG, Abuknesha RA. 2009. Characterization of bacterial proteases with a panel of fluorescent peptide substrates. Anal Biochem 384:321–328. https://doi.org/10.1016/j.ab.2008.10.004.

51. Togni G, Sanglard D, Falchetto R, Monod M. 1991. Isolation and nucleotide sequence of the extracellular acid protease gene (ACP) from the yeast Candida tropicalis. FEBS Lett 286:181–185. https://doi.org/10.1016/0014-5793(91)80969-A.

52. Wandersman C. 1989. Secretion, processing and activation of bacterial extracellular proteases. Mol Microbiol 3:1825–1831. https://doi.org/10.1111/j.1365-2958.1989.tb00169.x.

53. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res 46:D624–D632. https://doi.org/10.1093/nar/gkx1134.

54. Bridgham SD, Pastor J, Updegraff K, Malterer TJ, Johnson K, Harth C, Chen J. 1999. Ecosystem control over temperature and energy flux in northern peatlands. Ecol Appl 9:1345–1358.2.0.CO;2]. https://doi.org/10.1890/1051-0761(1999)009[1345:ECOTAE]2.0.CO;2.

55. Lin X, Tfaily MM, Steinweg JM, Chanton P, Esson K, Yang ZK, Chanton JP, Cooper W, Schadt CW, Kostka JE. 2014. Microbial community stratification linked to utilization of carbohydrates and phosphorus limitation in a boreal peatland at Marcell Experimental Forest, Minnesota, USA. Appl Environ Microbiol 80:3518–3530. https://doi.org/10.1128/AEM.00205-14.

56. Weltzin JF, Pastor J, Harth C, Bridgham SD, Updegraff K, Chapin CT. 2000. Response of bog and fen plant communities to warming and water-table manipulations. Ecology 81:3464–3478. https://doi.org/10.1890/0012-9658(2000)081[3464:ROBAFP]2.0.CO;2.

57. Lamit LJ, Romanowicz KJ, Potvin LR, Rivers AR, Singh K, Lennon JT, Tringe SG, Kane ES, Lilleskov EA. 2017. Patterns and drivers of fungal community depth stratification in Sphagnum peat. FEMS Microbiol Ecol 93:fix082. https://doi.org/10.1093/femsec/fix082.

58. Turetsky MR, Wieder RK, Williams CJ, Vitt DH. 2000. Organic matter accumulation, peat chemistry, and permafrost melting in peatlands of boreal Alberta. Ecoscience 7:115–122. https://doi.org/10.1080/11956860.2000.11682608.

59. Hofmockel KS, Schlesinger WH, Jackson RB. 2007. Effects of elevated atmospheric carbon dioxide on amino acid and NH4+-N cycling in a temperate pine ecosystem. Glob Chang Biol 13:1950–1959. https://doi.org/10.1111/j.1365-2486.2007.01411.x.

60. Kielland K, McFarland J, Ruess R, Olson K. 2007. Rapid cycling of organic nitrogen in taiga forest ecosystems. Ecosystems 10:360–368. https://doi.org/10.1007/s10021-007-9037-8.

61. Berg B, McClaugherty C. 2003. Plant litter. Springer, Berlin, Germany.

62. Dosoretz CG, Chen H-C, Grethlein HE. 1990. Effect of environmental conditions on extracellular protease activity in lignolytic cultures of Phanerochaete chrysosporium. Appl Environ Microbiol 56:395–400.

63. Sato S, Liu F, Koc H, Tien M. 2007. Expression analysis of extracellular proteins from Phanerochaete chrysosporium grown on different liquid and solid substrates. Microbiology 153:3023–3033. https://doi.org/10.1099/mic.0.2006/000513-0.

64. Thormann MN, Currah RS, Bayley SE. 2002. The relative ability of fungi from Sphagnum fuscum to decompose selected carbon substrates. Can J Microbiol 48:204–211. https://doi.org/10.1139/w02-010.

65. Gonzales T, Robert-Baudouy J. 1996. Bacterial aminopeptidases: properties and functions. FEMS Microbiol Rev 18:319–344. https://doi.org/10.1111/j.1574-6976.1996.tb00247.x.

66. Rao MB, Tanksale AM, Ghatge MS, Deshpande VV. 1998. Molecular and biotechnological aspects of microbial proteases. Microbiol Mol Biol Rev 62:597–635.

67. Addlagatta A, Gay L, Matthews BW. 2006. Structure of aminopeptidase N from Escherichia coli suggests a compartmentalized, gated active site. Proc Natl Acad Sci U S A 103:13339–13344. https://doi.org/10.1073/pnas.0606167103.

68. Weedon JT, Aerts R, Kowalchuk GA, van Logtestijn R, Andringa D, van

Bodegom PM. 2013. Temperature sensitivity of peatland C and N cycling: does substrate supply play a role? Soil Biol Biochem 61:109–120. https://doi.org/10.1016/j.soilbio.2013.02.019.

69. Tan PT, Poolman B, Konings WN. 1993. Proteolytic enzymes of Lactococcus lactis. J Dairy Res 60:269–286. https://doi.org/10.1017/S0022029900027606.

70. Niven G, Holder S, Strøman P. 1995. A study of the substrate specificity of aminopeptidase N from Lactococcus lactis subsp. cremoris Wg2. Appl Microbiol Biotechnol 44:100–105. https://doi.org/10.1007/BF00164487.

71. Law J, Haandrikman A. 1997. Proteolytic enzymes of lactic acid bacteria. Int Dairy J 7:1–11. https://doi.org/10.1016/0958-6946(95)00073-9.

72. Valdés-Stauber N, Scherer S. 1994. Isolation and characterization of Linocin M18, a bacteriocin produced by Brevibacterium linens. Appl Environ Microbiol 60:3809–3814.

73. Dedysh SN, Knief C, Dunfield PF. 2005. Methylocella species are facultatively methanotrophic. J Bacteriol 187:4665–4670. https://doi.org/10.1128/JB.187.13.4665-4670.2005.

74. Coenye T, Vandamme P. 2003. Diversity and significance of Burkholderia species occupying diverse ecological niches. Environ Microbiol 5:719–729. https://doi.org/10.1046/j.1462-2920.2003.00471.x.

75. Estrada-De Los Santos P, Bustillos-Cristales RO, Caballero-Mellado J. 2001. Burkholderia, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. Appl Environ Microbiol 67:2790–2798. https://doi.org/10.1128/AEM.67.6.2790-2798.2001.

76. Norman JS, Friesen ML. 2017. Complex N acquisition by soil diazotrophs: how the ability to release exoenzymes affects N fixation by terrestrial free-living diazotrophs. ISME J 11:315. https://doi.org/10.1038/ismej.2016.127.

77. Gaby JC, Buckley DH. 2011. A global census of nitrogenase diversity. Environ Microbiol 13:1790–1799. https://doi.org/10.1111/j.1462-2920.2011.02488.x.

78. Granhall U, Hofsten AV. 1976. Nitrogenase activity in relation to intracellular organisms in Sphagnum mosses. Physiol Plant 36:88–94. https://doi.org/10.1111/j.1399-3054.1976.tb05033.x.

79. Charman D. 2002. Peatlands and environmental change. John Wiley & Sons Ltd, New York, NY.

80. Sebestyen SD, Dorrance C, Olson DM, Verry ES, Kolka RK, Elling AE, Kyllander R. 2011. Long-term monitoring sites and trends at the Marcell Experimental Forest, p 15–72. In Kolka RK, Sebestyen SD, Verry ES, Brooks KN (ed), Peatland biogeochemistry and watershed hydrology at the Marcell Experimental Forest. CRC Press, Boca Raton, FL.

81. Iversen CM, Childs J, Norby RJ, Ontl TA, Kolka RK, Brice DJ, McFarlane KJ, Hanson PJ. 2018. Fine-root growth in a forested bog is seasonally dynamic, but shallowly distributed in nutrient-poor peat. Plant Soil 424:123–143. https://doi.org/10.1007/s11104-017-3231-z.

82. Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11:431. https://doi.org/10.1186/1471-2105-11-431.

83. Rubio LM, Ludden PW. 2002. The gene products of the nif regulon, p 101–136. In Leigh GJ (ed), Nitrogen fixation at the millenium. Elsevier, Amsterdam, The Netherlands.

84. Gaby JC, Buckley DH. 2012. A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase. PLoS One 7:e42149. https://doi.org/10.1371/journal.pone.0042149.

85. Young J. 2005. The phylogeny and evolution of nitrogenases, p 221–241. In Palacios R, Newton WE (ed), Genomes and genomics of nitrogen-fixing organisms. Springer, Heidelberg, Germany.

86. Fish JA, Chai B, Wang Q, Sun Y, Brown CT, Tiedje JM, Cole JR. 2013. FunGene: the functional gene pipeline and repository. Front Microbiol 4:291. https://doi.org/10.3389/fmicb.2013.00291.

87. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42:D633–D642. https://doi.org/10.1093/nar/gkt1244.

88. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

89. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357. https://doi.org/10.1038/nmeth.1923.

90. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

91. R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

92. Wickham H. 2011. The split-apply-combine strategy for data analysis. J Stat Softw 40:1–29.

93. Wickham H. 2009. ggplot2, p 115–137. Springer, New York, NY.