Supporting Results

Shifts in eggNOG genes with SSD

Among the genes annotated with eggNOG, replication, recombination and repair (L) was the most abundant category (relative abundance 16%) across all the SSDs, followed by energy production and conversion (C; 11.5%), translation, ribosomal structure and biogenesis (J; 11.4%), and amino acid transport and metabolism (E; 10.0%). The relative abundances of these categories did not differ between the SSDs (Figure S11; Table S12).

Differential abundance analysis conducted with DESeq2 was used to evaluate changes in the relative abundance of functional genes between two SSDs (only COG genes were considered here). Among the total 1939 functional genes annotated with eggNOG, there were more genes that differented in abundant (P<0.05) between barren soils and biological soil crusts, between barren and vegetated soils, and between biological soil crusts and vegetated soils than other comparisons (Table 5). The same was also true for the DESeq2 analysis of the genes annotated with CAZy and NCyc. Therefore, here we only report results for these three pairwise comparisons.

The functional category "post-translational modification, protein turnover, and chaperones (O)" did not differ significantly in abundance between the different SSDs (Table S6). However, at the gene level, many genes involved in this functional category differed significantly in abundance between the SSDs. For example, COG0068 (hydrogenase maturation protein) was overrepresented in the biological soil crusts compared with barren soils, while COG0225 (a repair enzyme for proteins that have been deactivated by oxidation) was overrepresented in the vegetated soils compared with the biological soil crusts (Table S6). Detailed results of the $\log_2$-fold changes (LFCs) of all genes and comparisons can be found in Table S6.

Shifts in CAZy genes with SSD

Among the genes annotated using CAZy, the most abundant enzyme class was the glycoside hydrolases (GH, 41.7%, a widespread group of enzymes that hydrolyze the glycosidic bond, which can degrade substrates such as cellulose, starch and hemicellulose), followed by glycosyl transferases (GT, 37%, responsible for catalyzing the formation of the glycosidic linkage to form a glycoside), carbohydrate-binding modules (CBM, 11.5%, previously classified as a cellulose-binding domain based on the initial discovery of several cellulose-binding modules), carbohydrate esterases (CE, 5.4%, a group of enzymes that catalyze the de-O or de-N-acylation of substituted saccharides), auxiliary activities (AA, 2.9%, responsible

for degrading lignin), polysaccharide lyases (PL, 1.5%, a group of enzymes that cleave polysaccharide chains containing uronic acid via a β-elimination mechanism to generate an unsaturated hexenuronic acid residue and a new reducing end [http://www.cazy.org/], can degrade starch and pectin) (Figure S12A).
In total we found 239 different CAZy families, including 110 GHs, 47 GTs, 44 CBMs, 14 CEs, 10 AAs and 14 PLs. At the family level, AAs were especially abundant with AA2 and AA3. CEs were especially abundant with CE4 and CE11. GHs were abundant with GH3, GH13 and GH94. The most abundant gene families were GT2, GT4, GT47 and GT51 (Figure S12B).

For significantly differently abundant CAZy classes between the biological soil crusts and vegetated soils, GHs were overrepresented in vegetated soils (LFC = 0.23, $P<0.05$), while CEs were overrepresented in the biological soil crusts (LFC = -0.18, $P<0.05$; Figure 3F; Table 6; Table S7). CAZy families involved in the degradation of starch/oligosaccharide (i.e., GH94, GH65 and GH88, but not CBM42, GH13_10, GH32, GH38, GH13_23, CBM51 and GH35), pectin (i.e., CBM32, CBM67 and CE8, but not PL4_2), hemicellulose (i.e., GH36, CE16 and CBM9), cellulose (i.e., GH12, GH8 and CBM56, but not GH5_13 and GH5_22), chitin (i.e., GH18, CE9 and CBM5) and lignin (AA1 but not AA6) were overrepresented in the biological soil crusts (Figure 5). Detailed results of the LFCs of CAZy classes/families and comparisons can be found in Table S7.

Shifts in N-cycling genes with SSD

Genes in the organic degradation and synthesis (OD&S) family were most abundant with *glnA* (17.0%), *gs_K00265* (11.5%), *gs_K00266* (10.0%) and *nmo* (9.2%). In the N-cycling gene family denitrification & DNR, the most abundant N-cycling genes were *norB* (2.6%), *narZ* (2.3%), *narG* (2.2%) and *napA* (2.0%; Figure S13B).

Comparisons at the gene family level between barren soils and biological soil crusts showed that anammox (LFC = 4.95, $P<0.05$) and nitrogen fixation (LFC = 2.15, $P<0.05$) were overrepresented in the biological soil crusts, while OD&S, nitrification and assimilatory N reduction (ANR) were overrepresented in the barren soils (Figure 3G; Table 6; Table S8).
By comparing barren and vegetated soils, DESeq2 analysis showed that the N-cycling gene family ANR (LFC = 0.91, $P<0.05$) was overrepresented in the vegetated soils, while nitrification (LFC= -1.37, $P<0.05$) was overrepresented in the barren soils (Figure 3H; Table 6; Table S8).

In the comparison of the N-cycling gene families between the biological soil crusts and vegetated soils, ANR (LFC = 1.96, $P<0.05$) and OD&S (LFC = 1.06, $P<0.05$) were overrepresented in the vegetated soils (Figure 3I; Table 6). Detailed results of the LFCs of N-cycling families/genes and comparisons can be found in Table S8.