AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

Year-Round Shotgun Metagenomes Reveal Stable Microbial Communities in Agricultural Soils and Novel Ammonia Oxidizers Responding to Fertilization

Luis H. Orellana,[a] Joanne C. Chee-Sanford,[b] Robert A. Sanford,[c] Frank E. Löffler,[d,e] Konstantinos T. Konstantinidis[a]

[a]Georgia Institute of Technology, Atlanta, Georgia, USA
[b]U.S. Department of Agriculture, Agricultural Research Service, Urbana, Illinois, USA
[c]University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[d]University of Tennessee, Knoxville, Tennessee, USA
[e]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

ABSTRACT The dynamics of individual microbial populations and their gene functions in agricultural soils, especially after major activities such as nitrogen (N) fertilization, remain elusive but are important for a better understanding of nutrient cycling. Here, we analyzed 20 short-read metagenomes collected at four time points during 1 year from two depths (0 to 5 and 20 to 30 cm) in two Midwestern agricultural sites representing contrasting soil textures (sandy versus silty loam) with similar cropping histories. Although the microbial community taxonomic and functional compositions differed between the two locations and depths, they were more stable within a depth/site throughout the year than communities in natural aquatic ecosystems. For example, among the 69 population genomes assembled from the metagenomes, 75% showed a less than 2-fold change in abundance between any two sampling points. Interestingly, six deep-branching *Thaumarchaeota* and three complete ammonia oxidizer (comammox) *Nitrospira* populations increased up to 5-fold in abundance upon the addition of N fertilizer. These results indicated that indigenous archaeal ammonia oxidizers may respond faster (are more copiotrophic) to N fertilization than previously thought. None of 29 recovered putative denitrifier genomes encoded the complete denitrification pathway, suggesting that denitrification is carried out by a collection of different populations. Altogether, our study identified novel microbial populations and genes responding to seasonal and human-induced perturbations in agricultural soils that should facilitate future monitoring efforts and N-related studies.

IMPORTANCE Even though the impact of agricultural management on the microbial community structure has been recognized, an understanding of the dynamics of individual microbial populations and what functions each population carries are limited. Yet, this information is important for a better understanding of nutrient cycling, with potentially important implications for preserving nitrogen in soils and sustainability. Here, we show that reconstructed metagenome-assembled genomes (MAGs) are relatively stable in their abundance and functional gene content year round, and seasonal nitrogen fertilization has selected for novel *Thaumarchaeota* and comammox *Nitrospira* nitrifiers that are potentially less oligotrophic than their marine counterparts previously studied.

KEYWORDS metagenomics, nitrifiers, soil microbiology, seasonal dynamics

Agricultural soils are characterized by a dynamic interplay between complex biotic and abiotic processes driving the nutrient cycling of the soil ecosystem (1, 2). Even though the central role of microorganisms participating in the cycling of nutrients in

Received 9 August 2017 **Accepted** 29 October 2017

Accepted manuscript posted online 3 November 2017

Citation Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. 2018. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl Environ Microbiol 84:e01646-17. https://doi.org/10.1128/AEM.01646-17.

Editor Marie A. Elliot, McMaster University

Address correspondence to Konstantinos T. Konstantinidis, kostas@ce.gatech.edu.

soil ecosystems has been extensively reported (3, 4), the dynamics of microbial communities responding to seasonal agricultural perturbations (e.g., nitrogen fertilization) have been mostly described using chemotaxonomic approaches or at higher taxonomical levels based on 16S rRNA gene amplification and sequencing (5–8) but not at the individual population level. These reports have revealed the effects of agricultural management on microbial community composition compared to that in uncultivated soils and identified the variable and core taxa in sites characterized by different land use (9). However, most of these approaches relied on 16S rRNA gene sequencing, which provides limited information at the species level and below (10, 11). Yet, the individual population (species) level is both traceable and a more relevant unit of diversity than 16S-defined operational taxonomic units (OTUs) based on metagenomics (12). The scarcity of information at the genome level limits our understanding of the abundance and functional gene content dynamics of individual populations generating and consuming key nutrients such as carbon (C) and nitrogen (N) in soils. The rates of synthetic N fertilizer addition often exceed crop requirements, resulting in unintended losses of N from the soil. For instance, higher atmospheric concentrations of nitrous oxide (N_2O), a potent greenhouse and ozone-depleting gas in the atmosphere (13), have been recorded compared to preindustrial-era levels, with soils contributing approximately 65% of the total N_2O emitted to the atmosphere (14). These elevated N_2O emissions result mainly from the activities of microorganisms controlling the N cycle (15, 16). Therefore, the identification and tracking of individual microbial populations responding to elevated N inputs can provide important new insights toward better modeling of N_2 and N_2O dynamics in soils, with implications for greenhouse gas emissions.

The advent of high-throughput sequencing technologies has expanded our understanding of natural microbial communities by unraveling the previously undetected microbial diversity in different ecosystems. Metagenomic approaches have been applied to examine highly diverse microbial ecosystems, providing a description of the taxonomic and genetic potentials of natural microbial communities that were previously elusive from cultivation-based approaches. For instance, short-read metagenomes obtained from grassland soils after a decade of induced warming revealed shifts in metabolic pathways related to major ecosystem processes, such as organic matter respiration (17). Further, the assembling and binning of contigs from metagenomes have enabled the recovery of genome sequences of abundant populations (metagenome-assembled genomes [MAGs]) from various soil samples. For example, the reconstruction of MAGs from an aquifer sediment provided a detailed picture of the individual populations participating in key biogeochemical pathways of the carbon, nitrogen, and sulfur cycles (18). However, little is known about the genomic diversity and abundance dynamics of individual populations undergoing recurrent external perturbations, such as from agricultural practices. Therefore, employing the new sequencing technologies to recover the genomes of uncultivated microbial species from agricultural soils holds great potential in providing new insights and data on the organisms controlling the fate of N compounds in agricultural soils.

Diverse microbial populations catalyze the transformation of soil N by performing complete or incomplete denitrification ($NO_3^-/NO_2^- \rightarrow N_2O$ and/or N_2), nitrification ($NH_4^+ \rightarrow NO_2^-/NO_3^-$), and ammonification (e.g., protein degradation and respiratory ammonification, NO_3^- or $NO_2^- \rightarrow NH_4^+$). Further, nitrification was originally described as a sequential two-step process mediated by ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB) (19). However, within the past decade, new genomic and cultivation advances have rapidly broadened our understanding of the microbial diversity participating in nitrification. For instance, studies of ammonia-oxidizing archaea (AOA) belonging to the *Thaumarchaeota* (20) phylum have revealed that there are differences in the ecophysiology and substrate affinity for these organisms compared to those of AOB (21) and that AOA are strongly enriched in specific soil horizons (22) and dominate at specific pH ranges compared to AOB (23). Although AOA numerically dominate AOB in many soils, the contribution of AOA to terrestrial nitrification and consequent N_2O production is not yet well established (24). Moreover, the

recent discovery of comammox (complete ammonia oxidizer) bacteria related to *Nitrospira*, harboring all necessary enzymes to perform complete nitrification in biofilms (25) and bioreactors (26), has initiated new interest into the relative contributions of these organisms to N cycling in nature. Because ammonia oxidation provides the main source of energy for these microbial groups, it has been proposed that a tight interplay among affinity, tolerance, and source of ammonia could control these microbial populations in nature (27, 28). Nevertheless, the relative responses of AOA, AOB, NOB, and the recently described comammox populations to agricultural fertilization are currently unclear but important for predicting their relative contributions to the N cycle and generation of by-products such as N_2O.

In the present work, we described the diversity and seasonal dynamics of metagenome-assembled genomes (MAGs) in two U.S. Midwest soil ecosystems with contrasting soil textures, i.e., sandy (93% sand, 7% clay) and silty loam (21% sand, 69% silt, 10% clay). We analyzed the total microbial community DNA (Illumina short-read metagenomes) from samples collected during the four seasons in 2012. Our findings showed that the recovered genomes, which represent abundant members of the natural microbial communities in these agricultural soils, are generally stable throughout the year. In fact, these microbial communities are not characterized by strong seasonal shifts in gene content diversity that typify other natural ecosystems such as freshwater lakes and the ocean. In addition, novel deep-branching ammonia-oxidizing *Nitrospirae* and *Thaumarchaeota* are among the most abundant nitrifiers in these agricultural soils, especially in the deeper soil layer, and are responsive to N fertilization. These findings highlight the power of shotgun metagenomic approaches in recovering and genotyping nitrifying populations that respond to the addition of fertilizers and can be used to advance soil nitrification models and expand the known nitrifier diversity.

RESULTS

Agricultural soil physicochemical characteristics and statistics of metagenomes. We focused our study on two sites with established legacies of the same agricultural management practices for at least 15 years. These sites represent distinct soil textures with opposing drainage characteristics and water-holding capacities located in Havana (sandy) and Urbana (silt loam), Illinois, USA. Soil cores obtained during 2012 revealed contrasting soil chemical parameters among the sites and depths (see Table S1 in the supplemental material). In particular, a higher organic matter (OM) content was observed in Urbana (average, 3.84%) compared to that in Havana (average, 0.7%) (two-tailed *t* test, $P < 0.01$). Whereas no statistically significant differences in OM were observed between soil layers in Urbana, a higher OM content was observed for the top soil layer in Havana across the year (two-tailed *t* test, $P < 0.01$). pH values were higher in Havana (average, 7.42) than in Urbana (average, 6.03) (two-tailed *t* test, $P < 0.01$), and substantial differences were observed between the soil layers in each site. The 0- to 5-cm soil layer in Havana was slightly more alkaline (average pH 7.57) than the 20- to 30-cm layer (average pH 7.32) (two-tailed *t* test, $P = 0.057$); a greater difference between top (average pH 5.7) and deep (average pH 6.26) soil layers was determined in Urbana (two-tailed *t* test, $P < 0.01$) (Table S1). Our analysis did not include the intermediate depth (i.e., 5 to 20 cm) because our previous studies of microbial community structure and diversity based on terminal restriction fragment length polymorphism (TRFLP) analysis of 16S rRNA and *nosZ* genes showed that this soil layer represented an intermediate between the top and the deeper layers (J. C. Chee-Sanford, A. K. Welsh, L. M. Connor, F. E. Löffler, and R. A. Sanford, unpublished). During the growing season, Havana was planted with maize and received fertilizer (UAN28: ammonium nitrate, 40%; urea, 30%; water, 30%) and herbicide (glyphosate), whereas Urbana was planted with soybean and received herbicide but not synthetic fertilizer (see Table S2).

A total of 20 shotgun metagenomes were obtained from both agricultural sites, ranging from 21 to 72 million reads per sample (or ~2.7 to 9.2 Gbp, 4.3 Gbp average)

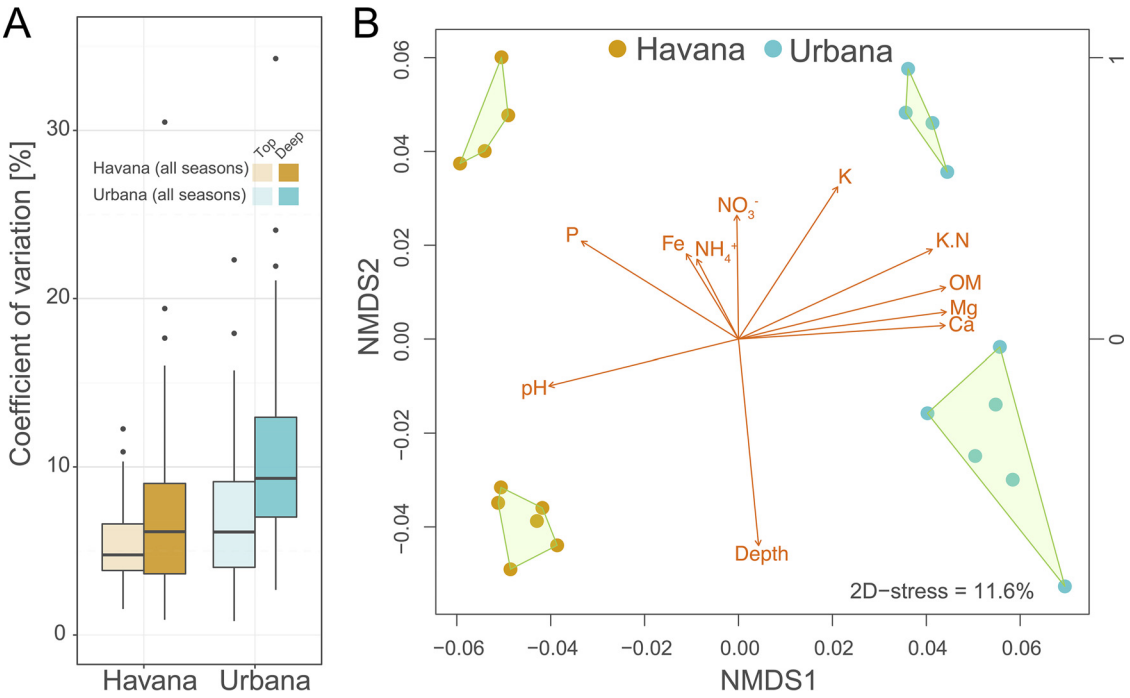

FIG 1 Sequence and functional compositional differences between two agricultural sites. (A) Distributions of coefficients of variation of SEED subsystems related to secondary metabolism in Havana and Urbana. (B) Nonmetric multidimensional scaling (NMDS) analysis based on MinHash distances determined by Mash showed independent clustering by site and depth. The lengths of the gold arrows are proportional to the correlations between measured metadata and determined ordination values. The directions of the arrows point to increasing changes in the values of the corresponding metadata.

and an average read length of ~125 bp after trimming (see Table S3). The estimated coverages based on the read redundancy value calculated by the Nonpareil algorithm (29) were averages of 0.15 and 0.14 for the metagenomes obtained from the 0- to 5-cm soil layers in Havana and Urbana, respectively. Higher coverage values of 0.21 and 0.26 were observed in the metagenomes obtained from the 20- to 30-cm soil layers in Havana and Urbana, respectively (Table S3). Sequence diversity values (a measure of alpha diversity derived from Nonpareil curves) were higher for the Havana 0- to 5-cm soil layer than for the 20- to 30-cm soil layer, but there were no differences between Urbana soil layers (see Fig. S1). Coassembly of the metagenomes by depth and location recovered over 1.1 million contigs of at least 500 bp in length each, comprising 1.59 Gbp, in total, for the four coassemblies and ~2.3 million predicted genes (see Table S4). The N_{50} values averaged from both 20- to 30-cm soil metagenomes were slightly higher than those for the 0- to 5-cm samples (~1,650 versus 1,240 bp), reflecting the lower sequence coverage determined by Nonpareil for the surface samples.

Microbial community structure and diversity. Throughout the year, stable abundances of functional genes were observed for both soil depths. For instance, the ~60 most abundant functional categories associated with secondary metabolism (details in supplemental material) showed, on average, 5.3% and 6.8% annual variations (measured as a coefficient of variation across all samples) for the top and deep layers of Havana, respectively. Similarly, 6.9% and 10.9% average variations for top and deep layers were observed for Urbana, respectively (Fig. 1). A much lower variation was observed when housekeeping genes and general metabolic functional categories were included in the analysis (see Fig. S2A), as expected for core functions harbored by almost every organism. To assess the extent of sample variation (i.e., within site), 7.2% and 14.6% variations in gene annotations related to secondary metabolism were observed among metagenomes obtained from three independent soil cores (replicates) from the 20- to 30-cm soil layer in June for Havana and Urbana, respectively (Fig. S2B). Smaller average variations were observed within Havana (2.01%) and Urbana (5.3%)

cores when all functional categories were considered. Thus, these consistent variation patterns between and within sites revealed a higher variation within Urbana soil metagenomes but largely stable variations values across the year for both sites.

Interestingly, a deep versus surface separation of samples was observed between metagenomes from each site based on annotation-independent MinHash similarity distances (nonmetric multidimensional scaling [NMDS], Mash similarity distances, and analysis of similarity [ANOSIM] $P \leq 0.001$, $R = 0.95$) (Fig. 1B). A similar spatial clustering was observed when functional gene annotations derived from short reads (SEED subsystems) were used for ordination (NMDS, Bray-Curtis distances, and ANOSIM $P \leq 0.001$, $R = 0.89$) (see Fig. S3A). We further investigated the functional features driving the separation between the soil layers in each site using the SEED subsystem information and best match analyses against publicly available genomes. Bacterial and archaeal genes encoding oxygen and light sensors, ferrous iron transporters, and photolyases, among others, were characteristic of the top layers in both soils ($\log_2$ fold change >2, adjusted $P < 0.05$) (see Fig. S4). In contrast, the strongest characteristic of the deeper layer corresponded to archaeal pathways accounting for 39.4% and 34.1% of the overrepresented SEED subsystems ($\log_2$ fold change >2, adjusted $P < 0.05$) in Havana and Urbana, respectively (Fig. S4). A significantly higher functional diversity (Chao-Shen entropy index using the number of gene functions recovered) was observed for Havana than for Urbana (two-tailed *t* test, $P < 0.05$) over the entire year of sampling. However, similar diversity values were observed among the top and deep soil samples from each site (Fig. S1C).

Similar to the functional annotations, the taxonomic affiliations of recovered 16S rRNA gene fragments from the soil metagenomes showed moderate changes in abundance across the year, but significant differences were detected between the soil depths (for further details, see Fig. S3B and supporting information in the supplemental material). We also explored the impacts of the microbial communities on the breakdown and recycling of plant biomass in soils. For instance, glycoside hydrolases (GH) are a group of enzymes that catalyze the hydrolysis of the glycoside bond and are key for the degradation of labile carbon compounds, e.g., starch and polysaccharides, as well as recalcitrant carbon compounds, e.g., lignocellulose, and other complex organic carbon compounds (30). Similar to the whole-community functions reported above, stable abundances for GH genes were found throughout the year (details in the supplemental material).

Effect of fertilization on nitrogen cycle gene abundances. We further examined the impact of agricultural practices on target N cycling genes throughout the growing season. We combined an accurate approach for detecting metagenomic reads for a gene of interest (ROCker) with a phylogenetics-based classification of detected short reads (see Table S5). For the nitrification process, we focused on the well-characterized genes that are responsible for the bacterial oxidation of ammonia via nitrite to nitrate and the archaeal oxidation of ammonia. The archaeal equivalent of the bacterial hydroxylamine dehydrogenase enzyme (HAO) is currently unknown (31); hence, only bacterial hydroxylamine oxidation was assessed in our study. The archaeal oxidation of ammonia was only assessed at the first step, catalyzed by the ammonia monooxygenase enzyme AmoA (Fig. 2A). Havana and Urbana showed 4.3-fold- and 3.2-fold-higher abundances on average, respectively, of archaeal *amoA*, bacterial hydroxylamine dehydrogenase (*haoA*), and nitrite oxidoreductase (*nxrA*) genes in the deep layers of soil relative to the top layers (Fig. 2A). Also, a urease gene (*ureC*) had the highest relative abundance (~0.3 genome equivalent, meaning that approximately 30% of the total cells sampled harbored the urease gene [see Materials and Methods for further details]) among all N genes detected in both sites but did not show substantial differences in abundance between soil layers (see Fig. S5). Abundances for other N cycle genes are available in the supporting information.

Havana received synthetic N fertilizer during late April 2012, which we hypothesized affected the abundance and distribution of N cycle genes associated with oxidation

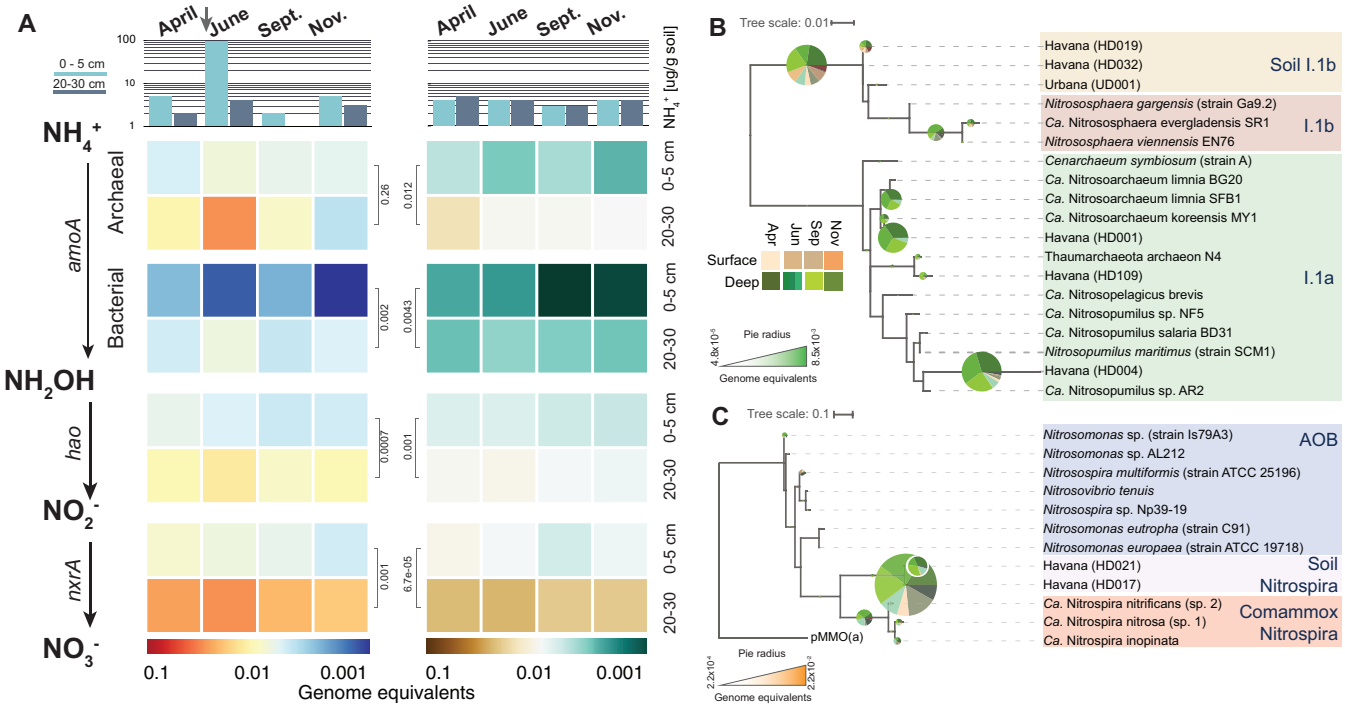

FIG 2 Abundance and diversity of nitrification genes in sandy (Havana) and silt loam (Urbana) soils. (A) The top shows the concentration of NH_4^+ at both sites and depths. The arrow shows the point in time when N fertilizer (UAN28, 180 lb N/acre) was applied to the Havana site. Heatmaps represent calculated relative abundances of nitrification genes (genome equivalents) for Havana (left) and Urbana (right) soil samples. Values for the 20- to 30-cm layer in June represent the averages of the three soil cores. Gene abundance comparisons between soil layers were performed using two-sided *t* tests, and *P* values are shown in black, to the side of the heatmaps. Right panels show the phylogenetic reconstruction of archaeal (B) and bacterial (C) AmoA protein sequences recovered from contigs. Names in parentheses indicate the corresponding MAGs. Both trees include reference protein sequences and assembled sequences from both soil metagenomes. The pie charts represent the placing of Havana metagenomic reads for archaeal and bacterial *amoA* genes using RAxML EPA. Pie chart radii represent the read abundances for each node (calculated as genome equivalents), and the colors of the slices represent the depths and months the metagenomic reads originated from.

processes (e.g., nitrification), such as *amoA*, *haoA*, and *nxrA* (Fig. 2A). Interestingly, ammonia oxidation genes showed the highest increase in abundance in June (20- to 30-cm soil layer), approximately 1 month after fertilization was performed in Havana (3.8- and 1.5-fold increases from April to July for archaea and bacteria, respectively). In addition, archaeal *amoA* gene fragments were ~6 times more abundant than those in bacteria, on average, during the year. On the other hand, Urbana, which relied on naturally fixed N, showed a slightly higher abundance of nitrification genes during April, but in general, these relative abundances were much more stable throughout the year compared to those in Havana.

In Havana, 60.5% of the archaeal *amoA* reads were placed within the I.1a clade (Fig. 2B). Notably, 55% of the bacterial *amoA* reads were placed within a clade closely related to the recently described comammox *Nitrospira* (Fig. 2C). Only 15% of the total bacterial *amoA* reads were placed within the *Betaproteobacteria* class, and these sequences were mostly derived from the top layer. As expected, similar results were observed for *haoA* and most *nxrA* gene fragments belonging to different NOB clades (see Fig. S6). On the other hand, in Urbana, the majority (87.3%) of the archaeal *amoA* gene fragments were placed within the I.1b clade, whereas less than 10% were placed in the I.1a clade (see Fig. S7). Compared to that in Havana, fewer bacterial *amoA* gene fragments were detected in Urbana (Fig. 2). A similar fraction (58.9%) of the *amoA* gene fragments were placed within the soil comammox clade, but a higher fraction (~39.7%) were placed inside *Betaproteobacteria* clades. Also, similar to those in Havana, the majority of the *haoA* and *nxrA* gene fragments were placed in soil comammox clades (Fig. S7).

Spatiotemporal abundance of MAGs. To precisely identify and quantify individual populations and their dynamics throughout the season, we performed genome binning

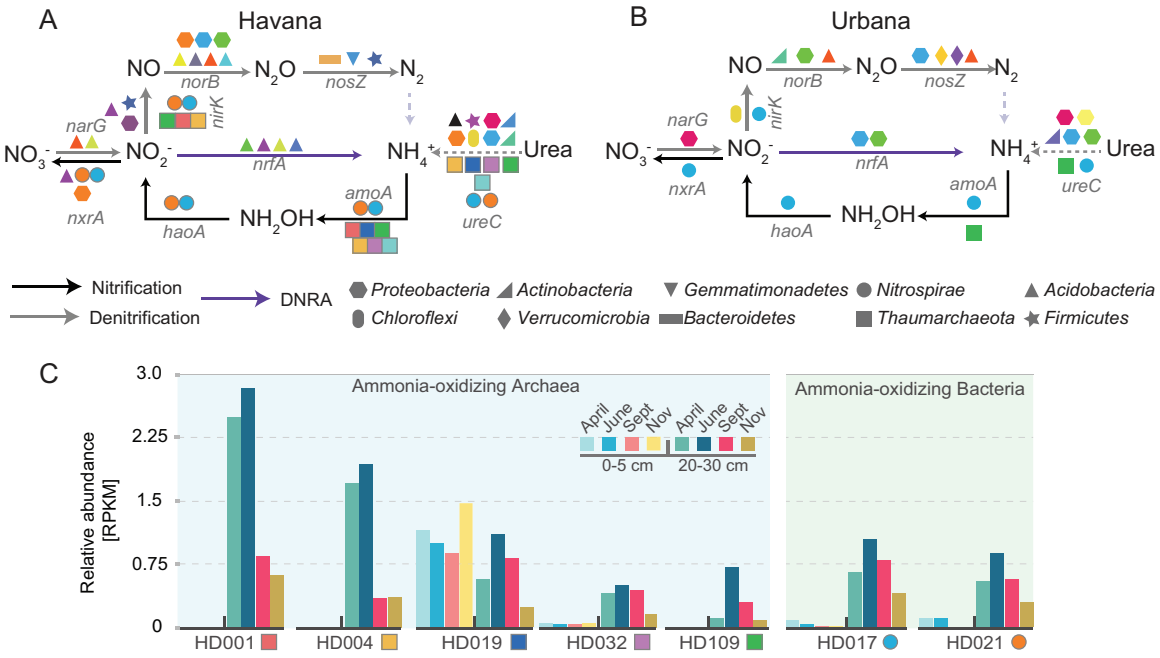

FIG 3 Nitrogen cycle genes present in selected metagenome-assembled genomes and population abundance dynamics during the year in Havana. (A and B) MAGs obtained from Havana (A) and Urbana (B) were queried for the presence of N cycle gene markers using HMM models. Colors represent individual MAGs and the shapes depict the predicted taxonomies at the phylum level. Arrows show predicted N cycle pathways. (C) Bar plots show the relative abundances (*y* axis, reads per kilobase per million reads [RPKM]) for nitrifier MAGs (*x* axis) obtained from Havana soil metagenomes. Bright colors represent samples from the 0- to 5-cm soil layer, whereas darker color equivalents correspond to samples from the 20- to 30-cm layer (see figure key).

analysis of the coassembled metagenomic data sets. The majority of the 69 recovered MAGs were obtained from the 20- to 30-cm layer (Havana, 45/47; Urbana, 18/22), presumably due to the lower sequence diversity and higher average coverage obtained (Fig. S1A). Additional details about recovered MAGs can be found in the supporting information. The abundance of each recovered genome bin, calculated as the fraction of total short reads mapping on all MAGs, was stable over time for most of MAGs from both field sites (see Fig. S8). In fact, only ~9% to 24% of the total MAGs ($n = 69$) showed 2-fold changes in abundance between any two sequential sampling points, and these MAGs were obtained mostly from the deep layer of Havana (HD MAGs) compared to that from Urbana (UD MAGs). For instance, from April to June, MAGs HD109 (nitrifier *Thaumarchaeota*), HD116 (*Bacteroidetes*), and HD098 (*Acidobacteria*) showed over 2-fold increases and only MAG UD001 showed a similar amount of decrease in abundance. Interestingly, *Thaumarchaeota* nitrifier MAGs HD109, HD001, and HD004, in addition to MAGs HD098, HD116, and HD051, showed more than 2-fold decreases from June to September. Lastly, from September to November, while only two MAGs (HD051 and HD103) showed slight increases over 2-fold, 14 MAGs showed more than 2-fold decreases, three of which were *Thaumarchaeota* MAGs (HD032, HD019, and HD109). Other comparisons between April and September (i.e., no crops versus mature crops) or between April and November (i.e., no crops) indicated that 22% and 26% of MAGs, respectively, experienced greater than 2-fold changes. In these two comparisons, *Thaumarchaeota* MAGs HD001 and HD004 showed more than 3-fold increases, indicating the prolonged effect of fertilization over these archaeal populations and/or the high persistence of these organisms during changing environmental conditions.

Diversity of MAGs involved in nitrogen cycling. The examination of key N cycling genes showed that 39 MAGs, representing different bacterial and archaeal lineages, encoded nitrification and denitrification enzymes (Fig. 3A and B). These MAGs, mostly recovered from Havana, represented almost 40% of the MAGs showing 2-fold increases

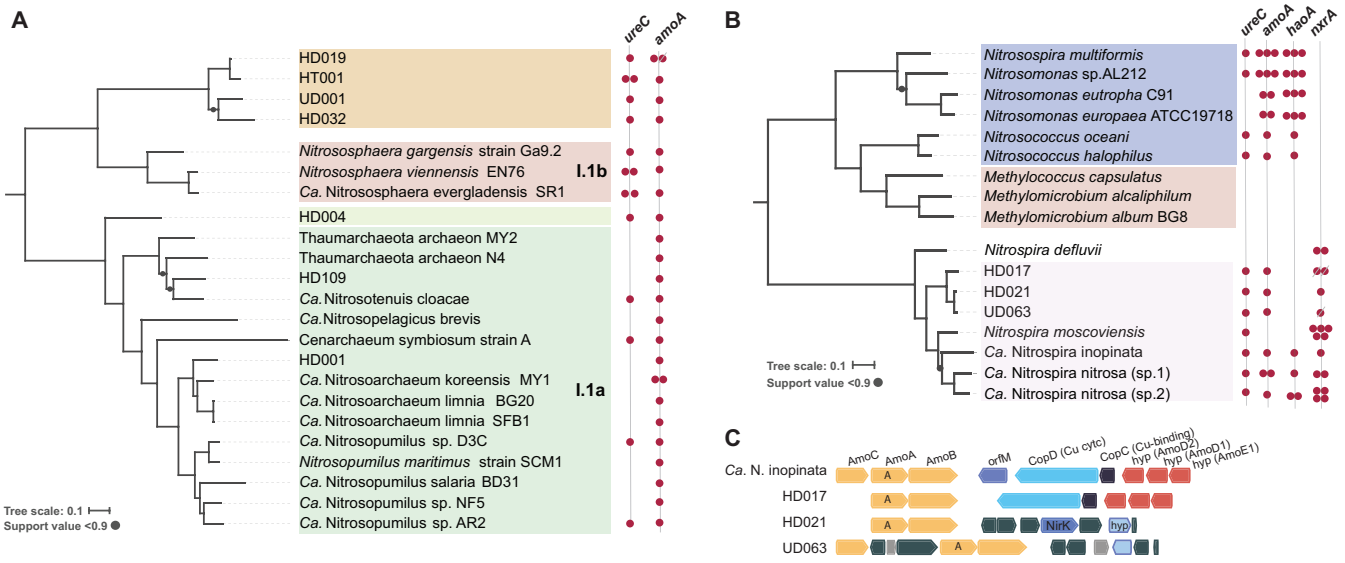

FIG 4 Recovery of indigenous archaeal and bacterial ammonia-oxidizing populations. (A and B) Phylogenetic reconstruction of archaeal (A) and bacterial (B) MAGs harboring *amoA* genes. Concatenated alignments of conserved genes for bacterial or archaeal genomes were used to build maximum-likelihood trees in RAxML. Colored circles on the right of each tree show the presence of selected nitrification genes. Strikethrough circles indicate incomplete sequences detected in MAGs. (C) Comparison of the genomic context for the *amoA* genes found in "*Ca.* Nitrospira inopinata" and nitrifier bacterial MAGs recovered from both sites. Colors denote different gene operons.

or decreases in abundance at any sampling point (Fig. S8). MAGs encoding nitrification enzymes mostly belonged to *Thaumarchaeota* and *Nitrospirae* phyla. All *Thaumarchaeota* MAGs had at least one copy of the *amoA* and *ureC* genes. Interestingly, all bacterial nitrifier MAGs also contained the *ureC* gene in addition to all the necessary genes associated with complete oxidation of ammonium to nitrate, i.e., *amoA*, *haoA*, and *nxrA*, similar to the recently described *Nitrospira* organisms capable of performing complete nitrification (25, 26). On the other hand, none of the MAGs carried all the genes to perform canonical or complete denitrification (i.e., reduction of NO_3^- or NO_2^- to N_2). Instead, most MAGs obtained from both sites encoded single steps of the denitrification pathway (i.e., Fig. 3A and B). The relatively high degree of completeness determined for the MAGs indicated that the denitrification genes were likely absent as opposed to not being assembled. Therefore, these results indicated that complete denitrifiers are not abundant in these soils. Notably, *Nitrospirae* and *Thaumarchaeota* MAGs, commonly associated with ammonia-oxidizing activity, showed increased abundance upon N fertilization in Havana (Fig. 3C), suggesting their participation in nitrification in the agricultural soils.

Whole-genome phylogenetic analyses using available genomes for isolates and *Candidatus* taxa, showed that these potential nitrifier MAGs represented novel taxa (Fig. 4). For instance, 10 MAGs harboring divergent archaeal and bacterial *amoA* genes were recovered from both sites. On the basis of average amino acid comparisons (AAI) (32), six of these ammonia-oxidizing MAGs most likely represent new genera within the *Thaumarchaeota* and *Nitrospirae* phyla (e.g., AAI lower than 57% compared to previously described genomes) (see Table S6). Consequently, phylogenetic reconstruction based on concatenated single-copy proteins revealed that archaeal MAGs HD032, HD019, and HT001 from Havana and UD001 from Urbana formed an independent sister clade of the *Thaumarchaeota* group I.1b, mostly consisting of *Thaumarchaeota* isolated from soil (Fig. 4A). The remaining archaeal MAGs were placed within the I.1a group, although MAG HD004 formed a deep branch within this group. Other ammonia-oxidizing archaeal MAGs from Havana, such as HD001, were placed close to "*Candidatus* Nitrosoarchaeum limnia" strains (79.7% AAI) whereas HD109 was clustered with "*Candidatus* Nitrosotenuis cloacae" (76.5% AAI). Bacterial MAGs from Havana HD017 and HD021 and Urbana UD063 formed an independent clade closely related to the recently

described comammox "*Candidatus* Nitrospira inopinata." In fact, AAI values between "*Ca*. Nitrospira inopinata" and MAGs HD017, HD021, and UD063 were 65.3%, 66.3%, and 65.8%, respectively, indicating their relatedness at the genus level. Even though the Havana MAG HD021 and the Urbana MAG UD063 were obtained from different agricultural sites, they shared 96.1% average nucleotide identity ([ANI] standard deviation [SD], 3.5%, from 5,838/12,667 1-kb-long fragments), revealing that they represent closely related populations at the level of (same) species (33). Independent phylogenetic trees using hallmark nitrification proteins AmoA, HAO, and NxrA (Fig. S6 and S7) showed a similar topology to the one observed when using concatenated alignment of multiple single-copy proteins (Fig. 4), indicating a limited recent horizontal gene transfer of the genes. Interestingly, the genetic contexts of the *amoCAB* operon differed in all MAGs (Fig. 4C). For instance, even though HD017 did not harbor a copy of the *amoC* gene in the same contig (probably missed during assembly), the synteny of *amoA*, *amoB*, the genes *copD* and *copC* encoding copper binding proteins, and *amoD2*, *amoD1*, and *amoE1* showed the same arrangement as that found in "*Ca*. Nitrospira inopinata." On the other hand, the *copD* and *copC* genes were absent in MAGs UD063 and HD021. Consistent with previous reports describing nitrification genes in AOB and AOA genomes (34), MAG HD021 possessed a copy of the nitrite reductase gene (*nirK*), a hallmark gene of denitrification, upstream of the *amoCAB* operon, which was surrounded by transposase and integration elements.

DISCUSSION

Temporal stability of natural microbial communities during the growing season. The soil metagenomes obtained at two different depths and four time points throughout the year from two agricultural fields, which received different management, provided new insights into the functional and community dynamics of indigenous microbial communities. Our results revealed remarkable composition stability for these microbial communities in their functional, taxonomic, and individual population components during the sampled period, especially compared to those with aquatic habitats (see below). Consistent with the findings reported here, previous reports focusing on single phylogenetic markers (e.g., 16S rRNA gene amplicons) have identified stable genetic compositions across soils having different land use or agricultural practices (9, 35). In addition, the recovered MAGs provided the almost complete functional genes harbored by the corresponding populations, as well as precise estimations of how novel the soil populations are compared to previously characterized organisms from other habitats and engineered wastewater systems. It is important to note that the metagenomic snapshots reported here might have missed short-term abundance dynamics or gene expression activity shifts. Nonetheless, the seasonal shifts in the soil microbial communities observed were much less profound than those from freshwater or ocean ecosystems, even when the differences in sampling procedures and sample heterogeneity were considered. For instance, freshwater metagenomes obtained during 1 year from Lake Lanier (Atlanta, GA) showed ~7.5-fold-higher coefficients of variation in gene functions or taxonomic compositions, on average, compared to those in the soil metagenomes (Fig. S2C in the supplemental material), highlighting that seasonality has a stronger effect on structuring the lake microbial communities (36, 37). In general, a larger fraction of the microbial community is presumably dormant in soils than in aquatic habitats (38), and low metabolic activity accounts, most certainly but not fully, for some of the year-round abundance patterns observed here (e.g., smaller abundance shifts compared to that in freshwater ecosystems). For instance, larger taxa and gene content differences were observed, in general, between spatial scales (e.g., different depths) than between temporal scales (e.g., four seasons) for both agricultural soils. These findings suggest that the biotic (e.g., plants) and edaphic soil physicochemical characteristics have stronger effects on structuring and modulating the variability of microbial soil communities than seasonal changes, in agreement with previous findings (39). Our findings also indicated that many soil organisms may withstand changing conditions by modulating specific gene expression rather than undergoing changes in their

abundance. In addition, soil organisms might be well adapted to the seasonal environmental fluctuations in soils (e.g., C or N inputs) compared to those from other environments. In fact, soil bacteria have larger genome sizes than those in other environments (40) and harbor a greater variety of metabolic pathways than their water or human-associated counterparts (41, 42), which is consistent with the latter interpretations.

Impact of N fertilizer on microbial soil communities. Even though most populations examined here showed steady abundances throughout the seasons, a fraction showed conspicuous responses to agricultural activities. These populations showed a 2-fold or higher change, which represents a substantial change in abundance for the slow growing conditions that prevail in bulk soil, e.g., 1 to 2 generations per year, on average (43). The largest changes in abundance observed at any time point sampled were clearly those in response to synthetic N fertilization at the Havana site. Previously unrecognized nitrifier populations belonging to the *Thaumarchaeota* phylum showed an increase and subsequent decrease in abundance upon the application of N fertilizers. These microbial populations might also contribute to N_2O emissions, as previously suggested (21, 31, 44). In Urbana, probably because this site did not receive synthetic fertilizer in the sampling year, many MAGs showed stable abundances, meaning that they did not exhibit changes in abundance greater than 2-fold throughout the year. In addition, the high relative abundance of MAGs encoding enzymes necessary for oxidation (i.e., nitrification) or reduction (i.e., denitrification) of N species underscored the effect of N fertilization on a selected fraction of the microbial communities involved in the N cycle. Interestingly, none of the recovered MAGs encoded all enzymes required for performing complete denitrification. Even though the 69 recovered MAGs are far from representing the extant soil microbial community diversity, they most likely better represent the most abundant organisms at the times sampled, whose genomes were recoverable by assembly and binning. Our previous work on organisms harboring clade II or atypical *nosZ* genes obtained from the same agricultural sites also proposed a modular assembly for denitrification pathways in these soils (45, 46), consistent with previous reports (47). Collectively, these findings suggest that a reduction of oxidized N species to N gas (N_2) would require the concerted participation of different N-reducing organisms and highlight the importance of accounting for the different organisms and their interactions to better understand denitrification processes and the N balance (e.g., N losses in the form of N_2O gas) in soil ecosystems.

Novel nitrifiers in agricultural soils. Even though previous reports detected a high abundance of common nitrification marker genes or 16S rRNA gene sequences assignable to known nitrification taxa in soils (27, 48), the genomic information of these taxa has been limited. Both agricultural sites examined here showed high abundances of *Nitrospirae* and *Thaumarchaeota* communities despite their different edaphic characteristics (but similar crop rotation management). The bacterial nitrifier genomes recovered from the two sites represented clades divergent from those of the well-characterized and canonical *Betaproteobacteria* and *Gammaproteobacteria* nitrifiers. In fact, these soil organisms were most closely related, yet distinct at the genus level based on their AAI relatedness, to the novel comammox *Nitrospira*, which was recently demonstrated to be capable of performing complete nitrification (25, 26). These comammox organisms were isolated from pipe and trickling filter biofilms, with environmental conditions and N inputs contrasting those of agricultural soils. Therefore, it appears that the agricultural management in the two soils has selected for discretely evolving nitrifier populations with ecophysiologies different from those of commonly known nitrifiers. In addition, the substantially different physicochemical characteristics of the two soils examined indicated that the novel *Nitrospirae* and *Thaumarchaeota* genomes recovered may represent ecologically successful populations, at least for the agricultural soils studied here, and could serve as models for future studies.

Our findings seem to support that recovered soil AOA populations might thrive in environments receiving large inputs of N and, perhaps more importantly, respond to N fertilization and do not necessarily follow the presumed oligotrophic lifestyle of AOA

compared to that of AOB. For instance, we observed that the abundance of archaeal MAGs changed upon nitrogen fertilization as much as those from their bacterial counterparts, which are thought to grow faster (r-strategists). Interestingly, recent findings have revealed that nonmarine AOA isolates showed an ammonia affinity similar to that of AOB but a low affinity compared to that of a comammox isolate, supporting the idea that the former isolates represent less-oligotrophic AOA (49). Future experiments examining activity rates and substrate affinities will elucidate if these patterns in ammonia utilization between different groups of nitrifiers also apply to indigenous soil comammox and AOA communities. Nonetheless, the present findings indicated that in soils receiving large yearly inputs of synthetic N fertilization, indigenous microbes, including AOA, have been adapted to these conditions and likely evolved to be less oligotrophic than their marine counterparts.

The five most abundant archaeal nitrifier genomes recovered in Havana also showed distinct depth distributions. For instance, MAG HD109 was the only AOA population showing elevated abundances in both the top and deeper soil layers, whereas the rest of the *Thaumarchaeota* MAGs showed high abundances only in deep soil layers. Apparently, additional niches and ecophysiologies that remain to be elucidated underlie the distribution patterns for the detected AOA. On the other hand, comammox nitrifier populations represented by MAGs from Havana (HD021) and Urbana (UD063) showed a high level of similarity (ANI $>$ 95%), revealing ecological successes in agricultural soils with contrasting characteristics. These nitrifier genes and genomes were likely missed by previous studies that employed probe designs based on available sequences of functional (e.g., *amoA*) or 16S rRNA gene sequences (50). Collectively, our results propose a role for novel AOA and comammox organisms in responding to high N inputs from fertilization. Furthermore, the gene and genome sequences reported here should facilitate primer design for PCR assays that can be used to precisely monitor the abundance dynamics of these novel populations in agricultural soils in the Midwest United States and further corroborate the results reported here.

Altogether, our study showed that stable microbial communities dwell in agricultural soils and identified key populations and genes responding to seasonal (e.g., fall biomass return) and human-induced (e.g., fertilization practices) perturbations. These findings also propose a much broader niche for the recently described comammox organisms and ammonia-oxidizing archaea controlling the fate of N in agricultural soils.

MATERIALS AND METHODS

Soil samples and DNA extraction and sequencing. Agricultural soil samples were collected in 2012 from field locations in Havana, IL (lat 40.296, long 89.944; elevation, 150 m), and Urbana, IL (lat 40.075, long 88.242; elevation, 222 m), both with long histories of conventionally managed corn and soybean rotations. The Havana field site is characterized as a sand (93% sand, 7% clay) with somewhat excessive drainage with no ponding duration or frequency. During the summer season, the field is irrigated with underlying groundwater from the Mahomet aquifer. The Urbana field site is situated on a slight slope profile, characterized as a silt loam (average content 21% sand, 69% silt, 10% clay). This site is classified as poorly drained, with brief and frequent ponding, and soil moisture is exclusively due to precipitation events. Once each season (April, June, late August/early September, and November), three soil cores (2.5-cm diameter by 30-cm length) were collected at three fixed locations 30 m apart (centroids) within each field plot (9 cores total per field, per sampling time), with each core then partitioned into two depths (0 to 5 cm and 20 to 30 cm). In sampling year 2012, corn was seeded in Havana and received UAN28 (28% N as urea-ammonium-nitrate) fertilizer (180 lb N/acre), whereas Urbana was planted with soybean and no fertilizer was applied (additional soil management events conducted in 2012 are described in Table S2 in the supplemental material). Soil physicochemical characteristics (organic matter, P, K, Mg, Ca, NO_3^--N, NH_4^+-N, total N, cation exchange capacity) were determined using a composite pool of soil combined for each depth range at each time of sampling (A&L Laboratories, Ft. Wayne, IN) (Table S1). Soil pH and gravimetric soil moisture content were measured using separate cross sections from each soil core. DNA was extracted from ~0.5 g of soil from each fraction using a modified phenol-chloroform and purification protocol, as previously described (46), and equal quantities of DNA from each sample on the basis of agarose gel quantification were pooled to create composite samples for each of the two soil depths (0 to 5 cm and 20 to 30 cm) for each soil type. The 20- to 30-cm soil samples taken from both sites in June were pooled according to centroids (Havana-E, Havana-M, and Havana-W; Urbana-N, Urbana-M, and Urbana-S), and these pools were independently sequenced. Sequencing of DNA samples was performed using the Illumina HiSeq 2000 platform and the NextEra DNA 150 by 150 library preparation protocol (Table S3), as described previously (46).

Short-read assembly and analyses. Metagenomic raw reads (fastq) for all samples were trimmed with SolexaQA (51) using a Phred score cutoff of 20 and minimum fragment length of 50 bp. The average coverage for each sequenced library was determined by Nonpareil (29) using default settings (Table S3). Protein-coding sequences were predicted from the short-read metagenomes using FragGeneScan (52), and functional annotation was performed by blastp v2.2.29+ searches (53) against UniProt (54) using default parameters. To decrease false-positive matches (stringent threshold), BLAST search outputs were filtered for best match and a minimum identity of ≥50% and a read alignment of ≥70%. Protein annotations were subsequently translated into SEED subsystems (55) for functional analyses. Calculated Mash distances (56) and annotation counts from SEED subsystems were visualized in ordination plots (NMDS) using the vegan (57) and ecodist (58) libraries in R v3.3.1. Differentially abundant SEED functional annotations or taxonomical levels were determined with the DESeq2 (59) package. The homogeneity of the variance (i.e., homoscedasticity) across groups was corroborated by using Levene's test implemented in the car package (60) and Bartlett's test available in R. Short-read metagenomes were coassembled as Havana top (four samples), Havana deep (six samples), Urbana top (four samples), and Urbana deep (six samples) using IDBA_UD v1.1.1 (61) (Table S4).

Identification of nitrogen cycle genes. To identify and quantify reads encoding specific proteins of interest, in-house databases were constructed and manually curated using sequences obtained from UniProt (54) for the archaeal and bacterial ammonia monooxygenase alpha subunit (AmoA), hydroxylamine dehydrogenase (HAO), nitrite oxidoreductase alpha subunit (NxrA), nitrate reductase (NarG), nitrite reductase (NirK and NirS), nitric oxide reductase beta subunit (NorB), nitrous oxide reductase (NosZ), nitrite reductase (NrfA), and DNA-directed RNA polymerase subunit beta (RpoB) (Table S5). Independent ROCker (62) models (length = 125 bp, as it was the average for all the metagenomes) were subsequently built based on these databases with the exception of NarG and NxrA, where the databases were combined as a single model (Table S5). Trimmed short reads from soil metagenomes were used as the query for blastx searches (E value, 0.01) against the latter protein databases, and outputs were filtered using the previously generated ROCker models. Target gene abundance was determined as genome equivalents by calculating the ratio between normalized target reads (counts divided by median protein length) and normalized RpoB reads (counts divided by median RpoB protein length). Protein databases and ROCker models are available at http://enve-omics.ce.gatech.edu. We also searched for assembled N cycle protein sequences in the four coassemblies and MAGs using precompiled hidden Markov models obtained from FunGene (63) and HMMer (64). Detected target N cycle proteins were manually curated by assessing the presence of a characteristic amino acid and phylogenetic congruency. Recovered AmoA sequences were also used for constructing phylogenetic trees. The construction of phylogenetic trees, short-read placement in the trees, and visualization were performed as previously reported (62). Additional details are available in supporting information in the supplemental material.

Recovery of MAGs and analyses. Assembled contigs larger than 1,000 bp were binned from each coassembly using an expectation-maximization algorithm as implemented in MaxBin v2.1.1 (65). MAGs with over 70% completion according to CheckM (66) were reassembled using the reads mapping on the contigs of the MAG (identity, ≥98%; fraction of aligned read, ≥70%) in Velvet v1.2.10 (67), which typically improved MAG quality, e.g., average contig length and contamination level. Taxonomic classification and novelty for the obtained MAGs were assessed in Microbial Genome Atlas (MiGA) (http://microbial-genomes.org/). Contigs with conflicting taxonomies (most likely representing binning misplacement) were manually inspected by comparing calculated coverages and inferred taxonomical classifications from MyTaxa (68) scan reports as implemented in MiGA and were subsequently manually discarded (Table S6). Reported MAG statistics (contamination and completeness) were determined using the "HMM.essential.rb" script from the enveomics collection and manually corrected for fragmented single-copy genes (e.g., a single-copy gene fragmented into two different contigs). MAG abundance was calculated as reads per kilobase per million reads (RPKM) using the matching reads to the binned contigs from blastn searches (minimum nucleotide identity, ≥98%; fraction of read aligned, ≥50% for a match) divided by the metagenomic sample sizes (in millions of reads) and the length of the MAG in kilobases. The phylogenetic reconstruction of MAGs was performed based on concatenated alignments of universal single-copy proteins identified for each MAG using the "HMM.essential.rb" script. Forty bacterial and nine archaeal proteins present in the corresponding MAGs were extracted, and multiple alignments for each protein were generated using ClustalΩ. Concatenated alignments without invariable sites were generated for archaeal and bacterial alignments using the script "Aln.cat.rb" from the enveomics collection. Phylogenetic reconstructions were determined using maximum-likelihood phylogeny employing PROTGAMMAAUTO modeling and 100 bootstraps in RAxML v8.0.19 (-f a, -m PROTGAMMAAUTO, –N 100) and visualized in iTol.

Accession number(s). Raw metagenomic soil datasets and MAGs are deposited in the European Nucleotide archive under study number PRJEB20068. MAGs are also available at http://enve-omics.ce.gatech.edu/data/.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.01646-17.

SUPPLEMENTAL FILE 1, PDF file, 1.1 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.5 MB.
SUPPLEMENTAL FILE 3, XLSX file, 0.5 MB.

ACKNOWLEDGMENTS

We thank Alissa Hooker and the anonymous reviewers for helpful suggestions regarding the manuscript.

This work was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, Genomic Science Program (award DE-SC0006662) and U.S. National Science Foundation (award 1356288).

The authors declare no conflict of interest.

REFERENCES

1. Dick RP. 1992. A review: long-term effects of agricultural systems on soil biochemical and microbial parameters. Agric Ecosyst Environ 40:25–36. https://doi.org/10.1016/0167-8809(92)90081-L.
2. Altieri MA. 1999. The ecological role of biodiversity in agroecosystems. Agric Ecosyst Environ 74:19–31. https://doi.org/10.1016/S0167-8809(99)00028-6.
3. Kennedy AC, Smith KL. 1995. Soil microbial diversity and the sustainability of agricultural soil. Plant soil 170:75–86. https://doi.org/10.1007/BF02183056.
4. Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95:6578–6583. https://doi.org/10.1073/pnas.95.12.6578.
5. McCaig AE, Glover LA, Prosser JI. 2001. Numerical analysis of grassland bacterial community structure under different land management regimens by using 16S ribosomal DNA sequence data and denaturing gradient gel electrophoresis banding patterns. Appl Environ Microbiol 67:4554–4559. https://doi.org/10.1128/AEM.67.10.4554-4559.2001.
6. Buckley DH, Schmidt TM. 2003. Diversity and dynamics of microbial communities in soils from agro-ecosystems. Environ Microbiol 5:441–452. https://doi.org/10.1046/j.1462-2920.2003.00404.x.
7. Bending GD, Turner MK, Rayns F, Marx MC, Wood M. 2004. Microbial and biochemical soil quality indicators and their potential for differentiating areas under contrasting agricultural management regimes. Soil Biol Biochem 36:1785–1792. https://doi.org/10.1016/j.soilbio.2004.04.035.
8. Garbeva P, van Veen JA, van Elsas JD. 2004. Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. Annu Rev Phytopathol 42:243–270. https://doi.org/10.1146/annurev.phyto.42.012604.135455.
9. Lauber CL, Ramirez KS, Aanderud Z, Lennon J, Fierer N. 2013. Temporal variability in soil microbial communities across land-use types. ISME J 7:1641–1650. https://doi.org/10.1038/ismej.2013.50.
10. Cole JR, Konstantinidis KT, Farris RJ, Tiedje JM. 2010. Microbial diversity and phylogeny: extending from rRNAs to genomes, p 1–20. *In* Liu WT, Jansson JK (ed), Environmental molecular microbiology. Horizon Scientific Press, Norwich, UK.
11. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635–645. https://doi.org/10.1038/nrmicro3330.
12. Caro-Quintero A, Konstantinidis KT. 2012. Bacterial species may exist, metagenomics reveal. Environ Microbiol 14:347–355. https://doi.org/10.1111/j.1462-2920.2011.02668.x.
13. Ravishankara AR, Daniel JS, Portmann RW. 2009. Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. Science 326:123–125. https://doi.org/10.1126/science.1176985.
14. Seitzinger SP, Kroeze C, Styles RV. 2000. Global distribution of N_2O emissions from aquatic systems: natural emissions and anthropogenic effects. Chemosphere Global Sci 2:267–279. https://doi.org/10.1016/S1465-9972(00)00015-5.
15. Firestone MK, Davidson EA. 1989. Microbiological basis of NO and N_2O production and consumption in soil, p 7–21. *In* Andreae MO, Schimel DS, Robertson GP (ed), Exchange of trace gases between terrestrial ecosystems and the atmosphere. Wiley and Sons, New York, NY.
16. Bremner JM. 1997. Sources of nitrous oxide in soils. Nutr Cycl in Agroecosys 49:7–16. https://doi.org/10.1023/A:1009798022569.
17. Luo C, Rodriguez-R LM, Johnston ER, Wu L, Cheng L, Xue K, Tu Q, Deng Y, He Z, Shi JZ, Yuan MM, Sherry RA, Li D, Luo Y, Schuur EAG, Chain P, Tiedje JM, Zhou J, Konstantinidis KT. 2014. Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. Appl Environ Microbiol 80:1777–1786. https://doi.org/10.1128/AEM.03712-13.
18. Hug LA, Thomas BC, Sharon I, Brown CT, Sharma R, Hettich RL, Wilkins MJ, Williams KH, Singh A, Banfield JF. 2016. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. Environ Microbiol 18:159–173. https://doi.org/10.1111/1462-2920.12930.
19. Prosser JI. 1989. Autotrophic nitrification in bacteria. Adv Microb Physiol 30:125–181.
20. Stahl DA, de la Torre JR. 2012. Physiology and diversity of ammonia-oxidizing archaea. Annu Rev Microbiol 66:83–101. https://doi.org/10.1146/annurev-micro-092611-150128.
21. Prosser JI, Nicol GW. 2012. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. Trends Microbiol 20:523–531. https://doi.org/10.1016/j.tim.2012.08.001.
22. Jung M-Y, Park S-J, Kim S-J, Kim J-G, Damsté JSS, Jeon CO, Rhee S-K. 2014. A mesophilic, autotrophic, ammonia-oxidizing archaeon of thaumarchaeal group I.1a cultivated from a deep oligotrophic soil horizon. Appl Environ Microbiol 80:3645–3655. https://doi.org/10.1128/AEM.03730-13.
23. Lu X, Bottomley PJ, Myrold DD. 2015. Contributions of ammonia-oxidizing archaea and bacteria to nitrification in Oregon forest soils. Soil Biol Biochem 85:54–62. https://doi.org/10.1016/j.soilbio.2015.02.034.
24. Jung M-Y, Well R, Min D, Giesemann A, Park S-J, Kim J-G, Kim S-J, Rhee S-K. 2014. Isotopic signatures of N_2O produced by ammonia-oxidizing archaea from soils. ISME J 8:1115–1125. https://doi.org/10.1038/ismej.2013.205.
25. Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, Jehmlich N, Palatinszky M, Vierheilig J, Bulaev A, Kirkegaard RH, von Bergen M, Rattei T, Bendinger B, Nielsen PH, Wagner M. 2015. Complete nitrification by *Nitrospira* bacteria. Nature 528:504–509.
26. van Kessel MAHJ, Speth DR, Albertsen M, Nielsen PH, Op den Camp HJM, Kartal B, Jetten MSM, Lücker S. 2015. Complete nitrification by a single microorganism. Nature 528:555–559. https://doi.org/10.1038/nature16459.
27. Verhamme DT, Prosser JI, Nicol GW. 2011. Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. ISME J 5:1067–1071. https://doi.org/10.1038/ismej.2010.191.
28. Levičnik-Höfferle S, Nicol GW, Ausec L, Mandić-Mulec I, Prosser JI. 2012. Stimulation of thaumarchaeal ammonia oxidation by ammonia derived from organic nitrogen but not added inorganic nitrogen. FEMS Microbiol Ecol 80:114–123. https://doi.org/10.1111/j.1574-6941.2011.01275.x.
29. Rodriguez-R LM, Konstantinidis KT. 2014. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics 30:629–635. https://doi.org/10.1093/bioinformatics/btt584.
30. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
31. Kozlowski JA, Stieglmeier M, Schleper C, Klotz MG, Stein LY. 2016. Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and *Thaumarchaeota*. ISME J 10:1836–1845. https://doi.org/10.1038/ismej.2016.2.
32. Konstantinidis KT, Tiedje JM. 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 10:504–509. https://doi.org/10.1016/j.mib.2007.08.006.
33. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0.
34. Treusch AH, Leininger S, Kletzin A, Schuster SC, Klenk H-P, Schleper C. 2005. Novel genes for nitrite reductase and Amo-related proteins

indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. Environ Microbiol 7:1985–1995. https://doi.org/10.1111/j.1462-2920.2005.00906.x.
35. Hartmann M, Frey B, Mayer J, Mäder P, Widmer F. 2015. Distinct soil microbial diversity under long-term organic and conventional farming. ISME J 9:1177–1194. https://doi.org/10.1038/ismej.2014.210.
36. Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD. 2007. Interannual dynamics and phenology of bacterial communities in a eutrophic lake. Limnol Oceanogr 52:487–494. https://doi.org/10.4319/lo.2007.52.2.0487.
37. Crump BC, Peterson BJ, Raymond PA, Amon RMW, Rinehart A, McClelland JW, Holmes RM. 2009. Circumpolar synchrony in big river bacterioplankton. Proc Natl Acad Sci U S A 106:21208–21212. https://doi.org/10.1073/pnas.0906149106.
38. Lennon JT, Jones SE. 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol 9:119–130. https://doi.org/10.1038/nrmicro2504.
39. Lauber CL, Hamady M, Knight R, Fierer N. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75:5111–5120. https://doi.org/10.1128/AEM.00335-09.
40. Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A 101:3160–3165. https://doi.org/10.1073/pnas.0308653100.
41. Bentley SD, Parkhill J. 2004. Comparative genomic structure of prokaryotes. Annu Rev Genet 38:771–792. https://doi.org/10.1146/annurev.genet.38.072902.094318.
42. Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124:837–848. https://doi.org/10.1016/j.cell.2006.02.017.
43. Gray T, Williams ST. 1971. Microbial productivity in soil, p 255–256. *In* Hughes D, Rose AH (ed), Microbes and biological productivity. 21st Symp. Soc. Gen Microbiol Cambridge. Cambridge University Press, London, UK.
44. Stieglmeier M, Mooshammer M, Kitzler B, Wanek W, Zechmeister-Boltenstern S, Richter A, Schleper C. 2014. Aerobic nitrous oxide production through N-nitrosating hybrid formation in ammonia-oxidizing archaea. ISME J 8:1135–1146. https://doi.org/10.1038/ismej.2013.220.
45. Sanford RA, Wagner DD, Wu Q, Chee-Sanford JC, Thomas SH, Cruz-García C, Rodríguez G, Massol-Deyá A, Krishnani KK, Ritalahti KM, Nissen S, Konstantinidis KT, Löffler FE. 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proc Natl Acad Sci U S A 109:19709–19714. https://doi.org/10.1073/pnas.1211238109.
46. Orellana LH, Rodriguez-R LM, Higgins S, Chee-Sanford JC, Sanford RA, Ritalahti KM, Löffler FE, Konstantinidis KT. 2014. Detecting nitrous oxide reductase (NosZ) genes in soil metagenomes: method development and implications for the nitrogen cycle. mBio 5:e01193-14. https://doi.org/10.1128/mBio.01193-14.
47. Graf DRH, Jones CM, Hallin S. 2014. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. PLoS One 9:e114118. https://doi.org/10.1371/journal.pone.0114118.
48. Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, Prosser JI, Schuster SC, Schleper C. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806–809. https://doi.org/10.1038/nature04983.
49. Kits KD, Sedlacek CJ, Lebedeva EV, Han P, Bulaev A, Pjevac P, Daebeler A, Romano S, Albertsen M, Stein LY, Daims H, Wagner M. 2017. Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. Nature 549:269–272. https://doi.org/10.1038/nature23679.
50. Santoro AE. 2016. The do-it-all nitrifier. Science 351:342–343. https://doi.org/10.1126/science.aad9839.
51. Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485. https://doi.org/10.1186/1471-2105-11-485.
52. Rho M, Tang H, Ye Y. 2010. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res 38:e191. https://doi.org/10.1093/nar/gkq747.
53. 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.
54. UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res 43:D204–D212. https://doi.org/10.1093/nar/gku989.
55. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1,000 genomes. Nucleic Acids Res 33:5691–5702. https://doi.org/10.1093/nar/gki866.
56. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132. https://doi.org/10.1186/s13059-016-0997-x.
57. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens HH. 2007. vegan: community ecology package. Community Ecol Pack 10:631–637. http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/packages/vegan.pdf.
58. Goslee SC, Urban DL. 2007. The ecodist package for dissimilarity-based analysis of ecological data. J Stat Softw 22:1–19. https://doi.org/10.18637/jss.v022.i07.
59. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.
60. Fox J, Weisberg S. 2011. An R companion to applied regression, 2nd ed. SAGE Publications, Los Angeles, CA.
61. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.
62. Orellana LH, Rodriguez-R LM, Konstantinidis KT. 2017. ROCker: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores. Nucleic Acids Res 45:e14.
63. 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.
64. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
65. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2:26. https://doi.org/10.1186/2049-2618-2-26.
66. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
67. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
68. Luo C, Rodriguez-R LM, Konstantinidis KT. 2014. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res 42:e73. https://doi.org/10.1093/nar/gku169.