Comparing DNA, RNA and protein levels for measuring microbial activity in nitrogen-amended soils

Luis H. Orellana,[a] Janet K. Hatt,[a] Ramsunder Iyer,[b,c] Karuna Chourey,[b] Robert L. Hettich,[b] Jim C. Spain,[d] Wendy H. Yang,[e] Joanne C. Chee-Sanford,[f] Robert A. Sanford,[g] Frank E. Löffler,[h,i] Konstantinos T. Konstantinidis[a,#]

[a] School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[b] Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

[c] Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, Tennessee, USA

[d] Center for Environmental Diagnostics & Bioremediation, University of West Florida, Pensacola, Florida, USA

[e] Department of Geology, University of Illinois at Urbana-Champaign, Urbana, Illinois USA

[f] U.S. Department of Agriculture, Agricultural Research Service, Urbana, Illinois, USA

[g] University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[h] University of Tennessee, Knoxville, Tennessee, USA

[i] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Running head: Multi-omics of soil microcosms

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

ABSTRACT

Multi-omic techniques can offer a comprehensive overview of microbial communities at the gene, transcript and protein levels. However, to what extent these levels reflect *in situ* process rates is less clear, especially in highly complex habitats such as soils. Here we performed microcosm incubations using soil from a site with a history of agricultural management. Microcosms, amended with isotopically labelled ammonium and urea to simulate a fertilization event, showed nitrification (up to 4.1 ± 0.87 µg $N\text{-}NO_3^-$ g^{-1} dry soil d^{-1}) and accumulation of N_2O after 192 hours of incubation. Nitrification activity ($NH_4^+ \rightarrow NH_2OH \rightarrow NO_2^- \rightarrow NO_3^-$) was accompanied by a 6-fold increase in relative expression of the 16S rRNA gene (RNA/DNA) between 10 and 192 hours of incubation for ammonia-oxidizing bacteria (AOB) *Nitrosomonas* and *Nitrosospira*. In contrast, ammonia-oxidizing archaea (AOA) and complete ammonia oxidizer (comammox) nitrifiers showed stable gene expression during incubations but were generally more abundant (DNA level) than their *Betaproteobacteria* AOB counterparts. A strong relationship between nitrification activity and (mostly) betaproteobacterial ammonia monooxygenase (*amoA*; $NH_4^+ \rightarrow NH_2OH$) and nitrite oxidoreductase (*nxrA*; $NO_2^- \rightarrow NO_3^-$) transcript abundances revealed that mRNA levels quantitatively reflected measured activity and were generally more sensitive than the DNA level in the microcosm incubations. Although peptides related to housekeeping proteins from nitrite-oxidizing microorganisms were detected, their abundance was not significantly correlated with activity, revealing that meta-proteomics provided only a qualitative assessment of activity. Altogether, these findings underscore the strengths and limitations of multi-omic approaches for assessing complex microbial communities and provide the molecular means to assess nitrification processes in soils.

IMPORTANCE

Even though the use of omic approaches has expanded our knowledge of the diversity of microbial communities in natural and engineered systems, it is less clear how well the use of whole community DNA-, RNA- or protein-based approaches reflect microbial activities. To this end, we directly compared the different levels of molecular information (i.e., DNA, RNA or proteins) in order to assess which level best correlated with isotope-based measurements of nitrification activity in agricultural soils after fertilization. This work reveals the strengths as well as the associated limitations of metagenomic, metatranscriptomic, and metaproteomic approaches in serving as reliable proxies for examining microbial activities in highly diverse environments like soils.

INTRODUCTION

Even though the central role of microbes in the cycling of nitrogen is recognized, the dynamics and controls of the interrelated microbial nitrogen pathways in agricultural soils are still poorly understood. This scarcity of information limits the development of more accurate, predictive models of nitrogen flux that encompasses the role of microbes in the generation and consumption of nitrogen substrates, as well as the emission of greenhouse gases, including nitrous oxide (N_2O) (1). In agricultural soils receiving large inputs of nitrogen fertilizer, ammonia-oxidizing bacteria (AOB), ammonia-oxidizing archaea (AOA) and nitrite-oxidizing bacteria (NOB) collectively are responsible for the conversion of ammonium to nitrate. In addition, the recent discovery of *Nitrospira* bacteria capable of complete oxidation of ammonia to nitrate (comammox) has revealed that the process of nitrification in natural environments might be carried out by a single taxon (2, 3). It has also been reported that nitrification is a major N_2O source under low oxygen concentrations (4), although detailed mechanistic understanding is lacking (5). Alternatively, under anoxic conditions, nitrate (NO_3^-) can be reduced to gaseous forms such as dinitrogen (N_2), nitric oxide (NO) or N_2O by denitrifying organisms and consequently be lost to the atmosphere. Despite the apparent importance of nitrification in the generation of N_2O and NO_3^-, the relative contributions of comammox, AOA, AOB and NOB populations in this process, especially during soil fertilization events, is less clear (6). Advancing this issue is essential for better prediction of the contributions of these microbial taxa to the nitrogen cycle and the modeling of the corresponding activities and products. High-throughput sequencing and proteomic approaches offer the means to characterize the nitrogen pathways in the environment. However, to what extent these omic approaches reflect process rates is still unclear.

Although DNA, RNA, and protein abundances all reflect microbial potential and responses to environmental changes and thus, can be used to study nitrogen cycling in soils, each measurement generally offers different types of information. For instance, metagenomics

(DNA level) offers a comprehensive overview of the functional potential of microbial communities but does not generally reflect active community members or functions. Short-term microbial responses to external changes (e.g., nitrogen addition) can be tracked by analyzing the actively expressed genes (i.e., metatranscriptomics). For instance, the relationship between measured nitrification processes and the ammonia monooxygenase (*amoA*) transcripts have revealed differences between archaeal and bacterial activity in acidic soils (7). Proteomics provides a third level of molecular information much closer to the metabolic processes by reflecting synthesized enzymes that catalyze reactions. Although proteomics has been applied to only a limited number of natural microbial communities, the results have provided new insights about metabolic pathways and interdependencies among microbial groups [reviewed in (8)]. Furthermore, recent advances in metagenomics and metaproteomics techniques as well as integration with isotope-based technologies (e.g., NanoSIMS) have disentangled the role of previously elusive keystone microbial populations. The combined application of metagenomics and metaproteomics has provided new understanding of novel not yet cultured microorganisms participating in the cycling of sulfur, nitrogen, and carbon in the terrestrial subsurface (9).

Only a few studies have examined how the above approaches correlate with process rates, especially in soil ecosystems that are characterized by low metabolic activity along with high microbial diversity and heterogeneity. Thus far, almost all studies have provided only qualitative results from applications of omics to soils (10). Quantitative results in a few recent reports have focused mostly on systems with reduced diversity or specific functions and taxa (as opposed to community-wide activities). For instance, metatranscriptomic approaches examining the degradation of the herbicide atrazine by *Escherichia coli* in bioreactors revealed a linear relationship between the measured enzymatic activity and the transcripts encoding the associated enzyme (11). Additionally, in microbial leaf litter decomposition incubations, cellulase and xylanase protein abundances were positively correlated with their corresponding enzymatic activities (12). On the other hand, even though the combination of multi-omic datasets provided

new insights into diversity and gene potential of microbial communities of permafrost ecosystems, the datasets were less predictive of measured process rates (13). Therefore, to what extent the omic measurements correlate with each other and with process rates in soils remain unclear.

Toward closing this knowledge gap, we examined nitrogen-amended sandy soils obtained from a site with a history of agricultural management and application of synthetic nitrogen fertilizer. A prior year-round characterization of field samples from the same agricultural site revealed increased abundance of novel *Thaumarchaeota* and comammox nitrifiers, but the findings were limited to metagenomics (14). Here, our goal was to assess the strengths and limitations of multi-omics in detecting microbial activity by correlating measurements of DNA, RNA, and protein abundances with measured rates of nitrate formation and N_2O production in soils incubated under controlled conditions in the laboratory. The results reveal that metatranscriptomic data best reflected the measured nitrification rates under the tested experimental conditions.

RESULTS

Nitrification activity in soil microcosms

We first examined nitrification activity in nitrogen-amended microcosms with an equimolar mixture of NH_4^+ and urea during an eight-day period by following NO_3^- formation and NH_4^+ disappearance. Based on the NH_4^+ concentration patterns, urea quickly hydrolyzed to release NH_4^+ within the first two days of incubation (Figure 1a). Specifically, the NH_4^+ concentrations peaked at 48 hours of incubation (18.02 ± 1.5 μgN-NH_4^+ g^{-1} dry soil) from urea hydrolysis, and decreased to 5.4 ± 2.5 μgN-NH_4^+ g^{-1} dry soil by 192 hours of incubation due to nitrification. Nitrification activity increased five to eight days after the addition of the NH_4^+ and urea mixture, reaching an average rate of 4.1 ± 0.87 μg N-NO_3^- g^{-1} dry soil d^{-1} (n = 6) after 192 hours of the incubation (Figure 1b). The NO_3^- concentrations gradually increased from an initial

value of 0.81 ±0.28 μgN-NO_3 g^{-1} dry soil to 1.91 ± 0.5 at 120 hours of incubation, and then increased at a faster rate to 15.06 ± 2.7 μgN-NO_3^- g^{-1} dry soil at 192 hours of incubation (Figure 1a). As a result of nitrification activity, pH values decreased across replicated nitrogen-amended microcosms during the incubation (Supplementary Table 1). In order to examine the generation of N_2O possibly generated as a by-product of oxidation reactions during nitrification, we measured the production of N_2O in nitrogen-amended incubations. Net N_2O production rates in the incubation headspace increased from 0.08 ± 0.006 ng N-N_2O g^{-1} dry soil d^{-1} after 24 hours to 0.71 ± 0.57 ng N-N_2O g^{-1} dry soil h^{-1} at the end of the incubations (Figure 1c). Control microcosms receiving only irrigation water (i.e., no nitrogen amendment) did not show net NH_4^+oxidation.

To evaluate possible differences between the use of NH_4^+ or urea in the nitrifying activity, we determined $^{15}NO_3^-$ production rates using nitrogen stable isotopes in the microcosms. In general, $^{15}NO_3^-$ production was similar between $^{15}NH_4^+$ and ^{15}N-urea microcosms, although rates were higher after 10 and 48 hours of incubation (two tailed *t*-test, $P<0.01$) in $^{15}NH_4^+$ and ^{15}N-urea microcosms, respectively, but converged thereafter (Supplementary Figure 1). By the end of the incubations, approximately half of the added ^{15}N was converted to ^{15}N-NO_3^- (49-55% for both labelled solutions/treatments), and only a small fraction converted to ^{15}N-N_2O (0.006-0.01%). The remaining added nitrogen was presumably converted to N_2, assimilated into microbial biomass, or adsorbed to soil particles.

Soil metagenomes and metatranscriptomes

To explore the genetic potential of microbial communities in control and nitrogen-amended microcosms, we examined the metagenomes and metatranscriptomes obtained from the incubated soils. Metagenomes ranged from 23.7 to 53.4 and metatranscriptomes from 10.1 to 31.3 million short-reads per sample (Supplementary Tables 2 and 3). The estimated average coverage based on read redundancy using Nonpareil (15) ranged from 0.27 to 0.42 for the soil

metagenomes (values range from 0 to 1). The co-assembly of selected soil metagenomes generated 1.52 million contigs over 500 bp (assembly N50=1,176) and 1.56 million predicted protein-coding genes.

A high fraction of ribosomal RNA was detected for all metatranscriptomes ranging from 94% to 98% of the total sequences (Supplementary Table 4). No rRNA depletion step was performed during our metatranscriptomic protocol due to overall low total RNA yields from the soils. As expected based on the length of the rRNA genes, 23S rRNA/16S rRNA ratios ranged from 1.7 to 1.9, indicating adequate RNA quality. Bacterial 16S rRNA (16S) was the most abundant, ranging between 30.6% and 35.9% of total transcripts per sample. Archaeal 16S and eukaryotic 18S rRNA molecules were less abundant, with values ranging from 0.09% to 0.15% and 0.55 to 2.9%, respectively.

Taxonomy of microbial soil populations based on 16S rRNA gene sequences

The taxonomic composition and abundances of the main microbial groups determined from recovered 16S rRNA (16S) gene sequences (DNA level) from nitrogen-amended incubations, were generally stable during incubations. At the class taxonomic level, *Actinobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* were the most abundant groups in metagenomes, accounting for more than 57% of the total community in nitrogen-amended incubations (Supplementary Figure 2). The taxonomic composition derived from metatranscriptomes (cDNA reads) was also stable during the incubations but the abundances for main taxonomical groups were substantially different from the metagenomes. For instance, *Betaproteobacteria*, *Gammaproteobacteria*, and *Flavobacteria* were among the most abundant groups in cDNA samples, accounting for an average of 77.5% of the 16S transcripts. In agreement with our previous results based on field samples from the same agricultural site (14), bacterial and archaeal groups associated with the nitrification processes were comparatively less abundant than the aforementioned groups in both DNA and cDNA datasets. For instance,

known AOB and NOB genera such as *Nitrosomonas* and *Nitrospira* had average relative abundances of 0.01% and 1.6% of the total populations in the metagenomes from incubated soils. Additionally, the relative abundances of the AOA genera related to *Nitrososphaera* and *Nitrosopumilus* were 0.9% and 0.3% in the microcosm metagenomes. Similar low abundances were determined for known nitrifier genera in metatranscriptomes. Notably, 16S transcript abundances for AOB conspicuously increased during the incubation period (Supplementary Figure 3). In fact, relative 16S gene expression ratios (cDNA/DNA) for AOB and NOB belonging to *Nitrosospira*, *Nitrosomonas* and *Nitrospira* increased 3-,6-, and 14-fold between 10 and 192 hours. In contrast, the 16S gene expression levels for the archaeal groups *Nitrososphaera* and *Nitrosopumilus* were stable during the same incubation period, although with a slight increase in relative expression at 48h of incubation (Supplementary Figure 3).

Individual populations from microcosm metagenomes

The assembly and binning of the soil metagenomes recovered 11 metagenome-assembled genomes (MAGs) mostly representing *Proteobacteria*, *Acidobacteria*, *Actinobacteria* and *Nitrospirae* phyla. Most of the recovered MAGs represented novel genera (n=7) and species (n=5) when the taxonomic novelty was evaluated against 10,487 reference genomes (taxonomically classified at the species level) using genome-aggregate amino acid identity (AAI) thresholds for taxonomic rank delineation (16) (Supplementary Table 5). Given that none of the MAGs represented AOA, AOB, NOB, or comammox populations, we included MAGs obtained from a previous analysis of field samples from the same site (Havana county, Illinois, USA) and depth as the soil used in the soil microcosms in the present study (14). MAGs potentially involved in nitrification processes were likely missed in the microcosm metagenomes due to comparatively lower sequencing effort or because of sample heterogeneity (e.g., lower population abundance) but had relatively higher abundance in the previous field samples. The MAGs (designated with the letter F at the end of their name for Field metagenomes) consisted

of two complete ammonia oxidizer (comammox) *Nitrospira* MAGs (MAG021F and MAG017F) and five ammonia-oxidizing archaea MAGs representing the *Thaumarchaeota* lineages I.1b (MAG032F and MAG019F) and I.1a (MAG004F, MAG109F, and MAG001F) (Supplementary Figure 4b). The *Nitrospira* MAG007, obtained from the microcosm metagenomes, was closely related to previously described soil comammox (e.g., MAG017F) organisms, sharing 67.8% AAI (SD: 18% based on 2201 shared proteins). However, the MAG007 only encoded a hydroxylamine oxidoreductase (*haoA*) gene and lacked *amoA* and *nxrA* genes (Supplementary Figure 4b). Furthermore, the *Nitrospira* MAG007 formed an independent but related cluster to the soil comammox organisms when reconstructed phylogenies using concatenated single-copy genes were evaluated (Supplementary Figure 5). Thus, AAI values and phylogenetic reconstruction supported the affiliation of MAG007 to *Nitrospira*, but the lack of genes involved in ammonia and nitrite oxidation (possibly due to low sequencing coverage) made it inconclusive whether this taxon is involved in nitrification processes and might indicate divergence from previously described soil comammox organisms.

Relative expression values of MAGs (measured as transcripts or reads per kilobase million, RPKM) were used as a proxy for comparing the response and metabolic activity among nitrifying bacteria and archaea during incubations. Even though expression values for most nitrifying MAGs belonging to *Nitrospira* and *Thaumarchaeota* were stable and relatively low, AOA MAGs 004F, 019F and comammox MAG017F, had, on average, the highest expression values throughout the incubations (Supplementary Figure 4a). For instance, the increase in expression values for AOA MAGs belonging to the I.1b clade, 004F and 032F, were 39% and 50% after 48 hours of incubation (compared to expression levels at 10 hours incubation), respectively. In contrast, gene expression of comammox MAG017F and *Nitrospira* MAG007 increased by 59% and 68% after 120 and 192 hours of incubation, respectively (Supplementary Figure 4a). Note that AOB and NOB were not included in the RPKM analysis due to lack of recovered MAGs representing these populations (see above). Nonetheless, a gene-based

245 approach allowed the analysis of changes in transcript abundances of genes involved in
246 nitrification activity for AOB and NOB nitrifiers (see below).
247
248 **Quantification of nitrification genes in microcosms**
249 To further explore the microbial nitrification processes in incubated soils at the gene
250 level, we specifically quantified gene fragments and transcripts directly involved in nitrification
251 reactions. Relative expression values belonging to the gene encoding urease subunit c (*ureC*)
252 were stable throughout the incubation but average abundances were relatively low compared to
253 other nitrification genes (Figure 2a). The relative expression of the bacterial gene encoding
254 ammonia monooxygenase subunit alpha (*amoA*) was 53.5-fold higher compared to the
255 expression values at 10h of incubation. Most of the detected *amoA* transcripts (cDNA) were
256 phylogenetically affiliated with *Betaproteobacteria* and corresponded to up to 90% of the total
257 detected bacterial *amoA* transcripts at 192 hours of incubation (Figure 2a). Unlike
258 betaproteobacterial *amoA*, abundance of transcripts belonging to comammox were stable
259 throughout the incubation. Furthermore, transcripts belonging to comammox were more
260 abundant compared to *Betaproteobacteria amoA* transcripts after 48 hours of incubation;
261 however, at 192 hours of incubation, betaproteobacterial *amoA* DNA abundance increased 66-
262 fold, whereas comammox *amoA* gene fragments remained stable (Figure 2b). The latter results
263 indicated that the comammox *amoA* may be more abundant under field conditions but
264 betaproteobacterial *amoA* might show a faster response upon ammonia addition, which was
265 also consistent with a previous study (14). Although the relative expression for the archaeal
266 *amoA* was more stable throughout the incubation compared to its betaproteobacterial
267 counterparts, a maximum expression was reached after 120 hours of incubation, suggesting
268 that archaeal AmoA activity temporarily increased at later time points during the incubation.
269 Archaeal *amoA* transcripts belonging to the group I.1b were ~7 times more abundant than their
270 I.1a counterpart across the incubations (Figure 2b). Similar to *amoA* patterns, the relative

271 expression for the betaproteobacterial hydroxylamine oxidoreductase (*haoA;* $NH_2OH \rightarrow NO_2^-$)
272 steadily increased during the incubations, whereas comammox *haoA* transcripts represented
273 the remaining smaller transcript fraction and were stable throughout the incubations (Figure 2b).
274 Expression values for the nitrite oxidoreductase subunit alpha (*nxrA;* $NO_2^- \rightarrow NO_3^-$) had a 12.4-
275 fold increase compared to the 10-hour time point, consistent with the patterns observed for the
276 previous nitrification genes and NO_3^- accumulation. Unexpectedly, expression values for *nirK*
277 ($NO_2^- \rightarrow NO$) affiliated to *Thaumarchaeota* were higher compared to *nirK* transcripts assigned to
278 the *Nitrospira* clade. In fact, *Thaumarchaeota nirK* transcripts had a 3.4-fold increase after 192
279 hours of incubation relative to earlier sampling points, indicating that *Thaumarchaeota* might
280 have been more active in the reduction of nitrite compared to other steps of nitrification.
281 Specifically, there was a 3.4-fold increase for clade I.1b *nirK* transcripts during the 10 to 192
282 hours of incubation period, whereas the abundance of transcripts from clade I.1a were stable
283 throughout the incubations (Figure 2b).

284 In summary, the metatranscriptomic profiles suggested that AOB, but not comammox,
285 responded rapidly to the nitrogen amendment, whereas AOA followed with less pronounced
286 transcriptome shifts. The response of AOB, and to a lesser extent AOA, was also reflected at
287 the DNA level, albeit with a substantial time delay. For instance, shifts were observed early at
288 the transcript level while at the DNA level, changes were mostly evident 192 hours after the start
289 of incubation (Supplementary Figure 6a, b). These results were consistent across the individual
290 nitrification steps and indicated that at least the AOB nitrifiers grew in response to nitrogen
291 addition.

292

293 **A proteomic perspective in soil microcosms**

294 A metaproteomic analysis of the control and nitrogen-amended microcosms at 192
295 hours of incubation detected a total of 2,892 and 1,629 non-redundant peptides, respectively. A
296 total of 844 peptides were shared among control and nitrogen-amended incubations, whereas

297 2,048 and 785 were exclusively present in each microcosm, respectively. Most of peptides
298 detected in control and nitrogen-amended incubations matched protein sequences predicted
299 from metagenomic assemblies (89.4% and 88.2%, respectively) and the remaining fraction
300 matched reference proteomes (Supplementary Table 6). The top 20 most abundant proteins in
301 control and nitrogen-amended treatment microcosms were related to housekeeping and
302 transport proteins whereas in the latter incubation, oxidoreductases for small carbon and alcohol
303 molecules and ATP synthesis were among the most abundant proteins detected
304 (Supplementary Table 7). The taxonomic affiliation, at the class level, for the most abundant
305 annotated peptides belonged to *Alphaproteobacteria*, *Betaprotebacteria*, and *Acidobacteria* in
306 control and nitrogen-amended incubations. Although there were major compositional changes
307 for abundant groups such as *Betaproteobacteria* (40% decrease) and *Gammaproteobacteria*
308 (50% decrease) (Figure 3a), increased abundance was detected for less abundant groups
309 commonly associated with the nitrification process. For instance, close to a 2.2-fold increased
310 abundance for nitrogen-amended incubations were detected for peptides belonging to
311 *Nitrospira*. Detected peptides related to folding and synthesis were the most abundant and had
312 similar abundances in the control and nitrogen-amended microcosms after 192 hours of
313 incubation. However, the relative abundance of ATP synthases and transcription categories
314 were higher in the nitrogen-amended samples relative to the control, presumably as a
315 consequence of a higher microbial activity generated after the nitrogen input. On the other hand,
316 heat-shock and degradation proteins were more abundant in the control incubation, probably
317 reflecting a more prevailing dormant state for the microbial communities in these samples
318 (Figure 3b). However, unlike the metagenomic and metatranscriptomic datasets, only some
319 peptides involved in nitrification were identified using metaproteomics. For instance, the
320 detected peptides directly involved in nitrification pathways corresponded to the nitrite
321 oxidoreductase subunit B (NxrB), which had a 31.3% abundance increase in the nitrogen-
322 amended samples compared to the control.

DISCUSSION

Using multi-omic approaches for examining process rates

Measuring nitrification rates in incubated soils allowed us to evaluate the explanatory and predictive power of omic approaches in a highly diverse soil system. Despite all three omic approaches revealed increased abundance for target genes, transcripts and proteins related to nitrification pathways, they differed in temporal resolution and quantitative capabilities. For instance, the strongest agreement to the observed nitrification processes (i.e., ammonia or nitrite oxidation) was for the metatranscriptomic data within the first days of incubations (e.g., Figure 2b), whereas metagenomes lagged behind and only reflected the ongoing nitrification process after 192 hours of incubation (e.g., Supplementary Figure 6a, b). These data were presumably attributed to the fact that growth (e.g., at least a few replication cycles) should occur before metagenomics can reveal shifts in relative abundance over time. Note that microbial growth was not explicitly measured by our study to further corroborate the above conclusions and interpretations. Therefore, metagenomics could also reflect underlying microbial processes if the processes are ongoing for a period of time and are coupled with the growth of the corresponding organisms. In contrast, if the goal is to see immediate responses to a perturbation or the perturbation is short-lived (e.g., lasting a few hours), metatranscriptomic data will be preferable. We also observed that metatranscriptomes were as good as metagenomics, if not better, at reflecting microbial activity for nitrification processes even at later incubation time points. In contrast, the metaproteomes offered, at most, a qualitative glimpse at nitrification processes and were less definitive in identifying common nitrification markers. The latter was largely attributable to the computational challenges associated with proteomic data such as high peptide redundancy and the requirement of high-quality assemblies which are still challenging for highly complex soil metagenomes. Furthermore, many challenges remain for efficient

348 extraction of membrane proteins from low abundance organisms such as nitrifiers. Ultimately,
349 these technical limitations could be reflected in a lower number of detected proteins compared
350 to the number of metagenomic and metatranscriptomic reads recovered that encoded the
351 proteins of interest.

352 While shifts in 16S gene ratios (cDNA/DNA) were relatively small for AOA, the 16S and
353 functional gene ratio shifts (e.g., *amoA*) for AOB/NOB were much more pronounced throughout
354 the incubations (Supplementary Figure 3). Nonetheless, there were changes in transcript
355 abundances for nitrification genes from both microbial groups in the microcosms (Figure 2).
356 These results might reflect an active and growing state for AOB/NOB and mostly active AOA
357 communities as observed before for agricultural soil microcosms (17). The differences observed
358 between target gene abundances and 16S gene ratios from AOA could reflect a limitation of the
359 latter data when used as a proxy for assessing microbial activity (18). However, more frequent
360 sampling and incubations under different physicochemical conditions will be required for more
361 robust conclusions to emerge on the exact relationship(s) between molecular level information
362 and process rates. The results reported here provided an overview of this relationship for soils
363 and are highly promising for the future.

364 In terms of the ecological adaptation of the nitrifiers analyzed here, the Havana
365 agricultural site has had a long history of cyclical seasonal inputs (e.g., fertilizers) that have
366 shaped the structure of microbial communities in different soil layers. The AOA and AOB
367 communities in the Havana site have legacy establishments at the 20-30 cm soil depth and are
368 under relatively stable environmental conditions compared to the top soil layer (14). Thus,
369 nitrogen amendments tested in our experiment and experimental conditions might not represent
370 closely the conditions usually experienced by the examined AOA and AOB communities. The
371 rapid response of AOB observed here might be a reflection of physiological adaptations of AOB
372 to thrive under high nitrogen content as reported previously (17). In contrast, the low response
373 observed for comammox and some AOA communities might reflect their limited physiological

374 capabilities to respond to high nitrogen concentrations (2, 3) that were assayed in our
375 experimental setup.

376 Previous authors have also found metatranscriptomic approaches to be better predictors
377 of measured microbial activity (11) in controlled laboratory systems amended with exogenous
378 organic compounds, but have been more limited in providing insights into the whole-microbial
379 community response to the amendment. For instance, the changes in transcripts observed at
380 early incubation points for specific lineages (e.g., comammox vs. betaproteobacterial *amoA*)
381 suggested ongoing microbial activity that became evident only at the DNA level (relative
382 abundance) at the last incubation point in the metagenomes (Figure 2b). Future incubation
383 studies could shed light on the intrinsic differences between nitrifier (and denitrifier) communities
384 by testing variables such as oxygen availability (i.e., water saturation) and different agricultural
385 soil types. For instance, the incubation conditions used in our study deliberately promoted
386 nitrification over denitrification processes and as a result, the N_2O production was detected due
387 to the former process. Consequently, nitric oxide (e.g., *norB*) and nitrous oxide reductases (e.g.,
388 *nosZ*) transcripts, which are responsible for N_2O production and consumption during
389 denitrification, respectively, were not detected in our metatranscriptomes datasets (i.e.,
390 abundance below detection limit). Also, the use of nitrification inhibitors could help to elucidate
391 the origin of the measured N_2O whether production was biotic or abiotic, for which our data are
392 limited in predicting. Thus, the integration of *in situ* rates along with the microbial dynamics
393 examined by metatranscriptomes and metagenomes could provide the means to better
394 understand and predict nitrification and N_2O emission in agricultural soils.

395 **New insights into nitrification pathways**

396 The metagenomic and metatranscriptomic datasets combined with phylogenetic
397 approaches provided a closer examination of the poorly studied microbial diversity in agricultural
398 soils. Assessing the individual gene level, as opposed to whole genome transcript level,

399 provided more robust results for relating population response to measured nitrification reactions,
400 presumably due to higher sequence coverage (less noise). Even though a direct comparison at
401 the genome level between AOB, NOB and comammox AOB was not possible due to the lack of
402 recovered MAGs representing AOB and NOB populations, analysis of individual 16S and
403 nitrogen cycling genes elucidated the importance of AOB and NOB in the microcosm
404 experiments. Our results showed that even though betaproteobacterial *amoA* transcripts
405 responded to the addition of ammonium and urea, the relative abundance of comammox *amoA*
406 transcripts was stable (i.e., not responding to the nitrogen amendment), although comammox
407 populations were relatively more abundant than AOB in the microcosms. This observation is
408 consistent with previous metagenomic results from the same agricultural soil, where comammox
409 *amoA* genes and the organisms encoding these genes represented the highest fraction of
410 nitrifying bacteria (14). The differences between measured genes and transcripts indicated that
411 the incubation conditions favored the activity of *Betaproteobacteria* over comammox nitrifying
412 bacteria, suggesting ecophysiological differences among these taxa for the incubation
413 conditions or added substrates compared to field conditions.

414 The sequencing of isolates and environmental AOA genomes has shown that even
415 though they encode an AmoA protein, they lack a canonical hydroxylamine oxidation pathway
416 (19). Previous studies have proposed that nitric oxide is essential for hydroxylamine oxidation to
417 nitrite in archaea (20). The proposed mechanism involves oxidation of ammonium to
418 hydroxylamine followed by oxidation to nitrite catalyzed by a putative Cu-protein that uses nitric
419 oxide as co-reactant for the oxidation of hydroxylamine. Interestingly, nitric oxide has been
420 proposed to be derived from the activity of the NirK enzyme present in all AOA sequenced
421 genomes. Our results show that unlike AOA *amoA* or bacterial *nirK* transcripts,
422 *Thaumarchaeota nirK* transcripts increased in abundance in the incubated soils, supporting the
423 abovementioned hypothesis. Therefore, even though AOA *amoA* transcripts did not show clear
424 changes in abundances compared to their bacterial counterparts, these results might be in

agreement with the previous hypothesis, and likely denote an unaccounted role for *Thaumarchaeota nirK* in nitrification in agricultural soils.

Multi-omic limitations

Soil samples are challenging to analyze not only because of their heterogeneous structure and chemical composition, but also because of the highly diverse microbial communities and slow growth kinetics. Despite the advancements presented here, there are still opportunities for further improvements. For instance, here we analyzed total RNA extractions from soils where ribosomal rRNA transcripts represented 94-98% of the total sample, limiting our study to a small fraction of transcripts related to functional genes. Current experimental approaches offer successful rRNA depletion for environmental samples, when RNA yields are not limiting (21). Additionally, all the results represented here provide only relative abundances for measured microbial markers. For instance, approaches such as qPCR or internal standards spiked into the DNA or cDNA library for sequencing (21) can strengthen and provide improved quantification compared to those presented here.

Metaproteomics offered an additional layer of information for the microbial activity, but it was less comprehensive compared to metagenomes and metatranscriptomes. Even though our database for proteomic analyses included a high fraction of nitrification proteins predicted from these agricultural soils, only peptides belonging to the NxrB were detected. The results obtained were attributable, at least partially, to the low biomass, especially for the low abundance nitrifiers targeted here. Further, possible protein extraction biases due to the complexity of soil matrices as well as limited extraction of membrane proteins, such AmoA, might have also influenced the outcome of our efforts (22). Nonetheless, the abundances for several peptides belonging to housekeeping proteins of nitrifier organisms were increased during the incubation time, consistent with the results from metagenomic and metatranscriptomic approaches. Therefore, metaproteomics provided a qualitative confirmation of the underlying nitrification

processes ongoing during our incubations and of the responsible taxa. Alternative proteomic approaches focused on a preselected set of proteins (i.e., selected reaction monitoring or target proteomics) could be used to explore low abundance nitrification proteins. For instance, targeted proteomic approaches have been used to study proteins in low abundance involved in bioremediation pathways in highly-diverse environmental systems (23). Therefore, targeted proteomics might offer new opportunities for researchers interested in detecting low-abundance peptides and prediction of process rates in complex samples (24).

The analyses of different omic levels obtained from the incubations showed a high correspondence between nitrification gene markers and nitrification process rates. The gene fragments and transcripts were mostly affiliated to novel nitrifier populations similar to those previously described in field soil metagenomes from the same agricultural site (14). Therefore, the gene and genome sequences reported here could facilitate future investigations of nitrogen cycling in agricultural fields; for instance, by applying qPCR assay targeting the key taxa and biomarker genes and transcripts. The combination of metagenomic and metatranscriptomic approaches used in our study provided a promising strategy for examining microbial activity in agricultural soil environments. Therefore, the findings presented here highlighted the potential of omics data to serve as reliable proxies for examining microbial processes *in situ*, especially in soils, which has been proven to be among the most challenging tasks for environmental studies.

MATERIALS AND METHODS

Soil Sampling

Our study was focused on an agricultural plot located in the Havana County, Illinois, USA (lat 40.296, long 89.944; elevation, 150 m). The site is representative of the US Midwest and has a long history of conventionally managed corn and soybean crop rotation. In October 2014, we collected ~2 kg of bulk soil from a 20-30 cm soil depth as previous results have shown significant presence of ammonia-oxidizing microorganisms in this layer (14).

476

477 **Soil Incubations, Gas and Chemical Analyses**

478 Soil microcosms were established in triplicates, using ~120 g of soil (~8% moisture
479 content) in 500 ml gas-tight canning jars equipped with gas sampling ports, and were sampled
480 at six time points (0, 10, 24, 48, 120, and 192 hours). To set up the microcosms, 6 ml of 40 mM
481 NH_4Cl and 20 mM urea (80 mM N) in water used for irrigation at the site was added to two
482 separate batches of 400 g of soil (Final concentration= 1.2 µmoles-N/g or 18.3 µg-N/g dry
483 weight). Two stable isotope treatments were done, one for NH_4Cl (50% ^{15}N-NH_4Cl and 50% ^{14}N-
484 NH_4Cl) and one for urea (50% ^{15}N-NH_2CONH_2 and 50% ^{14}N- NH_2CONH_2). The two treatments
485 allowed for differentiating how the products of nitrification differed between urea and NH_4Cl
486 when both were present. After vigorously mixing, 120 g were dispensed into three separate
487 microcosm jars and incubated in a dark growth chamber with diurnal temperature fluctuation of
488 22-24 ºC as observed in Havana field soil at 20-30 cm during the spring fertilization period (early
489 June). Triplicate microcosms each receiving 6 ml of filtered irrigation water (no nitrogen
490 amendment) served as controls. After each sampling point, headspace gas was collected from
491 closed jars and the N_2O concentration was measured on a Shimadzu GC-2014 gas
492 chromatograph (Columbia, MD) equipped with an electron capture detector. Jars were opened
493 for soil sampling and to reestablish equilibration with atmospheric air before being resealed until
494 the next sampling. Residual ammonium and nitrate in soil subsamples (20 g) were extracted in
495 2 M KCL and the concentrations were determined using colorimetric analysis on a flow injection
496 auto-analyzer (Lachat Instruments, Milwaukee, WI) (25). Soil pH (1:1 in water) and gravimetric
497 water content were measured at each time point (Supplementary Table 1). ^{15}N isotopic
498 composition of N_2O in collected jar headspace samples was determined using an IsoPrime 100
499 isotope ratio mass spectrometer interfaced with an IsoPrime trace gas analyzer (Cheadle
500 Hulme, UK) at the University of Illinois at Urbana-Champaign. The ^{15}N atom % enrichment of
501 the NO_3^- pool was determined using acid trap diffusion (26) and analysis of the diffusion disks

on a Vario Micro Cube elemental analyzer (Elementar, Hanau, Germany) interfaced to an IsoPrime 100 continuous flow isotope ratio mass spectrometer (Cheadle Hulme, UK). $^{15}NO_3^-$ and $^{15}N_2O$ production rates were calculated from the change in $^{15}NO_3^-$ and $^{15}N_2O$ concentrations, respectively, from one time point to the following sampling time point. NO_3^- and N_2O production rates were estimated from the $^{15}NO_3^-$ and $^{15}N_2O$ production rates based on the mean ^{15}N excess atom % of the NH_4^+ source pool (27). No inhibitors of nitrogen cycle pathways were used in the incubations.

Nucleic Acid Extractions

DNA was extracted from ~0.5 g of soil using a modified phenol-chloroform and purification protocol as previously described (28). For RNA extraction, 2 gr of soil was preserved in LifeGuard (MoBio) and stored at ⁻80°C. A modified protocol derived from the PowerMax Soil DNA kit for extracting RNA was used for total RNA extractions (MoBio). TURBO DNAse (Ambion) was used to remove DNA according to the recommendations of the manufacturer. Nucleic acid extracts were quantified using Quant-it ds DNA HS and HS RNA assays (Invitrogen) according to the instructions of the manufacturer. RNA quality was assessed using Agilent RNA 6000 pico kit (Agilent Technologies) and samples having RNA integrity number (RIN) above 7 were used.

Nucleic Acid Sequencing

For metagenomes, dual-indexed DNA sequencing libraries were prepared using the Illumina Nextera XT DNA library prep kit according to manufacturer's instructions, except that the protocol was terminated after isolation of cleaned amplified double stranded libraries. For metatranscriptomes, single-indexed cDNA sequencing libraries were prepared using ScriptSeq v2 protocol using ~25 ng of total RNA as input. All DNA and cDNA library concentrations were

527 determined by fluorescent quantification using a Qubit HS DNA kit and Qubit 2.0 fluorometer
528 (ThermoFisher Scientific) according to manufacturer's instructions and samples were run on a
529 High Sensitivity DNA chip using the Bioanalyzer 2100 instrument (Agilent) to determine quality
530 and average library insert sizes. An equimolar mixture of the libraries was sequenced on an
531 Illumina HiSEQ 2500 instrument (School of Biological Sciences, Georgia Institute of
532 Technology) for a rapid run of 300 cycles (2 x 150 bp paired end) using the HiSeq Rapid PE
533 Cluster Kit v2 and HiSeq Rapid SBS Kit v2 (Illumina). Adapter trimming and demultiplexing of
534 sequenced samples was carried out by the Illumina software, according to the
535 recommendations of the manufacturer.
536
537 **Short-read Analyses**
538 Metagenomic and metatranscriptomic raw reads (FASTQ) for all samples were trimmed
539 using SolexaQA (29) using a Phred score cutoff of 20 and minimum fragment length of 50 bp.
540 Short-reads derived from metatranscriptomes were merged using PEAR using default
541 parameters (30). Average coverage for each sequenced metagenome was determined by
542 Nonpareil (15) using default settings except that 2,000 reads were used as query (-X option)
543 (Supplementary Tables 3 and 4).
544 Short-read sequences encoding 16S rRNA gene fragments were extracted from each
545 metagenome and metatranscriptome by SortMeRNA (31) and their taxonomy was assigned
546 using RDP classifier (cutoff 50) (32).
547 To identify and quantify reads encoding specific protein sequences of interest, we used
548 the previously published protein sequences as references (14) for the archaeal ammonia
549 monooxygenase alpha subunit (AmoA), bacterial AmoA, hydroxylamine oxidase (HaoA), nitrite
550 oxidoreductase alpha subunit (NxrA), nitrite reductase (NirK), nitric oxide reductase beta subunit
551 (NorB), nitrous oxide (NosZ), nitrite reductase (NrfA) and DNA-directed RNA polymerase
552 subunit beta (RpoB). Independent ROCker (33) models (length=125 bp) were subsequently built

based on these reference protein sequences with the exception of NarG and NxrA, where the sequences were combined into a single model. Trimmed short-reads from soil metagenomes were used as query for BLASTx searches (e-value 0.01) against the latter protein databases and outputs were filtered using the previously generated ROCker models. For metagenomes, target gene abundance in metagenomes was determined as genome equivalents by calculating the ratio between normalized target reads (number of reads matching divided by median protein length) and normalized RpoB reads (number of reads matching divided by median RpoB protein length), a universal single-copy gene. For metatranscriptomes, target transcripts abundance was calculated as reads per kilobase of transcript per million mapped reads (RPKM). Protein databases and ROCker models are available through http://enve-omics.ce.gatech.edu/.

Assembly and Binning of Metagenomic Populations

Short-read metagenomes from control and treatments (t=0,120 and 192 hours) were co-assembled using IDBA_UD v1.1.1 (34) and binning was performed as previously described (14). Taxonomic classification and degree of novelty (novel species, genus, etc) of the MAGs were obtained from the Microbial Genomes Atlas (MiGA) webserver (35). MAG abundance was determined as the total length of all matching metagenomic or metatranscriptomic reads to the binned contigs from BLASTn searches (identity >=98% and fraction of read aligned >= 50%) divided by the metagenomic or metatranscriptomic sample sizes (in millions of reads) and the length of the bin genomes in Kbp (Kilo base pairs). Reads encoding rRNA sequences (such as 5S, 5.8S, 16S, and 23S) were identified by SortMeRNA, and removed for non-rRNA analyses in order to avoid overestimating abundances.

Phylogenetic reconstruction of MAGs was performed based on the concatenated alignment of universal single-copy proteins identified for each bin using the “HMM.essential.rb” script of the enveomics collection (36). For this, thirty bacterial proteins present in the corresponding bins 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". Phylogenetic reconstructions were determined using in RAxML v8.0.19 (-f a, -m PROTGAMMAAUTO, –N 100) and visualized in iTol.

N cycle protein sequences in the co-assembly and MAGs were detected using hidden Markov models obtained from FUNGENE (37), using HMMer (38). Detected target N cycle proteins were manually curated, when necessary, by assessing the presence of characteristic amino acid and phylogenetic congruency.

Phylogenetic Trees and Placement of Short-reads

To assess the phylogenetic affiliation of metagenomic or metatranscriptomic reads, reference and fully assembled protein sequences were aligned using ClustalΩ (39) with default parameters. Resulting alignments were used to build phylogenetic trees in RAxML v8.0.19 (40). Short-reads encoding the protein of interest were extracted from metagenomes or metatranscriptomes using ROCker (BLASTx) and placed in their corresponding phylogenetic tree using the methodology previously described (14). Quantification of the number of reads assigned to a specific clade (e.g., to distinguish between *nxrA* or *narG* reads) was done using the "JPlace.distances.rb" script, also available in the enveomics collection. To quantify *nirK* gene fragments assigned to specific clades, the same process as described above was repeated except that all reads detected by multiple ROCker models to previously described clades (41) (clades I+II, III and *Thaumarchaeotea*) were used.

Shotgun Metaproteomics

Approximately 10 g of soil were collected from the 192 hours control and ^{15}N-NH_4^+ amended microcosms and stored at -80°C. Frozen soil (5 g) was thawed and suspended in lysis buffer and boiled for 15 minutes as described previously (42). The supernatant was

retained and amended with 100% chilled TCA to final concentration of 25% (vol/vol) and kept at -20°C overnight. Samples were centrifuged at 21,000 x *g* for 20 min and the protein pellets processed as described previously (43) and solubilized in 6 M guanidine buffer (6 M guanidine; 10 mM dithiothreitol [DTT] in Tris-$CaCl_2$ buffer (10 mM Tris; , pH 7.8) with 3 hr incubation at 60°C. An aliquot of 25 µl/ sample was retained for protein estimation and the rest of the protein sample was digested, peptides desalted and solvent exchanged as described earlier (44). The amount of protein extracted from each sample was calculated using the RC/DC protein estimation kit (Bio-Rad Laboratories, Hercules, CA, USA) as per the manufacturer's instructions. Bovine serum albumin (supplied with the kit) was used as standard for the assay.

All chemicals were obtained from Sigma Chemical Co. (St. Louis, MO), unless specified otherwise. High performance liquid chromatography- (HPLC-) grade water and other solvents were obtained from Burdick & Jackson (Muskegon, MI), 99% formic acid was purchased from EM Science (Darmstadt, Germany) and sequencing-grade trypsin was acquired from Promega (Madison, WI).

NanoLC-MS/MS Analysis.

Peptides (75 ug) were loaded onto in-house prepared biphasic resin packed column [SCX (Luna, Phenomenex, Torrance, CA) and C18 (Aqua, Phenomenex, Torrance, CA)] as described earlier (44, 45) and subjected to an offline wash for 15 min as previously described (46). The sample column was aligned with an in-house C18 packed nanospray tip (New Objective, Woburn, MA) connected to a Proxeon (Odense, Denmark) nanospray source as previously detailed (46). Peptides were eluted and subjected to chromatographic separation and measurements via 24-hr Multi-Dimensional Protein Identification Technology (MuDPIT) approach as described earlier (44-46). Measurements were carried out using LTQ mass spectrometer (Thermo Fisher Scientific, Germany) coupled to the Ultimate 3000 HPLC system

(Dionex, USA) and operated in data dependent mode, via Thermo Xcalibur software V2.1.0 as described earlier (45).

For protein identification, the raw spectra from each run were searched against a custom database and was constructed using protein sequences predicted from metagenome assemblies obtained from the same soil and 20-30 cm depth (14), metagenome assemblies from incubations (Supplementary Table 2), and reference proteomes for 47 common soil organisms (Supplementary Table 5). These predicted proteins were used for constructing a database for metaproteomic searches (available through http:// http://enve-omics.ce.gatech.edu/data/multiomics-soil). Database matching was done via Myrimatch v2.1 algorithm (47) set to parameters described before (48) with minor modifications where static cysteine and dynamic oxidation modifications were not considered. Identification of at least two peptides per protein (one unique and one non-unique) sequence was a prerequisite for protein identifications. Common contaminant peptide sequences from trypsin and keratin were concatenated to the database. Reverse database sequences were also included in the database as decoy sequences to calculate false discovery rate (FDR). For data analysis, spectral counts of identified peptides was normalized as described before (49) to obtain the normalized spectral abundance factor (NSAF) and the NSAF values were multiplied by a constant number (100,000) for better visualization and referred to as normalized spectral counts (nSpc). The nSpc were used to compare expression of proteins across different samples and different time points. Detected proteins predicted from metagenomic assemblies were annotated using BLASTp (50) and UniProt database as reference (51) (downloaded in May of 2017).

Accession numbers

Raw metagenomic and metatranscriptomic soil datasets and MAGs are deposited in the European Nucleotide archive under study number PRJEB27434.

656

657 **ACKNOWLEDGMENTS**

658 We thank Joel Kostka and Alissa Hooker for helpful discussions related to the manuscript. This
659 work was supported in part by U.S. Department of Energy, Office of Biological and
660 Environmental Research, Genomic Science Program [award DE-SC0006662], US National
661 Science Foundation [Award 1831582], and the Chilean Fulbright-Conicyt doctoral scholarship
662 [L.H.O.].

663

664 **Figure Legends**

665

666 **Figure 1. Nitrogen pools and fluxes in soil incubations amended with NH_4^+ and urea.**
667 Mean NH_4^+ and NO_3^- concentrations (A), total NO_3^- production rate (B), and total N_2O production
668 rate (C) for the nitrogen-amended and control (irrigation water only) microcosms at each
669 incubation time point. Error bars represent the standard deviation from replicate samples (n=6
670 for nitrogen-amended and n=3 for control).

671

672 **Figure 2. Nitrification genes in incubated soils.** A. Relative expression ratios for each
673 nitrification step in incubated soils were determined at 10, 48, 120, and 192 hours incubation. B
674 and C show determined RPKM values for bacterial *amoA* (B), *hao* (C), thaumarchaeotal *amoA*
675 (D), and *nirK* (E) transcripts from metatranscriptomes.

676

677 **Figure 3. Metaproteomic analyses of incubated soils at 192 hours of incubation.** Panel A
678 shows taxonomic affiliation (class) and abundance (average spectral counts) for peptides
679 detected in control and N-amended incubations. Panel B shows summarized functional
680 annotation of detected peptides using SEED functional categories.

681

682 **REFERENCES**

683

684 1. **Ravishankara AR**, **Daniel JS**, **Portmann RW**. 2009. Nitrous oxide (N_2O): the dominant
685 ozone-depleting substance emitted in the 21st century. Science **326**:123–125.

686 2. **Daims H**, **Lebedeva EV**, **Pjevac P**, **Han P**, **Herbold C**, **Albertsen M**, **Jehmlich N**,
687 **Palatinszky M**, **Vierheilig J**, **Bulaev A**, **Kirkegaard RH**, **Bergen von M**, **Rattei T**,
688 **Bendinger B**, **Nielsen PH**, **Wagner M**. 2015. Complete nitrification by *Nitrospira*
689 bacteria. Nature **528**:504–509.

690 3. **van Kessel MAHJ**, **Speth DR**, **Albertsen M**, **Nielsen PH**, **Op den Camp HJM**, **Kartal B**,
691 **Jetten MSM**, **Lücker S**. 2015. Complete nitrification by a single microorganism. Nature
692 **528**:555–559.

4. **Kool DM**, **Dolfing J**, **Wrage N**, **Van Groenigen JW**. 2011. Nitrifier denitrification as a distinct and significant source of nitrous oxide from soil. Soil Biol Biochem **43**:174–178.

5. **Zhu X**, **Burger M**, **Doane TA**, **Horwath WR**. 2013. Ammonia oxidation pathways and nitrifier denitrification are significant sources of N2O and NO under low oxygen availability. Proc Natl Acad Sci USA **110**:6328–6333.

6. **Prosser JI**, **Nicol GW**. 2012. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. Trends Microbiol **20**:523–531.

7. **Gubry-Rangin C**, **Nicol GW**, **Prosser JI**. 2010. Archaea rather than bacteria control nitrification in two agricultural acidic soils. FEMS Microbiol Ecol **74**:566–574.

8. **Hettich RL**, **Sharma R**, **Chourey K**, **Giannone RJ**. 2012. Microbial metaproteomics: identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. Curr Opin Microbiol **15**:373–380.

9. **Hug LA**, **Thomas BC**, **Sharon I**, **Brown CT**, **Sharma R**, **Hettich RL**, **Wilkins MJ**, **Williams KH**, **Singh A**, **Banfield JF**. 2015. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. Environ Microbiol **18**:159–173.

10. **Jones SE**, **Lennon JT**. 2010. Dormancy contributes to the maintenance of microbial diversity. Proc Natl Acad Sci USA **107**:5881–5886.

11. **Helbling DE**, **Ackermann M**, **Fenner K**, **Kohler H-PE**, **Johnson DR**. 2012. The activity level of a microbial community function can be predicted from its metatranscriptome. ISME J **6**:902–904.

12. **Schneider T**, **Keiblinger KM**, **Schmid E**, **Sterflinger-Gleixner K**, **Ellersdorfer G**, **Roschitzki B**, **Richter A**, **Eberl L**, **Zechmeister-Boltenstern S**, **Riedel K**. 2012. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. ISME J **6**:1749–1762.

13. **Hultman J**, **Waldrop MP**, **Mackelprang R**, **David MM**, **McFarland J**, **Blazewicz SJ**, **Harden J**, **Turetsky MR**, **McGuire AD**, **Shah MB**, **VerBerkmoes NC**, **Lee LH**, **Mavrommatis K**, **Jansson JK**. 2015. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature **521**:208–212.

14. **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.

15. **Rodriguez-R LM**, **Konstantinidis KT**. 2014. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics **30**:629–635.

16. **Luo C**, **Rodriguez-R LM**, **Konstantinidis KT**. 2014. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res **42**:e73–e73.

17. **Jia Z**, **Conrad R**. 2009. Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. Environ Microbiol **11**:1658–1671.

18. **Blazewicz SJ**, **Barnard RL**, **Daly RA**, **Firestone MK**. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J **7**:2061–2068.

19. **Stahl DA**, **la Torre de JR**. 2012. Physiology and diversity of ammonia-oxidizing archaea. Annu Rev Microbiol **66**:83–101.

20. **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.

21. **Tsementzi D**, **Poretsky R**, **Rodriguez-R LM**, **Luo C**, **Konstantinidis KT**. 2014. Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. Environ Microbiol Rep **6**:640–655.

22. **VerBerkmoes NC**, **Denef VJ**, **Hettich RL**, **Banfield JF**. 2009. Systems Biology: Functional analysis of natural microbial consortia using community proteomics. Nat Rev Microbiol **7**:196–205.

23. **Werner JJ**, **Ptak AC**, **Rahm BG**, **Zhang S**, **Richardson RE**. 2009. Absolute quantification of *Dehalococcoides* proteins: enzyme bioindicators of chlorinated ethene dehalorespiration. Environ Microbiol **11**:2687–2697.

24. **Hood LE**, **Omenn GS**, **Moritz RL**, **Aebersold R**, **Yamamoto KR**, **Amos M**, **Hunter Cevera J**, **Locascio L**. 2012. New and improved proteomics technologies for understanding complex biological systems: Addressing a grand challenge in the life sciences. Proteomics **12**:2773–2783.

25. **Yang WH**, **Traut BH**, **Silver WL**. 2015. Microbially mediated nitrogen retention and loss in a salt marsh soil. Ecosphere **6**:art7.

26. **Herman DJ**, **Brooks PD**, **Ashraf M**, **Azam F**, **Mulvaney RL**. 1995. Evaluation of methods for nitrogen-15 analysis of inorganic nitrogen in soil extracts. II. Diffusion methods. Commun Soil Sci Plant Anal **26**:1675–1685.

27. **Templer PH**, **Silver WL**, **Pett-Ridge J**, **DeAngelis KM**, **Firestone MK**. 2008. Plant and microbial controls on nitrogen retention and loss in a humid tropical forest. Ecology **89**:3030–3040.

28. **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.

29. **Cox MP**, **Peterson DA**, **Biggs PJ**. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics **11**:485.

30. **Zhang J**, **Kobert K**, **Flouri T**, **Stamatakis A**. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics **30**:614–620.

31. **Kopylova E**, **Noé L**, **Touzet H**. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics **28**:3211–3217.

32. **Wang Q**, **Garrity GM**, **Tiedje JM**, **Cole JR**. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol **73**:5261–5267.

33. **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.

34. **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.

35. **Rodriguez-R LM**, **Gunturu S**, **Harvey WT**, **Rosselló-Móra R**, **Tiedje JM**, **Cole JR**, **Konstantinidis KT**. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res **84**:e00014.

36. **Rodriguez-R LM**, **Konstantinidis KT**. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints **4**:e1900v1.

37. **Fish JA**, **Chai B**, **Wang Q**, **Sun Y**, **Brown CT**, **Tiedje JM**, **Cole JR**. 2013. FunGene: the functional gene pipeline and repository. Frontiers in Microbiology **4**:291.

38. **Eddy SR**. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol **7**:e1002195.

39. **Sievers F**, **Wilm A**, **Dineen D**, **Gibson TJ**, **Karplus K**, **Li W**, **Lopez R**, **McWilliam H**, **Remmert M**, **Söding J**, **Thompson JD**, **Higgins DG**. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7**:539–539.

40. **Stamatakis A**. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688–2690.

41. **Wei W**, **Isobe K**, **Nishizawa T**, **Zhu L**, **Shiratori Y**, **Ohte N**, **Koba K**, **Otsuka S**, **Senoo K**. 2015. Higher diversity and abundance of denitrifying microorganisms in environments than considered previously. ISME J **9**:1–12.

42. **Chourey K**, **Jansson J**, **VerBerkmoes N**, **Shah M**, **Chavarria KL**, **Tom LM**, **Brodie EL**, **Hettich RL**. 2010. Direct cellular lysis/protein extraction protocol for soil metaproteomics. J Proteome Res **9**:6615–6622.

43. **Chourey K**, **Nissen S**, **Vishnivetskaya T**, **Shah M**, **Pfiffner S**, **Hettich RL**, **Löffler FE**. 2013. Environmental proteomics reveals early microbial community responses to biostimulation at a uranium- and nitrate-contaminated site. Proteomics **13**:2921–2930.

805 44. **Thompson MR**, **VerBerkmoes NC**, **Chourey K**, **Shah M**, **Thompson DK**, **Hettich RL**.
806 2007. Dosage-dependent proteome response of *Shewanella oneidensis* MR-1 to acute
807 chromate challenge. J Proteome Res **6**:1745–1757.

808 45. **Brown SD**, **Thompson MR**, **VerBerkmoes NC**, **Chourey K**, **Shah M**, **Zhou J**, **Hettich**
809 **RL**, **Thompson DK**. 2006. Molecular dynamics of the *Shewanella oneidensis* response
810 to chromate stress. Mol Cell Proteomics **5**:1054–1071.

811 46. **Sharma R**, **Dill BD**, **Chourey K**, **Shah M**, **VerBerkmoes NC**, **Hettich RL**. 2012.
812 Coupling a detergent lysis/cleanup methodology with intact protein fractionation for
813 enhanced proteome characterization. J Proteome Res **11**:6008–6018.

814 47. **Tabb DL**, **Fernando CG**, **Chambers MC**. 2007. MyriMatch: highly accurate tandem
815 mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome
816 Res **6**:654–661.

817 48. **Xiong W**, **Giannone RJ**, **Morowitz MJ**, **Banfield JF**, **Hettich RL**. 2015. Development of
818 an enhanced metaproteomic approach for deepening the microbiome characterization of
819 the human infant gut. J Proteome Res **14**:133–141.

820 49. **Paoletti AC**, **Parmely TJ**, **Tomomori-Sato C**, **Sato S**, **Zhu D**, **Conaway RC**, **Conaway**
821 **JW**, **Florens L**, **Washburn MP**. 2006. Quantitative proteomic analysis of distinct
822 mammalian Mediator complexes using normalized spectral abundance factors. Proc Natl
823 Acad Sci USA **103**:18928–18933.

824 50. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**, **Madden TL**.
825 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421

826 51. **UniProt Consortium**. 2015. UniProt: a hub for protein information. Nucleic Acids Res
827 **43**:D204–12.

828

Figure 1

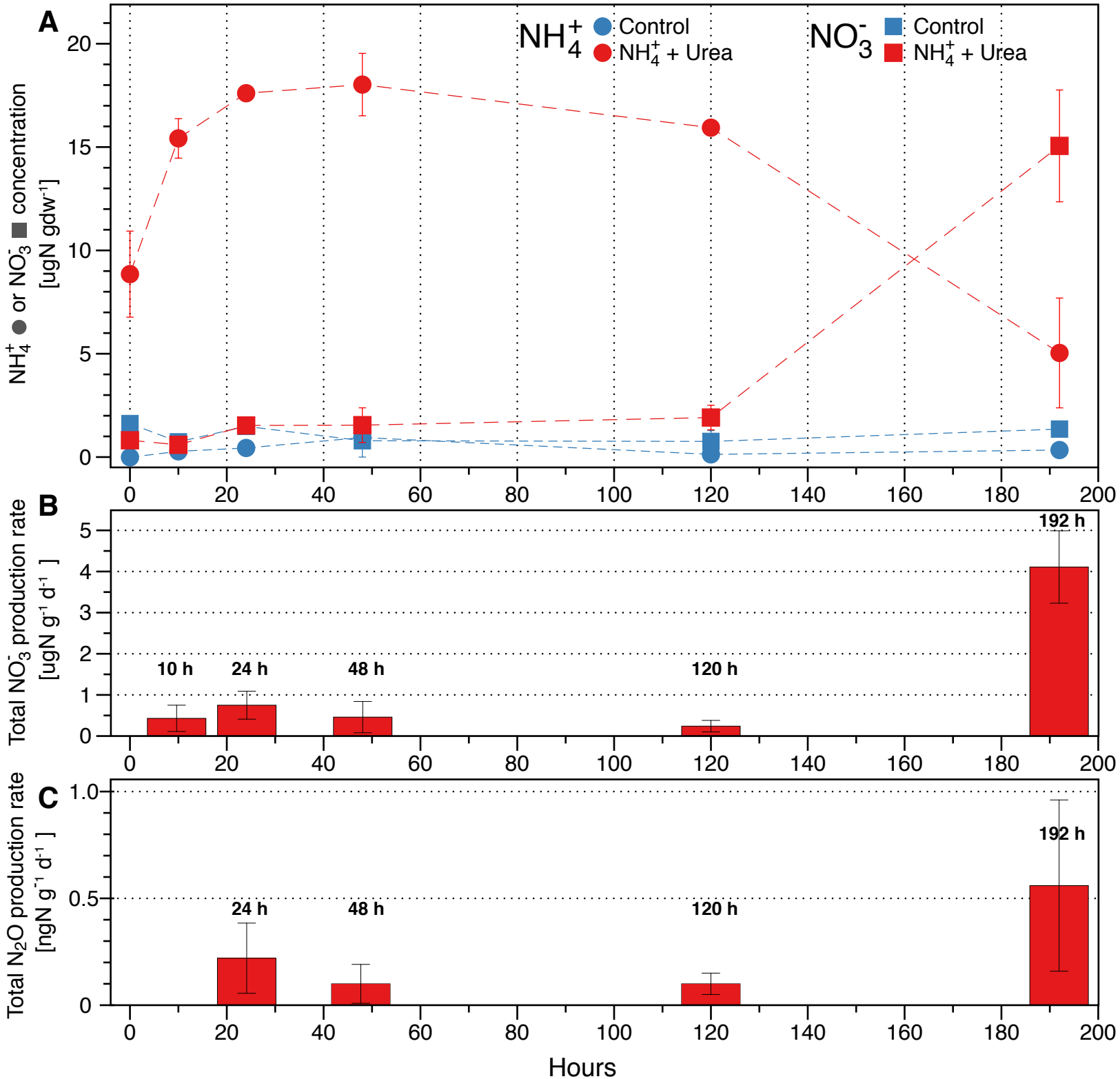
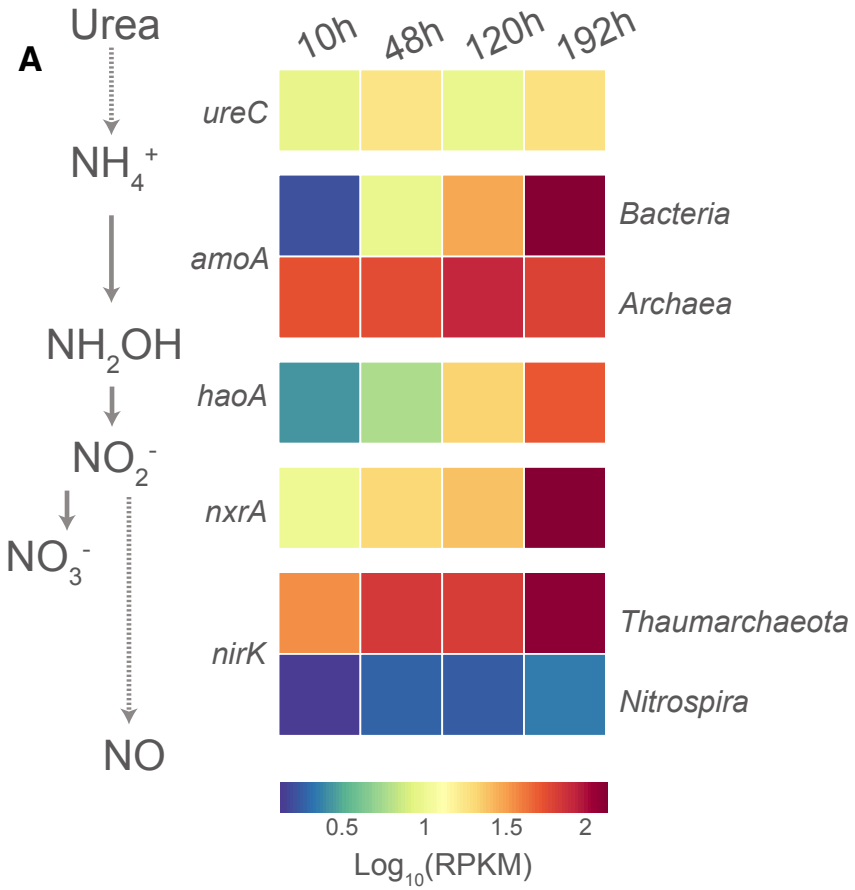

A
NH_4^+ Control
NH_4^+ + Urea
NO_3^- Control
NH_4^+ + Urea
NH_4^+ ● or NO_3^- ■ concentration [ugN gdw^{-1}]
B
Total NO_3^- production rate [ugN g^{-1} d^{-1}]
10 h
24 h
48 h
120 h
192 h
C
Total N_2O production rate [ngN g^{-1} d^{-1}]
24 h
48 h
120 h
192 h
Hours

Figure 2


A
Urea
NH_4^+
NH_2OH
NO_2^-
NO_3^-
NO
10h
48h
120h
192h
ureC
amoA
haoA
nxrA
nirK
Bacteria
Archaea
Thaumarchaeota
Nitrospira
0.5
1
1.5
2
Log_{10}(RPKM)

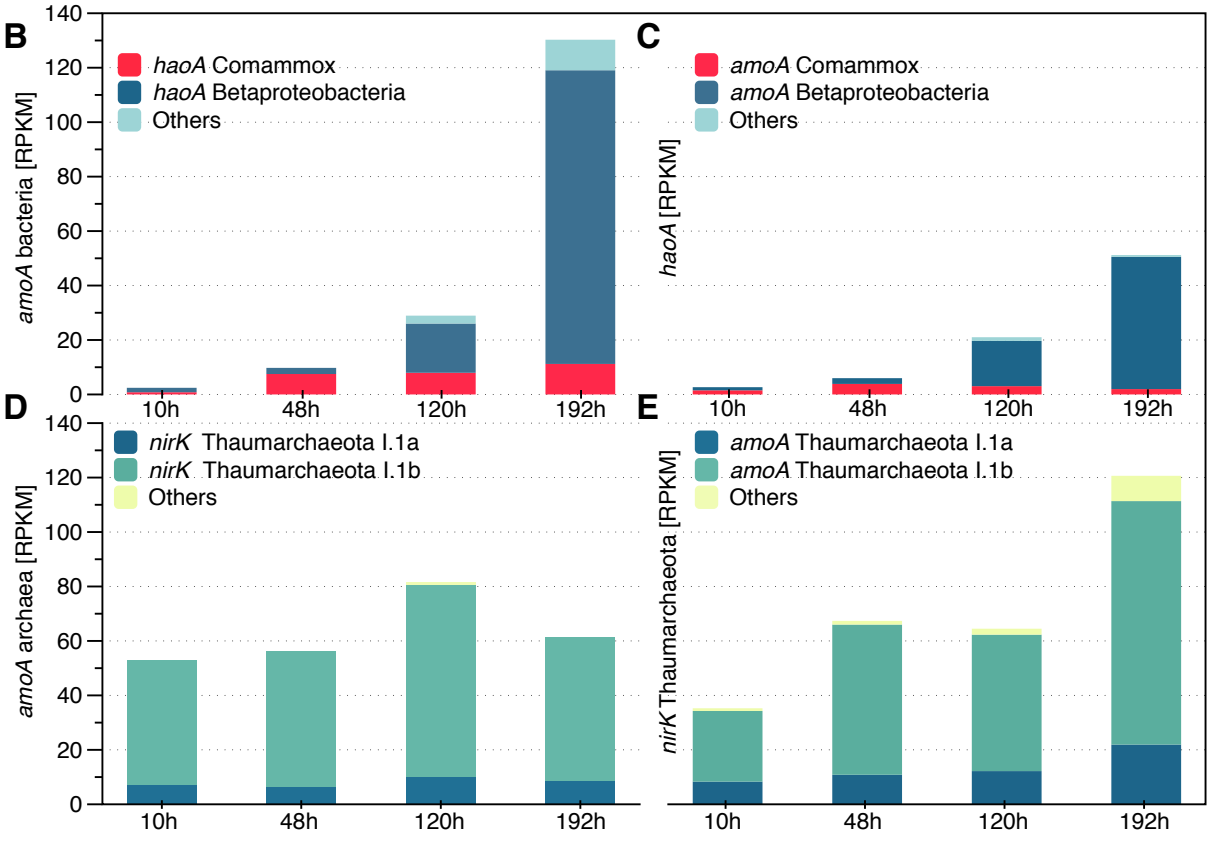
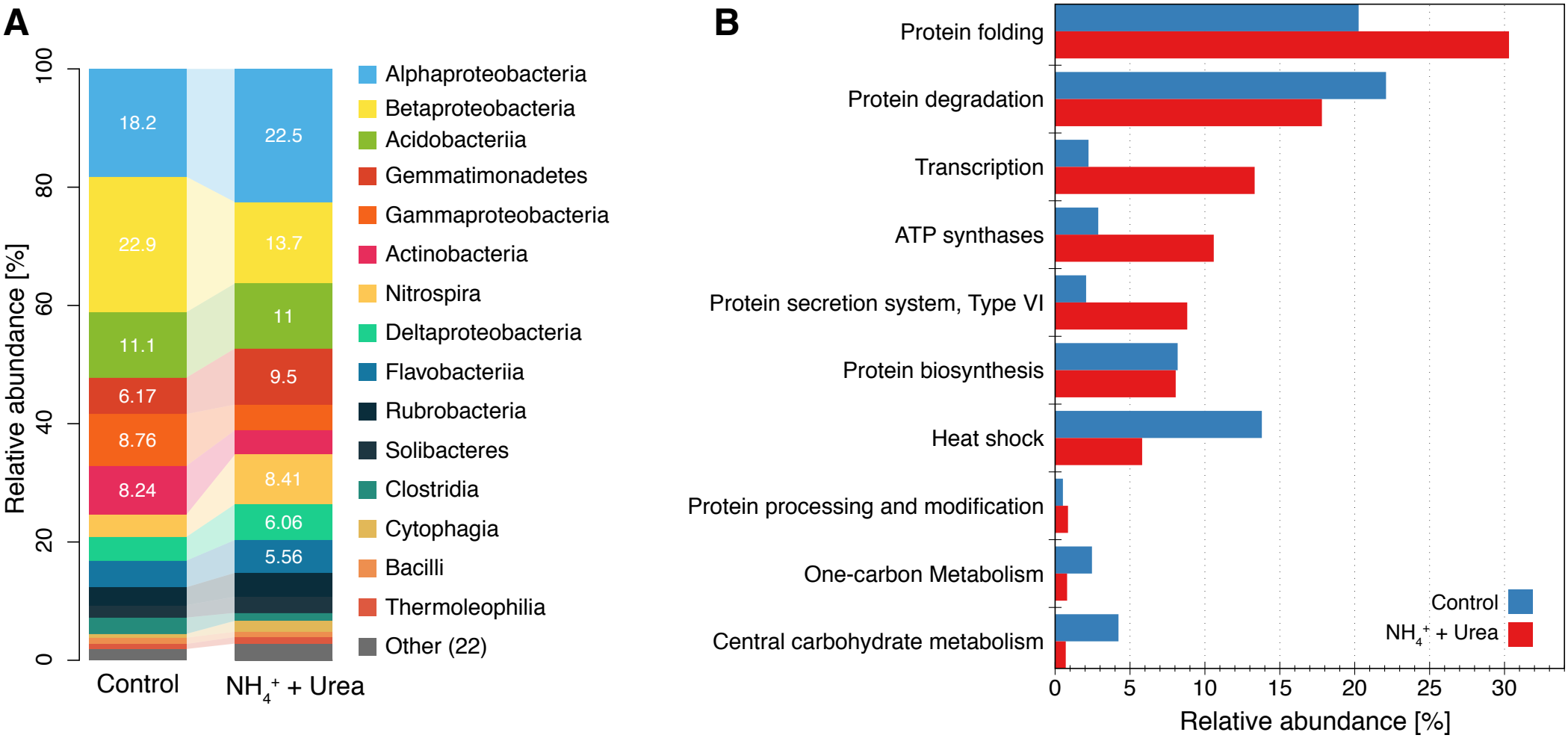

B
amoA bacteria [RPKM]
haoA Comammox
haoA Betaproteobacteria
Others
C
haoA [RPKM]
amoA Comammox
amoA Betaproteobacteria
Others
D
amoA archaea [RPKM]
nirK Thaumarchaeota I.1a
nirK Thaumarchaeota I.1b
Others
E
nirK Thaumarchaeota [RPKM]
amoA Thaumarchaeota I.1a
amoA Thaumarchaeota I.1b
Others
0
20
40
60
80
100
120
140
10h
48h
120h
192h

Figure 3


A
Relative abundance [%]
0
20
40
60
80
100
18.2
22.9
11.1
6.17
8.76
8.24
22.5
13.7
11
9.5
8.41
6.06
5.56
Control
NH4+ + Urea
Alphaproteobacteria
Betaproteobacteria
Acidobacteriia
Gemmatimonadetes
Gammaproteobacteria
Actinobacteria
Nitrospira
Deltaproteobacteria
Flavobacteriia
Rubrobacteria
Solibacteres
Clostridia
Cytophagia
Bacilli
Thermoleophilia
Other (22)
B
Protein folding
Protein degradation
Transcription
ATP synthases
Protein secretion system, Type VI
Protein biosynthesis
Heat shock
Protein processing and modification
One-carbon Metabolism
Central carbohydrate metabolism
Control
NH4+ + Urea
0
5
10
15
20
25
30
Relative abundance [%]