Rendering the metabolic wiring powering wetland soil methane production

Authors: Angela M. Oliverio[1,2+], Adrienne B. Narrowe[1+], Jorge A. Villa[3], Christian Rinke[4], David W. Hoyt[5], Pengfei Liu[1,6], Bridget B. McGivern[1], Emily K. Bechtold[1], Jared B. Ellenbogen[1], Rebecca A. Daly[1], Garrett J. Smith[7], Jordan C. Angle[7], Rory M. Flynn[1], Andrew P. Freiburger[8], Katherine B. Louie[9,10], Brooke Stemple[1], Trent R. Northen[9,10], Christopher Henry[8], Christopher S. Miller[11], Timothy H. Morin[12], Gil Bohrer[13], Mikayla A. Borton[1*], Kelly C. Wrighton[1*]

[1] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, 80521
[2] Department of Biology, Syracuse University, Syracuse, NY, 13210
[3] School of Geosciences, University of Louisiana at Lafayette, Lafayette, LA 70503
[4] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD, Australia
[5] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, 99352
[6] Center for Pan-third Pole Environment, Lanzhou University, Lanzhou 730000, China
[7] Department of Microbiology, The Ohio State University, Columbus, OH 43210
[8] Data Science and Learning, Argonne National Laboratory, Lemont, IL 60439
[9] Lawrence Berkeley National Lab, Berkeley, CA, 94720
[10] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720
[11] Department of Integrative Biology, University of Colorado Denver, Denver, CO, 80217
[12] Department of Environmental Resources Engineering, State University of New York College of Environmental Sciences and Forestry, Syracuse, New York, 13210
[13] Department of Civil, Environmental & Geodetic Engineering, The Ohio State University, Columbus, OH, 43210

[+] Authors contributed equally to this work
*Co-Corresponding Authors: Kelly C. Wrighton and Mikayla A. Borton

Email: wrighton@colostate.edu and mborton@colostate.edu

Keywords: genome-resolved metagenomics, metatranscriptomics, global change, greenhouse gases, methanogen

Abstract

Despite accounting for less than 8% of Earth's land coverage, freshwater wetlands are the foremost contributor to global methane emissions. Although methane in wetlands originates from soil microbial processes, discerning the primary active microbial lineages and their metabolic roles has proven elusive. Here we sampled one of the most prolific methane-emitting wetlands in the continental United States over a five-year span and constructed a microbial genome resource of 17,333 metagenome assembled genomes. This catalog encompasses 2,502 unique genomes from 72 bacterial and archaeal phyla with more than half representing novel lineages lacking genomic representation. Integration of these wetland microbial genomes with 133 spatially and temporally resolved metatranscriptomes, along with paired amplicon, geochemical and greenhouse gas data from over 700 samples, afforded new insights into the metabolic reservoir harbored in wetland soils. We delineate the transcriptional contributions of each genome to biogeochemically relevant traits to uncover the dominant axes of environmental variation in wetland soils. Despite extended flooding captured in our sampling, soil redox shifts failed to reconfigure the transcriptional profiles of 101 methanogen and methanotroph genomes, indicating microbial methane cycling resilience to climate perturbations. Employing regression-based modeling, we identified the integrated carbon and energy assignments responsible for porewater methane hotspots, exposing the microbial metabolic circuitry that underpins wetland methane production. Our findings show how microbial-scale biogeochemical interactions within freshwater wetland soils can amplify into significant ecosystem-scale fluxes.

Introduction

Wetland soils play a pivotal role in the global carbon cycle by storing nearly a third of soil carbon, surpassing even forests in their capacity for carbon sequestration (1, 2). Despite this crucial function, these predominantly saturated soils, characterized by high organic matter and low porewater oxygen concentrations, can become a source of microbially derived methane, the second most potent greenhouse gas after carbon dioxide (3–6). Notably, wetlands are the largest natural source of methane, contributing significantly to the uncertainties in global methane emissions (7). The microbial metabolic pathways responsible for transforming soil organic carbon into methane, along with their environmental and biological controls, remain inadequately understood. This knowledge gap is one of the barriers to incorporation of microbial processes into land surface and biogeochemical models, and is needed to improve predictions of methane flux from wetlands at local, regional, and global scales (5, 8, 9).

Over the last decade metagenomic sequencing approaches have described the microbial community membership and function associated with methane production across a range of habitats. In comparison to other high methane-emitting systems like wastewater (10), ruminants (11), or subsurface systems (12), wetland soils have historically been far less genomically tractable as they contain thousands of microbial species dispersed over heterogeneous geochemical gradients that change over centimeter-scale in depths, meter-scale laterally, and at daily-monthly scale temporally (5, 13–15). Moreover, despite the high rates of methane emitted from wetland soils, methanogens, the microorganisms responsible for methane production, are rare members in these soils (14, 16, 17) and thus remain under sampled with metagenomic approaches.

Only a handful of genome-resolved studies from freshwater wetlands exist to date, and nearly all are focused on organic soils from northern peatlands, with or without permafrost (18–21). A single permafrost-focused study that recovered ~1,500 metagenome assembled genomes (MAGs) (19) serves today as the most comprehensive microbial genomic catalog for all wetlands soil types. However, the highest methane-contributing wetland types, temperate and tropical marshes, have limited genomic representation with only a few genome-resolved studies to date (22–24). These studies do not focus on the breadth of metabolisms in wetlands, often highlighting single biogeochemical processes (22, 23, 25).

To meet this need, we conducted a sampling initiative within the Old Woman Creek (OWC) National Estuarine Research Reserve temperate wetland, cataloging hundreds of spatially, temporally, and depth resolved samples. Situated alongside Lake Erie, this freshwater, mineral soil marsh is susceptible to climate-induced flood events. Its selection for this study was informed by the observation of remarkably high methane fluxes, surpassing ten times the median for wetlands within a comparable annual temperature range (26, 27). Reconstructing thousands of genomes and pairing this information to soil metatranscriptomes and metabolomes unveiled new organisms, metabolisms, and interactions that sustain the methane cycle. The insights derived from this resource will advance comprehension of the complex microbial metabolisms governing biogeochemical processes within temperate wetland soils.

Results and Discussion

Freshwater wetland soils harbor thousands of novel and transcriptionally active genomes

We constructed a microbial greenhouse gas observatory at one of the highest methane-emitting wetlands in the AmeriFlux network (site ID US-OWC), an investigator network with

tailored instrumentation, data processing, and QA procedures for flux measurements (27–30) (**Fig. 1A**). Methane porewater concentrations were measured once per month in the summers from centimeter-depth resolved soil porewater samplers (peepers) at multiple locations throughout the wetland (**Fig. 1A**). These data were coupled to microbial communities and soil geochemistry from 705 samples (***SI Appendix*, Fig. S1, Table S1-2**). Over 5 years we repeatedly sampled ten sites spanning four land coverage types, with each site sampled at least 52 times (**Fig. 1A**). Screening for methanogen membership in these samples with 16S rRNA amplicon sequencing, we identified 42 samples for deep metagenomic sequencing, generating up to 300 Gbp per sample (3.1 Tbp total). Collectively, this effort doubled the amount of metagenomic data from wetland soils available in public repositories (***SI Appendix*, Fig. S1**). Multiple assembly and binning methods targeted the recovery of low abundant methanogens and other strains, recovering over 17,333 draft genomes, resulting in a final collection of 2,502 unique, high and medium quality metagenome assembled genomes (MAGs) from temperate wetland soils (**Fig. 1B**).

Here we present the Multi-omics for Understanding Climate Change (MUCC) database, a multi-omics and biogeochemical resource for temperate wetland soils. MUCC contains 60 bacterial (N=2,204 MAGs) and 12 archaeal phyla (n=298 MAGs). The most represented phyla include the archaeal Halobacteriota (n=73 MAGs) and the bacterial Proteobacteria (n=484), Acidobacteriota (n=316 MAGs), Chloroflexota (n=237 MAGs), and Desulfobacterota_H (n=140). A majority (57%) of the MUCC genomes were designated as new representatives at the genus level or higher, many of which constituted dominant members in these soils (**Fig. 1C**). For example, we recovered representatives from 1 new bacterial phylum (here referred to as New Phylum01_b), 8 classes, 40 orders, 93 families, and 639 genera (**Fig. 1B; *SI Appendix*, Table S3**). In addition, over a hundred genomes were from poorly or previously uncharacterized phyla often denoted only by an alphanumeric identifier or placeholder name (**Fig 1B**), such as QMZS01, CSP1-3, and TA06_A among others. This genomic catalog of new or poorly described lineages reflects the vast under sampling of temperate wetland soils even from the most comprehensive genomic repositories (31).

A valuable component of MUCC is the broad sampling of archaeal MAGs (n=298), exceeding the sampling from previous genomic compendiums from subsurface groundwater (24 MAGs) (32), permafrost wetland soil (95 MAGs) (19), human gut (28 MAGS) (33), or rumen (126 MAGS) (34) microbiomes. MUCC archaeal genomes spanned 12 phyla with 45 families including 98 genera, 10 of which were new family level designations identified here (**Fig. 1B**). Additionally, our data set included a total of 85 high and medium quality dereplicated methanogen MAGs. These genomes belonged to 4 methanogen phyla (Crenarchaeota, Euryarchaeota, Halobacteriota, Thermoplasmatota) that included 25 genera (15 of which were first identified here) from 8 families. One of these families was a new methanogen family (NewFamily01_m) in the Methanomicrobiales. Within this family, we reconstructed 25 MAGs assigned to 4 new genera. Lastly beyond canonical methanogens, MUCC also included members of presumed non-methanogenic, *mcrA*-containing taxa such as the anaerobic methanotrophic *Methanoperedens* genus and the alkane-oxidizing Helarchaeales.

We next mapped 2.7 Tbp of data from 133 metatranscriptomes to the MUCC database, finding three quarters of the MAGs recruited transcripts (***SI Appendix*, SI Methods, Table S4**). This finding reinforces the complexity and richness harbored in soil microbiomes. Broadly the mean transcription of genes within a genome was greater in the Archaea (26%) than the Bacteria (9%) (***SI Appendix*, Fig. S2**). The top 30 most active bacterial and archaeal genomes (max

transcription) were also prevalent across the site, detected in at least 35% of the metatranscriptomes. Notably, only 3 of these 30 most active MAGs had known taxonomic assignments (e.g., methanogen *Methanomethylicus* and *Methanoregula*) with the rest being newly discovered here (new Genus) or identified by alphanumeric indicator (**Fig. 1C**). Underscoring the need for comprehensive genomic sampling across a spectrum of soil ecosystems, these discoveries emphasize the previously enigmatic biogeochemical contributions originating from these inadequately characterized lineages.

We next assessed the relevance of these MUCC genomes to other wetland collections. First using 16S rRNA comparison, we show half of the taxa identified here were also dominant in the microbial communities from other North American wetlands (n=505 samples from 5 wetlands) (***SI Appendix*, Fig. S3**). We also evaluated the genomic sampling across saturated soil systems, comparing the membership and mean transcription of OWC wetland methanogen MAGs to those from arctic peatland soils in northern Sweden (19). We found that genomes for the same methanogenic genera were recovered from both sites and that OWC derived MAGs had detectable levels of transcription in Stordalen Mire metatranscriptomes (***SI Appendix*, Fig. S4**). Many of the dominant genera recovered from OWC are prevalent in wetlands systems more generally, indicating this resource will be valuable broadly for freshwater wetlands researchers (see ***Data availability***).

Metabolic trait expression is responsive to wetland environmental gradients

Leveraging MUCC, we placed microbial genomes into a broader geochemical framework. To do this, we developed a genome-based rule set for 14 microbial metabolic traits that impact the biogeochemical transformations of oxygen (aerobic respiration, microaerophilic respiration), methane (methanogenesis, aerobic methanotrophy, anaerobic methanotrophy), sulfur (sulfur oxidization, sulfur reduction), nitrogen (nitrification, nitrogen reducer, nitrogen fixation, and Dissimilatory Nitrate Reduction to Ammonium (DNRA)), iron (iron oxidation, iron reduction), and others (phototrophy, obligate fermentation) (***SI Appendix*, Fig. S5)**. Each genome within MUCC was assessed for these 14 metabolic traits based on their genomic potential and expressed genes resulting in 1,136 MAGs that expressed at least one trait (**Fig. 2**, **Table S5)**.

Trait assignments offered an opportunity to examine how genomic potential is translated to expressed function. The most dominant metabolic traits (i.e., detected in the greatest number of genomes) by potential assignments did not correspond to the most dominant traits in transcribed assignments. For example, the capacity for microaerophilic and aerobic respiration was detected in the largest percentage of genomes (53% and 51%), but only expressed in 8% and 6% of genomes respectively. In contrast, anaerobic metabolisms like fermentation, iron reduction, and nitrogen reduction were each more expressed than aerobic respiration (**Fig. 2A-C)**. This expressed enrichment for anaerobic metabolisms is consistent with porewater dissolved oxygen concentrations below 1uM in most of these samples (***SI Appendix*, Table S4**). Metatranscriptomes across finely resolved depth gradients accentuates the fine-tuning by microorganisms adapted to highly fluctuating redox landscapes in wetland soils.

We examined the distribution and expression of metabolic traits at the genome level to delineate the metabolic spectrum of specialization to versatility (**Fig. 2D)**. Focusing on the extreme ends of this continuum, “active specialists” encoded 2 or fewer metabolic functionalities and expressed these traits. Of 1,136 trait classifiable MAGs, 40% were assigned to this group, with obligate fermenters (e.g., the archaeal Bathyarchaeia and bacterial QNDG01) and

207 methanogens (e.g., *Methanomethylicus*, *Methanothri*x, and *Methanoregula*) constituting most
208 functionalities in this group. On the other end of the spectrum, versatility was defined as having
209 the genomic capacity for 6 metabolic traits. These were defined as "active and versatile",
210 expressing at least 4 of their encoded traits. The 12 active versatile genomes all expressed genes
211 for microaerophilic metabolism and included members of Gammaproteobacteria (e.g.,
212 Burkholderiales and Chromatiales) and Desulfobacterota (e.g., Geobacterales). The final group
213 was "versatile with specialized expression", which despite encoding 6 or more diverse traits
214 expressed 2 or less across the sample set. This category included 85 MAGs composed of
215 Nitrospirota (UBA9217), Gammaproteobacteria (e.g., Burkholderiales and Chromatiales), and
216 Acidobacteriales. In summary, the identification of specialized and versatile expression
217 categories contributes to our growing understanding of the metabolic strategies across redox-
218 diverse spatiotemporal wetland samples.
219 We quantified how environmental gradients shaped attributes of the microbiome.
220 Multiple regression models determined the environmental predictors explaining microbial (i)
221 community composition (16S rRNA relative abundance), (ii) community compositional activity
222 (genome-resolved transcription), and (iii) metabolic trait activity (genome-resolved trait
223 transcription) (***SI Appendix*, Table S6**). These models were constructed using 17 environmental
224 factors composed of spatiotemporal (depth, ecotype, sampling month, sampling year) and
225 geochemical predictors (% clay, % silt, total organic carbon, acetate, cation exchange capacity
226 (CEC), conductivity, DO, Fe(II), Mg, nitrate, nitrite, pH, sulfate). Depth and related features
227 (CEC, Fe(II)), were the major environmental drivers across all three microbiome metrics. The
228 variation in metabolic trait activity was best explained by the environmental factors, whereas the
229 same variables were least explanatory for community composition (***SI Appendix*, Fig. S6)**. These
230 findings show that the measured environmental attributes play a more direct role in shaping
231 genome transcription and expressed traits, as opposed to composition and distribution measured
232 at the individual gene level. Our findings underscore that DNA based inferences, whether
233 derived from 16S rRNA or genome resolved, fall short in capturing the finely tuned microbial
234 responses to environmental conditions. This emphasizes the potential importance of
235 incorporating expression-based metrics for the development of more accurate and predictive
236 biogeochemical frameworks.
237
238 **Newly defined archaeal roles in plant litter decomposition**
239 Although environmental metagenomic studies are yielding increasingly large sets of
240 archaeal MAGs, there is still a dearth of information sampling archaeal metabolic activity and
241 translating these to biogeochemical outcomes in terrestrial systems. Here we used the gene
242 expression data recruited to each archaeal MAG to reconstruct a trophic structure for carbon
243 degradation. We assigned archaeal genomes to roles in (1) polymer degradation (2) sugar
244 fermentation, (3) short chain fatty acid and methylotrophic conversions, (4) methane metabolism,
245 and (5) other C1 metabolisms (methanotrophy, carbon dioxide fixation) **(*SI Appendix*, Table**
246 **S7)**. Our expression-based findings revealed Archaea were not latent members of a soil seedbank
247 (35), but instead genomes from 34 of 45 archaeal families were active in 75% of the
248 metatranscriptomes, many across all depth gradients (**Fig. 3A**).
249 Plant litter in soils serves as a crucial source of polymeric carbon, which must be
250 hydrolyzed by microbial enzymes to unlock the soil carbon cycle (19, 36). Here we show that
251 archaeal MAGs, including the Thermoplasmatota order DHVEG-1 and Thermoproteota
252 Bathyarchaeia members UBA233, expressed genes for utilizing hemicellulose, cellulose, and

253 polyphenols, the three major plant litter polymeric constituents. Additionally, members of the
254 family Thorarchaeaceae, including genera JACEL101 and a new OWC defined genus (OWC5),
255 also expressed multiple genes for decomposing cellulose and hemicellulose **(Fig. 3A)**. Archaea
256 are recognized for roles methane cycling, yet our findings broaden the scope of their influence to
257 encompass plant litter decomposition, aligning archaea with their bacterial and fungal
258 counterparts in carbon depolymerization (37, 38).
259 Microbial depolymerization of plant litter results in the generation of smaller monomeric
260 carbon compounds like sugars. From our metabolomes, raffinose, sucrose, trehalose, and two
261 features putatively annotated as maltose/cellobiose were detected in all 33 LC-MS derived
262 metabolomes (***SI Appendix*, Table S8-S9**). Thorarchaeaceae and UBA233 expressed genes for
263 utilizing 6 or more carbon substrates, in both polymeric and sugar trophic groups. These lineages
264 were key carbon cycling members responsible for converting complex plant material to more
265 readily usable carbon sources, contributing to soil respiration in the process (**Fig. 3A**). We also
266 identified sugar utilizing specialist archaea (members of the TCS64 in the Bathyarcahaeia and a
267 new OWC7 family within the DTKX01 of the Thermoplasmata) that expressed genes for
268 utilizing at least 3 sugar substrates. We consider that the anoxic conditions in these saturated
269 soils limit aerobic depolymerization by fungi and bacterial members, allowing for an expanded
270 role for the largely fermentative archaeal members (**Fig. 3A**).
271 Downstream of sugar fermentation, microbial carbon processing results in short chain
272 fatty acids (e.g., acetate, butyrate, lactate) and small methylated compounds (e.g., methylamine,
273 methanol). Compared to the higher trophic levels, methylotrophy and SCFA conversions were
274 the most represented across archaeal families (**Fig. 3**). In addition to the aforementioned
275 Thorarchaeaceae and TCS64, MAGs within the Thermoplasmatota, especially UBA10834 and
276 UBA472 families, had the greatest number of genes expressed within the archaea. These findings
277 add to a growing body of literature demonstrating, non-methanogen based methylotrophy is
278 prevalent across the archaeal tree of life (39). It is also likely that active members of this trophic
279 level generate the substrates to cross feed methanogens.
280 The one carbon metabolizing trophic level contains 85 unique methanogen genomes.
281 These were assigned to 10 families, including a new family (NewFamily01_m) in the
282 Methanomicrobiales and 15 new methanogen genera across diverse families (***SI Appendix*, Fig.**
283 **S7**). The new family contains 25 MAGs assigned to 4 genera, which all expressed the capacity
284 for hydrogenotrophic methanogenesis, while new genus03 also had members that could utilize
285 methylsulfonated compounds (***SI Appendix*, Table S10**). Beyond novelty, our detailed metabolic
286 reconstruction of wetland methanogens offered new insights into the methane cycle. First, we
287 found that *Methanothrix* (16 MAGs) and closely related members of the Methanotrichacea (3
288 MAGs) were the only acetoclastic methanogens genomically recovered from this wetland. We
289 posit that these obligate acetogens outcompeted other acetoclasts due to their competitive
290 substrate acquisition under the low acetate concentrations (<1 mM, mean 45 μM) (40) found in
291 these wetland soils (***SI Appendix*, Figure S8**). Understanding acetate metabolism in
292 methanogens and its application to biogeochemical models predominantly relies on
293 *Methanosarcina* as a model organism. Nevertheless, the significance of *Methanothrix* identified
294 here emphasizes the need for future physiological investigations to ensure enhanced realism in
295 predictive, process-oriented models of methane flux in wetland soils (41, 42).
296 Methanogen expression patterns across our transcriptome set also indicated the
297 importance of methylotrophic methanogenesis. Our findings confirm that the 23
298 Methanoregulaceae MAGs are all hydrogenotrophic, as expected from the defined physiology of

this family (43). We also show that two genera (3 MAGs) encoded the capacity for utilizing methylsulfonated compounds. Additionally, four new genera in the obligately methylotrophic families of Methanomassiliicoccales (new genus 8, 9, 10) and Methanomethylicales (new genus 12) expressed genes for utilizing methyl-oxygen, methyl-sulfur, and methylamine compounds. While it is often cited that wetland methane is derived exclusively from acetoclastic and hydrogenotrophic methanogens (5), the discovery of over 19 MAGs with the capacity for methylotrophy, and their high levels of transcriptional activity (***SI Appendix*, Fig. S9**), indicate that hydrogen-supported methylotrophy of nitrogen, sulfur, and even oxygenic compounds likely represents an underappreciated source of methane in wetlands (24, 44).

Taken together, our findings demonstrate the holistic, domain wide contributions to carbon cycling in these climatically important soils. We provide some of the first demonstrations of gene expression for many archaeal members cosmopolitan across human gut (45), aquatic (46, 47), and terrestrial (48) microbiomes. Additionally, likely due to both sequencing depth and targeted assembly strategies, our study uncovers far more methanogen genomes than prior metagenomic efforts. Prior to this effort, the largest wetland (permafrost peatland) MAG database contained 21 methanogen representatives with 4-fold less genus level diversity (19). Comparison of the methanogen membership across these two studies indicates shared dominant and active orders like Bog-38, Methanomassiliicoccales, Methanotrichales, but also genome-resolved endemism for each wetland (***SI Appendix*, Fig. S4)**. Our assignments of archaeal lineages to a trophic framework expanded archaeal roles in soil respiration to further sustain methane production.

Methane cycling guilds reside in depth resolved, temporally stable 'neighborhoods'

To ascertain how methane cycling expression changed across spatial-temporal gradients, we collected triplicate metatranscript data from 3 depths, monthly over a three-month period from a flooded mud-flat in 2018. Based on coordinated genome-wide expression patterns over time and space (***SI Appendix*, Fig. S9**), we show core and depth resolved colocalization of active methanogen and methanotroph genomes. First, six core methanogen genomes were active across all soil depths and timepoints, including representatives from a newly discovered OWC family (newfamily01_m) and a new Methanoregulaceae genera (***SI Appendix*, Fig. S9**). The remaining methane cycling members were localized to specific soil depths, forming co-expressing clades of organisms that persisted over the seasonal scale. Substrate utilization was not restricted to specific depths or times, but generally a greater proportional enrichment of acetoclastic and aerobic methanotrophs was observed surface soils, while middle and deepest depths included the anaerobic methane oxidizing *Methanoperedens* genome and proportionally more methyl-utilizing methanogens. These metatranscript findings suggest that wetland soils harbor depth-stratified and persistent methane cycling 'neighborhoods', comprising diverse strains with functional overlap.

Subsequently, we scrutinized the transcriptional activity of the methane community over a three-year timescale, where our sampling encompassed changes to the physiochemical environment. We compared triplicate core metatranscriptome samples from both surface (0-5cm) and deep (25-30 cm) layers at the same mud-flat site, from August from 2018 and August 2015. Within this sampling period the wetland experienced a rapid shift in water elevation, where water levels increased by 1.6 m in 2018, flooding the previously exposed mud flat from 2015. We expected that flooding would result in different microbial neighborhood compositions due to altered pore water oxygen concentrations. We also expected this flood restructuring microbiome

response would be most strongly observed in the surface soils, rather than the deep soils, which had more consistent saturation over time.

We observed pore water oxygen concentrations that matched our flooding expectations, with only detectable oxygen concentrations in unflooded surface soils in 2015 (**Fig. 3B**). In contrast to our expectations, most methanogen MAGs (77/85) had stable transcription patterns that were not affected by flooding (**Fig. 3B**). This implies that the microbial neighborhoods that existed on a seasonal scale (***SI Appendix*, Fig. S9)** endured across years and redox landscapes. However, one of the most active MAGs in 2018, an obligate methylotrophic methanogen identified as *Methanomethylicus*, was not detected under non-flooded conditions in 2015. This observation implies some of level of reorganization or sorting of the methanogen community during this particular time period (**Fig. 3B**).

Contrary to our initial expectations, methane cycling genomes in the deeper soils displayed higher transcriptional responsiveness to flooding (**Fig. 3B**). This suggests a reduced temporal stability in the deeper methane cycling communities, which could reflect the diminished disturbance legacy of these soils compared to surface soils which are more accustomed to wet-dry fluctuations and thus more resistant to these perturbations (49). Across the functional types, acetoclastic methanogens were the most stable in both surface and deep soils over time, while somewhat surprisingly, oxygen-requiring methanotrophs had a statistically significant increase in the surface soils with flooding (**Fig. 3B**). This phenomenon could be explained by the strong oxygen affinity of these Gammaproteobacterial Methylococcacea members with methanotrophy occurring at sub μM levels (50). Therefore, our findings show methanotrophy occurs below the detection limit of commonly used oxygen probes used in wetland soils. This has important implications for process-based models (42) that rely on measurable oxygen as a constraint on methanotrophy, as these models may substantially underestimate the prevalence of this biological methane filtering process.

In a broader context, the stability observed over this multi-year period suggests that expression of methane cycling processes is less affected than anticipated by soil redox changes induced by flooding. This finding is in alignment with the fact that surface methanogens may already be insulated from flooding impacts through anoxic microsites, such as biofilms (23, 51). Further supporting the notion that carbon decomposition was not altered in these mudflat samples in response to flooding, metabolomics revealed methanogen substrate concentrations were not significantly changed with flooding in both surface and deep soil compartments (e.g. acetate, formate, methylamines, methanol) (***SI Appendix*, Fig. S8**). Further supporting this notion, a community wide stability assessment showed that over all microbial community transcription was more stable across years, and less stable across month or site level (***SI Appendix*, Table S11**). These findings support a growing body of research indicating that oxygen and other electron acceptor concentrations may not serve as the primary regulators of methane production or consumption activities in wetland soils, as previously proposed (23, 51). Consequently, it may be prudent to reconsider the widely accepted thermodynamic hierarchy theory, foundational to many biogeochemical models used in modeling wetland methane emissions.

Co-expression networks reveal key lineages modulating soil greenhouse gas production

Given the stability of the microbiome observed over time and space, we next sought to identify genomes with shared co-expression patterns that could inform soil greenhouse gas concentrations (***SI Appendix*, Table S4**). We analyzed 133 paired metatranscriptomes with key

environmental parameters to reveal two subnetworks that predicted porewater CH_4 and CO_2 concentrations. Specifically, one subnetwork (turquoise) was surface associated and negatively correlated with CH_4 and CO_2 concentrations, while the other subnetwork (brown) was deep associated and positively correlated to greenhouse gas concentrations (***SI Appendix*, Fig. S10, Table S12-S13**). The surface and deep subnetworks contained 551 and 372 active co-expressing genomes respectively. Transcriptionally active genomes in each subnetwork with the highest importance for methane prediction were highlighted (**Fig. 4AB**).

Based on the different membership and functional assignment of the co-expressing lineages in these subnetworks, we unraveled the process configurations that contribute to the methane cycle in these depth-defined zones. By graphing the carbon trophic levels in context of the energy metabolisms we visualized the expressed metabolic network structures in each of the compartments (**Fig. 4C**). While numerous functional roles remained consistent across depths, distinct organismal representatives occupied these roles, highlighting the functional equivalence encoded and expressed within wetland soils. For example, microaerophilic respiratory organisms were responsible for accessing polymeric carbon, catalyzed by Bacteroidales in the surface soils, but Anaerolineales and Sedimentisphaerales in the deep soils. Carbon decomposition in surface soils oxidized by oxygen and nitrate respiration, while iron respiration was predominant in deeper soils. The aerobic methanotrophs were active across surface soils (22), perhaps explaining the negative relationship between members of this surface derived subnetwork and methane concentrations. Notably with the methanogens, acetotrophs (*Methanothrix*), hydrogenotrophs (UBA9949 and new genus 13), and methylotrophs (Methylococcaceae UBA6136 and Methylomonadaceae KS41) were active in the surface network, while only hydrogenotrophs (new family 01, *Methanoregula*) were active in the deep subnetwork, indicating greater functional diversity conserved in surface soil methanogen communities.

Collectively, these observations paint a picture of population-level co-expression gene networks that contribute to greenhouse gas concentrations in wetland soils. Our study reveals that diverse microbial organisms execute similar carbon decomposing functions even in the presence of varying redox conditions. These intricate networks likely bestow ecosystem resilience in the face of environmental perturbations. Although we have observed the existence of these co-expressing assemblages across multiple years, sites, and conditions, the lingering question is whether a critical event or tipping point could prompt a reconfiguration of the metabolic structures and what that consequence would be for greenhouse gas emissions. Genome scale metabolic models based on these network architectures will offer opportunity for exploring the fortitude of this wiring to environmental perturbation common in the Anthropocene. Our findings illuminate the underestimated complexity of soil microbiomes, revealing the shared and distinct metabolic pathways in wetland soils that ultimately shape methane production.

Conclusion

In summary, this comprehensive investigation of the Old Woman Creek National Estuarine Research Reserve has significantly advanced understanding of the microbial communities orchestrating crucial biogeochemical cycles in a high-methane emitting temperate wetland. The establishment of the Multi-omics for Understanding Climate Change (MUCC), with its wealth of metagenome-assembled genomes, metatranscriptomes, and associated geochemical and emission datasets, provides an invaluable resource for unraveling the complexities of microbial activities across spatiotemporal scales in soils. Noteworthy revelations include the previously unrecognized roles of archaeal members in carbon decomposition, the

diverse substrate utilization patterns and resistance to redox changes exhibited by methanogens, and the delineation of integrated metabolic networks contributing to methane production. Moreover, our standardized sampling and sequencing design set the stage for future cross-site comparisons, enhancing the ability to generalize findings across temperate wetlands. These data underscore the necessity of integrating gene expression measurements with detailed genomic databases for a more profound comprehension of the intricate metabolic circuitry governing wetland methane. Through this research, we lay the groundwork for ongoing exploration of soil microbiomes impacting climate change, which play a crucial role in shaping the functioning and resilience of temperate wetland ecosystems amidst environmental change.

References

1. W. J. Mitsch, J. G. Gosselink, L. Zhang, C. J. Anderson, *Wetland Ecosystems* (John Wiley & Sons, 2009).

2. J. A. Villa, B. Bernal, Carbon sequestration in wetlands, from science to practice: An overview of the biogeochemical process, measurement methods, and policy framework. *Ecological Engineering* **114**, 115–128 (2018).

3. S. C. Neubauer, J. P. Megonigal, Moving Beyond Global Warming Potentials to Quantify the Climatic Role of Ecosystems. *Ecosystems* **18**, 1000–1013 (2015).

4. R. B. Jackson, *et al.*, Increasing anthropogenic methane emissions arise equally from agricultural and fossil fuel sources. *Environ. Res. Lett.* **15**, 071002 (2020).

5. S. D. Bridgham, H. Cadillo-Quiroz, J. K. Keller, Q. Zhuang, Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. *Glob Chang Biol* **19**, 1325–1346 (2013).

6. K. Calvin, *et al.*, "IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.," First (Intergovernmental Panel on Climate Change (IPCC), 2023) https:/doi.org/10.59327/IPCC/AR6-9789291691647 (January 19, 2024).

7. M. Saunois, *et al.*, The Global Methane Budget 2000–2017. *Earth System Science Data* **12**, 1561–1623 (2020).

8. R. Conrad, Importance of hydrogenotrophic, aceticlastic and methylotrophic methanogenesis for methane production in terrestrial, aquatic and other anoxic environments: A mini review. *Pedosphere* **30**, 25–39 (2020).

9. W. J. Riley, *et al.*, Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in CESM. *Biogeosciences* **8**, 1925–1953 (2011).

10. B. Kieft, *et al.*, Genome-resolved correlation mapping links microbial community structure to metabolic interactions driving methane production from wastewater. *Nat Commun* **14**, 5380 (2023).

11. L. M. Solden, *et al.*, Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat Microbiol* **3**, 1274–1284 (2018).

12. M. A. Borton, *et al.*, Coupled laboratory and field investigations resolve microbial interactions that underpin persistence in hydraulically fractured shales. *Proceedings of the National Academy of Sciences* **115**, E6585–E6594 (2018).

13. P. Bodelier, S. Dedysh, Microbiology of wetlands. *Frontiers in Microbiology* **4** (2013).

14. A. B. Narrowe, *et al.*, High-resolution sequencing reveals unexplored archaeal diversity in freshwater wetland soils. *Environ Microbiol* **19**, 2192–2209 (2017).

15. S. He, *et al.*, Patterns in Wetland Microbial Community Composition and Functional Gene Repertoire Associated with Methane Emissions. *mBio* **6**, 10.1128/mbio.00066-15 (2015).

16. G. Webster, *et al.*, Archaeal community diversity and abundance changes along a natural salinity gradient in estuarine sediments. *FEMS Microbiol Ecol* **91**, 1–18 (2015).

17. S. T. Bates, *et al.*, Examining the global distribution of dominant archaeal populations in soil. *ISME J* **5**, 908–917 (2011).

18. N. Taş, *et al.*, Landscape topography structures the soil microbiome in arctic polygonal tundra. *Nat Commun* **9**, 777 (2018).

19. B. J. Woodcroft, *et al.*, Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).

20. J. Hultman, *et al.*, Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **521**, 208–212 (2015).

21. R. M. Wilson, *et al.*, Soil metabolome response to whole-ecosystem warming at the Spruce and Peatland Responses under Changing Environments experiment. *Proceedings of the National Academy of Sciences* **118**, e2004192118 (2021).

22. G. J. Smith, *et al.*, Members of the Genus Methylobacter Are Inferred To Account for the Majority of Aerobic Methane Oxidation in Oxic Soils from a Freshwater Wetland. *mBio* **9**, e00815-18.

23. J. C. Angle, *et al.*, Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat Commun* **8**, 1567 (2017).

24. A. B. Narrowe, *et al.*, Uncovering the Diversity and Activity of Methylotrophic Methanogens in Freshwater Wetland Soils. *mSystems* **4**, e00320-19 (2019).

25. M. Bahram, *et al.*, Structure and function of the soil microbiome underlying N2O emissions from global wetlands. *Nat Commun* **13**, 1430 (2022).

26. G. Bohrer, *et al.*, Methane and CO2 chamber fluxes and porewater concentrations US-OWC Ameriflux wetland site, 2015-2018 (2019) https:/doi.org/10.15485/1568865 (December 18, 2023).

27. G. Bohrer, J. Kerns, AmeriFlux BASE US-OWC Old Woman Creek (2023) https:/doi.org/10.17190/AMF/1418679 (January 19, 2024).

28. H. Chu, *et al.*, AmeriFlux BASE data pipeline to support network growth and data sharing. *Sci Data* **10**, 614 (2023).

29. G. Pastorello, *et al.*, The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci Data* **7**, 225 (2020).

30. K. B. Delwiche, *et al.*, FLUXNET-CH4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. *Earth System Science Data* **13**, 3607–3689 (2021).

31. P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).

32. K. Anantharaman, *et al.*, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219 (2016).

33. A. Almeida, *et al.*, A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* **39**, 105–114 (2021).

34. R. D. Stewart, *et al.*, Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol* **37**, 953–961 (2019).

35. S. M. Gibbons, *et al.*, Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Sciences* **110**, 4651–4655 (2013).

36. B. B. McGivern, *et al.*, Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nat Commun* **12**, 2466 (2021).

37. F. Barbi, *et al.*, Fungal ecological strategies reflected in gene transcription - a case study of two litter decomposers. *Environ Microbiol* **22**, 1089–1103 (2020).

38. L. Žifčáková, P. Baldrian, Fungal polysaccharide monooxygenases: New players in the decomposition of cellulose. *Fungal Ecology* **5**, 481–489 (2012).

39. V. De Anda, *et al.*, Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. *Nat Commun* **12**, 2404 (2021).

40. M. S. M. Jetten, A. J. M. Stams, A. J. B. Zehnder, Methanogenesis from acetate: a comparison of the acetate metabolism in Methanothrix soehngenii and Methanosarcina spp. *FEMS Microbiology Letters* **88**, 181–197 (1992).

41. R. F. Grant, Simulation of methanogenesis in the mathematical model ecosys. *Soil Biology and Biochemistry* **30**, 883–896 (1998).

42. R. F. Grant, *et al.*, Mathematical Modelling of Arctic Polygonal Tundra with Ecosys: 1. Microtopography Determines How Active Layer Depths Respond to Changes in Temperature and Precipitation. *Journal of Geophysical Research: Biogeosciences* **122**, 3161–3173 (2017).

43. S. Sakai, *et al.*, Methanolinea mesophila sp. nov., a hydrogenotrophic methanogen isolated from rice field soil, and proposal of the archaeal family Methanoregulaceae fam. nov. within

the order Methanomicrobiales. *International Journal of Systematic and Evolutionary Microbiology* **62**, 1389–1395 (2012).

44. J. B. Ellenbogen, *et al.*, Methylotrophy in the Mire: direct and indirect routes for methane production in thawing permafrost. *mSystems* **0**, e00698-23 (2023).

45. C. M. Chibani, *et al.*, A catalogue of 1,167 genomes from the human gut archaeome. *Nat Microbiol* **7**, 48–61 (2022).

46. B. J. Baker, K. E. Appler, X. Gong, New Microbial Biodiversity in Marine Sediments. *Annual Review of Marine Science* **13**, 161–175 (2021).

47. J. A. Rodríguez-Ramos, *et al.*, Genome-Resolved Metaproteomics Decodes the Microbial and Viral Contributions to Coupled Carbon and Nitrogen Cycling in River Sediments. *mSystems* **7**, e00516-22 (2022).

48. B. J. Baker, *et al.*, Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat Microbiol* **1**, 1–9 (2016).

49. M. Doering, *et al.*, Microbial communities in floodplain ecosystems in relation to altered flow regimes and experimental flooding. *Science of The Total Environment* **788**, 147497 (2021).

50. Q. Hao, *et al.*, Methylobacter couples methane oxidation and N2O production in hypoxic wetland soil. *Soil Biology and Biochemistry* **175**, 108863 (2022).

51. M. J. Bogard, *et al.*, Oxic water column methanogenesis as a major component of aquatic CH4 fluxes. *Nat Commun* **5**, 5350 (2014).

52. E. Hassett, *et al.*, Carbon flux measurements from chambers collected between July to October 2022 at Old Woman Creek, Huron, Ohio. (2023) https:/doi.org/10.15485/2229438 (January 19, 2024).

53. J. Villa, G. Bohrer, L. Kinsman-Costello, Y. Ju, Nutrient and carbon concentrations in dated soil cores at US-OWC Ameriflux wetland site (OWC NERR) (2022) https:/doi.org/10.15485/1839288 (January 19, 2024).

Figure 1. Freshwater wetlands host thousands of novel, taxonomically diverse, and transcriptionally active microbial lineages. (A) Old Woman Creek (OWC) is located on the southern part of Lake Erie and is the highest US methane emitting wetland in the AmeriFlux Network. A site map of the OWC freshwater wetland shows where the 700 samples were collected from 10 major sampling sites that were historically defined as emergent vegetation, either *Nelumbo* (N) or *Typha* (T), mud cover (M), or location in open water channel (OW). Sites were sampled over multiple years and depths at 5 cm intervals from 0-30 cm, with sampling density proportional to circle size, 52 is the minimum sampling per location (***SI Appendix*, Fig. S1**). (B) The taxonomic distribution of the 2,502 MAGs summarized at the phylum level. Bold phylum label indicates a candidate phylum, and a blue text indicates the phylum contains a top 10 dominant genus. Each point represents a MAG and is colored by taxonomic novelty. Points are distributed by their log10 maximum transcription across the 133 metatranscriptome samples. MAGs to the right of the black line were transcriptionally active (N = 1,948), and MAGs to the right of the red line were most active (log10 max. transcription > 1.5) The gray bar chart to the right summarizes the number of genomes per phylum recovered, and red points overlaid represent the number of genera per phylum, using the same x-axis. (C) The most dominant archaeal and bacterial MAGs (top 15 for each domain) grouped by genus are assessed by prevalence (gray bar, the number (N) of samples in which MAGs from the genera were detected, bottom x-axis) and maximum transcription at any time point or depth (blue point for each MAG, top x-axis).

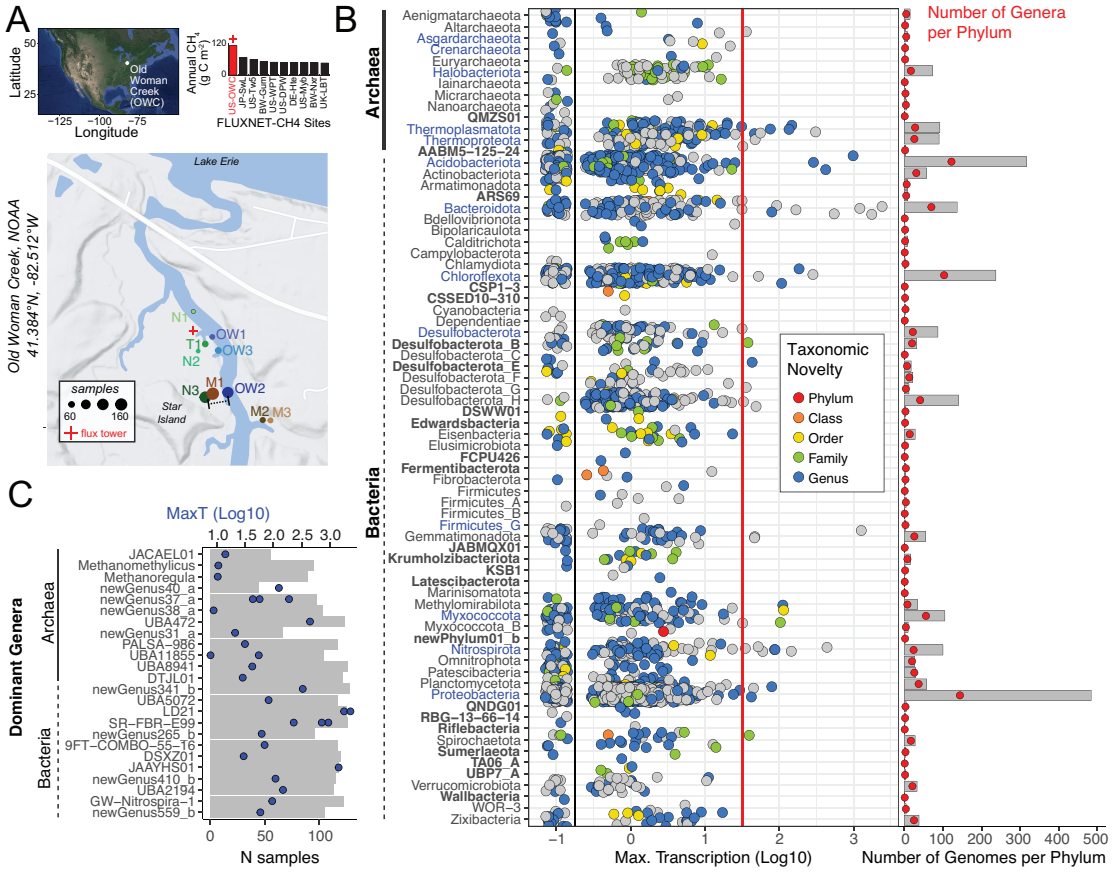


Figure 2. Microbial genomes from freshwater wetlands are phylogenetically and metabolically diverse. (A) The phylogenetic distribution of 2,204 bacterial MAGs and their corresponding potential (metagenome) and expressed (metatranscriptome) metabolic diversity. The phylogenetic tree is constructed from the 2,204 bacterial MAGs based on a subset of 30 genes from the PhyEco database. Nodes are colored by phylum level taxonomic assignment. Rings represent metabolic diversity at the trait level for each MAG. For each color, the inner ring represents potential (metagenome) metabolism, and the outer ring represents the expressed (metatranscriptome) metabolism. Bar plots summarize the genome counts across functional traits for potential (B) and expressed (C), respectively (n = 1,948 and includes bacteria and archaeal genomes). (D) The distribution of potential and expressed traits are shown by a scatterplot where each point is colored by phylum and represents a single MAG. Position on plot defines metabolic specialization (potential < 2 traits) or versatility (potential > 6 traits) as shown by arrows at the top. Gray boxes highlight extreme ends of each spectrum where 1) active specialized encoded 2 or less traits and had an expressed to potential ratio of 1, 2) active versatile encoded 6 or more traits and expressed more 4 or more traits, and 3) versatile with specialized expression encoded 6 or more traits and expressed 1 to 2 traits. Pie charts below summarize each 1-3 category and are colored by phylum.

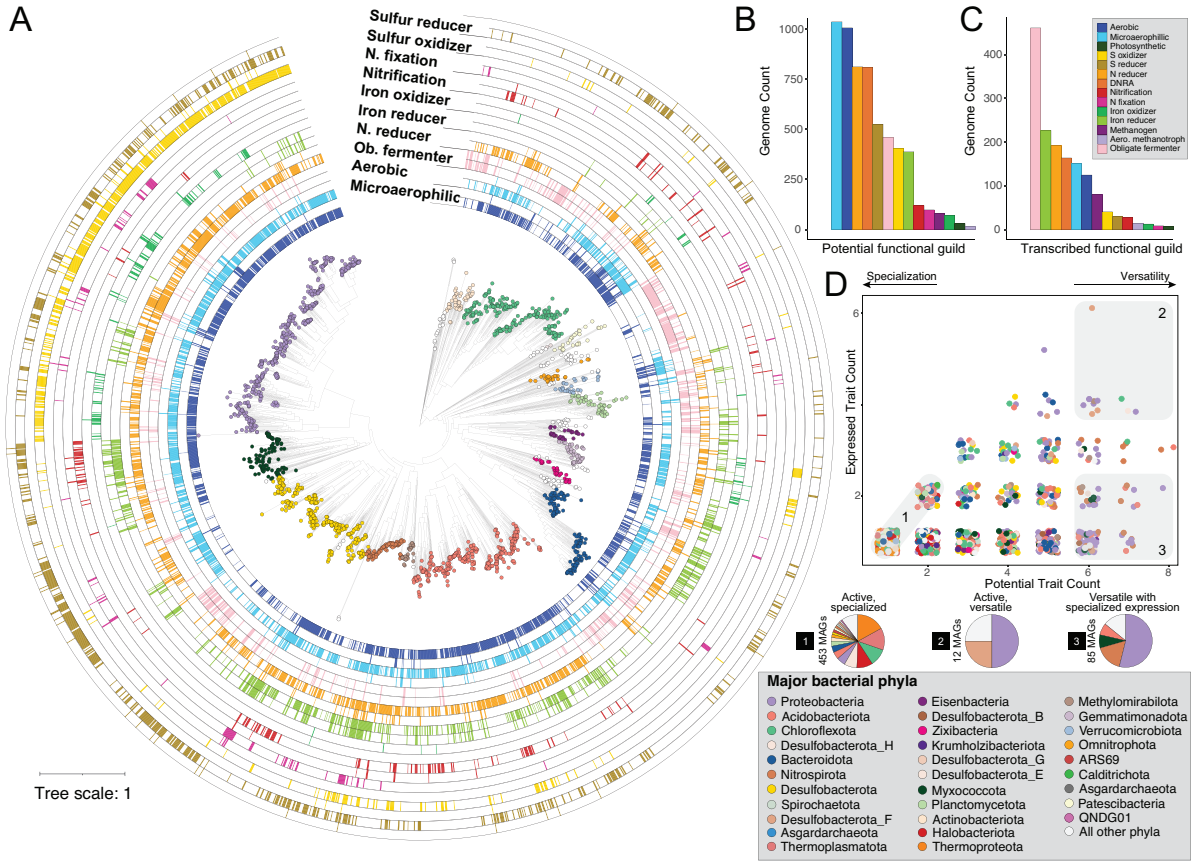


Figure 3. The metabolic diversity, expression, and temporal stability of recovered Archaeal MAGs. (**A**) (**Top right**) The schematic on the right visualizes the major trophic levels contributing to soil carbon decomposition including: polymers (polyphenolics, mannan, xyloglucan, xylan, cellulose, starch, pectin); oligomers and sugars (sucrose, lactose, mannose, galactose, fucose, xylose); organic acids, nitrogen, and alcohols (fermentation, methylotrophy, anaerobic alkane oxidation); methanogenesis (acetoclastic, methylotrophic, and hydrogenotrophic); and other 1 carbon metabolisms (anaerobic methane oxidation, CO_2 fixation). (**Top left**) For each archaeal family the bubble plot denotes gene expression (green), presence (black), or absence (white) of major carbon utilization pathways assigned to the trophic levels. (**Bottom**) The bar chart summarizes the number of transcriptionally active MAGs recovered within each archaeal phylum and family. (**B**) Dissolved porewater oxygen depth profiles from August 2015 and 2018 highlight the flooding of Old Woman Creek. Scatterplot shows the change of mean MAG transcription from 2015 to 2018 from the methane cycling community within surface (0-5 centimeters) and deep (20-30 centimeters) soils. Bubbles represent a single methane cycling MAG and are colored by metabolism and sized by average transcription. MAGs within the red dashed lines do not have log2 fold change in mean transcription and are considered temporally stable.

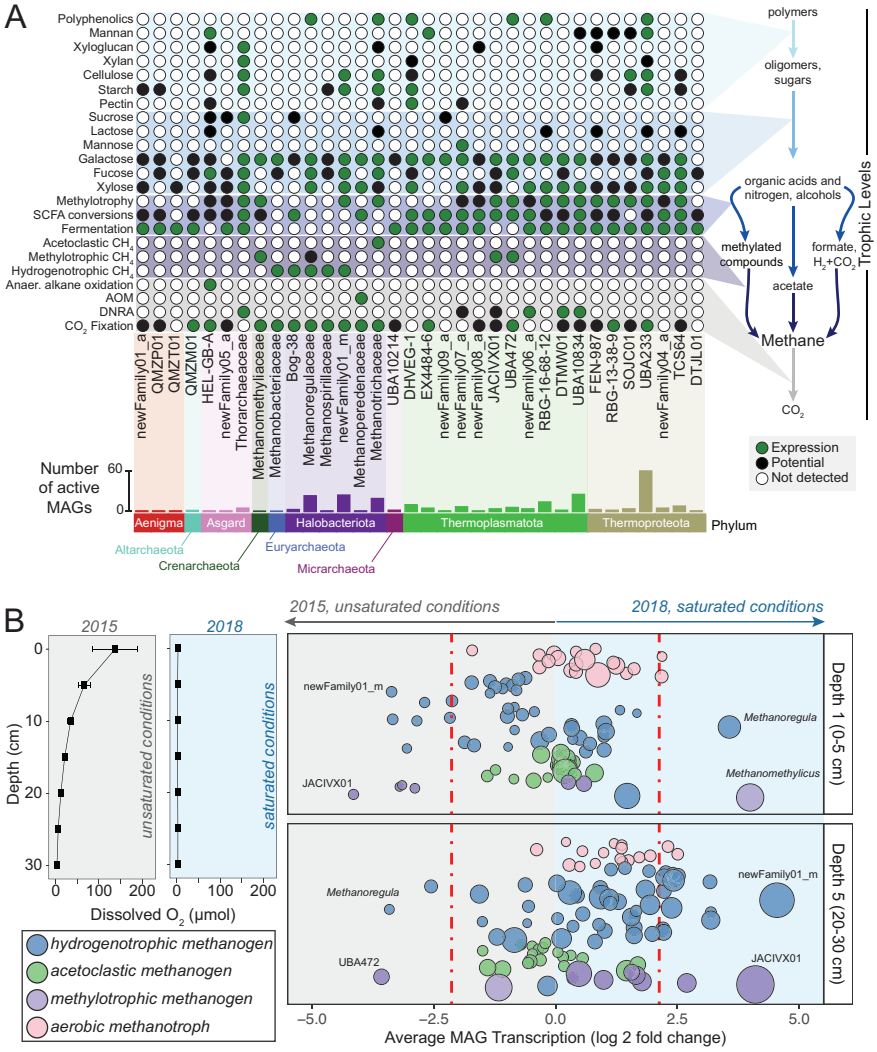
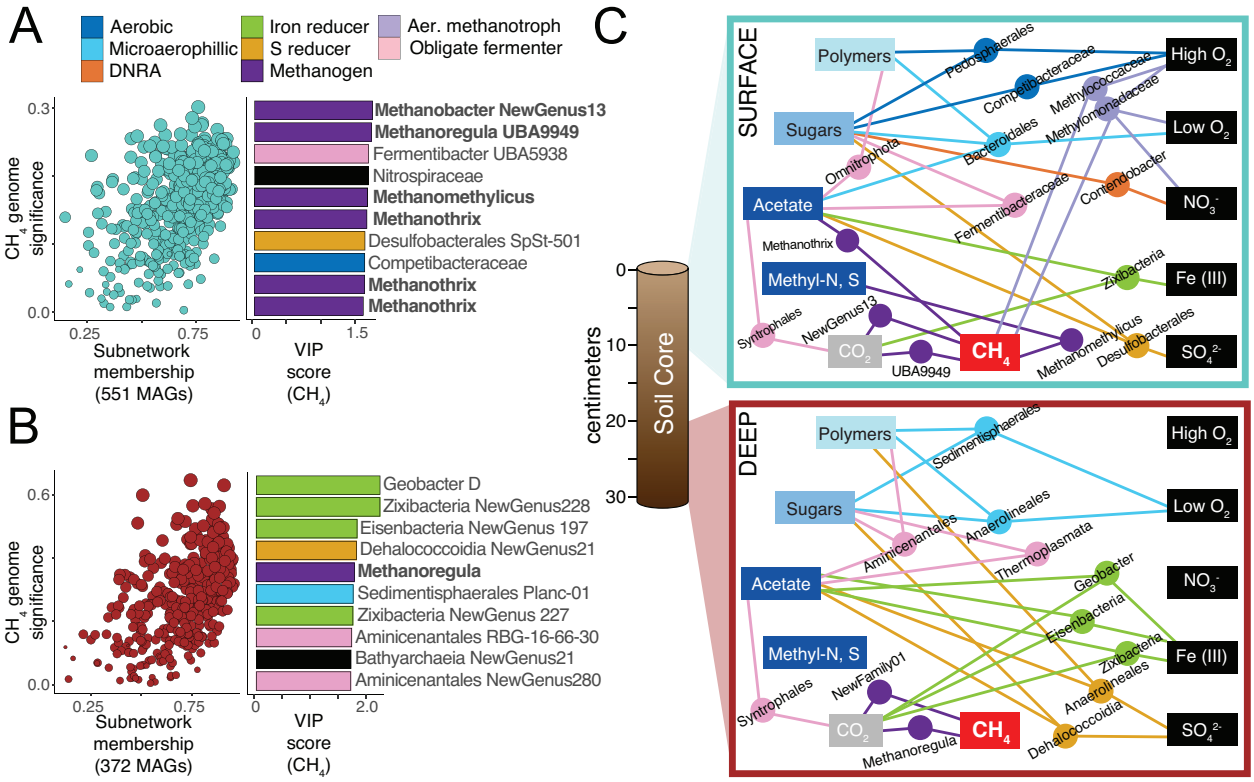


Figure 4. Coordinated MAG-resolved gene expression networks inform wetland carbon greenhouse gas concentrations in surface and deep soils. (A) The genome significance plot and highest variable important genomes related to CH_4 concentrations in the turquoise, surface derived subnetwork (n = 551 MAGs, Rho = 0.53, P < 0.001). (B) The genome significance plot and highest variable important genomes related to CH_4 concentrations from the brown, deep subnetwork (n = 372 MAGs, Rho = 0.58, P < 0.001). For panels A and B, the bar charts (middle) highlight the relative contribution of the top 10 variably important genomes (colored by metabolic classification). Methanogens are highlighted with bold text. (C) A process-based schematic based on genome-resolved expression data of the highest variably important taxa and their contribution to methane production in the surface subnetwork from A (top) and the deep subnetwork from B (bottom).



Supporting Information Text: SI Methods

SI Appendix Methods

Field campaign and sample collection. We extensively sampled the main wetland ecological sites (ecosites), or ecohydrological patch types, across two major field campaigns to capture the patch-level spatiotemporal heterogeneity. The patches included wetland plant cover (plant), predominantly *Typha latifolia*, exposed mud -flat (mud), and standing freshwater (open water). Our sampling was also depth-resolved every 5 cm from 0-25 cm, resulting in a total of 705 soils included in this study (**Fig. S1A**).

Sampling was conducted at the Old Woman Creek National Estuarine Research Reserve (OWC). OWC is a freshwater wetland located on the southern shore of Lake Erie, OH, USA (41˚22'N, 82˚30'W). Samples were collected during field campaigns in 2013, 2014, 2015 and 2018. Sample collection details for the 2013-2014 and 2014-2015 campaigns were previously described (1–3). Consistent collection methods were employed for the 2018 campaign, which is first reported in this publication. Briefly, in 2018 we selected 10 sampling sites that include locations sampled during prior years and represent a range of current and historic patch types of Typha (n=1), Nelumbo (n=3), Open Water Channel (n=3), and Temporal Mud flat (n=3) (**Fig. 1A**). At each location, cores were collected within a meter of each of the 10 installed pore-water dialysis peepers (n=1 core for May, June, and n=3 cores for July, August, and September, during peak methane production at the wetland). Prior to core collection, we measured soil dissolved oxygen (DO) and temperature within 5 cm increments from the soil surface using the Fibox 4 meter and DP-Pst3 probe (PreSens Precision Sensing, Germany). Cores were hydraulically extruded and immediately divided into 5 cm depth sections from 0-25 cm, as measured from the core surface. Each 5 cm section was further subsampled for molecular and geochemical analyses: subsamples were flash frozen in the field and then shipped on dry ice for storage at -80˚C prior to RNA extraction.

Geochemical analysis and collection of gasses. Concentrations of soil nitrate, nitrite, acetate, and sulfate were measured using ion chromatography as previously described (1). Briefly, 1:1 soil slurries were made using 5g of soil with 5 ml MilliQ water. Following pH measurement, slurries were filtered using a 0.2µM filter, with the filtrate quantified on a Dionex ICS-2100 Ion Chromatography System with an AS18 column. Using soil slurry filtrate, concentrations were measured for: formate, methanol, acetate, at the Pacific Northwest National Laboratory using proton magnetic resonance (^{1}H NMR) as previously described (4, 5). Fe^{2+} was measured in the field using Hach FerroVer packets per the manufacturer's directions (Hach #927-9). Surface emissions of CH_4 and CO_2 were measured by non-steady-state chambers in duplicate in each ecosite and season and in sync with peeper measurement times (1). We also measured *in situ* concentrations of dissolved CH_4 and CO_2 belowground monthly via porewater dialysis samplers (1). In brief, peepers with a vertical resolution of 2.8 cm measured concentrations via a 0.1 µm membrane, allowing water inside the peeper windows to equilibrate with dissolved gas concentrations, with minimal soil disturbance. Gas concentrations from peepers were quantified with a Shimadzu GC-2014 gas analyzer as previously described (1).

Soil metabolomics methods. Water-soluble metabolites were extracted from soil by adding 7 mL of autoclaved Milli-Q water to 1 g of wet soil in a sterile 15-mL centrifuge tube. To identify field-present metabolites in Old Woman Creek, including methanogenic substrates, we performed ^{1}H NMR on aliquots water extractions. Samples (180 µL) were combined with 2,2-dimethyl-2-silapentane-5-sulfonate-d_6 (DSS-d_6) in D_2O (20 µL, 5 mM) and thoroughly mixed prior to transfer to 3 mm NMR tubes. NMR spectra were acquired on a Bruker Avance III spectrometer operating at a field strength of 17.6 T (^{1}H ν_0 of 750.24 MHz) and equipped with a 5mm Bruker TCI/CP HCN (inverse) cryoprobe with Z-gradient and at a regulated temperature of 298 K. The 90° ^{1}H pulse was calibrated prior to the measurement of each sample. The one-dimensional ^{1}H spectra were acquired using a nuclear Overhauser effect spectroscopy (noesypr1d) pulse sequence with a spectral width of 12 ppm and 1024 transients. The NOESY mixing time was 100ms and the acquisition time was 4s followed by a relaxation delay of 1.5 s during which presaturation of the water signal was applied. Time domain free induction decays (72114 total points) were zero filled to 131072 total points prior to Fourier transform. Chemical shifts were referenced to the ^{1}H methyl signal in DSS-d_6 at 0 ppm. The 1D ^{1}H spectra were manually processed, assigned metabolite identifications and quantified using Chenomx NMR Suite 8.6. Metabolite identification was based on matching the chemical shift, J-coupling and intensity of experimental signals to compound signals in the Chenomx and custom in-house databases (***SI Appendix,* Fig. S8, Table S8**). Quantification was based on fitted metabolite signals relative to the internal standard (DSS-d_6). Signal to noise ratios (S/N) were measured using MestReNova 14 with the limit of quantification equal to a S/N of 10 and the limit of detection equal to a S/N of 3.

In preparation for LC-MS, water extracts were frozen and lyophilized dry (FreeZone 2.5 Plus, Labconco). Each extract was resuspended in 170 µL of 100% methanol containing isotopically labeled internal standards (5-50 µM of 13C,15N Cell Free Amino Acid Mixture, #767964, Sigma; 1 ug/mL 2-amino-3-bromo-5-methylbenzoic acid, ABMBA, #R435902, Sigma) and centrifuge-filtered (0.22 µm hydrophilic PVDF membrane, #UFC30GV00, Millipore) before transferring to glass LC-MS autosampler vials. To detect polar metabolites using LC-MS/MS, normal phase chromatography was performed using a QExactive HF Orbitrap MS (Thermo Scientific, San Jose, CA) in line with an Agilent 1290 LC stack. Chromatography was performed using a HILIC column (Agilent InfinityLab Poroshell 120 HILIC-Z, 2.1 x 150 mm, 2.7 µm, #673775-924) held at 40 ºC and mobile phase running at a flow rate of 0.45 mL/minute with 2 µL injection volumes for each sample. The column was first equilibrated with 100% buffer B (95:5 ACN:H2O with 5 mM ammonium acetate) for 1 minute, then diluting this down to 89% with buffer A (100% H2O with 5 mM ammonium acetate and 5 µM methylenediphosphonic acid) over 10 minutes, then down to 70% B over 4.75 minutes, then down to 20% B over 0.5 minutes, followed by isocratic elution in 80% buffer A for 2.25 minutes followed by re-equilibrating the column at 100% B prior to the next injection. Mass spectrometry data was collected in centroid mode in both positive and negative polarity, and included full MS spectra from m/z 70-1050 at 60,000 resolution as well as MS2 fragmentation spectra acquired using stepped then averaged 10, 20 and 40 eV collision energies at 17,500 resolution. Source settings included spray voltage of 3kV for both positive and negative ionization, sheath gas flow of 55 (au), auxiliary gas flow of 20 (au), sweep gas flow of 2 (au), and capillary temperature of 400 degrees C. Sample injection order was randomized, with an injection blank of 100% methanol run between each sample, and replaced by an internal standard

741 or quality control mixture only every 3 to 15 samples. Metabolomics data is available through
742 MassIVE under accession MSV000093935.
743
744 Metabolites were identified by comparing exact mass, retention time (RT) and MS/MS
745 fragmentation spectra to that of compound standards run using the same chromatography and
746 mass spectrometry methods. LC-MS spectra were analyzed using custom Python code (Yao et
747 al., 2015), with a score of 0 to 3 assigned to each feature (unique m/z and RT pair) that
748 represented the level of confidence in identifying the metabolite. A metabolite positively
749 identified had a detected m/z of ≤ 5 ppm or 0.001 Da from theoretical and RT ≤ 0.5 minutes
750 compared to that of a pure standard run using the same LC-MS conditions. Highest level of
751 positive identification (score of 3) was given for features having matching fragmentation spectra
752 to an internal database of standards run and collected on a Q Exactive Orbitrap MS, with
753 mismatching spectra invalidating a positive identification.
754
755 **DNA and RNA extraction and 16S rRNA gene amplicon sequencing.** DNA was extracted
756 from soil samples using the PowerSoil DNA kit (Qiagen). Total RNA was extracted using the
757 Takara NucleoBond RNA Soil kit (Takara Bio USA, CA) with 5g of soil as input. Extracted
758 RNA was purified and concentrated using the Zymo RNA Clean & Concentrator Kit (Zymo
759 Research) and quantified with the Qubit RNA HS Assay (Thermo Fisher Scientific). For low
760 biomass samples, up to 5 RNA extractions of 5g were pooled prior to concentration. RNA purity
761 and DNA removal were verified by PCR using universal V4 region 16S rRNA gene primers. 16S
762 rRNA genes in extracted DNA were amplified with the primers 515F/806R (6) and sequenced at
763 Argonne National Laboratory on the Illumina MiSeq with 251-bp paired end read. Read sets
764 (n=701) were processed with QIIME2 (release 2018.11, (7)) and ASVs were recovered with
765 DADA2 (8). Taxonomies were assigned with the q2-feature-classifier with the SILVA database
766 (release 132) (9). ASVs were only included in downstream analyses if they were detected in at
767 least 10% of samples. We set a minimum read threshold of at least 5,000 sequences per sample,
768 retaining 672 samples, and then normalized our ASV table with cumulative sum scaling prior to
769 downstream analyses (**Table S2**).
770
771 **Metagenomic and metatranscriptomic sequencing.** Of the 42 Metagenomes, 25 are newly
772 reported data here (**Table S1**). This includes 16 samples from the 2018 field sampling
773 representing a range of patch types, soil depths and time points to be used in building a MAG
774 database. For the 2018 field samples sequencing was performed by the Department of Energy
775 Joint Genome Institute following their established protocols. Briefly, plate-based DNA library
776 preparation for Illumina sequencing was performed on the PerkinElmer Sciclone NGS robotic
777 liquid handling system using Kapa Biosystems library preparation kit. 200.20 ng of sample DNA
778 was sheared to 472 bp using a Covaris LE220 focused-ultrasonicator. The sheared DNA
779 fragments were size selected by double-SPRI and then the selected fragments were end-repaired,
780 A-tailed, and ligated with Illumina compatible sequencing adaptors from IDT containing a
781 unique molecular index barcode for each sample library. The prepared libraries were quantified
782 using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche
783 LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the
784 Illumina NovaSeq sequencer using NovaSeq XP V1 reagent kits, S4 flowcell, following a 2x151
785 indexed run recipe.
786

787 We also sequenced 9 samples from laboratory enrichments that are released in this study. Two of
788 these enrichment samples were sequenced at The Ohio State University Genomic Shared
789 Resource Facility with library preparation Nextera XT library system and sequencing done on
790 HiSeq 2500 using methods described in more detail (10). An additional 7 laboratory enrichments
791 were sequenced in 2018 from Genomics and Microarray Shared Resource at University of
792 Colorado Denver Cancer Center, Denver, CO, USA. For these samples, library preparation was
793 done (via Tecan/Nugen Ovation Ultralow DNA Input Kit, No. 0344) and sequenced on the
794 Illumina NovaSeq 6000, with an S4 flow cell. These enrichments were used to try to enhance
795 methanogen recovery.
796
797 We obtained 133 samples from our metatranscriptomics sequencing across the multi-year study.
798 From the field in 2018 we obtained 109 metatransciptome samples from triplicate cores for each
799 of 3 depths (0-5 cm, 10-15 cm, and 20-25 cm) for July and September, and 5 depths (0-5 cm, 5-
800 10 cm, 10-15 cm, 15-20, and 20-25 cm) for August (**Table S4**). Library preparation and
801 sequencing were performed at the Department of Energy Joint Genome Institute using standard
802 methods. Plate-based RNA sample prep was performed on the PerkinElmer Sciclone NGS
803 robotic liquid handling system using Illumina Ribo-Zero rRNA Removal Kit (Bacteria) and the
804 TruSeq Stranded Total RNA HT sample prep kit following the protocol outlined by Illumina
805 with the following conditions: total RNA starting material of 100 ng per sample and 10 cycles of
806 PCR for library amplification. The prepared libraries were quantified using KAPA Biosystem's
807 next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR
808 instrument. Sequencing of the flow cell was performed on the Illumina NovaSeq sequencer using
809 NovaSeq XP v1 reagent kits, S4 flow cell, and following a 2x150 indexed run recipe. For the
810 remaining 24 samples, these were previously reported in (1, 3), with 12 collected in 2014 and 12
811 from 2015.
812
813 Two samples did not yield sufficient biomass to be sequenced at JGI. For these two samples and
814 nine additional samples from 2015, RNA-seq library construction and sequencing was performed
815 at The Genomics and Microarray Shared Resource at University of Colorado Denver Cancer
816 Center, Denver, CO, USA. Library preparation and rRNA depletion were performed using the
817 Zymo-Seq Ribo Free Total RNA Library Kit Cat No. R3000 and libraries were sequenced on the
818 Illumina NovaSeq 6000 using 2x150 paired end reads on an S4 flow cell.
819
820 **Genome assembly and binning.**
821 Metagenomic samples were assembled and binned using multiple approaches to maximize
822 assembly and metagenome assembled genome (MAG) recovery **(Table S3)**. Collectively, this
823 resulted in 42 samples for the Multiomics for Understanding Climate Change (MUCC) genome-
824 resolved database made up of 2,502 unique MAGs, with assembly and binning methods
825 described below. Described more extensively in Supplemental Table 1, but briefly here MAGs
826 were recovered from
827 1) 2013 soil field samples (n=9) (2, 11)
828 2) 2014 soil field samples (n=3) (1, 3)
829 3) 2015 soil field samples (n=3) (1, 3)
830 4) 2018 soil field samples *released in this study* (n=16)
831 5) lab enrichment samples – published previously (n=2) (4)
832 6) lab enrichment samples *released in this study* (n=9)

833
834 Raw metagenomic reads from the 2018 metagenomes were trimmed using Sickle (pe) (12) and
835 then full assemblies were generated using Megahit (v1.2.2) (13) with parameters --k-min 41 --k-
836 max 121 --k-step 10. Subsampled assemblies using 30% of sequencing reads were generated
837 using IDBA-UD v 1.1.3 (14) with default parameters. To generate more finely resolved contig
838 depth signatures as an aid to binning, reads from multiple samples were mapped to each
839 assembly using BBMap with minid=0.95 and ambig=random (v 38.70) (15) and these profiles
840 were used with assembled contigs for binning using Metabat2 (16). Bins from these assemblies
841 were combined with bins from metagenomic assemblies derived from earlier sampling of this
842 wetland and from metagenomic sequencing of wetland soil enrichments (1, 3, 4). This bin pool
843 was dereplicated using dRep (v 1.4.3) (17) at 99% identity with the 'winning' bins retained as a
844 preliminary wetland genome database. Finally, reads from the 2018 assemblies (n=16) that did
845 not map to contigs in the preliminary MAG database were grouped by depth (e.g. D1-D3 and
846 D4-D5) and co-assembled using Megahit. These two co-assemblies were binned as described
847 above and dereplicated with the preliminary database yielding a final database of 3217 wetland
848 MAGs. MAG completeness and contamination was estimated using CheckM (18) and taxonomy
849 assigned using GTDB-tk (v 1.5.0) with GDTB database release 202 (19, 20). Of these 3217
850 MAGs, 2502 were of medium or high quality based on adapted MIMARKS standards
851 (completeness >=50% and contamination <10%) (21).
852
853 **Metatranscriptomic table construction.** Raw metatranscriptomic reads were filtered and initial
854 QC performed using Joint Genome Institute standard protocols. Filtered reads were further
855 quality trimmed using sickle (pe) (12) and mapped to the full MAG database using BBMap (v
856 38.70) (15). Following mapping, bamfiles were filtered to retain mappings of at least 99% ID
857 and per-gene read counts were estimated using HT-Seq (v0.11.2) (22). The read count table was
858 first filtered to retain genes which have a minimum total count of 10 across all samples. The
859 filtered table was normalized for gene length and TMM normalized using the EdgeR 'cpm'
860 function with log2 normalization (23). MAG mean transcription values were calculated for those
861 MAGs having at least 20 genes transcribed.
862
863 **Taxonomic and functional assignments.** All MAGs were used as input for GTDB-tk taxonomy
864 assignment using the *de novo* workflow and Nanoarchaeota as the outgroup using release 202 of
865 GTDB (19). The resulting multiple sequence alignment was used to produce a guide tree using
866 IQ-TREE (24) and the LG+F+G model. A PMSF tree was generated using IQ-TREE with model
867 LG+C10+G+F and 100 bootstraps. The resulting tree was annotated with iTOL (25).
868 Bacterial and archaeal MAGs were annotated using DRAM v1(26). We also annotated bacterial
869 and archaeal metagenome-resolved transcriptomes *via* DRAM (26). To assign traits to MAGs,
870 we curated the MAG annotations made by DRAM (26) using a ruleset visualized in **Fig. S5**.
871 MAGs were assigned potential traits derived from metagenomics and expressed traits derived
872 from metatranscriptomics. From metagenomics, MAGs were first classified by respiratory
873 capacity based on the presence of >50% of the subunits required for Complex 1 of the electron
874 transport chain and the presence of at least one gene for an electron acceptor. Based on which
875 acceptor was predicted, a MAG would be assigned to a particular functional guild (e.g., N
876 reducer, aerobic, iron reducer). As an example, for a MAG to be classified as microaerophillic,
877 we required the MAG to have more than 50% of Complex 1 subunits and at least one subunit of
878 a high affinity cytochrome oxidase. Likewise, if a MAG did not have more than 50% of the

879 subunits required for Complex 1 of the electron transport chain or the potential for any electron
880 acceptor, it was classified as an obligate fermenter (**Table S5**).
881
882 Beyond respiratory metabolisms, other traits were assigned based on functional gene presence.
883 For example, N fixation was assigned based on the presence of at least one subunit of any
884 nitrogenase (**Table S5**). For specific metabolisms, such as methanogenesis and methanotrophy,
885 assignment was based on the presence of a key functional gene (*mcrA* and *pmoA*, respectively) or
886 by manual curation including taxonomic information given the conserved nature of these
887 metabolisms at the genus level and the tendency for these key functional genes to be on short
888 contigs, precluding them from binning (1, 3). From metatranscriptomics, trait calls were based
889 on potential calls from MAGs, but also required various levels of transcription to be considered
890 expressed. For example, to be considered an expressed microaerophile, a MAG had to 1) be
891 classified as a microaerophile based on metabolic potential (more than 50% of complex 1 subunit
892 and at least one subunit of a low affinity cytochrome oxidase) and 2) had to be expressing more
893 than 30% of complex 1 subunits and at least one subunit of a low affinity cytochrome oxidase.
894 For trait classified based on an absence of data (obligate fermenter) or taxonomy (methanogen,
895 methanotroph), MAGs needed to be considered active (more than 20 genes transcribed) to be
896 classified as expressed for a particular trait. As trait assignment was based on MAG annotation,
897 we also assigned MAG-level transcription. Likewise, MAGs could be assigned to more than one
898 functional guild, and totals were summed across samples.
899
900 Phylogenetic analyses were performed to refine the annotation of nitrogen related metabolism
901 including genes annotated as respiratory nitrate reductase (*nar*), nitrite oxidoreductase (*nxr*),
902 ammonia monooxygenase (*amo*), or methane monooxygenase (*pmo*) to improve the assignment
903 the nitrogen cycling capabilities of MUCC MAGs (**Table S5**). Specifically, Nxr/Nar and
904 PmoA/AmoA amino acid reference sequences were downloaded and this set of reference
905 sequences were combined with amino acid sequences of homologs from the MUCC, aligned
906 separately using MUSCLE (v3.8.31) (27), and run through an in-house script for generating
907 phylogenetic trees. This resulted in two phylogenies, one for *Nxr/Nar* genes and one for
908 *Pmo/Amo* genes, with this approach used to refine the homology-based gene annotations in the
909 MAG database.
910
911 For MAGs assigned to the methanogen functional guild, substrate classification was further
912 characterized, assigning each as hydrogenotrophic methanogen, acetoclastic methanogen, or
913 methylotrophic (methyl-N, methyl-S, methyl-O) methanogen, with genes required for each
914 detailed in Table S4. Gene calls were based on DRAM annotations. Specifically, all genes
915 involved in methanogenesis were inventoried per MAG and then based on those genes that were
916 specific for a substrate type, substrate calls were made. Archaeal metabolism calls were based on
917 DRAM annotations with specific genes required for each (detailed in **Tables S5**, **S7**).
918
919 **Phylogenomic analysis of OWC and Stordalen Mire methanogen MAGs.** Phylogenomic
920 analysis of the 99% dereplicated methanogen MAGs from OWC (n=85) and Stordalen Mire
921 (n=74) was performed using the GTDB-Tk v2.3.0 r214 (19). The alignment was based on 53
922 concatenated archaeal marker genes. A GTDB-derived MAG (GCA_002495465.1) from the
923 phylum Undinarchaeota was used as in outgroup. The tree was visually modified in R using the
924 ggtree package (28). Metatranscriptome data from Stordalen Mire included on the tree was

925 previously published (5). The Newick tree is available at on Zenodo at DOI
926 10.5281/zenodo.8194032.
927
928 **Cross-validation of wetlands taxa recovered from OWC.** We validated the ubiquity of our
929 identified wetlands microbial taxa at OWC with an independent 16S rRNA gene database
930 assembled from 505 soils collected from 5 unique wetlands systems that collectively represent a
931 broad range of geographic locations (see **Fig. S3**), span a range of geochemical and climatic
932 conditions, and represent different wetland types: marsh (P7, P8, LA2), swamp (JLA), and bog
933 (SPRUCE). All selected samples were previously amplified with the EMP primers 515F/806R
934 and sequenced on the Illumina Miseq. Using DADA2 (8) through QIIME2 (7) 16S rRNA gene
935 sequences were processed with a consistent methodology and used the SILVA v138 and GDTB
936 (r202) databases to assign taxonomy across all 505 samples. We then compared the
937 representative sequences of our wetlands 16S rRNA to those recovered in our assembled
938 reference database using BLAST to compare sequence similarity at 100, 99, 97, and 80%
939 similarity cutoffs. The percentages of ASVs from each site and within the entire wetland
940 database that were detected in our OWC dataset were calculated for each cutoff. The relative
941 abundances of the ASVs within these cutoff groups was calculated for each site and the entire
942 database to understand how much community composition from other wetland sites was shared
943 with OWC.
944
945 **Assessment of genome transcription across major lineages.** To calculate the percentage of a
946 genome transcribed for the 1,948 genomes recovered where at least 20 genes were transcribed,
947 we first counted the number of genes transcribed per genome, and then divided that count by the
948 number of predicted genes per genome as estimated by CoverM and multiplied by 100. We
949 performed a Mann-Whitney U test to assess if there was a significant difference in the
950 percentage of a genome expressed in bacterial versus archaeal genomes. We also classified phyla
951 by their median transcription and variability by calculating the median percentage of genomes
952 transcribed (summarized by phylum) and by the variability of that % transcription of MAGs
953 within a phylum, calculated as the median relative standard deviation (RSD, aka coefficient of
954 variation).
955
956 **Identification of major environmental axes driving wetlands microbiome variation.** We first
957 reduced collinearity by dropping a subset of geochemical and climatic variables that were highly
958 correlated with retained variables (Spearman's $r > 0.75$; Fig. S1) including Ethanol, Isopropanol,
959 extractable Na, extractable Al, % Sand, extractable K, and temperature. We log-transformed
960 variables where appropriate and standardized variables prior to analyses. Our final set of
961 environmental predictors included measured soil properties (Dissolved Oxygen, Acetate, Nitrate,
962 Nitrite, Sulfate, Fe2, pH, CEC, % Total Organic Carbon, % Silt, % Clay, Hydraulic
963 Conductivity, extractable Mg), site attributes (ecotype, depth, collection month, collection year),
964 associated geographic variables (latitude/longitude), and greenhouse gas concentrations (CH_4 and
965 CO_2 peeper concentrations).
966
967 To characterize how the composition and transcription of microbial communities differed across
968 time, space, and major geochemical gradients, we first assessed calculated dissimilarities with
969 Bray-Curtis metric on square root transformed relative abundances. We used three metrics of
970 microbiome communities: taxonomic composition (ASVs), relative mean MAG transcription,

971 and relative mean metabolic guild transcription, and visualized as ordinations with nonmetric
972 multidimensional scaling (k=3). We tested the correlation between pairwise distances in bacterial
973 and archaeal composition/transcription to pairwise distances in our suite of geochemical and
974 spatiotemporal variables using Mantel tests with 999 permutations. Then, to identify the best
975 overall explanatory models associated with differences in composition and expression we
976 conducted multiple regression analyses on distance matrices.
977
978 **Genome co-expression network analysis and process-based model generation.** Beyond
979 individual lineages correlations described above, we were interested in identifying if there were
980 community-level emergent properties (e.g., clusters of co-transcribed genomes) that were highly
981 correlated in their relative expression to relate to in situ measurements of greenhouse gas
982 concentrations (CH_4, CO_2). To test this, we used weighted correlation network analysis
983 (WGCNA) (29) across all of our metatranscripomes (n=132, 1 sample was excluded due to
984 outlier screening within the WGCNA pipeline). A signed adjacency measure for each pair of
985 features was calculated with the power threshold of 7, identified by plotting mean connectivity as
986 a function of the soft-thresholding power and selecting the power threshold above 0.80 to
987 optimize the scale-free topology network fit. The adjacency matrix was used to calculate
988 topological overall measure (TOM) and derive a co-expression network. We used a signed
989 hybrid network, with a minimum module size of 50, and a merge height of 0.3. Hierarchical
990 clustering was employed to resolve subnetworks (i.e., modules), using a distance based on the
991 calculated TOM. For each subnetwork, we calculate pairwise Pearson correlation coefficients
992 between the subnetwork principal components and key geochemical measurements (pH, sulfate,
993 CEC, nitrate, nitrite, DO, % TOC, conductivity, depth, Fe_2, acetate, Mg, and ecotypes) as well as
994 our *in situ* measurements of CO_2 and CH_4 concentrations.
995
996 We next assessed the two subnetworks that had the highest significant correlation scores
997 (correlations between module eigenvalues and CO_2/CH_4 concentrations, which we refer to as
998 'traits') with an intramodular analysis. Our first goal was to identify genomes with a high
999 genome trait significance (the absolute value of the correlation between the genome and the trait)
1000 and subnetwork membership (the correlation of the module eigenvalue and the genome
1001 expression profile). Second, we also computed VIP scores from sPLS (30) models (in relation to
1002 both CH_4 and CO_2) for all genomes within each of the two subnetworks to obtain an additional
1003 metric of their individual importance. VIP scores were highly correlated with genome trait
1004 significance scores (rho > 0.9 for all). Subnetwork #1 contained 372 genomes (the subnetwork–
1005 trait correlation to CH_4 and CO_2 was 0.38 and 0.36 respectively, with $P<0.0001$ for both).
1006 Likewise, for each subnetwork, we found that genome significance to be significantly correlated
1007 to subnetwork membership both for CH_4 (Rho = 0.54 and 0.52 for subnetworks #1 and 2,
1008 respectively; $P<0.001$) and for CO_2 (Rho = 0.54 and 0.29 for subnetworks #1 and 2, respectively;
1009 $P<0.001$).
1010
1011 MAGs within the process-based model were based on WGCNA and sPLS analyses, as well as
1012 MAG expression profiles. First, MAGs within each of the two subnetworks (brown and
1013 turquoise) were filtered to those with a VIP score ≥ 1 in relation to CH_4 and then clustered at the
1014 genus level. Then clusters were classified based on expressed traits. For clusters containing more
1015 than one MAG, functional guilds were consolidated. If trait assignments disagreed (e.g. aerobe
1016 and obligate fermenter) within a cluster, the functional guild from the most complete MAG

within a cluster was used. Clusters with no expressed adjectives or without at least one MAG expressing ≥ 20% of its genes were removed from subsequent analyses. Remaining clusters were retained in the process-based model. For MAGs within the remaining clusters, DRAM annotations were used to profile expressed carbon usage. Based on the expressed carbon utilization and functional guild of each cluster, a schematic representation of the process model (**Fig. 4**) was constructed in Adobe Illustrator, with all statistics and metabolic calls including VIP values, expression patterns, functional guilds, and carbon utilization provided in **(*SI Appendix*, Tables S12-S13**).

Figure S1. Experimental Design. (A) Upset plot showing the major data types obtained from samples and the number of samples obtained for each set (top bars) including from left to right: geochemistry, 16S rRNA amplicon sequencing, NMR metabolomics, metatranscriptomics, and metagenomics. The gray bars (right) represent the number of interesting sets across sample types. (B) Overview of the number of metagenome-resolved genomes (MAGs) recovered in this study. All 2,502 medium and high-quality (MQHQ) MAGs used in analyses were ≥50% complete, were <10% contaminated. (C) Schematic illustration depicting the soil core sampling as well as the greenhouse gas monitoring using a) porewater dialysis peepers for porewater production, b) chambers for emission quantification, and c) EC tower for site wide data. (D) Boxplot highlights the amount of sequencing in gigabasepairs (Gbp) for this study in Old Woman Creek (OWC) versus other classified wetlands in the Sequence Read Archive (SRA) for metagenomes and metatranscriptomes. Each point represents a single sample and boxplots show the median Gbp. (E) Overview of the genome-resolved metagenomics (top) and genome-resolved metatranscriptomics (bottom) bioinformatics pipeline we developed to obtain genome-resolved microbial community activity profiles.

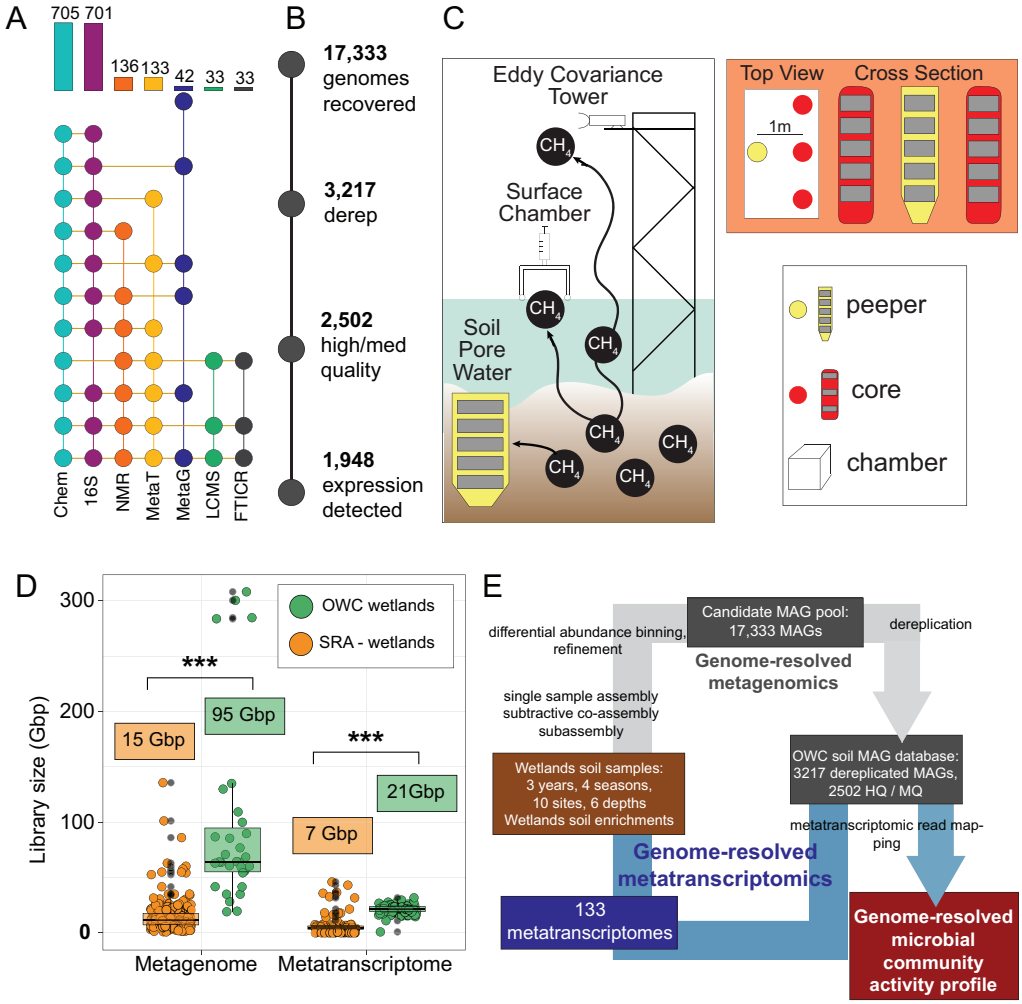


Figure S2. Genome expression is phylogenetically predictable, and the relative expression of major lineages varies across ecotype and depth. (A) Boxplots at the domain level summarizing the distribution of the percentage of the genome that is transcribed for all 1948 MAGs in which transcription was detected (i.e., the number of genes transcribed over the total number of predicted genes). (B) Boxplots as per A at the phylum level, ordered by percent of genome transcribed. Only phyla composed of at least 5 MAGs recovered are visualized here. Archaeal phyla are denoted with bolded labels. (C) A scatter plot summarizing the median percentage of genome transcribed within each phylum versus the median relative standard deviation (RSD). Points in the upper left-hand corner represent phyla comprised of MAGs with a low median % of overall genes transcribed and high variability across MAGs, and conversely, points in the lower right-hand corner represent phyla are comprised of MAGs with a consistently high % of the overall genome transcribed.

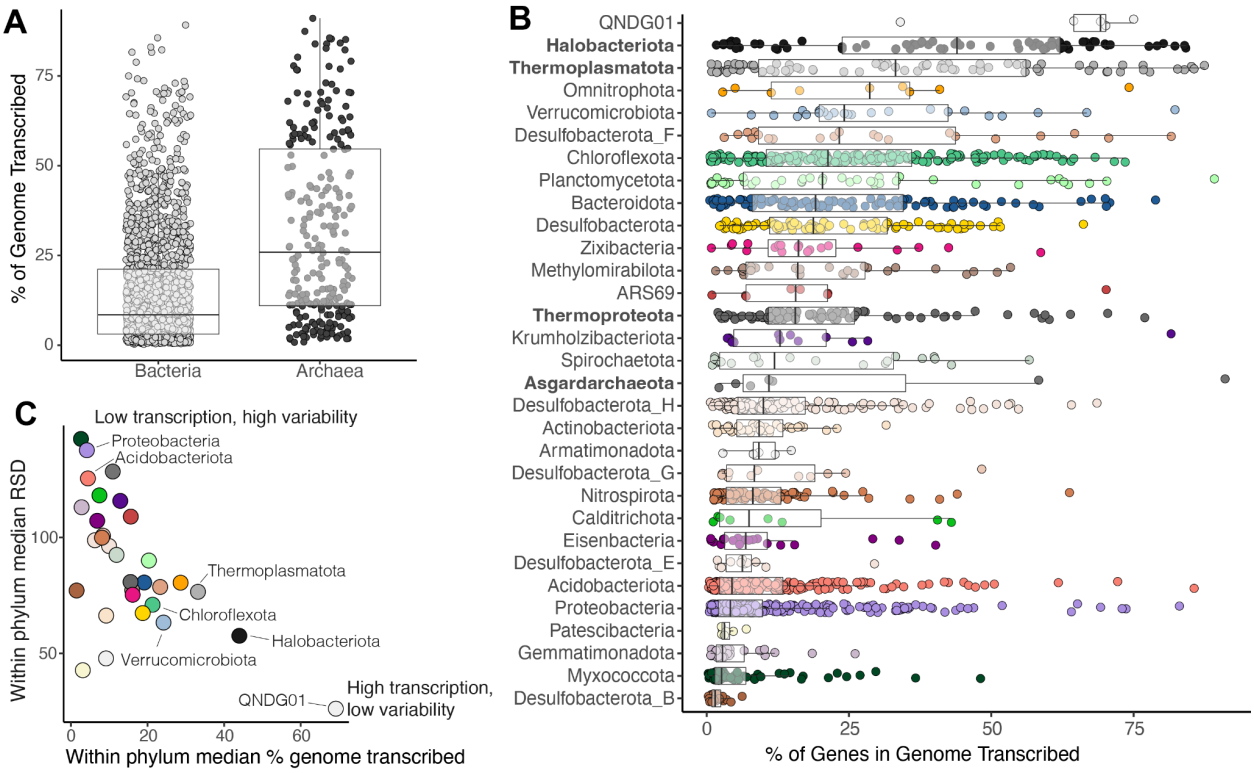


Figure S3. MAGs and ASVs recovered from this study (OWC) represent a large fraction of the microbiome of diverse North American wetland sites. (A) The geographic distribution of wetland sites for which we compiled a North American wetlands database from available 16S rRNA **(B)** A bar plot of the number of samples at each wetland site (ASV samples). The % composition **(C)** and the % of ASV matches **(D)** of the wetlands database covered by OWC ASVs at decreasing threshold cutoffs. OWC ASVs covered 36% and 51% of the composition of the wetlands database, at 99% and 97% threshold cutoffs, respectively.

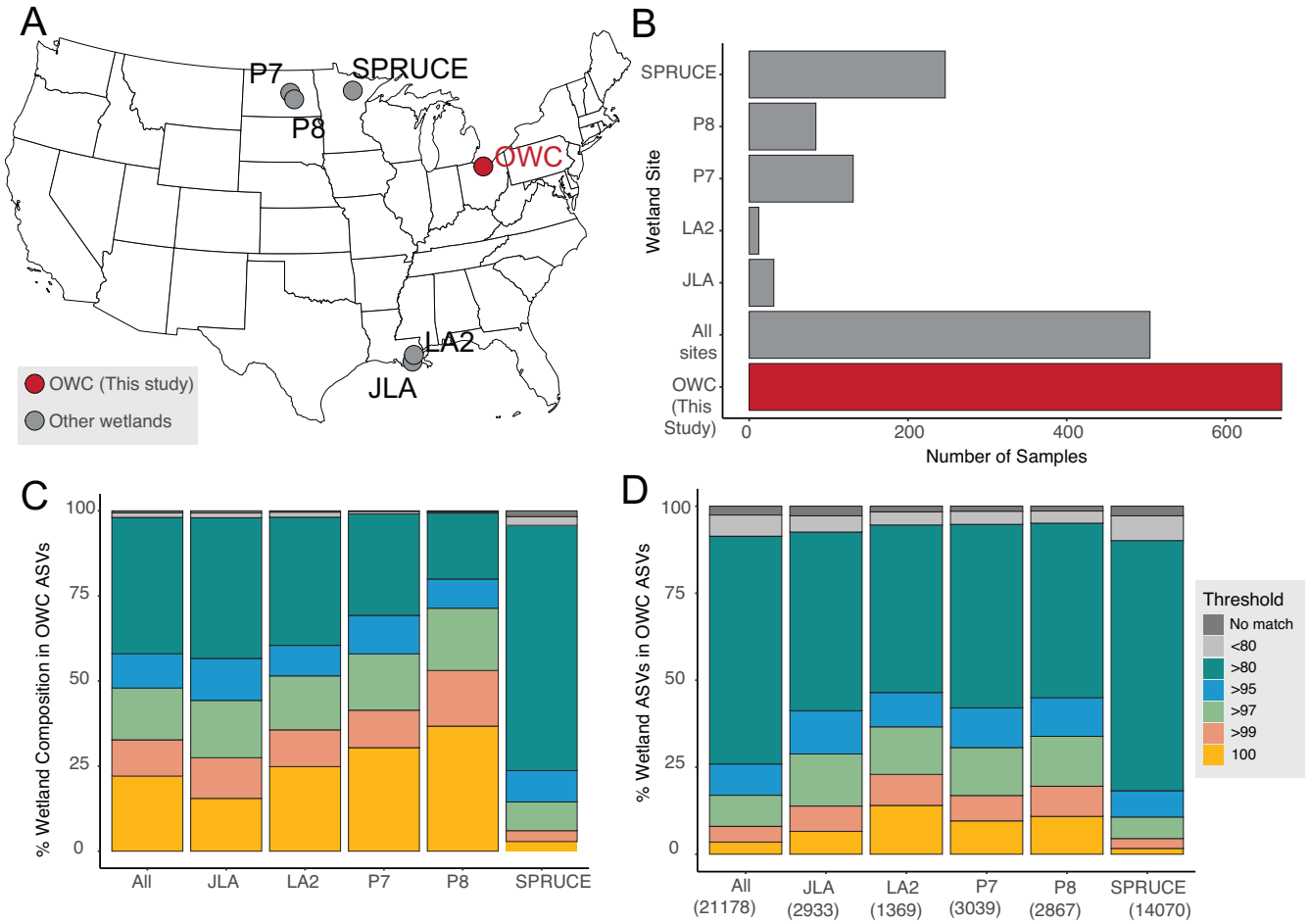


Figure S4. Phylogenetic tree of methanogens recovered from Old Woman Creek and Stordalen Mire shows taxonomic overlap while maintaining site specific MAGs. Phylogenetic tree was inferred using the GTDB-Tk de_novo_wf for methanogen MAGs derived from OWC (n=85) and Stordalen Mire (STM) (n=74), using p__Undinarchaeota as the outgroup. Leaves are colored by MAG taxonomy at the order level and MAG source is denoted by innermost ring. Two stacked barcharts highlight the average expression per MAG by depth and site in OWC (inner bar chart) and Stordalen Mire (outermost bar chart).

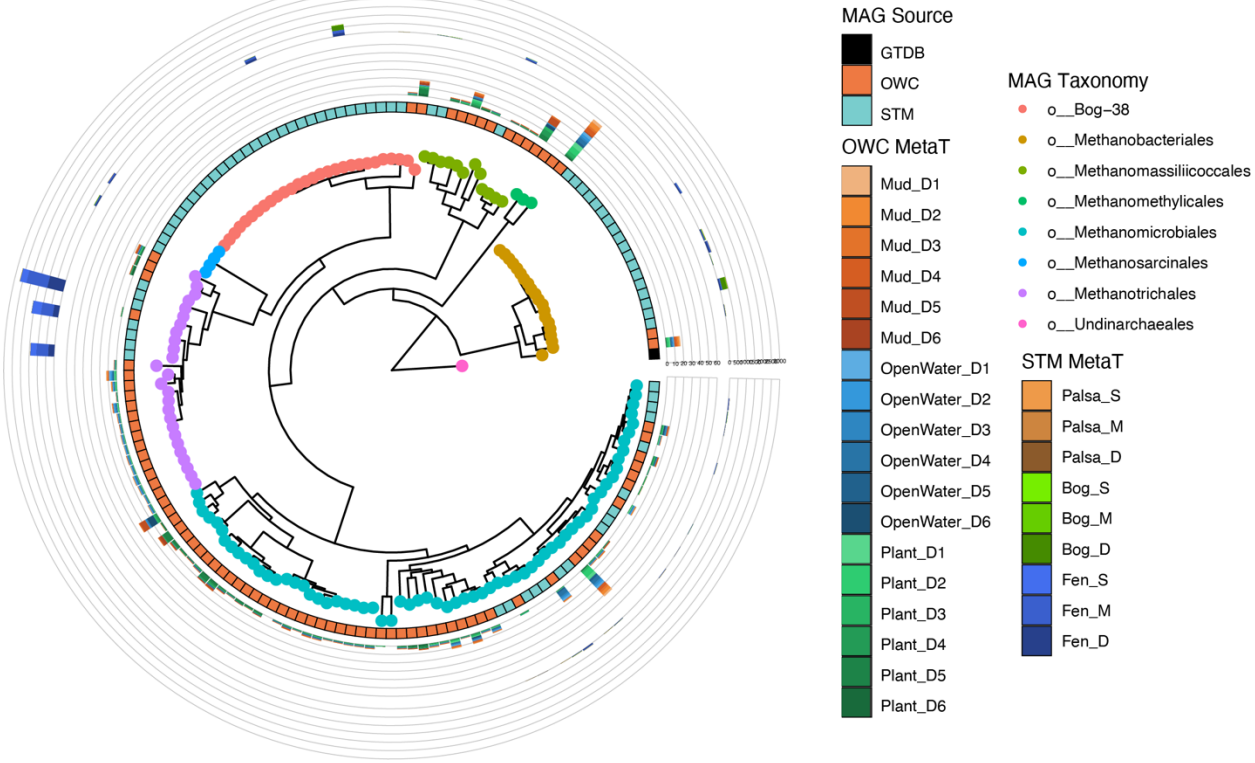


Figure S5. Overview of pathway for metabolic guild characterization for all MAGs recovered. The 2,502 metagenome assembled genomes were characterized for functional potential and expression of 14 major traits including: aerobic, microaerophilic, photosynthetic, N reducer, Iron reducer, S reducer, methanogen, obligate fermenter, DNRA, N fixation, nitrification, aerobic methanotroph, Iron oxidizer, and S oxidizer. The path diagram summarizes which gene suites were required for a trait assignment. For the subset of 1,948 transcribed genomes, we also assigned 'expressed' metabolism from transcripts recruited back to MAGs. For some genes (e.g., *pmoA*), phylogenetic analyses and other non-homology-based methods were used for confirmation. Additional details are provided in the methods.

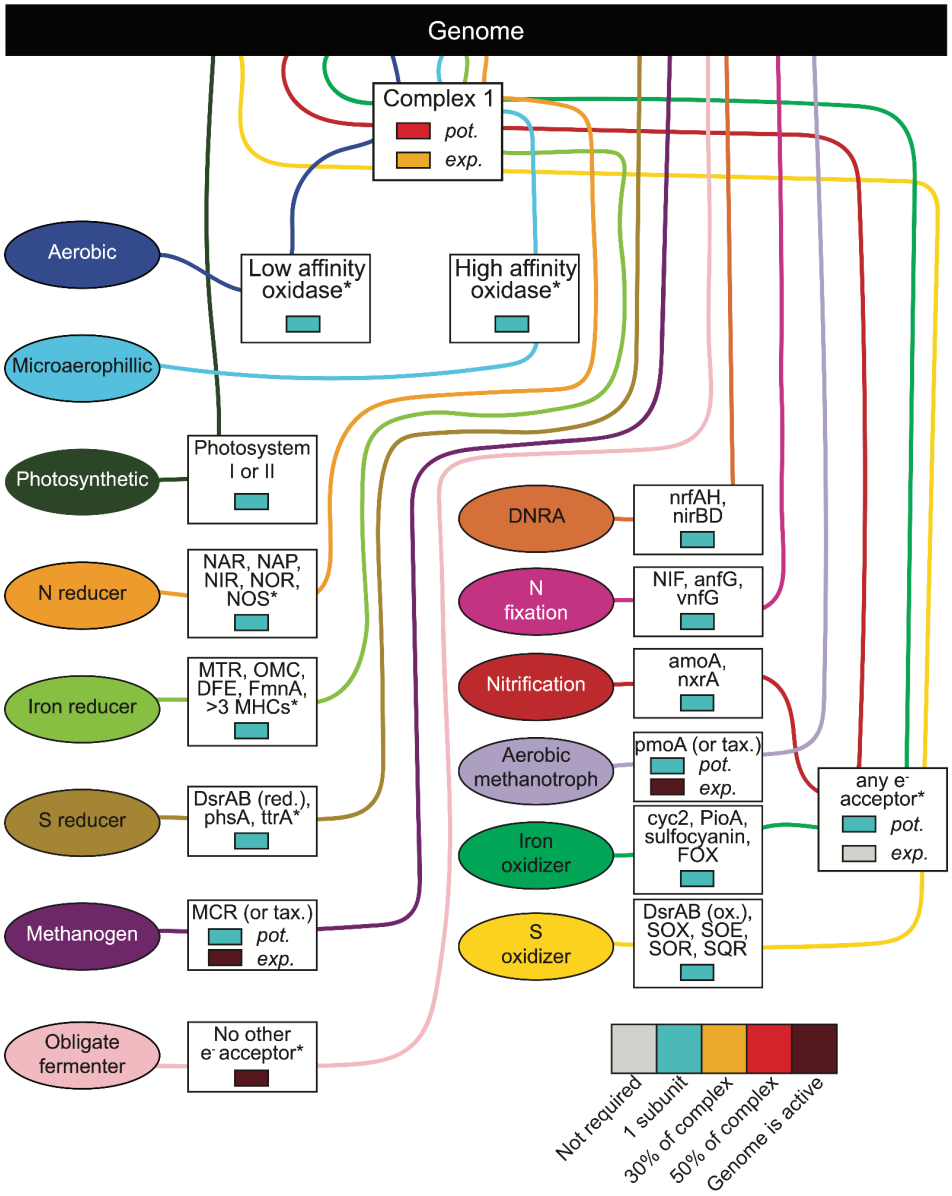


Figure S6. The dominant environmental drivers shaping wetlands microbiome are highly conserved across phylotype, genome, and guild level transcriptional profiles. Wetlands microbiomes are shaped by depth, ecotype, Fe2, acetate, and to a lesser extent, other spatial, temporal, and geochemical parameters. **(A)** Nonmetric multidimensional scaling (NMDS) ordinations of wetlands samples (left to right): overall composition (N=671), genome transcription (N=133), and guild transcription (N=133). The top row is colored by ecotype (brown=mud; green=plant; blue=open water) and the bottom row is colored by depth (yellow=surface to blue=deep), highlighting two of the major environmental predictors. **(B)** Bar plots visualizing the relative importance of spatiotemporal predictors included in models for explaining overall variation. **(C)** Bar plots visualizing the mantel correlations of geochemical predictors included in multiple regression models for explaining overall variation.

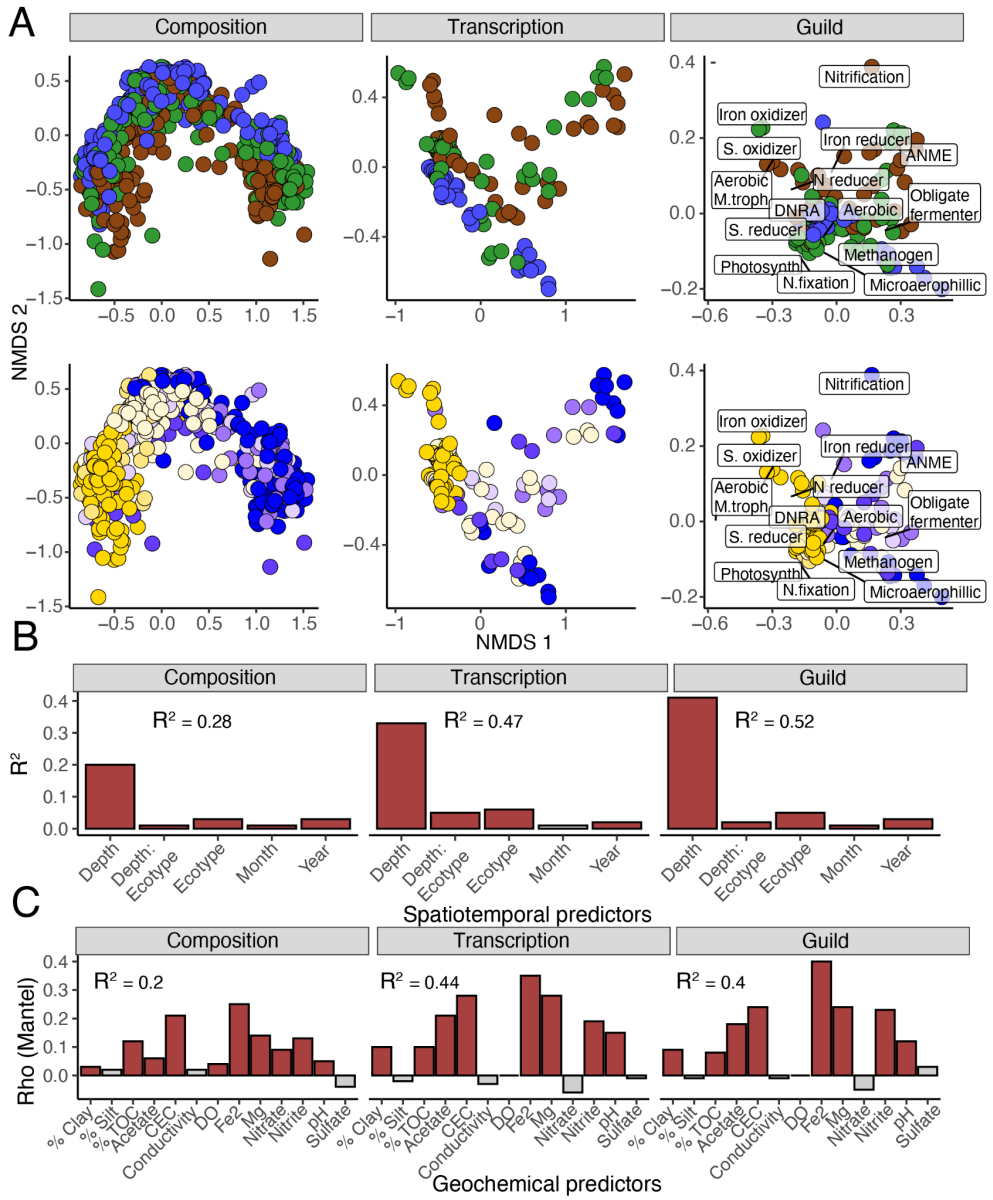


Figure S7. Inventory of methane cycling microorganisms. (A) Alluvial plot highlighting the taxonomic diversity of methanogens recovered. The first gray bar chart summarizes the number of genomes per genus recovered and the second gray bar chart summarizes the max transcription detected of a MAG within the genus. The MAG-resolved metabolisms show transcription-based substrate utilization. **(B)** Scatterplot highlighting the number of genes observed in a MAG versus the number of genes transcribed in the MAG and colored by metabolism. The dashed line represents a 1:1 line, points above the line signify MAGs that have a higher number of genes transcribed than average, and points below, fewer. **(C)** Boxplots summarizing the proportion of genes transcribed per genus and colored by metabolism.

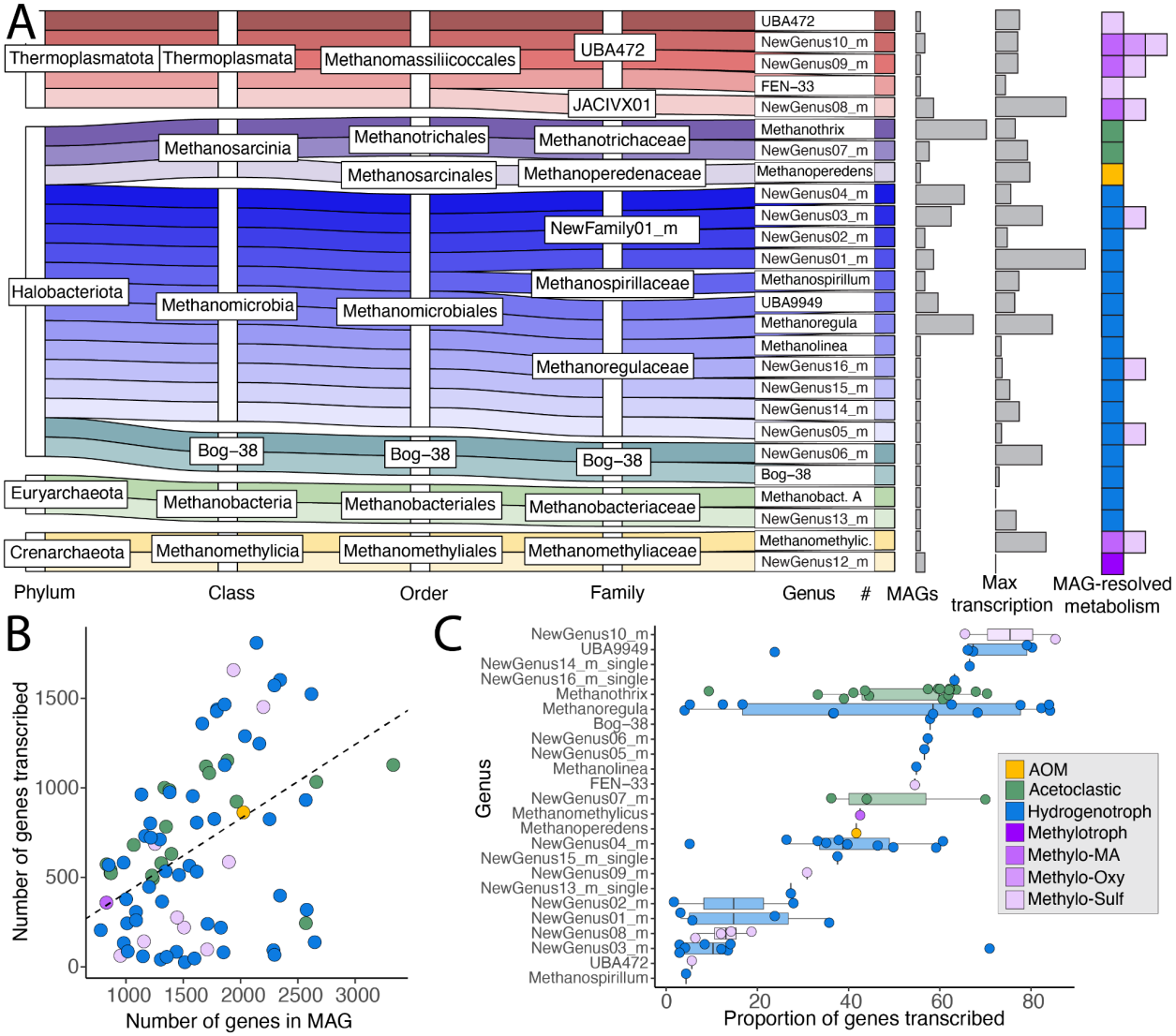


Figure S8. Methanogen substrate concentrations across years and depths. A) Heatmap displays paired LCMS and NMR metabolite data, with samples ordered from left to right first by patch type (mud, open water) then by depth. Compound text color indicates detection methods including LCMS (black) or NMR (grey). Bar chart shows the number of samples each compound was detected in for all samples analyzed including 109 samples for NMR and 33 samples for LCMS. B) and C) Boxplots show log transformed NMR concentrations of acetate, formate, and methanol by depth and by year, respectively.

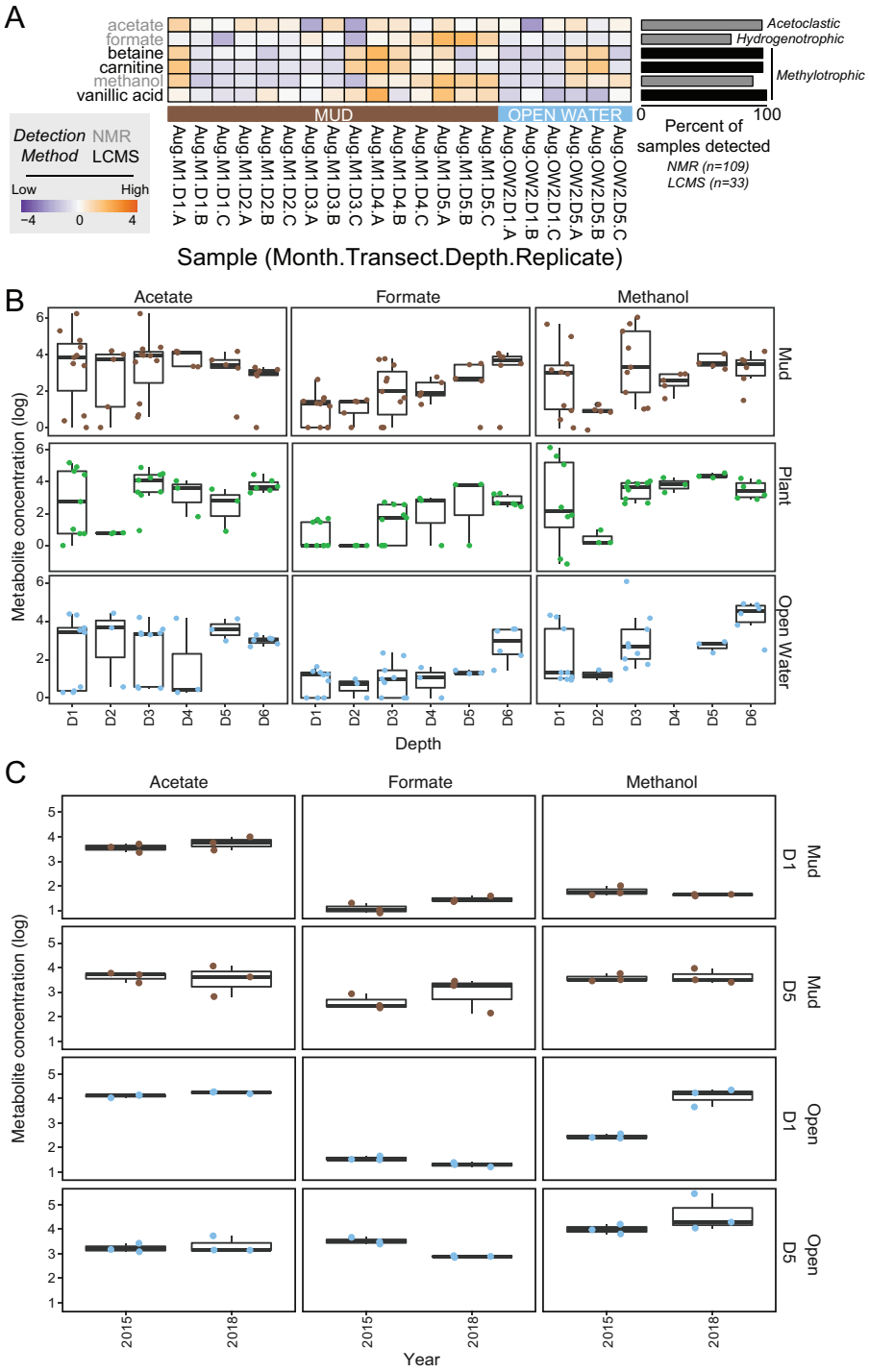
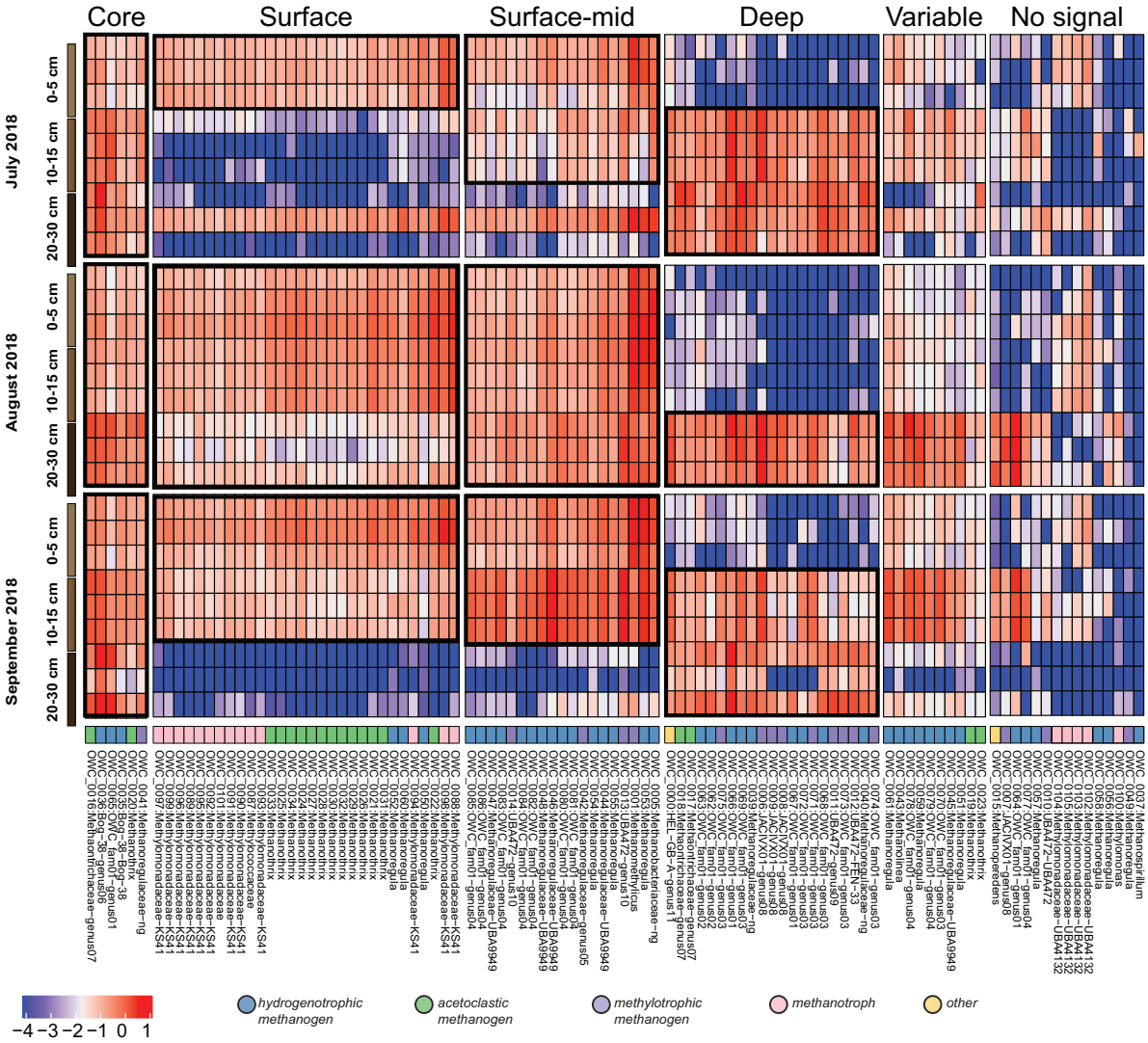


Figure S9. Methane cycling microorganisms show coordinated gene expression patterns, with groups co-localized by depth neighborhoods. By row metatranscriptome data from 3 cores at 3 depths (n=9) for each month (July, August, September) in blocks. Each column is the mean genome transcript for methanotroph or methanogen MAGs, with taxonomy and naming indicated on bottom and energy strategy denoted by box colors. Transcript response patterns are clustered by location, with some methanogens active in all depths in all seasons, some showing stronger surface, surface-middle, deep, deep-middle, and no clear depth signal.



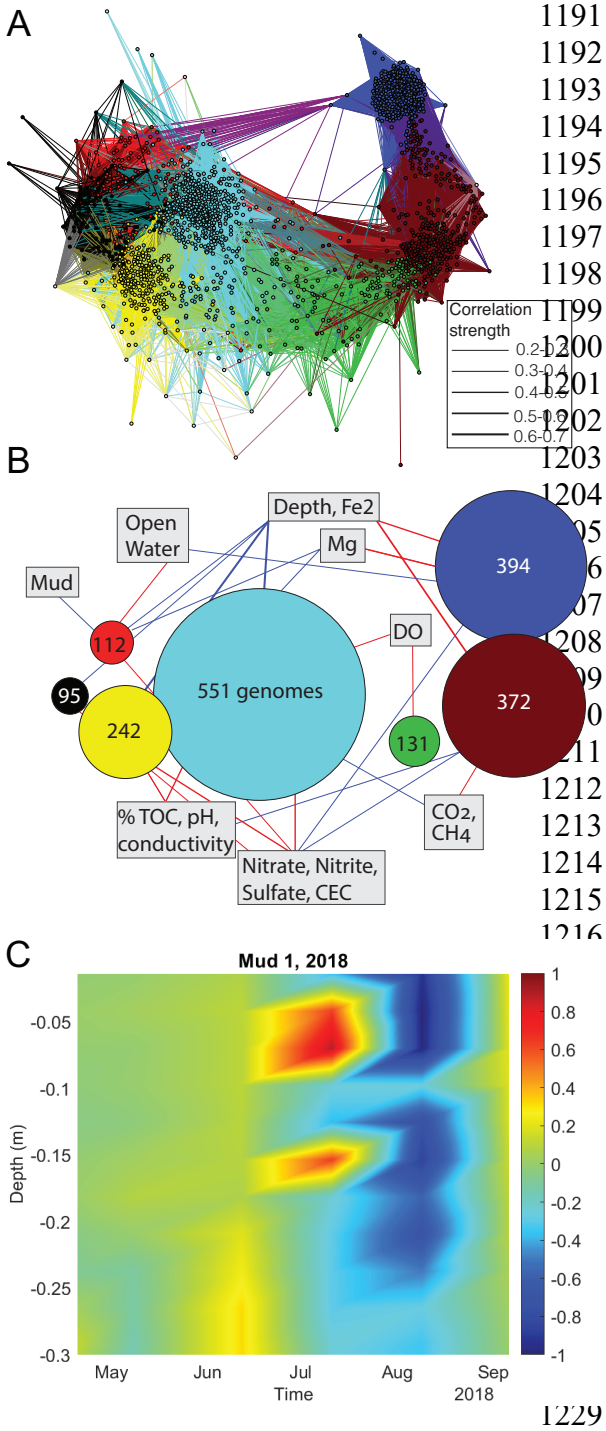


Figure S10. Coordinated co-transcription of key subnetworks drives CH_4 and CO_2 concentrations in wetlands systems. (A) Overall co-expression network from 1,948 MAGs across 132 metatranscriptomics samples. Node outline colors highlight subnetworks (where N > 100) identified with weighted gene co-expression analyses (WGCNA). **(B)** A simplified network diagram highlighting the associations between subnetworks (sized scaled by number of MAGs) and environmental features. Red and blue edges correspond to positive and negative associations, respectively (**C**) Methane production calculated as previously described (1) from porewater peeper measurements and chamber measurements (**Table S4**). This profile represents a mudflat across 5 months in 2018.

1237 **Captions for Tables S1-S11.**
1238
1239 **Table S1.** Overview and source details of the Old Woman Creek samples included in this study,
1240 including metagenomes, metatranscriptomes, and 16S rRNA gene amplicon sequences.
1241
1242 **Table S2.** Microbial abundance spanning space and time as determined by 16S rRNA amplicon
1243 sequencing for 705 samples.
1244
1245 **Table S3.** The 2,502 medium and high-quality MAGs recovered and their attributes, including:
1246 the RED taxonomy (including assigned novelty status), GTDB-Tk assigned taxonomy,
1247 contamination, completeness, genome size, number of scaffolds, number of contigs, N50
1248 scaffolds, N50 contigs, % GC, coding density, predicted genes.
1249
1250 **Table S4.** Overview of metatranscriptomics and metagenomics samples, associated geochemical
1251 and flux measurements, and metadata included in this study including accession numbers.
1252
1253 **Table S5.** The metabolic guild assignments for 2,502 MAGs in terms of both functional
1254 potential (e.g. metagenomic annotations) and transcribed potential (e.g. metatranscriptomic
1255 annotations, available for the subset of 1,948 MAGs that met our minimum transcription
1256 threshold), corresponding with Fig. 2.
1257
1258 **Table S6.** Model outputs (mantel tests and permanova models) assessing the relationships
1259 between wetlands microbial communities (including the composition, transcription, and guild
1260 level) and spatiotemporal and geochemical variables to explain variation in the distribution of
1261 relative abundance. The R^2 values reported refer to permanova predictors and rho values refer to
1262 mantel tests (see Methods for details).
1263
1264 **Table S7.** Functional assignments for Archaeal genomes, and gene-level requirements for
1265 assignments corresponding with Fig. 3.
1266
1267 **Table S8.** Soil metabolite concentrations as determined by LC-MS and NMR.
1268
1269 **Table S9.** Metabolite identification data, including specifics of LC-MS methodology.
1270
1271 **Table S10.** Genome metabolism for methanogens and methanotrophs expanded by gene content.
1272
1273 **Table S11.** Genome transcription stability across years, months, and depths.
1274
1275 **Table S12.** Correlations between transcription of bacterial and archaeal genomes and
1276 geochemical attributes detected via WGCNA network construction and analysis.
1277
1278 **Table S13.** Genomes included in focal WGCNA subnetworks used to build process models (Fig.
1279 4) including corresponding VIP scores and metabolism.
1280
1281

SI References

1. J. C. Angle, *et al.*, Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat Commun* **8**, 1567 (2017).

2. A. B. Narrowe, *et al.*, High-resolution sequencing reveals unexplored archaeal diversity in freshwater wetland soils. *Environ Microbiol* **19**, 2192–2209 (2017).

3. G. J. Smith, *et al.*, Members of the Genus Methylobacter Are Inferred To Account for the Majority of Aerobic Methane Oxidation in Oxic Soils from a Freshwater Wetland. *mBio* **9**, e00815-18.

4. A. B. Narrowe, *et al.*, Uncovering the Diversity and Activity of Methylotrophic Methanogens in Freshwater Wetland Soils. *mSystems* **4**, e00320-19 (2019).

5. J. B. Ellenbogen, *et al.*, Methylotrophy in the Mire: direct and indirect routes for methane production in thawing permafrost. *mSystems* **0**, e00698-23 (2023).

6. J. G. Caporaso, *et al.*, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621–1624 (2012).

7. E. Bolyen, *et al.*, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).

8. B. J. Callahan, *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).

9. C. Quast, *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596 (2013).

10. M. A. Borton, *et al.*, Targeted curation of the gut microbial gene content modulating human cardiovascular disease. 2022.06.20.496735 (2022).

11. A. B. Narrowe, *et al.*, Complex Evolutionary History of Translation Elongation Factor 2 and Diphthamide Biosynthesis in Archaea and Parabasalids. *Genome Biology and Evolution* **10**, 2380–2393 (2018).

12. najoshi, sickle - A windowed adaptive trimming tool for FASTQ files using quality (2023) (July 7, 2023).

13. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

14. Y. Peng, H. C. M. Leung, S. M. Yiu, F. Y. L. Chin, IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).

15. B. Bushnell, "BBMap: A Fast, Accurate, Splice-Aware Aligner" (Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), 2014) (December 20, 2023).

16. D. D. Kang, *et al.*, MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

1317 17. M. R. Olm, C. T. Brown, B. Brooks, J. F. Banfield, dRep: a tool for fast and accurate genomic
1318 comparisons that enables improved genome recovery from metagenomes through de-replication.
1319 *ISME J* **11**, 2864–2868 (2017).

1320 18. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: assessing the
1321 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.*
1322 **25**, 1043–1055 (2015).

1323 19. P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: a toolkit to classify genomes
1324 with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).

1325 20. D. H. Parks, *et al.*, Recovery of nearly 8,000 metagenome-assembled genomes substantially expands
1326 the tree of life. *Nat Microbiol* **2**, 1533–1542 (2017).

1327 21. R. M. Bowers, *et al.*, Minimum information about a single amplified genome (MISAG) and a
1328 metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725–731
1329 (2017).

1330 22. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput
1331 sequencing data. *Bioinformatics* **31**, 166–169 (2015).

1332 23. M. Smid, *et al.*, Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq
1333 data performs similarly in intersample analyses while improving intrasample comparisons. *BMC*
1334 *Bioinformatics* **19**, 236 (2018).

1335 24. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective
1336 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and*
1337 *Evolution* **32**, 268–274 (2015).

1338 25. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and
1339 annotation. *Bioinformatics* **23**, 127–128 (2007).

1340 26. M. Shaffer, *et al.*, DRAM for distilling microbial metabolism to automate the curation of microbiome
1341 function. *Nucleic Acids Research* **48**, 8883–8900 (2020).

1342 27. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput.
1343 *Nucleic Acids Res* **32**, 1792–1797 (2004).

1344 28. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: an r package for visualization and
1345 annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology*
1346 *and Evolution* **8**, 28–36 (2017).

1347 29. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC*
1348 *Bioinformatics* **9**, 559 (2008).

1349 30. H. Abdi, Partial least squares regression and projection on latent structure regression (PLS
1350 Regression). *WIREs Computational Statistics* **2**, 97–106 (2010).