Metabolic interactions underpinning high methane fluxes across terrestrial freshwater wetlands

Emily K. Bechtold[1], Jared B. Ellenbogen[1], Jorge A. Villa [2], Djennyfer K. de Melo Ferreira [1], Angela M. Oliverio [1,3], Joel E. Kostka [4], Virginia I. Rich [5], Ruth K. Varner [6], Sheel Bansal [7], Eric J. Ward [8], Gil Bohrer [9], Mikayla A. Borton [1], Kelly C. Wrighton [1] ✉ & Michael J. Wilkins [1] ✉

Current estimates of wetland contributions to the global methane budget carry high uncertainty, particularly in accurately predicting emissions from high methane-emitting wetlands. Microorganisms drive methane cycling, but little is known about their conservation across wetlands. To address this, we integrate 16S rRNA amplicon datasets, metagenomes, metatranscriptomes, and annual methane flux data across 9 wetlands, creating the Multi-Omics for Understanding Climate Change (MUCC) v2.0.0 database. This resource is used to link microbiome composition to function and methane emissions, focusing on methane-cycling microbes and the networks driving carbon decomposition. We identify eight methane-cycling genera shared across wetlands and show wetland-specific metabolic interactions in marshes, revealing low connections between methanogens and methanotrophs in high-emitting wetlands. *Methanoregula* emerged as a hub methanogen across networks and is a strong predictor of methane flux. In these wetlands it also displays the functional potential for methylotrophic methanogenesis, highlighting the importance of this pathway in these ecosystems. Collectively, our findings illuminate trends between microbial decomposition networks and methane flux while providing an extensive publicly available database to advance future wetland research.

Methane (CH_4) is a potent greenhouse gas (GHG) contributing to current atmospheric warming[1]. Despite accounting for less than 8% of the land coverage, natural wetlands represent the largest natural source of CH_4 and contribute between 20 and 50% of natural global CH_4 emissions[2–4]. Forecasting CH_4 flux from wetlands remains challenging due to complex interactions between environmental variables such as temperature, soil moisture, and vegetation type, as well as the spatial and temporal variability of CH_4 emissions from wetlands[3,5,6]. Furthermore, the wide array of wetland ecosystems, encompassing peatlands, marshes, swamps, and floodplains, adds complexity to the accurate quantification of CH_4 emissions at a global scale, as each wetland potentially harbors distinct CH_4 production processes and emission rates.

[1]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA. [2]School of Geosciences, University of Louisiana at Lafayette, Lafayette, LA, USA. [3]Department of Biology, Syracuse University, Syracuse, NY, USA. [4]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. [5]Department of Microbiology, The Ohio State University, Columbus, OH, USA. [6]Department of Earth Sciences and Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH, USA. [7]United States Geological Survey, Jamestown, ND, USA. [8]University of Maryland, College Park, MD, USA. [9]Department of Civil, Environmental & Geodetic Engineering, The Ohio State University, Columbus, OH, USA. ✉ e-mail: Kelly.Wrighton@colostate.edu; Mike.Wilkins@colostate.edu

In the saturated soil conditions typical of wetlands, CH_4 generation occurs through an interactive microbial decomposition network that hydrolyzes and ferments plant polymeric material into smaller molecular weight compounds (Supplemental Fig. 1). These compounds serve as substrates for methanogenic archaea, which canonically utilize three distinct metabolic pathways defined by their substrate preference - hydrogenotrophic, acetoclastic, and methylotrophic - for CH_4 production[7]. Microbially derived soil CH_4 can subsequently be emitted to the atmosphere or undergo further microbial oxidation by aerobic or anaerobic methanotrophic bacteria[8]. While this decomposition framework is well-theorized[9,10], the extent to which these microbial members, functional guilds, and overall trophic structure are conserved across different wetlands and their relationships to CH_4 emissions remain unclear.

To bridge this knowledge gap, genome-resolved metagenomics has begun to unveil the identity and metabolic capabilities of microbial communities in wetland soils. This information has uncovered new methanogen and methanotroph genera[11–14], pinpointed relevant functional pathways[15–19], and provided insights into their spatial and temporal relevance[20]. Moreover, metagenomic data from five distinct wetlands[9,10,21] was leveraged to construct microbial carbon decomposition networks, highlighting the microbial guilds and their constituent members involved in CH_4 cycling within these specific sites. While these studies laid valuable groundwork, it is imperative to complement site-specific knowledge with broader-scale analyses for a more comprehensive understanding of wetland microbiomes.

To address this broader sampling need, 16S rRNA gene amplicon sequencing characterizes bacterial and archaeal taxonomy and distribution across wetlands, albeit without providing functional content. This high throughput method allows for more extensive microbial sampling across wetland gradients, capturing microbial dynamics across wetland land coverage types, depth, and seasons[17,22–24]. Integrating knowledge from both marker gene analyses and metagenomics presents an opportunity to achieve comprehensive sampling of microbial conserved features, such as functional potential and network architecture across sites. Linking amplicon sequences to genomes from sampled wetland lineages would enable functional prediction, revealing the blueprints of complex wetland microbiomes at scale and transcending individual wetland boundaries.

We adopted this integrated approach to enable genomic functional predictions for marker gene-identified taxa, to uncover features of soil wetland communities and their association with CH_4 flux across an array of freshwater wetlands. We first analyzed paired amplicon and CH_4 flux data obtained from over a thousand samples collected across nine wetlands from the USA and Sweden (5 marshes, 1 swamp, 1 fen, and 2 bogs), representing a spectrum of CH_4 flux rates as well as ecological and climatic conditions. From this analysis, conserved wetland-wide microbial indicators were linked to a curated genomic catalog encompassing thousands of new and existing metagenome-assembled genomes (MAGs) from wetland soils. This cross-site endeavor revealed a core set of conserved wetland microorganisms, allowing us to elucidate the functional decomposition networks supporting their activity, and delve into the physiological drivers of specific methanogenic taxa associated with high CH_4-emitting wetlands. This study offers a comprehensive, multisite perspective on the microorganisms and processes dictating CH_4 dynamics in wetlands, thereby furnishing actionable insights for advancing scientific understanding and facilitating their translation and integration into climate-scale models.

Results and discussion
Models that rely on abiotic factors have increased uncertainty in high methane-emitting wetlands
Estimates of wetland contributions to the global methane (CH_4) budget often rely on ecosystem-scale models, which do not represent soil microbial metabolism, but instead use abiotic variables (such as air temperature, soil temperature and water content, water table depth/elevation, soil type, and vegetation properties)[25] to approximate environmental states conducive for soil carbon decomposition, methanogenesis, and methanotrophy[20]. A robust meta-analysis from 42 freshwater wetlands showed that air temperature partially accounted for mean annual CH_4 fluxes, explaining 51% of the variance across sites, and was the best individual variable tested[26] (Supplemental Fig. 2). Moreover, a recent analysis using multiple AI models that combined a wide array of environmental[27], hydrological, and ecological input variables to model methane fluxes showed that even the best site-specific model had errors of 20–150 nmol m^{-2} s^{-1} in the majority of the 23 sites studied[28]. This large discrepancy between CH_4 flux predictions and observations for many wetlands hints at a potential role for microbial contributions in explaining these variations[29], a feature we seek to examine in more detail in this study.

To understand unifying microbial features across wetlands and how microbial and geochemical properties relate to CH_4 flux, we conducted a meta-analysis using data from both published and unpublished wetland soil samples. Although differences between sample sizes and discrete depths existed between sites, all wetlands were sampled multiple times allowing for good microbial community representation that could be compared to CH_4 flux and other environmental variables. To qualify for inclusion in our study, sites had to have amplicon sequencing data from at least 12 samples obtained from a minimum of 2 sampling depths and have CH_4 flux measurements. From the original 42 wetlands[26] in the noted earlier study, we identified 16S rRNA gene amplicon microbial data for three of the sites (AmeriFlux, site-ID US-OWC (OWC), AmeriFlux site-ID US-Twt (TWI), AmeriFlux site-ID US-LA2 (LA2)), of which the amplicon data from LA2 is newly released in this study while OWC and TWI utilize previously published data[10,30]. We also expanded the dataset to include CH_4 flux, 16S rRNA gene amplicon, and temperature data from an additional 6 freshwater wetland sites (Jean Lafitte National Historical Park and Preserve (JLA), Prairie Potholes Region Permanent Wetland P7 (PPR P7), Prairie Potholes Region Permanent Wetland P8 (PPR P8), Stordalen Mire Fen (STM-fen), Stordalen Mire Bog (STM-bog), and Spruce and Peatland Responses Under Changing Environments (SPRUCE)). The incorporation of these additional sites reduced the predictive power of mean annual air temperature to explain 37% of the variability across sites (Fig. 1a). Notably, the addition of sites with the highest CH_4 fluxes (PPR P8, PPR P7) (Fig. 1a, b) reveals the limitations of mean annual air temperature in linear models as a predictor of CH_4 flux in high-emitting wetlands, such as Old Woman Creek (OWC) and those within the Prairie Pothole Regional complex (PPR).

We collated and analyzed microbial data from 1112 samples (10% is newly released in this study) from 9 wetlands to demonstrate how incorporating knowledge of CH_4-cycling microorganisms can contribute to improved predictive understanding of these ecosystems (Supplementary Data 1 & 2). Included data was derived from 5 marshes: OWC, PPR P7, PPR P8, LA2, and TWI; 1 swamp: JLA; 2 bogs: SPRUCE and STM-bog and 1 fen: STM-fen. To account for inter-study variability in depth fractions, we binned these samples into three categories: shallow (0–9 cm), mid (10–19 cm), and deep (20–39 cm) (Fig. 1c).

Additionally, we supplemented these data with genomic information creating a cross-wetland genomic catalog, Multi-omics for Understanding Climate Change (MUCC) v2.0.0 database. Here we expanded the original MUCC v1.0.0 genomic catalog, which was composed of 42 metagenome and 133 metatranscriptome samples obtained from a single, high CH_4 emitting marsh (OWC) (Fig. 1a)[10]. The 2507 medium and high-quality MAGs recovered from this wetland sampling were combined with 1529 additional MAGS from previously published palsa, bog, and fen metagenomes from a permafrost thaw gradient at Stordalen Mire (STM, Fig. 1a)[9]. Additionally, we added 50 publicly available MAGs derived from the PPR complex[31] and 43 publicly available MAGs from TWI[30]. Finally, we included 20 new
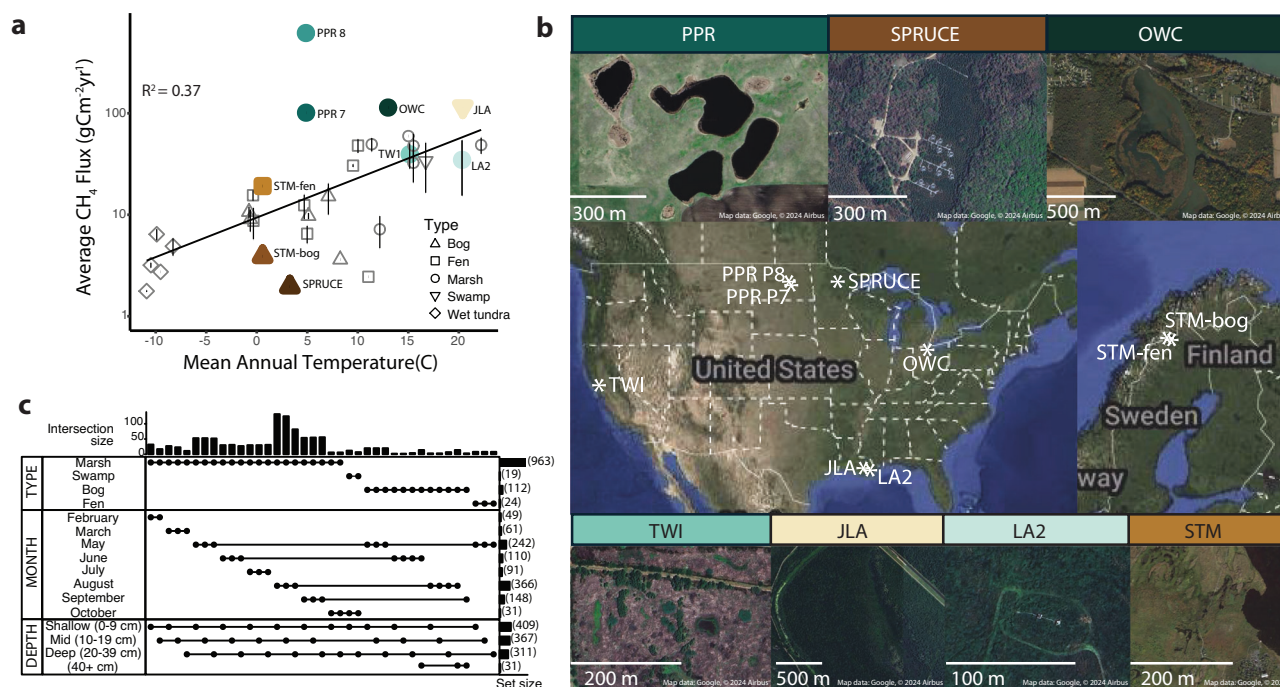
Fig. 1 | 9 freshwater wetlands were examined to determine linkages between microbial communities and predictions of methane flux. a Figure modified from Delwiche et al. [26] shows the mean annual methane (CH_4) flux from wetlands included in FLUXNET-CH_4. The deviation of the predictions from observations indicates this abiotic variable incompletely represented CH_4 flux, especially for the highest emitting wetlands. Colored points represent sites discussed in this study. Methane fluxes vary across an extremely large range, spanning many (3–5) orders of magnitude. For that reason, many statistical analyses of methane flux that compare multiple sites, some with high emissions and some with low emissions, use, and graph methane fluxes at a logarithmic scale. **b** Wetlands differ by type, size, geography, and climatic factors. In this study, we investigated 5 marsh sites (OWC, PPR P7, PPR P8, LA2, and TWI), 1 swamp (JLA), 1 fen (STM-fen), and 2 bogs (STM-bog and SPRUCE). 7 of the sites were found across the United States and 2 were in northern Sweden. Aerial images of each site were acquired from Google Earth. **c** Upset plot indicates the total number of 16S rRNA samples ($n = 1112$ samples) and their distribution across relevant categories including wetland type, sampling month, and sampling depth. Intersection size represents the number of samples found across each combination of wetland type, month, and sample depth scenario and set size represents the total number of samples found within each variable. Source data are provided as a Source Data file.

metagenomes from the PPR complex, LA2, and JLA (349 Gbp of new sequencing), resulting in 617 MAGs released as new data as part of this study. In total, MUCC 2.0.0 contains 3634 high and medium-quality, dereplicated (99% genome identity) MAGs derived from six wetland complexes totaling 8.9 Tb of sequence data (Supplementary Data S3). MUCC v2.0.0 compiles previous wetland genomic datasets and expands genome representation across wetland soils spanning diverse geographies, ultimately increasing database read recruitment and reducing the computational requirements for translating reads to functional content. This wetland-specific genomic resource database was used to connect microbial community profiles with functional potential.

High CH_4-emitting wetlands share microbial community composition and structure

Analyses across wetland sites revealed that wetland type, not geographical location, corresponded to microbial community composition and diversity. As might be expected by ecological wetland differences, bog samples derived from Sweden (STM) and Minnesota (SPRUCE), were more alike one another than bog and fen samples collected within the same wetland complex (STM). Wetlands categorized as marshes or swamps had higher bacterial and archaeal alpha diversity, higher pH, and higher CH_4 flux than bog and fen sites (Supplementary Fig. 3). Additionally, wetland type had a significant impact on community composition, and separation of communities was linked to pH (Fig. 2a & Supplementary Fig. 4, PERMANOVA, $p < 0.001$). Notably, communities in bogs with the lowest pH and CH_4 flux were most distinct from marsh/swamp communities with the

highest pH and CH_4 flux. Fens, with intermediate characteristics of bogs and marshes/swamps such as pH, vegetation, and nutrient levels, hosted microbial communities that were similarly intermediate of the bog and marsh communities[32].

CH_4 flux was loosely correlated with temperature across wetland types but this trend was absent at the level of individual wetland types. In marshes and swamps – the highest CH_4 emitting wetland types – no correlation to temperature was observed ($R^2 = 0.17$, $p = 0.16$) (Supplementary Fig. 5A), suggesting that other factors may be important for predicting CH_4 flux[3,26]. We next assessed the relationships between CH_4 flux and CH_4-cycling microbial community members including methanogens and methanotrophs across sites. Bog and marsh sites hosted different methanogen communities (Supplementary Fig. 6), with bog sites characterized by the dominance of a few methanogens and low relative abundances of acetoclastic methanogens[3,33–35]. For example, *Methanothrix*, an obligate acetoclastic methanogen was significantly more enriched in fen, marsh, and swamp samples than in bog samples. Overall, marsh and swamp sites contained a higher diversity and evenness of methanogen taxa and functional types compared to bog and fen sites. Collectively, the functional potential to utilize more diverse methanogenic substrates in high CH_4 emitting marsh sites could contribute to higher CH_4 fluxes.

To fully understand microbial contributions to the methane cycle, we also assessed the distribution of methanotroph communities across wetland types. Across all sites, aerobic low-affinity methanotrophs were dominant, while the anaerobic methanotrophs assigned to the genus *Methanoperedens* were found only in the three highest methane-emitting sites (OWC, PP7, PP8) (Supplementary Fig. 6). We
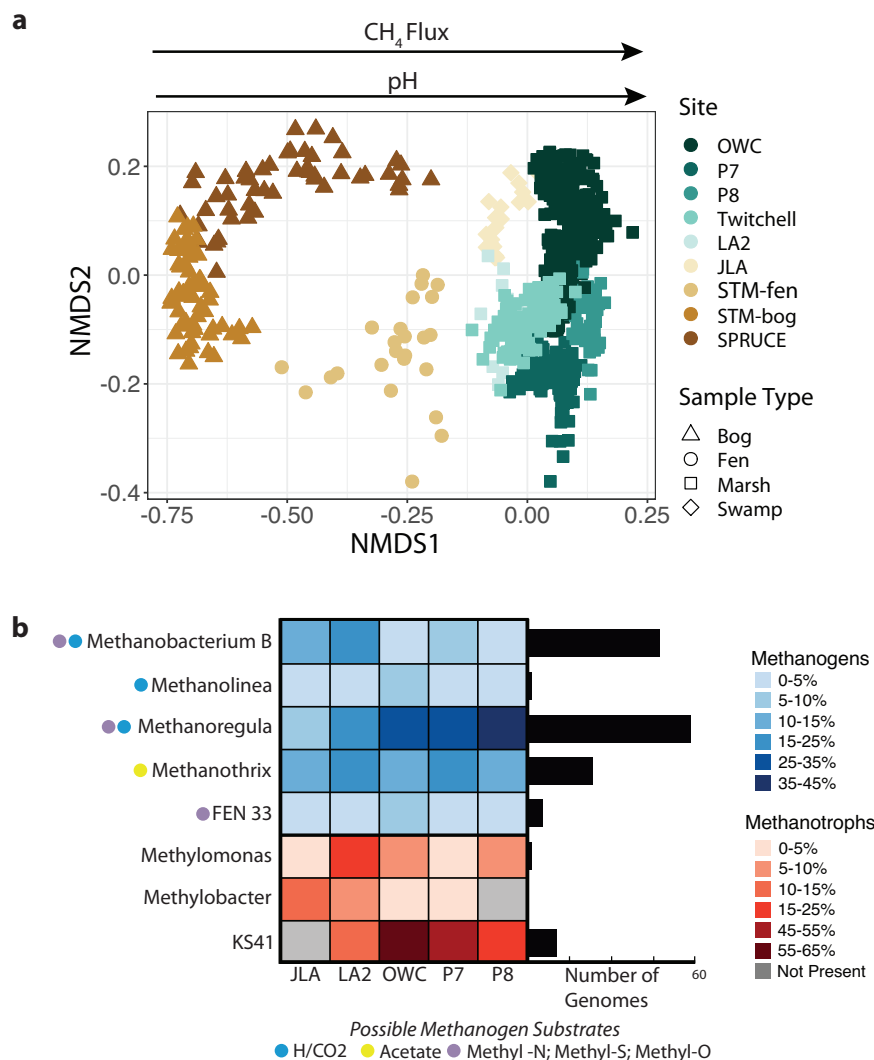
Fig. 2 | Methane-cycling microbes are conserved across wetlands. a Wetland type is an important control on microbiome membership and structure, despite differences in sampling strategies and geographic locations. 16S rRNA amplicon data ($n = 1112$ samples) on soil microbial communities from marsh and swamp samples cluster together (rectangles and diamonds, most right side) and are statistically distinct from fen (triangle, middle, and most left side) and bog (circle, middle) microbial communities. **b** Core methane-cycling members across distinct wetlands calculated from 16S rRNA data. Heatmap shows the relative abundance of each genus within the methanogen (blue) or methanotroph (red) community across wetlands. To illuminate the metabolic features of these core taxa in high methane-emitting wetlands, we utilized the Multi-Omics for Understanding Climate Change (MUCC) v 2.0.0 database, with 140 MAGs assigned to our core taxa. Genome counts per genus are shown in the bar chart (black). MAGs were identified as methanogens if they encoded any genes of the Methyl Coenzyme Reductase (*mcrABG*) and/or Heterodisulfide reductase (*hdrABCDE*) complexes and as methanotrophs if they contain a gene that encodes a methane monooxygenase (Supplementary Data S4). Source data are provided as a Source Data file.

found that the diversity of methanogens ($R^2 = 0.5$, $p = 0.034$), but not methanotrophs ($R^2 = 0.22$, $p = 0.2$), was significantly correlated to CH_4 flux. Additionally, the ratio of methanogen to methanotroph relative abundances was correlated to flux ($R^2 = 0.45$, $p = 0.047$) (Supplementary Fig. 5B), but the relative abundance of methanogens and methanotrophs alone was not. This suggests that the coupling of methanogens and methanotrophs acts as a control over CH_4 flux in wetland environments, highlighting how the balance between these microbial groups likely influences net methane emissions.

Microbial community features (e.g., Shannon diversity) and environmental variables were compared to annual CH_4 flux for the 9 wetlands in this study (Supplemental Fig. 7). The microbial features consistently performed better than the environmental variables tested except for pH, which had the highest correlation to CH_4 flux. Given that pH is known to be a major driver of microbial community structure[36] and correlated with wetland type (Supplemental Fig. 4), we expected pH and CH_4 flux to be correlated. CH_4 flux from wetlands has strong

seasonal trends with higher flux frequently being found in the late summer and early fall[26]. However, many environmental factors (soil moisture, redox, pH) are dynamic, changing rapidly over short timescales, and are thus not readily summarized at an annual scale. Similarly, microbial communities integrate abiotic environmental variables over biologically relevant timescales (moisture, redox, pH), and may therefore be better indicators of the mean conditions allowing for better predictions across coarse time scales.

Identification of a widespread, core group of CH_4 cycling organisms

Given that marshes and swamps have high CH_4 fluxes that are not well explained by methanogen and methanotroph diversity or temperature, we focused on understanding trends in microbial dynamics across 5 marsh and swamp sites (JLA, LA2, OWC, PPR P7, and PPR P8). Additionally, these sites had a more consistent sampling methodology (i.e., similar sampling/sequencing protocols processed by the same

team (see "Methods" section)). We first assessed occupancy patterns across sites using 16S rRNA gene data to identify if there were core methanogens and methanotrophs for these marsh samples, identifying five methanogen genera and three methanotroph genera in at least one sample from each site[37] (Fig. 2b). Despite wetland differences in site, depth, and time of year sampling (Fig. 1), five core methanogen genera were found in a majority of samples: *Methanothrix* (79.7%), *Fen 33* (order *Methanomassiliicoccales*) (72.6%), *Methanobacterium B* (50.9%), *Methanolinea* (55.5%), and *Methanoregula* (93.9%). Interestingly, each methanogenic pathway (hydrogenotrophic, acetoclastic, methylotrophic methanogenesis) was represented within the core community, indicating that all three pathways are consistently important and likely utilized for wetland CH_4 production in high-emitting marsh and swamp ecosystems (Fig. 2b). Three methanotrophs were identified as core but were found in a lower percentage of samples: *Methylomonas* (60.3%), *Methylobacter* (39.8%), and *KS41* (family *Methylomonadaceae*) (85.4%). However, because the core methanotrophs require oxygen for methane oxidation, these methanotrophs may not be as detectable in the deeper anoxic samples sampled here. Constraining our analyses to only the top 10 centimeters of sediment where oxygen might be more available, we found *Methylomonas* present in 75.1%, *Methylobacter* in 57.1%, and *KS41* in 95.2% of samples. Core microbiomes have become increasingly viewed as important because of their assumed role as critical to a given ecosystem's functioning[38,39] and in our study we found that the core members dominated the methane-cycling community. Collectively, these discoveries underscore the pivotal role of select organisms in actively shaping the methane cycle within freshwater marsh ecosystems. These insights carry implications for forthcoming research activities, highlighting these organisms as candidates for more thorough physiological validation and study, as well as focus organisms for scaling to modeling endeavors.

MUCC database enables deeper insight into trophic patterns from co-occurrence networks

For each of the 5 marsh sites, we performed network analysis based on co-occurrence patterns to help unravel possible microbial interactions within these complex, methanogen-oriented communities. We hypothesized that methanogen network structure in wetland communities would act as a predictor of CH_4 flux. To test this hypothesis, we built 16S rRNA gene-positive co-occurrence networks at each site using both the community-wide amplicon data and only the methanogen community data (Supplementary Fig. 8).

Although the network structure of the entire community did not relate to CH_4 flux (Fig. 3k), a more constrained network comprising the significant co-occurrences that included a methanogen member did uncover important trends (Fig. 3l). These networks revealed a negative correlation between the number of methanogen-related network nodes and CH_4 flux, indicating a relationship between less complex methanogen networks and higher annual CH_4 emissions. Furthermore, the number of methanotrophs associated with methanogens in these networks was greater in the lower methane-emitting sites (JLA, LA2), indicating that lower CH_4 fluxes are associated with communities where methanotrophs and methanogens co-occur. In contrast, while high CH_4-emitting sites (OWC, PPR7, PPR8) host methanotrophs and methanogens, they were generally linked by fewer connections (Fig. 3m). Methanotrophs can act as a filter, oxidizing anywhere from 20% to 60% of the CH_4 before it is released into the atmosphere[3,40,41] and these results indicate that their absence in wetland samples where methanogens are present could contribute to greater CH_4 fluxes.

To determine potential microbial interactions that underpin CH_4 production across these sites, we developed metabolic profiles for methanogen-connected taxa in our 16S rRNA gene networks. Utilizing the MUCC v2.0.0 database, we linked microbes present in the networks with MAG representatives and assigned them functional

categories: obligate fermenter, homoacetogen, demethylating, or none of these three criteria (Figs. 3a–e, 4 & Supplementary Data 4). We selected these criteria, as they are thought to cross-feed methanogens (Fig. 1b) and are traits that can be inferred from genomes clearly. Methanogen networks were composed of 699 unique co-associated genera, of which 131 genera had a genome representative in the MUCC database (Fig. 4). Summarizing these genome representatives within the methanogen networks, 12 were categorized as methanogens, 7 as methanotrophs, 23 as obligate fermenters, 8 as homoacetogens, 1 as both obligate fermenter and homoacetogen, and 75 demethylating (methyl-x), and 4 did not meet these criteria (Rules for assignment are found in Supplementary Data 4). Additionally, 6 methanogens and 10 methanotrophs identified based on 16S rRNA gene taxonomy alone (no matches to MUCC, but metabolism is defined in literature) were included in the networks (Fig. 4, Supplementary Fig. 8).

Specifically, obligate fermenters have the potential to produce acetate, formate, and H_2, which we hypothesized would directly promote methanogen activity[42,43] and thus be positively associated with our methanogen networks. As we expected, obligate fermenters were highly connected to hydrogenotrophic and acetoclastic methanogens, likely supporting cross-feeding. In total, obligate fermenters had 99 significant interactions with methanogens of which 73% were with hydrogenotrophic or acetoclastic methanogens (Fig. 3f–j). Additionally, obligate fermenters were found to highly co-occur with certain methylotrophic methanogens such as *Methanofastidiosum*, which requires H_2 to reduce methylated thiol to form methane. Compared to hydrogenotrophic methanogenesis, this form of methanogenesis is more thermodynamically favorable under low H_2 conditions and has been proposed to support H_2-producing syntrophs and fermenters by preventing the accumulation of H_2[12]. In summary, anoxic carbon exchanges between obligate fermenters and methanogens appear vital to carbon cycling in wetlands.

Syntrophy denotes a symbiotic interaction among diverse microorganisms, wherein the exchange of metabolic byproducts mutually supports each organism's metabolism. This phenomenon is particularly prominent in methanogenic environments, where methanogens play a crucial role in regulating product concentrations, thereby rendering otherwise endergonic processes thermodynamically favorable[44,45]. In our study, we investigated obligate fermenters to uncover evidence of secondary fermentative syntrophs, identifying two prevalent syntrophic genera across methanogen networks: *Smithella*, present in four marshes except PPR8, and *Syntrophorhabdus*, found across all five marsh networks. Previous research has demonstrated the capacity for acetate and hydrogen production by *Syntrophorhabdus*[46], aligning with our genome-based characterization of these 7 MAGs in MUCC. Notably, in our networks, *Syntrophorhabdus* exhibited multiple (8) connections to hydrogenotrophs and acetoclasts, further emphasizing its role in microbial metabolic exchanges. These genomic metabolic insights highlight the intricate connections harbored within these co-association networks, exchanges essential for maintaining metabolic efficiency in methanogenic environments.

Homoacetogens also interact with methanogens, as these microorganisms grow on $H_2/CO_2/CO$ and produce acetate as the main metabolic product. We hypothesized that these organisms could cross-feed acetoclastic methanogens[15] and could compete with hydrogenotrophic methanogens for substrates[47]. The 9 homoacetogen MAGs identified in the methanogen networks comprised 15 nodes and were closely related across sites, belonging to two main phyla, Desulfobacterota and Chloroflexota despite many other acetogens across other phyla existing in the MUCC database. We observed 32 associations between these acetogens and methanogens, with 50% to hydrogenotrophic, 28% to acetoclastic, and 22% to methylotrophic methanogens. Additionally, 6 of the 8 acetoclastic methanogens had at least one connection to an acetogen, supporting our hypothesis that
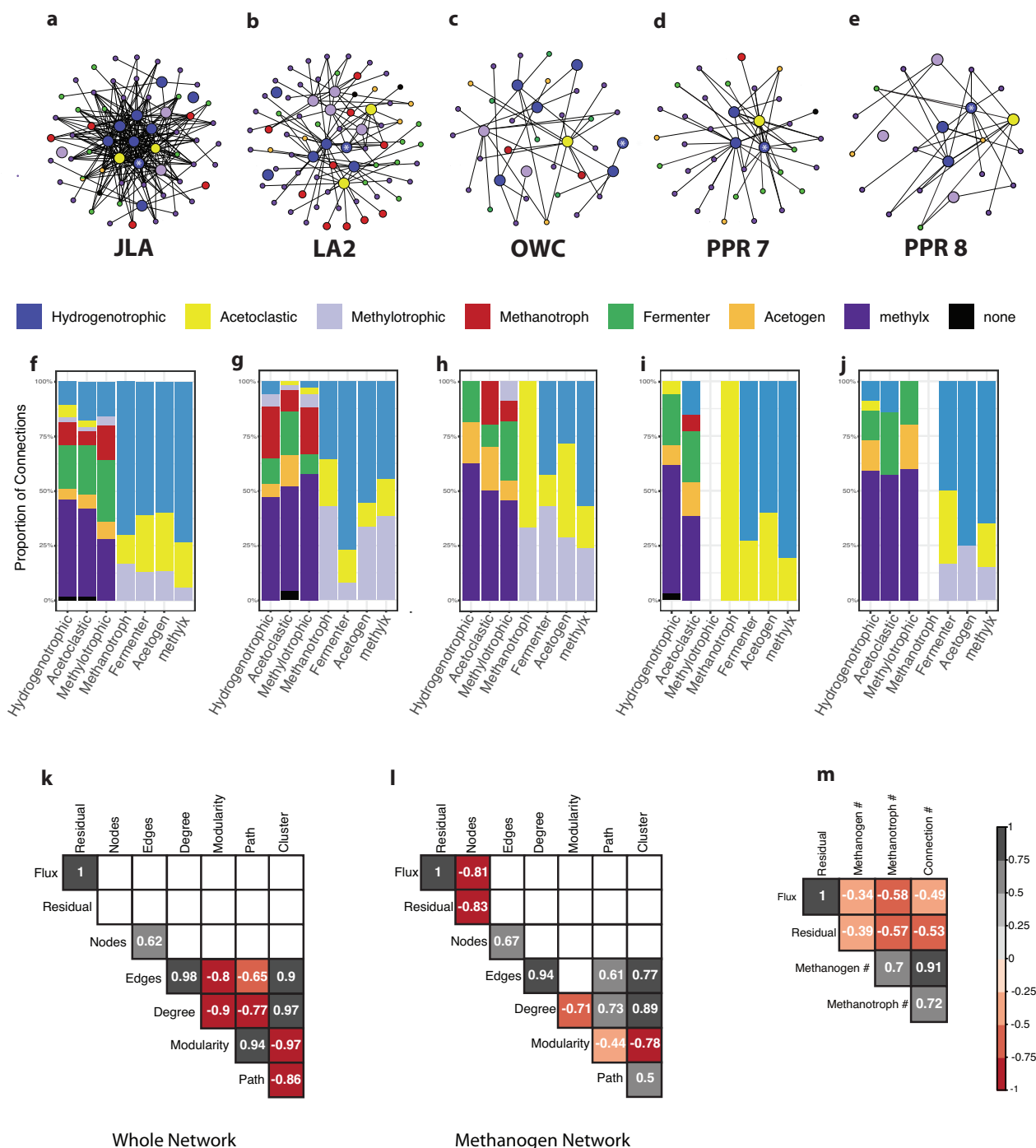
Fig. 3 | Methanogen co-occurrence patterns are related to methane flux.
a–**e** Co-occurrence network analysis revealed the network structure of methanogen-associated taxa across wetlands ($n = 12$ samples per site). Networks depicting site-specific co-occurrence analysis uncovered the network of microorganisms coordinated to methanogens across each site, with nodes representing microbial taxa. Larger nodes represent methanogens, while small nodes represent bacterial taxa. Nodes are colored by the inferred metabolic potential of 16S rRNA-linked MAGs within MUCC. White stars indicate the *Methanoregula* node in each network. Source data are provided as a Source Data file. **f**–**j** The Proportion of

connections between groups in each network is given in the bar charts and shows conserved patterns in network connections across sites. Missing bars indicate no connections. Correlation between network statistics and CH_4 flux measurements derived from the Ameriflux network was measured for **k** whole community networks and **l** methanogen networks. Only the number of nodes in the methanogen network was correlated with methane flux. **m** Additionally, a negative correlation between annual CH_4 residual and CH_4 flux (from Fig. 1) to the number of methanogens, methanotrophs, and connections between the two were observed.

acetogens were cross-feeding methanogens. While our finding does not preclude competition between hydrogenotrophs and other acetogens, these identified positive associations may reflect sufficient hydrogen production within the soil profile to support the co-existence of both guilds or the separation of guilds across microsites.

Finally, demethylating microorganisms, whether bacteria or archaea, are capable of removing methyl groups from oxygen, sulfur, and nitrogen (O, S, N) containing compounds. Unlike methylotrophic methanogens, these taxa do not produce methane directly; however, they may engage in cross-feeding or competition dynamics with
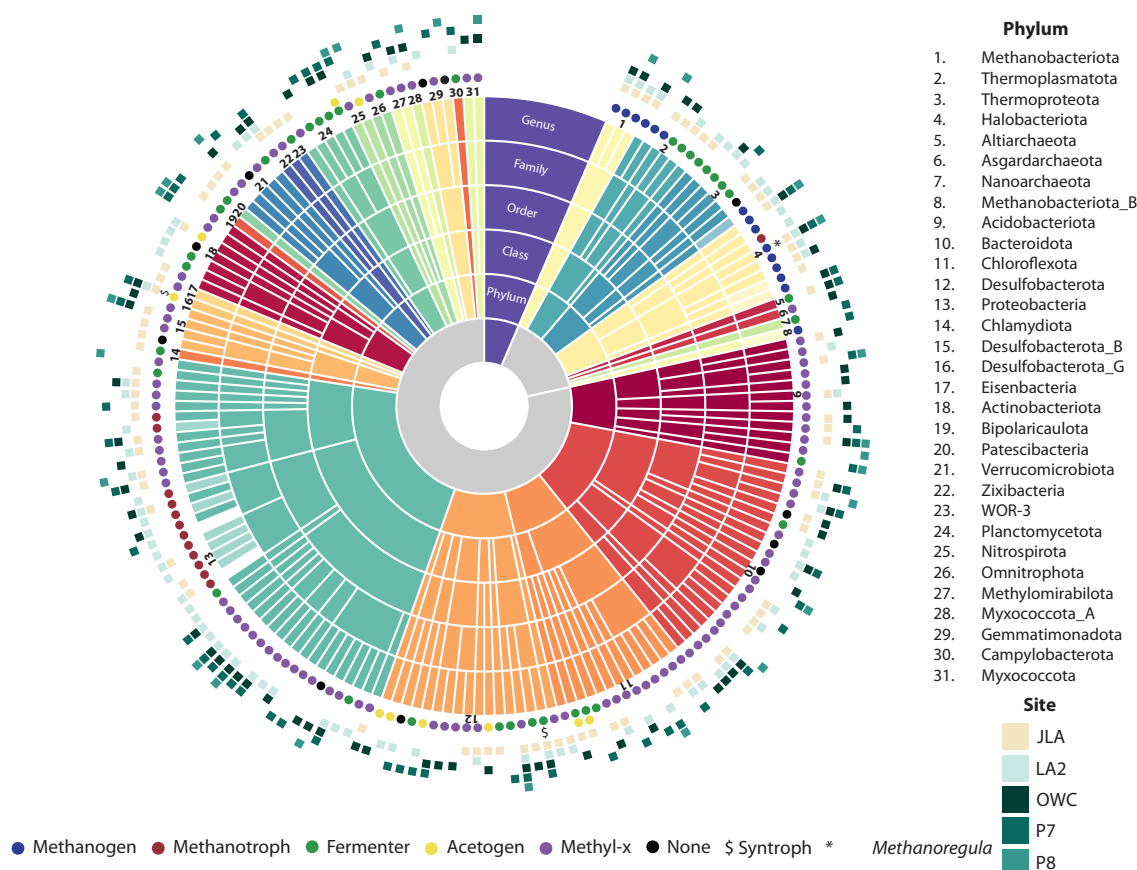
Fig. 4 | MUCC is a comprehensive cross-wetland database useful for multi-omics wetland studies. Taxonomy of the 158 genera represented in the networks that are found within the MUCC database. Additionally, 6 methanogens and 9 methanotrophs were identified based on 16S rRNA included in the networks and are shown in the network with reduced opacity at the genus level. Circles around the edge represent inferred metabolic potential and squares represent the sites where the genus had significant co-occurrence with a methanogen.

methylotrophic methanogens. Depending on the enzymatic systems they encode, these microorganisms can lead to several outcomes: (i) production of trimethylamine (TMA), a substrate for certain methanogens; (ii) formation of quaternary amines (QA), which can be utilized by select methylotrophic methanogens; or (iii) direct utilization of methylated O, N, or S compounds, which may (iiia) compete with methylotrophic methanogens or (iiib) generate acetate and hydrogen to support hydrogenotrophic or acetoclastic methanogens. The methyl-metabolism category exhibited substantial connectivity with methanogens, comprising nearly half of the connections across sites. Notably, 68% of these connections (comprised mostly of type iii demethylating microorganisms) were linked to acetoclastic and hydrogenotrophic methanogens rather than methylotrophs, suggesting that demethylating metabolisms in soils could indirectly bolster non-methylotrophic methane production. These findings underscore the complexity of microbial interactions beyond methane production and oxidation, thereby contributing to a more comprehensive understanding of microbial cross-feeding and its broader implications for methane emissions.

Methanoregula is critical for CH$_4$ production in wetlands

Two core methanogens (Fig. 2), *Methanothrix* and *Methanoregula*, were the only methanogens found in networks across every site indicating global importance in the wetland CH$_4$ cycle. *Methanothrix* is an obligate acetoclastic methanogen already shown to be globally distributed and an important contributor to CH$_4$ emissions in wetlands[16]. *Methanoregula* has been found in wetlands and other habitats around the world, and like at many of our sites, is a

prominent member of methanogenic networks and consistently a dominant methanogen[48,49]. We found that its dominance (proportion of methanogens that are *Methanoregula*) was related to CH$_4$ flux, such that the percent of methanogens that are *Methanoregula* significantly correlated to CH$_4$ flux and the residual values that were not well predicted from the temperature- CH$_4$ flux correlation in Fig. 1 (Fig. 5a). Additionally, we tested how well temperature, *Methanoregula* dominance, and the two combined explained methane flux. When looking at the 9 study sites, CH$_4$ flux was not predicted by temperature alone ($R^2 = 0.15$, $p = 0.30$,), and was predicted by *Methanoregula* dominance ($R^2 = 0.54$, $p = 0.02$), but that temperature combined with *Methanoregula* dominance was the best predictor ($R^2 = 0.84$, $p = 0.02$). This is one example of how incorporating biological insights with already existing abiotic data could improve the predictive power of climate models.

To understand potential physiological drivers that link *Methanoregula* and predications of CH$_4$ flux, we conducted a genomic analysis of 107 dereplicated MUCC-derived and publicly available (i.e., GTDB, JGI) MAGs. *Methanoregula* encoded diverse metabolic strategies, the capacity for fixing nitrogen (nitrogenase), viral defense (CRISPR-Cas), and mechanisms to respond to fluctuating redox conditions (reactive oxygen species) (Fig. 5b). *Methanoregula* is classically designated hydrogenotrophic[50], which we broadly confirmed here (Fig. 5b). We also report that some *Methanoregula* genomes encode genes for methylotrophic methanogenesis, specifically for the demethylation of methylated sulfides[51] and methoxylated[19] compounds that are prevalent in wetlands[10,15]. Although hydrogenotrophic methanogenesis is generally recognized as the dominant CH$_4$-generating
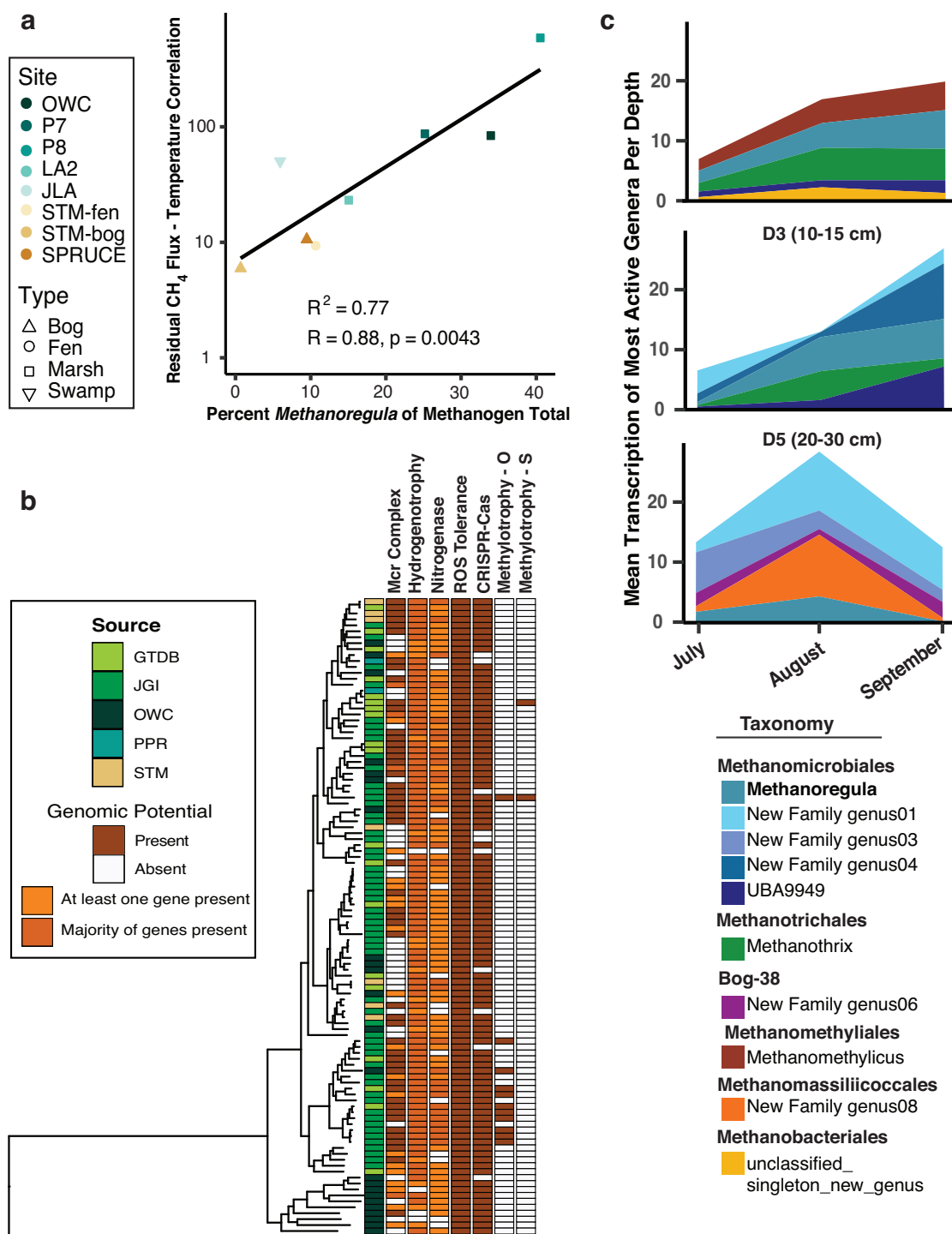
Fig. 5 | *Methanoregula* presence and activity are important predictors of methane flux. **a** A linear regression comparing the residual values from the methane flux to the temperature trend line showed a significant positive relationship to the relative abundance of *Methanoregula* within the methanogen community. **b** Genome tree of *Methanoregula* MAGs from MUCC (OWC, PPR, STM), plus available MAGS from JGI and GTDB (*n* = 149 *Methanoregula* MAGs). A pangenome analysis shows the largely conserved encoding of genes for key physiological features, as well as limited novel metabolic potential (e.g., methylotrophic genes) which may directly or indirectly support high methane fluxes from *Methanoregula* in wetlands. **c** Mean transcription of top five most active methanogenic genera at three depths (0–5 cm, 10–15 cm, 20–30 cm) in the mud site type across the 2018 sampling season predictive of CH_4 fluxes (*n* = 43 metatranscriptomes).

pathway in wetlands, recent studies have indicated that methylotrophic methanogenesis contributes more to CH_4 flux than previously realized[17,21,22,52]. Therefore, the apparent significance of *Methanoregula* in contributing to CH_4 emissions across diverse wetlands and within wetland gradients could partly be explained by a broader than previously understood ecological niche.

To investigate the role of *Methanoregula* within a high CH_4 emitting wetland, we mined a previously undefined role for *Methanoregula* from 39 paired metatranscriptome and metabolome datasets across spatial and temporal gradients from a single mudflat at OWC[10] (Supplementary Fig. 9A). At this mud-type site, a *Methanoregula* MAG (OWC-0053) was one of the transcriptionally most active methanogens

throughout the entire soil column across 3 months of peak CH_4 production (Fig. 5c). This genome was also one of the 9 genomes that predicted 78% of soil porewater CH_4 concentration (Supplementary Fig. 9B). In summary, our comprehensive analysis reveals *Methanoregula's* substantial contribution to CH_4 dynamics within a high-emission wetland, highlighting its prominent role as a key player in CH_4 production across spatial and temporal scales.

These findings help in part explain the significant correlation between *Methanoregula* abundance and CH_4 flux across wetlands, and its role in marsh CH_4 networks. Our results suggest that *Methanoregula* may possess a broader physiological capacity to produce CH_4 through both hydrogenotrophic and methylotrophic methanogenesis and adapt to various abiotic and biotic constraints present in marsh soils. By shedding light on the functional significance of *Methanoregula*, a core taxon across wetlands, our study contributes to advancing our understanding of wetland CH_4 emissions. Our findings use a cross-site analysis to identify core lineages, like *Methanoregula*, warranting further physiological exploration, as the metabolic assumptions may be constrained by prior strict substrate and redox capabilities. Ultimately our results show promise for biological knowledge to enhance predictive models of wetland emissions, ultimately facilitating more effective management and mitigation strategies.

In conclusion, microbial processes related to CH_4 flux have been well-characterized at a handful of individual sites. However, site-specific knowledge of wetland microbiomes suffers from limited generalizability, as wetland ecosystems vary widely. Therefore, insights gained from studying microbiomes in one wetland may not necessarily apply to others, restricting the broader understanding of wetland microbial communities and their roles in ecosystem processes. Here, we build on existing single-site studies by building a multisite wetlands database and synthesizing decomposer and CH_4-cycling networks and their relation to CH_4 flux data across multiple wetland ecosystems. Using data from several previously published and unpublished studies, we were able to link 16S rRNA gene data to genomes from the MUCC database and develop metabolic profiles for methanogen-connected taxa. Despite inconsistent sampling strategies and sequencing methods, we were still able to detect multisite trends across all wetland types and when focused exclusively on the higher CH_4 emitting marsh and swamp wetlands. We found microbial cross-feeding has broad implications for CH_4 emissions across wetland environments. Additionally, the highest CH_4-emitting wetlands had the fewest methanogen network connections, suggesting streamlined microbial metabolic circuits may contribute to enhanced CH_4 production across wetland soils. Finally, we revealed that *Methanoregula* is a key contributor to CH_4 flux in wetland environments, potentially due in part to previously unknown microbial metabolic versatility. This study shows that broad microbial trends across wetlands can be better predictors of CH_4 flux than environmental variables alone. However, due to the opportunistic nature of this study, trends in temporal, seasonal, and small-scale spatial variability across and within wetlands were not fully explored despite their importance in known variation of CH_4 flux. While this work cannot directly inform climate models, it demonstrates that microbial community features can explain variation in CH_4 flux across diverse wetlands. Future work spanning more wetlands across broader geographic regions and wetland types with conserved sampling strategies that account for seasonal dynamics will enable a better understanding of the microbial drivers of CH_4 flux in freshwater wetlands across time and space, helping to create better prediction models. The MUCC database will benefit these future studies, reducing the need to generate new multi-omic datasets for understanding microbial activity underlying CH_4 production in these environments. Ultimately, MUCC is a powerful microbiome tool enabling us to decode microbial organismal and metabolic patterns across multiple environments, with the goal of improving predictive modeling frameworks.

Methods

Multi-Omics for Understanding Climate Change (MUCC) v2.0.0 Database

Data was compiled from 9 different wetlands (5 marshes, 1 swamp, 1 fen, and 2 bogs), including both previously published and unpublished datasets. Published data were sourced from AmeriFlux, site-ID US-OWC (OWC), AmeriFlux site-ID US-Twt (TWI), and Spruce and Peatland Responses Under Changing Environments (SPRUCE); both published and unpublished data was compiled from Prairie Potholes Region Permanent Wetland P7 (PPR P7), Prairie Potholes Region Permanent Wetland P8 (PPR P8), and Stordalen Mire (STM-fen and STM-bog); and unpublished data were collected from Jean Lafitte National Historical Park and Preserve (JLA) and AmeriFlux site US-LA2 (LA2). The Multi-Omics for Understanding Climate Change (MUCC) v2.0.0 database combines 997 16S rRNA, 284 metagenomic, and 133 metatranscriptomic datasets from PPR7, PPR8, STM-bog, STM-fen, OWC, TWI, and SPRUCE, along with 115 newly analyzed 16S rRNA and 20 metagenomic samples from PPR P7, PPR P8, JLA, and LA2. Because samples are a mixture of previously published and new samples from several different research groups, extraction methods, sequencing primers, and sequencing facilities are not uniform across all sites. All methods for DNA extraction and amplicon sequencing information for all sites can be found in Supplementary Data 7. While methodological differences can result in biases, trends observed across sites with different extraction protocols and sequencing primers transcend potential methodological biases. Similar conclusions were validated in Delgado-Baquerizo et al. [53]. Additionally, in this study, we used eddy-covariance (EC) towers or chamber measurements depending on what was available at each site. However, our team has quantitatively compared chamber and EC-flux measurements in the same sites and found that the two methods are generally comparable, though chamber measurements have a higher degree of uncertainty[54].

Accession numbers for all samples can be found in Supplementary Data 1, while sample IDs and GTDBk v207 taxonomy for 16S rRNA data are in Supplementary Data 2, and the details of 4745 medium and high-quality Metagenome-Assembled Genomes (MAGs) are listed in Supplementary Data 3. The MAGs and 16S rRNA data from MUCC v2.0.0 are available on Zenodo (https://doi.org/10.5281/zenodo.14532347) and NCBI (PRJNA1007388).

Old Woman Creek (OWC). OWC National Estuarine Research Reserve (41° 22′N 82°30′W) is a 573-acre freshwater marsh located on the southern shore of Lake Erie in Ohio. It is composed of a permanently flooded channel surrounded by marsh, occasional mud flats (which are inundated most of the time), and an upland forested habitat[16]. In brief, sediment cores were collected from sites representing distinct eco-hydrological patch types (cattail plant, mud, and open water) in triplicate in May, June, July, August, and September of 2018 using a modified Mooring System soil corer[16]. Cores, sampled to a depth of 35 cm, were sub-sectioned into six depths using a hydraulic extruder: 0–5 cm, 5–10 cm, 10–15 cm, 15–20 cm, 20–25 cm, and 25–30 cm. Microbiome data from 626 samples included bacterial and archaeal 16S rRNA amplicon sequence data, metagenomes, and metatranscriptomes[10,16]. Meteorological and eddy-covariance flux data for the site are available through AmeriFlux, site-ID US-OWC[55]. Gap-filled and averaged data used in this analysis were obtained from FLUXNET-CH4[26].

Prairie Pothole Region (PPR). Cottonwood Lake Study Area (47° 05′N: 99° 06′W), located northwest of Jamestown, North Dakota, is a protected area owned by U. S. Fish and Wildlife Service and is a long-term research site (>30 years) for the U.S. Geological Survey (USGS). The 92-ha site consists of 17 distinct wetlands with permanent-to-temporary inundation. Samples were collected from open waters of two semi-permanent wetlands: P8 (47° 05′55.8″N 99°06′14.1″W) and 2 sub-

locations within P7 - Location 1 (47°05'43.7"N 99°06'00.8"W) and Location 2 (47°05'46.7"N 99°05'57.9"W). P7 and P8 are depressional wetlands with open waters (<1 m depth) in their centers and emergent vegetation (*Typha*) along their boundaries. Cores were collected in triplicate at each location in March, May, and September of 2015 using a modified Mooring System soil corer. Cores, sampled to a depth of 30 cm, were sub-sectioned using hydraulic extrusion in 3-cm increments. MUCC v 2.0.0 included 214 16S rRNA sequencing samples and 18 previously published metagenomes[24] combined with 18 new metagenomes from PPR.

Annual CH_4 flux data was averaged from 2011 to 2016[56]. Methane fluxes were measured using the static chamber method[57] every two weeks during the growing season (defined as soil temperature ≥5 °C). During each sampling event, chambers were floated in open waters of P7 and P8 for 30 min after which headspace gas samples were collected through a rubber septum and stored in evacuated 10-ml serum vials. Sample gases were analyzed for methane concentrations on a gas chromatograph equipped with electron capture and flame ionization detectors (SRI Model 8610, California) located at the USGS Northern Prairie Wildlife Research Center. Methane flux rates were calculated using the linear change in CH_4 concentration during the deployment, chamber dimensions and temperature, and the Ideal Gas Law. Biweekly flux rates were scaled to annual cumulative CH_4 flux by summing the mean flux rates between consecutive sampling events and multiplying by the time between events.

Louisiana Wetlands (JLA and LA2). Two distinct sites were sampled in Louisiana in October 2021. Jean Lafitte National Historical Park and Preserve (JLA) (29°48'06.6"N, 90°06'37.0"W) and AmeriFlux site-ID US-LA2[58] (LA2) (29°51'31.4" N, 90°17'11.3" W) on the Salvador Wildlife Management Area are located in coastal Louisiana. The JLA wetland is a Cypress-Tupelo swamp with distinct hollow and hummock features, and the LA2 wetland is a fresh flotant marsh vegetated by a mix of *Typha sp.* and *Sagittaria sp*. In JLA, triplicate soil cores were collected using a Russian Peat Corer, and 0–10 cm and 30–40 cm intervals were sampled at two distinct ecosites within the swamp. In LA2, triplicate slurry samples from 0 to 10 cm and 20 to 30 cm were collected using a sipper at two distinct locations within the marsh. In all ecosites, samples were collected in three 5 × 5 m plots.

Samples were kept on dry ice after processing. DNA was extracted using Zymo Research Quick-DNA™ Fecal/Soil Microbe Microprep Kit, following the manufacturer's protocol. Amplicon libraries were prepared using a single-step PCR to amplify the V4 region of the 16S rRNA gene with the primers 515F/806R[59] following the Earth Microbiome Project (EMP) PCR protocol. Pooled DNA products were sequenced on the Illumina MiSeq Platform using 251 bp paired-end sequencing chemistry at the Microbial Community Sequencing Lab (University of Colorado Boulder).

Gap-filled and averaged flux data for LA2 that were used here, were downloaded from FLUXNET-CH4[26], while JLA flux was measured in four field campaigns in June, August, October, and December of 2021. Measurements were conducted using a trace gas analyzer (LICOR 7810) coupled to a custom-made chamber in triplicate 2-min deployments in three hollow and three hummock locations. The concentration time series of each chamber was used to calculate diffusive and ebullitive fluxes separately[60]. Briefly, an empirical bubbling threshold was set using the maximum observed change in methane concentrations over ~1 s time-steps in the chambers that exhibited a monotonic increase, which is characteristic of diffusive-dominated measurements. Changes in concentration over the threshold were deemed as ebullitive events. We identified these in each chamber time series and isolated them. Then, the corresponding changes in concentration were added to produce the ebullitive flux during the chamber measurement. The other segments of the dataset, with a monotonic increase or decrease, were fitted on a one-dimensional diffusion model to

calculate the diffusive fluxes. In this study, we used the sum of the diffusive and ebullitive fluxes, accounting for the fluxes at the surface of the soil-water column.

Twitchell. AmeriFlux, site-ID US-Twt on Twitchell Island (TWI) (121.65°W, 38.11°N) is a marsh located in the Sacramento-San Joaquin River Delta, CA. TWI is a USGS wetland restoration site, which was created by flooding two agricultural plots in 1997[61]. All data used from the TWI site were previously published in He et al. [30]. In brief, samples were collected in February and August of 2011 at three distinct sites within the wetland. Cores were sampled to a depth of 25 cm and sectioned into two 12 cm subsections. Meteorological and flux data for the site are available through AmeriFlux, site-ID US-Twt[62]. The TWI experimental wetlands are categorized as freshwater marshes. Flux data was downloaded from FLUXNET-CH4[26].

SPRUCE. The SPRUCE experiment (47°30.4760'N; 93°27.1620'W), located in a forested temperate peat bog (S1 bog) of the US Department Agriculture (USDA) Forest Service's Marcell Experimental Forest, is located northeast of Grand Rapids, Minnesota. The S1 bog is a perched ombrotrophic bog that is acidic (pH 3.5–4.0), oligotrophic, and contains a hummock/hollow microtopography where the water table typically fluctuates within the top 30 cm of peat during a typical year. Vegetation at the S1 bog is dominated by peat mosses (*Sphagnum* spp.), shrubs, black spruce, eastern tamarack, some graminoids, and forbs[35,52]. All data used was published in Wilson et al. [22]. In brief, 200 cm soil core samples were collected and sectioned at 11 intervals (0–10, 10–20, 20–30, 30–40, 40–50, 50–75, 75–100, 100–125, 125–150, 150–175, and 175–200 cm) at each experimental plot. Only data from samples collected from +0 and ambient treatments and samples in the top 40 cm of the soil core were retained for the analyses conducted in this paper.

Stordalen Mire (Stm). Stordalen Mire (0°34'25.7"N; 37°34'30.1"E) located near Abisko, Sweden is an Arctic permafrost peatland that covers three main habitats across a discontinuous thaw gradient: palsa, bog, and fen. Palsa overlays intact permafrost and is well-drained and dominated by woody and ericaceous shrubs. Bog overlays partially thawed permafrost, with a perched water table and *Sphagnum* moss dominance. Fen is fully thawed, inundated, and sedge-dominated. The Mire was surveyed in May 2015 at a range of distributed palsas (18 sites), bogs (20 sites), and fens (8 sites); only bog and fen 16S rRNA gene amplicon data are used in this study. A serrated knife was used to cut vertically into the peat, and microbial samples were collected to fill 2 ml Eppendorf tubes from each depth: shallow (median of 2 cm, range 1–3 cm); middle (median of 12 cm, range 10–12 cm); and deep (median of 20 cm, range 18–20 cm). Sample tubes were stored on ice in the field and transferred to −80 °C within 10 h of collection. DNA was extracted with the PowerSoil 96-Well Soil DNA Isolation kit (MO BIO cat# 12955-4) following the manufacturer's protocol. 16S rRNA gene amplicon sequencing was performed by Argonne National Laboratory using the Earth Microbiome Project barcoded 515F-806R primer set and protocol and on an Illumina MiSeq sequencer. MAGs from 214 previously published metagenomes were also used (Supplemental Data 1)[9]. Methane flux data for Stordalen bogs and fens were annual averages from 2012 to 2018 of autochamber measurements (static, closed systems) that include three replicate measurements per cover type[63].

16S rRNA gene sequencing and analysis
All raw amplicon sequence data were processed using the QIIME2 (v2021.2) pipeline[64]. Data from OWC, PPR P8, PPR P7, LA2, JLA, STM-fen, STM-bog, and SPRUCE sites were independently processed through QIIME2 to account for sequencing run biases. Datasets were uniformly trimmed to the same length (195 bp), paired-end reads were merged, and ASVs were assigned using the naïve Bayes sklearn

classifier trained with GTDB database version 207[65], prior to merging at the ASV level across datasets. Because TWI was sequenced using a different primer set, sites were merged at the genus level. Due to a wide range in sequencing depth across sites, all samples were rarefied to 5000 reads resulting in a final dataset of 1112 samples (Fig. 1c). 43 samples were not retained because they fell below the minimum read depth. Across the 9 wetlands included in this study, core depth and interval sections varied. The compiled studies had different depth thresholds used to categorize shallow, middle, and deep sediments. To standardize depth measurements, we created 3 categories that encompassed the categories across studies: shallow included samples in the 0–9 cm horizon, middle included samples collected from 10 to 19 cm, and deep for samples collected from 20 to 40 cm.

Genome assembly and binning

Previously published metagenomic samples were combined with newly analyzed samples in this release of MUCC. 20 newly analyzed samples contributed 617 MAGs (Supplementary Data 1 & 3). MAGs were recovered from:

(1) 2021 LA2 Field Sample ($n = 1$)
(2) 2021 JLA Field Sample ($n = 1$)
(3) 2022 PPR Field Sediment Samples ($n = 7$)
(4) 2022 PPR Field Water Samples ($n = 2$)
(5) 2022 PPR Lab Enrichment Samples ($n = 9$)

LA and JLA metagenomes were processed separately from the PPR metagenomes. Raw metagenomic reads were trimmed using Sickle (pe)[66] and assemblies were generated using Megahit (v1.2.9)[67] with parameters --k-min 31 --k max 121 --k-step 10. Subsampled assemblies using 25% of sequencing reads were generated using IDBA-UD v 1.1.3[68] with default parameters. Reads were mapped to contigs greater than 2500 bp using BBMap (v 38.89)[69] and were subsequently binned using MetaBAT[70]. Only medium and high-quality bins based on adapted MIMARKS standards (completeness >=50% and contamination <10%) were retained[71]. PPR bins from these assemblies were combined with bins from metagenomic assemblies derived from the earlier sampling of PPR[31], were combined with the bins from LA2 and JLA, and with publicly available bins from OWC[10], STM[9], and TWI[30]. This bin pool was dereplicated using dRep (v 3.0.0)[72] at 99% identity. MAG completeness and contamination were estimated using CheckM[73] and taxonomy was assigned using GTDB-tk v2.3.0 with GDTB database release 207[65].

Community analysis

To determine the extent to which microbial community structure varied with both wetland type (marsh, swamp, fen, bog) and sample depth (shallow, mid, deep), we conducted a permutational analysis of variance (PERMANOVA) using Bray–Curtis distances. Results were visualized using non-metric multidimensional scaling (NMDS). PERMANOVA and NMDS were conducted using the *vegan* package[74] and visualized using *ggplot2*[75] in R Studio (version 4.0.2)[75]. We also correlated environmental parameters including pH, mean annual temperature, mean annual precipitation, latitude, longitude, and CH_4 flux with microbial community structure using the R-function "*envfit*" (as visualized in Supplementary Fig. 2). Alpha diversity of the entire microbial community, of methanotrophs and methanogens, of the methanogens only, and of the methanotrophs only was calculated using the Shannon diversity index. Differences in alpha diversity based on wetland type were calculated using a Kruskal–Wallis H test. Post hoc analysis was performed using the Pairwise Wilcox test with a Benjamini–Hochberg procedure for p-value correction using the function "*pairwise.wilcox.test*" in R. Marshes and swamps were grouped together because they have similar characteristics to each other such as pH while bog and fen were grouped because they are both types of peatland characterized by low pH and occur in similar climates[76]. Shannon diversity was correlated with individual

environmental parameters using linear regression and *corrplot* in R. Linear models were used to assess if the mean annual temperature (MAT) and/or relative abundance of *Methanoregula* was predictive of methane flux across wetlands using the "*lm*" function in R. MAT and *Methanoregula* relative abundance were also individually tested using a regression model conducted using the R package *ggpubr*[77].

To determine if certain methanogens and methanotrophs were widespread (found across all sites) or restricted to specific wetland types (i.e., marsh), we conducted a core community analysis. This analysis was conducted across all samples both regardless of sample depth, and within the depth categorization to understand if core members are more likely to be present in different depth zones. Because of the wide range of sampling schemes across sites, a microbe was determined to be a core member if it was present across all sites or all sites within a categorization (marsh/swamp or bog/fen). Core analysis was preformed using '*summarise*' and '*filter*' functions in *Tidyverse*[78]. The percentage that each core member was of the methanogen or methanotroph community was then calculated by dividing the 16S rRNA gene counts of each core member by the total count number of all methanogens or methanotrophs in each sample allowing us to determine if core community members were the dominant methanogen and methanotroph genera in each wetland.

Co-occurrence networks

To understand if co-occurrence patterns are related to methane flux, we created co-occurrence networks based on the entire community and significant co-occurrence patterns with methanogens from JLA, LA2, OWC, PPR P7, and PPR P8. We focused on these five marsh sites because we were interested in patterns within the highest methane-producing communities and because these all used the same amplicon primers. Because networks are sensitive to the number of input samples, each individual site's network was composed of 12 different community samples that were randomly sampled using the '*Sample_n*' function in *dplyr*[79] where each of the 3 sampling depths was represented by 4 samples. In order to minimize seasonal variation in methane-cycling communities, all samples in the networks came from samples collected at the same time of year (September or October) when CH_4 flux is historically highest[26]. The number of total samples represented in the networks from each site is as follows: LA2: 100%; JLA: 63%; OWC: 2%; P7: 9%; and PPR P8: 14%.

Network analyses were carried out in R using the packages *igraph*[80], *Hmisc*[81], and *Matrix*[82]. To determine co-occurrence patterns in the microbial communities, we used rarefied genus tables. Genera with less than 10 read counts were removed from the analysis. We used Spearman correlations to determine if genera were significantly correlated with a p-value cutoff of <0.05 and rho of >0.5. Gephi (0.10.1)[83] was used to visualize networks and calculate network parameters including the number of edges, nodes, average degrees, average path length, and modularity. Network parameters were correlated to methane flux using *corrplot* and linear regressions in R. Given our interest in the metabolic interactions of microbial taxa with methanogens, we focused downstream analyses on positive interactions.

To uncover the metabolic interaction patterns of the methanogens, co-occurrence networks were compared to MAGs in the MUCC database that had been assigned taxonomy using GTDB-Tk (v2.3.0 r207)[65]. Every MAG that appeared in the methanogen networks (determined if MAG and 16S ID matched at the genus level) was compiled and annotated using DRAM (v1.4.4)[84]. MAGs were further physiologically curated using DRAM curations and manual analyses, and subsequently put into one of the following categories: Methanogen, methanotroph, fermenter, acetogen, methyl-x, or other (Supplementary Data 5). Methanogens, methanotrophs, and fermenters were defined using the rules set published in Oliverio et al. [10]. Additional methanogens and methanotrophs were assigned if a MAG was not present for that genus but has been recognized in the literature.

Acetogens were assigned if they had at least 6 out of 10 steps of the Wood-Ljungdahl pathway. Methyl-x were assigned based on the presence of known substrate-specific methylotrophic genes including both aerobic and anaerobic metabolisms. All rules are outlined in Supplementary Data 4. If multiple MAGs existed for each genus, over 50% of the MAGs had to follow the rules laid out above for it to be classified within a given category.

Additionally, we conducted network analysis with a less conservative cutoff of $p < 0.06$ to understand if networks were dramatically different if we loosened the parameters. No new methanogen or methanotroph genera were added as nodes to any of the networks, thus the overall network structure remained unchanged. Additionally, while there were some new taxa added to each network, all connections were from taxa that had been seen in at least one network previously showing that the physiological relationships we found were robust and not altered under different parameters. Networks provided in the main text used the traditional cutoff value of $p < 0.05$, but the less stringent networks and the added connections are provided in Fig. S10 and Supplementary Data 9.

Phylogenomic and physiological analysis of *Methanoregula*
MAGs in the MUCC database were taxonomically assigned using GTDB-Tk (v2.1.1 r207)[65] and *Methanoregula* MAGs ($n = 37$) were parsed by genus from the full database. Further, publicly available *Methanoregula* MAGs were retrieved from GTDB ($n = 21$) and JGI ($n = 91$). These 149 MAGs were dereplicated at 99% using dRep[72] in 107 representative MAGs. All MAGs were annotated using DRAM (v1.4.4)[84].

Phylogenomic analysis of the 107 dereplicated *Methanoregula* MAGs was performed using GTDB-Tk v2.1.1 r207[65] run using the de novo workflow. The alignment was based on 53 concatenated archaeal marker genes, and a GTDB-derived genome from the phylum Undinarchaeota (GCA_002495465.1) was used as an outgroup to root the tree. The generated tree was read and visually modified, including the representation of physiological potential, in R using the *ggtree* package[85]. Newick tree is available at https://doi.org/10.5281/zenodo.14532347.

Methanoregula MAGs were screened for physiological potential for methanogenesis (*mcrABG*), hydrogenotrophy (genes encoding the Wood-Ljungdahl pathway), nitrogen fixation (nitrogenase), and CRISPR-Cas-associated proteins using DRAM. Meanwhile, to search for possession of genes encoding reactive oxygen species (ROS) detoxification enzymes, MAGs were searched via BLAST-P using a FASTA reference file (https://doi.org/10.5281/zenodo.14532347) of Uniprot and KEGG-derived reference sequences of ROS detox enzymes methanogens are known to encode[86]. The BLAST-P output was limited to include only hits with both a bitscore of ≥100 and ≥30% identity to the target sequence. Last, to curate methylotrophic potential, we carried out the strategy used by Ellenbogen et al. [15]. MAGs were searched via BLAST-P using a FASTA reference file[15] of known methylotrophic genes, namely those encoding substrate-specific corrinoid-dependent three-component methyltransferase systems comprised of a substrate:corrinoid methyltransferase, a corrinoid-binding protein, a methylcorrinoid:carbon-carrier methyltransferase, and a reductive activase. The BLAST-P output was limited to only include hits with a bitscore >60, and only genes from MAGs found to possess genes for directly substrate-interacting substrate:corrinoid methyltransferases were retained. Genes meeting these criteria were phylogenetically analyzed using ProtPipeliner to build RaxML trees (https://doi.org/10.5281/zenodo.14532347) relative to reference genes including those used in the BLAST-P search, plus other homologous sequences derived from UniProt from physiologically characterized methylotrophic methanogens and acetogens (Supplementary Data 2 tab FASTA_reference_for_genes_trees). Newick trees are available at https://doi.org/10.5281/zenodo.14532347. Trees were visually inspected in iTOL[87], and tree placement – plus gene synteny, as methylotrophic genes are

often co-encoded[88,89] - was used to confirm or refine the specific identification of genes.

Metatranscriptomic analyses
Metatransciptome analyses were performed using a previously published normalized read count table[10]. In brief, raw metatranscriptomic reads were quality trimmed, mapped to MUCC v 1.0.0, per gene read counts were estimated, and resulting read counts were normalized to gene length and TMM normalized using log2 normalization[90]. Mean geTMM values for all genes were summed for each MAG, to generate a total expression metric for each MAG activity within the 2018 OWC metatranscriptomes. Only metatranscriptome data from mud-type sites are included in these analyses. These MAG totals were further summed to the level of genus, and the methanogen data were parsed out of the full data set by taxonomy. It was manually determined which 5 methanogenic genera were most active in the D1 (0–5 cm), D3 (10–15 cm), and D6 (20–30 cm) samples independent of time. The genus-summed mean total transcription of these 5 methanogenic genera over time was plotted in R using *ggplot*[75]. To represent the activity of individual MAGs over time and depth, the mean MAG-level summed geTMM scores were plotted as a heatmap using *ggplot* in R.

Variable Importance in Projection (VIP) scores for Old Woman Creek (OWC)
Variable Importance in Projection scores (VIP) are used to estimate a variable's contribution to PLS regression, with predictors assigned high scores considered important for the PLS prediction of the tested response variable. Here, VIP were calculated as per Chong et al. [91] in R to correlate methanogen MAG activity – or genome expression- and methane production rates as the response variable in OWC. The production rates we used were calculated as described in Angle et al. In brief, data of chamber fluxes and peeper pore water concentrations for the 2015 growing season was combined with a diffusion model[16]. The model discretizes Fick's 2nd law in the vertical dimension and uses an implicit backward Euler method to account for diffusive transport within a 56-cm soil column. It also includes an empirical methane production/oxidation term calculated from the difference between concentrations measured each month. The fluxes are used as the top boundary condition. Raw VIP scores for MAG genome activity can be found in Supplementary Data 8. For MAG activity, the aforementioned summed average MAG activity table (see above) was used. Significant VIP scores (>2) were plotted using *ggplot* in R.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The data presented here in the MUCC v2.0.0 database is available across multiple platforms. The metagenomic reads, bacterial and archaeal MAGs, 16S rRNA gene sequencing reads, and metatranscriptomic reads utilized in this manuscript have been deposited in the National. Center for Biotechnology Information (NCBI) under BioProjects PRJNA330672, PRJNA1007388, PRJNA216952, PRJNA638786 and PRJNA216952. Supplementary Data 1 and 7 provide the specific BioProject IDs and BioSample numbers. The mean annual temperature and annual flux data for all sites, rarefied 16S rRNA sequencing data, subset 16S rRNA sequencing data used in network construction, and normalized metatranscriptomics activity data is available in the Source Data files. Additionally, all MAGs, processed 16S rRNA reads table, and metatranscriptomic normalized abundance table are available on Zenodo DOI 10.5281/zenodo.14532347. DRAM v1.4.4 annotations for all MAGs and the newick trees for phylogenomic and phylogenetic analyses are also available on Zenodo (https://doi.org/10.5281/zenodo.14532347). Flux data are available through AmeriFlux for site-ID US-

OWC[55], AmeriFlux site-ID US-LA2[58], and AmeriFlux, site-ID US-Twt[62]. Additional flux data is site specific and is given in previous publications for PPR[56], SPRUCE[22], and STM[63]. Source data are provided with this paper.

Code availability

Scripts involved in microbial data analysis, curation, and visualization are available at Github and Zenodo (https://github.com/ebechtold/Metabolic-interactions-underpinning-high-methane-fluxes-across-terrestrial-freshwater-wetlands/tree/main, https://doi.org/10.5281/zenodo.14532347).

References

1. Poulter, B. et al. Global wetland contribution to 2000–2012 atmospheric methane growth rate dynamics. *Environ. Res. Lett.* **12**, 094013 (2017).
2. Rosentreter, J. A. et al. Half of global methane emissions come from highly variable aquatic ecosystem sources. *Nat. Geosci.* **14**, 225–230 (2021).
3. Bridgham, S. D., Cadillo-Quiroz, H., Keller, J. K. & Zhuang, Q. Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. *Glob. Chang. Biol.* **19**, 1325–1346 (2013).
4. Laanbroek, H. J. Methane emission from natural wetlands: interplay between emergent macrophytes and soil microbial processes. A mini-review. *Ann. Bot.* **105**, 141–153 (2010).
5. Jackson, R. B. et al. Increasing anthropogenic methane emissions arise equally from agricultural and fossil fuel sources. *Environ. Res. Lett.* **15**, 071002 (2020).
6. Saunois, M. et al. The Global Methane Budget 2000–2017. *Earth Syst. Sci. Data* **12**, 1561–1623 (2020).
7. Conrad, R. 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).
8. Smith, G. J. & Wrighton, K. C. Metagenomic approaches unearth methanotroph phylogenetic and metabolic diversity. *Curr. Issues Mol. Biol.* **33**, 57–84 (2019).
9. Woodcroft, B. J. et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).
10. Oliverio, A. M. et al. Rendering the metabolic wiring powering wetland soil methane production. Preprint at *bioRxiv* https://doi.org/10.1101/2024.02.06.579222 (2024).
11. Vanwonterghem, I. et al. Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 1–9 (2016).
12. Nobu, M. K., Narihiro, T., Kuroda, K., Mei, R. & Liu, W.-T. Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. *ISME J.* **10**, 2478–2487 (2016).
13. Kumar, M. et al. Novel methanotrophic and methanogenic bacterial communities from diverse ecosystems and their impact on environment. *Biocatal. Agric. Biotechnol.* **33**, 102005 (2021).
14. Vaksmaa, A. et al. Distribution and activity of the anaerobic methanotrophic community in a nitrogen-fertilized Italian paddy soil. *FEMS Microbiol. Ecol.* **92**, fiw181 (2016).
15. Ellenbogen, J. B. et al. Methylotrophy in the Mire: direct and indirect routes for methane production in thawing permafrost. *mSystems* **9**, e0069823 (2023).
16. Angle, J. C. et al. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat. Commun.* **8**, 1567 (2017).
17. Narrowe, A. B. et al. Uncovering the diversity and activity of methylotrophic methanogens in freshwater wetland soils. *mSystems* **4**, e00320–19 (2019).
18. Liu, Y. & Whitman, W. B. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* **1125**, 171–189 (2008).
19. Kurth, J. M. et al. Methanogenic archaea use a bacteria-like methyltransferase system to demethoxylate aromatic compounds. *ISME J.* **15**, 3549–3565 (2021).
20. McCalley, C. K. et al. Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* **514**, 478–481 (2014).
21. Tveit, A. T., Urich, T., Frenzel, P. & Svenning, M. M. Metabolic and trophic interactions modulate methane production by Arctic peat microbiota in response to warming. *Proc. Natl Acad. Sci. USA* **112**, E2507–E2516 (2015).
22. Wilson, R. M. et al. Soil metabolome response to whole-ecosystem warming at the spruce and peatland responses under changing environments experiment. *Proc. Natl Acad. Sci. USA* **118**, e2004192118 (2021).
23. Wilson, R. M. et al. Microbial community analyses inform geochemical reaction network models for predicting pathways of greenhouse gas production. *Front. Earth Sci.* **7**, 59 (2019).
24. Dalcin Martins, P. et al. Abundant carbon substrates drive extremely high sulfate reduction rates and methane fluxes in Prairie Pothole Wetlands. *Glob. Chang. Biol.* **23**, 3107–3120 (2017).
25. Forbrich, I. et al. Three decades of wetland methane surface flux modeling by Earth system models-advances, applications, and challenges. *J. Geophys. Res. Biogeosci.* **129**, e2023JG007915 (2024).
26. Delwiche, K. B. et al. FLUXNET-CH4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. *Earth Syst. Sci. Data* **13**, 3607–3689 (2021).
27. Chang, K.-Y. et al. Substantial hysteresis in emergent temperature sensitivity of global wetland CH4 emissions. *Nat. Commun.* **12**, 2266 (2021).
28. Knox, S. H. et al. Identifying dominant environmental predictors of freshwater wetland methane fluxes across diurnal to seasonal time scales. *Glob. Change Biol.* **27**, 3582–3604 (2021).
29. Chadburn, S. E. et al. Modeled Microbial Dynamics Explain the Apparent Temperature Sensitivity of Wetland Methane Emissions. *Glob. Biogeochem. Cycles* **34**, e2020GB006678 (2020).
30. He, S. et al. Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. *mBio* **6**, e00066–00015 (2015).
31. Dalcin Martins, P. et al. Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome* **6**, 138 (2018).
32. Kellner, E. *Wetlands – Different Types, Their Properties and Functions*. SKB-TR–04-08 (IAEA, 2003).
33. Juottonen, H. et al. Methanogen communities and Bacteria along an ecohydrological gradient in a northern raised bog complex. *Environ. Microbiol.* **7**, 1547–1557 (2005).
34. Seward, J. et al. Peatland microbial community composition is driven by a natural climate gradient. *Micro. Ecol.* **80**, 593–602 (2020).
35. Lin, X. et al. Microbial community stratification linked to utilization of carbohydrates and phosphorus limitation in a boreal peatland at Marcell Experimental Forest, Minnesota, USA. *Appl Environ. Microbiol.* **80**, 3518–3530 (2014).
36. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103**, 626–631 (2006).
37. Shade, A. & Stopnisek, N. Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr. Opin. Microbiol.* **49**, 50–58 (2019).
38. Custer, G. F., Gans, M., van Diepen, L. T. A., Dini-Andreote, F. & Buerkle, C. A. Comparative analysis of core microbiome

assignments: implications for ecological synthesis. *mSystems* **8**, e01066–22 (2023).

39. Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12 (2012).

40. Esson, K. C. et al. Alpha- and gammaproteobacterial methanotrophs codominate the active methane-oxidizing communities in an acidic boreal peat bog. *Appl. Environ. Microbiol.* **82**, 2363–2371 (2016).

41. Singleton, C. M. et al. Methanotrophy across a natural permafrost thaw environment. *ISME J.* **12**, 2544–2558 (2018).

42. Wrighton, K. C. et al. Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. *ISME J.* **8**, 1452–1463 (2014).

43. Wüst, P. K., Horn, M. A. & Drake, H. L. Trophic links between fermenters and methanogens in a moderately acidic fen soil. *Environ. Microbiol.* **11**, 1395–1409 (2009).

44. Schink, B. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61**, 262–280 (1997).

45. Sieber, J. R., McInerney, M. J. & Gunsalus, R. P. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu. Rev. Microbiol.* **66**, 429–452 (2012).

46. Qiu, Y.-L. et al. Syntrophorhabdus aromaticivorans gen. nov., sp. nov., the first cultured anaerobe capable of degrading phenol to acetate in obligate syntrophic associations with a hydrogenotrophic methanogen. *Appl. Environ. Microbiol.* **74**, 2051–2058 (2008).

47. Karekar, S., Stefanini, R. & Ahring, B. Homo-acetogens: their metabolism and competitive relationship with hydrogenotrophic methanogens. *Microorganisms* **10**, 397 (2022).

48. Li, D. et al. Coexistence patterns of soil methanogens are closely tied to methane generation and community assembly in rice paddies. *Microbiome* **9**, 20 (2021).

49. Yang, S. et al. In-depth analysis of core methanogenic communities from high elevation permafrost-affected wetlands. *Soil Biol. Biochem.* **111**, 66–77 (2017).

50. Garcia, P. S., Gribaldo, S. & Borrel, G. Diversity and evolution of methane-related pathways in archaea. *Annu. Rev. Microbiol.* https://doi.org/10.1146/annurev-micro-041020-024935 (2022).

51. Fu, H. & Metcalf, W. W. Genetic basis for metabolism of methylated sulfur compounds in methanosarcina species. *J. Bacteriol.* **197**, 1515–1524 (2015).

52. Zalman, C. A. et al. Methylotrophic methanogenesis in Sphagnum-dominated peatland soils. *Soil Biol. Biochem.* **118**, 156–160 (2018).

53. Delgado-Baquerizo, M. et al. A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).

54. Morin, T. H. et al. Combining eddy-covariance and chamber measurements to determine the methane budget from a small, heterogeneous urban floodplain wetland park. *Agric. For. Meteorol.* **237–238**, 160–170 (2017).

55. Bohrer, G. & Kerns, J. AmeriFlux BASE US-OWC Old Woman Creek, Ver. 4-5, AmeriFlux AMP (Dataset) https://doi.org/10.17190/AMF/1418679 (2023).

56. Tangen, B. A. & Bansal, S. Soil properties and greenhouse gas fluxes of Prairie Pothole Region wetlands: a comprehensive data release: U.S. Geological Survey data release https://doi.org/10.5066/F7KS6QG2 (2019).

57. Bansal, S. et al. Practical guide to measuring wetland carbon pools and fluxes. *Wetlands* **43**, 105 (2023).

58. Ward, E., Merino, S., Stagg, C. & Krauss, K. AmeriFlux BASE US-LA2 Salvador WMA Freshwater Marsh, Ver. 3-5, AmeriFlux AMP (Dataset). https://doi.org/10.17190/AMF/1543387 (2023).

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

60. Villa, J. A. et al. Ebullition dominates methane fluxes from the water surface across different ecohydrological patches in a temperate freshwater marsh at the end of the growing season. *Sci. Total Environ.* **767**, 144498 (2021).

61. Miller, R. L., Fram, M., Fujii, R. & Wheeler, G. Subsidence reversal in a re-established wetland in the Sacramento-San Joaquin Delta, California, USA. *San Francisco Estuary Watershed Sci.* **6**, 3 (2008).

62. Knox, S., Matthes, J. H., Verfaillie, J. G. & Baldocchi, D. AmeriFlux BASE US-Twt Twitchell Island, Ver. 7-5, AmeriFlux AMP (Dataset) https://doi.org/10.17190/AMF/1246140 (2023).

63. Holmes, M. E. et al. Carbon accumulation, flux, and fate in Stordalen Mire, a permafrost peatland in transition. *Glob. Biogeochem. Cycles* **36**, e2021GB007113 (2022).

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

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

66. Joshi, N. A. & Fass, J. N. Sickle: a windowed adaptive trimming tool for FASTQ files using quality. Github https://github.com/najoshi/sickle (2011).

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

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

69. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. https://www.osti.gov/biblio/1241166 (2014).

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

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

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

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

74. Oksanen, J. et al. vegan: Community Ecology Package. R package version 2.7-0. github https://github.com/vegandevs/vegan (2025).

75. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

76. Clarkson, B. & Peterson, M. Wetland types. in Wetland restoration: a handbook for New Zealand freshwater systems 26–38 (Manaaki Whenua Press, Landcare Research, 2010).

77. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots .R package version 0.6.0. https://CRAN.R-project.org/package=ggpubr (2023).

78. Wickham, H. et al. Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

79. Wickham, H., Romain F., Henry, L., Müller, K. & Vaughan, D. dplyr: A Grammar of Data Manipulation. R package version 1.1.2. https://CRAN.R-project.org/package=dplyr (2023).

80. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006).

81. Harrell, F. et al. Hmisc: Harrell miscellaneous. R package version 4.5-0. https://CRAN.R-project.org/package=Hmisc (2021).

82. Douglas, B. & Martin M. *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.3-4. https://CRAN.R-project.org/package=Matrix (2021).

83. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Web Soc. Media* **3**, 361–362 (2009).

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

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

86. Jasso-Chávez, R. et al. Air-adapted methanosarcina acetivorans shows high methane production and develops resistance against oxygen stress. *PLoS ONE* **10**, e0117331 (2015).

87. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

88. Ellenbogen, J. B., Jiang, R., Kountz, D. J., Zhang, L. & Krzycki, J. A. The MttB superfamily member MtyB from the human gut symbiont Eubacterium limosum is a cobalamin-dependent γ-butyrobetaine methyltransferase. *J. Biol. Chem.* **297**, 101327 (2021).

89. Kountz, D. J., Behrman, E. J., Zhang, L. & Krzycki, J. A. MtcB, a member of the MttB superfamily from the human gut acetogen Eubacterium limosum, is a cobalamin-dependent carnitine demethylase. *J. Biol. Chem.* **295**, 11971–11981 (2020).

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

91. Chong, I.-G. & Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **78**, 103–112 (2005).

Acknowledgements

Author contributions

E.K.B., K.C.W., and M.J.W. conceptualized, designed, and supervised the study. E.K.B., J.B.E., J.A.V., D.K.M., A.M.O., J.E.K., V.I.R., R.K.V., S.B., E.J.W., G.B., and M.A.B. performed data collection and experimental work which created the data. E.K.B., J.A.V., A.M.O., G.B., S.B., J.B.E., and M.A.B. analyzed, curated, and visualized the data. E.K.B., J.B.E., K.C.W., and M.J.W. drafted the manuscript. All of the authors read, commented on and edited the manuscript, as well as approved its final form.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-56133-0.

Correspondence and requests for materials should be addressed to Kelly C. Wrighton or Michael J. Wilkins.

Peer review information *Nature Communications* thanks Cuijing Zhang and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.