Article

Discrepancies in lacustrine bacterial lipid temperature reconstructions explained by microbial ecology

Check for updates

Jie Liang [1,2] ✉, Manuel Chevalier [3], Keshao Liu[1], Amedea Perfumo[2], Mingda Wang [4], Haichao Xie [1] ✉, Juzhi Hou [1] ✉, Ulrike Herzschuh [2,5,6] ✉ & Fahu Chen [1]

Bacterial lipid branched glycerol dialkyl glycerol tetraethers (brGDGTs) are a valuable tool for reconstructing past temperatures. However, a gap remains regarding the influence of bacterial communities on brGDGT profiles. Here, we identified two distinct patterns of brGDGTs from the surface sediments of 38 Tibetan Plateau lakes using an unsupervised clustering technique. Further investigation revealed that salinity and pH significantly change bacterial community composition, affecting brGDGT profiles and causing brGDGT-based temperatures to be overestimated by up to 2.7 ± 0.7 °C in haloalkaline environments. We subsequently used the trained clustering model to examine the patterns of bacterial assemblages in the global lacustrine brGDGT dataset, confirming the global applicability of our approach. We finally applied our approach to Holocene brGDGT records from the Tibetan Plateau, showing that shifts in bacterial clusters amplified temperature variations over timescales. Our findings demonstrate that microbial ecology can robustly diagnose and constrain site-specific discrepancies in temperature reconstruction.

Quantitative estimates of terrestrial paleotemperature are crucial for deciphering Earth's past climate history[1]. Reconstructions of global temperature over the Holocene epoch remain a subject of debate, primarily due to discrepancies between proxy-based reconstructions and climate model simulations, known as the Holocene temperature conundrum[2,3]. Among the wide range of temperature proxies, bacterial lipids called branched glycerol dialkyl glycerol tetraethers (brGDGTs) have been extensively developed and attracted increasing attention in the last decade due to their potential to provide global, high-resolution, continuous records that are crucial for reconstructing paleotemperatures across a broad range of environments[4]. Despite the high potential of brGDGTs in paleotemperature reconstructions, challenges persist, particularly in mid-latitude regions where reconstructions are frequently inconsistent[5–7].

The unique and diverse environmental conditions of lakes, such as seasonality[8,9], in-situ production versus terrestrial input[10,11] and specific conditions like salinity[7], oxygen[12], and lake depth[13], are recognized as primary factors influencing lacustrine bacterial lipids, complicating temperature reconstruction efforts. For example, in mid- and high-latitude regions,

brGDGT production rates decrease in colder seasons[14–16]. Therefore, calibrating brGDGTs to the mean temperatures of months above freezing (MAF) instead of the mean annual air temperature may result in more accurate reconstructions[17,18]. The complexity of factors influencing brGDGT profiles highlights the need for innovative approaches to accurately interpret brGDGT-based temperature reconstructions. Recent advances in statistical and machine learning methods have effectively addressed these complexities[17–20]. Machine learning calibrations enhance the precision and reliability of temperature estimates by addressing the complex, non-linear relationships between brGDGT distributions and temperature that linear empirical calibrations may not fully capture[17,20,21]. For instance, the lacustrine brGDGT random forest calibration model, which categorizes brGDGTs into four distinct clusters based on temperature gradients and geographic locations, enhances predictive accuracy by accounting for environmental diversity[17]. Additionally, the mixed sources of brGDGTs (i.e., both from the lake and the surrounding catchment) also influence lacustrine brGDGT reconstructions. To this end, the recent application of machine learning methods has helped disentangle the sources

[1]Group of Alpine Paleoecology and Human Adaptation (ALPHA), State Key Laboratory of Tibetan Plateau Earth System, Resources and Environment (TPESRE), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China. [2]Polar Terrestrial Environmental Systems, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany. [3]Institute of Geosciences, Section of Meteorology, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany. [4]School of Geography, Liaoning Normal University, Dalian, China. [5]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany. [6]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany. ✉e-mail: jie.liang@awi.de; hcxie@itpcas.ac.cn; jzhou@itpcas.ac.cn; ulrike.herzschuh@awi.de

1

of brGDGTs from different depositional environments (soils, peats, and sediments) worldwide[18].

Advanced chromatography techniques have enabled the separation of 5-methyl and 6-methyl isomers of brGDGTs[22,23]. Studies have shown that the degree of methylation of 5-methyl brGDGTs (MBT$'_{5Me}$) has a stronger correlation with temperature compared to the traditional MBT index, which includes both 5- and 6-methyl isomers[22]. However, when the ratio of 6- over 5-methyl brGDGTs (IR$_{6Me}$) exceeds 0.5, the association between MBT$'_{5Me}$ or MBT' and temperature becomes weaker in global soil datasets[14,24]. Recent investigations have demonstrated that incorporating 6-methyl isomers in empirical temperature calibrations can improve the performance of the brGDGT paleothermometer in lacustrine[25–27] and soil[28] settings. The high IR$_{6Me}$ ratio has been associated with significant uncertainty in the application of the brGDGT proxy, which may be related to shifts in the bacterial communities responsible for the production of these lipids[29]. Nevertheless, the mechanisms underlying the effects of these isomers, particularly in relation to bacterial communities, on the paleothermometer remain unclear. One of the main challenges is our limited understanding of brGDGT-producing bacteria and their ecology, with only a few species having been cultured or incubated[30–33]. Understanding how bacterial communities respond to changing environmental factors beyond temperature is fundamental for accurately interpreting brGDGT distribution and developing reliable brGDGT-based temperature reconstructions.

The Tibetan Plateau, the highest and largest plateau on Earth, hosts numerous lakes that serve as unique archives for paleoclimate reconstructions[34]. Recent studies of Holocene temperature reconstructions using lacustrine brGDGTs have revealed diverse trends across this region[5,34], complicating our understanding of past temperature variations. More importantly, even after applying different brGDGT-temperature calibrations—including the latest machine learning-based random forest calibration (RF)[17]—regional Holocene temperature reconstructions continue to

show significant discrepancies (Fig. 1). For example, applying the subset-specific brGDGT[8] and Bayesian (BayMBT)[9] calibrations to lakes on the Tibetan Plateau, such as Bangong Co (BGC), Gahai (GH), and Ngamring Co (NG), reveals substantial temperature variations ranging from 9 °C to 12 °C over the Holocene. Similarly, the application of RF calibration to these lakes also indicates considerable variability, around 9 °C. In contrast, other lakes in the region exhibit significantly smaller fluctuations, regardless of the calibration method applied (Fig. 1). These discrepancies in temperature reconstructions persist across different lakes and calibrations, suggesting that the calibration approach alone may not be sufficient to account for the observed variability. This raises a critical question: even with advanced machine learning calibration, can brGDGTs reliably reconstruct past temperatures over long timescales? The variations observed across different lakes emphasize the need to explore additional factors, such as ecological shifts or environmental changes, that may be influencing these reconstructions. Crucially, this issue is not unique to the Tibetan Plateau, as similar discrepancies have been found in other regions[36,37], further emphasizing the complexity of brGDGT-based temperature reconstructions.

In this study, we investigate bacterial compositional changes upon environmental factors (e.g., lake depth, Secchi depth, water temperature, salinity, pH, oxygen, and chlorophyll a) by combining brGDGT and DNA analysis of modern samples ($n = 38$) from the Tibetan Plateau and assess their impact on Holocene temperature reconstructions. We focus on the MBT$'_{5Me}$ and simplified IR$_{6Me}$ (IR) indices as MBT$'_{5Me}$ reflects bacterial physiological adaptation to temperature changes[38], while IR accounts for variations in 6-methyl brGDGT production, which can complicate temperature reconstructions[29], particularly in alkaline or saline lakes where elevated concentrations of 6-methyl isomers are frequently observed[7,25,26]. We then integrate these data into an unsupervised probabilistic machine learning clustering model (Gaussian Mixture Models; GMM), which we
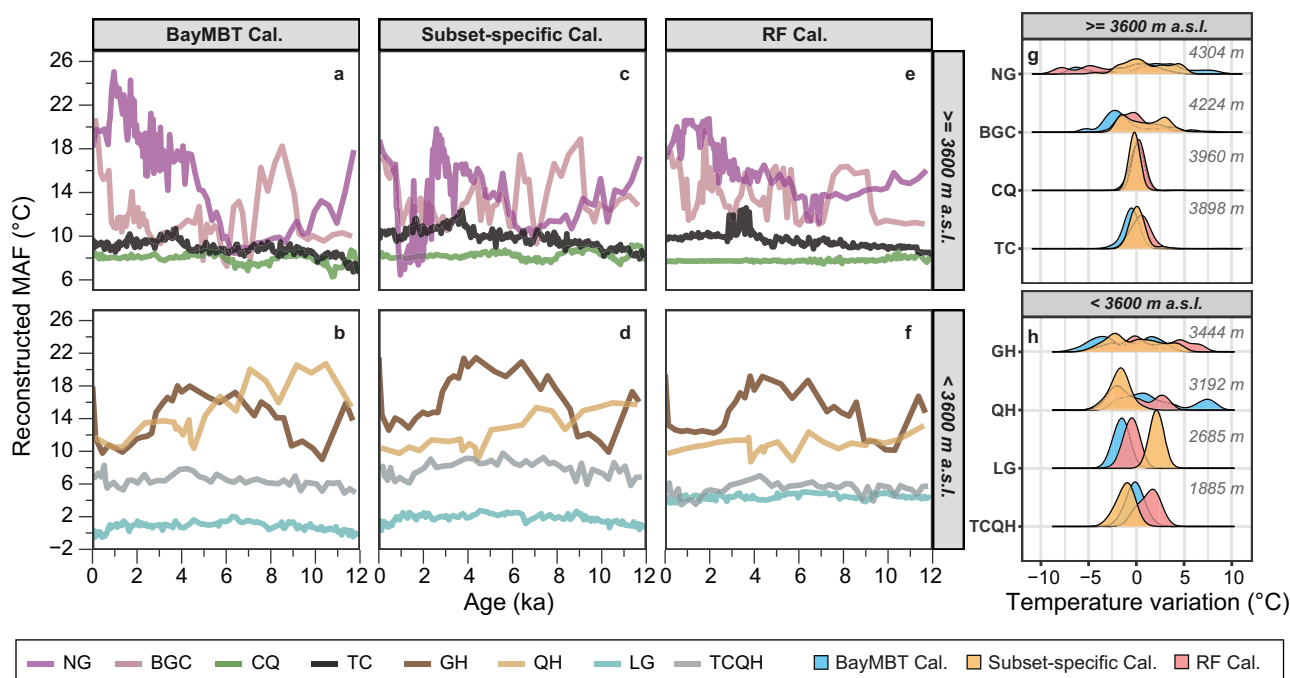

Fig. 1 | Comparison of Holocene mean temperatures of months above freezing (MAF) from lakes on the Tibetan Plateau. The three columns represent three different global lacustrine brGDGT-MAF calibrations: Bayesian calibration (BayMBT; **a**, **b**)[9], subset-specific brGDGT calibration (**c**, **d**)[8], and random forest calibration (RF; **e**, **f**)[17]. The top row (**a**, **c**, **e**) illustrates reconstructions from lakes above 3600 m a.s.l., while the bottom row (**b**, **d**, **f**) shows lakes below 3600 m a.s.l. Temperatures are adjusted to 3600 m a.s.l. using a lapse rate correction of 6 °C/km. Meteorological data from Lhasa (3600 m a.s.l.) indicate an MAF of 10.0 °C from 1981 to 2020. The right-side ridgeline plots (**g**, **h**) display temperature anomalies, calculated as the deviation of reconstructed temperatures from the averaged MAF for each lake, highlighting Holocene temperature variation among the different calibrations (BayMBT: blue; subset-specific: orange; RF: red). The numbers on the ridgeline plots represent the altitudes of the lakes. Lake abbreviations: Cuoqia lake (CQ), Tiancai lake (TC), Bangong Co (BGC), Ngamring Co (NG), Tengchong-qinghai lake (TCQH), Lugu lake (LG), Qinghai lake (QH), and Gahai lake (GH).

Fig. 2 | **Analysis of lake surface sediments from Tibetan Plateau. a** Map of 38 lake surface samples (yellow dots) and Holocene sample sites (blue stars) examined in this study (see Supplementary data for more details on the lakes); **b** Application of the Gaussian Mixture Model (GMM) clustering algorithm on isomer ratio (IR) and the degree of methylation (MBT'_{5Me}) data from Tibetan surface sediments, resulting in the identification of two distinct clusters. Kernel density plots illustrate the relationship between MBT'_{5Me} and IR ratio for the sediments, with cluster 1 in blue and cluster 2 in red. The regression lines represent the relationship between MBT'_{5Me} and IR for each cluster. The histograms at the top and right of the plot show the distributions of MBT'_{5Me} and IR, respectively. **c** The corresponding offsets (ΔMAF), defined as the difference between the reconstructed MAF based on brGDGTs and the observed MAF, were calculated for both clusters. The gray dashed lines indicate the median values for each cluster, with a difference of 2.7 °C between them.

(a) Map of sites

(b) GMM clusters

(c) MAF reconstruction bias
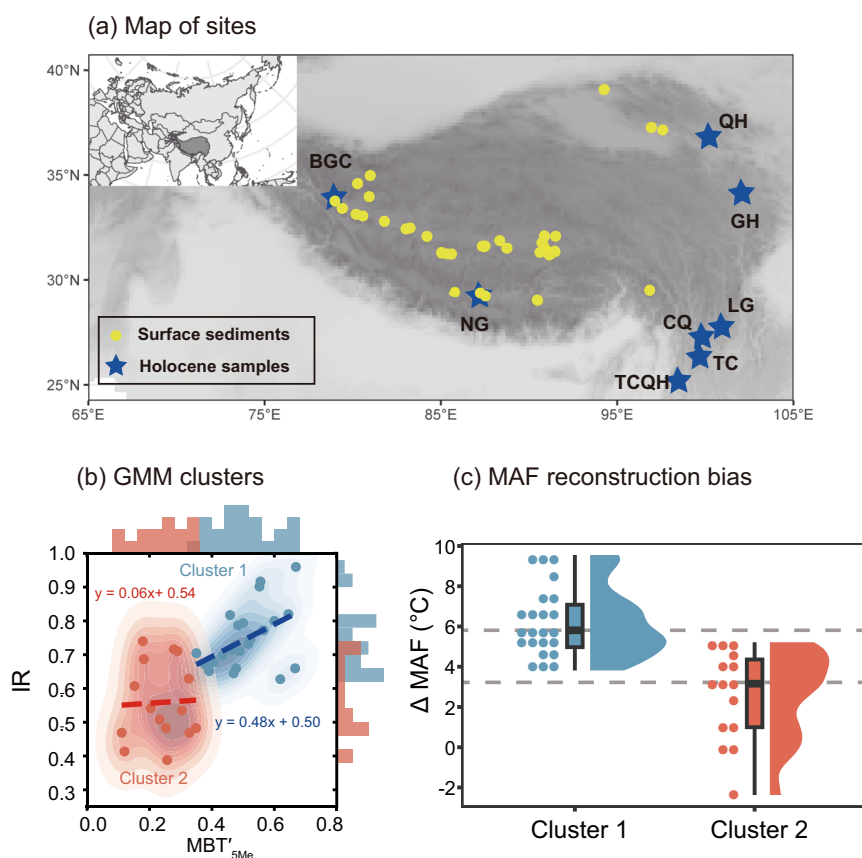
trained to examine the relationship between MBT'_{5Me} and IR indices. The GMM algorithm is an unsupervised learning method that clusters data based solely on brGDGT indices, without incorporating any environmental factors. This approach can yield a more accurate and unbiased classification of the data[39]. We identified two clusters indicating distinct biases in brGDGT-temperature reconstruction, each correlated with specific bacterial communities (i.e., haloalkaliphilic and freshwater). Therefore, we named this cluster model the brGDGT-GMM bacterial assemblages'estimator (brGMM-BAE). The brGMM-BAE model also accurately predicts the distribution patterns of bacterial assemblages in global modern surface sediments ($n = 415$), showing results that are consistent with lake environmental conditions. Finally, we apply the brGMM-BAE model to eight Holocene sediment cores from the Tibetan Plateau and find that the large discrepancies in temperature reconstruction over the past 12,000 years originate from shifts in bacterial assemblages due to changes in lake salinity and pH. Overall, our research introduces a novel and holistic study approach integrating advanced machine learning with brGDGTs and DNA analysis. This allows us to gain deeper insights into the environmental factors driving changes in brGDGT profiles, achieve a more precise interpretation of brGDGT temperature reconstruction, and enhance the accuracy of paleotemperature reconstruction.

Results

BrGDGT occurrence in lake sediments clusters with distinct bacterial communities

We performed brGDGT data analysis on 38 lake surface sediments from the Tibetan Plateau (Fig. 2a) to investigate the distribution of IR and MBT $'_{5Me}$ using Gaussian mixture model (GMM) algorithm. The GMM identified two distinct clusters (Fig. 2b). Cluster 1 is characterized by higher MBT $'_{5Me}$ (0.3–0.7) and IR values (0.6–1.0), with a positive correlation between MBT'_{5Me} and IR. In contrast, the indices in cluster 2 are only marginally positively correlated due to the large spread of IR values (0.4–0.7).

To reconstruct the mean temperatures of MAF, we applied three of the most used global lacustrine calibrations: RF calibration[17], Bayesian MAF calibration (BayMBT)[9], and subset-specific brGDGT calibration[8]. They all showed discrepancies between reconstructed and observed temperatures. We chose the RF calibration to reconstruct brGDGT-based temperature in lake sediments in this study because it resulted in smaller temperature offset between reconstruction and observation (ΔMAF) than the others (Supplementary Fig. 1). We observed that brGDGT-based MAF is overestimated in both clusters identified by the GMM, but with distinct differences between them. The warm bias is larger in cluster 1 with ΔMAF values ranging from 3.8 to 9.6 °C. The estimated bias of cluster 2 predominantly ranges from −2.4 to 5.2 °C. On average, cluster 1 overestimated the temperature by 2.7 ± 0.7 °C compared to cluster 2. The bias was calculated as the median difference in reconstructed temperature bias (ΔMAF) between the two clusters, with uncertainty (±0.7 °C) represented as the standard error (SE) of the median (Fig. 2c).

Bacterial DNA data support brGDGT clusters

We then integrated DNA analysis to assess bacterial community composition from all 38 lakes and its contribution to brGDGT profiles (Fig. 3). Comparing the 50 most abundant taxa at the family level in our two clusters showed that some bacteria in cluster 1 were absent or much less abundant in cluster 2. In particular, cluster 1 contained many bacteria known for their tolerance of high salinity and alkaline environments, including *Prochlorococcaceae*, *Synthrophomonadaceae*, *Saprospiraceae* and *Spirochaetaceae* (Fig. 3). In some genomes we identified the presence of GDGT synthase protein (tetraether synthase, Tes), a key enzyme in the biosynthesis of archaeal isoGDGTs[35] and then further applied to the study of bacterial brGDGTs[40]. This finding supports the hypothesis that these organisms, as potential brGDGT producers, could influence brGDGT profiles. The model, trained on Tibetan lake surface sediments, successfully identified two distinct bacterial communities and revealed the relationship between the IR
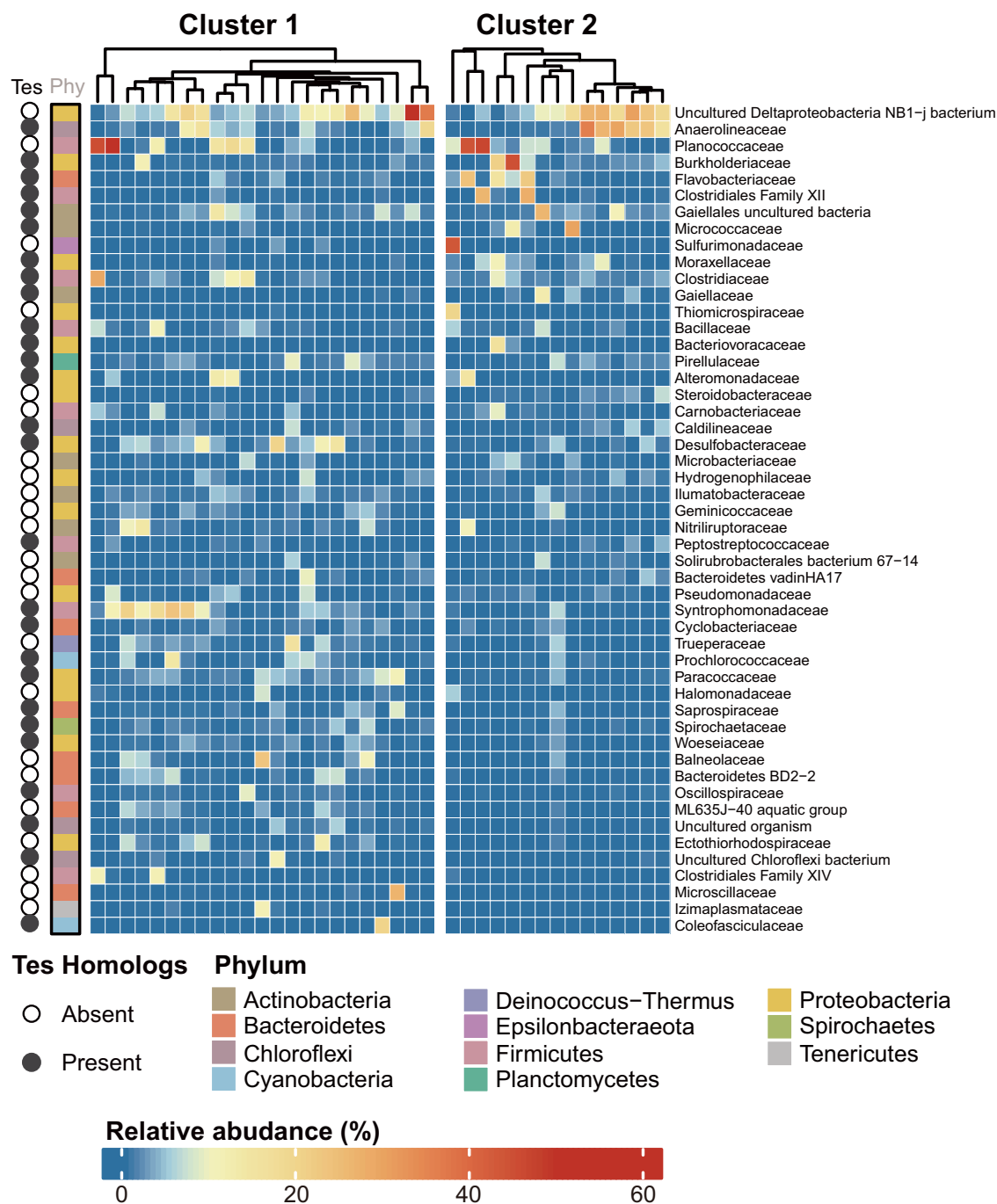
Fig. 3 | Bacterial diversity and Tes homolog presence in surface sediments. The top 50 most abundant bacterial species at the family level were identified in surface sediments from the Tibetan Plateau. The two clusters were identified using the brGMM-BAE model, based on the MBT′$_{5Me}$ and IR values of each sample. The leftmost section of the figure employs a binary color scheme to denote the presence (black) or absence (white) of Tes homolog genes within these taxa, providing insights into the genetic traits and potential metabolic capabilities of the microbial communities in these sediments. A color bar along the side labels each species by its corresponding phylum. The heatmap shows the relative abundance of bacteria at the family level.

and MBT′$_{5Me}$ indices. Consequently, we have named this trained GMM model the brGDGT-GMM bacterial assemblages' estimator (brGMM-BAE) model.
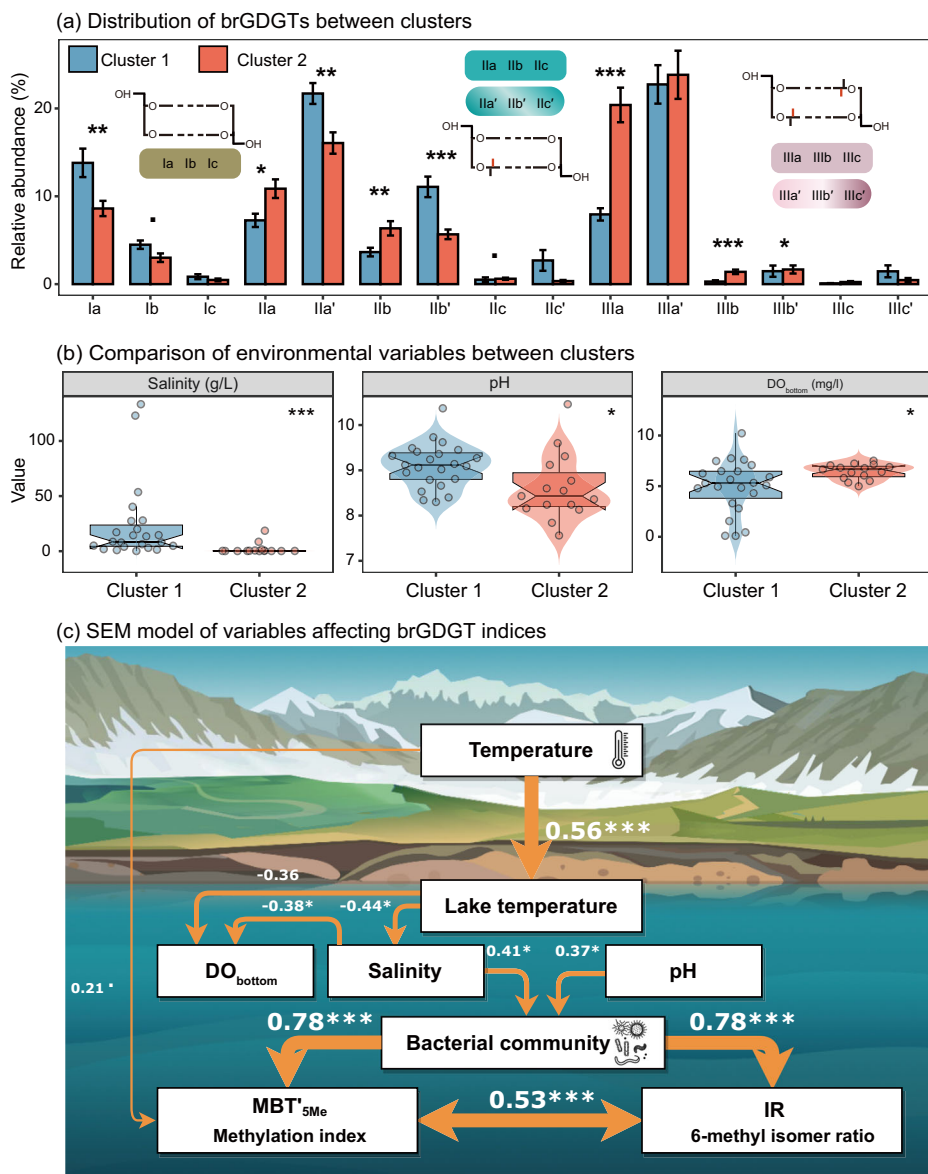
Environmental variables and bacterial communities influence brGDGT compositions

The distribution patterns of brGDGT compounds substantially differ between our two clusters (Fig. 4a). Tetramethylated brGDGTs (Ia, Ib, and Ic) are more abundant in cluster 1 (19%) than in cluster 2 (12%). Conversely, hexamethylated brGDGTs (IIIa, IIIb, IIIc, IIIa′, IIIb′, and IIIc′) are more

abundant in cluster 2 (48%) compared to cluster 1 (34%). Similarly, 5-methyl pentamethylated brGDGTs (IIa, IIb and IIc) also show a higher relative abundance in cluster 2 (18%) versus cluster 1 (11%). In contrast, 6-methyl pentamethylated brGDGTs (IIa′, IIb′, and IIc′) are more abundant in cluster 1 (35%) than in cluster 2 (22%).

Furthermore, the two clusters behave differently in response to salinity, pH, and dissolved oxygen levels at the bottom water of lakes (DO$_{bottom}$) (Fig. 4b). Cluster 1 shows preference for brackish to saline (1–133 g/L) and alkaline (pH from 8 to 10) conditions, with DO$_{bottom}$ ranging widely from 1 to 10 mg/L. In contrast, cluster 2 shows a tendency for freshwater lakes with

Fig. 4 | Profiles of brGDGTs, environmental variables, and modeling of biotic and abiotic variables affecting brGDGT indices in lakes from Tibetan Plateau. a The relative abundance of brGDGT compounds across two clusters identified in the Tibetan Plateau, with illustrations of brGDGT simplified structures. **b** A boxplot of environmental factors (salinity, pH, and dissolved oxygen of lake bottom water (DO_{bottom})) shows a significant impact on the clusters. The comparison of corresponding brGDGT compounds and environmental variables between the two clusters is made using the Kruskal-Wallis test, assessing whether samples from these clusters are derived from the same distribution. **c** Structural Equation Model (SEM) applied to the surface lake sediment datasets from Tibetan Plateau, showing standardized coefficients to illustrate the relationships among response variables and predictors. The relationships are represented by arrows, indicating the direction and strength of associations. The thickness of the paths (arrows) is scaled to the magnitude of their significance. Significance levels across all plots are denoted by symbols: '***' for $p < 0.001$, '**' for $p < 0.01$, '*' for $p < 0.05$, and '.' for $p < 0.1$.

(a) Distribution of brGDGTs between clusters

(b) Comparison of environmental variables between clusters

(c) SEM model of variables affecting brGDGT indices

low salinity (around 0.2 g/L) and low alkalinity (pH from 7 to 9), and dissolved oxygen levels around 5 mg/L.

To identify which environmental factors may influence brGDGT indices MBT'_{5Me} and IR, we employed a Structural Equation Model (SEM) (Fig. 4c). We found that the MBT'_{5Me} index is positively influenced by changes in the bacterial community (standardized effect size: $\beta = 0.78$, $p < 0.001$) and air temperature ($\beta = 0.21$, $p < 0.1$). While lake temperature, salinity, and pH impact MBT'_{5Me} indirectly through their effects on the bacterial community, both salinity ($\beta = 0.41$, $p < 0.05$) and pH ($\beta = 0.37$, $p < 0.05$) have direct effects on the bacterial community, explaining 32% of its variation. In contrast, the IR index is primarily influenced by the bacterial community ($\beta = 0.78$, $p < 0.001$). Air temperature directly impacts lake water temperature ($\beta = 0.56$, $p < 0.001$), which in turn affects salinity negatively ($\beta = -0.44$, $p < 0.05$).

Validating brGDGT clusters to global surface samples

We then applied our brGMM-BAE model to determine if similar patterns of bacterial assemblages could be predicted in global lake surface sediments based on MBT'_{5Me} and IR indices, which were used as input data for the model. Using these indices, the model predicted whether the bacterial communities were associated with haloalkaliphilic or freshwater

assemblages, and these predictions were then compared with measured salinity and pH values of lakes for validation. Results were classified into three categories: "both match" (both salinity and pH correctly predicted), "partial match" (either salinity or pH correctly predicted), or "no match" (both incorrectly predicted). Given the model's capability to accurately predict freshwater-low/neutral pH and saline-high pH environments, we further examined cases where only one parameter matched. For instance, a lake observed as freshwater-alkaline but predicted as a haloalkaliphilic cluster is considered a partial match. We obtained a "both match" in ~59% of the predictions, a "partial match" in 24%, and "no match" in only 17% compared to the measured lake conditions (Fig. 5a). Specifically, for salinity, the model correctly predicted freshwater and saline species with accuracies of 66% and 80%, respectively (Fig. 5b). For pH, the accuracies were 60% for acidic/neutral lakes and 74% for alkaline lakes (Fig. 5c).

The brGDGT-derived temperatures from global lake surface sediments using lacustrine RF calibration[17] indicate that the average ΔMAF across the four types (freshwater, brackish/saline, acidic/neutral, alkaline lakes) is approximately 0 °C (Fig. 5d). However, for haloalkaline lakes, temperatures reconstructed using brGDGTs show a large bias, with a standard deviation of ~3.1 °C (Fig. 5d). Additionally, the brGMM-BAE model predicted that lakes with haloalkaliphilic bacterial assemblages also
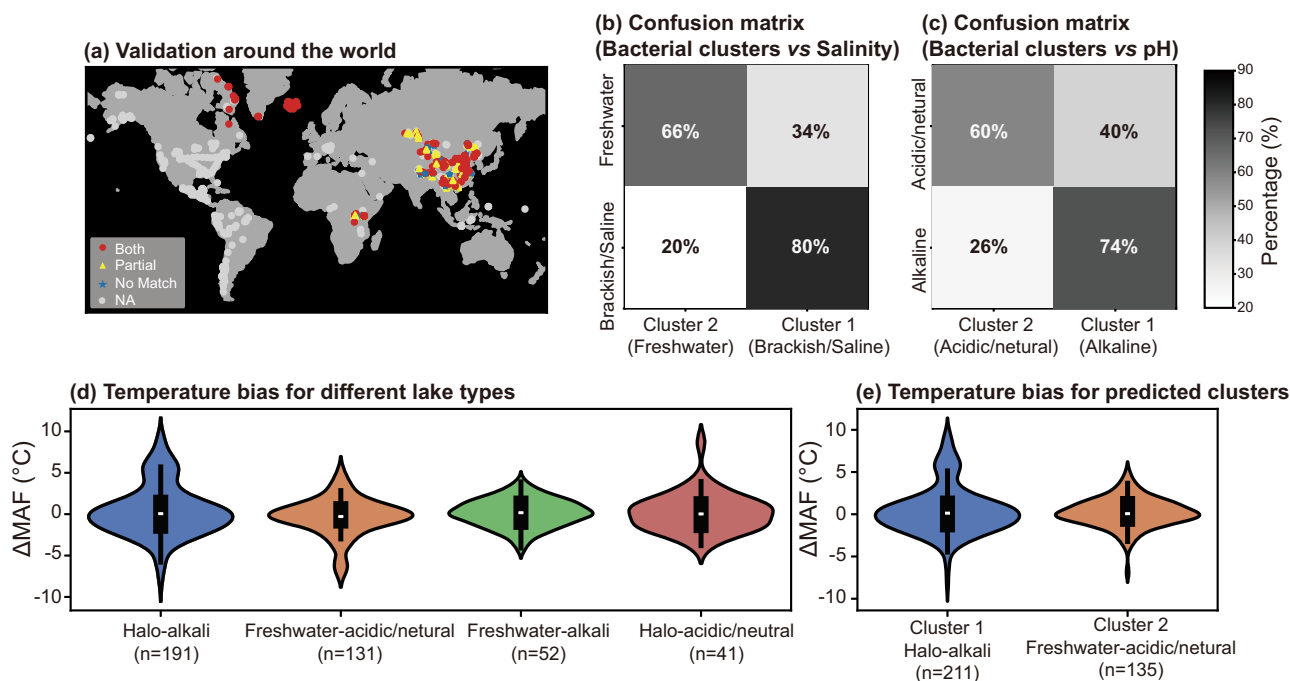
Fig. 5 | Global evidence of bacterial assemblages using the brGMM-BAE model compared to observed environmental conditions. a The accuracy of predicted microbial assemblages in relation to lake salinity and pH. Red dots (Both match): both salinity and pH were correctly predicted; Yellow triangles (Partial match): either salinity or pH was correctly predicted; Blue stars (No match): both were incorrectly predicted; Light gray dots (NA): salinity/pH data were not available. Confusion matrices provide a quantitative assessment of the model accuracy: **b** Confusion matrix for bacterial clusters with observed salinity, displaying how predicted bacterial assemblages correlate with observed salinity levels. **c** Confusion matrix for bacterial clusters with observed pH, illustrating the alignment of predicted assemblages with actual pH values in lake environments. Violin plots of temperature bias, defined as the difference between brGDGT-derived temperature and observed MAF, are analyzed for different types of lakes based on (**d**) observed lake conditions and **e** predicted bacterial clusters from both matched and partially matched sites. In the violin plots, black box plots are used to show the median and quartiles of temperature bias.

have a large temperature bias compared to those with freshwater bacterial assemblages, which have a standard deviation of ~2.9 °C (Fig. 5e).

We hypothesize that patterns predicted by modern brGDGT-based bacterial assemblages can be used to infer changes in bacterial community dynamics over time. To test this hypothesis, we extended our approach to Holocene lacustrine brGDGT records from the Tibetan Plateau (Fig. 2a). We analyzed brGDGTs from a Holocene lake sediment core (~11.6 ka) collected from Bangong Co (BGC), along with previously published paleorecords from other lakes in the Tibetan Plateau (n = 7; ~12 ka)[7,41–45]. Throughout the Holocene, lakes NG, QH, and LG were concentrated in the higher IR values, while lakes TCQH, GH, CQ, and TC were mainly found in the lower IR values. During the early and middle Holocene, lake BGC was primarily concentrated in the moderate IR values, whereas in the late Holocene, lake BGC was concentrated in the higher IR values (Fig. 6).

We input the brGDGT indices MBT$'_{5Me}$ and IR from these cores into the trained brGMM-BAE model to identify the different bacterial clusters in each lake over time. Reconstructions of paleo-bacterial assemblages indicated that lake QH was dominated by bacterial cluster 1 (haloalkaliphilic type) throughout the Holocene, while lakes GH, CQ, LG, TC, and TCQH were consistently dominated by cluster 2 akin to freshwater bacterial assemblages (Figs. 6, 7a). The reconstructed bacterial clusters from the topmost sample of these lakes closely matched the current salinity conditions (Fig. 6d). However, lake GH, which was incorrectly predicted to be a freshwater lake with neutral to slightly acidic conditions, is actually characterized by freshwater (salinity = 0.02 g/L) but with a high pH (pH = 9.3). It is clearly shown that the bacterial cluster in GH during the early and late Holocene shifted from a community similar to the slightly acidic TCQH (modern pH = 5.8) to communities more akin to the neutral TC (modern pH = 7.5) and slightly alkaline BGC (modern pH = 8.2) (Fig. 6b, d).

The brGMM-BAE model reconstructed BGC was predominantly inhabited by freshwater taxa (cluster 2) in the early and mid-Holocene,

shifting to more haloalkaliphilic taxa (cluster 1) in the late Holocene, and reverting back to freshwater ones in modern times (Figs. 6, 7a). NG lake was dominated by freshwater taxa (cluster 2) during the early and middle Holocene but completely changed to haloalkaliphilic taxa (cluster 1) after 5 ka (Figs. 6, 7a).

Constraining temperature on shifts in bacterial assemblages in Holocene records

Substantial variations were observed in Holocene temperature reconstructions across eight lakes (Fig. 1). Lakes NG, BGC, and GH exhibited temperature shifts up to ~9 °C contrast to the other five lakes, where changes were ~2 °C (Fig. 1). These findings were corroborated by predictions from the brGMM-BAE model, which highlighted significant shifts in bacterial communities specifically in NG and BGC. For example, NG showed a community shift from cluster 2 (freshwater type) in the middle Holocene to cluster 1 (haloalkaliphilic type) in the late Holocene (Fig. 6c, d), resulting in a temperature rise of 4.3 °C (Fig. 1e). Similarly, BGC demonstrated a notable community transition from cluster 2 (freshwater type) to cluster 1 (haloalkaliphilic type) ~1 ka (Fig. 6a), coinciding with a temperature increase from 7.5 °C to 13.7 °C (Fig. 1e). In the last century, GH transitioned from conditions resembling the more acidic lake TCQH to those akin to the more neutral and alkaline lakes TC and LG (Fig. 6d). Therefore, we assume that this bacterial community shift correlates with the 8.4 °C temperature rise observed in the last century (Fig. 1e). In contrast, the bacterial communities in the other five lakes (QH, TC, CQ, LG, and TCQH) displayed stability throughout the Holocene (Figs. 6, 7a), with temperature reconstructions derived from brGDGTs showing consistent trends and magnitudes of change (Fig. 1).

Substantial temperature fluctuations in the paleorecords were found to relate to shifts in bacterial assemblages, highlighting the need to account for the influence of bacterial communities on temperature reconstructions.
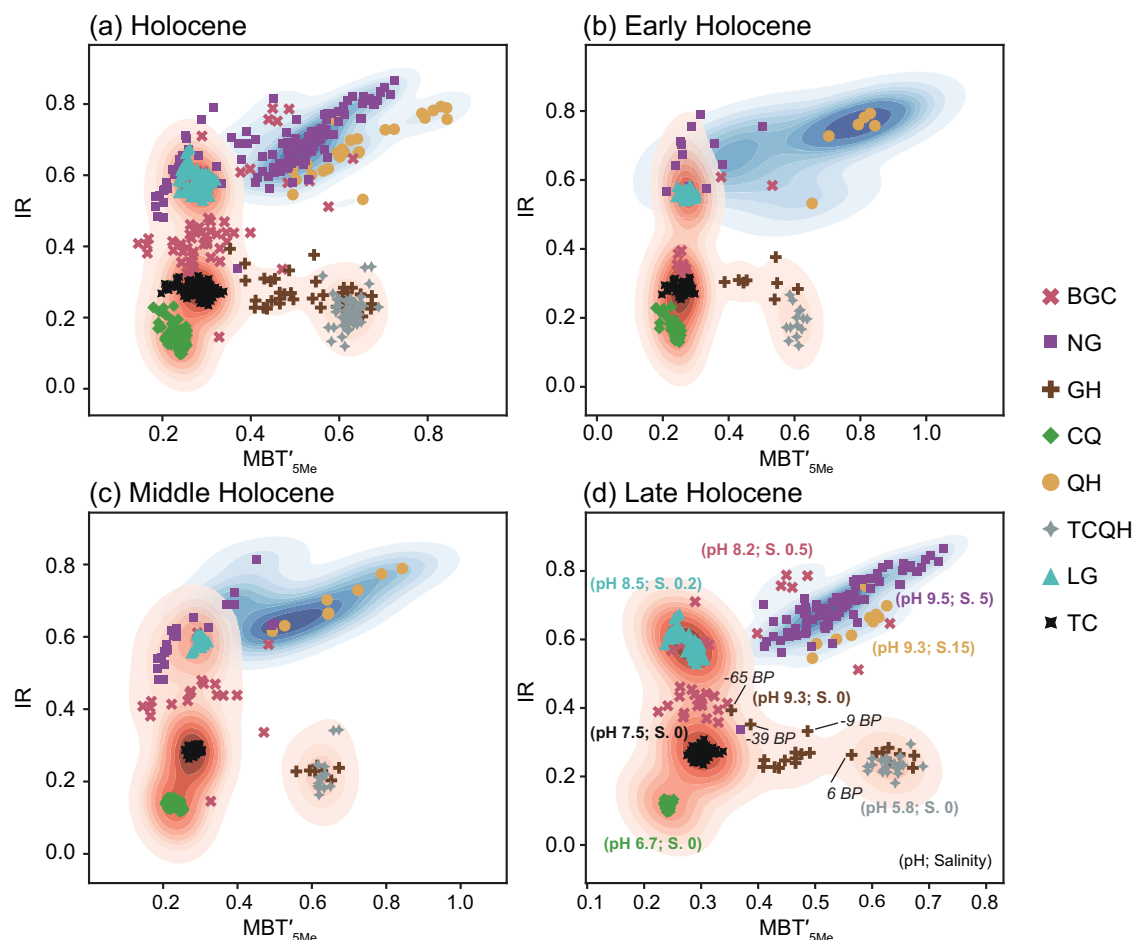
Fig. 6 | The relationship between MBT'$_{5Me}$ and IR for eight lakes across different periods of the Holocene from Tibetan Plateau. Each subplot represents a different Holocene period: **a** Holocene (all periods), **b** Early Holocene (12–8.2 ka), **c** Middle Holocene (8.2–4.2 ka), and **d** Late Holocene (after 4.2 ka). The scatter plots use specific colors and shapes to distinguish data points from different lakes: BGC (red, cross), NG (purple, square), GH (brown, plus), CQ (green, diamond), QH (yellow, circle), TCQH (gray, star), LG (light blue, triangle), and TC (black, asterisk). The color of kernel density plots is based on predicted bacterial assemblages using the brGMM-BAE model, with cluster 1 (haloalkaliphilic) in blue and cluster 2 (freshwater) in red. Numbers in (**d**) represent the corresponding ages (in BP) for lake GH, illustrating the shift in taxa composition over time. This transition reflects a change from bacterial communities associated with an acidic lake environment to those indicative of more neutral to alkaline conditions during the past century. Modern pH and salinity (g/L) values for each lake are provided in parentheses (pH; S).

Analysis of surface sediments from the Tibetan Plateau reveals that lakes dominated by haloalkaliphilic bacterial assemblages consistently yielded temperatures 2.7 ± 0.7 °C higher than those from freshwater-dominated lakes (Fig. 2c). While this temperature overestimation related to cluster shifts can be quantified, the community effects driven solely by pH- or salinity-induced changes remain difficult to quantify. Therefore, we applied a 2.7 °C constraint to the reconstructed temperatures only of BGC and NG lakes, which experienced cluster shifts during Holocene (Fig. 7a). Before constrained, the reconstructed temperatures were 13.9 °C at the topmost sample of NG (4304 m asl) and 10.9 °C at BGC (4224 m asl). By limiting the bias from shifts in bacterial assemblages, the temperatures at both NG and BGC were ~11 °C (Fig. 7f, g). Both sites exhibit low temperatures during the early Holocene. BGC shows temperature increases at ~8, 5, 3.5, 2, and 1 ka, while NG primarily exhibits warming at around 5.5, 2, and 0.6 ka.

Discussion
Clustering of MBT'$_{5Me}$ and IR indices reveals distinct temperature reconstructions linked to bacterial communities
Our study identified two distinct clusters associated to brGDGT distribution (based on the IR and MBT'$_{5Me}$ indices) and bacterial community composition (based on the 16S rRNA genes) in lake surface sediments from the Tibetan Plateau (Figs. 2, 3). Both clusters overestimated the reconstructed modern temperatures, with cluster 1 yielding much warmer temperatures

compared to observed temperatures than cluster 2. This implies that the distribution of brGDGTs may respond to factors other than air temperature, such as bacterial community, salinity or pH, as has also been reported for other saline lakes[7,46].

Our results show that more haloalkaliphilic bacterial taxa were enriched in cluster 1 (Fig. 3), suggesting that bacterial community composition may play a role in causing the discrepancies observed in brGDGT-based temperature reconstructions. Similarly, a study on soil brGDGT from Tibetan Plateau has found that brGDGT distributions are influenced by bacterial communities[47]. Also in marine sediments it was shown that bacterial communities exert a stronger influence on brGDGT compounds, with members of *Gemmatimonadetes*, *Planctomycetes*, and *Proteobacteria* identified as potential brGDGT producers[48].

Further support for the relationship between brGDGTs and bacterial communities is provided by the detection of tetraether synthase (Tes) gene homologs in the genomes of bacteria identified in the lake surface sediments (Fig. 3). The Tes enzyme, essential for the production of archaeal isoGDGTs through the condensation of two archaeal molecules[35], has been found in various bacterial genomes, including in *Acidobacteria*, highlighting its role in bacterial brGDGT biosynthesis[30,31,40]. Our results show that lakes in cluster 1 contain more haloalkaliphilic bacteria, and within these bacteria, we found the presence of the Tes gene (Fig. 3). This suggests that changes in bacterial communities are related to the distinct patterns of the two clusters
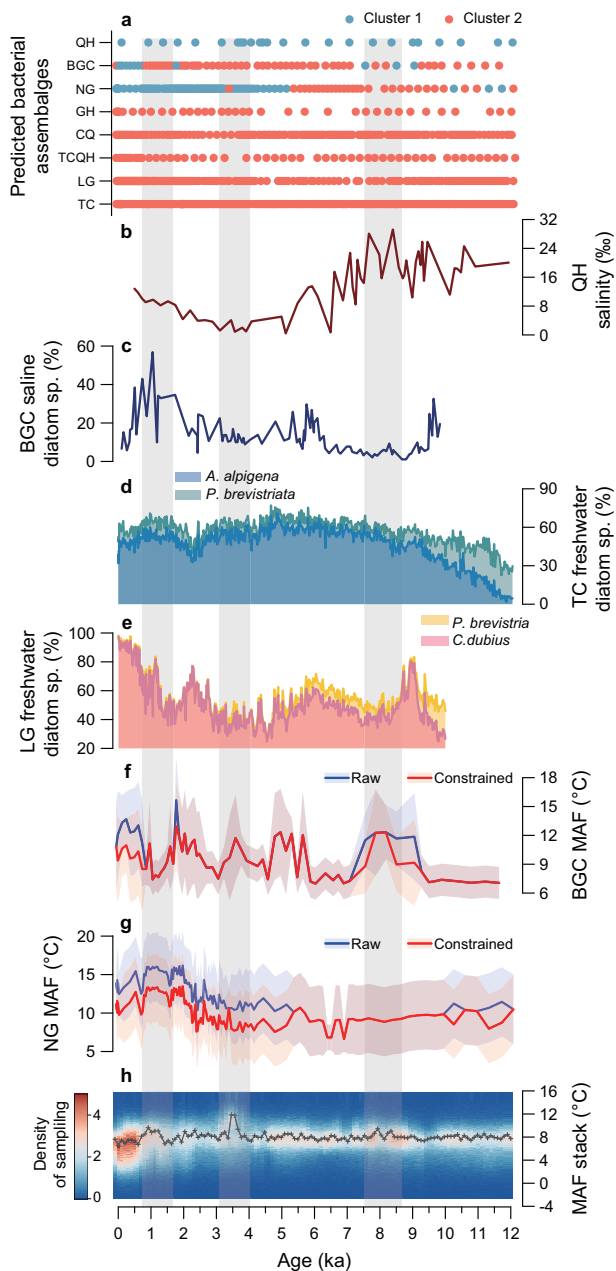
Fig. 7 | Holocene records including reconstructed bacterial assemblages, environmental conditions, and reconstructed temperature in lakes from Tibetan Plateau. a Reconstructed shifts in bacterial community clusters throughout the Holocene, inferred from the brGMM-BAE model. Published lake salinity and pH reconstructions employing other proxies: **b** QH lake, salinity reconstruction inferred from Ostracod Strontium/Calcium ratio[61]; **c** BGC lake, salinity inferred from oligosaline and mesosaline diatom species[64,101], including *Achnanthes affinis, Gyrosigma attenuatum, Navicula cincta, Navicula halophila, Navicula oblonga,* and *Nitzschia amphibia*; **d** Acidic and freshwater conditions inferred from diatom species (*Aulacoseira alpigena* and *Pseudostaurosira brevistriata*) in TC lake[63]; **e** LG lake freshwater conditions during the Holocene, inferred from the freshwater diatom species (*Pseudostaurosira brevistriata* and *Cyclostephanos dubius*)[62]. Holocene temperature reconstruction (blue line), based on RF calibration from lakes (**f**) BGC and (**g**) NG, with shaded area representing ±2σ uncertainty and temperature bias constraints derived from shifts in bacterial assemblages (red line). **h** BrGDGT-derived temperature stack based on RF calibration in lakes QH, TCQH, LG, TC, and CQ, where bacterial communities remained consistent throughout the Holocene. The gray shading bars highlight periods of temperature increase in the temperature stack.

predicted by the brGDGT indices from surface sediments. In addition, there is growing evidence that the accuracy of brGDGT-based temperature reconstruction is impacted by lake's abiotic and biotic factors such as salinity, pH, oxygen, and bacterial communities[12,31,47,49]. These findings emphasize the key contribution of bacterial community composition in shaping specific brGDGT profiles and prompt us to further consider the influence of environmental conditions, given bacteria's rapid response to environmental changes.

Understanding the interactions between environmental factors, bacterial community composition, and brGDGT profiles

The SEM model quantified the relative influence of abiotic and biotic factors on brGDGT indices (Fig. 4c). Air temperature is widely recognized as a primary driver of MBT'_{5Me} index[22,50]. The SEM analysis showed that air temperature not only affects MBT'_{5Me} directly but also indirectly through its influence on biotic components. Although we tested all variables in the SEM which may influence bacteria community or brGDGT indices, lake temperature did not exhibit a significantly direct effect on them. This could be due to salinity and pH that are dominant roles over other environmental factors in controlling both bacterial communities and brGDGT production. The SEM results also highlight a direct positive relationship between air temperature and lake surface water temperature, which in turn inversely affects lake salinity. Since meltwater is the primary source of these lakes in Tibetan Plateau, increased meltwater influx from rising temperatures reduces salinity, countering the traditional salinity increase from evaporation at high temperatures.

Furthermore, the SEM model demonstrates that salinity and pH impose selective pressure on the bacterial communities, with similar effects observed at both regional[51,52] and global scales[53,54]. Increased salinity and pH result in high relative abundance of haloalkaliphilic bacteria, such as members of the *Bacteroidetes* and *Gammaproteobacteria* phyla[51] and the families *Paracoccaceae, Saprospiraceae,* and *Spirochaetaceae*[55,56]. Our results are consistent with the other observations from Chinese lakes, which found that salinity and pH are crucial factors in both water and sediments, influencing bacterial diversity and community structure[57,58].

Significant differences between the two clusters in our samples were also found regarding the dissolved oxygen in lake bottom water (Fig. 4b). Our results revealed a phenomenon similar to that observed in culture experiments, where oxygen limitation can trigger the production of brGDGTs[31]. However, the SEM analysis did not show a significantly direct impact of dissolved oxygen on the bacterial communities or brGDGT indices. Instead, a direct correlation was identified between dissolved oxygen and salinity, which can be attributed to stratification driven primarily by salinity gradients in this region. This salinity-induced stratification leads to a decrease in dissolved oxygen at the bottom water of lake[59], suggesting that salinity gradients are the dominant driver influencing the ecological dynamics of lakes at mid-latitude. Therefore, the relationship between lake microbial communities and brGDGTs is more complex in lakes than observed in culture.

Our findings further indicate that changes in bacterial communities in response to lake salinity and pH, correlate with MBT'_{5Me} indices and IR values. This correlation with microbial trait distributions supports and extends previous studies[7,46], confirming that salinity and pH are critical factors controlling brGDGT distributions. A recent study of soil samples in the Okavango Delta found that all brGDGT concentrations increased with decreasing pH, and 5-methyl brGDGTs increased faster than 6-methyl brGDGTs in arid soils due to aridity-driven soil chemistry changes[60]. Taken together, these results highlight the crucial role of abiotic environmental factors, such as temperature, salinity, and pH, in shaping bacterial communities and, consequently, the brGDGT compounds they produce.

Inferring spatial and temporal ecological changes of bacterial assemblages using brGDGTs

The estimated bacterial assemblages based on the MBT'_{5Me} and IR indices from global surface sediments suggest spatial variations in lake salinity and

pH (Fig. 5a). Although the brGMM-BAE model shows strong potential in capturing spatial patterns of bacterial assemblages globally, it was primarily trained on data from lakes on the Tibetan Plateau. However, its accuracy could be further improved by incorporating more microbiological data from lakes across globally diverse environmental and ecological conditions. This is emphasized by the finding that around 24% of the sites showed partial matches, suggesting that some lakes may exhibit characteristics of either halophilic-acidic or freshwater-alkaline conditions. These partial matches highlight the intricate complexity and ecological diversity of lacustrine environments[52,57], which complicate the relationship between brGDGT indices and bacterial communities.

Over temporal scales, we observed changes in bacterial assemblages within lakes in the Tibetan Plateau, allowing us to reconstruct variations in bacterial communities across various lakes during the Holocene period. Lake QH was dominated by haloalkaliphilic bacterial assemblages, aligning with the salinity reconstruction inferred from the Sr/Ca ratios of fossil ostracods (Fig. 7b)[61]. Lakes TC and LG are characterized by freshwater bacterial assemblage (cluster 2) during the Holocene, which are consistent with freshwater diatom species also identified (Fig. 7d, e)[62,63].

Contrary to these stable conditions, lakes BGC and NG experienced significant shifts from freshwater to brackish conditions during the Holocene, which was reflected by the changes in bacterial communities derived from different brGDGT clusters (Fig. 7a). These transitions are supported by saline diatom assemblages from lake BGC[64] (Fig. 7c) and current chemical properties measurements—0.5 g/L salinity and pH 8.2 in BGC[65] and 5 g/L salinity and pH 9.5 in NG[42]. The topmost sample from lake BGC was reconstructed as freshwater, indicating that the lake is undergoing a transition from haloalkaline to freshwater conditions. Significant changes in the bacterial communities of the two lakes may be attributed to variations in lake conditions, shifting from freshwater conditions to haloalkaline conditions. It has been found that increases in salinity from freshwater habitats lead to more significant shifts in bacterial community composition compared to changes in habitats already characterized by high salinity[51]. Our results show strong correlations between bacterial assemblages and brGDGTs, revealing specific transitions in microbial communities that correspond with changes in freshwater and haloalkaline conditions. Modern spatial investigation of lakes across the Tibetan Plateau has similarly shown microbial communities adapting to changes in salinity[52]. These results highlight the adaptive capabilities of bacteria to their surroundings and emphasize the need to improve our understanding of the temporal dynamics that drive microbial adaptation to a changing environment.

A potential limitation in reconstructing past bacterial communities from sediment cores is the influence of active bacterial cells on ancient signals, potentially leading to overprinting[66]. Previous study has shown that lacustrine samples contain phosphohexose brGDGTs—a post-depositional source of brGDGTs that can persist in sediments for millennia[67]. This raises the possibility that in situ GDGT production in deeply buried sediments may be linked to the cell membranes of active prokaryotes, a process that cannot be entirely ruled out and could complicate the interpretation of paleoenvironmental signals[67,68]. In our study, we compared brGDGT-reconstructed bacterial assemblages with measured salinity and pH values across various spatial and temporal scales. We found consistent correlations between these environmental factors and the reconstructed microbial communities. This alignment supports the interpretation that the bacterial assemblages likely reflect past environmental conditions, though the potential influence of in situ microbial activity remains a consideration.

Linking microbial ecology to Holocene temperature reconstruction

The largest Holocene temperature fluctuations based on brGDGT data originated from lakes NG, BGC, and GH (Fig. 1). It has been found that temperature reconstructions based on soil brGDGTs in high-altitude regions (>3000 m) tend to overestimate temperatures[21]. To investigate whether the large variations in reconstructed temperatures result from the altitude amplification effect, we compared temperatures at similar altitudes.

For instance, a total temperature variation of 8.7 ± 2.1 °C was found in BGC at 4224 m throughout the Holocene, while other lakes at nearly the same altitudes, such as TC (3898 m) and CQ (3960 m), exhibit much smaller temperature changes of 4.2 ± 0.7 °C and 1.1 ± 0.2 °C, respectively (Fig. 1). Even though the lakes are at similar altitudes, their temperature changes are not the same. This suggests that the altitude amplification effect is unlikely to be the primary factor causing the observed site-specific discrepancies in reconstructed temperatures.

The reconstructed bacterial assemblages could explain the large brGDGT-based temperature fluctuations in lakes during the Holocene. Both BGC and NG exhibited pronounced species changes in response to salinity and pH shifts, leading to reconstructed temperature variations that were markedly greater than those reconstructed in other lakes (Figs. 1, 6). Despite the reconstruction of a single freshwater bacterial assemblage (cluster 2) in lake GH, large pH variations were still observed (Fig. 6d). This suggests that another bacterial assemblage, possibly a freshwater-alkaliphilic group, may have been present in the lake during the early and late Holocene (Fig. 6). The lake level reconstructed from grain sizes in the same core of lake GH reveals significant fluctuations throughout the Holocene[69]. Lake GH had particularly lower levels during the early and late Holocene compared to higher levels in the middle Holocene[69], likely contributing to changes in pH levels. It has been found that pH plays a critical role in freshwater lakes, determining the interplay between niche-related processes and influencing bacterial communities[57].

Constraining the temperature based on shifts in bacterial assemblages reduced the site-specific discrepancy of reconstructed temperatures in these two lakes (Fig. 7f, g). Compared with the temperature stack from lakes with a single bacterial assemblage (Fig. 7h), the NG and BGC lakes exhibited greater variation than the stack. BGC had a thermal maximum during the early Holocene around 8 ka, which also appeared in temperature stack. Most reconstructions have shown that there was a warming period during Bronze Age at around 3–4 ka, which was found also by summer temperature reconstructed using alkenone from lake Sayram at the northeastern Tibetan Plateau[70]. It has been found that humans settled at high altitudes on the northeastern Tibetan Plateau during this period (~3.6 ka)[71]. Our temperature stack and newly reconstructed records provide further evidence that favorable environmental conditions promoted expanded occupation of this high-altitude plateau. Another warm interval was found in our reconstructed temperature record around 1 ka, supported by pollen Artemisia/Chenopodiaceae records indicating a warm and dry climate during this period[72]. Additionally, diatom-based salinity records from BGC lake (Fig. 7c)[64], as well as ostracod-based (Fig. 7b)[61] and alkenone %37:4[73,74] records in QH lake showed increased salinity, further supporting the occurrence of higher evaporation caused by increased temperatures during this period.

Previous studies on globally distributed lake surface sediments have demonstrated substantial deviations in temperatures reconstructed from brGDGTs in mid-latitude lakes[5,7,9], raising concerns about the accuracy of lacustrine brGDGTs for quantitative temperature reconstructions in this region. Our findings suggest that lacustrine brGDGTs from mid-latitude lakes remain a valuable tool for reconstructing temperature. By constraining the biases caused by changes in lake bacterial ecology, we can effectively diagnose and constrain discrepancies in reconstructed temperatures, leading to a more accurate temperature record.

Perspectives of future lacustrine brGDGT reconstruction

Early efforts in brGDGT-temperature calibration primarily relied on advanced statistical methods, ranging from linear regression[22,50] to more sophisticated machine learning techniques[17,19,20], to better delineate the relationship between brGDGT compounds and temperature. For instance, loess reconstructions using global soil Bayesian neural network calibration have demonstrated improved accuracy, particularly during cold periods such as the Last Glacial Maximum[20]. However, these calibrations focus primarily on the correlation between temperature and brGDGTs, while overlooking other critical factors such as microbial community shifts and

environmental changes. For example, brGDGT-derived temperatures from Qinghai Lake were overestimated by as much as 10 °C during the early to mid-Holocene, likely due to changes in lake environments caused by shifts in brGDGT producers adapting to environmental variations[7].

To address these biases and improve the accuracy of lacustrine brGDGT-based temperature reconstructions, we recommend using the brGMM-BAE model, which is driven by the MBT'$_{5Me}$ and IR indices, to assess shifts in bacterial communities throughout the reconstruction time-scale. This model helps identify bacterial community shifts from freshwater to haloalkaliphilic assemblages, which influence temperature reconstructions. To mitigate this bias, we recommend applying a 2.7 ± 0.7 °C correction to the reconstructed temperatures during transition from freshwater to haloalkaliphilic assemblages. It should be noted that this method is limited by its focus on species shifts between haloalkaliphilic and freshwater assemblages, and we are still unable to isolate the specific community effects driven solely by salinity or pH on brGDGT-based reconstructions. Future research should not only focus on developing advanced statistical calibrations but also integrating ecological insights to refine the accuracy of paleoclimate reconstructions. Combining these ecological insights with AI-enhanced calibration offers a powerful approach to capturing the complex interactions between environmental variables and brGDGT distributions, significantly improving temperature reconstructions across diverse ecosystems.

Materials and methods

Surface sediments and temperature reconstructions from Tibetan Plateau and global sites

We collected 38 lake surface sediment samples in the Tibetan Plateau (longitude 63°-105°E and latitude 20°-45°N), each lake containing relative abundances of brGDGTs, limnological data, and DNA dataset (Supplementary data)[26,46,75]. Temperature reconstruction for these lake surface sediment samples was conducted using three calibrations: RF calibration[17], Bayesian MAF (BayMBT) calibration[9], and full-set MAF calibration[8].

To verify whether the pattern we discovered on the Tibetan Plateau is also present globally, we collected a total of 965 surface lake sediment brGDGT datasets from published papers spanning from 2017 to 2024 (Supplementary data)[6-9,11-13,16,17,25,26,43,50,76-86]. However, only 415 of these included both salinity (or conductivity) and pH data, so we used this dataset for global validation. Conductivity (μS/cm) was converted to salinity (g/L) using an empirical formula by multiplying it by 0.00065. We used a salinity threshold of 0.2 g/L to differentiate between brackish/saline and freshwater lakes[87], and a pH threshold of 8 to classify lakes as acidic/neutral or alkaline.

Regional Holocene temperature compilation

We spatially focused on Holocene records from the Tibetan Plateau, which were derived from this study and previously published studies. The lakes were carefully selected based on several criteria, including: 1) they were alpine lakes situated on the Tibetan Plateau, 2) the records from these lakes were analyzed both 5-methyl and 6-methyl brGDGTs, and 3) these records were continuous throughout the Holocene. Among the lakes that met these criteria, we analyzed eight of them: Bangong Co (BGC, this study), Cuoqia Lake (CQ)[45], Gahai Lake (GH)[41], Lugu Lake (LG)[43], Ngamring Co (NG)[42], Qinghai Lake (QH)[7], Tengchongqinghai Lake (TCQH)[43] and Tiancai Lake (TC)[44]. Geographical and limnological data for these eight lakes are provided in Supplementary Table 1. The chronology of BGC was based on studies of the same core[88,89], while the chronology of the other seven lakes was based on published papers[41-45]. These records were reanalyzed using the same global random forest calibration[17] to reconstruct Holocene temperatures.

To create the temperature stack used in this study, we re-ran the Bacon age model for these lakes. To ensure robust comparison and synthesis, we integrated the uncertainties from the age-depth models to interpolate reconstructed temperature values at regular 100-year intervals. Using a Monte Carlo framework, we sampled age and temperature values from their respective distributions for each sample, accounting for a wide range of possible scenarios. We then calculated the density of these scenarios for each 100-year time slice, to estimate climate reconstructions with associated uncertainties. Finally, we employed curve stacking to enhance the signal-to-noise ratio and reduce uncertainties from individual records[90]. The temperature stack was adjusted to the average altitude of 3600 m using an assumed constant lapse rate of 6 °C/km. Lhasa, a city in Tibet (3600 m a.s.l.), with a meteorological MAF of 10.0 °C (average from 1981 to 2020)[91] is used as the reference point.

Lipid extraction and analyses

Approximately 5–6 g of freeze-dried sediments from BGC were ultrasonically extracted for 15 min three times with dichloromethane (DCM): methanol (MeOH; 9:1, v/v) to obtain a total lipid extract (TLE). The TLE was separated over an Al$_2$O$_3$ column, employing n-hexane:DCM (9:1, v/v) and DCM:MeOH (1:1, v/v) as eluents for isolating nonpolar and polar fractions, respectively. The polar fraction, containing lipid GDGTs, was filtered using a 0.45 μm PTFE filter prior to analysis using high-performance liquid chromatography atmospheric pressure chemical ionization–mass spectrometry (HPLC-APCI–MS), according to previously published methods[92]. BrGDGTs were separated over three silica columns in sequence (100 × 2.1 mm, 1.9 μm; Thermo Fisher Scientific, USA) with a flow rate of 0.2 ml/min. For each sample, 10 μl was injected. The mobile phase consisted of 84% n-hexane and 16% ethyl acetate. The chromatography of separation was compared with two columns according to ref. 23 and provided in the Supplementary Fig. 5. Detection was performed in selected ion monitoring (SIM) mode via $[M + H]^+$ at m/z 744 for the internal standard, and 1050, 1048, 1046, 1036, 1034, 1032, 1022, 1020,1018 for brGDGTs. Peaks were identified manually, and peak areas were integrated manually.

The methylation index of branched tetraethers was developed (MBT'$_{5Me}$)[22]:

$$MBT'_{5Me} = \frac{Ia + Ib + Ic}{Ia + Ib + Ic + IIa + IIb + IIc + IIIa} \quad (1)$$

The ratio of 6- over 5-methyl brGDGTs was initially proposed as IR$_{6Me}$ ratio[22], which was later simplified as IR[93]:

$$IR = \frac{IIa' + IIIa'}{IIa + IIIa + IIa' + IIIa'} \quad (2)$$

Bioinformatics analyses

The taxonomic assignment was performed using the SILVA database (version 132 NR) and extract Operational Taxonomic Unit (OTUs) with >97% similarity[94]. To eliminate the influence of rare OTU in the analysis, we focused on the OTUs that were detected in at least two lakes and with a relative abundance above 0.01%. Approximately a total of 8000 OTUs were selected in each lake and used for further analysis.

BLASTP searches for tetraether synthase (Tes) homologs were conducted in the NCBI's non-redundant protein sequences (nr) database using the Tes (MA_1486) sequence of *Methanosarcina acetivorans* C2A as the search query and default parameters[40]. BLASTP was performed against the top 50 species we identified (e-value < 1e − 50, identity >20%, length >400 aa). The findings are presented in Fig. 3.

Statistic methods

GMM clustering, which is an unsupervised learning method capable of clustering without providing environmental factors other than brGDGT indices, was used to discover the distinct patterns in lake surface sediments from the Tibetan Plateau. The GMM accounts for the probabilistic distribution of the dataset by modeling the data as a combination of multiple Gaussian distributions, thereby capturing the data characteristics more effectively[95]. This property is especially suitable for nonlinear data distributions. GMM clustering method has been successfully employed in identifying thermal behaviors of archaeal glycerol dialkyl glycerol

tetraethers[39]. We used the GaussianMixture() function from the Scikit-learn package in Python to perform the GMM analysis[96]. The GMM process iteratively estimates multiple Gaussian distributions to model the joint distribution of IR and MBT$'_{5Me}$ indices, ensuring the best fit to the data. We assessed GMMs by testing cluster counts varying from 2 to 10. For each model, the Bayesian Information Criterion (BIC) and the silhouette score were computed to evaluate clustering quality. The optimal clustering distinctness, indicated by the highest silhouette score and the lowest BIC value, was observed at two components (Supplementary Fig. 2). Therefore, we selected to divide these lakes into two clusters based on the GMM clustering result for further analysis.

The non-parametric Kruskal-Wallis test was used to determine whether environmental variables and relative abundance of brGDGT compounds had significantly different distributions between the two clusters. We conducted a Kruskal-Wallis test on data from 38 lakes using the kruskal.test() function from the stats package in R. This test is recommended when predictor variables are not normally distributed or are spatially autocorrelated.

The structural equation model (SEM) analysis employed a piecewise SEM[97] to assess the direct and indirect influences of environmental factors and bacterial community on the brGDGT indices (MBT$'_{5Me}$ and IR). SEMs require the definition of a priori model that specifies the directionality of potential relationships between variables. Air temperature and lake environmental variables were considered as exogenous variables, as they directly influence both bacterial communities and brGDGT indices, with bacterial communities also acting as an intermediary that affects brGDGT indices. We initially characterized the environment with 15 variables, including depth, Secchi depth (SD), pH, salinity, surface water DO (DO$_{upper}$), bottom water DO (DO$_{bottom}$), averaged lake water column DO (DO$_{average}$), surface water Chlorophyll a (chl a$_{surface}$), bottom water Chlorophyll a (chl a$_{bottom}$), averaged water Chlorophyll a (chl a$_{average}$), lake surface water temperature (LST), lake bottom water temperature (LBT), averaged lake water temperature (LMT), mean annual air temperature (MAT), and mean temperatures of months with temperatures above 0 °C (MAF). To minimize collinearity, we reduced these to seven statistically independent variables using a pairwise correlation matrix (Supplementary Fig. 3). Further, we also evaluated these variables using the Kruskal-Wallis test, selecting those with significant differences for inclusion in our model. The environmental variables MAF, LST, salinity, pH, and DO$_{bottom}$ were thus included in our SEM model.

We employed linear discriminant analysis (LDA) to reduce the dimensionality of our bacterial dataset, which included the 50 most abundant bacterial species at the family level, and represented changes in the bacterial communities (Supplementary Fig. 4). Unlike principal component analysis (PCA), which identifies linear combinations of input variables that best explain the variance in the data without considering cluster labels, LDA explicitly uses these cluster labels to model and highlight differences between groups. This approach has been previously used to evaluate the relationship between environmental factors and bacterial community structure[98]. Our classifications were derived from GMM model, which grouped the species into two distinct clusters that reflect variations in bacterial community compositions. The LDA scores, calculated using the lda() function from the MASS package in R, facilitated a quantitative comparison of the community structures between the two distinct bacterial groups[99].

Finally, we utilized the gls() (Generalized Least Squares) function within the piecewiseSEM package in R[100] on strictly filtered data to establish SEM model. Our SEM model was developed with a series of structured equations, specified through linear modeling functions in R, enabling them to handle non-normal distributions, hierarchical structures, and a variety of estimation procedures. The lacustrine brGDGT indices were potentially affected by air temperature, lake water temperature, DO$_{bottom}$, pH, salinity, and bacterial communities. For the bacterial communities, we assumed a potential causal effect of DO$_{bottom}$, pH, and salinity. The goodness-of-fit tests for both the full and component models included calculating Akaike's information criterion (AIC) scores, returning scaled parameter estimates,

plotting partial correlations, and generating predictions. We evaluated model fit using Fisher's C and AIC statistic, which assesses all directed separation tests. The model also aided in identifying missing paths, leading to a more exploratory approach due to the absence of prior hypotheses for these paths. AIC-based model selection confirmed that our model, considering specific environmental factors and bacterial community, provided the best prediction of the mechanism of how these variables influence brGDGT indices.

Data availability
The dataset collected and compiled in this study is available in the Figshare repository: https://doi.org/10.6084/m9.figshare.26521102.v1.

Code availability
The code used for this manuscript is available in the following GitHub repository: https://github.com/jieliangbio/brGMM_BAE. For users unfamiliar with Python, we also provide a graphical user interface (GUI) for the brGMM-BAE model in the Figshare repository, enabling predictions without the need to install or run Python code: https://doi.org/10.6084/m9.figshare.27643836.v1.

References
1. Tierney, J. E. et al. Past climates inform our future. *Science* **370**, eaay3701 (2020).
2. Liu, Z. et al. The Holocene temperature conundrum. *Proc. Natl Acad. Sci.* **111**, E3501–E3505 (2014).
3. Kaufman, D. S. & Broadman, E. Revisiting the Holocene global temperature conundrum. *Nature* **614**, 425–435 (2023).
4. Raberg, J. H., Miller, G. H., Geirsdóttir, Á. & Sepúlveda, J. Near-universal trends in brGDGT lipid distributions in nature. *Sci. Adv.* **8**, eabm7625 (2022).
5. Lin, T. et al. Sedimentary brGDGTs in China: an overview of modern observations and proposed land Holocene paleotemperature records. *Earth-Sci. Rev.* **250**, 104694 (2024).
6. Zhao, B. et al. Evaluating global temperature calibrations for lacustrine branched GDGTs: seasonal variability, paleoclimate implications, and future directions. *Quat. Sci. Rev.* **310**, 108124 (2023).
7. Wang, H. et al. Salinity-controlled isomerization of lacustrine brGDGTs impacts the associated MBT5ME' terrestrial temperature index. *Geochim. Cosmochim Acta* **305**, 33–48 (2021).
8. Raberg, J. H. et al. Revised fractional abundances and warm-season temperatures substantially improve brGDGT calibrations in lake sediments. *Biogeosciences* **18**, 3579–3603 (2021).
9. Martínez-Sosa, P. et al. A global Bayesian temperature calibration for lacustrine brGDGTs. *Geochim. Cosmochim. Acta* **305**, 87–105 (2021).
10. Weber, Y. et al. Identification and carbon isotope composition of a novel branched GDGT isomer in lake sediments: Evidence for lacustrine branched GDGT production. *Geochim. Cosmochim. Acta* **154**, 118–129 (2015).
11. Wang, H. et al. Biomarker-based quantitative constraints on maximal soil-derived brGDGTs in modern lake sediments. *Earth Planet. Sci. Lett.* **602**, 117947 (2023).
12. Weber, Y. et al. Redox-dependent niche differentiation provides evidence for multiple bacterial sources of glycerol tetraether lipids in lakes. *Proc. Natl Acad. Sci.* **115**, 10926–10931 (2018).
13. Stefanescu, I. C., Shuman, B. N. & Tierney, J. E. Temperature and water depth effects on brGDGT distributions in sub-alpine lakes of mid-latitude North America. *Org. Geochem.* **152**, 104174 (2021).
14. Naafs, B. D. A., Gallego-Sala, A. V., Inglis, G. N. & Pancost, R. D. Refining the global branched glycerol dialkyl glycerol tetraether

(brGDGT) soil temperature calibration. *Org. Geochem.* **106**, 48–56 (2017).

15. Raberg, J. H. et al. BrGDGT lipids in cold regions reflect summer soil temperature and seasonal soil water chemistry. *Geochim. Cosmochim Acta* **369**, 111–125 (2024).

16. Cao, J., Rao, Z., Shi, F. & Jia, G. Ice formation on lake surfaces in winter causes warm-season bias of lacustrine brGDGT temperature estimates. *Biogeosciences* **17**, 2521–2536 (2020).

17. O'Beirne M. D., Scott W. P., Werne J. P. A critical assessment of lacustrine branched glycerol dialkyl glycerol tetraether (brGDGT) temperature calibration models. *Geochim. Cosmochim. Acta* (2023).

18. Martınez-Sosa P., et al. Development and application of the branched and isoprenoid GDGT machine learning classification algorithm (BIGMaC) for paleoenvironmental reconstruction. *Paleoceanogr. Paleoclimatol.* **38**, e2023PA004611(2023).

19. Véquaud, P. et al. FROG: A global machine-learning temperature calibration for branched GDGTs in soils and peats. *Geochim. Cosmochim. Acta* **318**, 468–494 (2022).

20. Häggi, C. et al. GDGT distribution in tropical soils and its potential as a terrestrial paleothermometer revealed by Bayesian deep-learning models. *Geochim. Cosmochim. Acta* **362**, 41–64 (2023).

21. Pérez-Angel, L. C. et al. Soil and air temperature calibrations using branched GDGTs for the tropical Andes of Colombia: toward a pan-tropical calibration. *Geochem. Geophys. Geosyst.* **21**, e2020GC008941 (2020).

22. De Jonge, C. et al. Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: implications for palaeoclimate reconstruction. *Geochim. Cosmochim. Acta* **141**, 97–112 (2014).

23. Hopmans, E. C., Schouten, S. & Sinninghe Damsté, J. S. The effect of improved chromatography on GDGT-based palaeoproxies. *Org. Geochem.* **93**, 1–6 (2016).

24. Dang, X., Yang, H., Naafs, B. D. A., Pancost, R. D. & Xie, S. Evidence of moisture control on the methylation of branched glycerol dialkyl glycerol tetraethers in semi-arid and arid soils. *Geochim. Cosmochim. Acta* **189**, 24–36 (2016).

25. Dang, X. et al. Different temperature dependence of the bacterial brGDGT isomers in 35 Chinese lake sediments compared to that in soils. *Org. Geochem.* **119**, 72–79 (2018).

26. Liang, J. et al. Calibration and application of branched GDGTs to Tibetan lake sediments: the influence of temperature on the fall of the Guge Kingdom in western Tibet, China. *Paleoceanogr. Paleoclimatol.* **37**, e2021PA004393 (2022).

27. Bittner, L. et al. A Holocene temperature (brGDGT) record from Garba Guracha, a high-altitude lake in Ethiopia. *Biogeosciences* **19**, 5357–5374 (2022).

28. Wang, H., An, Z., Lu, H., Zhao, Z. & Liu, W. Calibrating bacterial tetraether distributions towards in situ soil temperature and application to a loess-paleosol sequence. *Quat. Sci. Rev.* **231**, 106172 (2020).

29. Wang, H., Liu, Z., Zhao, H., Lu, H. & Liu, X. New calibration of terrestrial brGDGT paleothermometer deconvolves distinct temperature responses of two isomer sets. *Earth Planet. Sci. Lett.* **626**, 118497 (2024).

30. Chen, Y. et al. The production of diverse brGDGTs by an Acidobacterium providing a physiological basis for paleoclimate proxies. *Geochim. Cosmochim. Acta* **337**, 155–165 (2022).

31. Halamka, T. A. et al. Production of diverse brGDGTs by Acidobacterium Solibacter usitatus in response to temperature, pH, and O2 provides a culturing perspective on brGDGT proxies and biosynthesis. *Geobiology* **21**, 102–118 (2022).

32. Martínez-Sosa, P. & Tierney, J. E. Lacustrine brGDGT response to microcosm and mesocosm incubations. *Org. Geochem.* **127**, 12–22 (2019).

33. Martínez-Sosa, P., Tierney, J. E. & Meredith, L. K. Controlled lacustrine microcosms show a brGDGT response to environmental perturbations. *Org. Geochem.* **145**, 104041 (2020).

34. Chen, F. et al. Climate change, vegetation history, and landscape responses on the Tibetan Plateau during the Holocene: a comprehensive review. *Quat. Sci. Rev.* **243**, 106444 (2020).

35. Zeng, Z. et al. GDGT cyclization proteins identify the dominant archaeal sources of tetraether lipids in the ocean. *Proc. Natl Acad. Sci.* **116**, 22505–22511 (2019).

36. Acharya, S. et al. Environmental controls on the distribution of GDGT molecules in Lake Höglwörth, Southern Germany. *Org. Geochem.* **186**, 104689 (2023).

37. Martin, C. et al. Impact of human activities and vegetation changes on the tetraether sources in Lake St Front (Massif Central, France). *Org. Geochem.* **135**, 38–52 (2019).

38. Naafs, B. D. A., Oliveira, A. S. F. & Mulholland, A. J. Molecular dynamics simulations support the hypothesis that the brGDGT paleothermometer is based on homeoviscous adaptation. *Geochim. Cosmochim. Acta* **312**, 44–56 (2021).

39. Rattanasriampaipong, R., Zhang, Y. G., Pearson, A., Hedlund, B. P. & Zhang, S. Archaeal lipids trace ecology and evolution of marine ammonia-oxidizing archaea. *Proc. Natl Acad. Sci.* **119**, e2123193119 (2022).

40. Zeng, Z. et al. Identification of a protein responsible for the synthesis of archaeal membrane-spanning GDGT lipids. *Nat. Commun.* **13**, 1545 (2022).

41. Hou, X. et al. BrGDGTs-based seasonal paleotemperature reconstruction for the last 15,000 years from a shallow lake on the eastern Tibetan Plateau. *Clim. Discuss* **2023**, 1–34 (2023).

42. Sun, Z. et al. Potential winter-season bias of annual temperature variations in monsoonal Tibetan Plateau since the last deglaciation. *Quat. Sci. Rev.* **292**, 107690 (2022).

43. Zhao, C. et al. Possible obliquity-forced warmth in southern Asia during the last glacial stage. *Sci. Bull.* **66**, 1136–1145 (2021).

44. Feng, X. et al. Evidence for a Relatively warm mid-to late holocene on the Southeastern Tibetan Plateau. *Geophys. Research. Lett.* **49**, e2022GL098740 (2022).

45. Zhang C., et al. Seasonal imprint of Holocene temperature reconstruction on the Tibetan Plateau. *Earth-Sci. Rev.* **226**, 103927 (2022).

46. Kou, Q. et al. Influence of salinity on glycerol dialkyl glycerol tetraether-based indicators in Tibetan Plateau lakes: implications for paleotemperature and paleosalinity reconstructions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **601**, 111127 (2022).

47. Liang, J. et al. Branched glycerol dialkyl glycerol tetraether (brGDGT) distributions influenced by bacterial community composition in various vegetation soils on the Tibetan Plateau. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **611**, 111358 (2023).

48. Chen, Y. et al. Potential influence of bacterial community structure on the distribution of brGDGTs in surface sediments from Yangtze River Estuary to East China Sea. *Chem. Geol.* **647**, 121934 (2024).

49. Wu, J. et al. Variations in dissolved O_2 in a Chinese lake drive changes in microbial communities and impact sedimentary GDGT distributions. *Chem. Geol.* **579**, 120348 (2021).

50. Russell, J. M., Hopmans, E. C., Loomis, S. E., Liang, J. & SinningheDamsté, J. S. Distributions of 5-and 6-methyl branched glycerol dialkyl glycerol tetraethers (brGDGTs) in East African lake sediment: Effects of temperature, pH, and new lacustrine paleotemperature calibrations. *Org. Geochem.* **117**, 56–69 (2018).

51. Rath, K. M., Fierer, N., Murphy, D. V. & Rousk, J. Linking bacterial community composition to soil salinity along environmental gradients. *ISME J.* **13**, 836–846 (2019).

52. Liu, Y. et al. Salinity impact on bacterial community composition in five high-altitude lakes from the Tibetan Plateau, Western China. *Geomicrobiol. J.* **30**, 462–469 (2013).

53. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl Acad. Sci.* **104**, 11436–11440 (2007).

54. Shaffer, J. P. et al. Standardized multi-omics of Earth's microbiomes reveals microbial and metabolite diversity. *Nat. Microbiol.* **7**, 2128–2150 (2022).

55. Song, T. et al. Salinity gradient controls microbial community structure and assembly in coastal solar salterns. *Genes* **13**, 385 (2022).

56. Gao, L. et al. Paracoccus salsus sp. nov., a novel slightly halophilic bacterium isolated from saline lake sediment. *Int. J. System. Evol. Microbiol.* **72**, 005473 (2022).

57. Ren, L. et al. pH influences the importance of niche-related and neutral processes in lacustrine bacterioplankton assembly. *Appl. Environ. Microbiol.* **81**, 3104–3114 (2015).

58. Yang, J. et al. Distinct co-occurrence patterns of prokaryotic community between the waters and sediments in lakes with different salinity. *FEMS Microbiol. Ecol.* **97**, fiaa234 (2021).

59. Liang, C. et al. Salinization mechanism of lakes and controls on organic matter enrichment: from present to deep-time records. *Earth-Sci. Rev.* **251**, 104720 (2024).

60. Lattaud, J., Gondwe, M. J., Griepentrog, M., Helfter, C. & De Jonge, C. Soil chemistry effect on GDGT abundances and their proxies in soils of the Okavango Delta. *Org. Geochem.* **195**, 104847 (2024).

61. Zhang, P., Zhang, B., Qian, H., Li, G. & Xu, L. The study of paleoclimatic parameter of Qinghai Lake since Holocene. *Quat. Sci.* **14**, 225–238 (1994).

62. Wang, Q., Yang, X., Anderson, N. J. & Dong, X. Direct versus indirect climate controls on Holocene diatom assemblages in a sub-tropical deep, alpine lake (Lugu Hu, Yunnan, SW China). *Quat. Res.* **86**, 1–12 (2016).

63. Chen, X. et al. Diatom response to Asian monsoon variability during the Late Glacial to Holocene in a small treeline lake, SW China. *Holocene* **24**, 1369–1377 (2014).

64. Gasse, F., Fontes, J. C., Van Campo, E. & Wei, K. Holocene environmental changes in Bangong Co basin (Western Tibet). Part 4: discussion and conclusions. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **120**, 79–92 (1996).

65. Wang, M., Hou, J. & Lei, Y. Classification of Tibetan lakes based on variations in seasonal lake water temperature. *Chin. Sci. Bull.* **59**, 4847–4855 (2014).

66. Capo, E. et al. Environmental paleomicrobiology: using DNA preserved in aquatic sediments to its full potential. *Environ. Microbiol.* **24**, 2201–2209 (2022).

67. Raberg, J. H. et al. Intact Polar brGDGTs in Arctic lake catchments: implications for lipid sources and paleoclimate applications. *J. Geophys. Res.: Biogeosci.* **127**, e2022JG006969 (2022).

68. Lengger, S. K., Hopmans, E. C., Sinninghe Damsté, J. S. & Schouten, S. Fossilization and degradation of archaeal intact polar tetraether lipids in deeply buried marine sediments (Peru Margin). *Geobiology* **12**, 212–220 (2014).

69. Wang, N. et al. Palynological evidence reveals an arid early Holocene for the northeast Tibetan Plateau. *Clim* **18**, 2381–2399 (2022).

70. Jiang, Q. et al. Exceptional terrestrial warmth around 4200–2800 years ago in Northwest China. *Sci. Bull.* **67**, 427–436 (2022).

71. Chen, F. et al. Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 BP. *science* **347**, 248–250 (2015).

72. Jiang, Q. et al. Holocene vegetational and climatic variation in westerly-dominated areas of Central Asia inferred from the Sayram Lake in northern Xinjiang, China. *Sci. China Earth Sci.* **56**, 339–353 (2013).

73. Wang, Z., Liu, Z., Zhang, F., Fu, M. & An, Z. A new approach for reconstructing Holocene temperatures from a multi-species long chain alkenone record from Lake Qinghai on the northeastern Tibetan Plateau. *Org. Geochem.* **88**, 50–58 (2015).

74. Hou, J. et al. Large Holocene summer temperature oscillations and impact on the peopling of the northeastern Tibetan Plateau. *Geophys. Res. Lett.* **43**, 1323–1330 (2016).

75. Yan, Q. et al. Distinct strategies of the habitat generalists and specialists in sediment of Tibetan lakes. *Environ. Microbiol.* **24**, 4153–4166 (2022).

76. Bauersachs, T., Schubert, C. J., Mayr, C., Gilli, A. & Schwark, L. Branched GDGT-based temperature calibrations from Central European lakes. *Sci. Total Environ.* **906**, 167724 (2024).

77. Dang, X., Xue, J., Yang, H. & Xie, S. Environmental impacts on the distribution of microbial tetraether lipids in Chinese lakes with contrasting pH: implications for lacustrine paleoenvironmental reconstructions. *Sci. China Earth Sci.* **59**, 939–950 (2016).

78. Dugerdil, L. et al. Climate reconstructions based on GDGT and pollen surface datasets from Mongolia and Baikal area: calibrations and applicability to extremely cold–dry environments over the Late Holocene. *Climate* **17**, 1199–1226 (2021).

79. Li, J. et al. Distribution of branched tetraether lipids in ponds from Inner Mongolia, NE China: insight into the source of brGDGTs. *Org. Geochem.* **112**, 127–136 (2017).

80. Miller, D. R., Habicht, M. H., Keisling, B. A., Castañeda, I. S. & Bradley, R. S. A 900-year New England temperature reconstruction from in situ seasonally produced branched glycerol dialkyl glycerol tetraethers (brGDGTs). *Climate* **14**, 1653–1667 (2018).

81. Ning, D. et al. Holocene mean annual air temperature (MAAT) reconstruction based on branched glycerol dialkyl glycerol tetraethers from Lake Ximenglongtan, southwestern China. *Org. Geochem.* **133**, 65–76 (2019).

82. Qian, S. et al. Rapid response of fossil tetraether lipids in lake sediments to seasonal environmental variables in a shallow lake in central China: Implications for the use of tetraether-based proxies. *Org. Geochem.* **128**, 108–121 (2019).

83. Wu J., et al. BrGDGT-based quantitative reconstructions of paleotemperature in lakes: regional vs. site-specific calibrations. *Quat. Sci. Rev.* **322**, 108416 (2023).

84. Yao, Y. et al. Correlation between the ratio of 5-methyl hexamethylated to pentamethylated branched GDGTs (HP5) and water depth reflects redox variations in stratified lakes. *Org. Geochem.* **147**, 104076 (2020).

85. Zhao, B. et al. Development of an in situ branched GDGT calibration in Lake 578, southern Greenland. *Org. Geochem.* **152**, 104168 (2021).

86. Lei, Y. et al. Regional vs. global temperature calibrations for lacustrine BrGDGTs in the North American (sub)tropics: Implications for their application in paleotemperature reconstructions. *Org. Geochem.* **184**, 104660 (2023).

87. Hammer U. T. *Saline Lake Ecosystems of the World* (Springer, 1986).

88. Wang, M. et al. Internal feedback forced Middle Holocene cooling on the Qinghai-Tibetan Plateau. *Boreas* **50**, 1116–1130 (2021).

89. Hou, J., D'Andrea, W. J., Wang, M., He, Y. & Liang, J. Influence of the Indian monsoon and the subtropical jet on climate change on the Tibetan Plateau since the late Pleistocene. *Quat. Sci. Rev.* **163**, 84–94 (2017).

90. Chevalier, M. & Chase, B. M. Southeast African records reveal a coherent shift from high- to low-latitude forcing mechanisms along the East African margin across last glacial–interglacial transition. *Quat. Sci. Rev.* **125**, 117–130 (2015).

91. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *J. Atmos. Ocean. Technol.* **29**, 897–910 (2012).

92. Yang, H. et al. The 6-methyl branched tetraethers significantly affect the performance of the methylation index (MBT′) in soils from an altitudinal transect at Mount Shennongjia. *Org. Geochem.* **82**, 42–53 (2015).

93. De Jonge, C. et al. The influence of soil chemistry on branched tetraether lipids in mid- and high latitude soils: Implications for brGDGT-based paleothermometry. *Geochim. Cosmochim. Acta* **310**, 95–112 (2021).

94. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).

95. Bishop C. M., Nasrabadi N. M. *Pattern Recognition and Machine Learning* (Springer, 2006).

96. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

97. Lefcheck J., Byrnes J., Grace J. Package 'piecewiseSEM'. *R package version* **1** (2016).

98. He S., et al Nitrogen loading effects on nitrification and denitrification with functional gene quantity/transcription analysis in biochar packed reactors at 5 °C. *Sci. Rep.* **8**, 9844 (2018).

99. Ripley, B. et al. Package 'mass'. *Cran r.* **538**, 113–120 (2013).

100. Lefcheck, J. S. piecewiseSEM: piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).

101. Fan, H. et al. Holocene environmental changes in Bangong Co basin (Western Tibet). Part 3: biogenic remains. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **120**, 65–78 (1996).

Author contributions

J. Liang: Conceptualization, methodology, investigation & writing-original draft. M. Chevalier: Methodology & visualization. K. Liu: Resources & writing-original draft. A. Perfumo: Conceptualization & writing-original draft. M. Wang: Methodology, writing-original draft. J. Hou, H. Xie, U. Herzschuh and F. Chen: Conceptualization, writing-original draft, supervision, project administration & funding acquisition.

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/s43247-024-01925-3.

Correspondence and requests for materials should be addressed to Jie Liang, Haichao Xie, Juzhi Hou or Ulrike Herzschuh.

Peer review information *Communications Earth & Environment* thanks Lina Pérez-Angel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Yiming Wang and Carolina Ortiz Guerrero. [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.