Geochimica et Cosmochimica Acta 362 (2023) 41–64

Contents lists available at ScienceDirect

Geochimica et Cosmochimica Acta

journal homepage: www.elsevier.com/locate/gca

GDGT distribution in tropical soils and its potential as a terrestrial paleothermometer revealed by Bayesian deep-learning models

Christoph Häggi [a,b,*], B. David A. Naafs [c], Daniele Silvestro [d,e,f], Dailson J. Bertassoli Jr. [g,h], Thomas K. Akabane [h], Vinícius R. Mendes [i], André O. Sawakuchi [h], Cristiano M. Chiessi [g], Carlos A. Jaramillo [j], Sarah J. Feakins [a]

[a] Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA
[b] Geological Institute, ETH Zürich 8092, Zürich, Switzerland
[c] Organic Geochemistry Unit, School of Chemistry, School of Earth Sciences, University of Bristol, Bristol, UK
[d] Department of Biology, University of Fribourg, Fribourg, Switzerland
[e] Swiss Institute of Bioinformatics, Fribourg, Switzerland
[f] Department of Biological and Environmental Sciences and Gothenburg Global Biodiversity Centre, University of Gothenburg, Gothenburg, Sweden
[g] School of Arts, Sciences and Humanities, University of São Paulo, São Paulo SP, Brazil
[h] Institute of Geosciences, University of São Paulo, São Paulo SP, Brazil
[i] Institute of Marine Science, Federal University of São Paulo, Santos, Brazil
[j] Smithsonian Tropical Research Institute, Ancón, Panama

ARTICLE INFO

Associate editor: Thomas Wagner

Keywords:
GDGT
Temperature
Precipitation
Vegetation
Machine learning

ABSTRACT

Branched and isoprenoidal glycerol dialkyl glycerol tetraethers (br- and isoGDGTs) are membrane lipids produced by bacteria and archaea, respectively. These lipids form the basis of several frequently used paleoclimatic proxies. For example, the degree of methylation of brGDGTs (MBT'_{5Me}) preserved in mineral soils (as well as peats and lakes) is one of the most important terrestrial paleothermometers, but features substantial variability that is so far insufficiently constrained. The distribution of isoGDGTs in mineral soils has received less attention and applications have focused on the use of the relative abundance of the isoGDGT crenarchaeol versus brGDGTs (BIT index) as an indicator of aridity. To expand our knowledge of the factors that can impact the br- and isoGDGT distribution in mineral soils, including the MBT'_{5Me} index, and to improve isoGDGT-based precipitation reconstructions, we surveyed the GDGT distribution in a large collection of mineral surface soils (n = 229) and soil profiles (n = 22) across tropical South America. We find that the MBT'_{5Me} index is significantly higher in grassland compared to forest soils, even among sites with the same mean annual air temperature. This is likely a result of a lack of shading in grasslands, leading to warmer soils. We also find a relationship between MBT'_{5Me} and soil pH in tropical soils. Together with existing data from arid areas in mid-latitudes, we confirm the relationship between the BIT-index and aridity, but also find that the isoGDGT distribution alone is aridity dependent. The combined use of the BIT-index and isoGDGTs can strengthen reconstructions of past precipitation in terrestrial archives. In terms of site-specific variations, we find that the variability in BIT and MBT'_{5Me} is larger at sites that show on average lower BIT and MBT'_{5Me} values. In combination with modelling results, we suggest that this pattern arises from the mathematical formulation of these proxies that amplifies variability for intermediate values and mutes it for values close to saturation (value of 1). Soil profiles show relatively little variation with depth for the brGDGT indices. On the other hand, the isoGDGT distribution changes significantly with depth as does the relative abundance of br- versus isoGDGTs. This pattern is especially pronounced for the $isoGDGT_{IsomerIndex}$ where deeper soil horizons show a near absence of isoGDGT isomers. This might be driven by archaeal community changes in different soil horizons, potentially driven by the difference between aerobic and anaerobic archaeal communities. Finally, we use our extensive new dataset and Bayesian neural networks (BNNs) to establish new brGDGT-based temperature models. We provide a tropical soil calibration that removes the pH dependence of tropical soils (n = 404; RMSE = 2.0 °C) and global peat and soil models calibrated against the temperature of the months above freezing (n = 1740; RMSE = 2.4) and mean annual air temperature (n = 1740; RMSE = 3.6). All models correct for the bias found in arid samples. We also successfully test the new

* Corresponding author.
E-mail address: christoph.haeggi@gmx.ch (C. Häggi).

https://doi.org/10.1016/j.gca.2023.09.014
Received 4 September 2022; Accepted 17 September 2023
Available online 20 September 2023

calibrations on Chinese loess records and tropical river sediments. Overall, the new calibrations provide improved temperature reconstructions for terrestrial archives.

1. Introduction

Branched and isoprenoidal glycerol dialkyl glycerol tetraethers (br- and isoGDGTs) are membrane lipids produced by bacteria and archaea that are ubiquitously found in aquatic and terrestrial environments (Koga et al., 1993; Schouten et al., 2013; Sinninghe Damsté et al., 2000; Weijers et al., 2007). In soils, brGDGTs synthesized by bacteria are usually dominant, while isoGDGTs produced by archaea are typically dominant in marine settings (Schouten et al., 2013; Weijers et al., 2006). Although the isoGDGT crenarchaeol is thought to be synthesized uniquely by ammonia oxidizing Thaumarchaeota (Sinninghe Damsté et al., 2012; Sinninghe Damsté et al., 2002), the origin of the other isoGDGTs and brGDGTs and their isomers is debated and not always well constrained, especially in mineral soils and peats (Blewett et al., 2020; Chen et al., 2022; De Jonge et al., 2021; Naafs et al., 2019; Sinninghe Damsté et al., 2018).

The relative distribution of brGDGTs has been found to be temperature and pH dependent in mineral soils (De Jonge et al., 2014; Wang et al., 2018; Weijers et al., 2007), lakes (Pearson et al., 2011), speleothems (Blyth and Schouten, 2013; Yang et al., 2011) and peats (Naafs et al., 2017b). The temperature dependence of brGDGTs is likely the effect of homeoviscous adaption as indicated by molecular dynamic simulations of membranes consisting of brGDGTs (Naafs et al., 2021) and preliminary culture studies (Chen et al., 2022). Early temperature calibrations that used brGDGTs had to be corrected for pH (Weijers et al., 2007) and exhibited a bias to cooler temperatures in arid regions (Peterse et al., 2012; Weijers et al., 2007). Advances in chromatography allowed for the exclusion of pH and aridity dependent 6-methylated brGDGTs, using only the 5-methyl isomers in the MBT'$_{5Me}$ index (De Jonge et al., 2013; De Jonge et al., 2014). Despite these improvements, the MBT'$_{5Me}$ proxy for mean annual air temperature (MAAT) still carries large uncertainties, especially in mid-latitudes, suggesting additional influences other than MAAT (Dearing Crampton-Flood et al., 2020; Naafs et al., 2017a; Véquaud et al., 2022). While large temperature seasonality, varying pH conditions, vegetation type and temperatures below freezing are known confounding factors, the impact of changing microbial diversity has been suggested but remains speculative (De Jonge et al., 2021; Liang et al., 2019; Peterse et al., 2014; Raberg et al., 2022; Wang et al., 2020; Weijers et al., 2011).

IsoGDGTs are generally less abundant compared to brGDGTs in mineral soils and peats and have received less attention in soil proxy calibration studies, with some notable exceptions (Blewett et al., 2020;

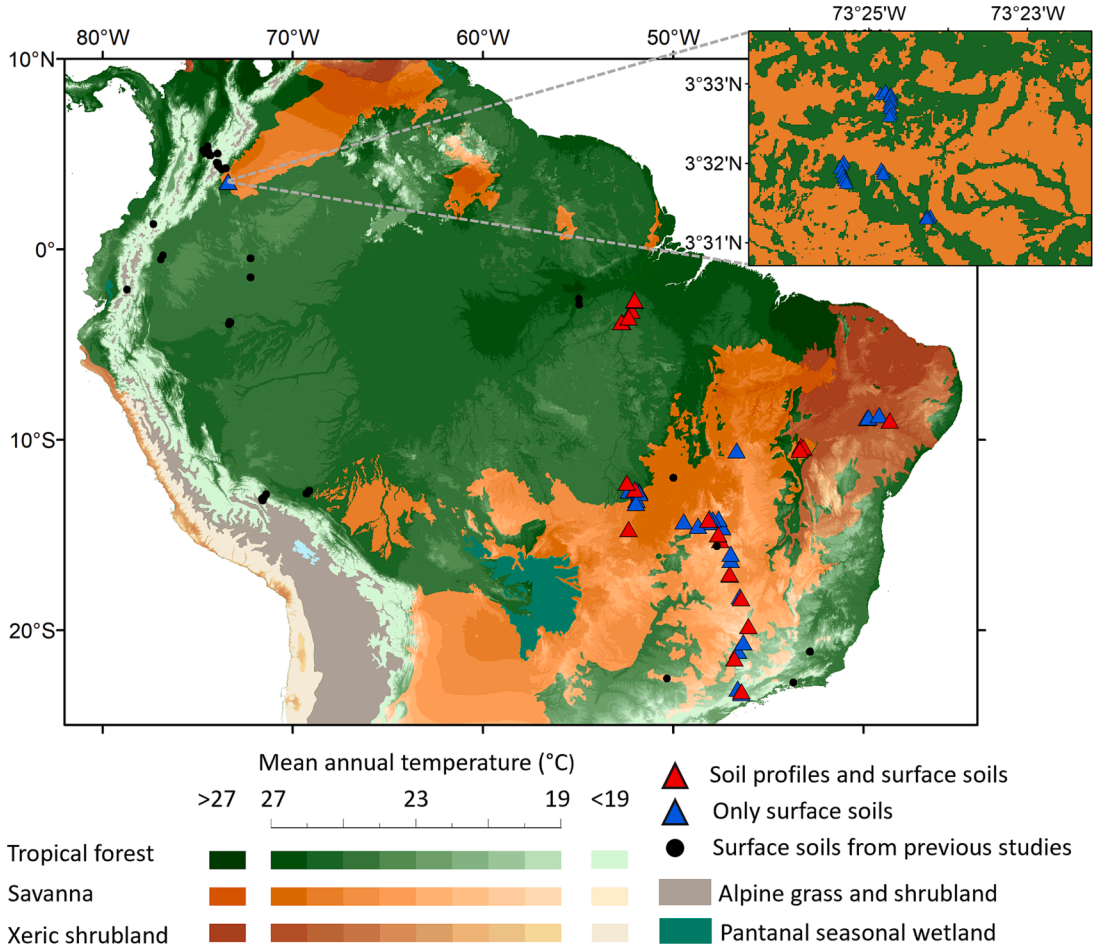

Fig. 1. Vegetation and mean annual temperature map of tropical South America (Karger et al., 2017; Olson et al., 2001). Colors (see legend) differentiate the major biomes (Olson et al., 2001), and intensity (see legend) refers to mean annual temperature. The small inset map provides a detailed overview of the sample locations in the Llanos basin. Sample locations are shown with triangles and locations of previous soil studies by De Jonge et al. (2014), Pérez-Angel et al. (2020) and Kirkels et al. (2020) are shown as black dots.

Dirghangi et al., 2013; Naafs et al., 2018; Xie et al., 2012). In the terrestrial realm, isoGDGTs have been mainly used in the BIT index and the $R_{i/b}$ ratio, both of which reflect the relative contribution of certain isoGDGTs versus brGDGTs and have been used as proxies for aridity in mineral soils (Dirghangi et al., 2013; Tang et al., 2017; Xie et al., 2012). Improved chromatography has led to the detection of isoGDGT isomers in certain soils (Becker et al., 2013; Blewett et al., 2020; Hopmans et al., 2016; Pitcher et al., 2009). In peat, the relative contribution of isoGDGT isomers appears to be pH dependent (Blewett et al., 2020) and Thaumarchaeota culture experiments also found a temperature dependence of the relative abundance of isoGDGT isomers (Bale et al., 2019). There are, however, no reports yet of these isoGDGTs isomers in mineral soils.

Here we undertake a survey of br- and isoGDGT distributions in mineral surface soils and soil profiles from across lowland tropical South America (Fig. 1). The vegetation mosaic of the Cerrado and the Llanos savannas and the adjacent rainforests allows for the study of the impact of pH and vegetation on the soil GDGT distribution. In addition, the region covers a major precipitation gradient from rainforest to the semi-arid Caatinga shrubland areas. This is important as open savanna and shrubland soils are generally understudied for GDGT-based proxies. So far, only a few samples from lowland rainforests and savanna soils have been analyzed for their brGDGT distributions and paleoenvironmental reconstructions have mainly focused on the Amazon Basin (Bendle et al., 2010; Häggi et al., 2019).

In our study, we test the influence of vegetation, climate, and soil properties among the major biomes on the GDGT distribution. As soils are known to be locally heterogeneous, we analyze multiple samples per site and samples from different soil horizons to study local soil heterogeneity. We also make use of advancements in numerical analysis methods to extract information from complex systems beyond linear regressions (Véquaud et al., 2022) and present Bayesian neural network (BNN) based models for improved brGDGT-based temperature reconstructions.

2. Materials and methods

2.1. Study area

The study area covers mineral soils across tropical South America between 5 °N and 25 °S and includes all major biomes of tropical South America (Fig. 1). Namely the Amazon and Atlantic rainforests, the Caatinga shrubland, and the savannas and grasslands of the Cerrado and the savanna and grassland mosaics of the Llanos basin. The Cerrado savanna features a continuum of savanna vegetation types varying between dry forests (Cerradão), mixed savanna with trees and interspersed grasses (Cerrado sensu stricto), savannas with small trees and shrubs (Campo Cerrado) and open grasslands of the Campo Sujo and Campo Limpo (Goodland, 1971). Likewise, the Llanos grassland consists of a variety of different types of vegetation density (Blydenstein, 1967). In this study we focus on the comparison between soils from open grassland versus those from riparian forest areas. Mean annual air temperatures (MAAT) in the study area varied between 19 °C in the mountainous areas (altitude up to 1000 m) of southeastern Brazil that are at the southern edge of our transect and 27 °C in the semi-arid Caatinga shrubland areas in northeastern Brazil (Karger et al., 2017) (Fig. 2). The highest mean annual precipitation (MAP) is found in the Llanos basin (>3000 mm y^{-1}), while the driest sites were sampled in the Caatinga shrubland (<500 mm y^{-1}) (Karger et al., 2017) (Fig. 2). The drivers of the dominance of forest and savanna vegetation in tropical South America are still debated as forest-savanna boundaries are not necessarily following clear climatological thresholds. On a local scale, water availability controlled by soil properties can play a role (Ruggiero et al., 2002), while hysteresis processes have been put forward as an explanation of the current rainforest-savanna boundary (Staver et al., 2011; Wuyts et al., 2017).

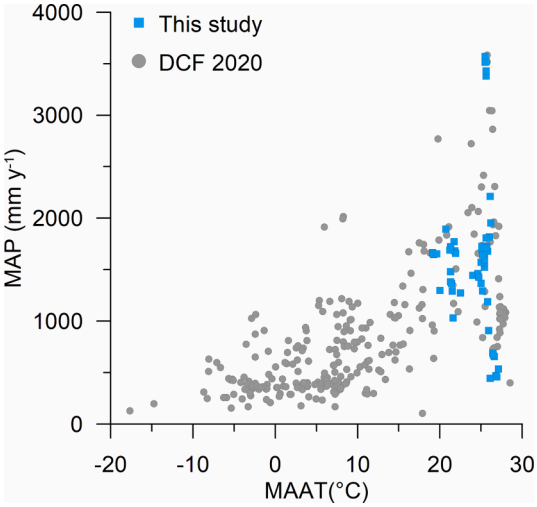

Fig. 2. Climate conditions sampled in the present study (blue squares) compared to the global dataset (grey dots) used in the MBT'_{5Me} calibration to mean annual air temperature (MAAT) in the global compilation by Dearing Crampton-Flood et al. (2020) (DFC 2020), and references therein. The temperatures from the present study are derived from a 30 arc sec climate temperature model by Karger et al. (2017).

2.2. Sample collection

Sampling was conducted during multiple field campaigns in 2016, 2019 and 2020 (Fig. 1). Surface soil samples and soil profiles from the Atlantic rainforest, the Cerrado (savanna) and the southern Amazon rainforest were collected on a north–south transect from São Paulo to the Amazon rainforest in April 2019 (Fig. 1), followed by collection from the Caatinga (shrubland) in northeastern Brazil in March 2020. These were supplemented by soil profiles from close to the boundary between Cerrado and Caatinga vegetation collected in March 2019; samples from the Llanos basin were collected from grasslands, riparian forests, and *Mauritia* swamps sampled in February 2019; and soil profile samples from the interior Amazon rainforest that were collected in September 2016. Sampling sites cover lowland and hill areas with an altitude between 200 and 1200 m.a.s.l. (meters above sea level). In the transect from the Atlantic rainforest across the Cerrado to the Amazon rainforest and the Caatinga transect, we took five surface samples within a quadrant of maximally 15 m × 15 m to analyze local variability. We took samples from the Llanos basin in transects connecting riparian forest and grassland areas. We collected surface samples from the top 5 cm of the soil using a shovel and took soil profile samples from the cleaned surface of soil profiles with a total depth between 55 and 277 cm. Samples collected from the long transects across the Cerrado and Caatinga of Brazil were transported on ice in coolers to avoid microbial activity and were freeze-dried at the University of São Paulo prior to shipping. The other samples were shipped to the University of Southern California and freeze-dried there.

2.3. Laboratory preparation

We homogenized freeze-dried soils with mortar and pestle and removed root fragments manually or using a 1 mm sieve. The homogenized soils were extracted by Accelerated Solvent Extraction (ASE 350) with a 9:1 *v/v* mixture of dichloromethane (DCM):methanol (MeOH) and two 15 min extraction cycles (100 °C, 1500 psi) and a rinse step. After removal of a 20 % split for archiving purposes, the total lipid extract was separated by column chromatography (5 cm × 40 mm Pasteur pipette, NH_2 sepra bulk packing, 60 Å) eluting the neutral fraction with 2:1 DCM:isopropanol (IPA) followed by elution of diethyl ether: formic acid 96:4 (the acid fraction; not used here). The neutral

fraction was further separated by open column chromatography (5% deactivated silica gel, 100–200 mesh) eluting a saturated fraction (not used here) with hexane, and then eluting the polar fraction, containing the GDGTs, with subsequent rinses of pure DCM and MeOH. 50% of the polar fraction (40% of the total extract) was then filtered using a syringe PTFE filter (0.45 μm) using hexane: IPA 99:1 and passively evaporated to dryness prior to HPLC-MS analysis.

2.4. GDGT analyses

For GDGT analysis, the polar fractions were redissolved in 300 or 100 μl Hex:IPA (99:1) and the GDGT distribution was analyzed by high performance liquid chromatography/atmospheric pressure chemical ionization – mass spectrometry (HPLC/APCI-MS) using a ThermoFisher Scientific Accela Quantum Access triplequadrupole MS at the University of Bristol. Injection volume was 15 μl and we used two ultra-high performance liquid chromatography silica columns and normal phase separation, following Hopmans et al. (2016). Analyses were performed in selective ion monitoring mode (SIM) to increase sensitivity and reproducibility and $M+H^+$ (protonated molecular ion) GDGT peaks were manually integrated using the Xcalibur software. The MS continuously scanned for *m/z*'s 1302 (isoGDGT-0), 1300 (isoGDGT-1 and isomers), 1298 (isoGDGT-2 and isomers), 1296 (iso GDGT-3 and isomers), 1294 (GDGT-4 and isomers), 1292 (crenarchaeol (cren.) and its regio isomer (cren.')), 1050 (brGDGT-IIIa and isomers), 1048 (brGDGT-IIIb and isomers), 1046 (brGDGT-IIIc and isomers), 1036 (brGDGT-IIa and isomers), 1034 (brGDGT-IIb and isomers), 1032 (brGDGT-IIc and isomers), 1022 (brGDGT-Ia), 1020 (brGDGT-Ib), 1018 (brGDGT-Ic), 744 (C_{46} standard), and 653 (archaeol).

Analytical stability was monitored by repeated analysis of a peat standard. MBT'$_{5Me}$ of this standard yielded an average value of 0.94 ± 0.004. BIT values yielded an average value of 0.998 ± 0.002.

The MBT'$_{5Me}$ index (De Jonge et al., 2014) reflects the degree of methylation of brGDGTs and is temperature dependent. It represents an adaption from the original MBT(') index (Peterse et al., 2012; Weijers et al., 2007):

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

There are multiple proxies that have been correlated with soil pH. The CBT_{5Me} index (De Jonge et al., 2014) is an adaptation for 5-methyl brGDGTs from the CBT index (Weijers et al., 2007), which reflects changes in the degree of cyclisation of brGDGTs:

$$\mathrm{CBT_{5Me}} = -\log_{10}\left(\frac{(Ib + IIb)}{(Ia + IIa)}\right) \quad (2)$$

The CBT' also includes 6-methyl brGDGTs (De Jonge et al., 2014):

$$\mathrm{CBT'} = \log_{10}\left(\frac{(Ic + IIa' + IIb' + IIc' + IIIa' + IIIb' + IIIc')}{(Ia + IIa + IIIa)}\right) \quad (3)$$

The degree of cyclization (DC) describes the relative contribution of compounds containing cyclopentane moieties (Sinninghe Damsté et al., 2009):

$$\mathrm{DC} = \frac{(Ib + IIb + IIb')}{(Ia + IIa + IIa' + Ib + IIb + IIb')} \quad (4)$$

The isomer ratio of 6-methyl brGDGTs (IR_{6Me}) reflects the relative abundance of 5 and 6-methyl brGDGTs and is also pH dependent in mineral soils (De Jonge et al., 2014) and peats (Naafs et al., 2017b):

$$\mathrm{IR_{6Me}} = \frac{(IIa' + IIb' + IIc' + IIIa' + IIIb' + IIIc')}{(IIa + IIb + IIc + IIa' + IIb' + IIc' + IIIa + IIIb + IIIc + IIIa' + IIIb' + IIIc')} \quad (5)$$

To determine the relative distribution of br- versus the specific isoGDGT crenarchaeol, we used the branched and isoprenoid tetraether index (BIT-index) (Hopmans et al., 2004) with the inclusion of both 5 and 6-methyl brGDGTs:

$$BIT = \frac{(Ia + IIa + IIIa + IIa' + IIIa')}{(Ia + IIa + IIIa + IIa' + IIIa' + cren.)} \quad (6)$$

The ratio of br-GDGT relative to *iso*-GDGT is determined using the $R_{i/b}$ ratio (Xie et al., 2012):

$$R_{i/b} = \frac{\Sigma iso - GDGT}{\Sigma br - GDGT} \quad (7)$$

To be able to compare our results to previously published data, we only included the non-isomer isoGDGTs compounds in the $R_{i/b}$, while we included both 5 and 6-methyl brGDGTs that coeluted in previous methods.

The methane index (MI) measures the relative abundance of crenarchaeol in relation to the other cyclic isoprenoid GDGTs (Zhang et al., 2011):

$$MI = \frac{(GDGT-1 + GDGT-2 + GDGT-3)}{(GDGT-1 + GDGT-2 + GDGT-3 + cren. + cren.')} \quad (8)$$

The $isoGDGT_{IsomerIndex}$ measures the relative abundance of the isomers of the cyclic isoGDGTs 1–3 (Blewett et al., 2020):

$$isoGDGT_{IsomerIndex} = \frac{\Sigma_1^{3'} GDGT}{\Sigma_1^{3'} GDGT + \Sigma_1^{3} GDGT} \quad (9)$$

Given the occasional presence of multiple early eluting isomers, we used the sum of all early eluting isomers to calculate the $isoGDGT_{IsomerIndex}$.

In addition, we define the isoGDGT humidity index:

$$isoGDGT_{HumidityIndex} \frac{(cren.)}{(GDGT-1 + cren.)} \quad (10)$$

2.5. Determination of environmental parameters

Soil pH was determined in the laboratory using Extech pH110 and Extech pH90 pH meters. Measurements were conducted on a 2.5:1 (v:v) water:soil mixture that reflects the same methodology that was used in previous calibrations studies for brGDGTs (Weijers et al., 2007). pH meters were regularly calibrated against buffer solutions (pH of 4 and 7). MAAT and MAP were derived from a global high-resolution 30 arc sec climate model by Karger et al. (2017), equivalent to 1 km^2 resolution at the equator. This resolution is sufficient to constrain temperature and precipitation at our sites between 200 and 1200 m.a.s.l. as we do not include areas of steep relief, where climatic variations finer than the resolution of the model could be a concern. Indeed, the GMTED2010 landscape model (Danielson and Gesch, 2011) underlying the precipitation and temperature models by Karger et al., (2017) has a standard deviation of <50 m for the sample locations. However, we note that local soil conditions depend on many variables and might be slightly different compared to those obtained from the climate model, potentially adding uncertainty. Both tropical and global data sets used for the new calibrations presented in this study include samples from mountainous areas, where the modelled data resolution may lead to mismatch with the sample conditions. Therefore, we adopted the same approach as Pérez-Angel et al. (2020) and use *in situ* measurements wherever available and used model data for the new samples with no other information. Given that the neural network models introduced in Section 2.7. can be re-trained with new data, alternative temperature models can be used according to user preference. We also used the mean temperature of the months above freezing (MAF) for calibration purposes to avoid the complications posed by large temperature seasonalities that involve freezing temperatures. Our new tropical data set does not

include samples from regions with months with temperatures below freezing and the MAAT and MAF values are therefore identical. However, for previously published data, we relied on the data provided in the global data set compiled by Raberg et al. (2022). Raberg et al. (2022) found that studies frequently report MAAT values but MAF data is less often available. Therefore, Raberg et al. (2022) employed model-generated data to fill the missing MAF values. This results in a pattern where some locations have higher MAAT temperatures than MAF temperatures, which is an artefact of the usage of model MAF data alongside analytical MAAT temperatures.

2.6. Statistical analyses

Statistical analyses were performed using the statistical software R (R_Core_Team, 2021). The violin plots shown were created using the R-package vioplot (Adler and Kelly, 2019). We further conducted Monte Carlo modelling to assess to what degree local variability found among the samples collected at the same site can be explained by pure stochastics. To that end we used global br- and isoGDGT data, added random normally distributed variability to each data point and modelled 5 samples per site to reflect our sampling strategy. We varied the σ of the added variability to test the best fit corresponding to our data set. The code used for the modeling can be found in the supplementary information and is available on https://github.com/chaeggi/gdgt and in the supplementary material. The matrixStats package was used for the modelling (Bengtsson, 2018).

2.7. Predicting climate variables using neural networks

We used supervised learning to train neural network models that predict MAAT and MAF from brGDGT distributions. On the one hand, we provide a tropical soil calibration based on the novel soil data presented in this study and tropical soil data curated by Pérez-Angel et al. (2020) calibrated to MAAT. On the other hand, we provide global soil and peat calibrations based on the new data and data compiled by Raberg et al. (2022) calibrated to MAAT and MAF. We used brGDGT data for which a corresponding climatic variable was available (labeled data) to train the models and used them to predict climate from unlabeled brGDGT data.

We rescaled the brGDGT data to a range between 0 and 1 and used them as features in the input layer of the neural network (NN) and trained independent NNs for MAAT and MAF. The NNs were configured with an output layer with one node predicting MAAT (or MAF) and a linear activation function. To train our models, we split the data into a training set (80% of the entries) and a validation set (20%) and performed 5-fold cross-validation, shifting the validation set at each fold. This allowed us to estimate a cross-validation prediction error, which we used to evaluate the performance of our models.

We used a Bayesian NN (BNN) as implemented in the Python module npBNN v.0.1.15 (Silvestro and Andermann, 2020) to obtain posterior estimates of MAAT (or MAF) and a quantification of the prediction error. The BNN used a gaussian likelihood function with a mean defined by the output of the network and a standard deviation set equal to the sample standard error between true and predicted values. We used standard-normal prior distributions N(0, 1) on the weights. The BNNs included two hidden layers of 12 and 4 nodes, respectively, using a tanh activation function. The parameters of the model were sampled from their posterior distribution via Markov Chain Monte Carlo, ran for 25 million iterations, from which we obtained 1,000 posterior samples of the weights. We therefore used the sampled weights combined across the 5 cross-validation sets to obtain posterior estimates of MAAT (MAF) for unlabeled data and we defined the expected prediction error as the standard error averaged across samples.

For comparison, we optimized NN regression models as implemented in Tensorflow v. 2.1 (Abadi et al., 2016) to minimize the mean squared error between predicted and true MAAT (or MAF). Unlike the BNN this model only returned point estimates of the temperature without quantification of uncertainty. We optimized the NN weights through a number of epochs determined by monitoring the validation loss with the patience parameter set to 10. We used three hidden layers 128, 64, and 32 nodes and ReLU activation function.

We evaluated the performance of the different NN and BNN architectures by computing the root mean squared error (RMSE) across the 5 cross-validation sets. For BNN estimates we also evaluated the coverage as the fraction of data for which the true value was included within the 95% credible interval of the predicted one (as determined by the sampled means and standard errors).

We ran npBNN and Tensorflow using Python v. 3.10 (python.org). All codes and scripts implementing NN and BNN analyses is available here: https://github.com/dsilvestro/gdgt-ai. The scripts include detailed model settings and random seeds that ensure the full reproducibility of our results.

We further tested the performance of our new models on published late Pleistocene to Holocene brGDGT records from the Chinese Loess Plateau. Based on their high temporal resolution, we selected the Xifeng record from the northern part of the Chinese Loess Plateau (Lu et al., 2019) as well as the Lantian and Weinan records, from the southern parts (Lu et al., 2016; Tang et al., 2017). Detailed information on the context and location of all studied loess sections can also be found in Lu et al. (2019). We focused our analysis on the last 50 thousand years (ka), where the age control allows for direct comparisons. Given the high temperature seasonality on the Chinese Loess Plateau, we compared the output of our BNN model calibrating the global data set to MAF (BNN MAF_{Global}) to the output of the FROG machine learning approach, using the $FROG_0$ calibration to MAF (Véquaud et al., 2022). We also compared a stepwise regression model ($MAST_{SR}$) based on a local Chinese soil-temperature calibration (Wang et al., 2020). We provide the MBT'_{5ME} index for all records, although we do not convert to absolute temperatures as the numerous available calibrations lead to different absolute temperatures but maintain the same overall trends (De Jonge et al., 2014; Dearing Crampton-Flood et al., 2020; Naafs et al., 2017a).

3. Results

3.1. Relative distribution of brGDGTs in surface soils

Measured soil pH across our tropical soils varied between 3.9 and 7.4. Open grassland soils from the Cerrado and Llanos had low soil pH < 6, while soils under forest and shrubland vegetation had a pH between 3.9 and 7.4. The relationship between soil pH and the pH-dependent indices like CBT', CBT_{5Me}, and IR_{6Me} yielded values in the range of previous global mineral soil calibrations (Fig. 3a-c; Fig. 4a-c) (De Jonge et al., 2014; Dearing Crampton-Flood et al., 2020; Naafs et al., 2017b; Raberg et al., 2022). The relationship between the MBT'_{5Me} proxy and MAAT also falls within that of global soil, peat and lake datasets, except for a few Campo Cerrado and Campo Sujo grassland samples (Figs. 3d, 4d) (De Jonge et al., 2014; Dearing Crampton-Flood et al., 2020; Martínez-Sosa et al., 2021; Naafs et al., 2017a; Raberg et al., 2022). Llanos Basin grasslands similarly have higher MBT'_{5Me} values than nearby forest soils despite constant climate conditions (Fig. 5b). If grassland and forest biomes are combined for the entire data set, it becomes evident that the grassland soils yield overall higher MBT'_{5Me} values compared to forest soils (Fig. 5a). Caatinga vegetation is only found in the warmest and driest parts of the study area. MBT'_{5Me} values for this biome were in the range of forest samples from the Amazon and the Llanos basin (Fig. 4d).

3.2. Relative distribution of isoGDGTs in surface soils

The relative abundance of brGDGT and isoGDGTs shows a distinct relationship with precipitation (Fig. 6). For example, BIT values < 0.7 and $R_{i/b}$ values > 1 are exclusively found in semi-arid areas with MAP <

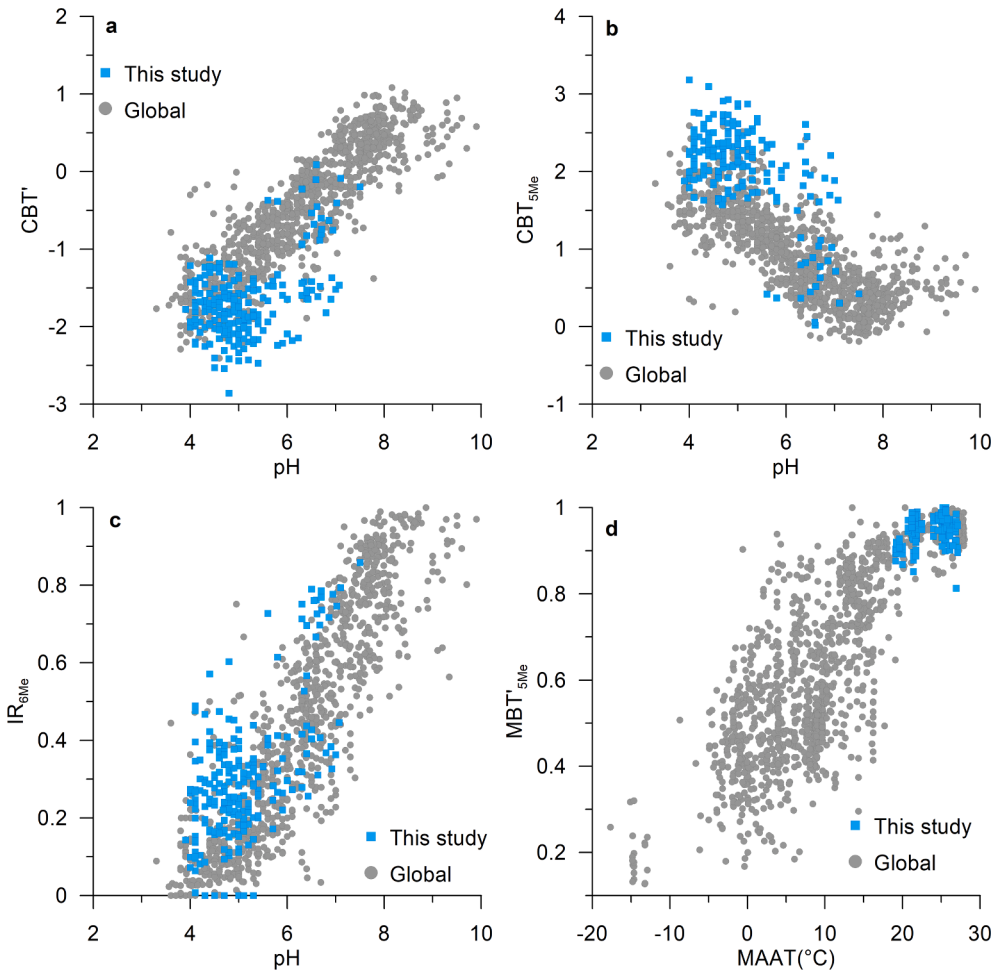

Fig. 3. BrGDGT-based proxies for ph and mean annual air temperature (MAAT) in South American surface soils compared to those from the global dataset compiled by Raberg et al. (2022). a) CBT' pH proxy. b) CBT_{5Me} pH proxy. c) IR_{6Me} pH proxy. d) MBT'_{5Me} temperature proxy. The Raberg et al. (2022) data set consists of data by Cao et al. (2020), Chen et al. (2021), De Jonge et al. (2014), De Jonge et al. (2019), De Jonge et al. (2021), Dearing Crampton-Flood et al. (2020), Ding et al. (2018), Dugerdil et al. (2021), Guo et al. (2020); Guo et al. (2021), Kirkels et al. (2020), Kusch et al. (2019), Li et al. (2017), Li et al. (2018), Miller et al. (2018), Naafs et al. (2017a), Pei et al. (2021), Pérez-Angel et al. (2020), Wang et al. (2018) Wang et al. (2019), Wang et al. (2020), Wang and Liu (2021), Weber et al. (2018), Zang et al. (2018).

750 mm/y. In addition, we also find a precipitation dependence of the relative distribution of isoGDGTs as indicated by the $isoGDGT_{HumidityIndex}$ (Fig. 6c). Here, values close to 1 are found throughout the entire dataset, while lower values < 0.7 are limited to regions with a MAP of >1000 mm y^{-1}. In addition to the response of the indices in (Fig. 6a-c), the relative distribution of isoGDGT-1 further reveals the precipitation dependence of isoGDGT distributions in soils with a relative abundance of isoGDGT-1 in excess of 0.1 only in areas with MAP > 1000 mm/y (Fig. 6d).

Principal component analysis (PCA) of isoGDGT distributions shows that the loadings of isoGDGT-0 and crenarchaeol are in opposite directions for principal component (PC)1 accounting for 78% of the variability, while both isoGDGT-0 and crenarchaeol show opposite loads to isoGDGTs 1–3 in PC2 accounting for 11% of the variability (Fig. 7a). Moreover, the PCA shows that samples from the semi-arid Caatinga shrublands cluster around positive values along PC1, while grass and savanna samples have negative values along PC1. Forest samples show intermediate values. Consistent with this, the ratio of isoGDGT-0 relative to crenarchaeol indicates distinct isoGDGT distributions among different vegetation types (Fig. 8). The $isoGDGT_{IsomerIndex}$ does not show a significant relationship with climate parameters or soil pH (Fig. 7c, d). Rather, it is related to the relative abundance of isoGDGT-1 to 3 indicating that a high $isoGDGT_{IsomerIndex}$ is only possible for samples where isoGDGT-0 and crenarchaeol dominate over isoGDGT 1 to 3 (Fig. 7b).

3.3. Local variability

The site-specific variability based on the analysis of five samples per site reveals that the variability of the climate dependent MBT'_{5Me}, BIT and $R_{i/b}$ proxies increases as the mean site value diverges from the endmember of the proxy (1 in the case of the MBT'_{5Me}, and BIT-index; 0 in case of the $R_{i/b}$; Fig. 9a-c). These results are broadly in line with our Monte Carlo modelled results that illustrate a purely stochastic explanation for the pattern (Fig. 9a-c). For all the climate dependent proxies, Caatinga samples have the largest variability (Fig. 9a-c). For the BIT and $R_{i/b}$, this is in line with the on average lowest BIT and highest $R_{i/b}$ values at these sites. For the MBT'_{5Me}, we also observe a weak, but significant (p-value < 0.05) relationship of variability with precipitation (Fig. 9e).

3.4. Down-profile variability

For the br-GDGT-based proxies, we do not observe systematic down-profile trends for most proxies (Fig. 10a-d). The IR_{6Me}, CBT_{5Me}, and the CBT' show values that fit the global surface soil pH calibrations in deeper soil horizons (Figs. 10a-c; 11a-c). Likewise, the MBT'_{5Me} index

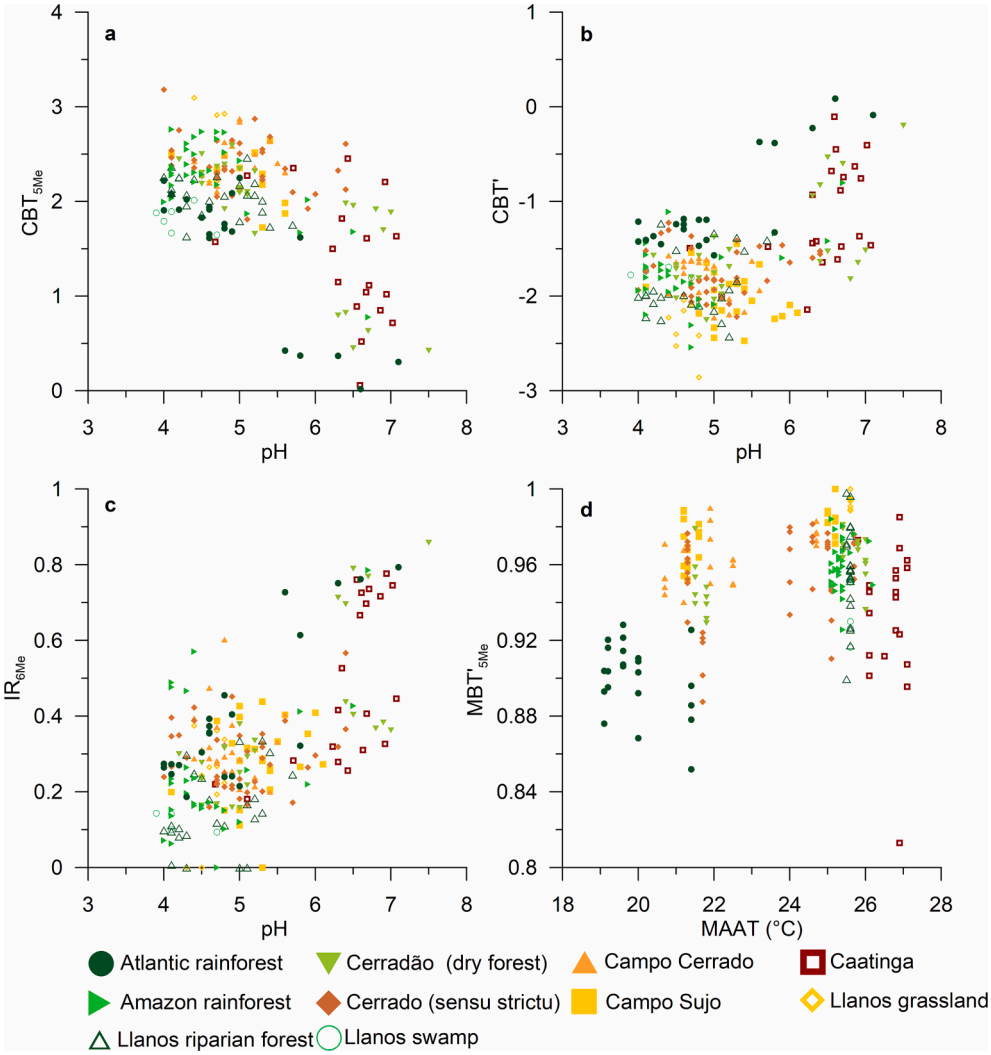

Fig. 4. Influence of vegetation type on brGDGT-based proxies for pH and MAAT in South American surface soils. a) Relationship of the CBT_{5Me} index with pH. b) Relationship of the CBT' index to pH. c) Relationship of the IR_{6Me} with pH. d) Relationship between the MBT'_{5Me} index and mean annual air temperature (MAAT).

shows little down-profile changes and samples from deeper soil horizons are still consistent with those expected from the global surface calibration (Figs. 10d, 11d). For the CBT_{5Me}, the CBT' and the MBT_{5Me} index, savanna samples show greater deviations from the surface than forest samples but values are still within the range of the global calibration (Figs. 10a,b,d; 11a,b,d). The absence of directional changes in the brGDGT-based proxies with depth is contrasted by pronounced directional down-profile shifts that are observed in the relative abundance of brGDGT and isoGDGT as well as in the isoGDGT-based proxies (Fig. 12a-d). For the BIT index, the Caatinga soil profiles that have surface samples with low BIT index indicative of arid conditions, show an increasing down-profile trend towards values of 1, while the profiles from more humid areas are stable around 1 (Fig. 12a). For the $R_{i/b}$, a comparable picture arises for the Caatinga samples, where elevated $R_{i/b}$ values decline in deeper soil profiles (Fig. 12b). In addition, the $R_{i/b}$ also features distinct trends towards increasing values for some of the savanna profiles (Fig. 12b). The $isoGDGT_{IsomerIndex}$, describing the relative abundance of isoGDGT isomers, features a down-profile trend towards 0 for all vegetation types indicating the low abundance of isoGDGT isomers in deeper soil profiles (Fig. 12d). The MI, describing the relative distribution of isoGDGTs, increases with depth towards 1 (Fig. 12c).

3.5. Temperature models using Bayesian neural networks

The tropical BNN and NN temperature models show both high fidelity to MAAT with a cross-validation root mean squared errors (RMSE) of 2.04 and 1.92, respectively (Fig. 13a, Fig. S2a). The coverage for BNN predictions was 0.95, indicating that 5% of the predictions might not include the true value in their 95% credible interval. The global tropical NN and BNN models showed nearly identical results (Fig. S2b). The following reporting of the results, therefore focuses on the BNN side (BNN $MAAT_{Tropical}$) but would also apply to the NN model. The residuals of the novel models show no correlation with precipitation and soil pH (Fig. 13e, Fig. 14b). The highest residuals are still found in samples from open grasslands, while samples from Caatinga and rainforests have lower residuals (Fig. 13e). Samples with a low degree of cyclization and MBT'_{5Me} values close to 1 show BNN $MAAT_{Tropical}$ temperatures of around 25 °C, while values for samples with a high degree of cyclization extend up to 28 °C (Fig. 14c). Likewise, the global BNN models calibrating the global data set to MAF (BNN MAF_{Global}) and MAAT (BNN $MAAT_{Global}$) also show high fidelity with MAF and MAAT with RMSEs of 2.42 and 3.6, respectively (Fig. 13c, Fig. S3a). Both models also show no increased residual for samples from arid areas (Fig. 13f, Fig. S3c).

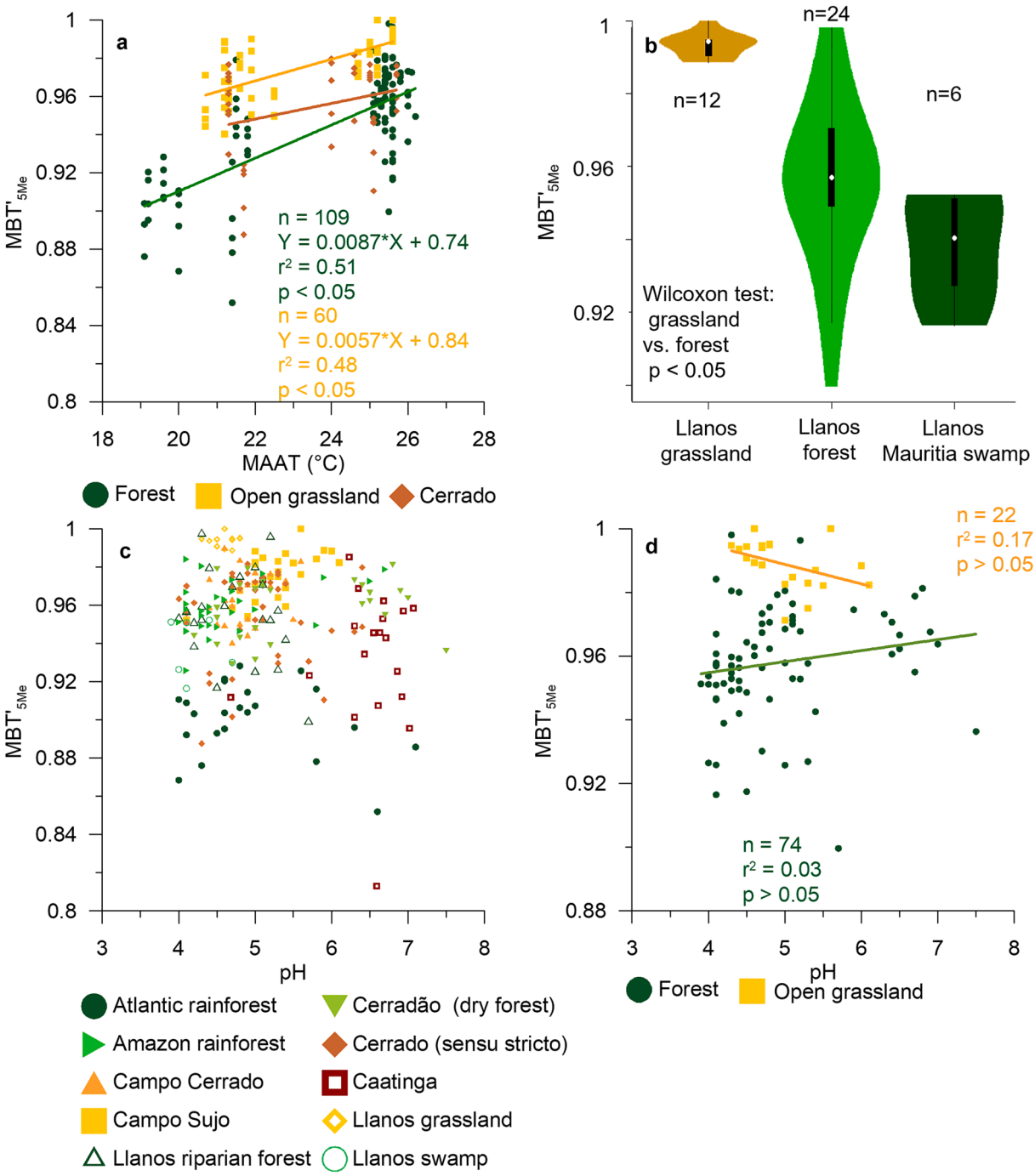

Fig. 5. Impact of vegetation and soil pH on the MBT'$_{5Me}$ temperature proxy in surface soils. a) MBT'$_{5Me}$-MAAT relationship for combined forest (Amazon rainforest, Atlantic rainforest, Cerradão, Llanos riparian forest and swamp), open grassland (Llanos grassland, Campo Sujo and Campo Cerrado) and mixed Cerrado savanna (Cerrado sensu stricto) vegetation types. Caatinga samples are only found at the upper end of the temperature range and are not shown here. b) MBT'$_{5Me}$ distributions under different vegetation in close proximity within the Llanos Basin. The violin plots show data density (kernel density plot) and range (cut off at the data limits). The box-whisker plots indicate interquartile range (IQR) (thick black bar), whiskers with a maximum of 1.5 IQR (thin black bar) and median (white point). c) Relationship of the MBT'$_{5Me}$ to soil pH for the entire data set. d) Relationship of the MBT'$_{5Me}$ to soil pH in forest and open grassland samples with a constant temperature between 25 and 26 °C. The panel illustrates the absence of a pH relationship even when other factors such as MAAT and vegetation are controlled.

3.6. Application to Chinese loess sections

The application of the BNN MAF$_{Global}$ model to the brGDGT records from the Xifeng, Lantian and Weinan loess sections yields consistent temperature trends for all three loess records over the last 50 ka (Fig. 15c; older parts of the records with lower temporal resolution and less precise age control are shown in Fig. S4). All three records show the lowest temperatures during the last glacial maximum (LGM), a temperature maximum during the late deglacial, and mid to late Holocene temperatures higher than during the LGM. Absolute reconstructed temperatures were highest for the Weinan section (LGM ~ 17 °C, late deglacial ~ 25 °C, late Holocene ~ 20 °C) and lower for the Lantian (LGM 10 °C, late deglacial ~ 20 °C, late Holocene ~ 15 °C) and Xifeng sections (LGM ~ 11 °C, late deglacial ~ 17 °C, late Holocene ~ 14 °C).

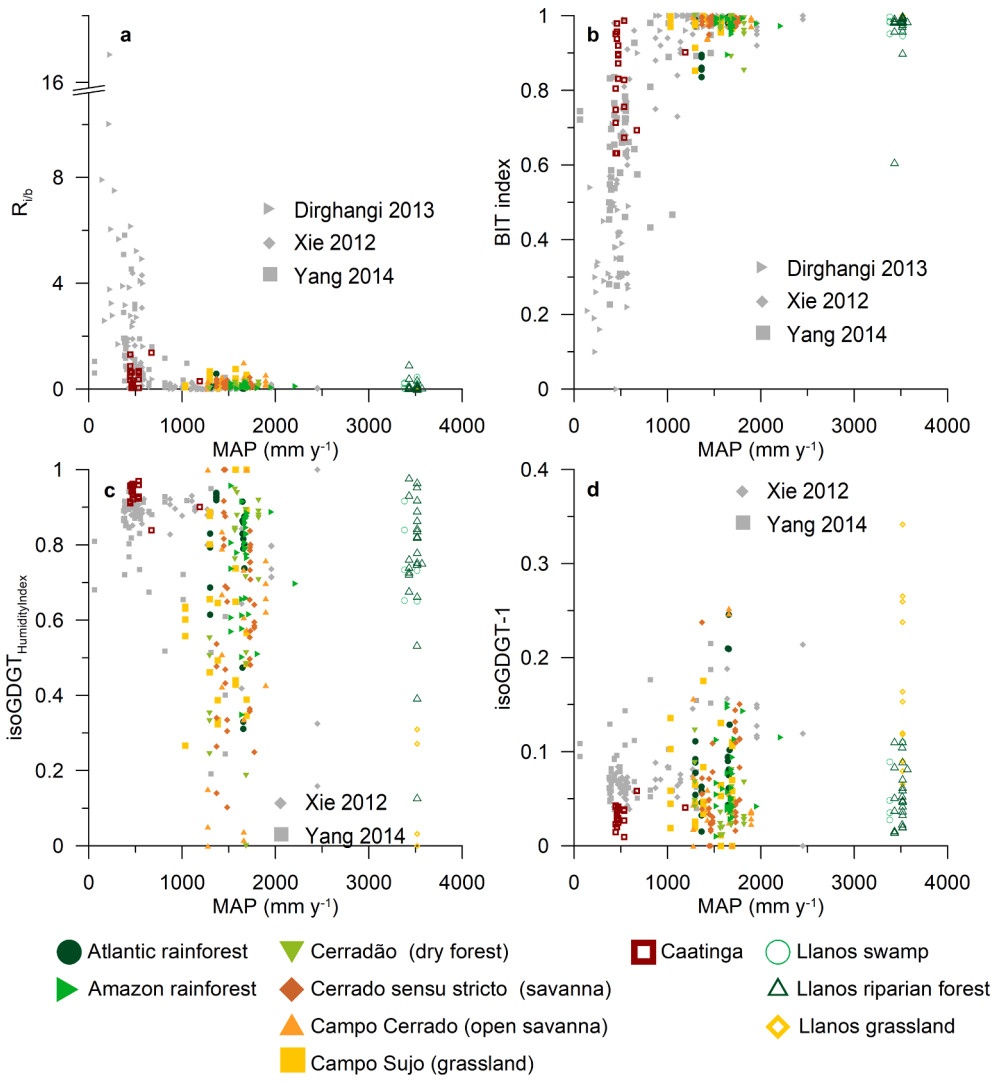

Fig. 6. Relationship of the relative distribution of isoprenoid and branched GDGTs with mean annual precipitation in comparison to previous studies on surface soils from China (Xie et al., 2012; Yang et al., 2014) and the United States (Dirghangi et al., 2013). A salt pond sample was excluded from the Xie et al. (2012) data set. a) $R_{i/b}$ ratio. b) BIT index. c) isoGDGT$_{HumidityIndex}$. Panel d) illustrates the relative abundance of isoGDGT-1 compared to the classical isoGDGTs (isoGDGTs-0 to 3; Cren, Cren').

4. Discussion

4.1. Impact of vegetation type and pH on brGDGT proxies

4.1.1. Impact of vegetation on MBT'$_{5Me}$

The MBT'$_{5Me}$ to temperature relationship observed in the mineral surface soils from tropical South America falls in the range of the global dataset (Fig. 3d). The exception are the Campo Sujo and Campo Cerrado samples at the lower end of the studied temperature range, where MBT'$_{5Me}$ values are higher than found in mineral soils with similar MAAT in previous studies (Figs. 3d, 4d). This deviation is part of a larger trend where savanna soil samples have systematically higher MBT'$_{5Me}$ values compared to forest soils under comparable MAAT-conditions (Fig. 5a, b). Higher MBT'$_{5Me}$ values in savanna soils are both observed for the MAAT- MBT'$_{5Me}$ regressions across the entire dataset as well as in the Llanos data set that was sampled under constant MAAT (Fig. 5a, b). As demonstrated in Fig. 5c-d, this observation is not an artefact of soil pH.

This impact of vegetation type on brGDGT-based temperatures might be attributed to two different factors. First, there is a direct impact of vegetation type on mean annual soil temperatures (MAST). Due to the lack of shading by canopy, savanna and shrubland soils typically have MAST that exceed MAAT by up to several degrees Celsius, while forest soils have MAST that closer match MAAT (Lembrechts et al., 2020; Pérez-Angel et al., 2020; Wang et al., 2020). Second, microbial communities have also been reported to differ between forest and savanna soils in the study area (Vieira et al., 2018), which might further impact MBT'$_{5Me}$ values. However, in the absence of detailed microbial analysis of the studied ecosystems this explanation remains speculative.

The divergence of air and soil temperatures has led to previous attempts of using *in situ* measured soil temperature to calibrate the MBT'$_{5Me}$ temperature proxy in soils. Pérez-Angel et al. (2020) used soil temperatures in a calibration study in the Colombian Andes and Wang et al. (2020) used this approach with Chinese soils. Pérez-Angel et al. (2020) sampled in forested regions and obtained consistent soil and air temperatures. Wang et al. (2020) studied shrublands, grasslands and forests and found that soil temperatures exceeded air temperatures under open vegetation cover, consistent with studies elsewhere (Lembrechts et al., 2020; Li, 1926; Zheng et al., 1993). Wang et al. (2020) did not report a unidirectional change in MBT'$_{5Me}$, but they reported an improvement in the proxy calibration when using MAST rather than MAAT.

We observe an offset of 0.04 in the MBT'$_{5Me}$ values for the Cerrado and Llanos grasslands compared to forest vegetation types, equivalent to

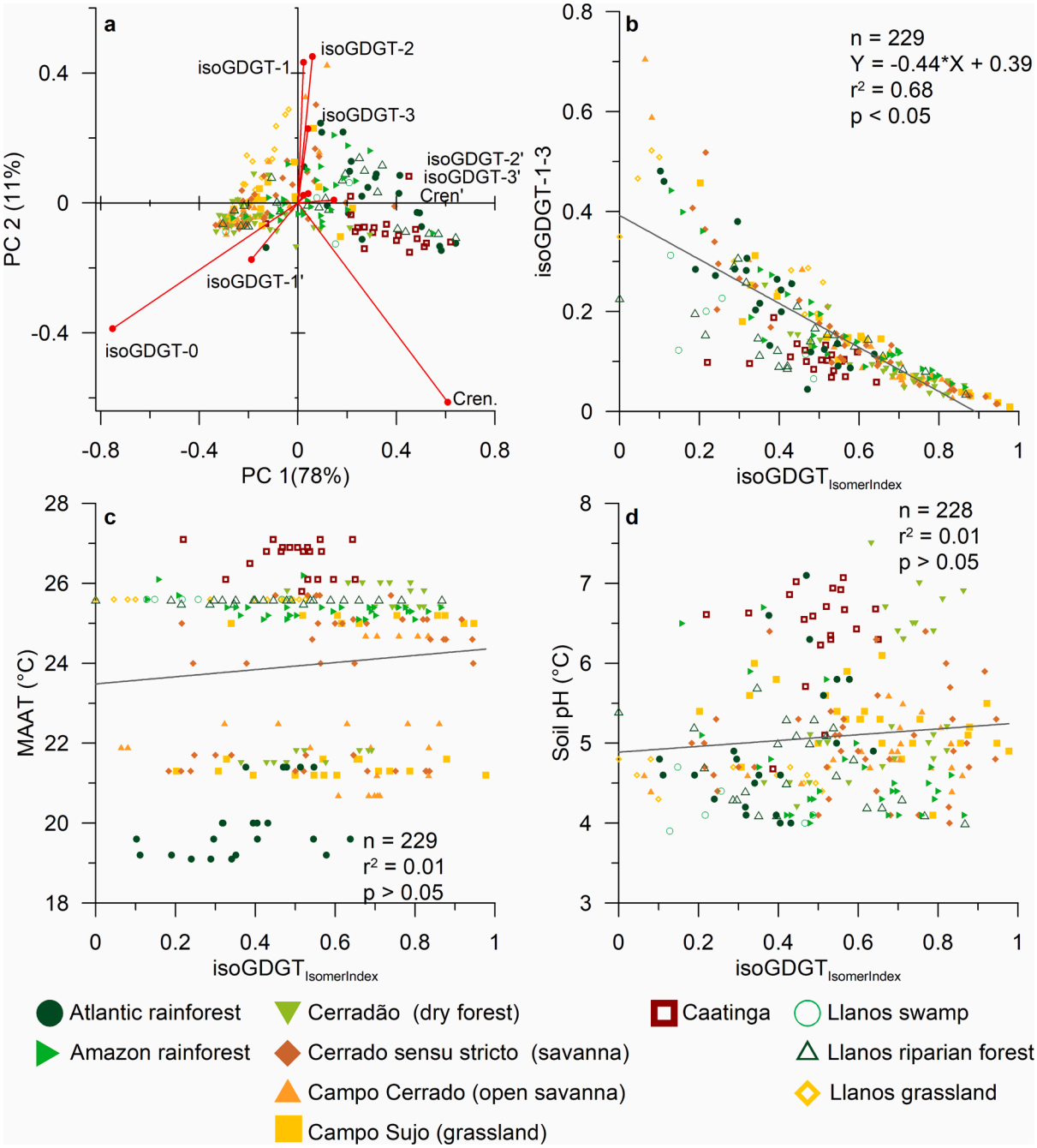

Fig. 7. IsoGDGT patterns in surface soils in tropical South America. a) Principal Component Analysis (PCA) of isoGDGTs. PC1 and PC2 are plotted on eigenvalue scale. Loadings for each compound are indicated by red points. b) Relationship of the $isoGDGT_{IsomerIndex}$ to the relative contribution of isoGDGTs-1 to 3 to the total isoGDGT yield (isoGDGTs-0 to 3; isoGDGT-1′ to 3′; Cren, Cren'). c) Relationship of the $isoGDGT_{IsomerIndex}$ to MAAT. d) Relationship of the $isoGDGT_{IsomerIndex}$ to soil pH.

an offset of 1–2 °C depending on the choice of MAAT calibration (Fig. 5b). Soils are warmer than air (MAST-MAAT) by around 1 °C in forests and 3 °C in grasslands and shrubland, yielding a difference between forests and savannas of around 2 °C, in line with the magnitude of the observed MBT'_{5Me} discrepancy between these vegetation zones (Lembrechts et al., 2020; Wang et al., 2020). Hence, the MAST-MAAT offset largely explains the observed MBT'_{5Me} discrepancy between soils under forest and savanna vegetation in the Llanos and Cerrado, where the large-scale climate is otherwise similar.

Since open shrublands have a MAST-MAAT discrepancy comparable to grasslands of around 3 °C, we would also expect to observe a bias in MBT'_{5Me} values in the semi-arid Caatinga shrublands, (Lembrechts et al., 2020; Wang et al., 2020). However, we find that Caatinga soils have similar MBT'_{5Me} values as forest samples with lower MAAT and have overall the lowest residuals of all biomes (Fig. 4d, Fig. 14a). We note the Caatinga shrublands are drier compared to the other studied biomes. Dry soils in arid climates have previously yielded lower than expected MBT'_{5Me} values (Dang et al., 2016). This has been linked to the distinct microbial communities in dry soils (Guo et al., 2021).

Hence, our findings suggest that vegetation induced MAST-MAAT discrepancies only lead to higher MBT'_{5Me} values in open vegetation biomes found under comparably humid conditions such as in the Llanos basin and the Cerrado. For semi-arid areas such as the Caatinga shrublands this effect is likely overprinted by the effect of aridity or not

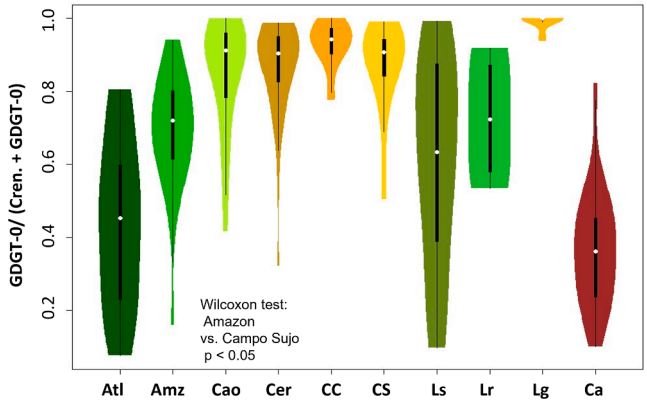

Fig. 8. Relative distribution of isoGDGTs among different vegetation types in tropical South America as indicated by the ratio GDGT-0 / (Crenarchaeol + GDGT-0). Abbreviations are as follows: Atl: Atlantic rain forest. Amz: Amazon rainforest. Cao: Cerradão dry forest. Cer: Cerrado sensu stricto. CC: Campo Cerrado. CS: Campo Sujo. Ls: Llanos swamp. Lr: Llanos riparian forest: Lg: Llanos grassland. Ca: Caatinga shrubland.

present in the first place in denser shrubland types. These results highlight the complex impact of vegetation on the MBT'$_{5Me}$ index.

4.1.2. *Impact of soil pH on MBT'$_{5Me}$*

While the separation of 5′ methylated and 6′ methylated brGDGTs was thought to eliminate the pH-dependent component of the MBT proxy, the impact of pH on the MBT'$_{5Me}$ temperature proxy is still debated (De Jonge et al., 2014; De Jonge et al., 2021; Dearing Crampton-Flood et al., 2020). Residuals of global temperature calibrations are not pH dependent (Dearing Crampton-Flood et al., 2020). However, in a field experiment, De Jonge et al. (2021) found an increase in MBT'$_{5Me}$ for a pH increase from 5 to 6.5, similar to the pH values in our dataset. MBT'$_{5Me}$ values also increased in alkaline soil profiles with pH values > 8, values beyond those found in our sample set (Peaple et al., 2022; Pei et al., 2021). In our sample set, we do not observe a clear influence of soil pH on MBT'$_{5Me}$ values (Fig. 5c, d), finding only a statistically insignificant trend towards lower MBT'$_{5Me}$ at higher pH (Fig. 5d). However, we do observe a negative relationship with pH in the compiled tropical dataset using available data from the literature, which covers a larger (from 3.3 to 9.1) pH range (Fig. 14a). This relationship is inverse to that noted by De Jonge et al., (2021) in their field experiment, yet the effects on GDGT distributions are largely the same: the relative contribution of the dominant brGDGT Ia decreases, while the contribution of 6-methylated and cyclized compounds increase (Fig. 4a-c). This can be explained as follows: low pH tropical brGDGT molecular abundance distributions are comprised almost exclusively of brGDGT Ia, and the addition of a more evenly distributed brGDGT assemblage introduced under high pH conditions inevitably lowers MBT'$_{5Me}$ values. For intermediate MBT'$_{5Me}$ values as observed in the field experiment of De Jonge et al. (2021) the brGDGT distribution is more even than in tropical counterparts. Hence, the addition of relatively evenly distributed brGDGT source can also lead to higher MBT'$_{5Me}$ values. The increase in auxiliary brGDGTs such as 6-methyl compounds might be caused by a change in bacterial communities in response to higher pH values (De Jonge et al., 2021; Guo et al., 2021), but could also be a metabolic response to pH or other covarying variables in soil chemistry or climate. In summary, the impact of soil pH on MBT'$_{5Me}$ is relatively limited for tropical sites.

4.1.3. *brGDGT-based pH proxies*

The brGDGT proxies for soil pH (IR$_{6ME}$, CBT' and the CBT$_{5Me}$) all show relationships that are consistent with previous findings from surface soils (Fig. 3a-c, Fig. 4a-c). Here, the detection of bias introduced by vegetation type is complicated by the observation that only forest vegetation covers most of the pH range observed in the study area (Fig. 4a-c). Semi-arid Caatinga shrubland soils mainly have pH values between 6 and 7, while savanna and grassland vegetation types are limited to pH values < 6 (Fig. 4a-c).

4.2. *Calibration of isoGDGT-based proxies*

4.2.1. *Use of isoGDGTs as precipitation proxy*

The use of isoGDGTs in mineral soils as climatic proxy has so far been largely limited to applications based on the relative distribution of brGDGTs and isoGDGTs as an indicator for aridity (e.g., Xie et al., 2012; Tang et al., 2017). There have been studies showing a temperature dependence of isoGDGT distributions (i.e., TEX$_{86}$) in soil altitude transects analogous to aquatic environments (Coffinet et al., 2014; Liu et al., 2013). These relationships have however been weaker than for brGDGT based temperature proxies and have not been regularly applied for paleoenvironmental studies. The results of our study confirm that increased R$_{i/b}$ values and lower BIT values occur in drier (semi-arid) areas, as reported in previous studies from Asia and the Western United States (Fig. 6a, b) (Dang et al., 2016; Dirghangi et al., 2013; Xie et al., 2012). This could be driven by the expansion of crenarchaeol producing ammonia oxidizing thaumarchaeota in well drained soils. Together these data demonstrate that the relative distribution of brGDGT to isoGDGTs, driven by changes in microbial community structure, can be used as a clear sign of aridity on a global scale (Fig. 6a, b). However, low R$_{i/b}$ values and BIT values close to 1 cannot be interpreted as the absence of arid conditions with a MAP < 500 mm y^{-1}, since these values are also found in semi-arid areas (Fig. 6a, b). This presents a limitation for these proxies (Tang et al., 2017; Xie et al., 2012).

To circumvent this limitation, we further analyzed the relative distribution of isoGDGTs. While the relative contribution of crenarchaeol to the isoGDGT pool is consistently high for semi-arid areas, likely driven by the preference of ammonia oxidizing Thaumarchaeota for (oxic) dry soils (Xie et al., 2012), we found that isoGDGT-1 shows consistently low values in semi-arid areas (Fig. 6d). IsoGDGTs-2 and 3 also show lower relative contributions for arid areas, but the trends are less pronounced than for isoGDGT-1 (Fig. S1a, b). Therefore, we focused on crenarchaeol and isoGDGT-1 to define the isoGDGT$_{HumidityIndex}$ (Eq. (10) and left out isoGDGT-2, isoGDGT-3 and cren' which are used in the MI (Fig. S1a-d). The isoGDGT$_{HumidityIndex}$ has an inverse relationship compared to the BIT index and the R$_{i/b}$. While elevated R$_{i/b}$ values and lowered BIT values are found in exclusively semi-arid samples, the isoGDGT$_{HumidityIndex}$ shows exclusively low values (<0.7) for soils from humid areas, while semi-arid soils all have values close to 1 (Fig. 6c). Thereby, a low (<0.7) isoGDGT$_{HumidityIndex}$ is a clear sign for humid conditions, while the index does not provide unequivocal evidence for aridity. Thus, combined with BIT and R$_{i/b}$ proxies, the isoGDGT$_{HumidityIndex}$ allows for improved precipitation reconstructions.

While the use of the proposed isoGDGT based precipitation proxies is limited to terrestrial archives due to aquatic isoGDGT production (Hopmans et al., 2004), our data compilation with samples from China and North America shows that isoGDGT based precipitation proxies are potentially useful in subtropical and mid-latitude areas around the globe, where the majority of arid and semi-arid regions and suitable archives such as loess can be found (Li et al., 2020).

4.2.2. *isoGDGT isomers in surface soils*

In contrast to previous studies that used peat and culture experiments (Bale et al., 2019; Blewett et al., 2020), our isoGDGT$_{IsomerIndex}$ data from mineral soils does not indicate any relationship between the relative abundance of isoGDGTs isomers and temperature or pH (Fig. 7c, d). Rather our results indicate that the isoGDGT$_{IsomerIndex}$ in mineral soils is shaped by the relative distribution of the classical isoGDGTs. PCA analysis of the isoGDGT distribution in surface soils indicates that the loading of isoGDGT-1′ is similar to the loading of isoGDGT-0, while the other, generally less abundant isoGDGT isomers do not show a

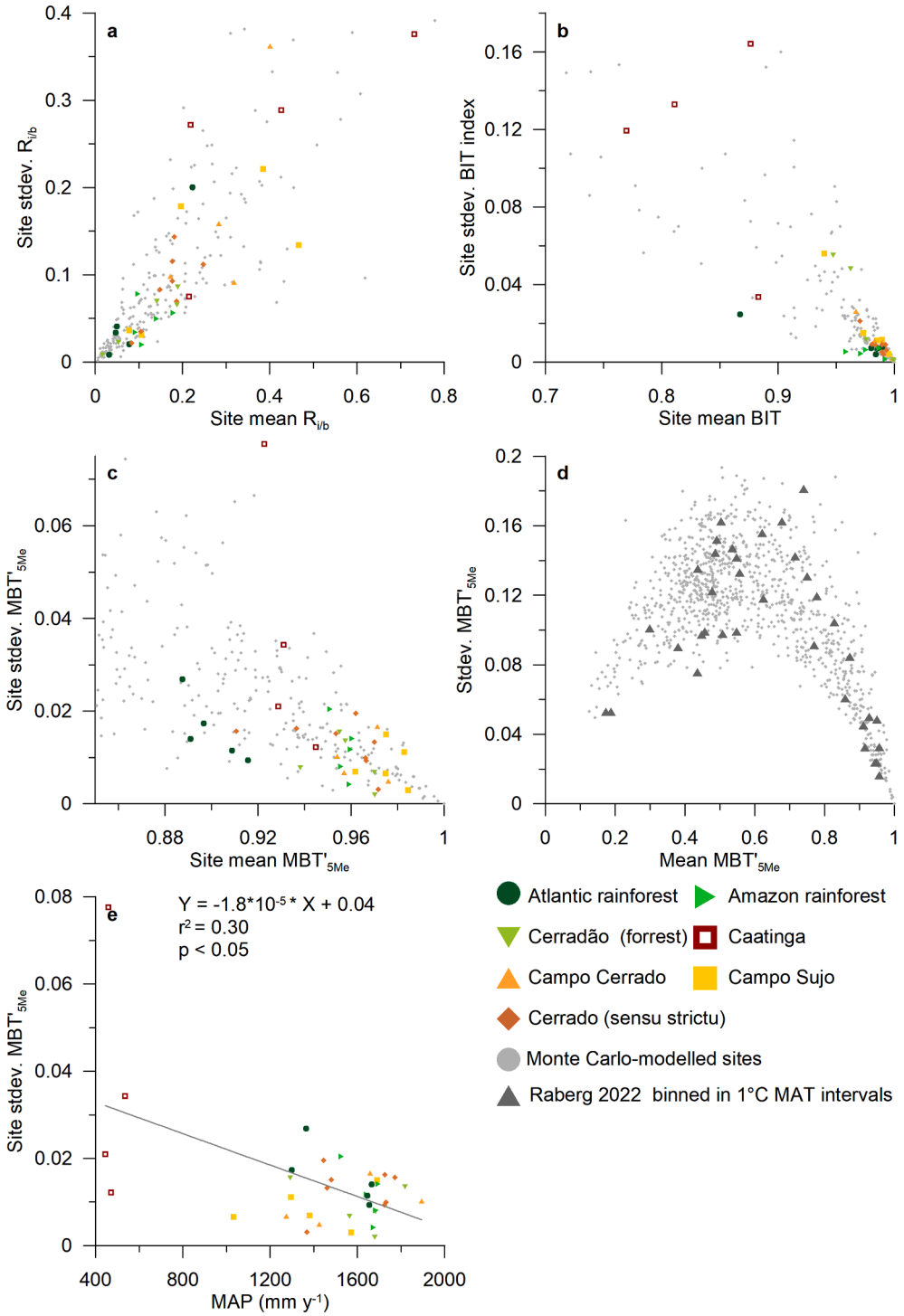

Fig. 9. Site specific variability of the $R_{i/b}$, BIT, MBT'$_{5Me}$ indices. The relationship between the site mean and the standard deviation of all data points from one site for a) $R_{i/b}$. b) BIT-index. c) MBT'$_{5Me}$. d) MBT'$_{5Me}$ values from the global soil data set by Raberg et al. (2022) (black triangles), binned into 1 °C intervals (e.g., 20–21 °C) showing the change in standard deviation across the temperature range. e) Correlation between the mean site standard deviation of the MBT'$_{5Me}$ and precipitation. The gray dots in panels a-d represents Monte Carlo modelled sites to illustrate the impact of variability on the proxies. Each modelled point in a), b) and c) represent the mean of 5 modeled samples (analogous to the 5 measured samples per site). The modelled points in d) represent 33 modelled samples per site to reflect the average bin size. Random variability was added to each datapoint (1σ = 0.5*compound abundance in panels a and b, 0.25*compound abundance in panel c and 0.4*compound abundance in panel d), showing that the proxy formulation amplifies variability for intermediate values and suppresses it close to 1 or 0.

pronounced loading (Fig. 7a). Crenarchaeol (in PC1) and isoGDGT-1 to isoGDGT-3 (in PC2) have loadings in opposite directions compared to isoGDGT-0. This indicates that isoGDGT-1′ is typically more abundant in soils with a higher isoGDGT-0 abundance and lower relative isoGDGT-1 to isoGDGTs-3 contributions (Fig. 7b). Thereby, our data indicates that soils do not reflect the patterns observed in cultures and in peat. This may be caused by the greater and distinct variability of archaeal species present in soils compared to pure cultures and peats (Bates et al., 2011; Ochsenreiter et al., 2003).

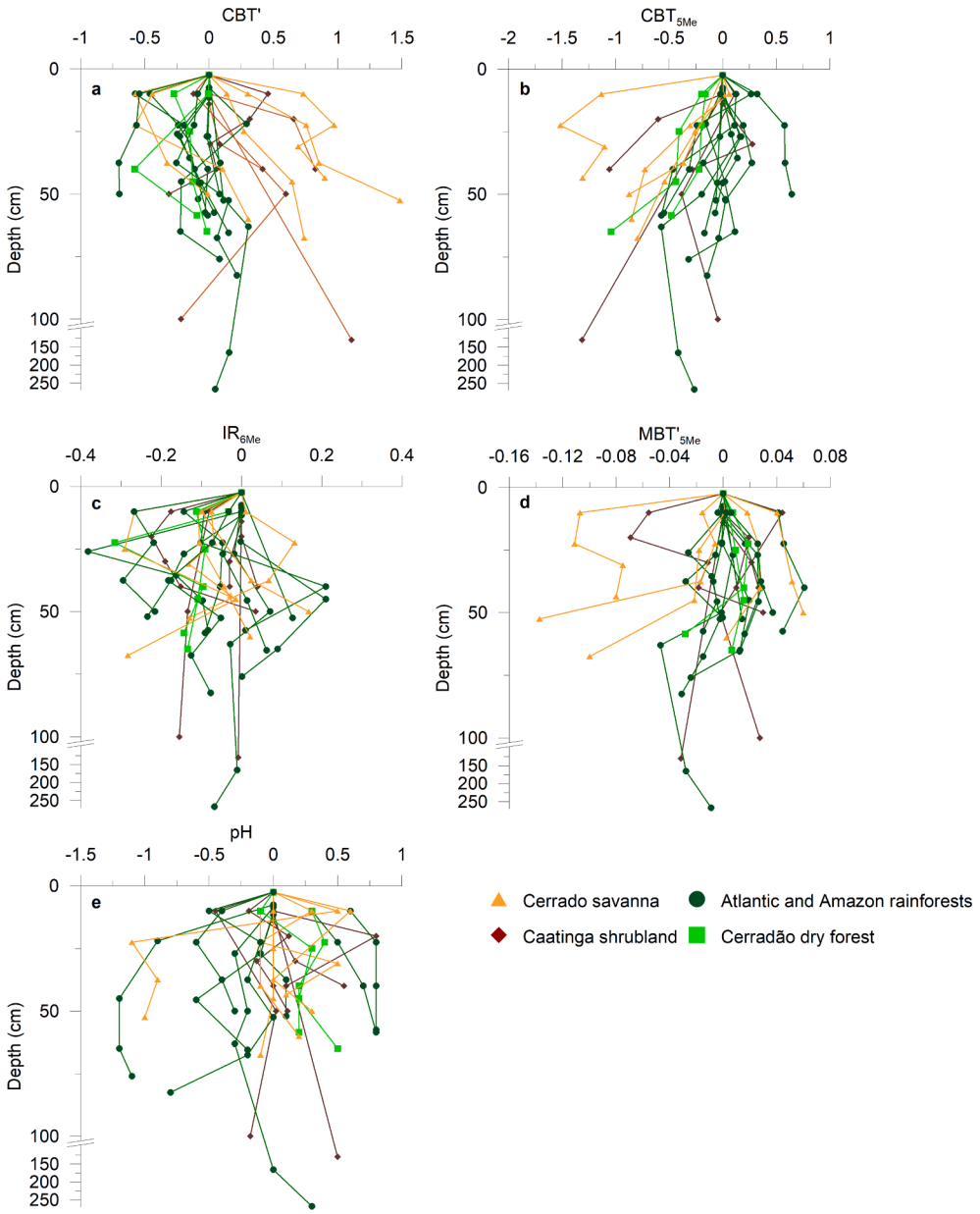

Fig. 10. Variability of brGDGT-based pH and temperature proxies in soil profiles as well as analytical soil pH values normalized to the topmost sample. a) CBT' pH proxy. b) CBT_{5Me} pH proxy. c) IR_{6Me} pH proxy. d) MBT'_{5Me} temperature proxy. e) Measured soil pH.

4.2.3. Vegetation dependence of isoGDGT distributions

While the relative abundance of isoGDGT isomers is not related to climate variables and pH, the PCA of the isoGDGT distributions shows patterns related to vegetation (Fig. 7a). While forest and shrubland samples mostly plot in the positive area of PC1, where crenarchaeol dominates, Cerrado savanna as well as grassland samples and Cerradão dry forest samples are predominantly found in the negative area where isoGDGT-0 is prevalent. Moreover, there is a secondary pattern in PC2, where most Caatinga samples plot in the negative range, while samples from the other vegetation types show greater variability, with open grassland vegetation types showing the most positive values. In particular, the effect of vegetation on isoGDGT distribution is highlighted when focusing on the relative distribution of isoGDGT-0 and crenarchaeol, which are dominant in most samples (Fig. 8). This further illustrates that savanna and grassland vegetation types (Cerrado and Llanos) have a low crenarchaeol contribution and are dominated by isoGDGT-0. This might reflect a difference in archaeal community between these soils with a greater contribution of isoGDGT-0 producing methanogens relative to crenarchaeal producing ammonia oxidizers in savanna and grassland soils. While the clustering of the Caatinga samples can be explained by the semi-arid climate conditions found in the area (e.g., Xie et al., 2012), the variability on the transect between Amazon and Atlantic rainforests is likely caused by vegetation. One mechanism that may explain the divergent patterns in different biomes might be varying nitrogen availability in soils underlying savanna and rainforest vegetation (e.g., Sugihara et al., 2015). Varying nitrogen availability would lead to higher relative abundances of ammonia oxidizing Thaumarchaeota synthesizing crenarchaeol, in rainforest soils with high nitrogen availability, while soils with lower nitrogen availability in savannas would have lower relative crenarchaeol abundances (e.g., Dirghangi et al., 2013; Sinninghe Damsté et al., 2012). Asides from the effects of nitrogen availability, oxygen availability might also play a role. Our results suggest that archaeal membrane lipid distributions in mineral soils can be related to vegetation type, but more work is needed to fully disentangle the factors influencing the distribution of isoGDGTs.

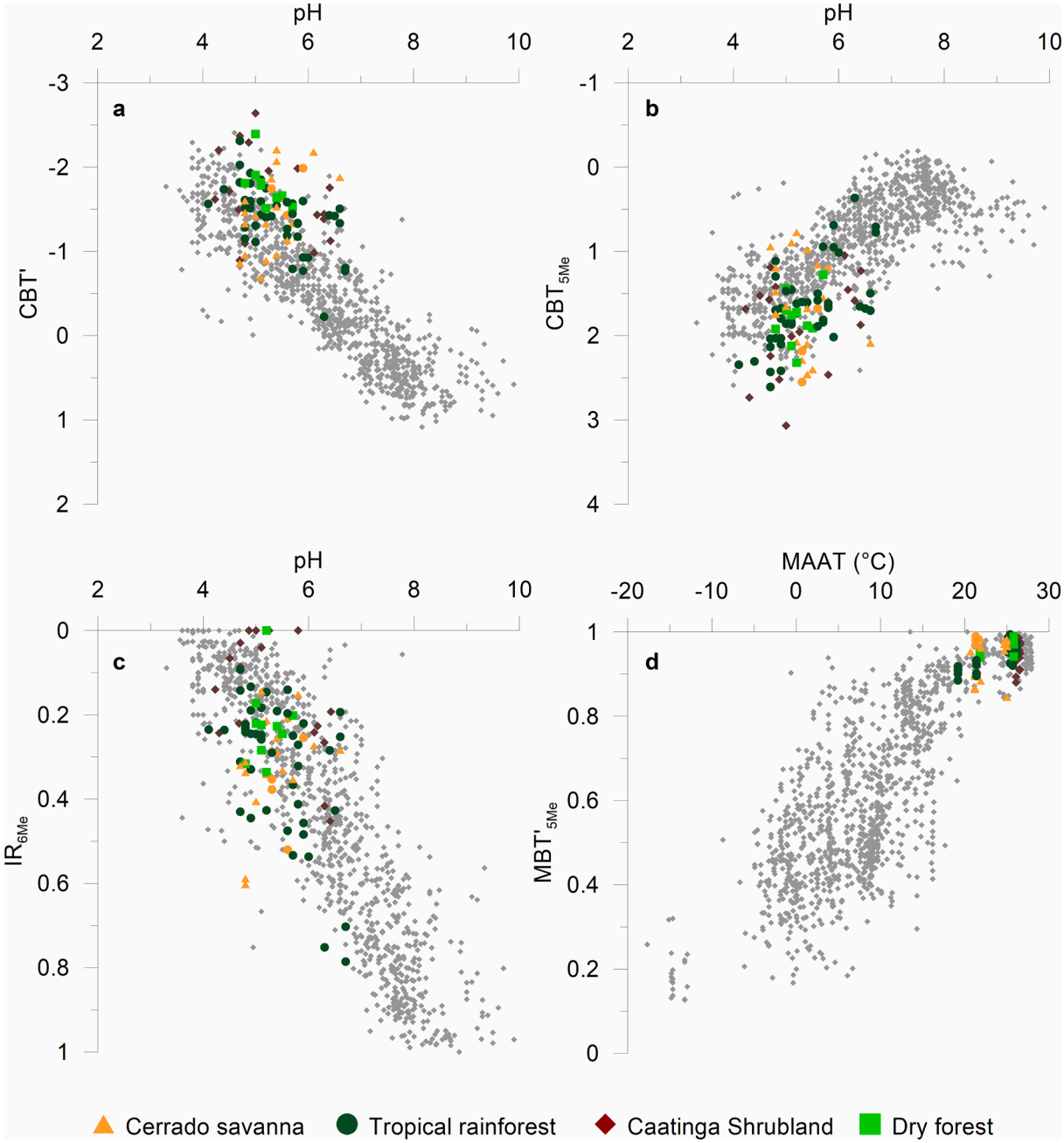

Fig. 11. Down-profile brGDGT-based pH and temperature proxies compared to the global surface soil data set by Raberg et al. (2022) a) CBT' pH proxy. b) CBT_{5Me} pH proxy. c) IR_{6Me} pH proxy. d) MBT'_{5Me} temperature proxy.

4.3. *Site specific proxy variability*

One of the main debates concerning the MBT'_{5Me} temperature proxy are the causes for the variability seen in mineral soils from temperate areas (De Jonge et al., 2021; Dearing Crampton-Flood et al., 2020; Naafs et al., 2017a). There are several factors such as more pronounced seasonality, shifts in microbial communities, vegetation, and soil pH that have been put forward (De Jonge et al., 2019; Wang et al., 2020). Our study demonstrates that vegetation type and soil pH indeed play a role (Fig. 5a, b). In addition, we also tested how consistent values are in local (15 m × 15 m) plots. Our results show that the climate sensitive MBT'_{5Me}, BIT and $R_{i/b}$ proxies all show a lower site-specific variability at sites with mean values closer to the maximal or minimal possible proxy value (i.e., 1 in the case of the MBT'_{5Me} and the BIT index and 0 in case of the $R_{i/b}$; Fig. 9a-c). In addition to this first order relationship, we observe that the semi-arid Caatinga shrubland sites have a higher MBT'_{5Me} variability, which can be explained by the comparably arid conditions in the Caatinga areas (Fig. 9e) (De Jonge et al., 2014). The finding that local MBT'_{5Me} variability is lower close to values of 1 is consistent with MBT'_{5Me} data from the global peat database (Naafs et al., 2017b) that highlights a greater variation in mid- and high-latitude compared to tropical peats.

Mathematically, proxy values for the MBT'_{5Me} index as well as the BIT index are more sensitive to small variations close to 0.5 and less sensitive for values close to the endmembers of 0 and 1. Let us assume for illustrative purposes that the BIT-index is a two-component system formed by crenarchaeol and branched GDGTs. If either component is doubled from a starting BIT value of 0.5, then this will result in a BIT value of 0.33 or 0.66. If either component is doubled from a starting point of 0.9, this will result only in a change to 0.95 or 0.82. Likewise, $R_{i/}$

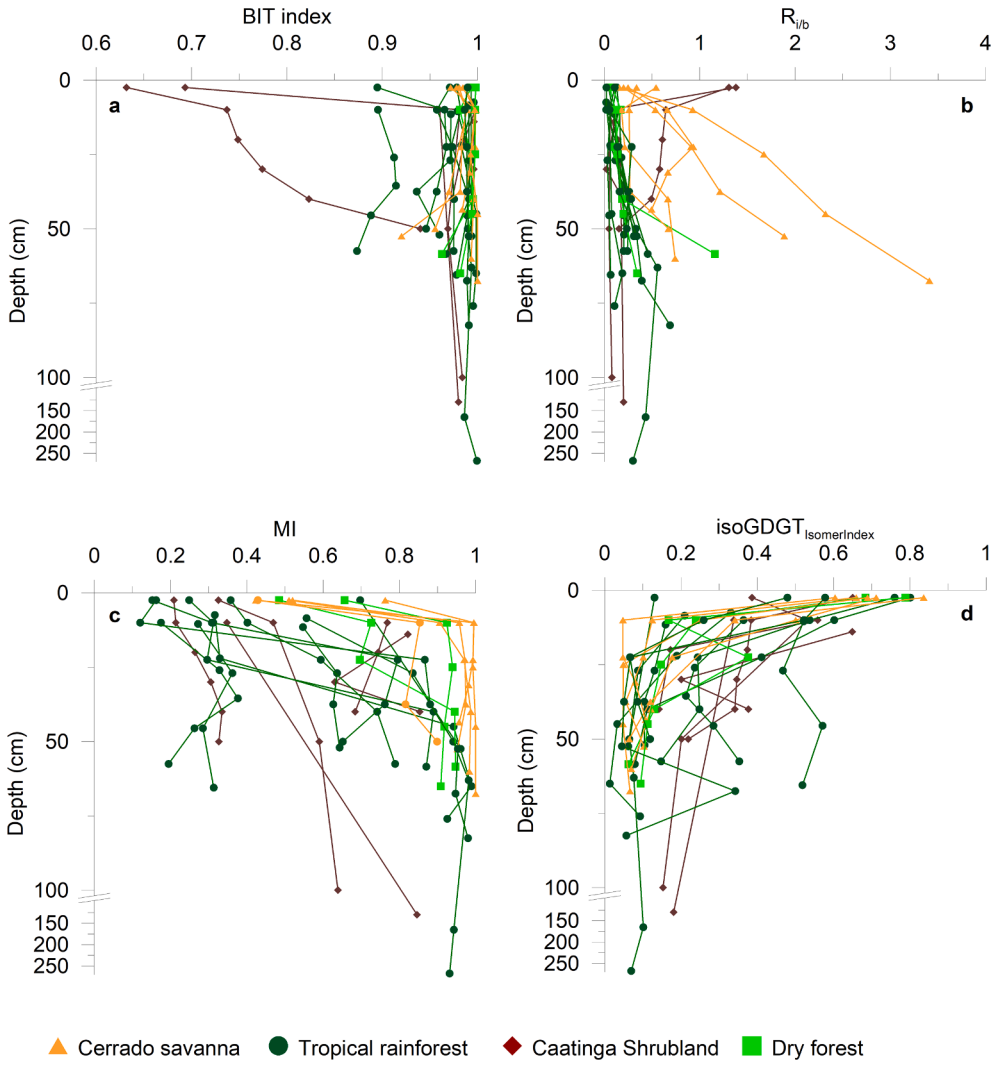

Fig. 12. Change of the relative contribution of branched and isoprenoid GDGTs and shifts in the relative distribution of isoprenoid GDGTs in soil profiles. a) BIT index. b) $R_{i/b}$ index. c) MI, measuring the relative contribution of crenarchaeol. d) $IsoGDGT_{IsomerIndex}$ measuring the relative contribution of isomerized isoGDGTs.

$_b$ values close to 0 are less sensitive to variability than values that are larger.

Using Monte Carlo modelling, we tested what 1σ variability would have to be added to the relative abundances of the constituent compounds of each proxy to produce the observed pattern in a purely stochastic way (Fig. 9). We find that the 1σ variability needed to explain the BIT and $R_{i/b}$ values is both 0.5 (Fig. 9a, b, gray dots), while the variability needed to explain the MBT'_{5Me} index is 0.25 (Fig. 9c).

In addition, to the purely stochastic explanation (i.e., that the constant relative variability added to the constituent compounds mostly explains the pattern), we also find a weak relationship of local MBT'_{5Me} variability with MAP (Fig. 6e). This observation is consistent with those from the global temperature calibration sets, where aridity also led to enhanced proxy variability (Dearing Crampton-Flood et al., 2020). The stochastic explanation for the observed local variability in the BIT, $R_{i/b}$ and MBT'_{5Me} indicates that the local variability in relative compound distribution is more or less constant across large climatic and vegetation gradients, demonstrating that these factors do not play the dominant role in explaining this variability. Our study area covers relatively homogeneous rainforest biomes, but also highly structured savanna mosaics making the constant variability a noteworthy finding. The lack of impact of vegetation structure on the variability might be explained by local homogenization of microbial communities due to dispersal by both macroscopic and microscopic soil fauna (Vieira et al., 2018; Vos et al., 2013).

Finally, we tested the impact of variability on the global dataset by Raberg et al. (2022). For this purpose, we binned the global dataset in 1 °C MAAT brackets of on average 33 samples each and analyzed their variability. The resulting variability also broadly follows the pattern predicted by a purely stochastic variation (Fig. 9d), highlighting the impact of the formulation of the MBT'_{5Me} index on driving the observed large variability in temperate regions. But, as demonstrated in this study, there are also additional factors such as vegetation and soil pH that play a role and can influence the MBT'_{5Me} index (Fig. 5a, b; Fig. 14c).

4.4. GDGT distributions in deeper soil horizons

4.4.1. brGDGT variations in soil profiles

The occurrence and causes of variability of brGDGT proxies in deeper horizons of soil and peat profiles has been inconclusive (Davtian et al., 2016; Huguet et al., 2010; Naafs et al., 2017b; Pei et al., 2021). Most studies have observed some degree of down-profile variability in the MBT'_{5Me} index, but these variations have not been unidirectional (Davtian et al., 2016; Pei et al., 2021). One consistent finding has been that alkaline soil profiles with a pH > 8 featured a trend towards higher MBT'_{5Me} indices with increasing soil depth (Peaple et al., 2022; Pei et al., 2021). In our soil profiles, we do not observe consistent MBT'_{5Me} trends

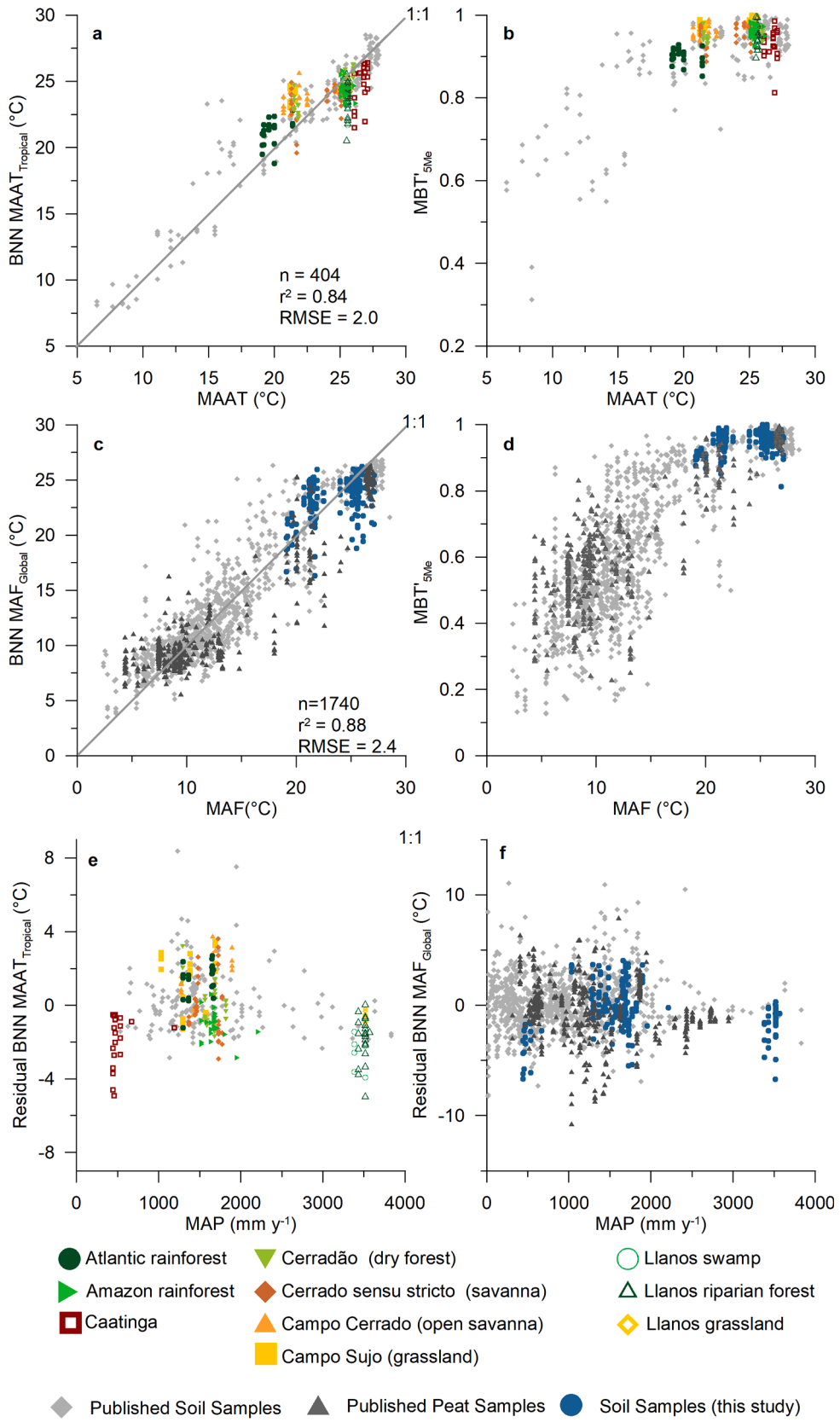

Fig. 13. Bayesian neural network (BNN) based temperature models. a) BNN tropical temperature model (BNN $MAAT_{Tropical}$) calibrated versus MAAT. The sample set includes previously published data from the tropics (23.4 °S – 23.4 °N) (Pérez-Angel et al., 2020). b) MBT'_{5Me} temperature proxy for the data shown in a) for comparison. c) BNN global temperature model calibrated versus MAF. The sample set includes previously published global soil and peat data (Raberg et al., 2022). d) MBT'_{5Me} temperature proxy for the data shown in c) for comparison. e) Residuals of the BNN $MAAT_{Tropical}$ model vs. MAP. f) Residuals of the global MAF model vs. MAP derived from Karger et al. (2017).

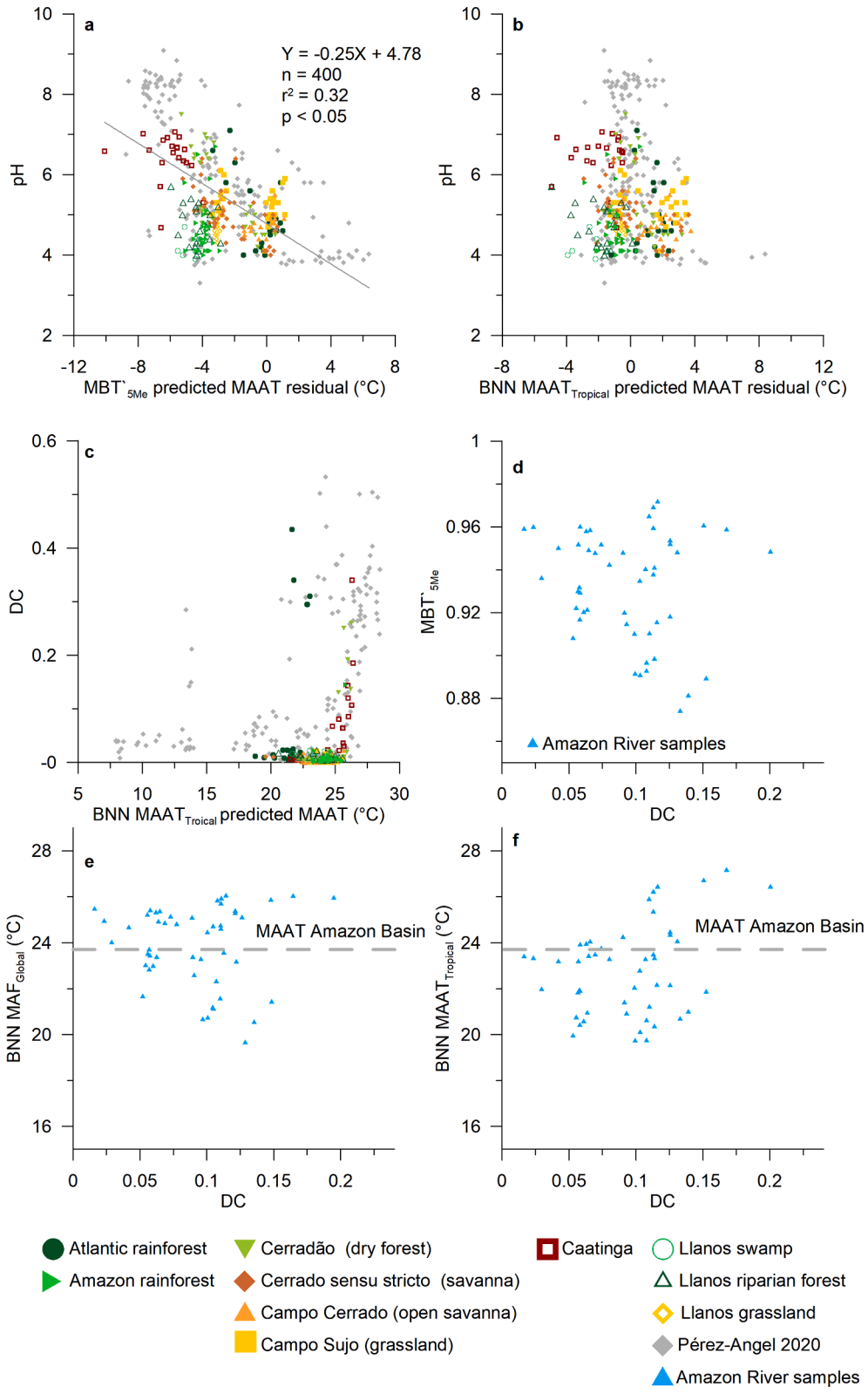

Fig. 14. Removal of pH bias through BNN modeling. a) Residuals of the MBT $_{5Me}$ temperature using the calibration by De Jonge et al. (2014) vs. pH. b) BNN MAAT$_{Tropical}$ predicted MAAT residuals vs. pH. c) BNN predicted MAAT vs. the degree of cyclization (DC). d) Relationship of DC and MBT'$_{5Me}$ in riverbed and suspended sediment samples from the lower Amazon River and its tributaries (Bertassoli et al., 2022). e) Comparison of BNN MAAT$_{Tropical}$ and DC in the same Amazonian samples (Bertassoli et al., 2022). f) Comparison of DC and BNN MAF$_{Global}$ from the Amazonian samples. The dashed lines in e) and f) represent the MAAT in the Amazon River Basin of 23.8 °C (Karger et al., 2017).

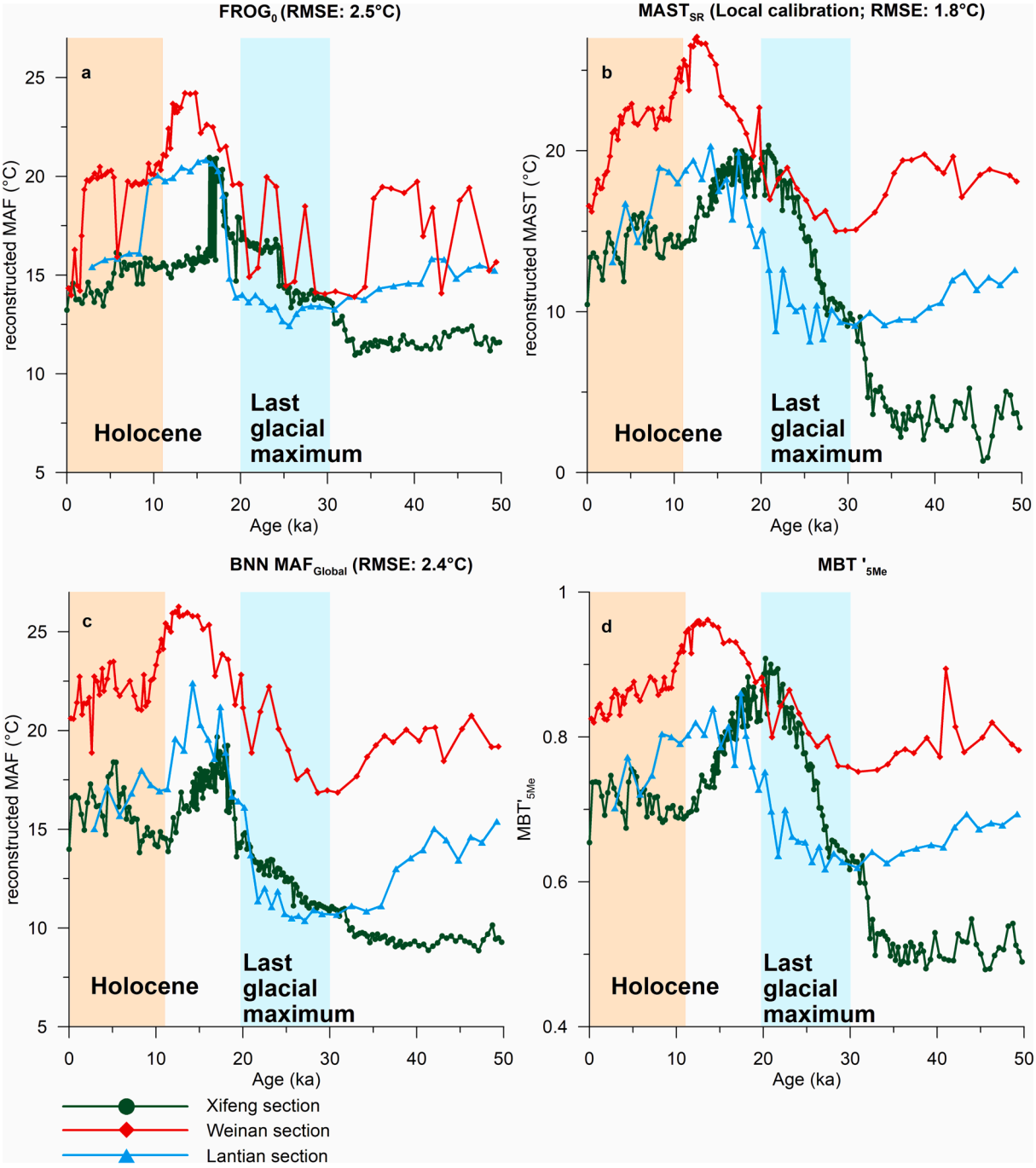

Fig. 15. Comparison of the BNN MAF_{Global} model to other calibrations using published records from the Xifeng (Lu et al., 2019), Lantian (Lu et al., 2016) and Weinan (Tang et al., 2017) loess sections. a) $FROG_0$ (Véquaud et al., 2022), b) $MAST_{SR}$, a local soil temperature calibration from China (Wang et al., 2020), c) BNN MAF_{Global} (this study) d) MBT'_{5Me} (De Jonge et al., 2014).

with depth for all studied vegetation types. Rather, we find that MBT'_{5Me} values are relatively stable within a profile and most values fall within a range of ±0.04 of their corresponding surface sample (Fig. 10d). A few savanna profiles have larger variability (up to 0.12; Fig. 10d). Consequently, we find that MBT'_{5Me} values fall within the range expected from the global calibration set (Fig. 11d). In mid-latitude soils radiocarbon ages of brGDGTs have been found to be at best centuries older than bulk organic carbon (Gies et al., 2021). Given that carbon turnover times in the tropics are faster than in mid latitude soils (Shi et al., 2020), brGDGTs in deeper soil horizons are likely reflecting Holocene temperature conditions (Gies et al., 2021). The consistency of down profile MBT'_{5Me} values is therefore indicating that the same processes controlling surface soils (e.g., homeoviscous adaptation (Naafs et al., 2021)) are also controlling MBT'_{5Me} distributions in subsurface horizons in tropical soil profiles with pH values < 7 (Fig. 10e).

Likewise, we find that the CBT_{5Me}, CBT' and the IR_{6Me} pH proxies do not show consistent unidirectional variations, which is consistent with the relatively stable soil pH (Fig. 10a-c, e). Moreover, their values also fall in the range of the global calibration data set (Fig. 11a-c). The most pronounced shifts can again be observed in the savanna samples. As an explanation for this pattern, we note low brGDGT concentrations in the deeper savanna profiles that could have rendered brGDGT distributions more susceptible to overprint at depth. The low brGDGT yields in deeper horizons also imply that these parts of the soil profiles are not expected

to have a significant contribution compared to the higher concentrations in the surface soils that will therefore dominate the biomarker fluxes in terrestrial erosion and fluvial transport to lacustrine and marine sedimentary archives in tropical South America.

4.4.2. Variations in the relative contribution of isoGDGT and brGDGTs in soil profiles

In contrast to the relatively stable downprofile brGDGT ratios, we observe significant variations in the isoGDGT distribution as well as the relative abundance of brGDGTs and isoGDGTs within soil profiles (Fig. 12). While we observe lowered BIT-values in semi-arid surface soils, down-profile samples show a convergence of BIT values towards 1 (Fig. 12a). This finding is consistent with results from other soil profiles (Peaple et al., 2022; Pei et al., 2021). Lowered BIT values in arid surface soils are likely caused by an enhanced relative contribution of ammonia oxidizing Thaumarchaeota archaea (Dirghangi et al., 2013; Yang et al., 2014). In deeper, less oxygenated soil horizons, this prevalence likely decreases and leads to the observed BIT values close to 1. Likewise, we find that elevated $R_{i/b}$ values in the semi-arid Caatinga shrubland decrease with depth (Fig. 12b), owing to the same explanation as the convergence of BIT values towards 1. Conversely, we find $R_{i/b}$ ratios increase with depth in some savanna profiles, a pattern undetected by BIT which uses crenarchaeol as the only archaeal compound. While we do not report absolute brGDGT concentrations, brGDGT yields in these savanna profile samples were either close to 0 or even absent. Hence, the high $R_{i/b}$ ratios in these soils were likely caused by low brGDGT concentrations rather than a decrease in isoGDGT concentrations. One explanation here could be that the typically less dense rooting of the dominant herbaceous strata in open savannas (Boonman et al., 2020) negatively affects the in-situ production of brGDGTs. Since brGDGTs producing microbes have been suggested to be closely associated to roots (Huguet et al., 2013), this might lead to the observed lower relative contribution of these compounds in deeper savanna soil.

4.4.3. IsoGDGT variations in soil profiles

Our results indicate a strong and directional change with depth in the soil profile isoGDGT distributions. The relative contribution of crenarchaeol as measured by the MI consistently decreases with depth and the $isoGDGT_{IsomerIndex}$ trends toward 0 in deeper soil profiles (Fig. 12c, d). While a decrease in crenarchaeol with depth has been observed before in mineral soils (Yang et al., 2019), the decrease in the $isoGDGT_{IsomerIndex}$ is a novel finding. Studying the $isoGDGT_{IsomerIndex}$ in peat profiles, Blewett et al. (2020) found a significant correlation with soil pH and, relatively stable values down-profile, contrasting with our findings. The gradient in $isoGDGT_{IsomerIndex}$ in our soil profiles suggests different source organisms contributing to the isoGDGT pools or that there is a metabolic response to conditions in deeper soil horizons. As deeper soil layers are less well aerated, their lower oxygen availability and associated redox impacts may drive this change in isoGDGT isomerization.

The down profile patterns with a decrease of the relative contribution of crenarchaeol and isoGDGT isomers and an increase in the relative contribution of isoGDGT- 1 to isoGDGT-3 is in line with the findings from the surface soils, where high $isoGDGT_{IsomerIndex}$ values were also only found in samples that had a low relative contribution of isoGDGT-1-3 (Fig. 7b).

4.4.4. Implications on paleo applications

The lack of significant downprofile changes in the brGDGT-based indices like MBT'_{5Me} demonstrates that brGDGT based proxies can be used to reconstruct environmental conditions over time and that there is limited production of brGDGTs at depth in mineral (tropical) soils.

The distinct down profile trends observed for isoGDGTs as well as the relative distribution of crenarchaeol and brGDGTs (i.e., BIT index) have implications for the applicability of proxies based on these compounds in terrestrial deposits.

The trend for the BIT index towards values close to 1 in deeper soil horizons as well as the similar trend of the MI also converging towards 1 both indicate a lower relative contribution of crenarchaeol in deeper soil horizons (Fig. 12a, c). Most soils analyzed in this study were developed through *in situ* weathering and not through sedimentation and the Caatinga soils developed on late Pleistocene palaeodunes did also not feature recent accumulation (Mescolotti et al., 2023). Therefore, our findings indicate that the low contribution of crenarchaeol in deeper soil horizons is likely the result of low production in these soil horizons, which might have to do with the lower oxygen availability.

This has implications for the application of both the BIT-index as well as the $isoGDGT_{HumidityIndex}$, where crenarchaeol are main constituent components. If crenarchaeol is produced in deeper soil horizons at a low rate, this would indicate that overprint during burial and sediment accumulation is negligible. Indeed, in accumulating terrestrial sediment deposits such as loess, lower BIT and higher $R_{i/b}$ values formed due to aridity in top soils are preserved after deposition and burial and are not overprinted by overlaying sediment layers (Tang et al., 2017; Xie et al., 2012).

4.5. Bayesian neural networks based temperature models

Our findings show that numerous factors such as vegetation, aridity and soil pH can impact brGDGT-based temperature reconstructions. To circumvent these limitations, previous studies have put forward calibrations for specific environments. Liang et al. (2019) for instance suggested vegetation specific calibrations for forests and grassland soils. However, this approach requires control of numerous factors that might not be fully constrainable for samples from sedimentary archives. In order to allow for independent reconstructions without reliance on further parameters, we used a Bayesian neural networks (BNN) approach to obtain novel temperature models. Machine learning approaches have shown to be advantageous for lipid biomarker based paleotemperature reconstructions including in calibrations using brGDGTs (Véquaud et al., 2022). Given that our study focuses on the tropics and more than doubles the GDGT data from tropical mineral soils, we opted to create a tropical temperature model including previously published mineral soil data between 23.3 °N and 23.3 °S (De Jonge et al., 2014; Jaeschke et al., 2018; Kirkels et al., 2020; Pérez-Angel et al., 2020; Wang et al., 2020). Creating a tropical soil model has the advantage that it does not cover areas that are subject to extensive seasonal temperature variability that complicate brGDGT-based MAAT reconstructions. To expand the applicability of the BNN approach to a global scale we further provide a global soil and peat BNN model calibrated against MAAT and MAF. Calibrations against MAF also allow to circumvent the issues regarding temperature seasonality and temperatures below freezing encountered in MAAT calibrations (De Jonge et al., 2014; Dearing Crampton-Flood et al., 2020). We note that further models specifically targeting data from other sources, such as from lakes can be easily trained using the provided code according to user preference.

The results of our BNN models yield strong temperature relationships ($MAAT_{Tropical}$: RMSE = 2.0 °C; coefficient of determination (r^2) = 0.84; MAF_{Global}: RMSE = 2.4; r^2 = 0.88; $MAAT_{Global}$: RMSE = 3.6; r^2 = 0.86; Fig. 13a, c; Fig. S3a). The models do not show a systematic relationship with MAP or pH and also do not feature a pattern that would indicate greater variability in arid areas with a MAP < 500 mm y^{-1} (Fig. 13e, f), a significant advance over previous calibrations.

The $MAAT_{Tropical}$ model also extends the temperature range compared to the MBT'_{5Me} calibration (Fig. 13a, Tab. S1). The upper limit of the MBT'_{5Me} proxy is, by definition, a value of 1, describing the exclusive occurrence of brGDGT Ia-c among the constituent proxy compounds (De Jonge et al., 2014). In the $MAAT_{Tropical}$ model an MBT'_{5Me} value of 1 does not necessarily correspond the highest temperature values. Rather, the compensation of the cold bias found for the MBT'_{5Me} in the tropical high pH samples can lead to higher BNN

$MAAT_{Tropical}$ temperatures (Fig. 14a, c). High pH samples contain compounds with a consistently higher degree of cyclization that also affect the MBT'_{5Me} proxy (Fig. 14c). The neural networks adapt to this pattern and correct for the lowered temperatures that would arise if the MBT'_{5Me} approach was used. Hence, the BNN temperatures reach the highest values for samples with a high pH and the resulting high degree of cyclization (Fig. 14a, b). Due to this adaption, the BNN temperature model removes the pH bias present in the MBT'_{5Me} data set (Fig. 14c, d).

Likewise, the results from the semi-arid Caatinga shrublands illustrate that the addition of 6-methylated and cyclopentane compounds can impact temperature reconstructions in arid areas (Fig. 4d). The observation that all BNN models show no trends towards more pronounced residuals in arid sites with a lower MAP < 500 mm y^{-1} illustrates that the BNN models also compensate for the bias introduced to arid samples that has been observed in previous calibrations (Fig. 13e, f, Fig. S3c) (e.g., Dearing Crampton-Flood et al., 2020).

The addition of a brGDGT component with enhanced contributions of 6-methylated and cyclopentane compounds has also been repeatedly reported from aquatic environments in the tropics such as the Amazon River system (Bertassoli et al., 2022; Zell et al., 2013). Given that the BNN approach successfully corrects for trends associated with increased cyclization, we also tested if the soil BNN models could also be applied to riverine samples from the Amazon Basin (Fig. 14e, f). Indeed, when looking at published samples from the lowland Amazon Basin (Bertassoli et al., 2022), one can observe lower MBT'_{5Me} values in some of the samples with higher DC (Fig. 14d). When applying the BNN temperature models, this is muted (Fig. 14e, f). In the BNN $MAAT_{Tropical}$ there are even signs of a slight overcorrection, as the samples with the highest DC have the overall highest BNN temperatures.

Our results show that the BNN approach is able to yield accurate temperatures in the Amazon River system. This opens the possibility for reconstructions in terrestrial records from the Amazon basin (Baker et al., 2015). There are however limitations of the approach when it comes to marine or lake sourced brGDGT. While the studied riverine samples have brGDGT distributions that are similar to the ones found in soils and have all a BIT value > 0.3 (Bertassoli et al., 2022), marine samples with a dominant aquatic overprint can have brGDGT distributions falling outside the ones found in modern soils and peats covered in the training data sets (Dearing Crampton-Flood et al., 2021; Sinninghe Damsté, 2016). Likewise, lake sediments have also distinct brGDGT distributions from soils, which also precludes the application of our novel soil calibrations to lacustrine records (Russell et al., 2018; Zhao et al., 2023). Indeed, when applying the novel calibrations to a recently published lacustrine surface sediment data set from tropical South America (Zhao et al., 2023), we found a consistent underestimation of reconstructed temperatures (Fig. S5). Hence the novel calibrations should not be applied to paleoclimatic estimations from samples with dominant marine or lacustrine contributions. For lacustrine samples the BNN models can be easily retrained to create lake-specific calibrations. Since marine and lacustrine sourced brGDGT distributions are distinct from soils and peats, a dominant marine or lacustrine contribution can be detected in paleoenvironmental records (Dearing Crampton-Flood et al., 2021; Sinninghe Damsté, 2016) and suitable calibrations applied accordingly.

Pretrained BNN and NN models are available on GitHub (https://github.com/dsilvestro/gdgt-ai) and can be used to predict MAAT and MAF from new user-provided brGDGT data.

4.6. Application to Chinese loess sections

We applied our BNN MAF_{Global} model to the interpretation of GDGT records previously reported from three Chinese loess-paleosol sections spanning the late Pleistocene to the Holocene. This is a rigorous test of the BNN approach as loess is typically a high-pH environment, which can compromise MBT-based thermometry, as demonstrated in this study. The results for the BNN MAF_{Global} show consistent trends for all three loess sections with the Weinan section being warmer than the other sites (Fig. 15c). The temperature minima for the Weinan and Lantian records occur during the last glacial maximum (LGM), whereas Xifeng has minimum temperatures at 40 ka BP, with warming during the LGM. For all sites temperatures peak during the late deglacial. Holocene temperatures are a few degrees lower than the peak, but substantially warmer than the glacial (Fig. 15c).

The early onset of warming at the Xifeng loess section before 30 ka BP is a common feature in all tested calibrations (Fig. 15a-d). Lu et al. (2019) ascribed this pattern to a shift to more open vegetation in the northern part of the Loess Plateau during this time period, an interpretation that is in line with vegetation effects identified in this study. In the BNN MAF_{Global} model we find the lowest temperatures at around 40 ka BP with a warming of 3 °C during the LGM (Fig. 15c). This warming could be consistent with a temperature increase purely caused by vegetation change, as the difference between open and closed forest vegetation observed in this and previous studies is of the same magnitude (Wang et al., 2020). In all other calibrations the LGM temperature rise is more pronounced and the timing for $MAST_{SR}$ and MBT'_{5Me} differs from the machine learning models (Fig. 15a-d). The $MAST_{SR}$ and MBT'_{5Me} reach the highest temperatures of the record towards the end of the LGM, while both machine learning approaches have later temperature maxima during the deglacial, consistent with the other two loess records (Fig. 15a-d).

Another distinction between the different calibrations lies in the degree and timing of the decrease in temperature during the Holocene in the Weinan section. While the BNN MAF_{Global} as well as MBT'_{5Me} show a gradual decrease or stable conditions during the Holocene, the $FROG_0$ and $MAST_{SR}$ show marked decreases in the youngest samples reaching similar-to-LGM temperatures.

Other studies consistently show that temperatures on the Chinese Loess Plateau were considerably lower during the LGM than during modern times: Clumped isotope analysis on snails from the Weinan loess section for instance show that summer temperatures were around 10 °C lower during the LGM than during modern times (Dong et al., 2020). Likewise, groundwater noble gas thermometry from two records to the north of the analyzed loess sections (38°N; analyzed loess sections are between 34 and 36°N) show a glacial cooling of 5–6 °C (Seltzer et al., 2021). Global model-proxy integration also shows by 3.5 °C lower temperatures in the region during the LGM (Osman et al., 2021; Tierney et al., 2020). Hence, the cooling during the Holocene to glacial temperature levels observed in some of the previous calibrations in the Weinan section and the temperature maxima during the LGM observed in the Xifeng section seem unlikely, when compared to other evidence (Dong et al., 2020; Osman et al., 2021).

The comparison to other proxy and model data also indicates that the difference between LGM and modern temperatures of 3 to 5 °C in the BNN MAF_{Global} calibration are within the expected range, further indicating the usefulness of the novel calibration. Given that both the noble gas as well as clumped isotope temperature estimates do not cover the deglacial (Dong et al., 2020; Seltzer et al., 2021), the veracity of the deglacial temperature maximum found for the BNN MAF_{Global} as well as the previous calibrations remains hard to assess.

Both temperature and precipitation at the studied sites on the Chinese Loess Plateau are subject to a high seasonality with temperatures varying up to 20 °C between January and July and precipitation predominantly taking place during summer (Lu et al., 2016; Lu et al., 2019; Tang et al., 2017). This can lead to seasonal bias towards summer temperatures in terms of reconstructed temperatures. For the Xifeng section all calibrations result in higher absolute temperatures for the most recent samples than the modern MAAT of 9.6 °C (Lu et al., 2016; Fig. 15). Furthermore, the deglacial temperature maxima in all calibrations reach temperatures of around 25 °C in the Weinan section, exceeding the modern MAAT of 13.8 °C by more than 10 °C (Tang et al., 2017; Fig. 15). Hence, both previous and our novel BNN calibration likely tend to show a bias towards summer temperatures in the studied

loess sections.

Overall, the BNN MAF_{Global} appears to perform well in generating expected temperature features across the glacial to Holocene transition (Fig. 15). The successful BNN MAF_{Global} application to loess-paleosol sections here demonstrates the broad utility and applicability of the approach. We also wish to emphasize that the BNN model can be retrained and extended to environments such as lakes and to other proxy systems with numerous compounds.

5. Conclusions

We studied brGDGT and isoGDGT distributions in mineral soils from across tropical South America to improve the understanding of their response to environmental variables such as vegetation, soil pH and climate parameters. We found that the MBT'_{5Me} temperature proxy is vegetation dependent, with open grassland vegetation leading to overestimations in reconstructed temperatures, which is likely caused by enhanced soil temperatures due to the sparser vegetation cover. Together with previously published tropical samples, we also find an effect of soil pH on MBT'_{5Me} in tropical soils, with higher soil pH leading to lower MBT'_{5Me} values. We further tested the local variability of these proxies in 15 m × 15 m plots and found that the variability follows a stochastic distribution controlled by the proxy formulation and that there is no enhanced variability in structurally heterogeneous vegetation types of the Brazilian Cerrado. In addition to the previously described relationship of BIT and $R_{i/b}$ values to aridity, we found that the broader distribution of isoGDGTs is precipitation dependent, which can be used to reconstruct precipitation in terrestrial archives like loess. In soil profiles, we found that bacterial brGDGT distributions remained relatively stable with depth, while archaeal isoGDGTs show strong down-profile variations. In particular, the $isoGDGT_{IsomerIndex}$ shows a strong trend towards 0 with increasing soil depth. We attribute this trend to a shift in archaeal communities that might be associated with sub-aerobic conditions in deeper soil horizons. To account for the impact of auxiliary factors such as soil pH on the MBT'_{5Me} proxy and to converge the information available in the brGDGT distribution, we used Bayesian neural networks to create improved tropical and global temperature calibrations. Our improved models remove the impact of soil pH and aridity and also provide accurate temperatures in tropical river samples. The models are also tested on timeseries from loess, generating temperature interpretations that appear to resolve realistic glacial to Holocene temperature transitions in three locations, resolving unexpected features when applying most other calibrations. The novel BNN methods have immediate application to reconstruct temperature and precipitation in terrestrial deposits, and the BNN model can be trained on new datasets and applied to new proxies and settings.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was funded by a Swiss National Science Foundation (SNF) mobility fellowship (grant P400P2_183856) to CH. We acknowledge undergraduate laboratory assistants at USC: Betelhem Assefa, Jonnie Dolan, Lindsay Luchinsky, Dea Kurti, Christopher Rincon and Sharon Tu. Soils were imported under USDA Permit P330-19-00164. We thank NERC for partial funding of the National Environmental Isotope Facility (NEIF; contract no. NE/V003917/1) and associated HPLC-MS capabilities at the University of Bristol. B.D.A.N. acknowledges a Royal Society Tata University Research Fellowship for funding. AOS is supported by CNPq (grant 307179/2021-4) and FAPESP (grant 2018/23899-2). Sampling of soils in eastern Amazon was funded by FAPESP (grant 2016/02656-9). DJB was financially supported by the São Paulo Research Foundation (FAPESP) (grants #2019/24977-0 and #2022/06440-1). TKA acknowledges the financial support from FAPESP (grants 2019/19948-0 and 2021/13129-8). D.S. received funding from the Swiss National Science Foundation (PCEFP3_187012) and from the Swedish Research Council (VR: 2019-04739), and the Swedish Foundation for Strategic Environmental Research MISTRA within the framework of the research programme BIOPATH (F 2022/1448). CMC acknowledges the financial support from FAPESP (grants 2018/15123-4 and 2019/24349-9), CNPq (grant 312458/2020-7) and the 2019-2020 BiodivERsA joint call for research proposals, under the BiodivClim ERA-Net COFUND programme. We thank Hongxuan Lu, Weiguo Liu, Hong Yang for making the data from the Xifeng and Lantian loess sections available and Changyan Tang and Shucheng Xie for providing the data for the Weinan loess section. We thank four anonymous reviewers for their helpful comments.

Research Data

Research data, including the brGDGT, isoGDGT abundances, derived indices and allied metadata corresponding to the soil samples can be accessed in the Supplementary Material and at www.pangaea.de.

BNN models presented in this study are available as Python code on the GITHUB following repository https://github.com/dsilvestro/gdgt-ai.

Appendix A. Supplementary material

The Supplementary Materials contains Supplementary Figure 1 illustrating the precipitation dependence of isoGDGT-2 and iso-GDGT-3 as well as the relationship between the Methane Index and the $isoGDGT_{HumidityIndex}$; Supplementary Figure 2 showing the relationship of the NN and BNN tropical temperature models; Supplementary Figure 3 illustrating the performance of the BNN $MAAT_{Global}$ model; Supplementary Figure 4 showing the full loess records shown in Figure 15; Supplementary Figure 5 showing the application of the BNN models to South American lake sediments; The R code for the Monte Carlo modelling presented in Figure 9; The research data. The file contains three sheets: One for the surface data, one for the soil-profile data and one for the loess data featured in Figure 15 and Figure S4. Supplementary material to this article can be found online at https://doi.org/10.1016/j.gca.2023.09.014.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I.J., Harp, A., Irving, G., Isard, M., Jia, Y., Józefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D.G., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P.A., Vanhoucke, V., Vasudevan, V., Viégas, F.B., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv: 1603.04467v2.

Adler, D., Kelly, S.T., 2019. vioplot: violin plot. R package version 0.3.4 https://github.com/TomKellyGenetics/vioplot.

Baker, P.A., Fritz, S.C., Silva, C.G., Rigsby, C.A., Absy, M.L., Almeida, R.P., Caputo, M., Chiessi, C.M., Cruz, F.W., Dick, C.W., Feakins, S.J., Figueiredo, J., Freeman, K.H., Hoorn, C., Jaramillo, C., Kern, A.K., Latrubesse, E.M., Ledru, M.P., Marzoli, A., Myrbo, A., Noren, A., Piller, W.E., Ramos, M.I.F., Ribas, C.C., Trnadade, R., West, A. J., Wahnfried, I., Willard, D.A., 2015. Trans-Amazon Drilling Project (TADP): origins and evolution of the forests, climate, and hydrology of the South American tropics. Sci. Dril. 20, 41–49.

Bale, N.J., Palatinszky, M., Rijpstra, W.I.C., Herbold, C.W., Wagner, M., Sinninghe Damsté, J.S., 2019. Membrane lipid composition of the moderately thermophilic ammonia-oxidizing archaeon "candidatus nitrosotenuis uzonensis" at different growth temperatures. Appl. Environ. Microbiol. 85, e01332–01319.

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., Fierer, N., 2011. Examining the global distribution of dominant archaeal populations in soil. ISME J. 5, 908–917.

Becker, K.W., Lipp, J.S., Zhu, C., Liu, X.-L., Hinrichs, K.-U., 2013. An improved method for the analysis of archaeal and bacterial ether core lipids. Org Geochem. 61, 34–44.

Bendle, J.A., Weijers, J.W.H., Maslin, M.A., Sinninghe Damsté, J.S., Schouten, S., Hopmans, E.C., Boot, C.S., Pancost, R.D., 2010. Major changes in glacial and Holocene terrestrial temperatures and sources of organic carbon recorded in the Amazon fan by tetraether lipids. Geochem. Geophys. Geosyst. 11.

Bengtsson, H., 2018. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.54.0. https://CRAN.R-project.org/package=matrixStats.

Bertassoli, D.J., Häggi, C., Chiessi, C.M., Schefuß, E., Hefter, J., Akabane, T.K., Sawakuchi, A.O., 2022. Controls on the distributions of GDGTs and n-alkane isotopic compositions in sediments of the Amazon River Basin. Chem. Geol. 594, 120777.

Blewett, J., Naafs, B.D.A., Gallego-Sala, A.V., Pancost, R.D., 2020. Effects of temperature and pH on archaeal membrane lipid distributions in freshwater wetlands. Org Geochem. 148, 104080.

Blydenstein, J., 1967. Tropical Savanna Vegetation of the Llanos of Colombia. Ecology 48, 1–15.

Blyth, A.J., Schouten, S., 2013. Calibrating the glycerol dialkyl glycerol tetraether temperature signal in speleothems. Geochim. Cosmochim. Acta 109, 312–328.

Boonman, C.C.F., van Langevelde, F., Oliveras, I., Couédon, J., Luijken, N., Martini, D., Veenendaal, E.M., 2020. On the importance of root traits in seedlings of tropical tree species. New Phytol. 227, 156–167.

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

Chen, C., Bai, Y., Fang, X., Zhuang, G., Khodzhiev, A., Bai, X., Murodov, A., 2021. Evaluating the potential of soil bacterial tetraether proxies in westerlies dominating western Pamirs, Tajikistan and implications for paleoenvironmental reconstructions. Chem. Geol. 559, 119908.

Chen, Y., Zheng, F., Yang, H., Yang, W., Wu, R., Liu, X., Liang, H., Chen, H., Pei, H., Zhang, C., Pancost, R.D., Zeng, Z., 2022. The production of diverse brGDGTs by an Acidobacterium providing a physiological basis for paleoclimate proxies. Geochim. Cosmochim. Acta 337, 155–165.

Coffinet, S., Huguet, A., Williamson, D., Fosse, C., Derenne, S., 2014. Potential of GDGTs as a temperature proxy along an altitudinal transect at Mount Rungwe (Tanzania). Org Geochem. 68, 82–89.

Dang, X., Yang, H., Naafs, B.D.A., Pancost, R.D., Xie, S., 2016. 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.

Danielson, J.J., Gesch, D.B., 2011, Global multi-resolution terrain elevation data 2010 (GMTED2010), U.S. Geological Survey Open-File Report 2011-1073, 26p.

Davtian, N., Ménot, G., Bard, E., Poulenard, J., Podwojewski, P., 2016. Consideration of soil types for the calibration of molecular proxies for soil pH and temperature using global soil datasets and Vietnamese soil profiles. Org Geochem. 101, 140–153.

De Jonge, C., Hopmans, E.C., Stadnitskaia, A., Rijpstra, W.I.C., Hofland, R., Tegelaar, E., Sinninghe Damsté, J.S., 2013. Identification of novel penta- and hexamethylated branched glycerol dialkyl glycerol tetraethers in peat using HPLC–MS2, GC–MS and GC–SMB-MS. Org Geochem. 54, 78–82.

De Jonge, C., Hopmans, E.C., Zell, C.I., Kim, J.-H., Schouten, S., Sinninghe Damsté, J.S., 2014. Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction. Geochim. Cosmochim. Acta 141, 97–112.

De Jonge, C., Radujković, D., Sigurdsson, B.D., Weedon, J.T., Janssens, I., Peterse, F., 2019. Lipid biomarker temperature proxy responds to abrupt shift in the bacterial community composition in geothermally heated soils. Org Geochem. 137, 103897.

De Jonge, C., Kuramae, E.E., Radujković, D., Weedon, J.T., Janssens, I.A., Peterse, F., 2021. The influence of soil chemistry on branched tetraether lipids in mid- and high latitude soils: implications for brGDGT- based paleothermometry. Geochim. Cosmochim. Acta.

Dearing Crampton-Flood, E., Tierney, J.E., Peterse, F., Kirkels, F.M.S.A., Sinninghe Damsté, J.S., 2020. BayMBT: A Bayesian calibration model for branched glycerol dialkyl glycerol tetraethers in soils and peats. Geochim. Cosmochim. Acta 268, 142–159.

Dearing Crampton-Flood, E., van der Weijst, C.M.H., van der Molen, G., Bouquet, M., Yedema, Y., Donders, T.H., Sangiorgi, F., Sluijs, A., Sinninghe Damsté, J.S., Peterse, F., 2021. Identifying marine and freshwater overprints on soil-derived branched GDGT temperature signals in Pliocene Mississippi and Amazon River fan sediments. Org Geochem. 154, 104200.

Ding, S., Kohlhepp, B., Trumbore, S., Küsel, K., Totsche, K.-U., Pohnert, G., Gleixner, G., Schwab, V.F., 2018. In situ production of core and intact bacterial and archaeal tetraether lipids in groundwater. Org Geochem. 126, 1–12.

Dirghangi, S.S., Pagani, M., Hren, M.T., Tipple, B.J., 2013. Distribution of glycerol dialkyl glycerol tetraethers in soils from two environmental transects in the USA. Org Geochem. 59, 49–60.

Dong, J., Eiler, J., An, Z., Wu, N., Liu, W., Li, X., Kitchen, N., Lu, F., 2020. Clumped and stable isotopes of land snail shells on the Chinese Loess Plateau and their climatic implications. Chem. Geol. 533, 119414.

Dugerdil, L., Ménot, G., Peyron, O., Jouffroy-Bapicot, I., Ansanay-Alex, S., Antheaume, I., Behling, H., Boldgiv, B., Develle, A.-L., Grossi, V., Magail, J., Makou, M., Robles, M., Unkelbach, J., Vannière, B., Joannin, S., 2021. Late Holocene Mongolian climate and environment reconstructions from brGDGTs, NPPs and pollen transfer functions for Lake Ayrag: Paleoclimate implications for Arid Central Asia. Quat. Sci. Rev. 273, 107235.

Gies, H., Hagedorn, F., Lupker, M., Montluçon, D., Haghipour, N., van der Voort, T.S., Eglinton, T.I., 2021. Millennial-age glycerol dialkyl glycerol tetraethers (GDGTs) in forested mineral soils: 14C-based evidence for stabilization of microbial necromass. Biogeosciences 18, 189–205.

Goodland, R., 1971. A physiognomic analysis of the 'Cerrado' vegetation of central Brasil. J. Ecol. 59, 411–419.

Guo, J., Glendell, M., Meersmans, J., Kirkels, F., Middelburg, J.J., Peterse, F., 2020. Assessing branched tetraether lipids as tracers of soil organic carbon transport through the Carminowe Creek catchment (southwest England). Biogeosciences 17, 3183–3201.

Guo, J., Ma, T., Liu, N., Zhang, X., Hu, H., Ma, W., Wang, Z., Feng, X., Peterse, F., 2021. Soil pH and aridity influence distributions of branched tetraether lipids in grassland soils along an aridity transect. Org Geochem. 104347.

Häggi, C., Schefuß, E., Sawakuchi, A.O., Chiessi, C.M., Mulitza, S., Bertassoli, D.J., Hefter, J., Zabel, M., Baker, P.A., Schouten, S., 2019. Modern and late Pleistocene particulate organic carbon transport by the Amazon River: Insights from long-chain alkyl diols. Geochim. Cosmochim. Acta 262, 1–19.

Hopmans, E.C., Weijers, J.W.H., Schefuß, E., Herfort, L., Sinninghe Damsté, J.S., Schouten, S., 2004. A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. Earth Planet. Sci. Lett. 224, 107–116.

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

Huguet, A., Fosse, C., Metzger, P., Fritsch, E., Derenne, S., 2010. Occurrence and distribution of extractable glycerol dialkyl glycerol tetraethers in podzols. Org Geochem. 41, 291–301.

Huguet, A., Gocke, M., Derenne, S., Fosse, C., Wiesenberg, G.L.B., 2013. Root-associated branched tetraether source microorganisms may reduce estimated paleotemperatures in subsoil. Chem. Geol. 356, 1–10.

Jaeschke, A., Rethemeyer, J., Lappé, M., Schouten, S., Boeckx, P., Schefuß, E., 2018. Influence of land use on distribution of soil n-alkane δD and brGDGTs along an altitudinal transect in Ethiopia: Implications for (paleo)environmental studies. Org Geochem. 124, 77–87.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4, 170122.

Kirkels, F.M.S.A., Ponton, C., Galy, V., West, A.J., Feakins, S.J., Peterse, F., 2020. From andes to amazon: assessing branched tetraether lipids as tracers for soil organic carbon in the Madre de Dios river system. J. Geophys. Res. Biogeosci. 125, e2019JG005270.

Koga, Y., Nishihara, M., Morii, H., Akagawa-Matsushita, M., 1993. Ether polar lipids of methanogenic bacteria: Structures, comparative aspects, and biosyntheses. Microbiol. Rev. 57, 164–182.

Kusch, S., Winterfeld, M., Mollenhauer, G., Höfle, S.T., Schirrmeister, L., Schwamborn, G., Rethemeyer, J., 2019. Glycerol dialkyl glycerol tetraethers (GDGTs) in high latitude Siberian permafrost: Diversity, environmental controls, and implications for proxy applications. Org Geochem. 136, 103888.

Lembrechts, J.J., Aalto, J., Ashcroft, M.B., De Frenne, P., Kopecký, M., Lenoir, J., Luoto, M., Maclean, I.M.D., Roupsard, O., Fuentes-Lillo, E., García, R.A., Pellissier, L., Pitteloud, C., Alatalo, J.M., Smith, S.W., Björk, R.G., Muffler, L., Ratier Backes, A., Cesarz, S., Gottschall, F., Okello, J., Urban, J., Plichta, R., Svátek, M., Phartyal, S.S., Wipf, S., Eisenhauer, N., Puşcaş, M., Turtureanu, P.D., Varlagin, A., Dimarco, R.D., Jump, A.S., Randall, K., Dorrepaal, E., Larson, K., Walz, J., Vitale, L., Svoboda, M., Finger Higgens, R., Halbritter, A.H., Curasi, S.R., Klupar, I., Koontz, A., Pearse, W.D., Simpson, E., Stemkovski, M., Jessen Graae, B., Vedel Sørensen, M., Høye, T.T., Fernández Calzado, M.R., Lorite, J., Carbognani, M., Tomaselli, M., Forte, T.A.G.W., Petraglia, A., Haesen, S., Somers, B., Van Meerbeek, K., Björkman, M.P., Hylander, K., Merinero, S., Gharun, M., Buchmann, N., Dolezal, J., Matula, R., Thomas, A.D., Bailey, J.J., Ghosn, D., Kazakis, G., de Pablo, M.A., Kemppinen, J., Niittynen, P., Rew, L., Seipel, T., Larson, C., Speed, J.D.M., Ardö, J., Cannone, N., Guglielmin, M., Malfasi, F., Bader, M.Y., Canessa, R., Stanisci, A., Kreyling, J., Schmeddes, J., Teuber, L., Aschero, V., Čiliak, M., Máliš, F., De Smedt, P., Govaert, S., Meeussen, C., Vangansbeke, P., Gigauri, K., Lamprecht, A., Pauli, H., Steinbauer, K., Winkler, M., Ueyama, M., Nuñez, M.A., et al., 2020. SoilTemp: A global database of near-surface temperature. Glob. Chang. Biol. 26, 6616–6629.

Li, T.-T., 1926. Soil temperature as influenced by forest cover. Yale School of Forest. Bull. 18, 1–92.

Li, Y., Zhang, C.Q., Wang, N.A., Han, Q., Zhang, X.Z., Liu, Y., Xu, L.M., Ye, W.T., 2017. Substantial inorganic carbon sink in closed drainage basins globally. Nat. Geosci. 10, 501–506.

Li, Y., Zhao, S., Pei, H., Qian, S., Zang, J., Dang, X., Yang, H., 2018. Distribution of glycerol dialkyl glycerol tetraethers in surface soils along an altitudinal transect at cold and humid Mountain Changbai: Implications for the reconstruction of paleoaltimetry and paleoclimate. Sci. China Earth Sci. 61, 925–939.

Li, Y., Shi, W., Aydin, A., Beroya-Eitner, M.A., Gao, G., 2020. Loess genesis and worldwide distribution. Earth Sci. Rev. 201, 102947.

Liang, J., Russell, J.M., Xie, H., Lupien, R.L., Si, G., Wang, J., Hou, J., Zhang, G., 2019. Vegetation effects on temperature calibrations of branched glycerol dialkyl glycerol tetraether (brGDGTs) in soils. Org Geochem. 127, 1–11.

Liu, W., Wang, H., Zhang, C.L., Liu, Z., He, Y., 2013. Distribution of glycerol dialkyl glycerol tetraether lipids along an altitudinal transect on Mt. Xiangpi, NE Qinghai-Tibetan Plateau, China. Org. Geochem. 57, 76–83.

Lu, H., Liu, W., Wang, H., Wang, Z., 2016. Variation in 6-methyl branched glycerol dialkyl glycerol tetraethers in Lantian loess–paleosol sequence and effect on paleotemperature reconstruction. Org Geochem. 100, 10–17.

Lu, H., Liu, W., Yang, H., Wang, H., Liu, Z., Leng, Q., Sun, Y., Zhou, W., An, Z., 2019. 800-kyr land temperature variations modulated by vegetation changes on Chinese Loess Plateau. Nat. Commun. 10, 1958.

Martínez-Sosa, P., Tierney, J.E., Stefanescu, I.C., Dearing Crampton-Flood, E., Shuman, B.N., Routson, C., 2021. A global Bayesian temperature calibration for lacustrine brGDGTs. Geochim. Cosmochim. Acta 305, 87–105.
Mescolotti, P.C., Giannini, P.C.F., Pupim, F.d.N., Sawakuchi, A.O., Ladeira, F.S.B., Assine, M.L., 2023. The largest Quaternary inland eolian system in Brazil: Eolian landforms and activation/stabilization phases of the Xique-Xique dune field. Geomorphology 420, 108516.
Miller, D.R., Habicht, M.H., Keisling, B.A., Castañeda, I.S., Bradley, R.S., 2018. A 900-year New England temperature reconstruction from in situ seasonally produced branched glycerol dialkyl glycerol tetraethers (brGDGTs). Clim. Past 14, 1653–1667.
Naafs, B.D.A., Gallego-Sala, A.V., Inglis, G.N., Pancost, R.D., 2017a. Refining the global branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration. Org Geochem. 106, 48–56.
Naafs, B.D.A., Inglis, G.N., Zheng, Y., Amesbury, M.J., Biester, H., Bindler, R., Blewett, J., Burrows, M.A., del Castillo Torres, D., Chambers, F.M., Cohen, A.D., Evershed, R.P., Feakins, S.J., Gałka, M., Gallego-Sala, A., Gandois, L., Gray, D.M., Hatcher, P.G., Honorio Coronado, E.N., Hughes, P.D.M., Huguet, A., Könönen, M., Laggoun-Défarge, F., Lähteenoja, O., Lamentowicz, M., Marchant, R., McClymont, E., Pontevedra-Pombal, X., Ponton, C., Pourmand, A., Rizzuti, A.M., Rochefort, L., Schellekens, J., De Vleeschouwer, F., Pancost, R.D., 2017b. Introducing global peat-specific temperature and pH calibrations based on brGDGT bacterial lipids. Geochim. Cosmochim. Acta 208, 285–301.
Naafs, B.D.A., Rohrssen, M., Inglis, G.N., Lähteenoja, O., Feakins, S.J., Collinson, M.E., Kennedy, E.M., Singh, P.K., Singh, M.P., Lunt, D.J., Pancost, R.D., 2018. High temperatures in the terrestrial mid-latitudes during the early Palaeogene. Nat. Geosci. 11, 766–771.
Naafs, B.D.A., Inglis, G.N., Blewett, J., McClymont, E.L., Lauretano, V., Xie, S., Evershed, R.P., Pancost, R.D., 2019. The potential of biomarker proxies to trace climate, vegetation, and biogeochemical processes in peat: A review. Glob. Planet. Change 179, 57–79.
Naafs, B.D.A., Oliveira, A.S.F., Mulholland, A.J., 2021. Molecular dynamics simulations support the hypothesis that the brGDGT paleothermometer is based on homeoviscous adaptation. Geochim. Cosmochim. Acta 312, 44–56.
Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L., Schleper, C., 2003. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. Environ. Microbiol. 5, 787–797.
Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial ecoregions of the worlds: A new map of life on Earth. Bioscience 51, 933–938.
Osman, M.B., Tierney, J.E., Zhu, J., Tardif, R., Hakim, G.J., King, J., Poulsen, C.J., 2021. Globally resolved surface temperatures since the Last Glacial Maximum. Nature 599, 239–244.
Peaple, M.D., Beverly, E.J., Garza, B., Baker, S., Levin, N.E., Tierney, J.E., Häggi, C., Feakins, S.J., 2022. Identifying the drivers of GDGT distributions in alkaline soil profiles within the Serengeti ecosystem. Org Geochem. 169, 104433.
Pearson, E.J., Juggins, S., Talbot, H.M., Weckström, J., Rosén, P., Ryves, D.B., Roberts, S. J., Schmidt, R., 2011. A lacustrine GDGT-temperature calibration from the Scandinavian Arctic to Antarctic: Renewed potential for the application of GDGT-paleothermometry in lakes. Geochim. Cosmochim. Acta 75, 6225–6238.
Pei, H., Zhao, S., Yang, H., Xie, S., 2021. Variation of branched tetraethers with soil depth in relation to non-temperature factors: Implications for paleoclimate reconstruction. Chem. Geol. 572, 120211.
Pérez-Angel, L.C., Sepúlveda, J., Molnar, P., Montes, C., Rajagopalan, B., Snell, K., Gonzalez-Arango, C., Dildar, N., 2020. Soil and air temperature calibrations using branched GDGTs for the tropical andes of colombia: toward a pan-tropical calibration. Geochem. Geophys. Geosyst. 21, e2020GC008941.
Peterse, F., van der Meer, J., Schouten, S., Weijers, J.W.H., Fierer, N., Jackson, R.B., Kim, J.-H., Sinninghe Damsté, J.S., 2012. Revised calibration of the MBT–CBT paleotemperature proxy based on branched tetraether membrane lipids in surface soils. Geochim. Cosmochim. Acta 96, 215–229.
Peterse, F., Vonk, J.E., Holmes, R.M., Giosan, L., Zimov, N., Eglinton, T.I., 2014. Branched glycerol dialkyl glycerol tetraethers in Arctic lake sediments: Sources and implications for paleothermometry at high latitudes. J. Geophys. Res. Biogeo. 119, 1738–1754.
Pitcher, A., Schouten, S., Sinninghe Damsté, J.S., 2009. In situ production of crenarchaeol in two California hot springs. Appl. Environ. Microbiol. 75, 4443–4451.
R_Core_Team, 2021, R: A Language and Environment for Statistical Computing: Vienna, Austria, R Foundation for Statistical Computing.
Raberg, J.H., Miller, G.H., Geirsdóttir, Á., Sepúlveda, J., 2022. Near-universal trends in brGDGT lipid distributions in nature. Sci. Adv. 8, eabm7625.
Ruggiero, P.G.C., Batalha, M.A., Pivello, V.R., Meirelles, S.T., 2002. Soil-vegetation relationships in cerrado (Brazilian savanna) and semideciduous forest, Southeastern Brazil. Plant Ecol. 160, 1–16.
Russell, J.M., Hopmans, E.C., Loomis, S.E., Liang, J., Sinninghe Damsté, J.S., 2018. 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.
Schouten, S., Hopmans, E.C., Sinninghe Damsté, J.S., 2013. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. Org Geochem. 54, 19–61.
Seltzer, A.M., Ng, J., Aeschbach, W., Kipfer, R., Kulongoski, J.T., Severinghaus, J.P., Stute, M., 2021. Widespread six degrees Celsius cooling on land during the Last Glacial Maximum. Nature 593, 228–232.
Shi, Z., Allison, S.D., He, Y., Levine, P.A., Hoyt, A.M., Beem-Miller, J., Zhu, Q., Wieder, W.R., Trumbore, S., Randerson, J.T., 2020. The age distribution of global soil carbon inferred from radiocarbon measurements. Nat. Geosci. 13, 555–559.
Silvestro, D., Andermann, T., 2020, Prior choice affects ability of Bayesian neural networks to identify unknowns. arXiv:2005.04987v1.
Sinninghe Damsté, J.S., 2016. Spatial heterogeneity of sources of branched tetraethers in shelf systems: The geochemistry of tetraethers in the Berau River delta (Kalimantan, Indonesia). Geochim. Cosmochim. Acta 186, 13–31.
Sinninghe Damsté, J.S., Hopmans, E.C., Pancost, R.D., Schouten, S., Geenevasen, J.A.J., 2000. Newly discovered non-isoprenoid glycerol dialkyl glycerol tetraether lipids in sediments. Chem. Commun. 1683–1684.
Sinninghe Damsté, J.S., Schouten, S., Hopmans, E.C., van Duin, A.C., Geenevasen, J.A., 2002. Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. J. Lipid Res. 43, 1641–1651.
Sinninghe Damsté, J.S., Ossebaar, J., Abbas, B., Schouten, S., Verschuren, D., 2009. Fluxes and distribution of tetraether lipids in an equatorial African lake: Constraints on the application of the TEX86 palaeothermometer and BIT index in lacustrine settings. Geochim. Cosmochim. Acta 73, 4232–4249.
Sinninghe Damsté, J.S., Rijpstra, W.I.C., Hopmans, E.C., Jung, M.-Y., Kim, J.-G., Rhee, S.-K., Stieglmeier, M., Schleper, C., 2012. Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids of group I.1a and I.1b thaumarchaeota in soil. Appl. Environ. Microbiol. 78, 6866–6874.
Sinninghe Damsté, J.S., Rijpstra, W.I.C., Foesel, B.U., Huber, K.J., Overmann, J., Nakagawa, S., Kim, J.J., Dunfield, P.F., Dedysh, S.N., Villanueva, L., 2018. An overview of the occurrence of ether- and ester-linked iso-diabolic acid membrane lipids in microbial cultures of the Acidobacteria: Implications for brGDGT paleoproxies for temperature and pH. Org Geochem. 124, 63–76.
Staver, A.C., Archibald, S., Levin, S.A., 2011. The global extent and determinants of savanna and forest as alternative biome states. Science 334, 230–232.
Sugihara, S., Shibata, M., Mvondo Ze, A.D., Araki, S., Funakawa, S., 2015. Effects of vegetation on soil microbial C, N, and P dynamics in a tropical forest and savanna of Central Africa. Appl. Soil Ecol. 87, 91–98.
Tang, C., Yang, H., Pancost, R.D., Griffiths, M.L., Xiao, G., Dang, X., Xie, S., 2017. Tropical and high latitude forcing of enhanced megadroughts in Northern China during the last four terminations. Earth Planet. Sci. Lett. 479, 98–107.
Tierney, J.E., Zhu, J., King, J., Malevich, S.B., Hakim, G.J., Poulsen, C.J., 2020. Glacial cooling and climate sensitivity revisited. Nature 584, 569–573.
Véquaud, P., Thibault, A., Derenne, S., Anquetil, C., Collin, S., Contreras, S., Nottingham, A.T., Sabatier, P., Werne, J.P., Huguet, A., 2022. FROG: A global machine-learning temperature calibration for branched GDGTs in soils and peats. Geochim. Cosmochim. Acta 318, 468–494.
Vieira, C.K., Borges, L.G.D., Marconatto, L., Giongo, A., Sturmer, S.L., 2018. Microbiome of a revegetated iron-mining site and pristine ecosystems from the Brazilian Cerrado. Appl. Soil Ecol. 131, 55–65.
Vos, M., Wolf, A.B., Jennings, S.J., Kowalchuk, G.A., 2013. Micro-scale determinants of bacterial diversity in soil. FEMS Microbiol. Rev. 37, 936–954.
Wang, H., An, Z., Lu, H., Zhao, Z., Liu, W., 2020. Calibrating bacterial tetraether distributions towards in situ soil temperature and application to a loess-paleosol sequence. Quat. Sci. Rev. 231, 106172.
Wang, H., Liu, W., 2021. Soil temperature and brGDGTs along an elevation gradient on the northeastern Tibetan Plateau: A test of soil brGDGTs as a proxy for paleoelevation. Chem. Geol. 566, 120079.
Wang, M., Zong, Y., Zheng, Z., Man, M., Hu, J., Tian, L., 2018. Utility of brGDGTs as temperature and precipitation proxies in subtropical China. Sci. Rep. 8, 194.
Wang, M., Zheng, Z., Zong, Y., Man, M., Tian, L., 2019. Distributions of soil branched glycerol dialkyl glycerol tetraethers from different climate regions of China. Sci. Rep. 9, 2761.
Weber, Y., Damsté, J.S.S., Zopfi, J., Jonge, C.D., Gilli, A., Schubert, C.J., Lepori, F., Lehmann, M.F., Niemann, H., 2018. Redox-dependent niche differentiation provides evidence for multiple bacterial sources of glycerol tetraether lipids in lakes. Proc. Nat. Acad. Sci. 115, 10926–10931.
Weijers, J.W.H., Schouten, S., Spaargaren, O.C., Damste, J.S.S., 2006. Occurrence and distribution of tetraether membrane lipids in soils: Implications for the use of the TEX86 proxy and the BIT index. Org Geochem. 37, 1680–1693.
Weijers, J.W.H., Schouten, S., van den Donker, J.C., Hopmans, E.C., Damste, J.S.S., 2007. Environmental controls on bacterial tetraether membrane lipid distribution in soils. Geochim. Cosmochim. Acta 71, 703–713.
Weijers, J.W.H., Bernhardt, B., Peterse, F., Werne, J.P., Dungait, J.A.J., Schouten, S., Sinninghe Damsté, J.S., 2011. Absence of seasonal patterns in MBT–CBT indices in mid-latitude soils. Geochim. Cosmochim. Acta 75, 3179–3190.
Wuyts, B., Champneys, A.R., House, J.I., 2017. Amazonian forest-savanna bistability and human impact. Nat. Commun. 8, 15519.
Xie, S.C., Pancost, R.D., Chen, L., Evershed, R.P., Yang, H., Zhang, K.X., Huang, J.H., Xu, Y.D., 2012. Microbial lipid records of highly alkaline deposits and enhanced aridity associated with significant uplift of the Tibetan Plateau in the Late Miocene. Geology 40, 291–294.
Yang, H., Ding, W., Zhang, C.L., Wu, X., Ma, X., He, G., Huang, J., Xie, S., 2011. Occurrence of tetraether lipids in stalagmites: Implications for sources and GDGT-based proxies. Org Geochem. 42, 108–115.
Yang, H., Pancost, R.D., Dang, X., Zhou, X., Evershed, R.P., Xiao, G., Tang, C., Gao, L., Guo, Z., Xie, S., 2014. Correlations between microbial tetraether lipids and environmental variables in Chinese soils: Optimizing the paleo-reconstructions in semi-arid and arid regions. Geochim. Cosmochim. Acta 126, 49–69.
Yang, H., Xiao, W., Słowakiewicz, M., Ding, W., Ayari, A., Dang, X., Pei, H., 2019. Depth-dependent variation of archaeal ether lipids along soil and peat profiles from

southern China: Implications for the use of isoprenoidal GDGTs as environmental tracers. Org Geochem. 128, 42–56.
Zang, J., Lei, Y., Yang, H., 2018. Distribution of glycerol ethers in Turpan soils: implications for use of GDGT-based proxies in hot and dry regions. Front. Earth Sci. 12, 862–876.
Zell, C., Kim, J.H., Moreira-Turcq, P., Abril, G., Hopmans, E.C., Bonnet, M.P., Sobrinho, R.L., Sinnighe Damsté, J.S., 2013. Disentangling the origins of branched tetraether lipids and crenarchaeol in the lower Amazon River: Implications for GDGT-based proxies. Limnol. Oceanogr. 58, 343–353.
Zhang, Y.G., Zhang, C.L., Liu, X.-L., Li, L., Hinrichs, K.-U., Noakes, J.E., 2011. Methane Index: A tetraether archaeal lipid biomarker indicator for detecting the instability of marine gas hydrates. Earth Planet. Sci. Lett. 307, 525–534.
Zhao, B., Russell, J.M., Tsai, V.C., Blaus, A., Parish, M.C., Liang, J., Wilk, A., Du, X., Bush, M.B., 2023. Evaluating global temperature calibrations for lacustrine branched GDGTs: Seasonal variability, paleoclimate implications, and future directions. Quat. Sci. Rev. 310, 108124.
Zheng, D., Hunt, E.R., Running, S.W., 1993. A daily soil temperature model based on air temperature and precipitation for continental applications. Climate Res. 2, 183–191.