University of Birmingham
Research at Birmingham

Global calibration of novel 3-hydroxy fatty acid based temperature and pH proxies

Wang, Canfa; Bendle, James A.; Yang, Huan; Yang, Yi; Hardman, Alice; Yamoah, Afrifa; Thorpe, Amy; Mandel, Ilya; Greene, Sarah E.; Huang, Junhua; Xie, Shucheng

DOI:
10.1016/j.gca.2021.03.010

License:
Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Wang, C, Bendle, JA, Yang, H, Yang, Y, Hardman, A, Yamoah, A, Thorpe, A, Mandel, I, Greene, SE, Huang, J & Xie, S 2021, 'Global calibration of novel 3-hydroxy fatty acid based temperature and pH proxies', *Geochimica et Cosmochimica Acta*, vol. 302, pp. 101-119. https://doi.org/10.1016/j.gca.2021.03.010

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 23. Apr. 2024

Global calibration of novel 3-hydroxy fatty acid based temperature and pH proxies

Canfa Wang[a], James A. Bendle[b], Huan Yang[c], Yi Yang[c], Alice Hardman[b], Afrifa Yamoah[b], Amy Thorpe[b], Ilya Mandel[d,e,f], Sarah E. Greene[b], Junhua Huang[g], Shucheng Xie[a]

[a] State Key Laboratory of Biogeology and Environmental Geology, Hubei Key Laboratory of Critical Zone Evolution, China University of Geosciences, Wuhan, 430074, China

[b] School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

[c] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China

[d] School of Physics and Astronomy, Monash University, VIC 3800, Australia

[e] OzGrav: The Australian Research Council Centre of Excellence for Gravitational Wave Discovery, Clayton, VIC 3800, Australia

[f] Birmingham Institute for Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK

[g] State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan, 430074, China

Abstract

3-Hydroxy fatty acids (3-OH-FAs), derived from Gram-negative bacterial outer membranes, have received recent attention for their potential as new terrestrial pH and temperature proxies for palaeoclimate studies. Initial studies from altitudinal transects of contemporary soils - correlating bacterial 3-OH-FA compositions to air temperature and pH - have shown promising results. But the geographical extent of recent calibrations is limited. In this study, we analyse 3-OH-FA lipid distributions in 186 globally distributed soil samples to study the environmental factors controlling the relative distribution of the 3-OH-FA isomers. Our sample-set covers a wide range of temperatures (-0.4 to 27°C) and pH (3.6 to 9.2). For the global compilation we find that the ratio of *anteiso* to *normal* 3-OH-FAs of the C_{15} or C_{17} homologues (RAN_{15} or RAN_{17}) shows a strong linear relationship with mean annual air temperature (MAAT) ($R^2=0.48$, $p < 0.001$ and $R^2 = 0.41$, $p < 0.001$, respectively). Additionally, the negative logarithm of the ratio of the summed *iso* and *anteiso* to the total amount of *normal* 3-OH-FAs (RIAN) is also strongly anticorrelated with the soil pH ($R^2 = 0.66$, $p < 0.001$). However, we find that for our 3-OH-FA based proxies there are significant differences in slope and intercept of the linear corrections at regional scales. Thus local or regional calibrations are likely preferable (at this stage of 3-OH-FA proxy development) for application to specific palaeoclimate archives. We also explore the relationship of 3-OH-FA isomer fractional abundances to environmental parameters using machine learning tools (a Gaussian Process (GP) emulator). This confirms the first order relationships to environmental parameters highlighted by the empirical equations and also derives several alternative GP emulator models for reconstructing MAAT and pH which give higher R^2 values (0.66 for MAAT; 0.63 for pH) and lower RSME values (3.5°C for MAAT; 0.76 for pH) compared to simple linear regressions at the global scale. We compare our 3-OH-FA based indices with bacterial branched glycerol dialkyl glycerol tetraethers (brGDGTs) based indices from the same soil samples. At a global scale RAN_{15} and RAN_{17} show negative correlations with the $\mathrm{MBT'}_{5\mathrm{ME}}$-MAAT ($\mathrm{MBT'}_{5\mathrm{ME}}$, methylation index of 5-methyl branched tetraethers) ($r = -0.59$, $p < 0.001$ and $r = -0.42$, $p < 0.001$, respectively), whilst RIAN shows strong linear correlations with the cyclisation ratio of branched tetraethers (CBT) ($r = 0.77$, $p < 0.001$). Similar to 3-OH-FA based temperature proxies, GDGT based temperature proxy $\mathrm{MBT'}_{5\mathrm{ME}}$ also showed different regional calibrations. Our new field-based correlations demonstrate the broad physiological response of Gram-negative bacterial cell membranes to external environmental changes on a global scale. We suggest that 3-OH-FA based proxies have widespread potential for palaeoenvironmental studies to estimate past MAAT and soil pH, but that regional/ local and context specific calibrations may need to be applied.

Keywords: 3-Hydroxy fatty acid; 3-OH-FA; Soils; Proxies; Temperature; pH; Palaeoclimate; Biomarkers

1. Introduction

Instrumental records, satellite observations and laboratory studies do not cover the likely amplitude or patterns of response of Earth’s climate and carbon system to the extreme climate forcing expected this century (IPCC, 2014). Reconstruction of past climate change, beyond the scope of meteorological records, is critical for providing natural baselines, improving understanding of the Earth system and predicting future change. A wide range of environmental information from both terrestrial and marine realms is required from palaeoclimate archives for this endeavour. Microbial lipids are sensitive to ambient environmental changes. A number of organic geochemical proxies based on microbial lipids have been developed for palaeoclimate reconstruction (Eglinton and Eglinton, 2008; Luo et al., 2019; Meyers, 1997; Schouten et al., 2013). Three lipid biomarker based indices, TEX$_{86}$ (Kim et al., 2008; Schouten et al., 2002), $U^{K'}_{37}$ (Brassell et al., 1986; Haug et al., 2005; Prahl and Wakeham, 1987; Sachs et al., 2001) and LDI (de Bar et al., 2020; Naafs et al., 2012; Rampen et al., 2012) have become important tools for determination of past sea surface temperature (SST). However, the above-mentioned proxies are generally applied in marine settings and biomarker based proxies for terrestrial environments, especially for temperature, remain relatively scarce. This is unfortunate as the terrestrial environment is where the climate change impacts will most affect human societies. Bacterial branched glycerol dialkyl glycerol tetraethers (brGDGTs) are the primary biomarker based proxy for temperature and pH (Peterse et al., 2012; Weijers et al., 2007) currently applied to terrestrial archives (Schouten et al., 2013 and references therein). Using improved chromatographic separation, a new temperature proxy MBT'_{5ME} was defined, which is pH independent and reduces the residual mean error (RMSE) for mean annual air temperature (MAAT) reconstructions (De Jonge et al., 2013; De Jonge et al., 2014; Hopmans et al., 2016). However, the utility of GDGT based approaches is still limited by uncertainties over the biological source (Weber et al., 2015), *in-situ* production and transport of brGDGTs in lake settings (Blaga et al., 2010). We note that several novel terrestrial bacterial biomarker based proxies have been recently proposed, namely the branched fatty alcohol ratio BNA$_{15}$ (Huang et al., 2013) and several proxies based on heterocyst glycolipids (HG$_{28}$ and HG$_{30}$) (Bauersachs et al., 2015; Klages et al., 2020). The BNA$_{15}$, HG$_{28}$ and HG$_{30}$ proxies show promise but have yet to be globally calibrated and widely applied. Finally, neither GDGTs nor HGs are readily amenable to isotopic analyses using standard methods, limiting potential insights to the terrestrial carbon and hydrological cycles. We seek to overcome these limitations by developing a new suite of terrestrial palaeoclimatic proxies that can reconstruct temperature and pH independently (and which have the future potential to yield isotopic information using routine analytical approaches). Thus, further development of novel terrestrial proxies, independent and complementary to GDGTs, is needed to expand applications and improve the reliability and accuracy of terrestrial environmental reconstructions.

Gram-negative bacterial membrane derived 3-hydroxy fatty acids (3-OH-FAs) have the potential to be developed as environmental proxies. 3-OH-FAs with carbon numbers from C_{10} to C_{18} are primarily derived from lipid A, a constituent of lipopolysaccharide (LPS), the main component of the outer membrane of Gram-negative bacteria (Szponar et al., 2003; Szponar et al., 2002; Wollenweber and Rietschel, 1990). Gram-negative derived 3-OH-FAs are bound to the glucosamine unit of lipid A either by ester bonds or amide bonds (Kumar et al., 2002; Raetz et al., 2007; Wollenweber and Rietschel, 1990). Acid digestion is a more appropriate method than saponification to extract them from soil and stalagmite samples (Wang et al., 2016; Yang et al., 2016). So far 3-OH-FAs have been found in soils (Huguet et al., 2019; Wang et al., 2016; Zelles, 1999), speleothems (Blyth et al., 2006; Huang et al., 2008; Wang et al., 2018; Wang et al., 2012), snow (Tyagi et al., 2016; Tyagi et al., 2015), aerosols (Lee et al., 2004), marine dissolved organic matter (DOM) (Wakeham et al., 2003), marine and lake sediments (Kawamura and Ishiwatari, 1984; Volkman et al., 1980; Wakeham, 1999; Wang et al., 2016; Yang et al., 2020; Zhang et al., 2014), and a 3-OH-FA based proxy for sea surface temperature (RAN_{13}) has recently been proposed (Yang et al., 2020) suggesting the potential for wide application if proxies based on 3-

OH-FA are available. Because Gram-negative bacteria are ubiquitous, 3-OH-FAs proxies could be applied to diverse archives, providing cross-correlation between speleothems (Wang et al., 2018), lake sediments, palaeosols and marine records (Yang et al., 2020). Proxies that span this environmental range are essential for elucidating links between marine and terrestrial climate change.

Even though the wide environmental occurrence of 3-OH-FAs has been known for some time, the development of 3-OH-FA based independent terrestrial environmental proxies was only recently initiated by Wang et al. (2016). Specifically, two temperature proxies, the ratio of *anteiso* to *normal* C_{15} 3-OH-FA (RAN_{15}, see Fig. 1 for example structures) and the ratio of *anteiso* to *normal* C_{17} 3-OH-FA (RAN_{17}), were proposed as novel and independent temperature proxies (Wang et al., 2016). Several pH proxies, such as the ratio of the total sum of *iso* and *anteiso* 3-OH-FAs to the total amount of *normal* 3-OH-FAs (Branching Ratio) and the negative logarithm of Branching Ratio (RIAN), were proposed as novel pH proxies (Wang et al., 2016). The 3-OH-FA based proxies for temperature (RAN_{15}) and pH proxy (RIAN) were successfully applied to a stalagmite to produce the first biomarker based temperature and hydrological reconstructions from a speleothem archive (Wang et al., 2018). Studies of 3-OH-FAs from two altitudinal transects have confirmed the promise of these temperature and pH proxies (Huguet et al., 2019). Initial calibrations were limited to altitudinal soil transects from Mt. Shennongjia (central China), Mt. Rungwe (SW Tanzania) and Mt. Majella (central Italy), with a limited number of samples (Huguet et al., 2019; Wang et al., 2016). Recent work on additional altitudinal transects in Italy, Tibet and the Andes expands the number of sites investigated globally (Véquaud et al., 2020). Strong linear relationships between 3-OH FA-derived indices (RAN_{15}, RAN_{17} and RIAN) and MAAT/pH were obtained locally, but also highlighted variation in calibration slopes and intercepts between discreet altitudinal transects (Véquaud et al., 2020). Another recent study from the French Alps found a high degree of scatter in the relationship between $RAN_{15}/_{17}$ and MAAT and taken together with the relatively weak relationships found on Mt. Majella suggests the relative abundance of these lipids maybe influenced by factors other than temperature and pH (Véquaud et al., 2020). Thus investigation based on a globally distributed soil sample set, including lowland samples and samples distributed at continental scales is needed to further explore the widespread applicability and constraint the accuracy of 3-OH-FA based proxies.

Here we aim to improve the accuracy and representativeness of the 3-OH-FA based proxies, extending the sample set of Wang et al. (2016) by adding 112 new surface soil samples globally located, and combining recently reported 3-OH-FAs distributions in soils from central China (Wang et al., 2018), NW Tanzania and central Italy (Huguet et al., 2019). The updated dataset confirms the first-order physiological response of Gram-negative bacterial membrane lipids to environmental drivers, but also finds significant differences in slopes and incepts of correlations and regional scales. Suggesting 3-OH-FA based proxies have great potential for widespread environmental applications, but that regional/ local calibrations and context will likely be required.

2. Materials and methods

2.1 Soil sample collection and compilation

Surface soils (0-10cm) used for this study are predominantly obtained from the soil sample repository of the International Soil Reference and Information Centre (ISRIC) in Wageningen, Netherlands, and from China and US. We obtained as many samples as possible (83) from the ISRIC repository that were previously studied for GDGT analysis by Weijers et al. (2007), Peterse et al. (2012) and De Jonge et al. (2014), and from China which GDGT analysis were conducted previously by Yang et al. (2014) and Lei et al. (2016). In addition to the samples previously studied by Weijers et al., 2007 (and others), we collected a number of new samples in the field. The final sample dataset is composed of 186 globally distributed surface soils (Figs. 2 and S1;Supplementary Data), with 112 soil samples analysed for 3-OH-FAs in this study and 26 soil samples reported by Wang et al. (2016), 9 soil samples reported by Wang et al. (2018) and 39 soil samples reported by Huguet et al. (2019). The MAAT for the soil sampling sites ranged from -0.4 to 27°C. The soil pH of all soil samples ranged from 3.60 to 9.20.

2.2 Determination of environmental parameters

If available, soil pH data either comes from Weijers et al. (2007) (which is originally obtained from the ISRIC Soil Information System database), Yang et al. (2014), Lei et al. (2016) and Huguet et al. (2019). The pH of the remaining soils were measured following the method of Yang et al. (2014), specifically, soil samples were mixed with ultrapure water in a ratio of 1:2.5 (g/mL). After standing for 30 min, the supernatant pH was measured, using a pH meter with a precision of ±0.01. The pH was measured three times and the mean value was taken as the final pH.

The mean annual air temperature (MAAT) and mean annual precipitation (MAP) are from meteorological stations nearest to the sample locations. The climatic data for soil samples from ISRIC represents a 30-year average over the period 1961–1990 (Weijers et al., 2007), for the soil samples from the US a 20-year average over the period of 1998 to 2017, for the rest of the soil samples a 30-year average over the period

1970-2000. If necessary, a temperature correction was performed for differences in altitude between the sample location and the weather station.

2.3 Extraction of 3-OH-FAs

The soil samples were freeze dried and ground with a mortar and pestle prior to extraction. The samples were subjected to acid hydrolysis following an optimized acid digestion method (Blyth et al., 2006; Wang et al., 2012). 10g of homogenized sample was mixed with 30 mL pre-cleaned HCl (3M), and then refluxed at 130 °C for 3h. After cooling, the solution was extracted x3 with DCM, to yield the Total Lipid Extract (TLE). The TLE was methylated by BF_3-MeOH solution at 70 °C for 1.5h. The resulting fatty acid methyl esters (FAMEs) were separated into non-OH-FAMEs and OH-FAMEs by silica gel column following the method described by Jenske and Vetter (2008). Non-OH-FAMEs were eluted in the first fraction with a solvent mixture of *n*-hexane and ethyl acetate (v:v, 98:2), whereas OH-FAMEs were obtained by elution with 100% ethyl acetate. The OH-FAME fraction was further derivatised by BSTFA (N, O-bis (trimethylsilyl) trifluoroacetamide) at 70 °C for 1.5 h before further analysis by gas chromatogram-mass spectrometer (GC-MS).

2.4 GC-MS analysis of 3-OH-FAs

The 3-OH-FAs were analysed by an Agilent 7890A gas chromatogram and 5975C mass spectrometer (GC-MS) equipped with a DB-5MS fused silica capillary column (60 m × 0.25 mm × 0.25 μm). The GC oven temperature was ramped from 70 °C to 200 °C at 10 °C /min, then to 310 °C at 3 °C /min, held at 310 °C for 30 min. The carrier gas was Helium (99.999%) and the gas flow was 1.0 mL/min. The ionization energy of the mass spectrometer was set at 70 eV. The 3-OH-FAs were identified based on their mass spectra and relative retention times, 3-OH-FA isomers with same carbon number come out in order of *iso*, *anteiso* and *normal* (Fig. 3). All the 3-OH-FAs TMSi esters show diagnostic fragment ions, m/z 175 ($[CH_3]_3SiO= CHCH_2CO_2CH_3^+$), due to the cleavage between C_3 and C_4, and M^+-15 (base peak) results from a loss of a CH_3 group. Other characteristic ions include m/z 103, 89, 133, 159, and M^+-31 (Eglinton et al., 1968; Mielniczuk et al., 1993; Volkman et al., 1999; Wang et al., 2016).

2.5 3-OH-FA based indices and mathematical analysis

2.5.1 Calculation of 3-OH-FA based indices

3-OH-FA based indices, in particular the RAN_{15}, RAN_{17} and RIAN, were calculated using the following equations, which were previously developed by Wang et al. (2016):

$$RAN_{15}= a\text{-}C_{15} / n\text{-}C_{15} \text{ 3-OH-FA} \quad (1)$$

$$RAN_{17}= a\text{-}C_{17} / n\text{-}C_{17} \text{ 3-OH-FA} \quad (2)$$

Where *a*- represents the *anteiso* homologue of 3-OH-FA, *n*- represents the *normal* homologue of 3-OH-FA.

$$\text{RIAN} = -\log ((I + A)/N) \quad (3)$$

Where I represents the sum of all the *iso* 3-OH-FAs, A represents the sum of all the *anteiso* 3-OH-FAs, and N represents the sum of all the *normal* 3-OH-FAs. Only 3-OH-FAs with carbon number range from C_{10} to C_{18} (derived from Gram-negative bacteria) were involved in the calculations. For the calibration of the other 3-OH-FAs based pH proxies, please refer to the Supplementary Information.

Analytical error bars are based on a) 14% of the soil samples being extracted and processed in duplicate or triplicate, e.g. ‘process duplicates’ and the average s.d. being applied to the samples that were not processed in duplicate (for this study, Wang et al., 2016, 2018), or b) triplicate injections e.g. Huguet et al., 2019. Errors for this study were 0.03 for RIAN, 0.29 for RAN_{15} and 0.10 for RAN_{17}. Errors for samples from Huguet et al., (2019) data were 0.006 for RIAN, 0.18 for RAN_{15} and 0.05 for RAN_{17}. The process duplicate errors are somewhat higher than the injection triplicates as would be expected. E.g. the process duplicates include variability from the entire process (extraction, column chromatography) as well as the GC-MS analysis.

2.5.2 Statistical analysis

We used the Canoco and Origin software to conduct the statistical analysis. Canoco 5 software was employed to determine the relationship of the fractional abundance of 3-OH-FAs and 3-OH-FA based indices to environmental factors. Firstly, a detrended correspondence analysis (DCA) was conducted to assess which model (linear or unimodal) was better suited to our dataset based on the length of gradient. If the length of gradient is below 2, a linear model analysis is suggested, while the length of gradient is above 2, a unimodal is suggested. The input data should be centered and standardised for linear model analysis. RDA, a type of linear model analysis, is a multivariate analogue of regression, and can be used to test the relationship of the 3-OH-FAs with one or more explanatory variables (in this case MAAT, pH, MAP and altitude).

Origin 2018 software was applied to test the Pearson correlation coefficient among the 3-OH-FA based indices (and their residuals) and environmental parameters.

2.5.3 Machine Learning

We used a Gaussian process (GP) emulator to make predictions for the environmental temperature and pH (outputs) based on the 3-OH-FA (input) data. A Gaussian process emulator is a machine learning tool that weighs a set of observations with known outputs (calibration data) in order to make predictions. The weights themselves are learned from the calibration. Typically, the GP will give greater weight to closer points in the input space. The training step thus consists of learning the appropriate distance metric on the multi-dimensional input space. A GP is able to handle high-dimensional inputs and find the best combinations, which allows for non-linear dependencies. It also provides quantified uncertainties on the output predictions (for technical details on GP regression refer to Rasmussen and Nickisch (2010) and Rasmussen and Williams (2006)). Our approach in applying GP regression to palaeo-proxy calibration builds on work by Dunkley Jones et al. (2020) who explore in detail the advantages of this approach versus pre-existing methods. Only samples with detectable quantities for all 3-OH-FA homologues (from C_{10} to C_{18}) were analysed for machine learning – resulting in a sample set of 158 (rather than 186). See Section 5 for results and further discussion.

Model code and introduction for the calculation of D_{nearest} values and OPT3MAL MAAT and pH estimates (MATLAB script) are available at https://github.com/carbonatefan/OPT3MAL. MAAT and pH can be predicted using the full global (or a regional) data-set provided here or with any use defined data-set of 3-OH-FA fractional abundances (e.g. future regional or global datasets). The code is also archived in the Zenodo repository https://doi.org/xxxxxxxx.

3. Results

3.1 Composition and Distribution of 3-OH-FAs in soil samples

Data from a total of 186 globally distributed surface soil samples were compiled, including new 112 soil samples analysed in this study (see Section 2.1). The complete results for each sample are provided in the Supplementary Data. The MAAT for the soil sampling sites ranged from -0.4 to 27.0°C (Fig. 2). The soil pH of all soil samples ranged from 3.60 to 9.20 (see Fig. 7). The range of pH is extended by 2 pH units (ca. 1 pH unit at both ends of the spectrum) compared to previously reported data sets (Huguet et al., 2019; Wang et al., 2016). The MAP ranged from 374 to 3313 mm (Supplementary Data).

The molecular fingerprint of 3-OH-FAs in soil samples is akin to that derived from the LPS component of the outer membrane of Gram-negative bacteria (Klok et al., 1988; Lee et al., 2004; Tyagi et al., 2015; Wakeham et al., 2003; Wang et al., 2018). 3-OH-FAs were present in every soil sample analysed, supporting earlier studies on the widespread occurrence of 3-OH-FAs in widely distributed altitudinal transects (Huguet, et al., 2019, Wang, et al., 2016) and suggesting a ubiquitous distribution of these membrane lipids in soils. Thus we assume that the 3-OH-FAs measured in the soils originate from the soil dwelling consortia of Gram-negative bacteria (Wang et al., 2016).

Large differences in the relative concentration of different 3-OH-FA homologues occurred throughout the sample set, displaying distinctive changes in chemical homologue distributions along environmental gradients (Figs. 2&3; S1& S2). The carbon number of 3-OH-FAs ranged from C_{10} to C_{18}, including *iso* C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{18} and *anteiso* C_{11}, C_{13}, C_{15}, C_{17} 3-OH-FAs, with the *normal* C_{12}, C_{14},

C_{16} and C_{18} homologues being typically most abundant (Fig. 3). The summed *normal* 3-OH-FAs are the most abundant, followed by the *iso* 3-OH-FAs, then the *anteiso* 3-OH-FAs. Observations apparent from the chromatograms are the visible differences in distribution in the dominant 3-OH-FA homologue, and the relative abundance of the *normal* vs *iso* and *anteiso* isomers in the different soil samples (Figs. 2&3; S1& S2). Especially apparent is the relative increase in the *anteiso* isomers of the C_{15} and C_{17} homologues in soil samples with colder MAATs (Figs. 2 & S2).

The dominant compound in the global soil samples is the *normal* C_{14} (155 out of 186). In the other samples, the dominant compound is either the *normal* C_{12}, C_{16}, C_{18} or *iso* C_{17}. Similar variations in the predominant compounds were reported in soils by Wang et al. (2016), Huguet et al. (2019) and Véquaud et al. (2020), and in snow pit samples reported by Tyagi et al. (2016). Laboratory culture experiments show that the dominant compounds varied among C_{10}, C_{12}, C_{14}, C_{16} within different Gram-negative genera and species (Goossens et al., 1986; Hedrick et al., 2009; Oyaizu and Komagata, 1983). For example, species of Gammaproteobacteria such as *Pseudomonas* appear to produce mainly even carbon numbered 3-OH-FAs, particularly C_{10}, C_{12} and C_{14} (Humphreys et al., 1972; Ikemoto et al., 1978; Oyaizu and Komagata, 1983; Wilkinson et al., 1973; Wollenweber et al., 1984). A large number of species in the phylum Bacteroidetes seem to have a dominance of C_{15}, C_{16} and C_{17}, compounds not commonly identified in Gammaproteobacteria (Bernardet et al., 1996; Lee et al., 2007; Miyagawa et al., 1979; Wollenweber et al., 1980). Thus the changes of the dominant compound in soil samples (and regional differences in $RAN_{15}/_{17}$ and RIAN calibration slopes and intercepts) may be due to the variation of Gram-negative bacterial community composition. However, we found no systematic variation of the predominant compound with changes in environmental parameters. Future study on the Gram-negative bacteria community composition of soils using genomic methods in representative soil samples will give insights into this. Furthermore, a comprehensive evaluation of the 3-OH-FA compounds produced by a wide diversity of Gram-negative bacteria is required to identify the main producers of 3-OH-FAs in different environments as previous research focuses on more readily culturable species of Gammaproteobacteria, and reports on the 3-OH-FA composition for phyla such as Acidobacteria, Chloroflexi, Planctomycetes and Verrucomicrobia appear to be much more limited.

3.2 Correlation of 3-OH-FA based indices and environmental proxies

Below we explore correlations of previously published 3-OH-FA based proxies to environmental parameters in the new global soil compilation dataset. RAN_{15} ranged from 0.54 to 10.18, RAN_{17} ranged from 0.26 to 4.75. Within the MAAT range of this study (-0.4 to 27°C), both the RAN_{15} and RAN_{17} showed negative linear correlations with MAAT (r = -0.69, p<0.001 and r = -0.64, p<0.001, respectively) (Figs. 4 and 5). The 3-OH-FAs based pH proxies, including the Branching Ratio, RIAN, Branched Index and RIN, were calculated for all the soil samples. Here, in the main text, we focus on the RIAN proxy but we present the results of the other pH proxies in the Supplementary Information. The RIAN index ranges from 0.11 to 0.98 with soil pH ranging between 3.60 and 9.20 and shows a negative linear correlation with the soil pH (r = -0.81, p < 0.001) (Figs. 4 and 5).

Statistical analyses were performed using Canoco software to explore the impacts of environmental parameters on the distribution of 3-OH-FAs and 3-OH-FAs based indices (See Supplementary Section 2). The DCA analysis revealed that a linear model was more appropriate for our dataset as the length of gradient is less than 2, then redundancy analysis (RDA) was performed. The RDA results confirm that soil pH and MAAT are the dominant controls on the distribution of 3-OH-FAs, while the other two environmental parameters, MAP and altitude, show insignificant effects on the distribution of 3-OH-FAs (Table S1 and Fig. S4A). Soil pH explains 24% of the variation of the 3-OH-FAs distribution and MAAT explains 5.8% (Table S1). Soil pH, MAAT and MAP are the dominant controls on the 3-OH-FA based indices (Table S2 and Fig. S4B). Soil pH explains 45.6% variation of 3-OH-FA based indices and MAAT explains 12.4% (Table S2). Further exploration of the data using machine learning was conducted and is discussed in section 5.

4. Discussion

4.1 Effect of temperature on the distribution of 3-OH-FAs

In our global soil compilation, RAN_{15} and RAN_{17} vary from 0.54 to 10.18 and 0.26 to 4.75 respectively, covering greater cumulative ranges than reported previously for initial altitudinal transect studies (Huguet et al., 2019; Wang et al., 2016; Supplementary Datasheet). In our global calibration, RAN_{15} shows a significant linear relationship with MAAT ranging from -0.4 to 27.0 °C ($r = -0.69$, $p < 0.001$; Fig. 4), RAN_{17} shows a linear correlation with MAAT as well but the correlation coefficient is relatively lower ($r = -0.64$, $p < 0.001$; Fig. 4).

Based on the global soil calibration, the updated MAAT equations based on RAN_{15} and RAN_{17} are (Fig. 5):

$$\text{MAAT} = 36.29 - 5.88 \times \text{RAN}_{15}\ (n = 186,\ R^2 = 0.47,\ p<0.001,\ \text{RMSE} = 4.9\ °\text{C}) \tag{4}$$

$$\text{MAAT} = 37.68 - 14.49 \times \text{RAN}_{17}\ (n = 185,\ R^2 = 0.39,\ p<0.001,\ \text{RMSE} = 5.2\ °\text{C}) \tag{5}$$

The above equations show that MAAT has a significant effect on the distribution of C_{15} and C_{17} 3-OH-FAs in the globally distributed soil samples. Both RAN_{15} and RAN_{17} increased with decreasing temperature. This is supported by the general principle of membrane adaptation to temperature, such that bacteria increase the proportion of *anteiso* 3-OH-FAs (increasing the RAN indices) with decreasing temperature in order to maintain membrane fluidity (see inset boxes in Fig. 2). *Anteiso* fatty acids have a lower melting point than *normal* and *iso* fatty acids (Kaneda, 1991; Suutari and Laakso, 1994). Specifically, Kaneda (1991) found that the melting point of the *a*-C_{15} (23.0°C) and *a*-C_{17} (36.8°C) fatty acids were 29.5°C and 24.5°C lower than the melting points of the *n*-C_{15} (52.5°C) and *n*-C_{17} (61.3°C) fatty acids, respectively. Phase transition temperature is even more closely related to membrane fluidity than the average melting temperature of compounds (Kaneda, 1991) and is defined as the temperature required to induce a change in the lipid physical state from the ordered gel phase, where the hydrocarbon chains are fully extended and closely packed, to the disordered liquid crystalline phase, where the hydrocarbon chains are randomly oriented and fluid. Kaneda (1991) found the phase transition temperature for the *a*-C_{15} (-16.5°C) and *a*-C_{17} (7.6°C) were 50.7°C and 41.2°C lower than the equivalent points for the *n*-C_{15} (34.2°C) and *n*-C_{17} (48.8°C). Furthermore, the *anteiso*-positioned fatty acids have a greater disturbance of the packing order of the hydrocarbon chains (Russell, 1995). All of these changes may contribute to maintaining permeability and a liquid crystalline phase of the plasma membrane at different environmental temperatures (Koga, 2012; Siliakus et al., 2017).

It is worth noting that, as well as having a slightly higher R^2 value (and lower RMSE), the RAN_{15} index undergoes a greater absolute change in index value (0.54 to 10.18) compared to RAN_{17} (0.26 to 4.75). This indicates a fundamentally higher amplitude response in the distribution of the C_{15} 3-OH-FA homologues compared to the C_{17} 3-OH-FA homologues along our global MAAT gradient and is illustrated by comparing Figs. 2 and S2. E.g. the proportional increase in relative abundance of *a*-C_{15} vs *n*-C_{15} produced at colder temperatures is ca. double that of the increase in *a*-C_{17} vs *n*-C_{17}. This may be due to the larger variation range and relatively higher abundances of *a*-C_{15} 3-OH-FA in our global soil samples (Fig. S3). This apparently greater physio-chemical response of the C_{15} 3-OH-FA homologues would appear to recommend RAN_{15} as a potentially better palaeo-temperature proxy over RAN_{17}. Moreover, the residuals of RAN_{15} showed no correlation with pH or precipitation which shows the residuals or RAN_{15} are truly random (Supplementary Fig. S9). However, we note that RAN_{15} has relatively more scatter than the RAN_{17} proxy when MAAT is below 10 ℃, possibly indicating that RAN_{17} may be more suitable for low temperature reconstructions. Further study including genomic analyses, insights to bacterial producer populations and culture experiments are required to confirm this.

We note that at a global scale, the relationship between RAN_{15}/RAN_{17} and MAAT contains significant scatter, likely highlighting how other environmental parameters, bacterial biogeography and physical soil effects may affect the variation of the RAN_{15}/RAN_{17} proxies. For instance, we take the recent 30-year average air temperature as representative of the soil temperature, which may be not accurate. This is due to the inherently heterogenous nature of soils, whereby near surface soil conditions and temperatures which bacteria experience may be offset from the boundary layer MAAT estimated from interpolating weather station data. This offset between soil and air temperatures is also not constant, varying with changes of vegetation type, vegetation coverage, soil moisture and texture, etc. (Chudinova et al., 2006; Wang et al., 2020).. Furthermore, the weak anticorrelation between MAAT and soil pH (r = -0.34, $p < 0.05$;

Fig. 4) may add scatter to the correlation between 3-OH-FA based temperature proxies and MAAT. However, we notice that the correlation coefficients of 3-OH-FA based temperature proxies with pH ($r = 0.15$, $p < 0.05$ and 0.34, $p < 0.05$, respectively) are much lower than those with the MAAT ($r = -0.69$, p< 0.001 and $r = -0.64$, p< 0.001, respectively; Fig. 4). As discussed in the next section, pH is the dominant environmental control on bacterial biogeographies at regional scales (Griffiths et al., 2011). Shifting bacterial compositions may in turn affect the distribution of 3-OH-FAs in soils, as some bacterial taxa with distinctive 3-OH-FA signatures may dominate in a particular region (Goossens et al., 1986; Hedrick et al., 2009; Oyaizu and Komagata, 1983). We note that a recent re-evaluation of GDGT based temperature and pH proxies in global soils shows that soil type may bias MAAT and pH estimates (Davtian et al., 2016) and that vegetation cover in the sample site may also influence the community structure of the Gram-negative bacteria, as Gram-negative bacteria prefer to utilise more plant-derived C sources that are relatively labile (Fanin et al., 2019). The combinations of environmental factors driving bacterial community structure are complex, and the major determinants may be region and taxa-specific (Oliverio et al., 2017; Singh et al., 2013; Yao et al., 2017). Both these effects will require further study in the development of 3-OH-FA based proxies.

Because of the scatter in the global calibration we investigated the correlations for RAN_{15} and RAN_{17} for discrete regions (Fig 6). We find that for the RAN_{15} and RAN_{17} temperature proxies there are significant differences in slope and intercept of the linear corrections at regional scales. The coefficient of determinations between RAN_{15} and MAAT varied from 0.30 to 0.79 in regional calibrations (Fig. 6A, Supplementary Data). The strongest correlation between RAN_{15} and MAAT were observed in Mount Rungwe ($R^2 = 0.79$, $p < 0.001$), strong correlation was also found in Northern America ($R^2 = 0.74$,

$p < 0.001$). Moderate correlations were observed in Mount Shennongjia, Mount Majella and Africa & Europe ($R^2 = 0.50$, $p < 0.001$, $R^2 = 0.44$, $p < 0.05$, $R^2 = 0.49$, $p < 0.001$, respectively). The samples from China had the most scatter ($R^2 = 0.30$, $p < 0.001$). For the RAN_{17} proxy, the coefficient of determinations varied from 0.28 to 0.74 in regional calibrations, except the Mount Majella and Africa & Europe where no significant correlations were found (Fig. 6B, Supplementary Data). The strongest correlation between RAN_{17} and MAAT were observed in Northern America ($R^2 = 0.74$, $p < 0.001$), moderate to weak correlations were observed in Mount Rungwe, Mount Shennongjia and China ($R^2 = 0.48$, $p < 0.001$, $R^2 = 0.52$, $p < 0.001$, $R^2 = 0.28$, $p < 0.001$, respectively). The lack of correlation in Africa & Europe may be due to the RAN_{17} being relatively less sensitive to temperature changes when MAAT is above 20 ℃; a similar feature is also found in brGDGT based MBT'_{5ME} proxy (De Jonge et al., 2014; Naafs et al., 2017). Thus local or regional calibrations are likely preferable (at this stage of 3-OH-FA proxy development) for application to palaeoclimate archives. For example, applying our global linear calibration to the only available 3-OH-FA based palaeo-record from a Chinese speleothem would result in a large overestimation of temperature (compared to the existing local calibration used by Wang et al., 2018).

4.2 Effect of pH on the distribution of 3-OH-FAs

pH is an important environmental parameter which affects the soil bacterial community structure and diversity (Bååth and Anderson, 2003; Delgado-Baquerizo et al., 2018; Griffiths et al., 2011; Lauber et al., 2009; Rousk et al., 2010). Acidic soils have commonly been found to support a lower diversity of bacteria, with a dominance of low pH specialists such as Acidobacteria in soils with a pH below 5 (Cho et al., 2019;

Jones et al., 2009; Lauber et al., 2009; Zhang et al., 2015). More importantly, pH can influence membrane fluidity, and lead to the changes in membrane lipids (Wang et al., 2016). For example, culture experiments on a strain of Gram-negative bacteria showed increased/decreased relative abundance of branched-chain fatty acids in higher pH/lower pH (Giotis et al., 2007). Our results from the global soil samples indicate that the proportion of Gram-negative bacteria derived branched 3-OH-FA homologues is affected by soil pH. These are illustrated by the correlations between the 3-OH-FA based indices and soil pH (Fig. 7 & S5).

In accordance with previous findings (Wang et al., 2016), the Branching Ratio showed an exponential relationship with soil pH (Fig. S5A). Since soil pH has a logarithmic relationship with the concentration of H^+, this suggests the variation of the Branching Ratio may be influenced by the concentration of soil H^+. Lower pH corresponds to a larger concentration of H^+, and thus steeper proton gradients across bacterial cell membranes. We suggest that the observation of a decreasing Branching Ratio at lower pH reflects chemiosmotic coupling, i.e., the production of fewer branched homologues, producing a less fluid or more impermeable membrane to counteract steeper proton gradients (Denich et al., 2003; McElhaney et al., 1973; Russell and Fukunaga, 1990). The existence and maintenance of a proton gradient over bacterial cell membranes is vital for the energy supply of a cell (Mitchell, 1966) and involves the trapping of proton-conducting water molecules in the lipid core of the membranes (Nagle and Morowitz, 1978; Wikström et al., 2015). Given the logarithmic relationship between pH and the Branching Ratio (Fig. S5A) and the definition of pH as the negative logarithm of the proton concentration, it is possible to obtain a linear relationship between the two by using the previously defined RIAN index:

$$\mathrm{RIAN} = -\log(\text{Branching Ratio}) \qquad (6)$$

The linear relationship between the RIAN and global soil pH is best fit by:

$$\text{RIAN} = 1.12 - 0.11 \times \text{pH} \ (n = 186, R^2 = 0.66, p<0.001, \text{RMSE} = 0.10) \quad (7)$$

This relationship between the RIAN index and soil pH is similar to what was previously reported (Huguet et al., 2019; Wang et al., 2016). The global calibration is consistent with previous local/regional calibration suggesting a wider applicability of the proxy in global soil samples. This is consistent with previous research on bacterial brGDGTs indicating that soil pH has a significant impact on the global soil brGDGTs distribution (Peterse et al., 2012; Weijers et al., 2007; Yang et al., 2012).

Based on the above correlation, we propose new global transfer equation for soil pH calibration:

$$\text{pH} = 10.18 - 9.09 \times \text{RIAN} \ (n= 186, R^2 = 0.66, p<0.001, \text{RMSE}= 0.78) \quad (8)$$

The pH proxies developed by Wang et al. (2016) were only based on 26 soil samples along an altitudinal transect of Mt. Shennongjia. In this paper, we have used 186 globally located soil samples, which greatly extended sample size and locations. Moreover, the pH range in our updated calibration ranges from 3.60 to 9.20, significantly extending the pH range compared to previous calibrations (Fig. 7), further confirming the applicability of RIAN (and other 3-OH-FA based indices) as a novel pH proxy.

Regional calibrations were also conducted to test the consistency of global and regional calibrations (Fig. 8). The results showed that samples from China, Northern America and Mt. Shennongjia shared identical slopes and intercepts with the global calibration. But samples from Mount Rungwe showed no significant correlation between RIAN and soil pH, this may be due to the narrowed pH changes in that region (Huguet et al., 2019). Samples from Africa & Europe also showed no significant correlations. Interestingly, a reversed correlation was found in Mount Majella (R^2 = 0.65, $p < 0.05$) which is completely different from the other regional and global calibrations in this study (Fig. 8 and Supplementary Data). Thus regional calibration may be more appropriate in some site specific settings.

4.3 Effect of precipitation on the distribution of 3-OH-FAs

Mean Annual Precipitation (MAP) varies from 374 to 3313 mm in our global soil compilation, covering samples from semi-arid to tropical zones. Despite a generally observed relationship between effective precipitation and pH in global soils (Slessarev et al., 2016; Yang et al., 2014), MAP for our soil samples shows low correlation with pH (r = -0.47, $p < 0.001$; Figs. 4 and S6). In our global soil dataset, we found weak correlations between the MAP and 3-OH-FAs based proxies (Figs. 4 and S7). Weak correlation between MAP and 3-OH-FAs were also found in Mt. Majella (Huguet et al., 2019), but no correlation was found in the samples from our original study on Mt. Shennongjia (Wang et al., 2016). Notably, we found no correlations between MAP and soil pH in Mt. Shennongjia (r = -0.27, $p > 0.05$) (Wang et al., 2016) and weak correlation in the global soil dataset (r= -0.47, $p < 0.001$; Fig. 4 and S6). The weak correlation between the MAP and 3-OH-FAs based proxies in the global soil samples suggests that MAP may affect the community composition of Gram-negative bacteria, and thus the distribution of 3-OH-FAs, although this appears to be a secondary effect compared to pH. Manipulative experiments in different steppes along a precipitation gradient in northern China showed that precipitation regime controls microbial activity and biomass, possibly by regulating soil moisture and substrate availability (Liu et al., 2016). Metagenomics of global topsoil samples show that bacterial global niche differentiation is associated with contrasting diversity responses to precipitation and soil pH (Bahram et al., 2018).

We found no linear correlations between precipitation and RAN_{15}/RAN_{17}, suggesting that precipitation likely does not affect the values of our 3-OH-FA based temperature proxies (Fig. S8). This independence of the 3-OH-FA based temperature proxies may be because only the *anteiso* and *normal* C_{15} or C_{17} homologues are utilised in these proxies. In comparison GDGT analysis of soil transects from the US highlights a substantial increase in the offset between measured MAAT and MBT/CBT-based MAAT below an annual precipitation of 700–800 mm yr^{-1}, implying an impact of precipitation amount on MBT/CBT-based temperature reconstruction (possibly related to soil aeration and pH) (Dirghangi et al., 2013). The study of bacterial GDGTs (brGDGTs) from global surface soils samples shows the relative abundance of some brGDGTs, but not all correlate with MAP (De Jonge et al., 2014; Peterse et al., 2012; Weijers et al., 2007). Our observation that MAP shows some impact on 3-OH-FA based pH proxies (Branching Ratio: $r = -0.51$, $p < 0.001$; RIAN: $r = 0.49$, $p < 0.001$; Branched Index: $r = -0.51$, $p < 0.001$; RIN: $r = -0.51$, $p < 0.001$; Figs. 4 and S7), but no impact on 3-OH-FA based temperature proxies may reflect changes in bacterial community composition and diversity between different precipitation regimes. Our pH indices, including RIAN, incorporate up to 21 different 3-OH-FA homologues and thus are more likely to reflect an aggregate change of 3-OH-FAs resulting from any differences in Gram-negative bacteria community between higher and lower precipitation regime soils. Whereas the more limited use of only 2 different homologues in the RAN_{15} and RAN_{17} indices must be inherently more specific to particular classes of Gram-negative bacteria.

4.4 Comparison with GDGT data

GDGT based proxies are well established for palaeoenvironmental reconstructions. In our new global dataset, MBT'$_{5ME}$-MAAT shows linear correlations with RAN$_{15}$ and RAN$_{17}$ proxies, but the correlation coefficient is relatively low (r = -0.59, p< 0.001 and r = -0.42, p < 0.001, respectively; Fig. 9). The relatively low correlation between the 3-OH-FA based RAN$_{15}$/ RAN$_{17}$ and GDGT based MBT'$_{5ME}$ may be partly due to the intrinsic relatively lower correlations between the 3-OH-FA based temperature proxies and MAAT, or due to different responses of Gram-negative bacteria and brGDGT-producing bacteria to other environmental factors (Huguet et al., 2019). Interestingly, the MBT'$_{5ME}$ data which are available for the samples in this study also showed different slopes and intercepts in global and regional calibrations (Fig. S10). This may add the scatter to the correlation of 3-OH-FA based RAN$_{15}$/ RAN$_{17}$ and GDGT based MBT'$_{5ME}$ indices. The cyclisation ratio of branched tetraethers (CBT) is an established pH proxy, first proposed by Weijers et al. (2007). We find that our 3-OH-FA based pH proxies show significant correlation with CBT (Branching Ratio: r = -0.72, p < 0.001; RIAN: r = 0.77, p < 0.001; Branched Index: r = -0.75, p < 0.001; RIN: r = -0.70, p < 0.001; Fig. 10), further confirming that these bacterial derived membrane lipids are both controlled by soil pH.

5. Further examination and calibration of relationships between 3-OH-FAs distributions with MAAT and soil pH using machine learning

The linear regression based indices above are defined by empirically linking environmental controls with a presumed, but unproven, physiological mechanism of membrane adaptation by the soil bacteria producing the 3-OH-FAs, i.e. an increase in the percentage of *anteiso* isomers with decreasing MAATs, and an increase in the percentage of branched isomers with increasing pH. There are a number of options to improve predictions based on linear regressions using machine learning techniques such as artificial neural networks, random forests and Gaussian Process emulators. These flexible, non-parametric models are all based on the idea of training a predictor by fitting a set of coefficients in a sufficiently complex, often multi-layer, model in order to minimise residuals on the calibration data set (Fig. 11). The objective is to search, agnostically, among a large space of smoothly varying functions of 3-OH-FA compositions for those functions which adequately describe temperature and pH variability. This, essentially, is a way of combining information from all calibration data points, not just the nearest neighbours, assigning different weights to different calibration points depending on their utility in predicting the temperature or pH at the input of interest.

GP regressions were applied to both the full input range of 3-OH-FA homologues and to the subset of compounds, which have previously demonstrated the clearest sensitivity to MAAT (the *i*-C_{15}, *a*-C_{15}, *n*-C_{15}, *i*-C_{17}, *a*-C_{17}, *n*-C_{17} isomers as utilized in the RAN_{15} and RAN_{17} indices). 90% of data points were used for calibration. Validation and performance were tested using the remaining 10% of data points, repeating the process 10 times with a random choice of which data fall into the calibration (90%) and validation (10%) groups.

By using all data the GP regression approach gives superior results compared to the simple linear regressions (Section 4.1) for both temperature (Fig. 12A: RSME = 3.5°C; $R^2 = 0.66$) and pH (Fig. 12B: RSME = 0.76 pH units; $R^2 = 0.63$). GP regression provides a confidence interval on the prediction (see Fig. 12), which can be used to test the self-consistency of the prediction: for example, we expect that the true value should fall into the 90% confidence interval 90% of the time. When using all of the isomers from C_{10} to C_{18} the validation value is contained within the 5 to 95% confidence interval of GP predictions only 80% of the time for temperature and 77% of the time for pH, rather than the expected 90%. This indicates the possibility of a systematic bias, perhaps because the large dimensionality of the input data means that there is often no calibration data sufficiently nearby (in parameter space) and the model is forced to extrapolate instead of interpolating.

GP regression using just the C_{15} and C_{17} *iso, anteiso* and *normal* isomers yields superior results compared to the simple linear regressions for both temperature (Fig. 12B: RSME = 3.9°C; $R^2 = 0.61$) and pH (Fig. 12C: RSME = 0.64 pH units; $R^2 = 0.74$). Moreover, unlike GP regression based on all isomers from C_{10} to C_{18}, when using only the C_{15} and C_{17} *iso, anteiso* and *normal* isomers, the validation values were contained within the 5 to 95% interval 93% of the time for temperature and 91% of the time for pH, statistically consistent with the expected 90%.

In addition to naturally yielding confidence limits on predictions, GP regression has the benefit of providing estimates of the relative importance of the inputs in predicting the output. By examining the learned GP kernel, we find that *anteiso* and *normal* C_{15} and *iso* and *anteiso* C_{17} play significant roles in temperature prediction, while *iso*, *anteiso* and *normal* C_{17} isomers and *anteiso* C_{15} play comparable roles in pH prediction.

In Fig.12 we illustrate the GP regressions for all the available global soils samples. But it should be noted that our code can be run on regional (or user defined) data-sets.

This may be desirable for specific applications, due to regional differences observed in the empirical linear regressions.

The superior performance of machine-learning on the sufficiently complex, multi-dimensional data set is not unexpected. It is able to effectively consider a much broader range of possible dependencies than those analysed in Section 4.1, including possible non-linear behaviour of the output as a function of inputs. Therefore, in the absence of a robust physical model, machine learning yields a preferred approach to making accurate predictions. It does suffer from an inability to extrapolate to input data regimes that are far from the available calibration data, though caution is always warranted for such extrapolation in the absence of a robust model (and if such a model does exist, it can be readily incorporated into the machine learning tools). The machine learning predictions are also challenging to translate into a human-readable model, though at least in the case of a GP emulator, the learned metric on the parameter space can be useful for interpreting which input parameters play the most significant roles in determining temperature and pH outputs. These limitations are generally more than compensated by increased prediction accuracy (e.g., Dunkley Jones et al., 2020), and by the availability of prediction uncertainties along with best-guess estimators.

6. Conclusion

Based on an extensive new global compilation (*n* = 186), we tested the performance of 3-OH-FA based proxies for MAAT and pH in global soil samples. We find that the 3-OH-FA based temperature proxies RAN_{15} and RAN_{17} show significant correlations with MAAT and the 3-OH-FA based pH proxy RIAN shows a significant correlation with soil pH. Machine learning based GP emulator models confirm that environmental signals are recorded by 3-OH-FAs. Moreover, the GP regressions give higher R^2 values and reduce RSME; they also provide confidence intervals on the predictons. We recommend that workers explore and apply both the simple linear regressions and machine-learning based models to palaeoclimate data-sets during this nascent stage of 3-OH-FA development for palaeoclimate. Moreover, we find that for our 3-OH-FA based proxies there are significant differences in slope and intercept of the linear corrections at regional scales. Thus local or regional calibrations are likely preferable at this stage of 3-OH-FA proxy development for application to specific palaeoclimate archives. While this manuscript was under review, Véquaud et al. (2020) applied other machine learning tools, including random forests, to this problem, achieving broadly similar results. Our empirical, global scale, compilation of 3-OH-FA based proxies builds on the promise of initial altitudinal calibrations (and a Holocene stalagmite climate reconstruction) and has wide implications for palaeoclimatic and environmental studies. Gram-negative bacteria are ubiquitous in natural environments, and 3-OH-FA based proxies are now developed for both terrestrial and marine settings. These compounds are easy to extract using a simple acid digestion and to analyse using GC–MS and GC–FID systems. This makes it possible to obtain high-resolution palaeo-records using a relatively small sample mass. We hope this investigation open up new avenues of research on 3-OH-FAs, including culture studies and DNA sequencing to constrain 3-OH-FA bacterial precursors, to investigate the underlying response mechanisms to environmental parameters, and applications to an array of palaeoclimatic archives (e.g., palaeosols, lakes, speleothems, marine records).

Acknowledgements

We thank Stephan Mantel and Gerard Heuvelink at ISRIC World Soil Information (Wageningen, Netherlands), for sampling and provision of archived soil samples. We thank Dr. Linfeng Gong, Mr. Shijin Zhao, Dr. Xinyue Dang and Dr. Jiantao Xue from China University of Geosciences (Wuhan) for their help during the field sampling in China. Huub Zwart, Vinothan Sivapalan, Caroline Bendle, George Bendle and Ryan Bendle are thanked for their participation in the USA field sampling campaign and Jessica Conway (University of Illinois Urbana Champaign), Chad Broyles and colleagues (International Ocean Discovery Programme - Texas A&M University) and Dr Amelia Shevenell (University of Southern Florida) and David Tubbs and Eimear Orgill (University of Birmingham) for help with equipment, logistics, shipping, sample handling and soil licenses. Thanks to Dr. Xing Xiang (Shangrao Normal University) for helping with the global map plotting. This work was supported by funding bodies in China and the UK, namely the National Natural Science Foundation of China (Grant Nos. 41830319, 41807419, 41821001, U20A2094, 91951208, 41773135), the Key R&D Project of Ministry of Science and Technology (grant no. 2016YFA0601100), the 111 project (National Bureau of Foreign Experts and the Ministry of Education of China; Grant No. BP20004), the Fundamental Research Funds for the National Universities, China University of Geosciences, Wuhan (Grant No. CUGL170815), the Leverhulme Trust (RPG-2018-110), the Birmingham-Illinois Partnership Fund, the UoB School of GEES Pump Priming Fund. We thank the China Scholarship Council (CSC) (Grant No. 201306410031) for supporting Canfa Wang’s studies at the University of Birmingham and the UK’s Natural Environmental Research Council (NERC) and the NERC CENTA-DTP (Central England NERC Training Alliance – Doctorial Training Partnership) for funding Alice Hardman’s PhD studies and research training grant (NE/L002493/1) and for an independent research fellowship to Sarah Greene (NE/L011050/1). Ilya Mandel is a recipient of the Australian Research Council Future Fellowship FT190100574.

Table and figure captions

Fig. 1. Molecular structure of *normal*, *iso* and *anteiso* C_{15} 3-OH-FAs.

Fig. 2. Maps showing the locations of soil samples used in this study. The colour spectrum of the dots illustrates the mean annual air temperature (MAAT) of each sampling site. A) Global overview map showing the locations of soil samples, with examples of C_{15} 3-OH-FAs distributions in three soils, with markedly different MAATs, from Greenland (Sample GL005-2), China (Sample TJ-3), and Ghana (Sample GH002-02). The peaks in green in the inset chromatograph represent *normal* 3-OH-FA, the peaks in blue represent *anteiso* 3-OH-FA, the peaks in orange represent *iso* 3-OH-FA. B) Map showing the locations of soil samples from the eastern USA. C) Map showing the locations of soil samples from Southern Africa. D) Map showing the locations of soil samples from eastern China.

Fig. 3. Examples of distribution of 3-OH-FAs in soils from different mean annual air temperature (MAAT) and pH. The peaks in green represent *normal* 3-OH-FAs, the peaks in blue represent *anteiso* 3-OH-FAs, the peaks in red represent *iso* 3-OH-FAs.

Fig. 4. Heat map showing the Pearson correlation coefficients of 3-OH-FA based proxies and environmental parameters.

Fig. 5. Scatter-plots showing the relationship of 3-OH-FA based indices and mean annual air temperature (MAAT) and residuals. A) Global RAN_{15} vs MAAT; B) Global RAN_{17} vs MAAT. 95% observational and functional bounds are also shown. These represent a 95% probability that: a) a new observation and; b) the true function without observational errors will lie within the respective bounds.

Fig. 6. Scatter-plot showing the regional data points and regional linear calibrations for 3-OH-FA based proxies vs MAAT (with the global linear regression line for comparison). Regression lines are not shown for regions where correlation is not significant ($p > 0.05$). A) RAN_{15} vs MAAT; B) RAN_{17} vs MAAT.

Fig. 7. Scatter-plot showing the global relationship between 3-OH-FA based RIAN proxy vs soil pH and residuals. 95% observational and functional bounds are also shown. These represent a 95% probability that: a) a new observation and; b) the true function without observational errors will lie within the respective bounds.

Fig. 8 Scatter-plots showing the regional and global calibrations between RIAN and soil pH. Regression lines are not shown for regions where correlation is not significant ($p > 0.05$).

Fig. 9. Comparison of 3-OH-FA based temperature proxies with GDGT based temperature proxies. A) RAN_{15} and MBT'_{5ME}-MAAT; B) RAN_{17} and MBT'_{5ME}-MAAT.

95% observational and functional bounds are also shown. These represent a 95% probability that: a) a new observation and; b) the true function without observational errors will lie within the respective bounds.

Fig. 10. Comparison of 3-OH-FA based pH proxies with GDGT based pH proxies. A) The linear correlation between Branching Ratio and CBT. B) The linear correlation between RIAN and CBT. C) The linear correlation between Branched Index and CBT. D) The linear correlation between RIN and CBT.

Fig. 11. A) Schematic of a Gaussian Process emulator (showing just 1 dimension of many); B) GP regression temperature predictions based on 3-OH-FA distributions vs true temperature in our new global soil data-set (see Fig. 12). The GP reduces the root mean square uncertainty on predictions compared to empirical regressions.

Fig. 12. Gaussian Process (GP) regression approach using all the 3-OH-FAs isomers (C_{10} -C_{18}) and just the C_{15} and C_{17} *iso, anteiso* and *normal* isomers for both temperature and pH. A) The GP regression temperature predictor as a function of the true temperature using all the isomers from C_{10} to C_{18}. B) The GP regression pH predictor as a function of the true pH using all the isomers from C_{10} to C_{18}. C) The GP regression temperature predictor as a function of the true temperature using just the C_{15} and C_{17} *iso, anteiso* and *normal* isomers. D) The GP regression pH predictor as a function of the true pH using just the C_{15} and C_{17} *iso, anteiso* and *normal* isomers.

References

Bååth, E. and Anderson, T.H. (2003) Comparison of soil fungal/bacterial ratios in a pH gradient using physiological and PLFA-based techniques. Soil Biology and Biochemistry 35, 955-963.

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H., Maltz, M.R., Mundra, S., Olsson, P.A., Pent, M., Polme, S., Sunagawa, S., Ryberg, M., Tedersoo, L. and Bork, P. (2018) Structure and function of the global topsoil microbiome. Nature 560, 233-237.

Bauersachs, T., Rochelmeier, J. and Schwark, L. (2015) Seasonal lake surface water temperature trends reflected by heterocyst glycolipid-based molecular thermometers. Biogeosciences 12, 3741-3751.

Bernardet, J.-F., Segers, P., Vancanneyt, M., Berthe, F., Kersters, K. and Vandamme, P. (1996) Cutting a Gordian knot: emended classification and description of the genus *Flavobacterium*, emended description of the family *Flavobacteriaceae*, and proposal of *Flavobacterium hydatis* nom. nov.(basonym, *Cytophaga aquatilis* Strohl and Tait 1978). International Journal of Systematic and Evolutionary Microbiology 46, 128-148.

Blaga, C.I., Reichart, G.-J., Schouten, S., Lotter, A.F., Werne, J.P., Kosten, S., Mazzeo, N., Lacerot, G. and Sinninghe Damsté, J.S. (2010) Branched glycerol dialkyl glycerol tetraethers in lake sediments: Can they be used as temperature and pH proxies? Organic Geochemistry 41, 1225-1234.

Blyth, A.J., Farrimond, P. and Jones, M. (2006) An optimised method for the extraction and analysis of lipid biomarkers from stalagmites. Organic Geochemistry 37, 882-890.

Brassell, S., Eglinton, G., Marlowe, I., Pflaumann, U. and Sarnthein, M. (1986) Molecular stratigraphy: A new tool for climatic assessment. Nature 320, 129-133.

Cho, H., Tripathi, B.M., Moroenyane, I., Takahashi, K., Kerfahi, D., Dong, K. and Adams, J.M. (2019) Soil pH rather than elevation determines bacterial phylogenetic community assembly on Mt. Norikura. FEMS Microbiology Ecology 95, fiy216.

Chudinova, S.M., Frauenfeld, O.W., Barry, R.G., Zhang, T. and Sorokovikov, V.A. (2006) Relationship between air and soil temperature trends and periodicities in the permafrost regions of Russia. Journal of Geophysical Research: Earth Surface 111.

Davtian, N., Ménot, G., Bard, E., Poulenard, J. and 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. Organic Geochemistry 101, 140-153.

de Bar, M.W., Weiss, G., Rampen, S., Lattaud, J., Bale, N.J., Brummer, G.-J.A., Schulz, H., Rush, D., Kim, J.-H. and Donner, B. (2020) Global temperature calibration of the Long chain Diol Index in marine surface sediments. Organic Geochemistry, 103983.

De Jonge, C., Hopmans, E.C., Stadnitskaia, A., Rijpstra, W.I.C., Hofland, R., Tegelaar, E. and 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. Organic Geochemistry 54, 78-82.

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

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K. and Fierer, N. (2018) A global atlas of the dominant bacteria found in soil. Science 359, 320-325.

Denich, T.J., Beaudette, L.A., Lee, H. and Trevors, J.T. (2003) Effect of selected environmental and physico-chemical factors on bacterial cytoplasmic membranes. Journal of Microbiological Methods 52, 149-182.

Dirghangi, S.S., Pagani, M., Hren, M.T. and Tipple, B.J. (2013) Distribution of glycerol dialkyl glycerol tetraethers in soils from two environmental transects in the USA. Organic Geochemistry 59, 49-60.

Dunkley Jones, T., Eley, Y.L., Thomson, W., Greene, S.E., Mandel, I., Edgar, K. and Bendle, J.A. (2020) OPTiMAL: a new machine learning approach for GDGT-based palaeothermometry. Clim. Past 16, 2599-2617.

Eglinton, G., Hunneman, D.H. and McCormick, A. (1968) Gas chromatographic—mass spectrometric studies of long chain hydroxy acids.—III.The mass spectra of the methyl esters trimethylsilyl ethers of aliphatic hydroxy acids. A facile method of double bond location. Organic Mass Spectrometry 1, 593-611.

Eglinton, T. and Eglinton, G. (2008) Molecular proxies for paleoclimatology. Earth and Planetary Science Letters 275, 1-16.

Fanin, N., Kardol, P., Farrell, M., Nilsson, M.-C., Gundale, M.J. and Wardle, D.A. (2019) The ratio of Gram-positive to Gram-negative bacterial PLFA markers as an indicator of carbon availability in organic soils. Soil Biology and Biochemistry 128, 111-114.

Giotis, E.S., McDowell, D.A., Blair, I.S. and Wilkinson, B.J. (2007) Role of branched-chain fatty acids in pH stress tolerance in *Listeria monocytogenes*. Applied and Environmental Microbiology 73, 997-1001.

Goossens, H., Irene, W., Rijpstra, C., Düren, R., De Leeuw, J. and Schenck, P. (1986) Bacterial contribution to sedimentary organic matter; a comparative study of lipid moieties in bacteria and Recent sediments. Organic Geochemistry 10, 683-696.

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M. and Whiteley, A.S. (2011) The bacterial biogeography of British soils. Environmental microbiology 13, 1642-1654.

Haug, G.H., Ganopolski, A., Sigman, D.M., Rosell-Mele, A., Swann, G.E., Tiedemann, R., Jaccard, S.L., Bollmann, J., Maslin, M.A. and Leng, M.J. (2005) North Pacific seasonality and the glaciation of North America 2.7 million years ago. Nature 433, 821-825.

Hedrick, D.B., Peacock, A.D., Lovley, D.R., Woodard, T.L., Nevin, K.P., Long, P.E. and White, D.C. (2009) Polar lipid fatty acids, LPS-hydroxy fatty acids, and respiratory quinones of three Geobacter strains, and variation with electron acceptor. Journal of Industrial Microbiology & Biotechnology 36, 205-209.

Hopmans, E.C., Schouten, S. and Sinninghe Damsté, J.S. (2016) The effect of improved chromatography on GDGT-based palaeoproxies. Organic Geochemistry 93, 1-6.

Huang, X., Cui, J., Pu, Y., Huang, J. and Blyth, A.J. (2008) Identifying "free" and "bound" lipid fractions in stalagmite samples: An example from Heshang Cave, Southern China. Applied Geochemistry 23, 2589-2595.

Huang, X., Meyers, P.A., Jia, C., Zheng, M., Xue, J., Wang, X. and Xie, S. (2013) Paleotemperature variability in central China during the last 13 ka recorded by a novel microbial lipid proxy in the Dajiuhu peat deposit. The Holocene 23, 1123-1129.

Huguet, A., Coffinet, S., Roussel, A., Gayraud, F., Anquetil, C., Bergonzini, L., Bonanomi, G., Williamson, D., Majule, A. and Derenne, S. (2019) Evaluation of 3-hydroxy fatty acids as a pH and temperature proxy in soils from temperate and tropical altitudinal gradients. Organic Geochemistry 129, 1-13.

Humphreys, G.O., Hancock, I.C. and Meadow, P.M. (1972) Synthesis of the hydroxyacids in lipid A of *Pseudomonas aeruginosa*. Microbiology 71, 221-230.

Ikemoto, S., Kuraishi, H., Komagata, K., Azuma, R., Suto, T. and Murooka, H. (1978) Cellular fatty acid composition in *Pseudomonas* species. The Journal of General and Applied Microbiology 24, 199-213.

IPCC (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.

Jenske, R. and Vetter, W. (2008) Gas chromatography/electron-capture negative ion mass spectrometry for the quantitative determination of 2-and 3-hydroxy fatty acids in bovine milk fat. Journal of Agricultural and Food Chemistry 56, 5500-5505.

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R. and Fierer, N. (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. The ISME Journal 3, 442-453.

Kaneda, T. (1991) Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. Microbiology and Molecular Biology Reviews 55, 288-302.

Kawamura, K. and Ishiwatari, R. (1984) Tightly bound aliphatic acids in Lake Biwa sediments: Their origin and stability. Organic Geochemistry 7, 121-126.

Kim, J.-H., Schouten, S., Hopmans, E.C., Donner, B. and Sinninghe Damsté, J.S. (2008) Global sediment core-top calibration of the TEX_{86} paleothermometer in the ocean. Geochimica et Cosmochimica Acta 72, 1154-1173.

Klages, J.P., Salzmann, U., Bickert, T., Hillenbrand, C.D., Gohl, K., Kuhn, G., Bohaty, S.M., Titschack, J., Muller, J., Frederichs, T., Bauersachs, T., Ehrmann, W., van de Flierdt, T., Pereira, P.S., Larter, R.D., Lohmann, G., Niezgodzki, I., Uenzelmann-Neben, G., Zundel, M., Spiegel, C., Mark, C., Chew, D., Francis, J.E., Nehrke, G., Schwarz, F., Smith, J.A., Freudenthal, T., Esper, O., Palike, H., Ronge, T.A., Dziadek, R. and Science Team of Expedition, P.S. (2020) Temperate rainforests near the South Pole during peak Cretaceous warmth. Nature 580, 81-86.

Klok, J., Baas, M., Cox, H.C., de Leeuw, J.W., Rijpstra, W.I.C. and Schenck, P.A. (1988) The mode of occurrence of lipids in a Namibian Shelf diatomaceous ooze with emphasis on the β-hydroxy fatty acids. Organic Geochemistry 12, 75-80.

Koga, Y. (2012) Thermal adaptation of the archaeal and bacterial lipid membranes. Archaea 2012, 789652.

Kumar, G.S., Jagannadham, M.V. and Ray, M.K. (2002) Low-temperature-induced changes in composition and fluidity of lipopolysaccharides in the antarctic psychrotrophic bacterium Pseudomonas syringae. Journal of Bacteriology 184, 6746-6749.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Applied and Environmental Microbiology 75, 5111-5120.

Lee, A.K.Y., Chan, C.K., Fang, M. and Lau, A.P.S. (2004) The 3-hydroxy fatty acids as biomarkers for quantification and characterization of endotoxins and Gram-negative bacteria in atmospheric aerosols in Hong Kong. Atmospheric Environment 38, 6307-6317.

Lee, H.-G., An, D.-S., Im, W.-T., Liu, Q.-M., Na, J.-R., Cho, D.H., Jin, C.W., Lee, S.-T. and Yang, D.-C. (2007) *Chitinophaga ginsengisegetis* sp. nov. and *Chitinophaga ginsengisoli* sp. nov., isolated from soil of a ginseng field in South Korea. International Journal of Systematic and Evolutionary Microbiology 57, 1396-1401.

Lei, Y., Yang, H., Dang, X., Zhao, S. and Xie, S. (2016) Absence of a significant bias towards summer temperature in branched tetraether-based paleothermometer at two soil sites with contrasting temperature seasonality. Organic Geochemistry 94, 83-94.

Liu, W., Allison, S.D., Xia, J., Liu, L. and Wan, S. (2016) Precipitation regime drives warming responses of microbial biomass and activity in temperate steppe soils. Biology and Fertility of Soils 52, 469-477.

Luo, G., Yang, H., Algeo, T.J., Hallmann, C. and Xie, S. (2019) Lipid biomarkers for the reconstruction of deep-time environmental conditions. Earth-Science Reviews 189, 99-124.

McElhaney, R.N., De Gier, J. and Van der Neut-Kok, E. (1973) The effect of alterations in fatty acid composition and cholesterol content on the nonelectrolyte permeability of *Acholeplasma laidlawii* B cells and derived liposomes. Biochimica et Biophysica Acta (BBA)-Biomembranes 298, 500-512.

Meyers, P.A. (1997) Organic geochemical proxies of paleoceanographic, paleolimnologic, and paleoclimatic processes. Organic Geochemistry 27, 213-250.

Mielniczuk, Z., Mielniczuk, E. and Larsson, L. (1993) Gas chromatography-mass spectrometry methods for analysis of 2- and 3-hydroxylated fatty acids: Application for endotoxin measurement. Journal of Microbiological Methods 17, 91-102.

Mitchell, P. (1966) Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. Biological Reviews 41, 445-501.

Miyagawa, E., Azuma, R. and Suto, T. (1979) Cellular fatty acid composition in Gram-negative obligately anaerobic rods. The Journal of General and Applied Microbiology 25, 41-51.

Naafs, B.D.A., Gallego-Sala, A.V., Inglis, G.N. and Pancost, R.D. (2017) Refining the global branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration. Organic Geochemistry 106, 48-56.

Naafs, B.D.A., Hefter, J. and Stein, R. (2012) Application of the long chain diol index (LDI) paleothermometer to the early Pleistocene (MIS 96). Organic Geochemistry 49, 83-85.

Nagle, J. and Morowitz, H. (1978) Molecular mechanisms for proton transport in membranes. Proceedings of the National Academy of Sciences 75, 298-302.

Oliverio, A.M., Bradford, M.A. and Fierer, N. (2017) Identifying the microbial taxa that consistently respond to soil warming across time and space. Global Change Biology 23, 2117-2129.

Oyaizu, H. and Komagata, K. (1983) Grouping of Pseudomonas species on the basis of cellular fatty acid composition and the quinone system with special reference to the existence of 3-hydroxy fatty acids. The Journal of General and Applied Microbiology 29, 17-40.

Peterse, F., van der Meer, J., Schouten, S., Weijers, J.W.H., Fierer, N., Jackson, R.B., Kim, J.-H. and Sinninghe Damsté, J.S. (2012) Revised calibration of the MBT–CBT paleotemperature proxy based on branched tetraether membrane lipids in surface soils. Geochimica et Cosmochimica Acta 96, 215-229.

Prahl, F.G. and Wakeham, S.G. (1987) Calibration of unsaturation patterns in long-chain ketone compositions for palaeotemperature assessment. Nature 330, 367-369.

Raetz, C.R., Reynolds, C.M., Trent, M.S. and Bishop, R.E. (2007) Lipid A modification systems in gram-negative bacteria. Annu. Rev. Biochem. 76, 295-329.

Rampen, S.W., Willmott, V., Kim, J.-H., Uliana, E., Mollenhauer, G., Schefuß, E., Sinninghe Damsté, J.S. and Schouten, S. (2012) Long chain 1,13- and 1,15-diols as a potential proxy for palaeotemperature reconstruction. Geochimica et Cosmochimica Acta 84, 204-216.

Rasmussen, C.E. and Nickisch, H. (2010) Gaussian processes for machine learning (GPML) toolbox. Journal of Machine Learning Research 11, 3011-3015.

Rasmussen, C.E. and Williams, C.K.I. (2006) Gaussian processes for machine learning. The MIT Press, Massachusetts.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., Knight, R. and Fierer, N. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. The ISME Journal 4, 1340-1351.

Russell, N. (1995) Psychrotrophy and adaptation to low temperatures: microbial membrane lipids, Proceedings of the 19th International Congress on Refrigeration: Workshop Refrigeration and Microbiology: Health, Food, Drinks and Flowers, pp. 359-365.

Russell, N. and Fukunaga, N. (1990) A comparison of thermal adaptation of membrane lipids in psychrophilic and thermophilic bacteria. FEMS Microbiology Letters 75, 171-182.

Sachs, J.P., Anderson, R.F. and Lehman, S.J. (2001) Glacial surface temperatures of the southeast Atlantic Ocean. Science 293, 2077-2079.

Schouten, S., Hopmans, E., Schefuβ , E. and Sinninghe Damsté, J. (2002) Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? Earth and Planetary Science Letters 204, 265-274.

Schouten, S., Hopmans, E.C. and Sinninghe Damsté, J.S. (2013) The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. Organic Geochemistry 54, 19-61.

Siliakus, M.F., van der Oost, J. and Kengen, S.W.M. (2017) Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. Extremophiles 21, 651-670.

Singh, D., Shi, L. and Adams, J.M. (2013) Bacterial diversity in the mountains of south-west China: climate dominates over soil parameters. Journal of Microbiology 51, 439-447.

Suutari, M. and Laakso, S. (1994) Microbial fatty acids and thermal adaptation. Critical Reviews in Microbiology 20, 285-328.

Szponar, B., Krasnik, L., Hryniewiecki, T., Gamian, A. and Larsson, L. (2003) Distribution of 3-hydroxy fatty acids in tissues after intraperitoneal injection of endotoxin. Clinical Chemistry 49, 1149-1153.

Szponar, B., Norin, E., Midtvedt, T. and Larsson, L. (2002) Limitations in the use of 3-hydroxy fatty acid analysis to determine endotoxin in mammalian samples. Journal of Microbiological Methods 50, 283-289.

Tyagi, P., Kawamura, K., Bikkina, S., Mochizuki, T. and Aoki, K. (2016) Hydroxy fatty acids in snow pit samples from Mount Tateyama in central Japan: Implications for atmospheric transport of microorganisms and plant waxes associated with Asian dust. Journal of Geophysical Research: Atmospheres 121, 13641-13660.

Tyagi, P., Yamamoto, S. and Kawamura, K. (2015) Hydroxy fatty acids in fresh snow samples from northern Japan: long-range atmospheric transport of Gram-negative bacteria by Asian winter monsoon. Biogeosciences 12, 7071-7080.

Véquaud, P., Derenne, S., Thibault, A., Anquetil, C., Bonanomi, G., Collin, S., Contreras, S., Nottingham, A., Sabatier, P., Salinas, N., Scott, W.P., Werne, J.P. and Huguet, A. (2020) Development of global temperature and pH calibrations based on bacterial 3-hydroxy fatty acids in soils. Biogeosciences Discussion 2020, 1-40.

Volkman, J.K., Barrett, S.M. and Blackburn, S.I. (1999) Fatty acids and hydroxy fatty acids in three species of freshwater eustigmatophytes. Journal of Phycology 35, 1005-1012.

Volkman, J.K., Johns, R.B., Gillan, F.T., Perry, G.J. and Bavor Jr, H.J. (1980) Microbial lipids of an intertidal sediment--I. Fatty acids and hydrocarbons. Geochimica et Cosmochimica Acta 44, 1133-1143.

Wakeham, S.G. (1999) Monocarboxylic, dicarboxylic and hydroxy acids released by sequential treatments of suspended particles and sediments of the Black Sea. Organic Geochemistry 30, 1059-1074.

Wakeham, S.G., Pease, T.K. and Benner, R. (2003) Hydroxy fatty acids in marine dissolved organic matter as indicators of bacterial membrane material. Organic Geochemistry 34, 857-868.

Wang, C., Bendle, J., Yang, Y., Yang, H., Sun, H., Huang, J. and Xie, S. (2016) Impacts of pH and temperature on soil bacterial 3-hydroxy fatty acids: Development of novel terrestrial proxies. Organic Geochemistry 94, 21-31.

Wang, C., Bendle, J.A., Zhang, H., Yang, Y., Liu, D., Huang, J., Cui, J. and Xie, S. (2018) Holocene temperature and hydrological changes reconstructed by bacterial 3-hydroxy fatty acids in a stalagmite from central China. Quaternary Science Reviews 192, 97-105.

Wang, C., Zhang, H., Huang, X., Huang, J. and Xie, S. (2012) Optimization of acid digestion conditions on the extraction of fatty acids from stalagmites. Frontiers of Earth Science 6, 109-114.

Wang, H., An, Z., Lu, H., Zhao, Z. and Liu, W. (2020) Calibrating bacterial tetraether distributions towards in situ soil temperature and application to a loess-paleosol sequence. Quaternary Science Reviews 231.

Weber, Y., De Jonge, C., Rijpstra, W.I.C., Hopmans, E.C., Stadnitskaia, A., Schubert, C.J., Lehmann, M.F., Sinninghe Damsté, J.S. and Niemann, H. (2015) Identification and carbon isotope composition of a novel branched GDGT isomer in lake sediments: Evidence for lacustrine branched GDGT production. Geochimica et Cosmochimica Acta 154, 118-129.

Weijers, J.W.H., Schouten, S., van den Donker, J.C., Hopmans, E.C. and Sinninghe Damsté, J.S. (2007) Environmental controls on bacterial tetraether membrane lipid distribution in soils. Geochimica et Cosmochimica Acta 71, 703-713.

Wikström, M., Sharma, V., Kaila, V.R.I., Hosler, J.P. and Hummer, G. (2015) New perspectives on proton pumping in cellular respiration. Chem Rev 115, 2196-2221.

Wilkinson, S.G., Galbraith, L. and Lightfoot, G.A. (1973) Cell walls, lipids, and lipopolysaccharides of *Pseudomonas* species. European Journal of Biochemistry 33, 158-174.

Wollenweber, H.-W. and Rietschel, E.T. (1990) Analysis of lipopolysaccharide (lipid A) fatty acids. Journal of Microbiological Methods 11, 195-211.

Wollenweber, H., Rietschel, E.T., Hofstad, T., Weintraub, A. and Lindberg, A. (1980) Nature, type of linkage, quantity, and absolute configuration of (3-hydroxy) fatty acids in lipopolysaccharides from *Bacteroides fragilis* NCTC 9343 and related strains. Journal of Bacteriology 144, 898-903.

Wollenweber, H.W., Seydel, U., Lindner, B., Lüderitz, O. and Rietschel, E.T. (1984) Nature and location of amide-bound (R)-3-acyloxyacyl groups in lipid A of lipopolysaccharides from various Gram-negative bacteria. European Journal of Biochemistry 145, 265-272.

Yang, H., Ding, W., Wang, J., Jin, C., He, G., Qin, Y. and Xie, S. (2012) Soil pH impact on microbial tetraether lipids and terrestrial input index (BIT) in China. Science China Earth Sciences 55, 236-245.

Yang, H., Pancost, R.D., Dang, X., Zhou, X., Evershed, R.P., Xiao, G., Tang, C., Gao, L., Guo, Z. and Xie, S. (2014) Correlations between microbial tetraether lipids and environmental variables in Chinese soils: Optimizing the paleo-reconstructions in semi-arid and arid regions. Geochimica et Cosmochimica Acta 126, 49-69.

Yang, Y., Wang, C., Bendle, J.A., Yu, X., Gao, C., Lü, X., Ruan, X., Wang, R. and Xie, S. (2020) A new sea surface temperature proxy based on bacterial 3-hydroxy fatty acids. Organic Geochemistry 141, 103975.

Yang, Y., Wang, C., Zhang, H., Huang, J. and Xie, S. (2016) Influence of extraction methods on the distribution pattern and concentration of fatty acids and hydroxy fatty acids in soil samples: Acid digestion versus saponification. Geochemical Journal 50, 439-443.

Yao, M., Rui, J., Niu, H., Heděnec, P., Li, J., He, Z., Wang, J., Cao, W. and Li, X. (2017) The differentiation of soil bacterial communities along a precipitation and temperature gradient in the eastern Inner Mongolia steppe. Catena 152, 47-56.

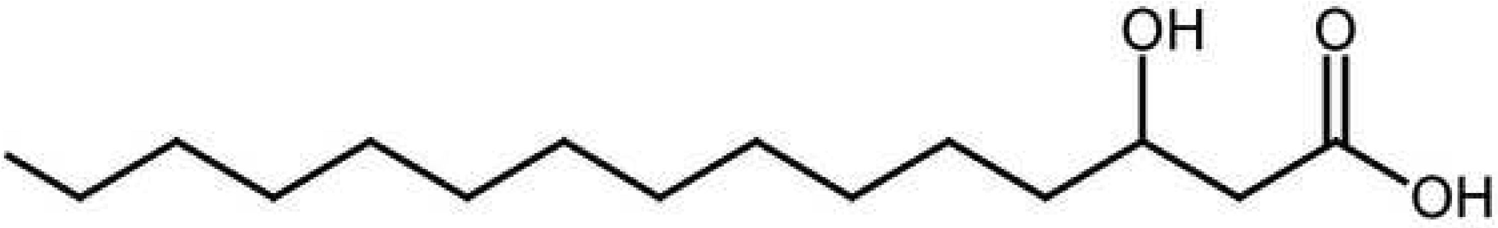
Zelles, L. (1999) Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation of microbial communities in soil: a review. Biology and Fertility of Soils 29, 111-129.

Zhang, Y., Cong, J., Lu, H., Li, G., Xue, Y., Deng, Y., Li, H., Zhou, J. and Li, D. (2015) Soil bacterial diversity patterns and drivers along an elevational gradient on Shennongjia Mountain, China. Microbial Biotechnology 8, 739-746.

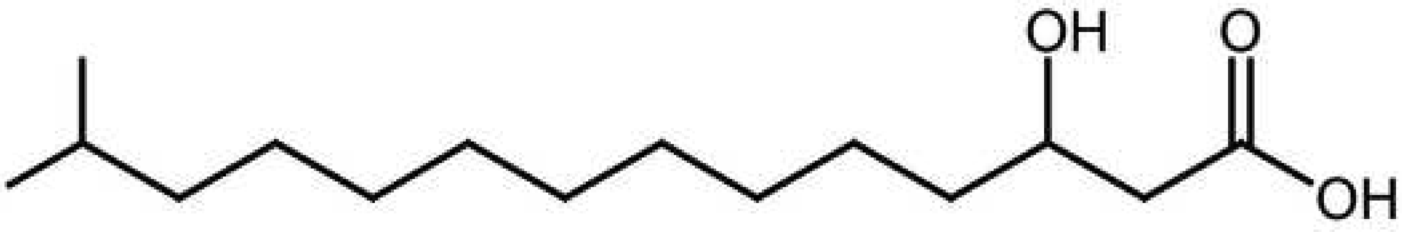
Zhang, Z., Metzger, P. and Sachs, J.P. (2014) Bound lipid biomarkers in sediments from El Junco Lake, Galápagos Islands. Organic Geochemistry 75, 122-128.

Figure 1

Normal (*n*-)



Iso (*i*-)



Anteiso (*a*-)

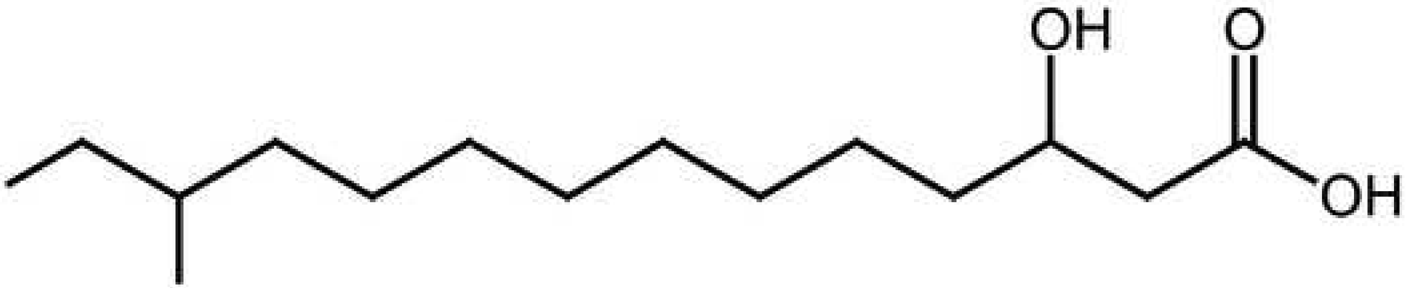


Figure 2

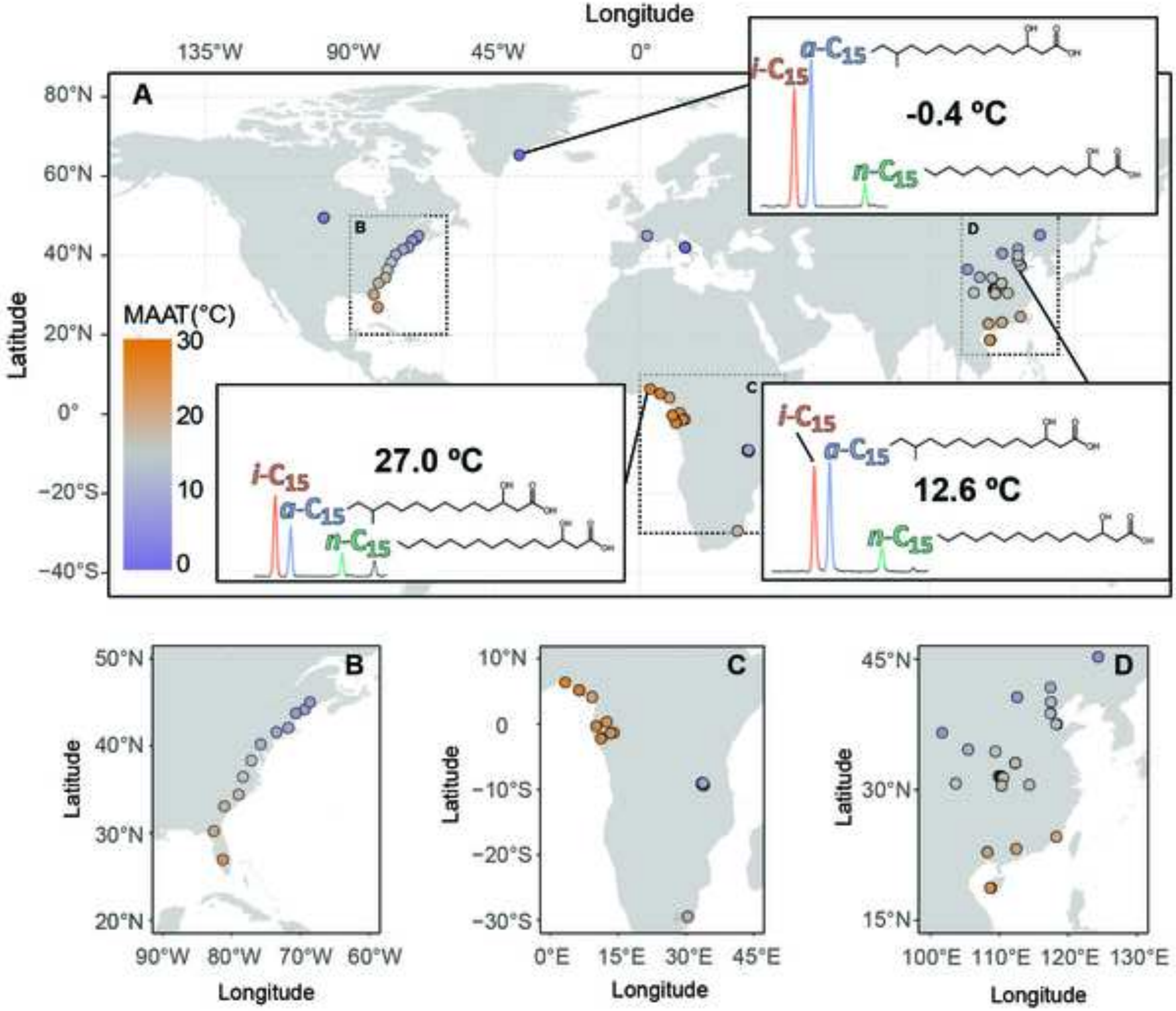

Longitude
135°W
90°W
45°W
0°
A
80°N
60°N
40°N
20°N
0°
-20°S
-40°S
Latitude
MAAT(°C)
30
20
10
0
B
C
D
i-C15
a-C15
n-C15
-0.4 °C
27.0 °C
12.6 °C
50°N
40°N
30°N
20°N
90°W
80°W
70°W
60°W
Longitude
Latitude
10°N
0
-10°S
-20°S
-30°S
0°E
15°E
30°E
45°E
45°N
30°N
15°N
100°E
110°E
120°E
130°E

Figure 3

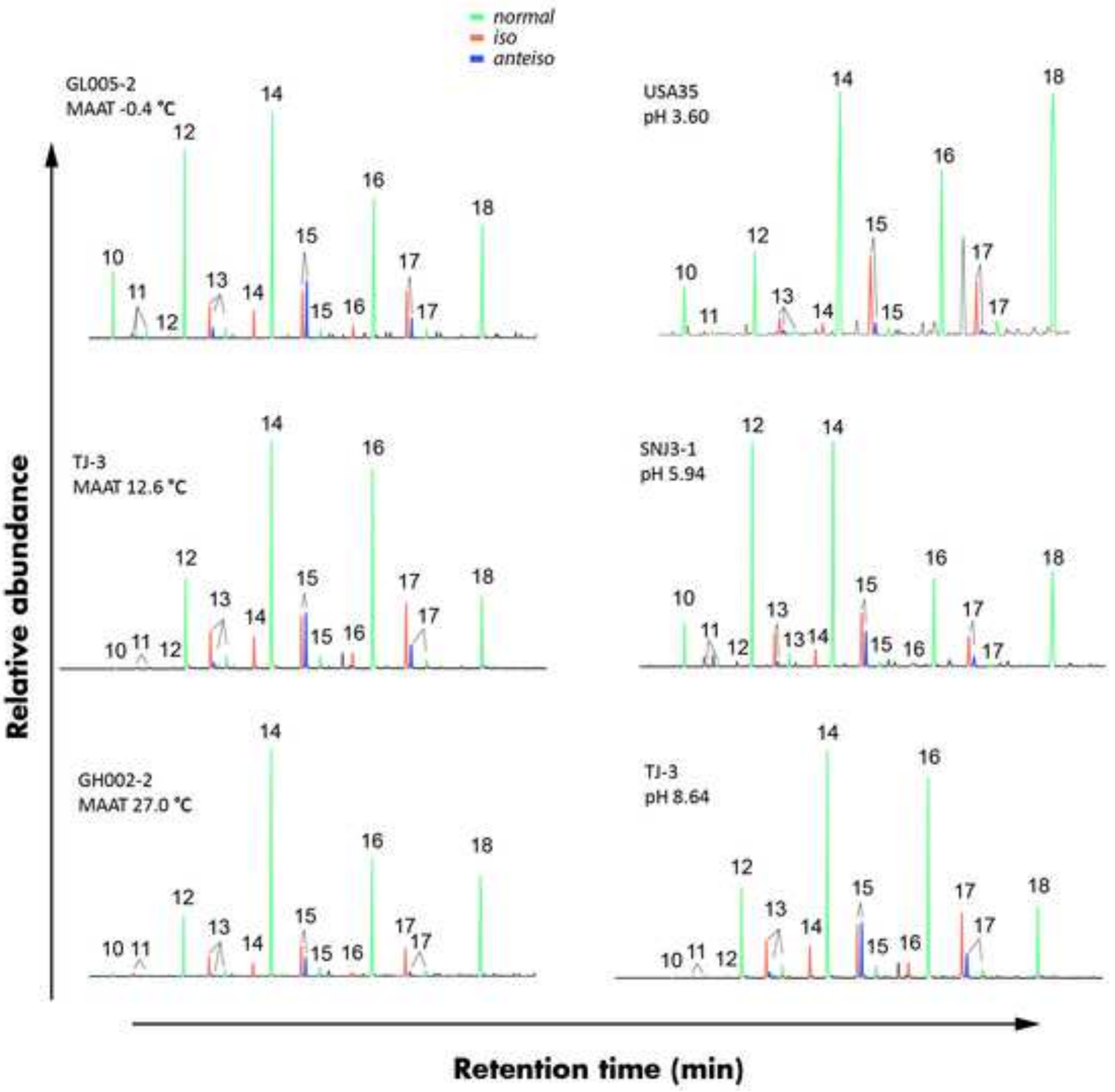

🟩 normal
🟥 iso
🟦 anteiso
GL005-2
MAAT -0.4 °C
10 11 12 12 13 14 14 15 15 16 16 17 17 18
USA35
pH 3.60
10 11 12 13 14 14 15 15 16 17 17 18
TJ-3
MAAT 12.6 °C
10 11 12 12 13 14 14 15 15 16 16 17 17 18
SNJ3-1
pH 5.94
10 11 12 12 13 13 14 14 15 15 16 16 17 17 18
GH002-2
MAAT 27.0 °C
10 11 12 13 14 14 15 15 16 16 17 17 18
TJ-3
pH 8.64
10 11 12 12 13 14 14 15 15 16 16 17 17 18
Relative abundance
Retention time (min)

Figure 4

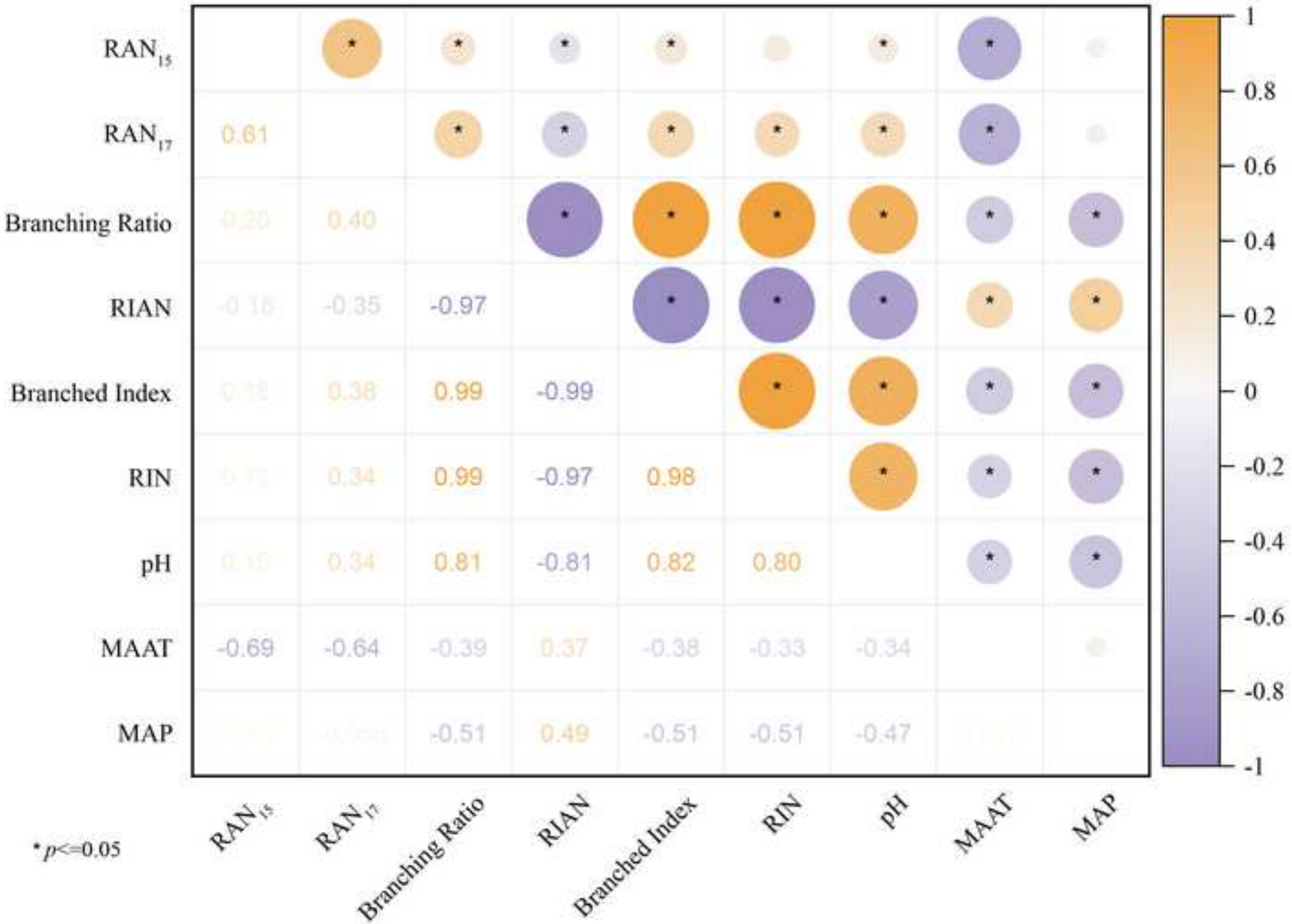
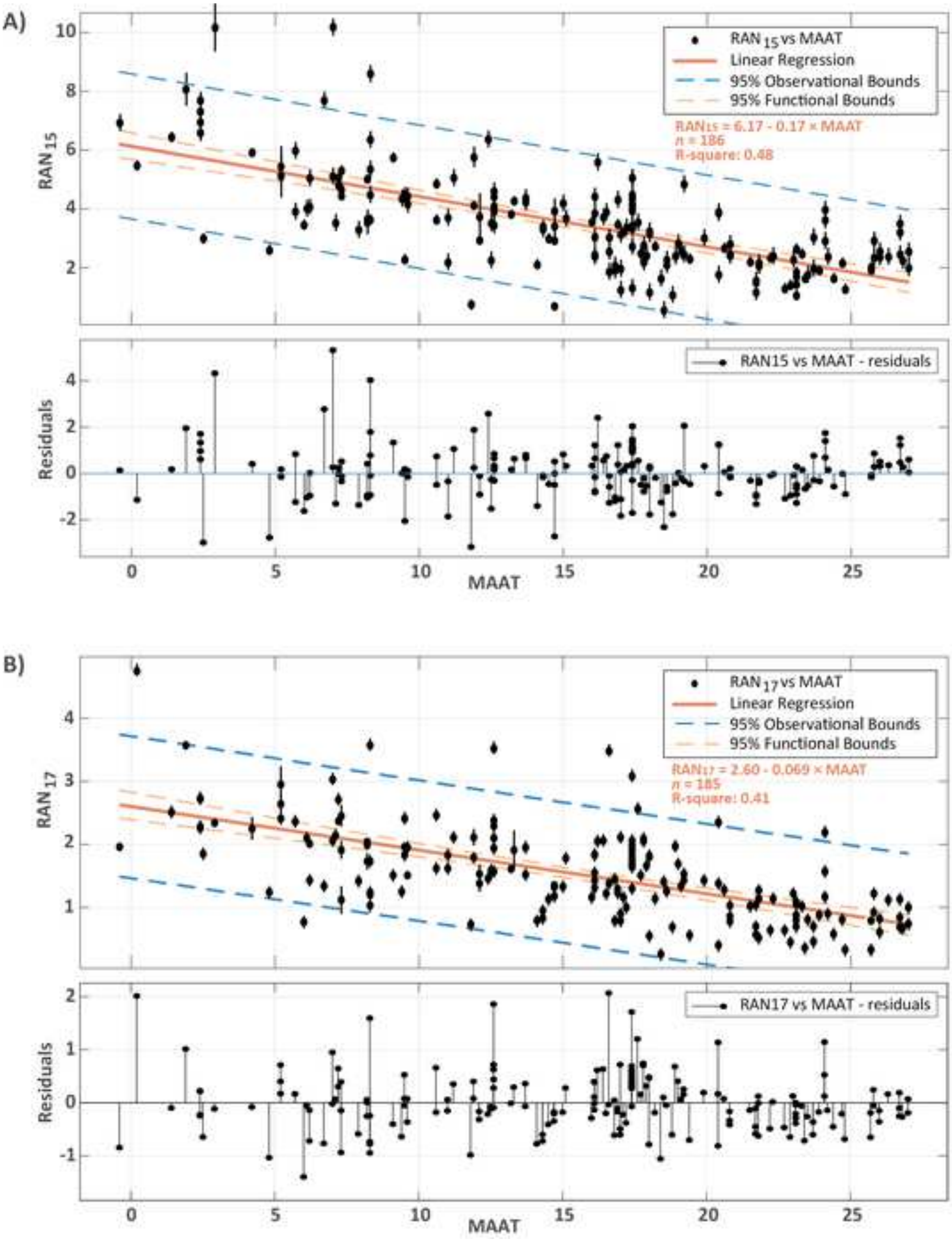
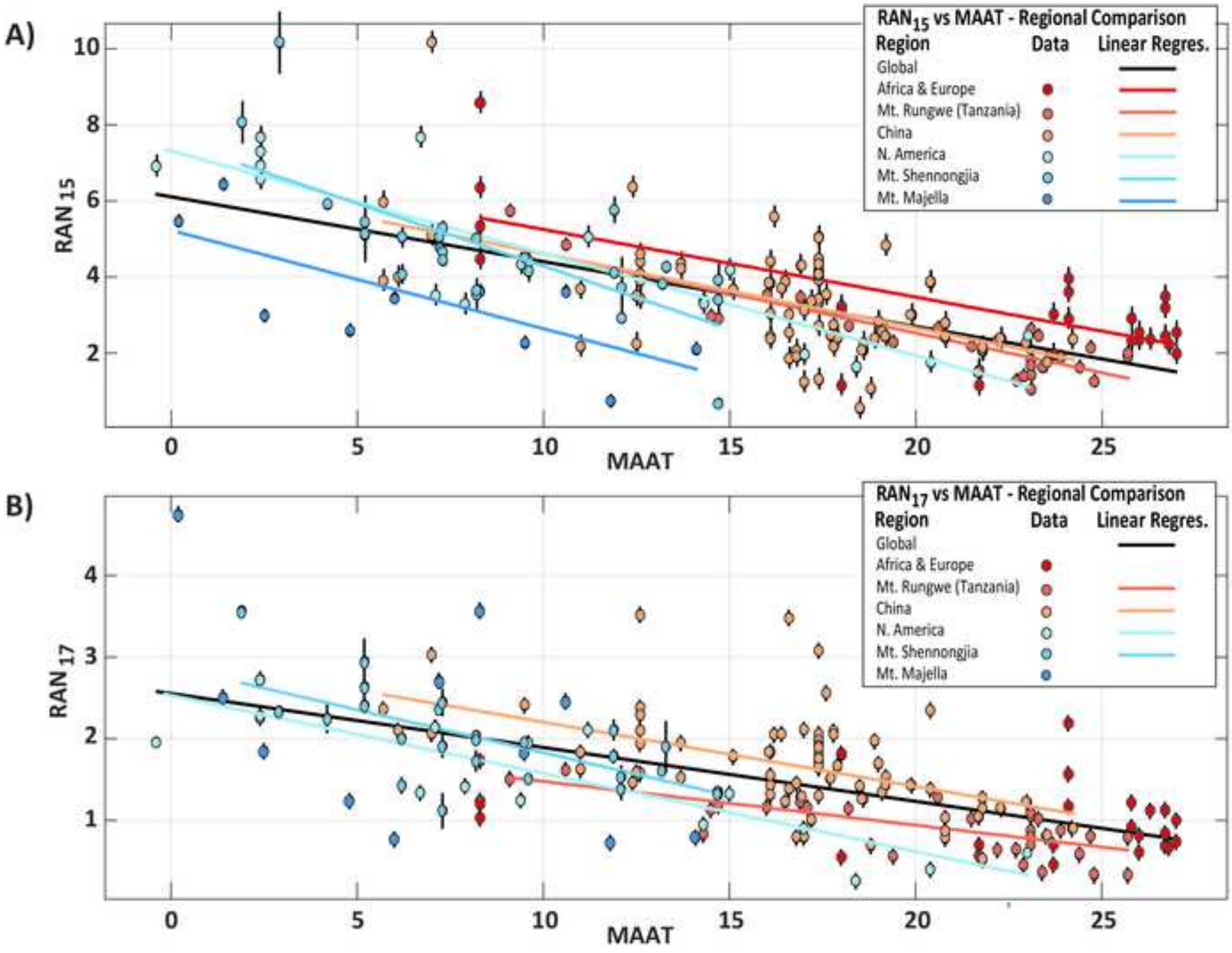

RAN15
RAN17
Branching Ratio
RIAN
Branched Index
RIN
pH
MAAT
MAP
0.61
0.20
0.40
-0.16
-0.35
-0.97
0.15
0.38
0.99
-0.99
0.34
0.99
-0.97
0.98
0.34
0.81
-0.81
0.82
0.80
-0.69
-0.64
-0.39
0.37
-0.38
-0.33
-0.34
-0.51
0.49
-0.51
-0.51
-0.47
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
* $p<=0.05$

Figure 5


A)
RAN$_{15}$ vs MAAT
Linear Regression
95% Observational Bounds
95% Functional Bounds
RAN$_{15}$ = 6.17 - 0.17 × MAAT
n = 186
R-square: 0.48
RAN$_{15}$
10
8
6
4
2
RAN15 vs MAAT - residuals
Residuals
4
2
0
-2
0
5
10
15
20
25
MAAT
B)
RAN$_{17}$ vs MAAT
Linear Regression
95% Observational Bounds
95% Functional Bounds
RAN$_{17}$ = 2.60 - 0.069 × MAAT
n = 185
R-square: 0.41
RAN$_{17}$
4
3
2
1
RAN17 vs MAAT - residuals
Residuals
2
1
0
-1
0
5
10
15
20
25
MAAT


A)
RAN 15
MAAT
0
5
10
15
20
25
2
4
6
8
10
RAN$_{15}$ vs MAAT - Regional Comparison
Region
Data
Linear Regres.
Global
Africa & Europe
Mt. Rungwe (Tanzania)
China
N. America
Mt. Shennongjia
Mt. Majella
B)
RAN$_{17}$
MAAT
1
2
3
4
RAN$_{17}$ vs MAAT - Regional Comparison
Region
Data
Linear Regres.
Global
Africa & Europe
Mt. Rungwe (Tanzania)
China
N. America
Mt. Shennongjia
Mt. Majella

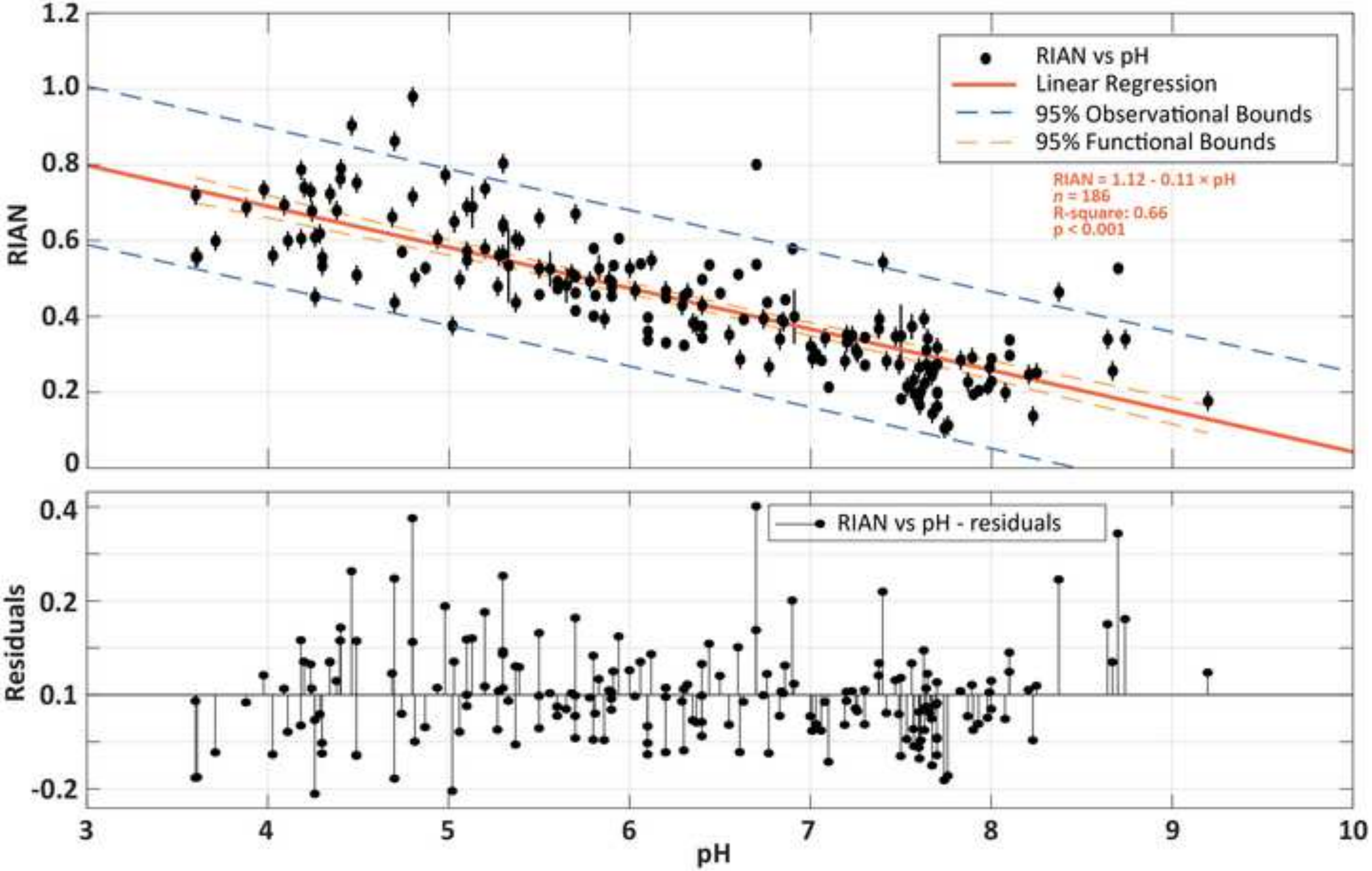
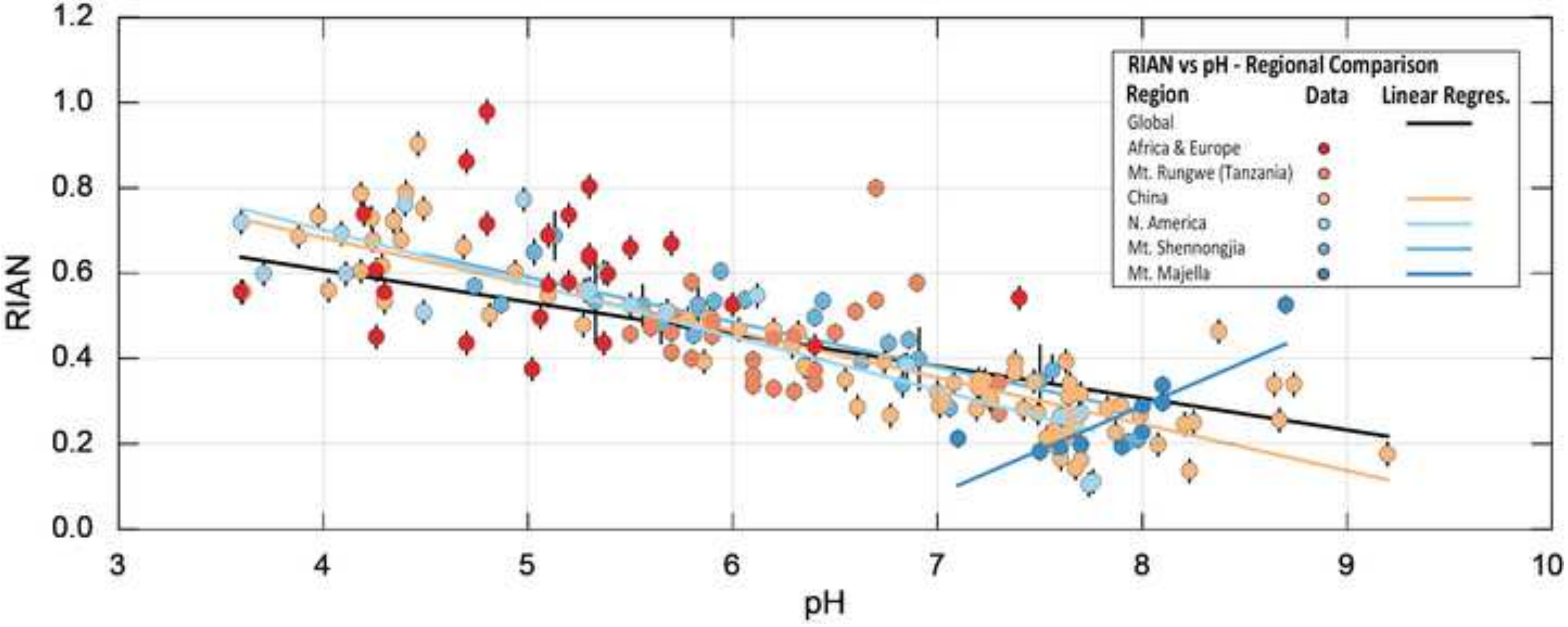
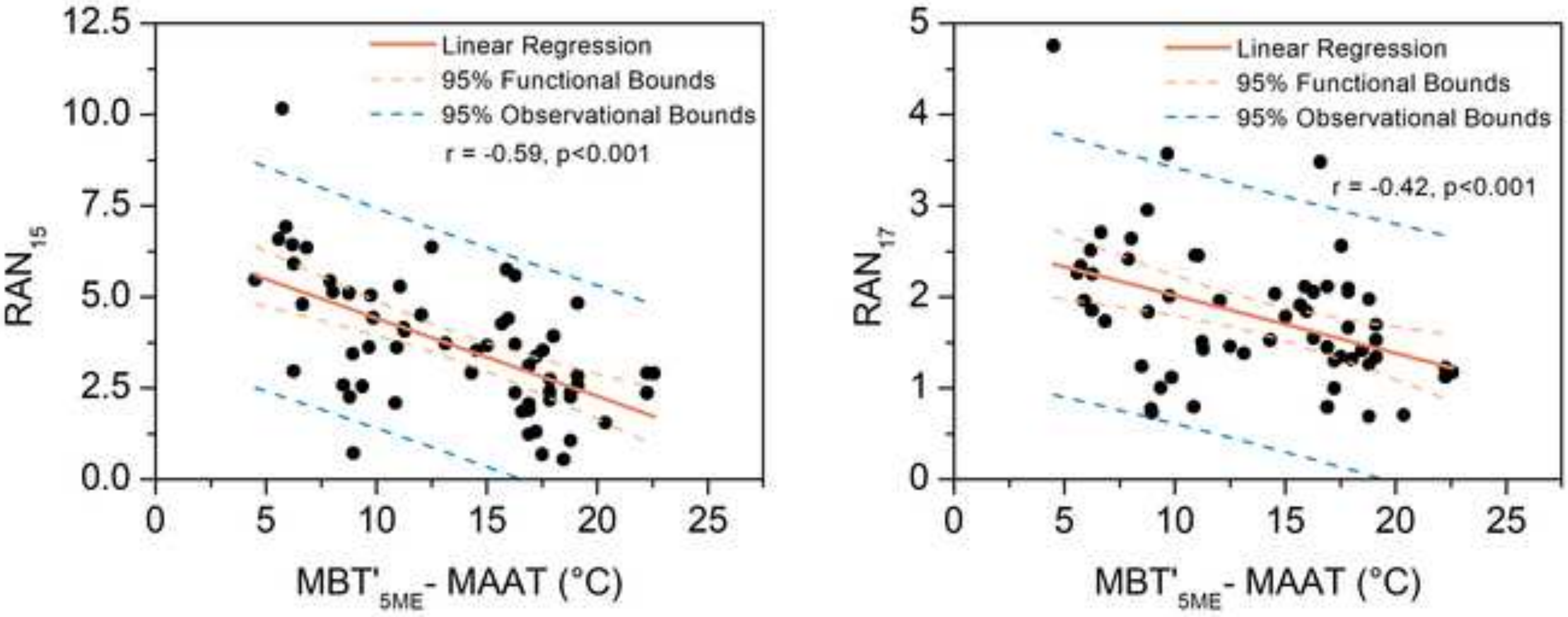

RIAN
1.2
1.0
0.8
0.6
0.4
0.2
0
● RIAN vs pH
Linear Regression
95% Observational Bounds
95% Functional Bounds
RIAN = 1.12 - 0.11 × pH
n = 186
R-square: 0.66
p < 0.001
Residuals
0.4
0.2
0.1
-0.2
RIAN vs pH - residuals
3
4
5
6
7
8
9
10
pH

Figure 8

Click here to access/download;Figure;Figure 8 Regional comparison-RIAN-V3-without shading.png




12.5
10.0
7.5
5.0
2.5
0.0
RAN15
0
5
10
15
20
25
MBT'5ME - MAAT (°C)
Linear Regression
95% Functional Bounds
95% Observational Bounds
r = -0.59, p<0.001
5
4
3
2
1
0
RAN17
MBT'5ME - MAAT (°C)
r = -0.42, p<0.001

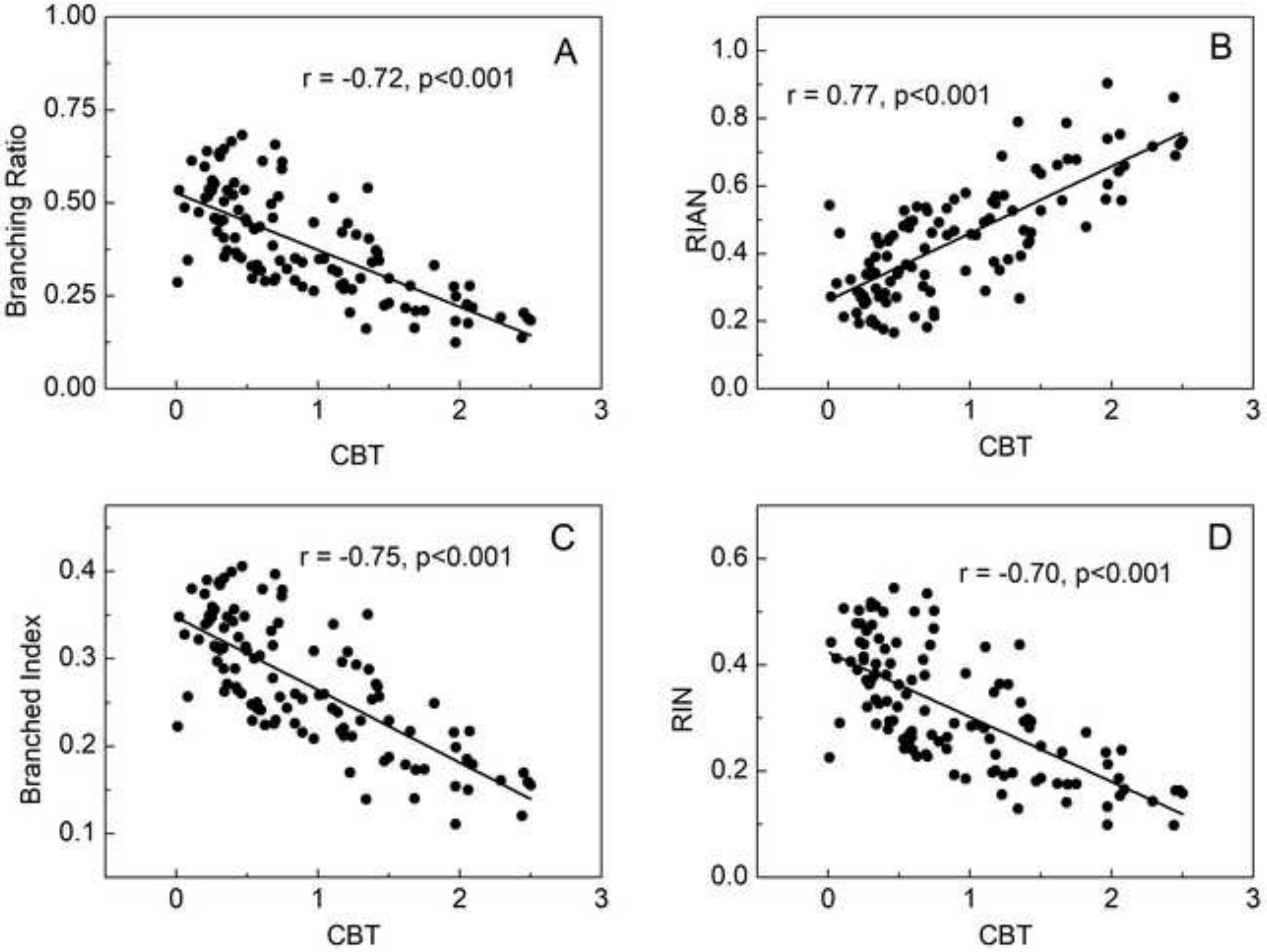

A
r = -0.72, p<0.001
Branching Ratio
1.00
0.75
0.50
0.25
0.00
0
1
2
3
CBT
B
r = 0.77, p<0.001
RIAN
1.0
0.8
0.6
0.4
0.2
0.0
CBT
C
r = -0.75, p<0.001
Branched Index
0.4
0.3
0.2
0.1
CBT
D
r = -0.70, p<0.001
RIN
0.6
0.4
0.2
0.0
CBT

Figure 11

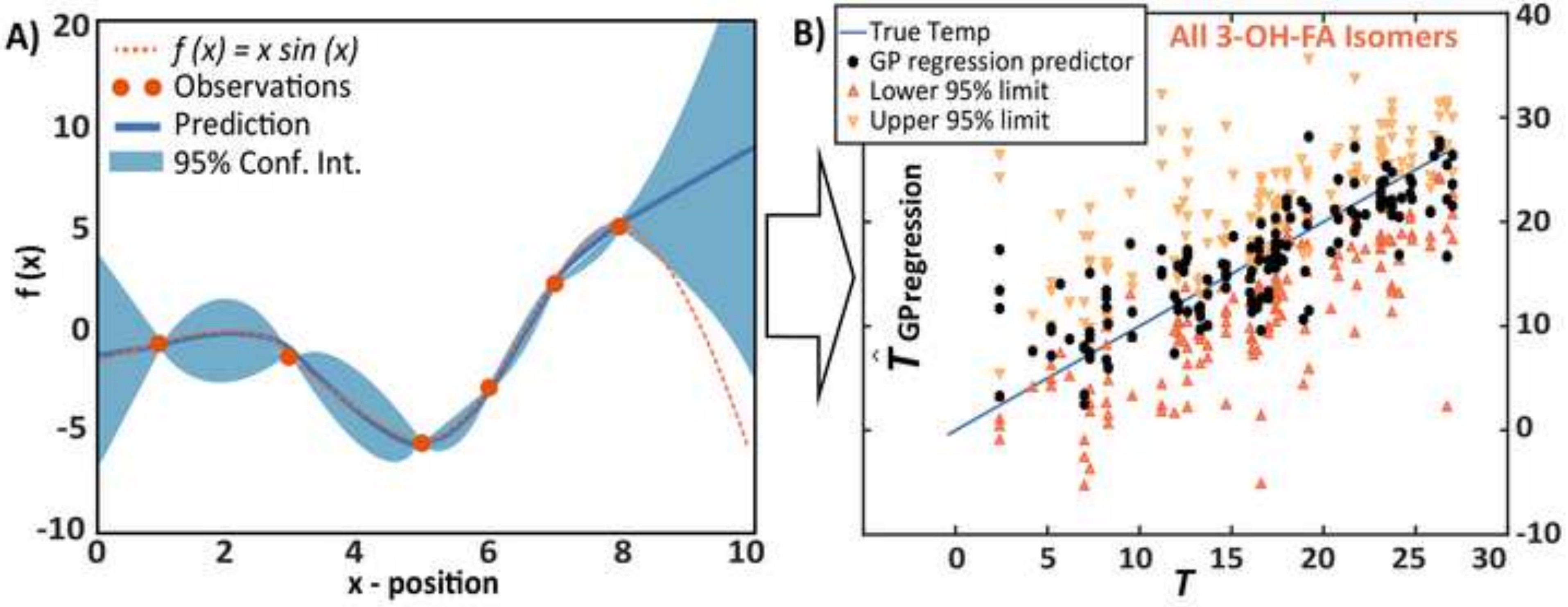
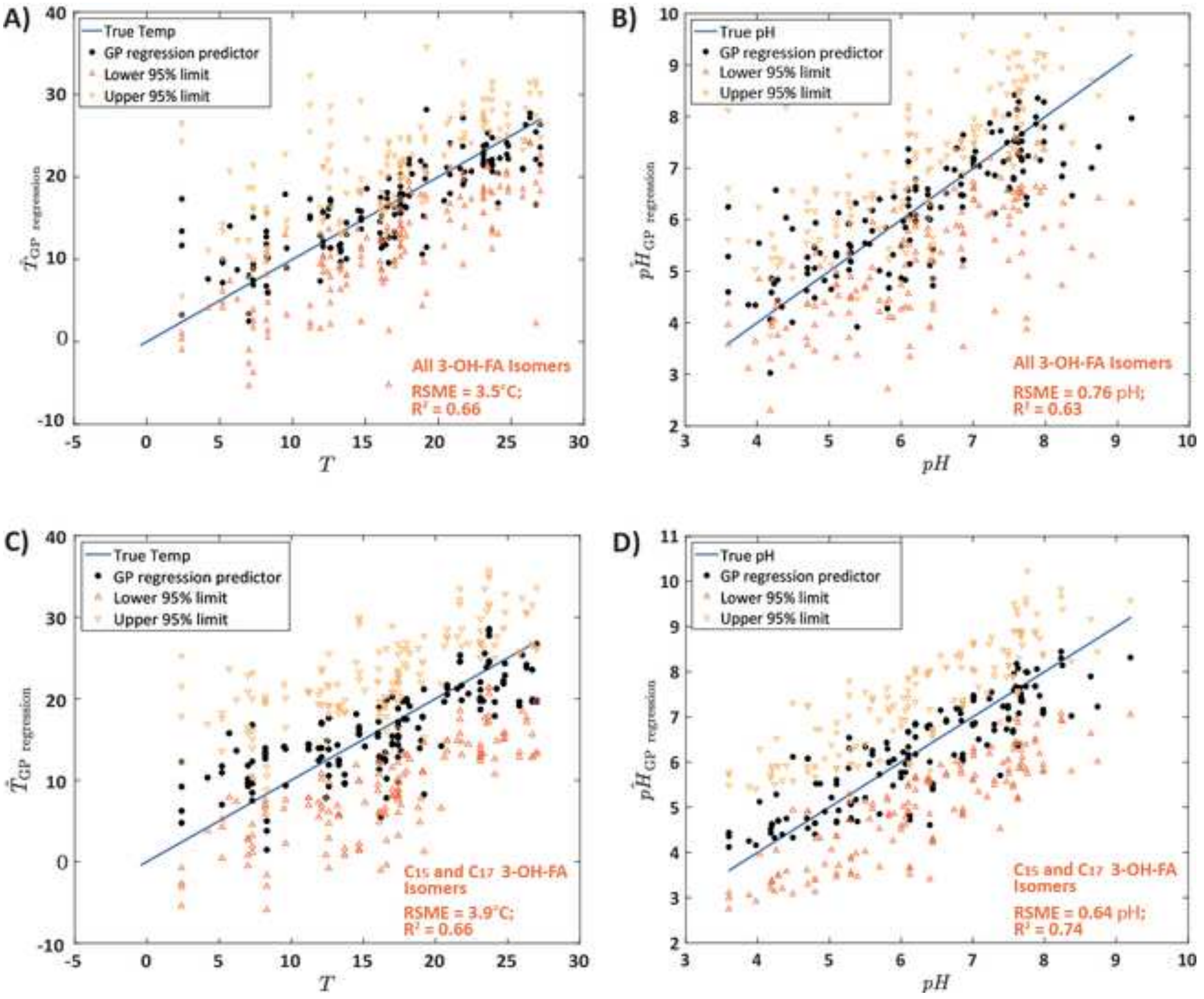

A)
f (x) = x sin (x)
Observations
Prediction
95% Conf. Int.
f (x)
x - position
20
10
5
0
-5
-10
0
2
4
6
8
10
B)
True Temp
GP regression predictor
Lower 95% limit
Upper 95% limit
All 3-OH-FA Isomers
$\hat{T}_{GPregression}$
T
40
30
20
10
0
-10
0
5
10
15
20
25
30

Figure 12


A)
True Temp
GP regression predictor
Lower 95% limit
Upper 95% limit
All 3-OH-FA Isomers
RSME = 3.5°C;
R² = 0.66
$\hat{T}_{\mathrm{GP\ regression}}$
T
40
30
20
10
0
-10
-5
0
5
10
15
20
25
30
B)
True pH
GP regression predictor
Lower 95% limit
Upper 95% limit
All 3-OH-FA Isomers
RSME = 0.76 pH;
R² = 0.63
$p\hat{H}_{\mathrm{GP\ regression}}$
pH
10
9
8
7
6
5
4
3
2
3
4
5
6
7
8
9
10
C)
True Temp
GP regression predictor
Lower 95% limit
Upper 95% limit
C15 and C17 3-OH-FA Isomers
RSME = 3.9°C;
R² = 0.66
$\hat{T}_{\mathrm{GP\ regression}}$
T
40
30
20
10
0
-10
-5
0
5
10
15
20
25
30
D)
True pH
GP regression predictor
Lower 95% limit
Upper 95% limit
C15 and C17 3-OH-FA Isomers
RSME = 0.64 pH;
R² = 0.74
$p\hat{H}_{\mathrm{GP\ regression}}$
pH
11
10
9
8
7
6
5
4
3
2
3
4
5
6
7
8
9
10

Click here to access/download

Appendix

Supplementary Data.xlsx

Declaration of Interest Statement

The authors declare no competing financial or non-financial interests.