



HAL
open science

Développement et application de nouveaux marqueurs d'environnement d'origine bactérienne en milieu terrestre

Pierre Vequaud

► **To cite this version:**

Pierre Vequaud. Développement et application de nouveaux marqueurs d'environnement d'origine bactérienne en milieu terrestre. Sciences de la Terre. Sorbonne Université, 2021. Français. NNT : 2021SORUS475 . tel-03719458

HAL Id: tel-03719458

<https://theses.hal.science/tel-03719458>

Submitted on 11 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

Ecole doctorale 398 Géosciences, Ressources Naturelles et Environnement

UMR 7619 METIS, Département de Biogéochimie

Développement et application de nouveaux marqueurs d'environnement d'origine bactérienne en milieu terrestre

Par Pierre Véquaud

Thèse de doctorat de géochimie organique

Dirigée par Arnaud Huguet, Sylvie Derenne et Sylvie Collin

Présentée et soutenue publiquement le 28 mai 2021

Devant un jury composé de :

M. Fabien ARNAUD	Directeur de Recherche CNRS	Rapporteur
M. Jérémie JACOB	Directeur de Recherche CNRS	Rapporteur
M. Bruno TURCQ	Directeur de Recherche IRD	Examineur
Mme Armelle RIBOULLEAU	Maître de Conférences	Examinatrice
M. Vincent GROSSI	Directeur de Recherche CNRS	Examineur
M. Arnaud HUGUET	Chargé de Recherche CNRS	Directeur de thèse
Mme Sylvie DERENNE	Directrice de Recherche CNRS	Co-encadrante de thèse
Mme Sylvie COLLIN	Maître de Conférences	Co-encadrante de thèse

*A mes parents,
ma sœur,
ma famille,
mes amis
Et pour toi, Jeanne...*

Certaines vérités exigent parfois qu'on reste dans le droit chemin, mais les voies de la connaissance scientifique sont aussi tortueuses et complexes que l'esprit humain

Stephen Jay Gould, Quand les poules auront des dents

Il faut reconnaître que la nature est très artiste. La loi des contrastes, elle la pratique ; en grand d'ailleurs, comme tout ce qu'elle fait.

Jules Verne, Le secret de Wilhelm Storitz

REMERCIEMENTS

Je souhaite tout d'abord remercier l'ensemble des membres du jury d'avoir accepté d'évaluer ce travail et tout particulièrement Jérémy Jacob et Fabien Arnaud qui ont accepté de le rapporter. J'aimerais aussi remercier Jean-Marie Mouchel pour m'avoir accueilli au l'UMR METIS. Enfin j'aimerais adresser de chaleureux remerciement aux membres de mon comité de suivi de thèse : Anniet Laverman, Vincent Grossi et François Baudin. Je tiens aussi à remercier Loïc Labrousse pour m'avoir accueilli au sein de l'ED GRNE.

Une thèse est un travail de recherche, long, parfois périlleux mais toujours enrichissant. C'est aussi une aventure humaine incroyable. Je tiens à adresser de grands remerciements à Arnaud Huguet, Sylvie Derenne et Sylvie Collin pour m'avoir encadré, guidé et accompagné tout au long de cette aventure. Un grand merci pour votre patience, votre gentillesse et votre confiance. Grâce à vous, j'espère avoir mûri en tant que chercheur, mais je sais avoir mûri en tant qu'homme...alors merci.

Cette thèse a été l'occasion de faire de nombreuses rencontres, auprès de personnes qui m'ont partagé leurs passions et leurs expertises avec une grande gentillesse. Ainsi je tiens à remercier Jérôme Poulenard et Pierre Sabatier pour leur accueil chaleureux au sein des locaux de l'UMR EDYTEM, et un immense merci pour cette belle découverte des Alpes françaises ! Merci aussi pour vos conseils, nos discussions et votre accompagnement tout au long de cette aventure. Je reviendrai vite vous voir dans ce joli coin des Alpes !

Je tiens aussi à remercier Magloire Madeng-Yogo et Mercedes Mendez pour leur accueil au sein des laboratoires de l'IRD, m'avoir fait partager leurs passions et leurs travaux et pour les sublimes analyses isotopiques et élémentaires !

Merci à Adrienne Kish pour nos discussions sur le vivant, et sa bonne humeur communicative !

Merci à Sergio Contreras pour ses messages d'encouragement, sa passion et nos discussions ! Espero poder visitarte algún día en Chile.

Lors de ces dernières années, mon second chez moi a été le couloir 56-66. Plus qu'une équipe de recherche, c'est une petite famille qui m'a accompagné lors de cette

thèse. Alors je tiens à remercier chaleureusement tous les occupants du couloir, Marie, Thanh, Maryse, Katell, Christelle, Véronique, Amélie, Emmanuel et Loïc, pour nos fous rires le midi, nos discussions et votre immense gentillesse. Je reviendrai très vite vous embêter au sein du couloir, promis ! Tout particulièrement, je tiens à remercier Maryse pour m'avoir accompagné et guidé lors de ma mission d'enseignement. Je tiens aussi à remercier Marie pour nos discussions passionnantes et de m'avoir prêté (pendant une durée indéterminée...ahem) un livre que j'ai fini par lui rendre ! Enfin, merci à toi Katell pour ta générosité, ta gentillesse et nos discussions ! Tous à un moment ou un autre avait su être là et me guider au cours de cette aventure. Alors un immense merci !

Merci à Fred, pour son franc-parler inimitable, son humour et sa passion pour la science. Je garde un superbe souvenir de cette soirée à Lille, où à l'occasion des RST, nous avons partagé ce plat fin et léger qu'est le Welsh avec Jérémy Jacob !

Je tiens à remercier tous les doctorants et post-doctorants que j'ai eu la chance de croiser au cours ma thèse, et plus particulièrement Flore, Noëlie et Mohamed avec qui nous avons pu partager nos aventures de thésards !

Un grand merci à Aliénor...Sa bonne humeur et sa gentillesse m'ont accompagné durant deux années. Deux années au cours desquelles elle a gentiment redécoré l'écran de mon ordinateur et envoyé de nombreux post-it. Merci pour tous ces bons moments !

Mais qu'aurait été cette thèse sans la reine de la GC, celle qui murmure à l'oreille de la LC, qui dompte la FID..., l'incroyable et unique... Christelle ! Alalala...que de souvenirs et de fous rires !!! A ton contact, j'ai pu découvrir le monde merveilleux des chromatogrammes, et du développement analytique...Je ne te remercierai jamais assez pour tout ce que tu m'as apporté au plan professionnel et surtout humain ! Sans oublier Louis bien sûr, toujours présent et d'une gentillesse sans égale...quel trio de choc !!

Et bien entendu un grand merci à Thierry qui au cours de son stage a aidé à la préparation et à l'analyse de nombreux échantillons !

Avant de commencer cette thèse, je répétais souvent que je n'aimais pas les statistiques...et bien les choses n'ont pas vraiment changé, mais grâce à toi Yannick j'ai pu découvrir la beauté et l'utilité des modèles statistiques et mathématiques. Je tiens aussi à te remercier pour ta gentillesse et l'amitié qui s'est tissée entre nous. J'espère te rendre visite très vite en Belgique ! Et bien entendu merci à toi Alexandre !!! Que d'après-midi passés à refaire le monde, à discuter de sciences (et de sujets beaucoup moins profonds !). Tu as été mon « Senseï » en statistiques, et je te dois énormément ! Merci à Guillaume

qui a dû cohabiter dans le même bureau de moi, d'avoir supporté mes blagues douteuses. Notre amitié vaut sans doute plus que tous les remerciements que je pourrai te faire...Enfin merci à toi Eve. Une rencontre bien trop tardive ! J'ai hâte d'apprendre à parler breton et venir visiter ton pays natal ! On a tendance à croire que c'est la durée d'une relation qui fait son intensité...Je ne suis pas d'accord, la preuve en est de notre rencontre avec Sharon, une personne incroyable (ou plutôt parfaite selon ses propres dires (!)). Hâte de partager de futures aventures avec toi. And last but not least, Clara, one of those rare encounters that will make you question your view of the world. You are a beautiful person and I know what I owe you...forever.

Mais cette aventure n'aurait pas été possible sans mes débuts au sein de l'UMR EPOC, et je tiens donc à remercier toutes les personnes qui m'ont accompagnées et guidées au cours de cette période ! J'aimerais aussi remercier une personne sans qui rien de tout cela n'aurait été possible, Frédérique Eynaud qui m'a initié au monde de la recherche. Je tiens aussi à remercier Bruno Malaizé et Thibaut Caley, pour leur accompagnement au cours de mes stages de recherches, leurs passions et leurs gentillesse. Enfin je tiens à remercier Xavier Crosta pour son soutien tout au long de mon parcours, son amitié et sa gentillesse !

Enfin, merci à ma famille et mes amis de m'avoir accompagné et soutenu tout au long de cette thèse. Merci à ma maman qui a toujours su croire en moi, à ma petite sœur Marie qui a illuminé les derniers mois de ma thèse avec mon neveu et ma nièce. Merci à vous Quentin et Eva-Marie...vous êtes là depuis le début (aaaah le bâtiment A22 du campus de Peixotto...) plus que des amis, vous êtes ma famille.

Enfin, Jeanne...je ne pourrai jamais assez te remercier de m'avoir accompagné, soutenu (et supporté) au cours de cette thèse. Sans toi rien n'aurait été possible...Tu es mon monde.

J'oublie énormément de monde, sachez que je ne vous remercierai jamais assez et que vous avez ma plus profonde gratitude !

TABLE DES MATIERES

REMERCIEMENTS	5
TABLE DES MATIERES	9
INTRODUCTION	17
CHAPITRE 1 – ETAT DE L'ART	23
1.1. INTERET ET ETUDE DES PAEOCLIMATS ET DES PAEOENVIRONNEMENTS	25
1.1.1. LE SYSTEME CLIMATIQUE	25
1.1.2. CHANGEMENTS CLIMATIQUES	28
1.1.2.1. Origines et conséquences des changements climatiques	28
1.1.2.2. Effets du changement climatique et prévisions futures	30
1.1.3. ETUDE DES PAEOCLIMATS ET PAEOENVIRONNEMENTS	35
1.1.3.1. Des outils pour reconstituer les climats passés	35
1.1.3.2. Concept de biomarqueur	38
1.2. DIVERSITE ET ADAPTABILITE DU VIVANT	40
1.2.1. DOMAINE DES EUCARYOTES	42
1.2.2. DOMAINE DES ARCHEES	43
1.2.3. DOMAINE DES BACTERIES	45
• Bactéries Gram-négatif vs. Gram-positif	47
1.2.4. ROLE ET STRUCTURE DES MEMBRANES CYTOPLASMIQUES	49
1.2.4.1. Un composant essentiel des membranes eucaryotes et bactériennes : les glycérophospholipides	49
1.2.4.2. Les éthers de glycérol archéens et bactériens	51
1.2.4.3. Notion de lipides polaires intacts et lipides de coeur	52
1.2.5. ADAPTATION HOMEO-VISQUEUSE DES MEMBRANES CYTOPLASMIQUES	54
1.2.5.1. Une mosaïque fluide...	54
1.2.5.2. ...qui s'adapte aux variations environnementales...	54
1.2.5.3. ...notamment de température et de pH	56
1.3. LES LIPIDES COMME MARQUEURS D'ENVIRONNEMENT	58
1.3.1. EXEMPLES DE BIOMARQUEURS DE TEMPERATURE EN MILIEU AQUATIQUE	58

1.3.1.1. Les alcénones insaturés (Indice U_{37}^k)	58
1.3.1.2. Les diols à longue chaîne (indice LDI)	62
1.3.1.3. Les alkyl tetraéthers de glycérol à chaînes isopréniques (indice TEX_{86})	65
1.3.1.4. Conclusion partielle	69
1.3.2. MARQUEURS DE TEMPERATURE ET DE PH EN MILIEU TERRESTRE	70
1.3.2.1. Les tetraéthers de glycérol bactériens	70
• Structures et sources des tetraéthers bactériens	70
• Utilisation des tetraéthers bactériens comme marqueurs de température et de pH en milieu terrestre	73
• Tetraéthers isopréniques et ramifiés : Indice BIT	77
1.3.2.2. Validation et application des brGDGTs comme marqueurs de température et de pH en milieu continental	78
• Validation des brGDGTs comme marqueurs de température et de pH	78
• Applications paléoenvironnementales	79
1.3.2.3. Les acides gras 3-hydroxylés	81
• Structures et sources des acides gras 3-hydroxylés	81
• Les 3-AGH comme marqueurs de température et de pH en milieu terrestre	83
1.3.3. CONCLUSION PARTIELLE	87
CHAPITRE 2 – MATERIELS ET METHODES	89

2.1. SITES D'ETUDE ET ECHANTILLONNAGE	91
2.1.1. MASSIFS DES BAUGES ET DU LAUTARET-GALIBIER	93
2.1.1.1. Massif des Bauges	94
2.1.1.2. Massif du Lautaret-Galibier	95
2.1.1.3. Echantillonnage	96
2.1.1.4. Données de température	97
2.1.2. MONT POLLINO (ITALIE)	99
2.1.3. MONT SHEGYLA (TIBET)	100
2.1.4. ANDES, TRANSECT DU KOSŃIPATA (PEROU)	101
2.1.5. ARAUCANOS (CHILI)	102
2.1.6. JEU DE DONNEES GLOBAL (BRGDGTs)	102
2.2. ETUDES DE SOLS EN ENVIRONNEMENTS CONTROLES (MICROCOSMES)	104
2.2.1. CHOIX DES SOLS	104

2.2.2. PROTOCOLE D'INCUBATION	106
2.3. ANALYSES ELEMENTAIRES, PH ET HUMIDITE	108
2.3.1. HUMIDITE RELATIVE	108
2.3.2. ANALYSES ELEMENTAIRES ET ISOTOPIQUES	108
2.3.3. MESURES DE PH	109
2.3.4. GRANULOMETRIE ET CAPACITE D'ECHANGE CATIONIQUE	109
2.4. ANALYSE DES LIPIDES BACTERIENS : BRGDGT ET 3-AGH	109
2.4.1. EXTRACTION, SEPARATION ET ANALYSE DES BRGDGTS	110
2.4.1.1. Protocole d'extraction et de séparation des tétraéthers bactériens	110
2.4.1.2. Protocole d'analyse des tétraéthers bactériens	110
2.4.2. EXTRACTION, SEPARATION ET ANALYSE DES ACIDES GRAS 3-HYDROXYLES	111
2.4.2.1. Protocole d'extraction et de séparation des acides gras 3-hydroxylés	111
2.4.2.2. Protocole initial d'analyse des acides gras 3-hydroxylés	112
2.4.2.3. Développement analytique	113
• Effet du séchage sous azote	113
• Répétabilité et reproductibilité des analyses	115
• Désilylation des échantillons au cours de l'analyse	115
• Présence de coélutions	116
2.4.2.4. Nouvelles conditions d'analyse des acides gras 3-hydroxylés	118
2.5. ANALYSES STATISTIQUES	119
2.5.1. ANALYSES UNIES ET BIVARIEES	119
2.5.2. ANALYSES MULTIFACTORIELLES	119
2.5.3. MACHINE LEARNING	121

CHAPITRE 3 – ETUDE DE MARQUEURS LIPIDIQUES BACTERIENS AU SEIN DE SOLS DES ALPES FRANÇAISES 125

3.1. INFLUENCE OF ENVIRONMENTAL PARAMETERS ON THE DISTRIBUTION OF BACTERIAL LIPIDS IN SOILS FROM THE FRENCH ALPS: IMPLICATIONS FOR PALEO-RECONSTRUCTIONS	127
3.1.1. INTRODUCTION	127
3.1.2. RESULTS	131
3.1.2.1. Bulk soil properties	131
3.1.2.2. BrGDGT and 3-OH FA distribution	133
3.1.3. DISCUSSION	134

3.1.3.1. BrGDGTs and 3-OH FAs as proxies of pH and temperature	134
• Relationship between pH and bacterial lipid distribution	134
• Relationship between MAAT and bacterial lipid distribution	136
3.1.3.2. Impact of other environmental variables on bacterial lipid distribution	139
• Selection of non-redundant environmental variables for statistical analyses	139
• RDA analysis of brGDGT distribution and environmental variables	141
• RDA analysis of 3-OH FA distribution and environmental variables	144
• Influence of vegetation and soil type on 3-OH FA and brGDGT distribution	145
3.1.3.3. Development of local calibrations between the relative abundance of bacterial lipids and MAAT/pH	148
• BrGDGTs	148
• 3-OH FAs	149
3.1.3.4. Constraints on the applicability of brGDGTs and 3-OH FAs as environmental proxies	151
3.1.4. CONCLUSIONS	153
3.2. ETUDE DE L'EFFET DE LA TEMPERATURE SUR LA DISTRIBUTION DES ACIDES GRAS 3-HYDROXYLES ET DES TETRAETHERS BACTERIENS AU SEIN DE SOLS DES ALPES FRANÇAISES EN CONDITIONS CONTROLEES.	155
3.2.1. INTRODUCTION	155
3.2.2. RESULTATS ET DISCUSSION	155
3.2.2.1. Viabilité des organismes au sein des microcosmes	155
3.2.2.2. Concentrations et abondances relatives des lipides bactériens	158
• <i>Les acides gras 3-hydroxylés (3-AGH)</i>	158
• <i>Les tétraéthers bactériens (brGDGT)</i>	165
3.2.2.3. Adaptation des microorganismes et production des lipides membranaires en réponse aux variations de températures	170
3.2.3. CONCLUSIONS	172
<u>CHAPITRE 4 – DEVELOPMENT OF GLOBAL TEMPERATURE AND PH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS</u>	<u>175</u>
4.1. INTRODUCTION	177
4.2. RESULTS	180
4.2.1. DISTRIBUTION OF BACTERIAL LIPIDS	180
4.2.1.1. 3-OH-FAs	180
4.2.1.2. brGDGTs	181

4.2.2. 3-OH FA- AND BRGDGT-DERIVED INDICES	182
4.2.2.1. 3-OH FAs	182
4.2.2.2. brGDGTs	183
4.2.3. PRINCIPAL COMPONENT ANALYSIS AND CLUSTERING OF 3-OH FA AND BRGDGT DISTRIBUTION	185
4.2.3.1. 3-OH FAs	185
4.2.3.2. brGDGTs	185
4.3. DISCUSSION	187
4.3.1. 3-OH FA- AND BRGDGT-DERIVED PROXIES	187
4.3.1.1. Relationships between pH- and bacterial lipid-derived proxies	187
4.3.1.2. Relationships between MAAT- and bacterial lipid-derived proxies	190
4.3.2. DEVELOPMENT OF NEW MODELS FOR THE RECONSTRUCTION OF MAAT AND PH FROM 3-OH FAS	194
4.3.2.1. Temperature calibrations	196
4.3.2.2. pH calibrations	198
4.3.3. PALAEOCLIMATE APPLICATION OF THE NEW 3-OH FA AND MAAT MODELS	201
4.3.3.1. Comparison of the multiple linear regression, k-NN and random forest global MAAT calibrations	202
4.3.3.2. Comparison of the global random forest and local RAN ₁₅ calibrations for MAAT reconstruction	203
4.4. CONCLUSIONS	207

CHAPITRE 5 – FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING **209**

5.1. INTRODUCTION	211
5.2. RESULTS	215
5.2.1. PRINCIPAL COMPONENT ANALYSIS AND CLUSTERING ON THE GLOBAL DATASET	215
5.2.2. RELATIONSHIPS BETWEEN MBT' _{5ME} AND MAAT	218
5.3. DISCUSSION	223
5.3.1. CONSTRAINTS ON THE MBT' _{5ME} -MAAT RELATIONSHIP IN SOILS	223
5.3.2. A NOVEL GLOBAL CALIBRATION BETWEEN MAAT AND BRGDGTs USING RANDOM FOREST ALGORITHM	227
5.3.2.1. Model development	227
5.3.2.2. Comparison of the FROG model with previous global soil calibrations	230
5.3.3. DEVELOPMENT OF ALTERNATIVE MAAT CALIBRATIONS	232

5.3.3.1. Alternative models considering the influence of thermal regime	233
5.3.3.2. Alternative model considering the influence of MAP	234
5.3.4. PALEO APPLICATION OF THE FROG GLOBAL CALIBRATION	235
5.3.4.1. Application of the FROG model to a peatland covering the Holocene	235
5.3.4.2. Application of the FROG model to sediments from the Pliocene	238
5.4. CONCLUSION	241
CONCLUSIONS & PERSPECTIVES	243
RÉFÉRENCES BIBLIOGRAPHIQUES	257
LISTE DES FIGURES	295
LISTE DES TABLEAUX	305
ANNEXES	309
ANNEXE 1 – DONNEES SUPPLEMENTAIRES DU CHAPITRE 3 –	311
SUPPLEMENTARY TABLE 3.1. RELATIVE ABUNDANCES (%) OF THE <i>NORMAL</i> , <i>ISO</i> AND <i>ANTEISO</i> 3-OH FAS IN SOILS FROM THE FRENCH ALPS AND CORRESPONDING INDICES.	311
SUPPLEMENTARY TABLE 3.2. RELATIVE ABUNDANCES (%) OF THE BRGDGTs IN SOILS FROM THE FRENCH ALPS AND CORRESPONDING INDICES.	314
SUPPLEMENTARY FIGURE 3.1. LOCATION OF THE SAMPLING SITES ALONG THE A) BAUGES AND B) LAUTARET-GALIBIER MASSIFS.	316
TABLEAU SUPPLEMENTAIRE 3.3. CONCENTRATIONS ET ABONDANCES RELATIVES (%) DES 3-AGH POUR LES 4 ECHANTILLONS DE SOLS PLACES EN MICROCOSME, AUX DIFFERENTS PAS DE TEMPS ET AUX DIFFERENTES TEMPERATURES	317
TABLEAU SUPPLEMENTAIRE 3.4. CONCENTRATIONS ET ABONDANCES RELATIVES (%) DES BRGDGT POUR LES 4 ECHANTILLONS DE SOLS PLACES EN MICROCOSME, AUX DIFFERENTS PAS DE TEMPS ET AUX DIFFERENTES TEMPERATURES	322
ANNEXE 2 – DONNEES SUPPLEMENTAIRES DU CHAPITRE 4 –	325
SUPPLEMENTARY TABLE 4.1. RELATIVE ABUNDANCES (%) OF THE <i>NORMAL</i> , <i>ISO</i> AND <i>ANTEISO</i> 3-OH FAS IN SOILS FROM THE PERUVIAN ANDES, CHILEAN ANDES, MTS POLLINO AND SHEGYLA.	325

SUPPLEMENTARY TABLE 4.2. RELATIVE ABUNDANCES (%) OF THE BRGDGTs (WITH SEPARATION OF THE 5- AND 6-METHYL ISOMERS) FROM THE PERUVIAN ANDES, CHILEAN ANDES, MTS. POLLINO AND SHEGYLA (N.D., NOT DETECTED).	328
SUPPLEMENTARY TABLE 4.3. CHARACTERISTICS OF THE LINEAR REGRESSIONS (R^2 , RMSE, P-VALUE, EQUATION) BETWEEN PH AND RIAN/CBT'; MAAT AND MBT' _{5ME} /RAN ₁₅ /RAN ₁₇ FOR THE DIFFERENT ALTITUDINAL TRANSECT INVESTIGATED	331
SUPPLEMENTARY FIGURE 4.1. PCA BETWEEN PH, MAAT AND 3-OH FA-DERIVED PROXIES (RAN ₁₅ , RAN ₁₇ , RIAN) ALONG MTS RUNGWE, POLLINO, MAJELLA, SHEGYLA, SHENNONGJIA, LAUTARET-GALIBIER PERUVIAN ANDES AND CHILEAN ANDES.	332
SUPPLEMENTARY FIGURE 4.2. PARTIAL PLOTS OF THE INDIVIDUAL 3-OH FA VARIATIONS IN MULTIPLE REGRESSION, RANDOM FOREST AND K-NN MODELS PROPOSED TO ESTIMATE MAAT.	333
SUPPLEMENTARY FIGURE 4.3. PARTIAL PLOTS OF THE INDIVIDUAL 3-OH FA VARIATIONS IN MULTIPLE REGRESSION, RANDOM FOREST AND K-NN MODELS PROPOSED TO ESTIMATE PH.	334
SUPPLEMENTARY FIGURE 4.4. APPLICATION OF THE THREE MAAT MODELS PROPOSED IN THIS STUDY AND APPLIED TO THE HS4 SPELEOTHEM: RANDOM FOREST CALIBRATION (GREEN); K-NN CALIBRATION (PURPLE); MULTIPLE LINEAR REGRESSION (BROWN).	335
ANNEXE 3 – DONNEES SUPPLEMENTAIRES DU CHAPITRE 5 –	336
SUPPLEMENTARY FIGURE 5.1 LINEAR RELATIONSHIP BETWEEN MAAT (°C) AND MBT' _{5ME} IN THE GLOBAL DATASET USED IN THIS STUDY (N=775).	336
SUPPLEMENTARY FIGURE 5.2. COMPARISON OF THE LINEAR REGRESSIONS BETWEEN MAAT (°C) AND MBT' _{5ME} IN THE GLOBAL DATASET FOR SAMPLES WITH IR _{6ME} >0.5 AND IR _{6ME} <0.5.	337
SUPPLEMENTARY FIGURE 5.3 ESTIMATION OF THE COMMUNITY INDEX THRESHOLDS, FOR EACH LINEAR REGRESSION BETWEEN MAAT AND MBT' _{5ME}	338
SUPPLEMENTARY FIGURE 5.4. PARTIAL PLOTS OF THE INDIVIDUAL BRGDGT VARIATIONS IN THE FROG MODEL PROPOSED TO ESTIMATE MAAT.	339
ANNEXE 4 – LISTES DES PUBLICATIONS ET COMMUNICATIONS EN LIEN AVEC CETTE THESE	340
ANNEXE 5 – ENSEIGNEMENTS	343

INTRODUCTION

Cette thèse prend place dans un contexte très particulier, la pandémie mondiale de SARS-CoV-2. Cette crise révèle l'importance de la recherche fondamentale afin de permettre à nos sociétés de relever de grands défis. Outre cette pandémie, notre civilisation est confrontée aujourd'hui à une crise environnementale et climatique sans précédent dans l'histoire récente. Les activités anthropiques sont à l'origine de cette crise globale qui se décline à la fois avec des changements globaux (vagues de sécheresses, augmentation des événements climatiques extrêmes, réchauffement du pergélisol, fonte des inlandsis...) et des enjeux géopolitiques (famines, tensions frontalières, migrations de masse, ...). Il est donc nécessaire de pouvoir appréhender ces problématiques par la compréhension de la variabilité climatique. La communauté des paléoclimatologues, paléocéanographes et paléo-environmentalistes peut mieux contraindre cette variabilité climatique, par l'étude des climats et environnements anciens. Pour atteindre cet objectif, il faut avoir accès aux données environnementales passées. Néanmoins, si cet aspect « utilitaire » de la recherche sur le climat revient souvent, il apparaît que celle-ci peut aussi être menée dans l'objectif simple d'acquérir des connaissances sur le fonctionnement de notre planète, par curiosité et amour de la connaissance, fondements même de toutes recherches fondamentales.

Des méthodes de mesures directes des données environnementales sont disponibles pour les deux derniers siècles (voire plus généralement quelques décennies), soit la période dite « instrumentale ». Au-delà de cette période, il faut recourir à des approches indirectes pour obtenir des renseignements sur les différents paramètres environnementaux (physiques, chimiques). Des indicateurs « indirects » ou « proxies » de ces paramètres ont ainsi été développés et utilisés en routine depuis quelques décennies.

Parmi ces indicateurs, certains sont basés sur des familles de molécules organiques synthétisées par des microorganismes. Les microorganismes ont la faculté de pouvoir modifier la composition lipidique de leurs membranes, en réponse aux variations de certains paramètres environnementaux (température, pH...) afin de conserver une fluidité appropriée à leur survie de leur membrane (Hazel et Williams, 1990). L'étude des lipides issus de la dégradation de ces membranes, lors de la mort de ces organismes, permet d'estimer les paramètres environnementaux à l'origine de la variabilité de la structure de ces composés. Ces derniers possèdent une relative résistance et peuvent donc être

INTRODUCTION

conservés dans les sédiments ou les sols, et servir de marqueurs des conditions environnementales passées. Les marqueurs moléculaires suscitent un grand intérêt dans la communauté des scientifiques depuis les années 1980.

Néanmoins, la grande majorité des biomarqueurs a été développée en milieu océanique, comme par exemple les alkyls tétraéthers de glycérol isopréniques (iGDGTs). Ces lipides sont produits par des archées qui peuvent faire varier la structure de leurs GDGTs membranaires en fonction des variations de paramètres environnementaux (Schouten et al., 2002). Un indice de températures de surface océaniques, le TEX₈₆, a ainsi pu être développé en étudiant l'abondance relative de tétraéthers produit par les *Thaumarcheota* (phylum d'archées) en milieu océanique. Le milieu continental, s'il apparaît plus accessible que les fonds abyssaux, présente des hétérogénéités de milieux et des problématiques propres aux surfaces terrestres : présence de reliefs, présence de végétation... Les biomarqueurs ont par conséquent été moins développés pour les milieux terrestres. Ainsi, les seuls paléomarqueurs moléculaires disponibles à ce jour en milieu terrestre sont les tétraéthers de glycérol ramifiés ou brGDGTs (Branched Glycerol Dialkyl Glycerol Tetraethers), différents des tétraéthers produits par les archées et produits par des bactéries non encore identifiées. Ces composés organiques sont des marqueurs du pH des sols et de la température atmosphérique. Ce sont des marqueurs ubiquistes, largement répandus dans tous les milieux, aquatiques et terrestres (Schouten et al., 2002 ; Weijers et al., 2007). Cependant, même si les brGDGTs sont de nos jours fréquemment utilisés dans le cadre de reconstructions paléo environnementales, les connaissances sur les microorganismes produisant ces composés et donc sur les paramètres qui influencent leur synthèse sont très limitées. Le développement de nouveaux marqueurs, indépendants et complémentaires des brGDGTs, est donc essentiel pour améliorer la fiabilité des reconstructions paléoclimatiques en milieu continental.

Récemment, il a été montré que d'autres lipides d'origine bactérienne, les acides gras 3-hydroxylés (3-AGH), pourraient également servir de marqueurs environnementaux dans les sols. Ces composés, qui contiennent 10 à 18 carbones et un groupement hydroxyle en 3^{ème} position, sont des constituants des lipopolysaccharides des bactéries Gram-négatif, ubiquistes dans des environnements variés, terrestres et aquatiques. Néanmoins, peu d'informations sont disponibles quant à l'effet d'une

variation des paramètres environnementaux sur la structure des 3-AGHs et sur la diversité et la réponse de leurs microorganismes sources. Dans ce contexte, mon travail de thèse avait pour objectif d'étudier l'applicabilité des 3-AGHs comme marqueurs de température et de pH en milieu terrestre, ainsi que de comprendre l'influence des paramètres environnementaux sur ces composés. Ces différents résultats ont été mis en perspective avec ceux obtenus sur les brGDGTs, utilisés comme marqueurs moléculaires de référence. Mon projet de thèse s'inscrit dans le cadre du projet EC2CO/CNRS **SHAPE (Les acides graS 3-Hydroxylés d'origine bActérienne : développEment et application de nouveaux marqueurs d'environnement en milieu continental ;** coordonné par Arnaud Huguet).

Ce manuscrit est divisé en 5 chapitres afin de répondre aux différents objectifs et enjeux de ce travail de thèse :

Le chapitre 1 présente dans un premier temps le système climatique et la nécessité de mieux le comprendre, notamment grâce à l'utilisation de marqueurs indirects d'environnements/de climats, dont les biomarqueurs organiques issus des trois domaines du vivant. En outre, les mécanismes d'adaptation membranaire des organismes face aux variations de différents paramètres de leur milieu de vie sont brièvement décrits. Enfin, dans une dernière partie, sont présentés spécifiquement les principaux biomarqueurs lipidiques de température et de pH en milieu aquatique et en milieu terrestre, en lien direct avec ce travail de thèse.

Le chapitre 2 présente les différents sites d'étude de cette thèse, ainsi que le matériel et les méthodes utilisés lors de ce travail (analyses et traitements statistiques).

Le chapitre 3 est dédié à l'étude des 3-AGH et des brGDGT au sein d'échantillons de sols des Alpes françaises. Ce chapitre est divisé en deux sous parties. Dans une première partie en anglais (article publié dans *Organic Geochemistry*, présenté en fin de manuscrit), l'influence des paramètres environnementaux (température, pH, végétation, granulométrie...) sur la distribution des 3-AGH et des brGDGTs a été étudiée dans 49 sols prélevés le long de transects altitudinaux dans les Alpes françaises (Massif du Galibier-Lautaret et des Bauges). Dans la deuxième partie (en français), sont

INTRODUCTION

présentées des expériences d'incubations en microcosmes réalisées sur quatre sols des Alpes, pendant un an à différentes températures. Ces expériences avaient pour but de fournir des informations sur la vitesse globale d'adaptation des communautés microbiennes et de leurs lipides membranaires face à un changement de température dans les sols.

Le chapitre 4 présenté en anglais (article publié dans *Biogeosciences*), vise à tester l'applicabilité des 3-AGH comme marqueurs de température et de pH dans des sols du monde entier ($n=168$). La réponse des 3-AGH aux changements de températures et de pH dans ces sols a été comparée à celle des brGDGTs, utilisés comme marqueurs «de références». L'utilisation de différents modèles de machine-learning a permis de développer les premières calibrations globales de température et de pH basées sur la distribution des 3-AGH

Le chapitre 5, en anglais (article en préparation), est basé sur les données de brGDGTs disponibles dans la littérature pour l'ensemble des échantillons de sols et de tourbières ($n=767$). Les outils statistiques développés dans les chapitres précédents ont été utilisés afin de mieux comprendre les paramètres environnementaux influençant la distribution des brGDGTs en milieu terrestre et pour développer une nouvelle calibration globale permettant d'améliorer la précision des reconstructions de température en milieu terrestre.

Enfin dans une dernière partie sont présentées les conclusions de ce travail de thèse et les perspectives qui en résultent.

Chapitre 1 – ETAT DE L'ART

1.1. INTERET ET ETUDE DES PALEOCLIMATS ET DES PALEOENVIRONNEMENTS

1.1.1. Le système climatique

Le climat est défini comme étant « l'état moyen de l'atmosphère sur une période donnée (au moins 30 ans) et au sein d'un lieu donné » (Deconinck, 2014; Renard et al., 2015). L'état moyen de l'atmosphère est caractérisé par des variables telles que les températures atmosphériques ou encore les précipitations (Deconinck, 2014; Renard et al., 2015). L'étude du système climatique peut-être réalisée dans le simple mais noble objectif d'acquérir des connaissances sur notre environnement, sur le fonctionnement global de notre planète, et participe donc d'une recherche fondamentalement désintéressée.

Néanmoins, le climat et ses variations est une des composantes essentielles du fonctionnement de notre planète, et possède donc une influence majeure sur nos sociétés. L'exemple le plus prégnant de cette influence est le réchauffement climatique associé aux changements globaux d'origine anthropique. Ainsi, l'évolution des températures depuis les années 1880 (période instrumentale, i.e. période au cours de laquelle les mesures directes des variables climatiques sont possibles et archivées) montre un réchauffement rapide et brutal des températures moyennes à la surface de notre planète (IPCC, Intergovernmental Panel on Climate Change, 5^{ème} rapport, 2014 -en français : GIEC, Groupe d'Experts Intergouvernemental sur l'Evolution du Climat). Cette élévation des températures est due à l'accroissement des émissions de gaz à effet de serre (GES) d'origine anthropique au sein de l'atmosphère. Les conséquences de cette élévation des températures sont nombreuses – fonte des glaces, élévation du niveau marin, diminution de la biodiversité, augmentation des événements météorologiques extrêmes – et affectent fortement nos sociétés : agriculture, sécheresse et famine, migrations de populations, épidémies...

La compréhension du système climatique est donc un enjeu sociétal majeur. Les mesures directes des paramètres climatiques ne sont disponibles que pour les dernières décennies. Au-delà, il est nécessaire d'utiliser des approches indirectes pour connaître et comprendre le climat. Cette compréhension des climats passés permet d'appréhender et de comprendre les climats futurs, notamment dans le cadre du changement climatique en cours.

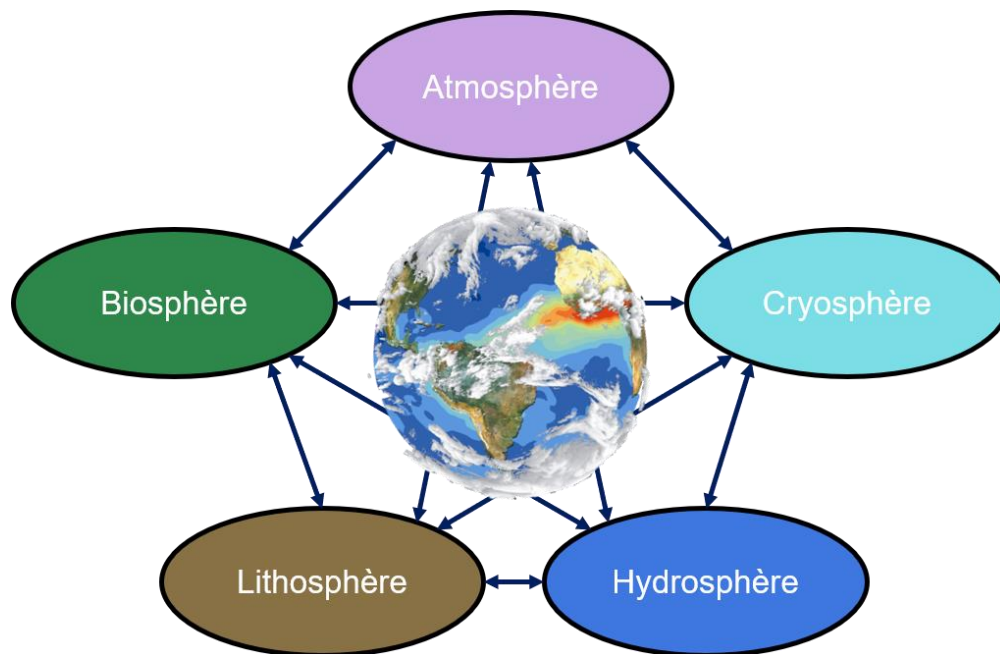


Figure 1.1 Les différents compartiments du système climatique terrestre (d'après Ruddiman, 2001)

Le climat résulte de l'interaction entre l'ensemble des compartiments du système Terre : Atmosphère, Hydrosphère, Biosphère, Cryosphère et Lithosphère. Les changements tectoniques, les variations de l'orbite de la Terre, et les variations de l'intensité solaire (Fig. 1.1) agissent sur les différents compartiments du système climatique, et sont donc à l'origine des variations du climat terrestre (Ruddiman, 2001; Deconinck, 2014; Renard et al., 2015).

Le moteur principal du climat est le Soleil, qui participe de manière prépondérante à l'apport d'énergie à la surface de notre planète (Deconinck, 2014; Renard et al., 2015). De manière plus précise, l'énergie répartie à la surface de la Terre possède deux origines distinctes (Fig. 1.2) :

- Le flux géothermique ou interne ($4,2 \times 10^{13} \text{ W m}^{-2}$)
- L'énergie solaire transmise sous forme de rayonnements électromagnétiques ($1,7 \times 10^{17} \text{ W}$)

Le bilan énergétique radiatif de la Terre est à l'équilibre, flux sortant et flux entrant étant égaux : 342 W m^{-2} (Ruddiman, 2001). D'après la loi de Planck, la température d'équilibre de la Terre devrait être en moyenne de -18°C . Pourtant, la température effective moyenne de notre planète est bien plus élevée : 15°C et ce, grâce à l'effet de serre (Fig. 1.2 ; Ruddiman, 2001). En effet, la surface terrestre réémet l'énergie reçue

sous forme de rayons infrarouges. Ces rayonnements sont absorbés par l'atmosphère puis retransmis vers la surface de la Terre, ce qui constitue le mécanisme de l'effet de serre.

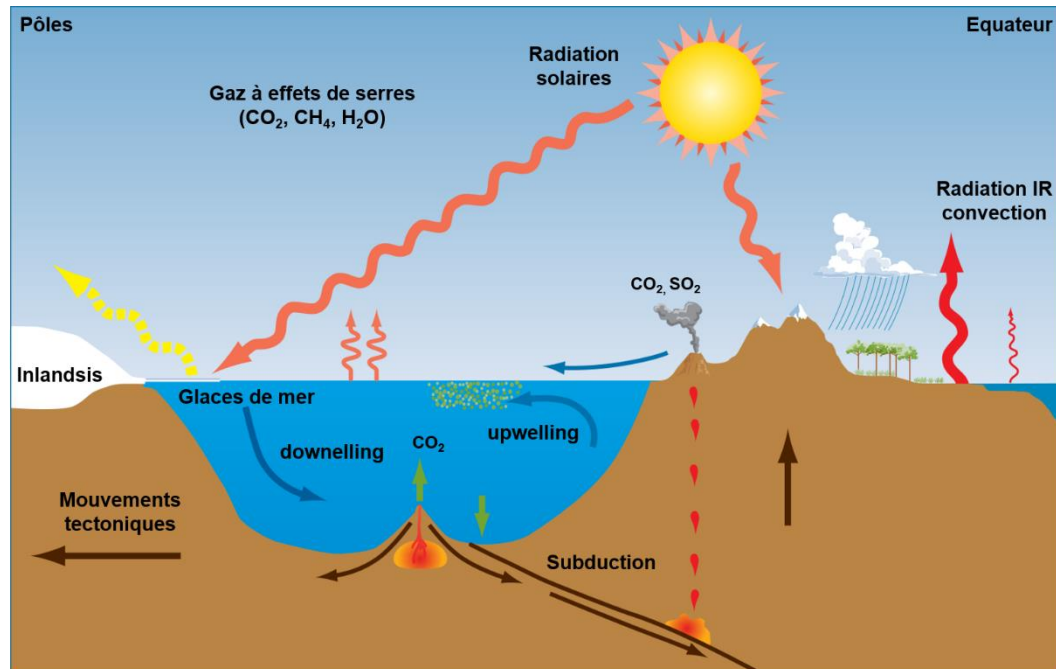


Figure 1.2 Bilan radiatif global du système Terre (d'après Ruddiman, 2001).

L'énergie reçue à la surface de la Terre au cours d'une année n'est pas la même partout : il existe des déséquilibres régionaux importants en fonction de la latitude. Cette différence d'intensité d'insolation dépend globalement de l'angle d'incidence des rayons solaires à la surface de la Terre. Ainsi, une même quantité de rayonnement se répartit sur une surface plus grande aux hautes latitudes qu'aux basses latitudes. Le déséquilibre thermique qui en résulte est régulé par une redistribution énergétique. Cette redistribution est permise par les circulations atmosphériques et océaniques entre les différentes régions du globe.

1.1.2. Changements climatiques

1.1.2.1. Origines et conséquences des changements climatiques

Il est admis que les changements globaux, et notamment les changements climatiques, ont majoritairement débuté durant la période industrielle (essor important de l'utilisation des énergies fossiles (charbons, pétrole), industrialisation et l'urbanisation croissante...). Cependant, selon plusieurs études, ces changements possèderaient une origine antique, voire préhistorique (Fig. 1.3) (IPCC, Climate Change 2014, the physical science basis). Des analyses des paléoclimats prouvent l'ampleur du changement climatique et global durant ces périodes (Fig. 1.3). Ainsi, se pose la question du début de l'augmentation des GES au sein de l'atmosphère terrestre, en particulier le méthane (CH₄) et le dioxyde de carbone (CO₂).

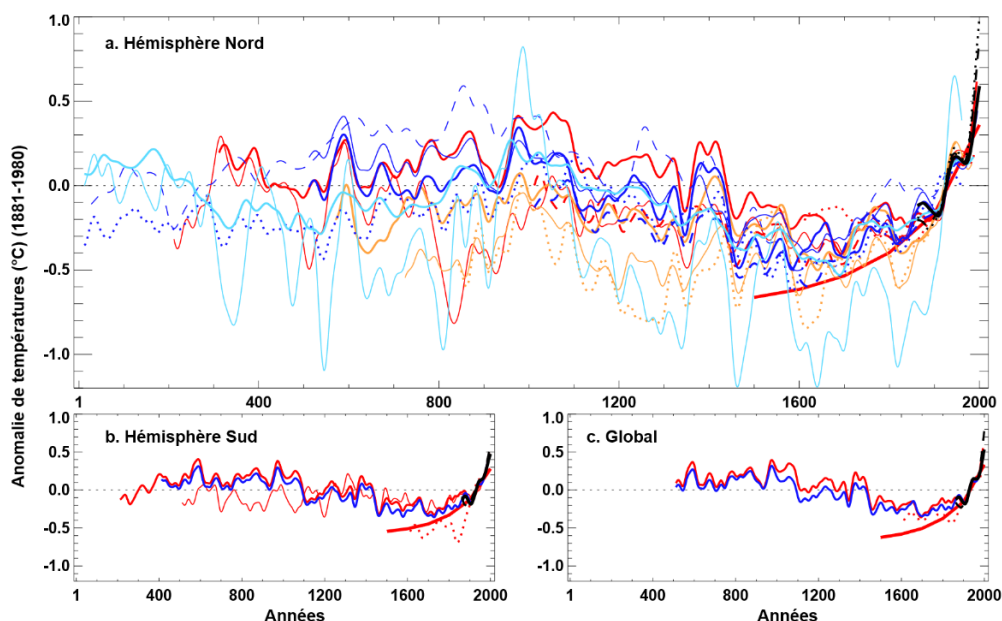


Figure 1.3 Températures annuelles au cours des 2000 dernières années pour (a) l'hémisphère nord et (b) l'hémisphère sud, et (c) les températures annuelles moyennes globales. Les différentes reconstitutions utilisées sont présentées comme indiqué dans les légendes, et regroupées par couleur en fonction de leur représentation spatiale (rouge : terre seulement – toutes latitudes ; orange : latitudes extratropicales uniquement terrestres ; bleu clair : latitudes extratropicales terrestres et maritimes ; bleu foncé: latitudes terrestres et maritimes toutes les latitudes) et les températures instrumentales sont indiquées en noir (d'après le 5^{ème} rapport de l'IPCC, *Climate Change 2014, the physical science basis*).

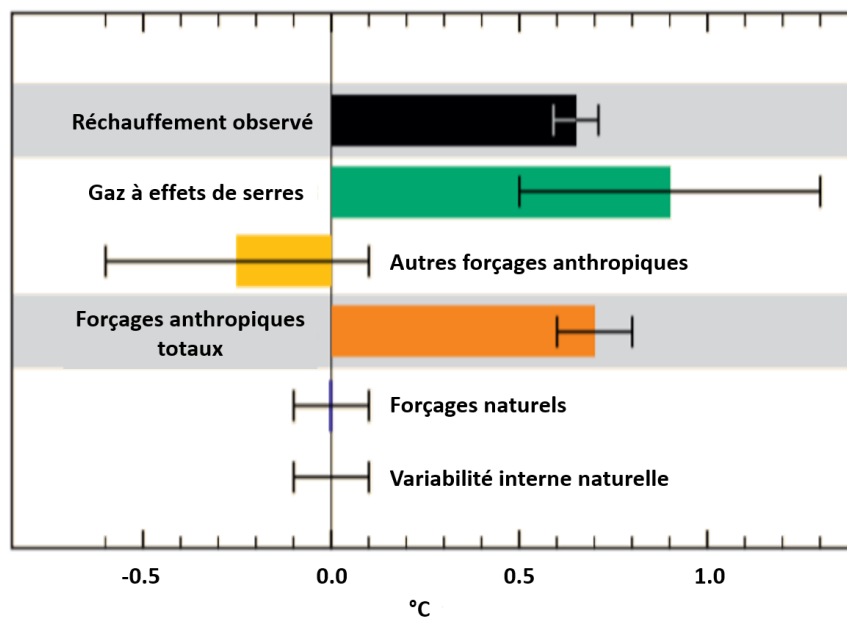


Figure 1.4 Evaluation des contributions au changement global (hausse des températures) pour la période 1951-2010 et erreurs associées. Adapté du rapport IPCC, 2014 : *Climate Change 2014 : Synthesis Report*.

Un consensus est aujourd'hui établi au sein de la communauté scientifique sur l'origine anthropiques des changements globaux (Groupe d'experts intergouvernemental sur l'évolution du climat, 2014 : *Climate Change 2014: Synthesis Report*) (Fig. 1.4). La croissance économique et démographique de nos sociétés depuis l'ère industrielle a provoqué une augmentation des émissions de GES, dont la rapidité et l'intensité sont sans précédent. Ainsi, les concentrations atmosphériques de dioxyde de carbone (CO₂), de méthane (CH₄), et d'oxyde nitreux (N₂O) connaissent des extrêmes jamais atteints depuis plusieurs centaines de milliers d'années (au moins 800 000 ans). Globalement, les émissions de GES ont augmenté au cours de la période comprise entre 1970 et 2010 et n'ont jamais été aussi importantes qu'entre 2000 et 2010. Les émissions de GES anthropiques en 2010 ont atteint $49 \pm 4,5$ GtCO₂-éq/an, tandis que la concentration en CO₂ atmosphérique mesurée au sein de l'observatoire de Mauna Loa (Hawaï) atteint une teneur record de 416,39 ppm en juin 2020 (Keeling et al., 1976 ; <https://gml.noaa.gov/>) (Fig 1.5). Les GES sont considérés comme étant une des causes principales du réchauffement climatique observé et mesuré depuis le 20^{ème} siècle.

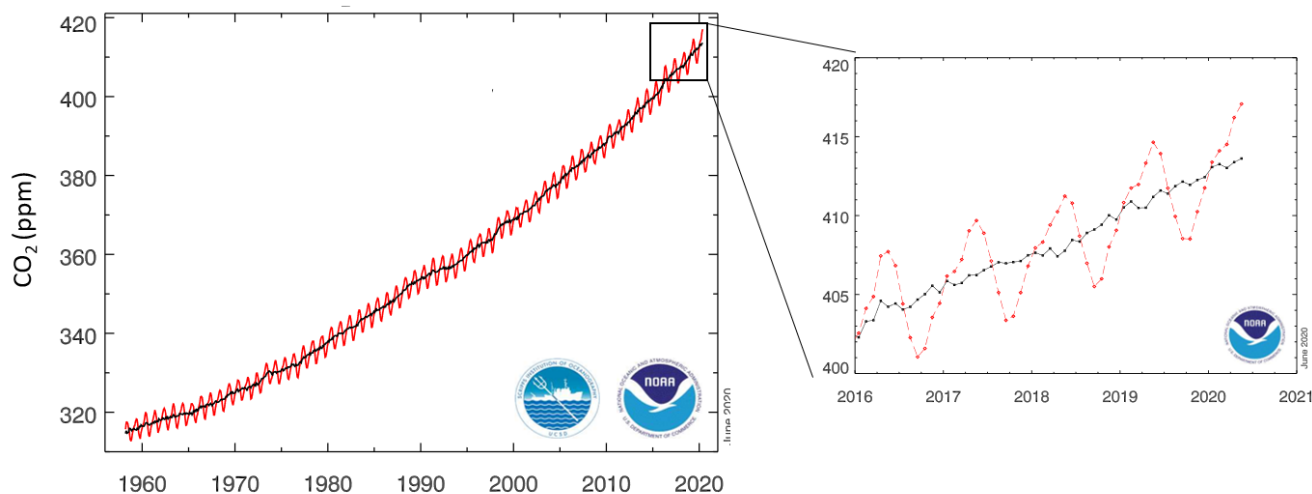


Figure 1.5 Concentrations mensuelles en CO₂ atmosphérique mesurées à l'observatoire de Mauna Loa (Hawaï) depuis 1958 jusqu'en juin 2020 (NOAA ; <https://gml.noaa.gov/>).

1.1.2.2. Effets du changement climatique et prévisions futures

D'après le 5^{ème} rapport du GIEC, édité en 2014, chacune des trois dernières décennies a été successivement plus chaude à la surface de la Terre que toutes les décennies précédentes depuis 1850. Cette période est probablement la plus chaude qu'ait connue l'hémisphère Nord au cours des 1400 dernières années. Ainsi, une moyenne globale des mesures de températures de surface atmosphérique et océanique met en lumière une hausse significative des températures de 0,85°C entre les années 1880 et 2012, sans précédent dans l'histoire récente (Fig. 1.6 a).

En parallèle de cette importante augmentation des températures, on observe une diminution des précipitations neigeuses, et donc une réduction du volume des calottes de glaces continentales et de la glace de mer (banquise). L'étendue moyenne de la banquise en Arctique a globalement diminué au cours de la période récente (1979-2012), avec un taux de 3,5 à 4,1% par période de 10 ans (IPCC, Climate Change 2014, the physical science basis). Au cours des trois dernières décennies, les calottes glaciaires du Groenland et de l'Antarctique ont également diminué, tandis que les glaciers continentaux ont reculé sur l'ensemble des continents (IPCC, Climate Change 2014, the physical science basis).

Une des conséquences majeures de la fonte des glaces continentales est l'élévation du niveau marin global. Entre 1901 et 2010, le niveau moyen mondial de la mer a augmenté de 0,19 m, et est supérieur au taux moyen des deux derniers millénaires (U.S. Global Change Research Program (2017). "Climate Science Special Report. Chapter 12: Sea Level Rise»; IPCC, *Climate Change 2014: Synthesis Report*) (Fig 1.6 b).

De plus, l'augmentation des températures atmosphériques participe de façon majeure à celle des océans. En effet, il est estimé que 90% de l'énergie thermique accumulée au cours des dernières décennies est stockée au sein des océans (IPCC, *Climate Change 2014 : Synthesis Report*). Ce réchauffement océanique global a de nombreuses conséquences néfastes : acidification, fonte des glaciers, hausse du niveau marin, relargage du CO₂ ... (IPCC, *Climate Change 2014 : Synthesis Report*). La dilatation thermique contribue à l'élévation du niveau océanique global à hauteur de 42% (pour la période entre 1993 et 2018 ; (WCRP Global Sea Level Budget Group, 2018)), suivie par la fonte des glaciers tempérés, du Groenland, et de l'Antarctique (WCRP Global Sea Level Budget Group, 2018) .

Ensuite, depuis le début de l'ère industrielle, l'absorption de CO₂ par les océans a augmenté. A l'heure actuelle, les océans ont un rôle majeur dans le stockage du carbone, notamment du CO₂ atmosphérique. L'augmentation du taux de CO₂ atmosphérique (Fig. 1.6c et d) a provoqué une acidification des océans : le pH de l'eau de surface a diminué de 0,1 (Howes et al., 2015), ce qui peut avoir des conséquences néfastes, notamment sur la biodiversité marine (Shaw et al., 2015; Zeng et al., 2015; Yang et Hansson, 2016). Néanmoins, l'augmentation de la température des eaux de surfaces océaniques diminue le potentiel de stockage du carbone par les eaux et augmente le relargage du CO₂ par l'océan (IPCC, *Climate Change 2014 : Synthesis Report*).

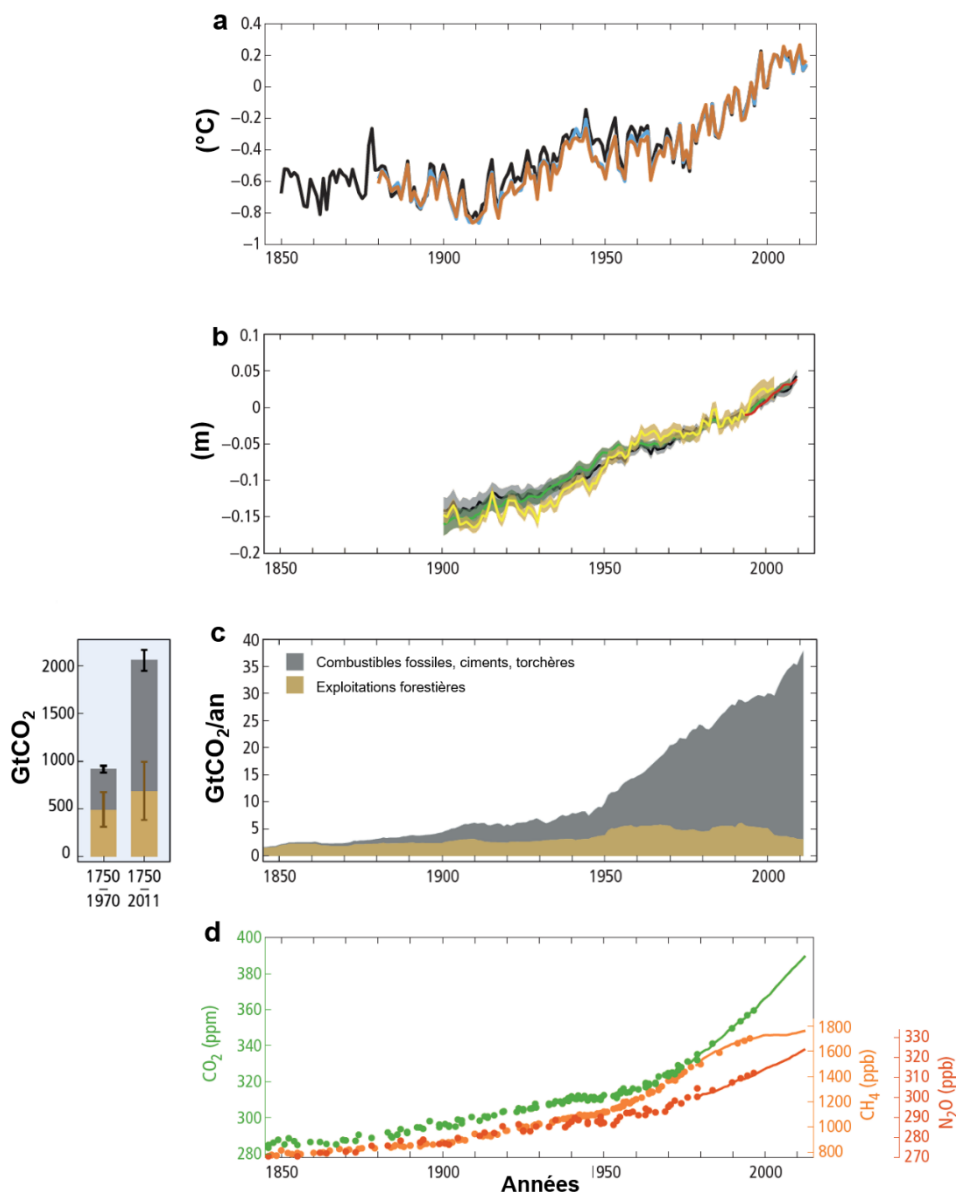


Figure 1.6 Observations du changement global et climatique. a) Anomalies de la température de surface des terres et des océans combinées, moyennes annuelles et mondiales par rapport à la moyenne sur la période 1986-2005. b) Variation annuelle et mondiale moyenne du niveau de la mer par rapport à la moyenne sur la période 1986-2005. c) Émissions mondiales de CO₂ anthropique provenant de la sylviculture et d'autres utilisations des terres ainsi que de la combustion de combustibles fossiles, de la production de ciment et des torchères. d) Concentrations atmosphériques des gaz à effet de serre dioxyde de carbone (CO₂, vert), le méthane (CH₄, orange) et l'oxyde nitreux (N₂O, rouge) déterminées à partir de données de carottes de glace (points) et de mesures atmosphériques directes (lignes). Adapté du rapport IPCC, 2014: *Climate Change 2014: Synthesis Report*.

Face aux changements d'origine anthropique, les membres du GIEC proposent plusieurs scénarios d'évolution du climat de notre planète (modèles RCP pour Representative Concentration Pathways), dans chaque édition du rapport (Fig. 1.7) :

- RCP 2.6 : contrôle des émissions des GES et autres changements globaux
- RCP 4.5 et 6.0 : stabilisation des émissions
- RCP 8.5 : fortes émissions

Le changement de température moyenne du globe en surface pour la fin du 21^{ème} siècle dépassera 1,5°C pour l'ensemble des modèles, sauf le scénario 2.6 (Fig. 1.7a). Les données paléoclimatiques soutiennent cette hypothèse, en montrant des réchauffements, notamment au niveau des pôles lors de périodes d'augmentation de la concentration atmosphérique en CO₂. Comme pour la température, tous les modèles RCP, sauf le 2.6, montrent une diminution des étendues de banquises au sein de l'hémisphère Nord (Fig. 1.7b), tandis que le niveau moyen des mers continuera probablement d'augmenter au cours du 21^{ème} siècle (Fig. 1.7c).

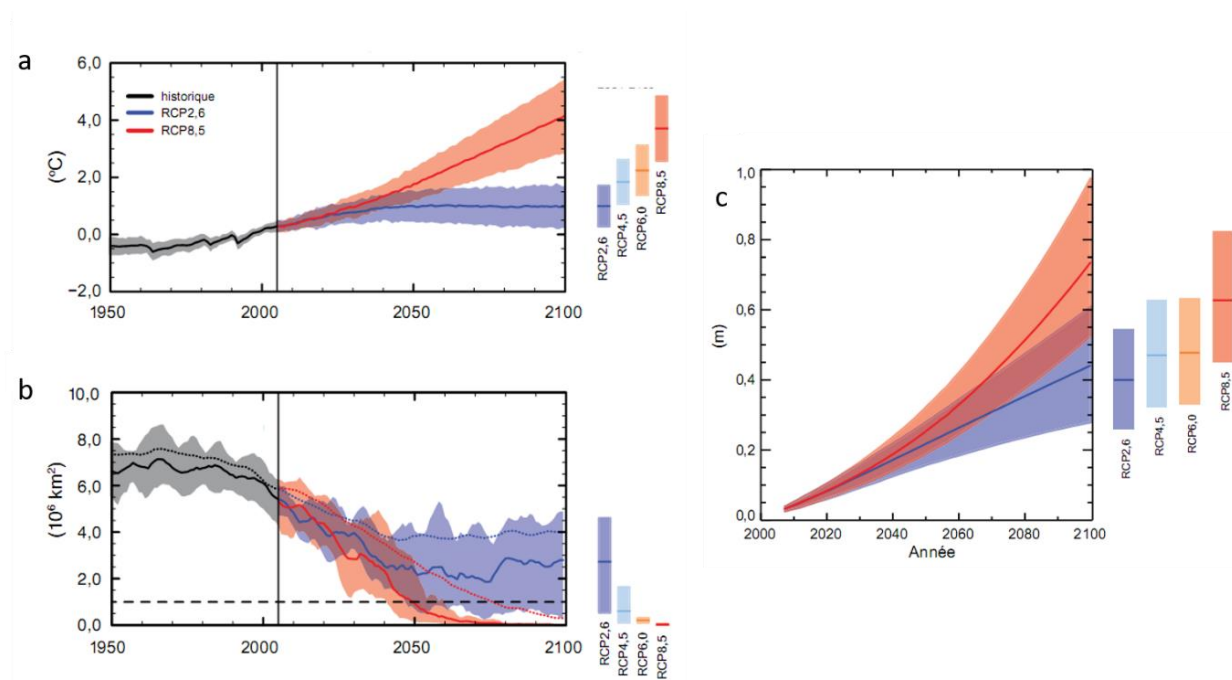


Figure 1.7 Différents modèles de l'évolution possible des a) la température moyenne à la surface du globe ; b) l'étendue de la banquise de l'hémisphère Nord en septembre ; c) l'élévation du niveau moyen des mers à l'échelle du globe. Adapté d'après IPCC 2014, the physical science basis, 5^{ème} rapport.

L'ensemble de ces données prouve sans équivoque l'ampleur du changement global et son impact sur nos civilisations. C'est donc une nécessité de le comprendre et d'y remédier. S'il est aujourd'hui établi que le changement global a pour principales causes

les activités anthropiques, il est encore nécessaire d'étudier les changements passés, afin de mieux contraindre les changements à venir. En effet, les différents scénarios du GIEC reposent sur l'utilisation de modèles, principalement alimentés par les données environnementales et climatiques disponibles des périodes passées. Comprendre la variabilité climatique et environnementale passée est donc capital pour pouvoir appréhender les changements futurs.

1.1.3. Etude des Paléoclimats et Paléoenvironnements

La communauté des paléoclimatologues, paléocéanographes et paléo-environmentalistes cherche à mieux contraindre la variabilité climatique, par l'étude des climats et environnements anciens. Dans ce but, il est nécessaire d'avoir accès aux données environnementales passées. Mais comment reconstituer les climats passés ? Comment avoir accès aux différentes variables environnementales, permettant de comprendre et de décrire les paléoclimats et paléoenvironnements ? Comment connaître les variations de la température moyenne annuelle de l'air (TMAA), de l'humidité, des précipitations en milieu continental, ou encore de la température de surface des océans et de la salinité en milieu océanique ?

Des mesures directes de ces données sont disponibles dans le meilleur des cas pour les deux derniers siècles, et plus généralement pour les dernières décennies. Au-delà de cette période, il est nécessaire d'utiliser des approches indirectes pour obtenir ces renseignements. A l'aide de différents supports, ou archives, il est possible de reconstruire ces différents paramètres. Deux grands types d'archives peuvent être distingués :

- des archives biologiques, c'est-à-dire des restes d'organismes, indicateurs des conditions environnementales régnant au cours de leurs vies,
- des archives sédimentaires/glaciaires, permettant de reconstituer les variations paléoclimatiques et paléoenvironnementales grâce à leur type de dépôt (continu ou discontinu), la nature de ce dépôt (marin, lacustre...), leur contenu (restes fossilisés, éléments terrigènes...), leur nature minéralogique (carbonatés, siliceuses...) ou leur contenu géochimique... (Ruddiman, 2001; Deconinck, 2014; Renard et al., 2015) .

1.1.3.1. *Des outils pour reconstituer les climats passés*

Le terme anglais « proxy » se réfère aux indicateurs indirects des conditions climatiques et environnementales, et est défini comme suit : “something that you use to represent something else that you are trying to measure or calculate” (Oxford dictionary). Dans la suite de ce manuscrit, pour désigner ces indicateurs, les termes « marqueur » ou « indicateur » sont utilisés dans les parties en français, tandis que dans les parties en anglais sera utilisé le terme « proxy ».

Les climatologues utilisent majoritairement deux grands types d'indicateurs pour les paléoreconstructions : (1) les marqueurs d'origine biologique, qui sont des restes préservés ou fossilisés d'organismes végétaux ou animaux, et (2) des marqueurs « géochimiques » (Fig 1.8).

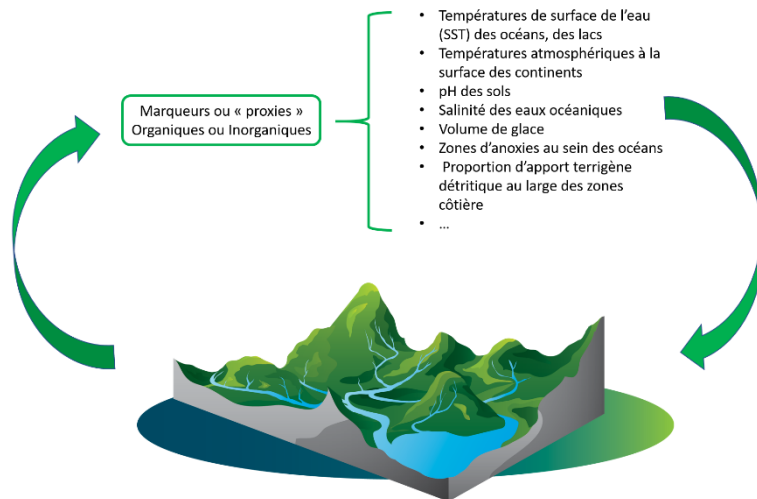


Figure 1.8 De l'intérêt des marqueurs : reconstituer de façon indirecte des conditions environnementales et/ ou des paléoenvironnements (source de l'image : <http://www.bvsm.ca>)

Les marqueurs d'origine biologique sont de très bons candidats pour renseigner les climatologues sur les variations climatiques et environnementales passées, en milieu océanique ou continental. En effet, la diversité écologique des organismes végétaux et animaux est très fortement dépendante des paramètres climatiques et environnementaux (Ruddiman, 2001; Gornitz, 2008). Néanmoins la plupart des organismes qui ont existé sur Terre sont aujourd'hui éteints, et plus on remonte dans le temps, moins les restes fossilisés sont reconnaissables. Ainsi, l'utilisation des marqueurs « biologiques » sur de longues échelles de temps (millions à plusieurs dizaines de millions d'années) nécessite souvent de se fier à la ressemblance des formes passées avec leurs homologues modernes (Ruddiman, 2001).

Au sein des milieux océaniques, quatre grands groupes de plancton animal et végétal formant des exosquelettes (coquilles ou tests) sont utilisés majoritairement pour

les reconstitutions des climats passés. Ces quatre groupes peuvent former des tests constitués de carbonate de calcium (CaCO_3), comme les foraminifères (planctoniques ou benthiques ; Ravelo et Hillaire-Marcel, 2007) et les coccolithophoridés, ou des tests siliceux comme les diatomées (par exemple : Hasle et al., 1996; Crosta et Koç, 2007; Round et al., 2007) et les radiolaires (par exemple : Kling, 1998; Cortese et Abelmann, 2002).

En milieu continental, les reconstitutions paléoenvironnementales sont effectuées majoritairement à l'aide des grains de pollens préservés dans les archives sédimentaires lacustres ou terrestres (par exemple : (Reille and de Beaulieu, 1990; Sánchez Goñi, 2006), des macrorestes végétaux ou des restes fossilisés d'insectes, les chironimides (par exemple : Gandouin et al., 2007). Les restes végétaux jouent donc un rôle central dans les paléoreconstructions, notamment au travers de l'étude de macrorestes (cônes, graines, feuilles) dont les caractéristiques peuvent renseigner sur les conditions climatiques passées. Par exemple, les climats plus chauds d'il y a des dizaines de millions d'années sont déduits de la présence de palmiers fossilisés aux hautes latitudes nordiques (Greenwood et West, 2016). Les climatologues utilisent aussi les assemblages de grains de pollen déposés dans les sédiments comme indicateurs de l'abondance relative des types de végétation (arbres, herbes...). Les grains de pollen produits par la végétation, sont déposés puis préservés dans des sédiments lacustres ou marins, à la suite de leur transport éolien ou aquatique (Ruddiman, 2001). En milieu lacustre, des exosquelettes de microorganismes peuvent fournir des informations sur les conditions paléoenvironnementales, comme par exemple les ostracodes (Holmes, 2000; Horne et al., 2002; Martens et al., 2008) ou encore certaines espèces de diatomées présentes en milieux lacustres (Leng and Barker, 2006). Enfin, l'étude de restes carbonisés de végétaux peut nous renseigner sur les régimes de feux de forêts passés et leurs fréquences (Power et al., 2008).

Néanmoins, l'étude et l'observation des restes « biologiques » fossilisés n'est pas la seule méthode utilisée pour reconstruire les variations climatiques et environnementales passées. Les techniques d'analyses géochimiques permettent d'apporter des informations complémentaires sur l'étude des sédiments, réels témoins des variations des conditions physico-chimiques de l'atmosphère et de l'hydrosphère. L'analyse de rapports d'isotopes stables ($^{18}\text{O}/^{16}\text{O}$, $^{13}\text{C}/^{12}\text{C}$, $^2\text{H}/^1\text{H}$) au sein des restes

fossilisés d'organismes, permet d'obtenir des informations paléoenvironnementales/climatiques mais aussi sur les conditions de vie et de biosynthèse des organismes. Ainsi, le rapport des isotopes stables de l'oxygène $\delta^{18}\text{O}$ (Emiliani, 1955) dans les restes carbonatés de microorganismes (par exemple, les foraminifères), en équilibre avec les conditions physico-chimiques de leurs milieux, constitue un des marqueurs les plus utilisés pour reconstituer les variations de températures passées. Dans le même ordre d'idée, les fluctuations de teneur en éléments traces (magnésium, strontium, manganèse...) au sein de ces restes carbonatés peuvent renseigner sur les variations des conditions environnementales au niveau local comme au niveau global (Brown and Elderfield, 1996; Mitsuguchi et al., 1996; Anand et al., 2003).

L'étude à l'échelle globale et moléculaire de la matière organique permet d'obtenir des informations complémentaires à celles obtenues à partir des marqueurs basés sur les restes fossilisés d'organismes. Ainsi, l'analyse isotopique ($\delta^{13}\text{C}$; $\delta^{15}\text{N}$) et élémentaire (C, N) de la fraction organique des sédiments renseigne sur l'origine de la matière organique, mais aussi sur son degré de dégradation (Meyers, 2003). Par ailleurs, l'analyse de molécules organiques spécifiques, appelés biomarqueurs, permet d'avoir accès à des informations sur les organismes sources de ces composés, sur leurs conditions de vie ou encore sur leurs voies de biosynthèses.

1.1.3.2. *Concept de biomarqueur*

Les biomarqueurs moléculaires sont des molécules organiques biosynthétisées par des organismes avant d'être libérées dans l'environnement, par exemple lors de la mort de l'organisme (Killops, 2005; Eglinton and Eglinton, 2008). Elles ont l'avantage de présenter une spécificité biologique et de pouvoir renseigner sur la nature des organismes source de ces composés mais aussi sur leurs conditions de vie et les conditions environnementales et climatiques associées. Néanmoins, la conservation de ces informations dans le temps est liée à la préservation des molécules.

Ainsi, un « bon » biomarqueur moléculaire doit présenter une bonne stabilité diagenétique, i.e. être résistant à la dégradation. Les biomarqueurs moléculaires peuvent être comparés à des « fossiles ou empreintes » moléculaires. Comme pour les fossiles, dont la conservation va dépendre du lieu d'enfouissement de l'organisme, de la matière

minérale environnante et des conditions environnementales, les biomarqueurs moléculaires et leur stabilité dans l'environnement dépendent des conditions environnementales de transport et d'enfouissement (transport le long de bassins versants, parfois jusqu'à l'océan, transport éolien, piégeage au sein d'archives sédimentaires ou de sols, liaisons avec des complexes macromoléculaires (parois cellulaires, matière organique, matière minérale...)).

L'étude des molécules organiques préservées dans les archives sédimentaires permet de reconstituer les variations de différents paramètres environnementaux. Dans le cadre de cette thèse, un regard particulier sera porté sur les indicateurs organiques de température et de pH en milieu continental, et plus particulièrement au sein des archives terrestres (sols, loess, tourbières...).

1.2. Diversité et adaptabilité du vivant

L'enjeu majeur de cette thèse est l'étude et le développement de marqueurs d'environnement d'origine microbienne. Après une brève introduction à la classification du vivant, les caractéristiques des membranes microbiennes et les mécanismes d'adaptation membranaire en jeu face aux variations de différents paramètres environnementaux (notamment la température et le pH) seront présentées.

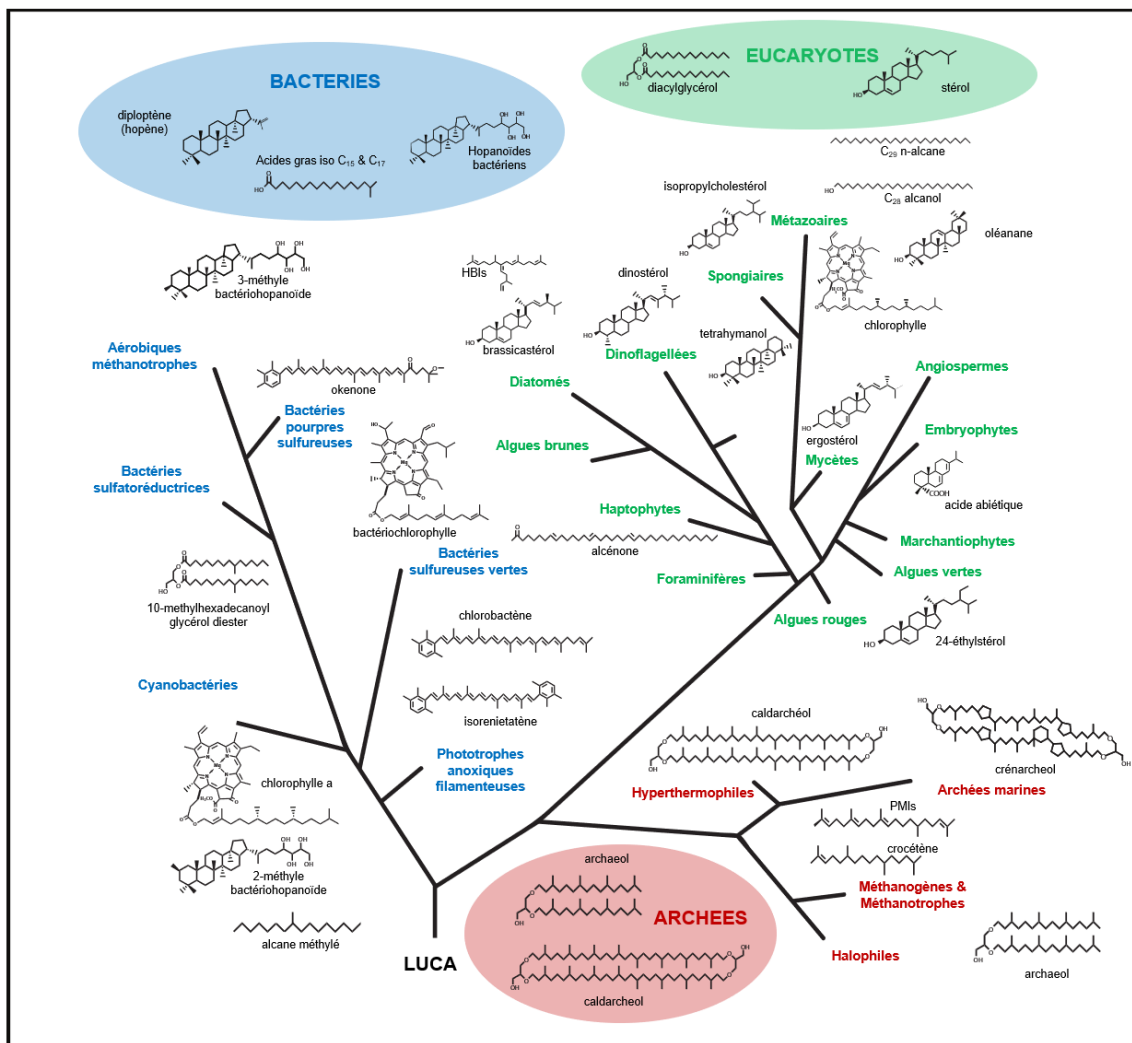


Figure 1.9 Représentation de l'arbre du vivant, basé sur l'étude des séquences de nucléotides des sous unités d'ARN ribosomique. Cet arbre illustre les trois grands domaines du vivant et leurs composés lipidiques caractéristiques. Adapté de Bianchi et Canuel, 2011 ; Briggs et Summons, 2014 et références incluses.

Une première tentative de classification moderne du Vivant (Whittaker et Margulis, 1978) proposait de rassembler les organismes en deux grands domaines : Eucaryotes et Procaryotes. Au sein des Eucaryotes, étaient alors rassemblés quatre grands règnes : les plantes, les animaux, les champignons et les protistes, et au sein des Procaryotes se trouvaient les bactéries. Néanmoins, cette classification ne faisait pas consensus au sein de la communauté scientifique. Le développement de l'étude du génome, et plus globalement de la biologie moléculaire a permis des avancées majeures sur la compréhension de la classification du vivant (Zuckerandl et Pauling, 1965; Margulis, 1967; Woese et Fox, 1977; Woese et al., 1990; Gray, 2017). De nos jours, la classification phylogénique propose une répartition de l'ensemble des organismes en 3 grands domaines : Archées, Bactéries et Eucaryotes (Bianchi and Canuel, 2011; Briggs and Summons, 2014) (Fig. 1.9) et en 7 grands règnes (archées, bactéries, protozoaires, chromistes, plantes, champignons et animaux). Cependant, la complexité du vivant (fusion, phagocytose, symbiose, transferts horizontaux de gènes), ainsi que la nature des protagonistes bactériens et archéens complexifie l'établissement d'une classification du vivant exhaustive et définitive. Cette question cruciale fait l'objet de très nombreux débats dans la communauté scientifique depuis plus de 20 ans et est loin d'être résolue. La plus grande prudence et ouverture d'esprit sont donc requises (Fitch and Upper, 1987; Lombard et al., 2012; Villanueva et al., 2020). Cette courte introduction montre la grande complexité du vivant, et la nécessité de le comprendre. Outre les études phylogénétiques, les études des biomolécules fossilisées (biomarqueurs) sont un outil puissant pour révéler l'histoire du Vivant, acquérir des connaissances sur les environnements passés, ou encore des indices sur les processus microbiens anciens.

Par exemple, la plupart des eucaryotes synthétisent des stérols au sein de leurs membranes, composés pouvant être d'une grande spécificité vis-à-vis d'un groupe taxonomique (Briggs et Summons, 2014) (Fig.1.9). Autre exemple, les membranes des bactéries et des eucaryotes simples sont composées de lipides constitués de chaînes d'hydrocarbures linéaires ou ramifiées liées à un groupement glycérol par des liaisons ester (i.e. acides gras). Les archées, en revanche, synthétisent des structures équivalentes en utilisant les chaînes isopréniques qui sont liées au glycérol par des liaisons éther. Beaucoup de ces lipides peuvent être conservés dans les roches sédimentaires et peuvent alors offrir des renseignements sur les conditions environnementales passées.

L'enjeu majeur de cette thèse étant le développement de marqueurs microbiens d'environnements (notamment de température et de pH), une courte présentation des 3 grands domaines du vivant est effectué dans la partie suivante.

1.2.1. Domaine des Eucaryotes

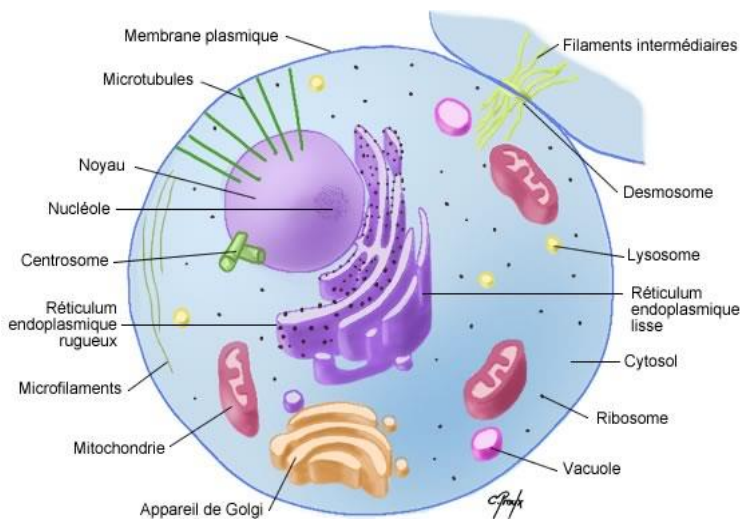


Figure 1.10 Schéma d'une cellule eucaryote (source : Chantal Proulx).

Le terme eucaryote, à opposer à celui de procaryote, est issu du grec (*eu* = « vrai » ; *karyos* = « noyau ») et qualifie donc les organismes dont les cellules possèdent un noyau renfermant le matériel génétique de la cellule, les chromosomes (Whittaker and Margulis, 1978; Fig 1.10).

Basée sur les découvertes en phylogénie moléculaire, la classification des eucaryotes reste l'objet de nombreux débats (transferts horizontaux de gènes et de processus d'endosymbioses depuis des bactéries et des archées (Vellai et Vida, 1999; Margulis, 1967; Zilber-Rosenberg et Rosenberg, 2008; Gray, 2017) et plusieurs auteurs ont proposé au fil des années des classifications différentes de ces organismes (e.g. Cavalier-Smith, 1998; Adl et al., 2005; Adl et al., 2012; Boudouresque, 2015; Adl et al., 2019).

Les eucaryotes sont des organismes uni- ou multicellulaires. Les cellules des organismes eucaryotes sont généralement plus grandes que celles des procaryotes, en moyenne de 2 à 20 μm , certaines cellules pouvant mesurer jusqu'à 600 μm de diamètre (Madigan et Martinko, 2018 ; Reece et al., 2014 ; Fig. 9a). La principale caractéristique de ces cellules est la structuration complexe de leur cytoplasme avec la présence d'un cytosquelette constitué de filaments d'actines et de microtubules. De plus, le cytoplasme contient des structures spécialisées – des organites – délimitées du reste de la cellule par une membrane phospholipidique. Il existe de nombreux types d'organites au sein des

cellules eucaryotes, comme les mitochondries chez les métazoaires, ou les chloroplastes chez les végétaux, responsables de la production d'énergie des cellules. Les cellules eucaryotes végétales peuvent posséder une paroi autour de la membrane, constituée de cellulose et d'autres polysaccharides. Les cellules des champignons possèdent une paroi cellulosique contenant aussi d'autres polysaccharides et notamment de la chitine. Enfin les organismes eucaryotes peuvent présenter des appendices associés à la motilité, comme des cils ou des flagelles.

1.2.2. Domaine des Archées

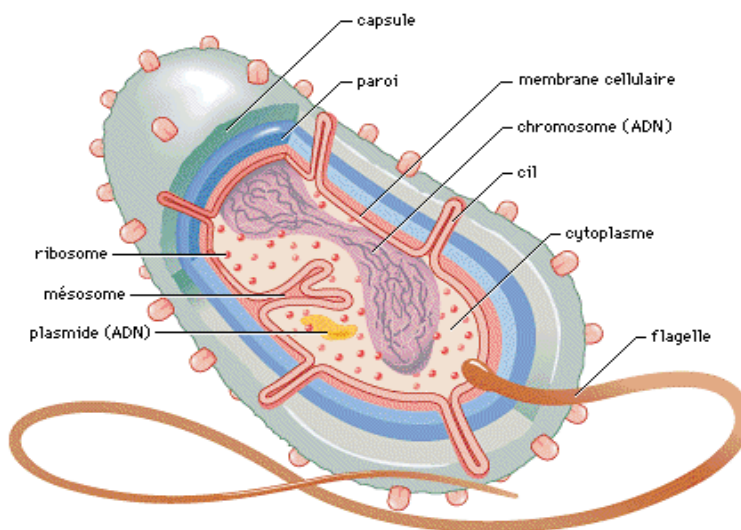


Figure 1.11 Schéma d'une cellule procaryote (source : Chantal Proulx).

Les archées sont des organismes procaryotes (Fig. 1.11), ubiquistes et unicellulaires.

Initialement, ces organismes étaient confondus avec les bactéries du fait de leur absence de noyau. Des travaux sur les procaryotes réalisés dans les années 60-70 ont permis de

scinder les bactéries et les archées en deux embranchements bien distincts (Woese et Fox, 1977). Initialement considérées comme extrémophiles strictes, les Archées ont ensuite été identifiées dans des environnements très variés, comme des sols, des tourbières mais aussi des environnements lacustres et marins (Schouten et al., 2013 et références incluses).

La classification communément admise propose que les archées soient réparties en 5 grands phyla principaux, à partir de souches d'archées cultivables (Guy and Ettema, 2011) :

- *Euryarchaeota* (Makarova et al., 1999) (halophiles et méthanogènes)
- *Crenarchaeota* (hyperthermophiles)
- *Thaumarchaeota* (Brochier-Armanet et al., 2008) (organismes mésophiles)
- *Korarchaeota* (Barns et al., 1996; Elkins et al., 2008)

- *Nonarchaeota* (Huber et al., 2003).

De nouvelles souches d'archées, proches des eucaryotes (Spang et al., 2015, 2018; Zaremba-Niedzwiedzka et al., 2017), voire formant un seul ensemble monophylétique avec les eucaryotes (Fournier et Poole, 2018) ont récemment été découvertes, remettant ainsi en question cette vision du vivant et de ses origines.

Les archées partagent des points communs avec les eucaryotes et les bactéries. En effet, les archées possèdent de nombreuses similitudes morphologiques avec les bactéries, comme l'absence de noyau, des structures cellulaires similaires (en cocci, bâtonnets, spirales...) ou encore la présence de flagelle assurant la motilité. Les archées, comme les bactéries, peuvent posséder une paroi assurant la protection de la cellule, notamment contre le stress osmotique. De la même façon, les eucaryotes et les archées présentent des mécanismes de réplication, de transcription de l'ADN en ARN et de traduction de l'ARN message en protéines similaires. Cependant, les archées possèdent aussi des caractéristiques uniques, absentes des bactéries ou des eucaryotes, comme certaines voies métaboliques (méthanogenèse), l'absence d'acides gras synthase, la structure en monocouche de leurs membranes plasmiques, constituées essentiellement de lipides de type éther de glycérol (au lieu d'esters de glycérol pour les eucaryotes ou la plupart des bactéries), et enfin leur paroi cellulaire constituée de pseudopeptidoglycane (pseudomuréine) (Madigan et Martinko, 2018 ; Reece et al., 2014).

1.2.3. Domaine des Bactéries

Les Bactéries ou Eubactéries sont des micro-organismes ubiquistes procaryotes (Fig. 1.11) et dans la majorité des cas unicellulaires, dont une des caractéristiques principales est l'absence de noyau renfermant l'information génétique. L'information génétique des bactéries est contenue dans un seul chromosome et dans des plasmides (ADN circulaire). Les bactéries « vraies » ont été renommées Eubactéries pour établir leur différence avec les Archées. A ce jour, au moins 84 phyla de bactéries ont été découverts par séquençage du gène codant l'ARN ribosomique 16S (ARNr 16S), unité ribosomale spécifique des procaryotes, et donc des bactéries (Woese and Fox, 1977). L'ensemble de ces phyla recouvre une très grande diversité d'espèces, présentant une importante diversité de morphologies

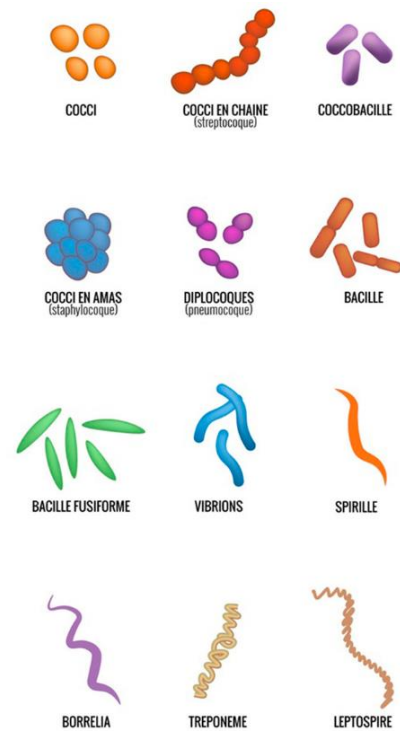


Figure 1.12 Quelques exemples de la diversité morphologiques des organismes procaryotes

(Fig. 1.12) et de métabolismes. Néanmoins il existe des caractéristiques communes à ce domaine. La plupart des bactéries ont une taille comprise entre 0,5 μm et 5 μm (Lecointre et Guyader, 2006), même si des espèces de bactéries de grande taille, pouvant mesurer jusqu'à 750 μm , ont été découvertes (Schulz et Jørgensen, 2001). La majorité des bactéries présente des formes sphériques ou en bâtonnet, mais une très grande diversité de formes peut être rencontrée comme des formes incurvées, spiralées ou hélicoïdales. De plus, si les bactéries peuvent être observées sous forme unicellulaire, elles peuvent aussi se regrouper en amas ou former des chaînes ou des feuillets (Madigan et Martinko, 2018 ; Fig. 1.12). Ces quelques exemples de caractéristiques bactériennes suggèrent la difficulté à discriminer ces organismes. Néanmoins, il existe différentes méthodes de classification des bactéries. Une première classification proposée par Carl von Linné (1735) répartit ces organismes en différents niveaux (règne, embranchement, famille, genre, espèce). Elle repose principalement sur les différences morphologiques des bactéries. Il est aussi possible de classifier les bactéries en fonction de leurs besoins

métaboliques, notamment ceux en oxygène, divisant ainsi les bactéries en organismes aérobies ou anaérobies (Madigan et Martinko, 2018).

Une autre classification possible, utilisée pour caractériser les eubactéries, repose sur la réaction de ces organismes au contact du test de coloration de Gram (Gram, 1884). Ce test permet d'identifier deux grands types de parois chez l'ensemble des bactéries : les bactéries Gram-positif et les bactéries Gram-négatif. Cependant, il convient de préciser que les groupes ainsi formés ne sont pas des ensembles monophylétiques, et il ne s'agit donc pas d'un classement phylogénétique des bactéries.

Le cytosol (espace intramembranaire) des bactéries est soumis à une pression osmotique importante (5 à 20 atmosphères ; Fig. 1.10). La paroi des bactéries possède donc une double fonction : empêcher l'éclatement de l'organisme sous l'effet de l'importante pression osmotique, et protéger la bactérie du milieu extérieur. Pour assurer ces fonctions, la paroi des bactéries est constituée de différentes familles de molécules. On retrouve notamment dans la paroi cellulaire des bactéries un polymère, la muréine, aussi appelé peptidoglycane. De plus, certaines bactéries peuvent biosynthétiser des couches moléculaires supplémentaires à la surface de la paroi, constitués par exemple de polysaccharides ou encore de protéines (couche S). Il est également possible d'observer chez certaines bactéries des organites extracellulaires comme des flagelles ou des cils servant à la mobilité de l'organisme (Madigan et Martinko, 2018).

Dans le cadre de cette étude, un intérêt particulier sera porté à l'étude des bactéries Gram-négatif, organismes sources des acides gras 3-hydroxylés. Cependant, afin de mieux appréhender les caractéristiques de ce groupe non monophylétique, une description des bactéries Gram-positif sera aussi effectuée.

• Bactéries Gram-négatif vs. Gram-positif

La classification des bactéries Gram-positif et Gram-négatif repose sur la structure de l'enveloppe cellulaire de ces organismes. Le peptidoglycane, présent dans la paroi des bactéries (Fig. 1.13), est un hétéropolymère composé de polysaccharides (les unités N-acetylglucosamine et les N-acetylmuramique) et de chaînes peptidiques assurant la liaison entre les unités glucidiques. Cette macromolécule unique forme un maillage qui n'a jamais été retrouvée dans les parois cellulaires des archées ou des eucaryotes. Son abondance varie selon les bactéries, ce qui conduit à la différenciation entre bactéries Gram-positif et Gram-négatif.

En effet, la paroi des bactéries Gram-positif est composée d'une épaisse couche de peptidoglycane (jusqu'à 90% de la paroi ; Fig. 1.13a) ou même d'un empilement de feuillet de peptidoglycanes (Navarre et Schneewind, 1999). Des acides téichoïques, qui sont des polymères de glycérol et de ribitol, associés à des unités phosphates sont également ancrés dans la membrane.

La structure de la paroi cellulaire des bactéries Gram-négatif est plus complexe que celle des bactéries Gram-positif (Madigan et Martinko, 2018). Celle-ci possède de l'intérieur vers l'extérieur (Fig. 1.13b) :

- une membrane externe (7-8 nm d'épaisseur), en contact avec le cytoplasme de l'organisme
- une mince couche de peptidoglycane (3 nm d'épaisseur)
- une membrane externe (7-8 nm d'épaisseur), en contact direct avec le milieu extérieur, principalement composée de phospholipides organisés en bicouche, et qui contient de nombreux complexes protéiques intrinsèques. Ces protéines membranaires assurent la cohésion de la membrane et des fonctions diverses, comme la perméabilité sélective des nutriments indispensables à la survie de la cellule (porine). La membrane externe est successivement hydrophile (polysaccharides), hydrophobe (lipide A + phospholipides) et enfin hydrophile (têtes polaires des phospholipides). A la surface de la membrane externe d'une grande majorité des bactéries Gram-négatif, on retrouve des lipopolysaccharide (LPS). Les LPS sont le composant principal de la membrane externe des bactéries Gram-négatif.

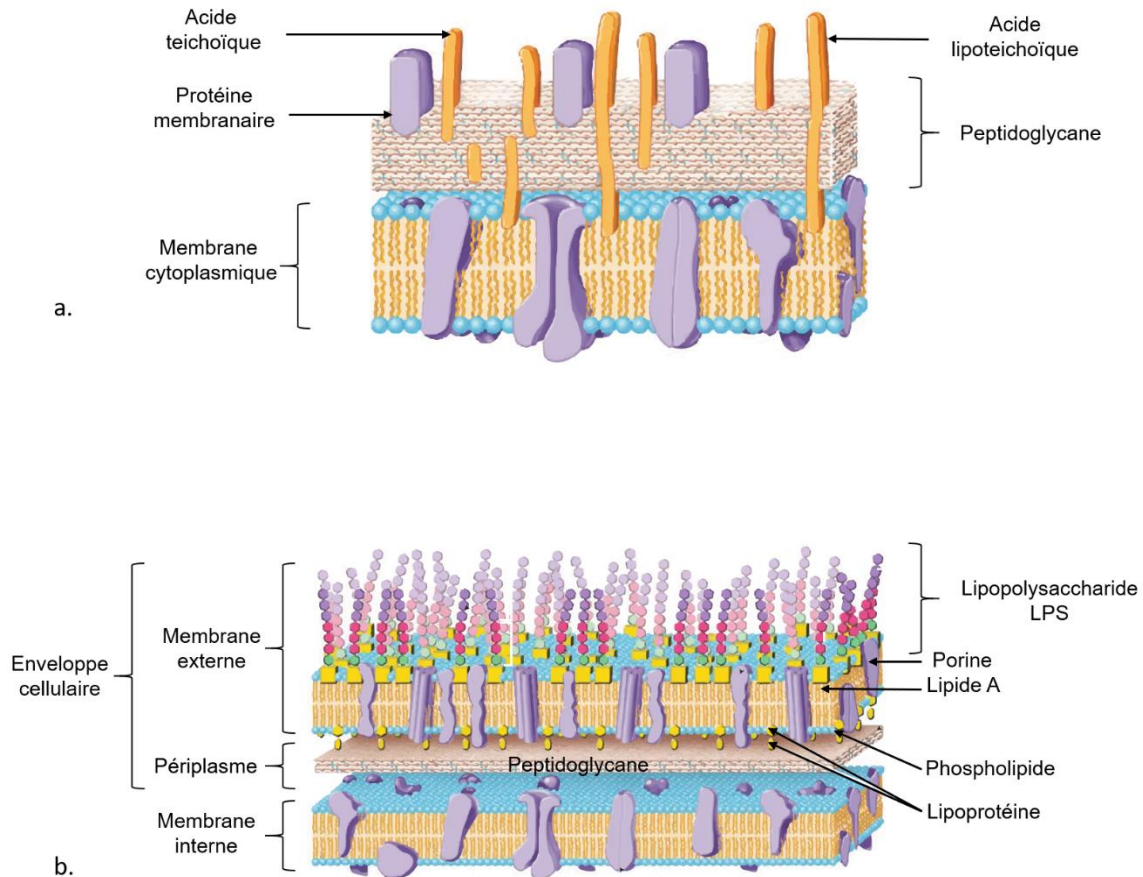


Figure 1.13 Schéma de la structure de la membrane des bactéries a) Gram-positif et b) Gram-négatif (d'après Madigan et Martinko 2018).

Le test de coloration de Gram repose sur une coloration différentielle des bactéries selon la composition de leur enveloppe cellulaire notamment selon l'abondance en peptidoglycane au sein de la membrane (Gram, 1884). La première étape est une coloration au violet de gentiane et au lugol des bactéries. Une deuxième étape à l'éthanol décolore les bactéries possédant peu de peptidoglycane (bactéries Gram-négatif), l'éthanol pouvant franchir cette barrière contrairement aux bactéries Gram-positif dans lesquelles la couche de peptidoglycane est plus épaisse. Enfin la dernière étape est une coloration à la safranine (rose) des bactéries précédemment décolorées (bactéries Gram-négatif). Les bactéries Gram positif apparaissent colorées en violet (violet de gentiane).

La survie des microorganismes dépend notamment de leur membrane cytoplasmique qui les isole du milieu extérieur. Cette membrane présente plusieurs caractéristiques afin d'assurer cette fonction. Celles-ci sont présentés brièvement dans la suite de ce manuscrit.

1.2.4. Rôle et structure des membranes cytoplasmiques

La membrane cytoplasmique possède plusieurs fonctions nécessaires à la survie de l'organisme. C'est une première barrière de protection de la cellule contre le milieu extérieur, et elle assure l'intégrité de l'organisme. De plus, la membrane est une barrière perméable permettant le transport d'ions et de solutés nécessaires à la survie de l'organisme, vers l'intérieur de la cellule ou, au contraire, vers le milieu extracellulaire (Reece et al., 2014 ; Madigan et Martinko, 2018). Ces transports sont permis notamment grâce aux protéines membranaires. La membrane cytoplasmique peut exister sous une forme électriquement neutre, mais peut aussi porter une charge électrique. En cela, la membrane possède un rôle majeur de conservation de l'énergie pour l'organisme via la force « proton motrice ». Celle-ci qui nécessaire au maintien de nombreuses activités cellulaires au sein de l'organisme, comme des réactions de transports de solutés, d'ions (par l'intermédiaire de protéines spécialisées), des mécanismes de motilité cellulaire (fonctionnement des cils et des flagelles), ou encore la biosynthèse de l'Adénosine Triphosphate (ATP ; vecteur énergétique universel nécessaire pour la survie de la cellule et impliqué dans de nombreux mécanismes cellulaires) (Knowles, 1980 ; Madigan et Martinko, 2018). Ainsi, la membrane plasmique n'est pas seulement une barrière physique, mais apparaît absolument indispensable au maintien du bon fonctionnement (métabolique et physiologique) et à l'intégrité de la cellule. La composition de la membrane plasmique diffère au sein des différents organismes. Les principales différences sont brièvement présentées ci-dessous.

1.2.4.1. Un composant essentiel des membranes eucaryotes et bactériennes : les glycérophospholipides

La membrane plasmique des bactéries est très similaire à celle des eucaryotes. Il s'agit d'une membrane lipidique, structurée en bicouche, dont les composants majoritaires sont les glycérophospholipides (GPL), constitués d'acides gras qui peuvent être ramifiés, insaturés, et posséder des longueurs de chaîne variable (14 à 20 atomes de carbone) (Fig 1.14). Ces acides gras sont reliés à une molécule de glycérol par l'intermédiaire de liaisons esters dans le cas des bactéries ou des eucaryotes (Cooper, 2000 ; Madigan et Martinko, 2018 ; Reece et al., 2014 ; Fig 1.14).

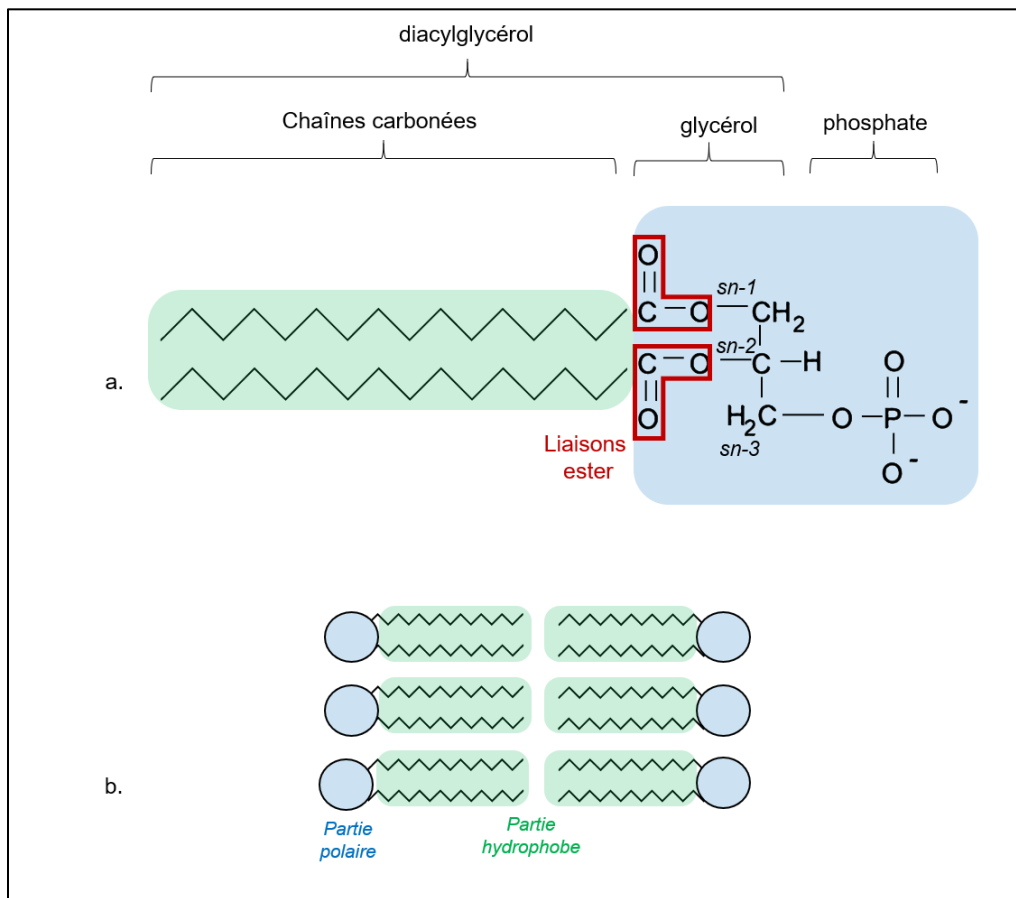


Figure 1.14 Structure des lipides membranaires des bactéries et eucaryotes. a) Acide phosphatidique. b) Configuration en bicouche de la membrane des bactéries et des eucaryotes.

Ainsi, deux acides gras liés à une molécule de glycérol forment un complexe moléculaire appelé diacylglycérol. Les GPL des eucaryotes et des bactéries possèdent aussi un groupement phosphate lié au glycérol. Les eucaryotes et les bactéries présentent une configuration stéréochimique similaire de leur GPL, avec les acides gras liés en position *sn-1* et *sn-2* sur le glycérol et le groupement phosphate lié en position *sn-3* (Koga, 2008; Albers et Meyer, 2011; Lombard et al., 2012) (Fig 1.14a). Ce groupement moléculaire est appelé acide phosphatidique et constitue l'élément de base des GPL. L'acide phosphatidique lié à un alcool par une liaison phosphoester sur le groupement phosphate forme le GPL. La structure des GPL peut varier en fonction de la nature des chaînes d'acides gras estérifiés sur le glycérol et de celle du radical estérifié sur l'alcool du groupement phosphate. On peut ainsi trouver des phosphatidylcholine (GPL + acide

aminé), phosphatidylsérine, phosphatidyléthanolamine, phosphatidylglycérol etc. (Madigan et Martinko, 2018).

En général, les membranes cytoplasmiques des eucaryotes et des bactéries sont composées à parts égales de lipides et de protéines, bien que ces proportions puissent varier en fonction des organismes. Les complexes protéiques permettent d'assurer l'intégrité de la membrane cellulaire et assurent différentes fonctions biologiques (par exemple, la synthèse d'ATP grâce à l'enzyme ATPsynthase), le transport de solutés et d'ions (par l'intermédiaire des porines, des protéines spécialisées) et la reconnaissance cellulaire (Madigan et Martinko, 2018 ; Reece et al., 2014).

La configuration moléculaire des GPL est à l'origine de plusieurs caractéristiques physico-chimiques nécessaires à la structure de la membrane cellulaire. Ainsi, les GPL sont des molécules amphiphiles, ce qui permet leur association en différentes structures : en micelle, liposome (Bangham et Horne, 1964) ou encore en bicouche lipidique stable (Madigan et Martinko, 2018 ; Reece et al., 2014) (Fig. 1.14b). Les GPL peuvent aussi jouer un rôle, notamment chez les eucaryotes, dans l'ancrage des protéines membranaires ou la signalisation cellulaire (Glycosylphosphatidylinositol, ancrage GPI).

D'autres lipides que les GPL peuvent être retrouvés dans la membrane des eucaryotes et des procaryotes, comme les stérols (chez les eucaryotes) et les hopanes (chez les bactéries). Ce sont des molécules planes, qui vont s'insérer au sein des membranes et avoir une influence sur la fluidité et l'épaisseur de cette dernière (Ernst et al., 2016 et références incluses).

1.2.4.2. Les éthers de glycérol archéens et bactériens

Comme évoqué précédemment, la plus grande différence entre les cellules d'archées et les cellules d'eucaryotes et de bactéries est la conformation en monocouche de leurs membranes plasmiques. De plus, les lipides membranaires de bactéries et d'eucaryotes sont liés au fragment glycérol par l'intermédiaire de liaisons ester, alors que les lipides membranaires d'archées sont constitués de chaînes carbonées isopréniques reliées à une ou deux molécules de glycérol par des liaisons éther (Sprott et al., 1991; Valentine, 2007 ; Fig 1.15 a et b). Enfin, en raison de différences évolutives (Koga et al., 1998; Koga et Morii, 2007; Koga, 2008), la stéréochimie des GPL d'archées diffère de celle des bactéries et des eucaryotes : les chaînes carbonées sont reliées en position *sn*-2

et *sn*-3 à un ou deux *sn*-glycérol-*I*-phosphate (Albers et Meyer, 2011; Lombard et al., 2012).

Néanmoins, il existe des lipides membranaires bactériens possédant des liaisons éther (Sinninghe Damsté et al., 2000): les tetraéthers de glycérol à chaînes alkyles ramifiées. Ils seront présentés en détail dans la partie 1.3.2.1.

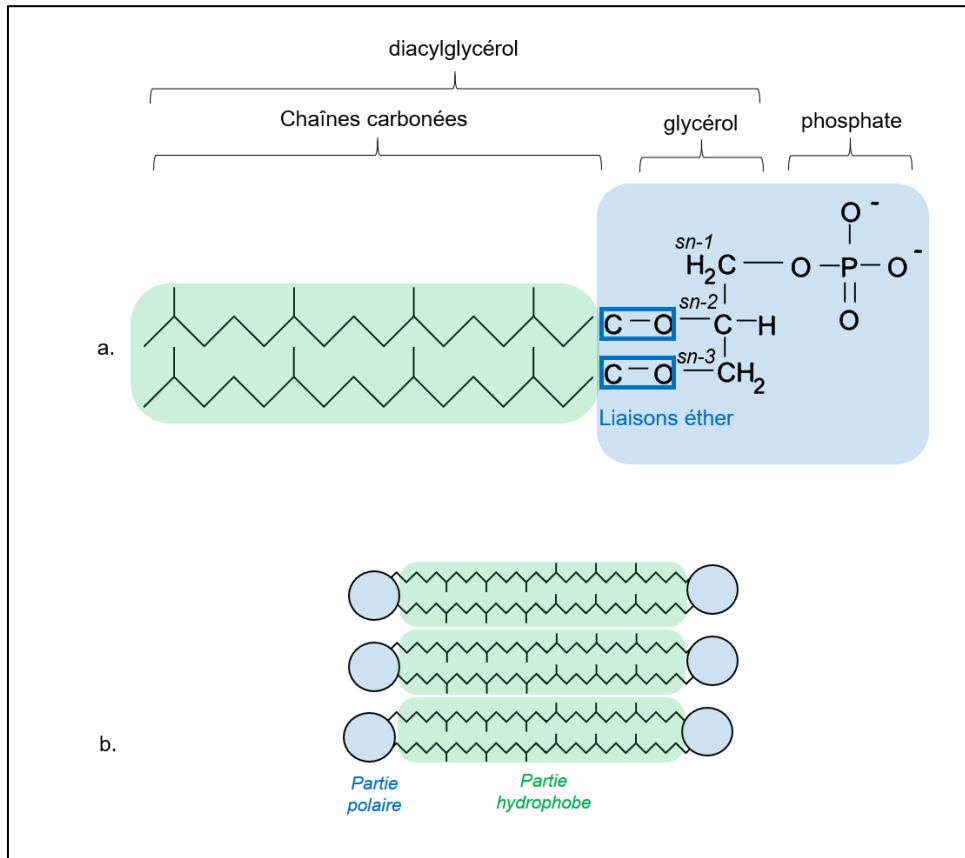


Figure 1.15 Structure des lipides membranaires archéens a) acide archaetidique b) Configuration en monocouche de la membrane des archées.

1.2.4.3. Notion de lipides polaires intacts et lipides de coeur

Les composés lipidiques sont naturellement présents sous forme dite « intacte » au sein des membranes des cellules vivantes (Pitcher et al., 2011; Elling et al., 2014). Ils sont caractérisés par la présence de groupements polaires (sucres, phosphates, acides aminés), fixés aux fragments de glycérol (Nishihara and Koga, 1987). En première approximation, les lipides polaires intacts sont considérés comme des indicateurs de biomasse vivante. Suite à la mort de l'organisme et à la lyse de la membrane cellulaire,

les groupements polaires des lipides intacts sont hydrolysés enzymatiquement (Schouten et al., 2010) (Fig. 1.16).

Les composés persistent alors dans l'environnement sous la forme de lipides dits « de cœur », considérés comme des indicateurs de biomasse morte. Néanmoins, il convient de considérer cette hypothèse avec précaution. En effet, les lipides membranaires peuvent aussi être préservés « tels quels » au sein de l'environnement après la mort de l'organisme. Ceci dépend de la nature des têtes polaires et des conditions environnementales (Fig. 1.16).

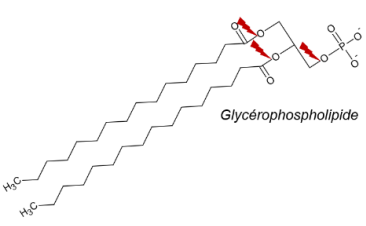
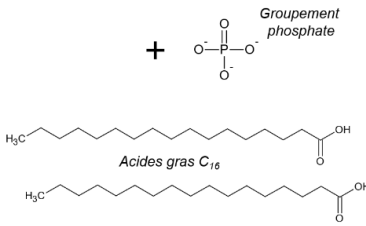
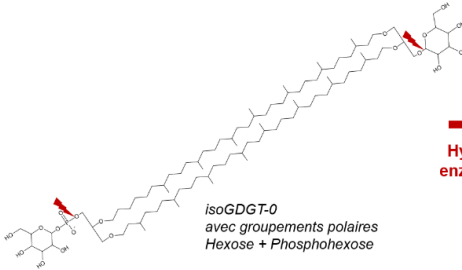
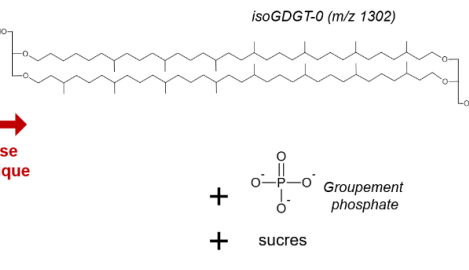
Exemples	Lipides intacts (IPL) Biomasse vivante	Lipides de cœur (CL) Biomasse « fossile »
Lipides membranaires bactériens et eucaryotes	 <p>Glycérophospholipide</p>	<p>Hydrolyse enzymatique</p>  <p>Acides gras C₁₈</p> <p>Groupement phosphate</p>
Lipides membranaires archéens	 <p>isoGDGT-0 avec groupements polaires Hexose + Phosphohexose</p>	<p>Hydrolyse enzymatique</p>  <p>isoGDGT-0 (m/z 1302)</p> <p>Groupement phosphate</p> <p>sucres</p>

Figure 1.16 Illustration de la transformation de lipides intacts en lipides de cœur par hydrolyse enzymatique. L'exemple donné pour les lipides bactériens ne concerne pas toutes les bactéries, certaines peuvent produire des tétraéthers de glycérol à chaînes alkyles ramifiées (partie 1.3.2.1)

Les lipides de cœur et les lipides intacts analysés au sein de l'environnement permettent l'obtention d'informations sur leurs organismes sources et leur milieu de vie. En effet, les microorganismes synthétisent des structures et des distributions de lipides différentes afin de conserver une fluidité appropriée de leurs membranes (Ernst et al., 2016).

1.2.5. Adaptation homéo-visqueuse des membranes cytoplasmiques

1.2.5.1. *Une mosaïque fluide...*

Des molécules amphiphiles, au sein d'un milieu aqueux, vont s'arranger spontanément en différentes structures pour que la partie polaire des molécules soit au contact du milieu environnant, tout en préservant la partie apolaire et hydrophobe du contact avec la phase aqueuse. Au sein des membranes biologiques, les GPL présentent une forme globalement cylindrique et s'arrangent spontanément afin de former des bicouches lipidiques stables (Mykytczuk et al., 2007; Ernst et al., 2016 et références incluses). Cette bicouche change de configuration en fonction de son contenu lipidique et protéique et de la conformation de ces molécules. La configuration « classique » de cette bicouche est celle d'une « mosaïque fluide » (modèle proposé par Singer et Nicolson, 1972). En effet, les lipides, sous l'effet de l'agitation thermique, ne sont pas « fixes » au sein de la membrane mais vont pouvoir se déplacer plus ou moins librement en son sein (Ladbrooke et Chapman, 1969; de Mendoza et Cronan, 1983; Russell, 1984; Beales, 2004;...). Les parties hydrophobes des lipides sont toujours orientées vers l'intérieur de la membrane mais les lipides vont pouvoir se déplacer librement autour de leurs axes (mouvement de rotation), ou diffuser horizontalement au sein de la membrane. Ainsi, cette bicouche lipidique se comporte comme un fluide, et ce, au sein des deux couches de lipides la constituant (Denich et al., 2003; Mykytczuk et al., 2007; Ernst et al., 2016).

1.2.5.2. *...qui s'adapte aux variations environnementales...*

Une modification de la distribution des acides gras membranaires d'une souche d'*Escherichia coli* a été observée en réponse à des variations de la température environnante (Marr and Ingraham, 1962) et ce mécanisme a été qualifié d'adaptation homéovisqueuse des membranes cellulaires (Sinensky, 1974). Des facteurs de stress extérieurs, comme des changements de conditions de température, de pH, de pression osmotique peuvent modifier l'état fluide de la membrane, et la faire évoluer vers une phase dite de « gel » cristallin (Mykytczuk et al., 2007). Cette transition de phase, entre un état fluide-liquide et un état de gel dépend de la température de fusion des lipides constituant majoritairement la membrane (Denich et al., 2003; Ernst et al., 2016). Ainsi, la température est un facteur d'influence majeur sur l'état de fluidité des membranes

biologiques (Denich et al., 2003). En outre, la fluidité de la membrane influence sa perméabilité aux différents solutés ioniques et neutres pouvant entrer ou sortir de la cellule. Plus la membrane est fluide, et donc perméable, plus les ions et molécules vont pouvoir transiter au sein de la cellule (Denich et al., 2003; Ernst et al., 2016).

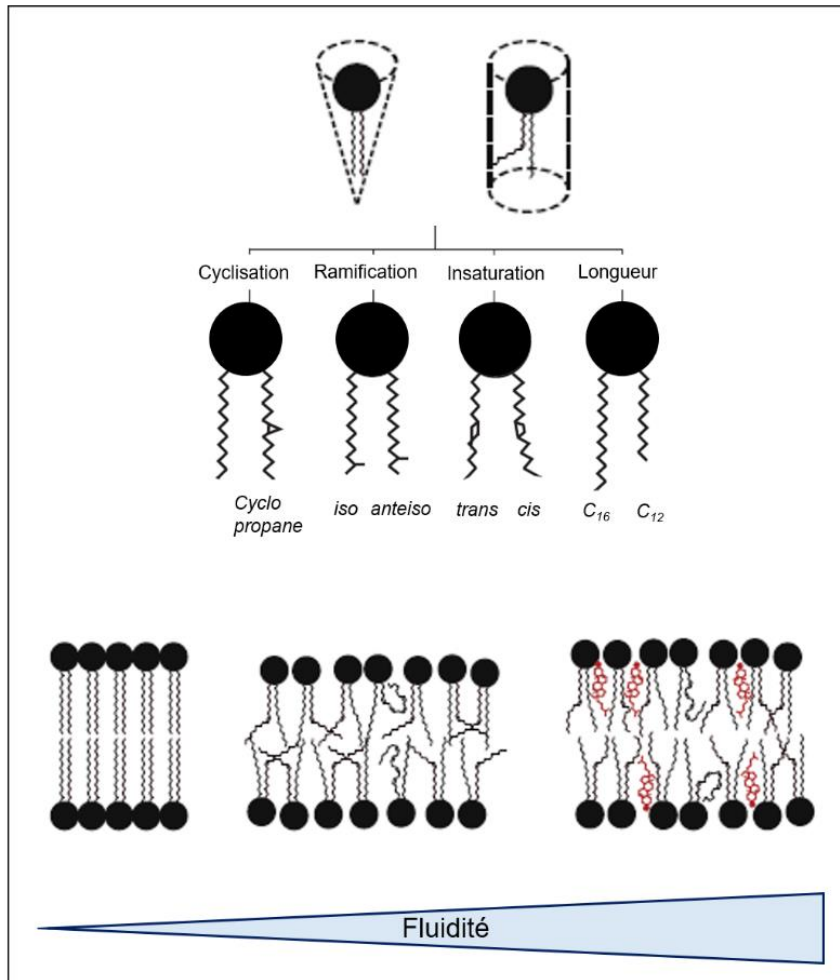


Figure 1.17 Schéma illustrant l'influence de la structure, de la composition et de l'agencement des lipides sur la fluidité de la membrane cytoplasmique (adapté de Ernst et al., 2016)

La fluidité de la membrane dépend ainsi de sa structure, de la composition et de l'agencement des lipides présents en son sein. La longueur de la chaîne carbonée des lipides a un effet sur la stabilité de la membrane. Ainsi, des chaînes carbonées plus longues favorisent un état « gel » de la membrane, en augmentant le nombre de d'interactions de « Van der Waals » entre les deux couches de la membrane lipidique et en stabilisant la structure membranaire. A l'inverse, des chaînes plus courtes limitent les

interactions entre les deux couches de la membrane et favorisent donc une structure plus fluide de cette dernière (Ernst et al., 2016; Fig 1. 17).

Par ailleurs, les insaturations et les ramifications éventuellement présentes sur les chaînes carbonées des GPL perturbent les interactions entre les lipides et augmentent l'encombrement stérique au sein de la membrane, favorisant les mouvements (rotation et diffusion) et donc la fluidité de la membrane (Denich et al., 2003; Ernst et al., 2016; Fig.1.17).

La modification de la proportion des acides gras saturés par rapport aux insaturés est le mécanisme d'adaptation le plus utilisé par les bactéries face aux stress environnementaux (Russell, 1984; Ernst et al., 2016). De plus, la position des ramifications sur les chaînes carbonées joue aussi un rôle important sur la fluidité de la membrane. Plus une ramification est haute sur la chaîne carbonée, plus l'encombrement stérique entre les GPL adjacents augmente, et donc plus la fluidité de la membrane augmente (Denich et al., 2003; Poger et al., 2014; Ernst et al., 2016; Fig.1.17). En outre, la présence de cycles propane sur les chaînes carbonées peut, de façon similaire, augmenter la fluidité des membranes, tout en améliorant leur stabilité (Zhang and Rock, 2008) (Fig.1.17).

1.2.5.3. *...notamment de température et de pH*

L'influence de la température sur les membranes bactériennes est à ce jour le facteur le plus étudié. Il a ainsi été mis en évidence qu'une température plus élevée que l'optimum de température de l'organisme provoque une fluidisation de la membrane par fusion des GPL, tandis qu'une température plus faible fait entrer la membrane dans une phase « gel », en restreignant les mouvements des GPL (Marr and Ingraham, 1962; Mykytczuk et al., 2007). Par exemple, une augmentation de la température du milieu de vie entraîne la biosynthèse de GPL à chaînes plus longues, saturées et non ramifiées limitant la fluidisation de la membrane (Denich et al., 2003; Haque and Russell, 2004).

Cependant, la température n'est pas le seul facteur pouvant influencer la structure des membranes lipidiques des bactéries et autres microorganismes. Afin de pouvoir survivre dans des milieux à pH variable, en présence de différents solutés, les organismes doivent pouvoir conserver une fluidité, mais aussi une perméabilité adaptée de leurs membranes (Ernst et al., 2016). Dans des environnements acides ou basiques, les

organismes peuvent aussi modifier les charges de leurs GPL (Russell et al., 1995; Dowhan, 1997), augmenter la proportion de composés saturés (Beales, 2004), ou stabiliser/ perméabiliser leurs membranes en intégrant des composés cycliques (Fig. 1.17; Zhang and Rock, 2008).

1.3. Les lipides comme marqueurs d'environnement

La modification de la composition de la membrane lipidique en réponse aux variations de différents paramètres environnementaux, notamment de température ou de pH, a conduit au développement de plusieurs marqueurs d'environnement. Parmi les lipides à la base de tels marqueurs, on peut citer les alcénones (indice U^k_{37}), les diols à longue chaîne (LDI), les tétraéthers d'origine archéenne (TEX₈₆), les tétraéthers d'origine bactérienne (MBT'_{5Me}/CBT) ou plus récemment les acides gras 3-hydroxylés (RAN₁₅, RAN₁₇, RIAN). Ces composés et les marqueurs associés sont présentés dans les paragraphes suivants afin d'apporter une compréhension sur leur développement, applications possibles et limites d'utilisation.

1.3.1. Exemples de biomarqueurs de température en milieu aquatique

1.3.1.1. Les alcénones insaturés (Indice U^k_{37})

Le paléothermomètre organique le plus utilisé dans le cadre de reconstitutions paléoenvironnementales en milieu océanique est l'indice d'insaturation des alcénones qui est basé sur la distribution des alcénones.

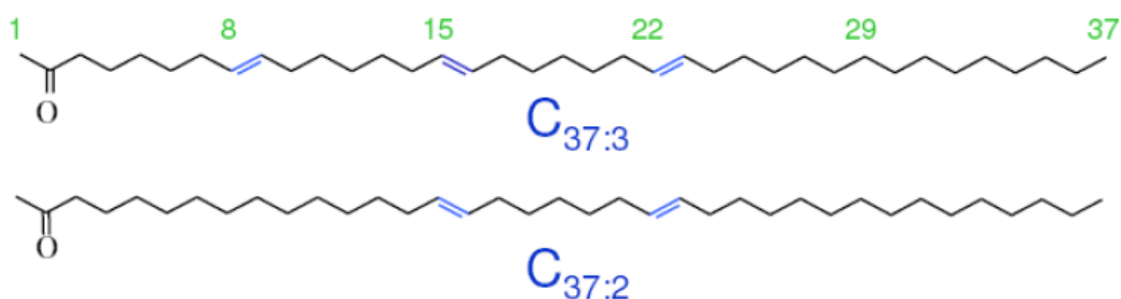


Figure 1.18 Structures des alcénones en $C_{37:2}$ et $C_{37:3}$

Ces molécules sont des cétones à longues chaînes (de 35 à 41 atomes de carbone) qui peuvent posséder de 1 à 5 insaturations (Fig. 1.18). Elles sont biosynthétisées par des microalgues unicellulaires majoritairement marines, les coccolithophoridés (Fig. 1.19). Ces organismes représentent environ 50 % de la production primaire mondiale, et constituent le premier producteur de carbonate de calcium de l'océan ouvert (Milliman, 1993). Deux espèces sont largement responsables de la production des alcénones en milieu océanique : *Emiliana huxleyi* et *Gephyrocapsa oceanica* (Fig 1.19).

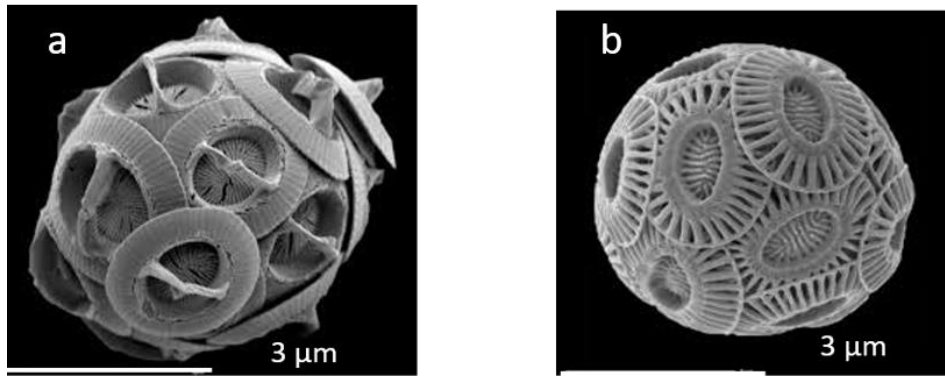


Figure 1.19 Photographies au microscope électronique de a) *Emiliana huxleyi* et b) *Gephyrocapsa oceanica*. (source : <https://www.mikrotax.org/>)

Dans un premier temps, des cultures en milieu contrôlé de coccolithophoridés, de l'espèce *Emiliana huxleyi*, ont permis d'établir une relation linéaire entre la température du milieu de vie de ces organismes et le degré d'insaturation des alcénones qu'elles produisent (indice U^k_{37} ; Marlowe et al., 1984; Brassell et al., 1986). Par la suite, l'indice a été simplifié en supprimant le composé $C_{37:4}$ souvent très peu abondant (Prahl and Wakeham, 1987). Lorsque la température augmente, la proportion relative des homologues en C_{37} à trois insaturations ($C_{37:3}$) diminue par rapport aux homologues en $C_{37:2}$ (Fig 1.20 ; Prahl et Wakeham, 1987), ce qui se reflète dans l'indice U^k_{37} (Eq.1.1 ; Brassell et al., 1986), qui varie linéairement avec la température:

$$U^k_{37} = \frac{C_{37:2}}{C_{37:2} + C_{37:3}} \quad (\text{Eq. 1.1})$$

où $C_{37:2}$ et $C_{37:3}$ sont les abondances relatives des alcénones représentées en Fig 1.18.

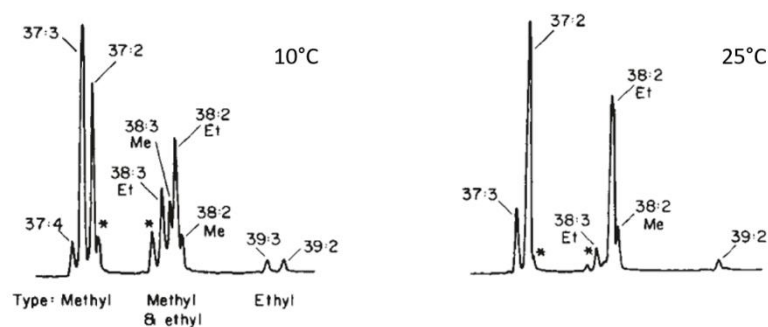


Figure 1.20 Chromatogrammes des alcénones détectées au sein des cultures d'*Emiliana huxleyi* à 10°C et 25°C (Prahl et Wakeham 1987).

Des analyses de sédiments de surface du monde entier ont ainsi permis d'établir des droites de calibrations globales entre l'indice $U^{k'}_{37}$ et la température de surface de l'eau de mer (Sea Surface Temperature ou SST; Brassell et al., 1986). Le nombre d'échantillons analysés a augmenté au cours des ans, avec par exemple pratiquement 600 échantillons analysés dans l'étude de Conte et al. (2006 ; Fig. 1.21) et la relation suivante entre l' $U^{k'}_{37}$ et la SST (Eq. 1.2) :

$$SST (^{\circ}C) = 29.876 \times U^{k'}_{37} - 1.334, R^2 = 0.97, n = 592 \quad (\text{Eq. 1.2})$$

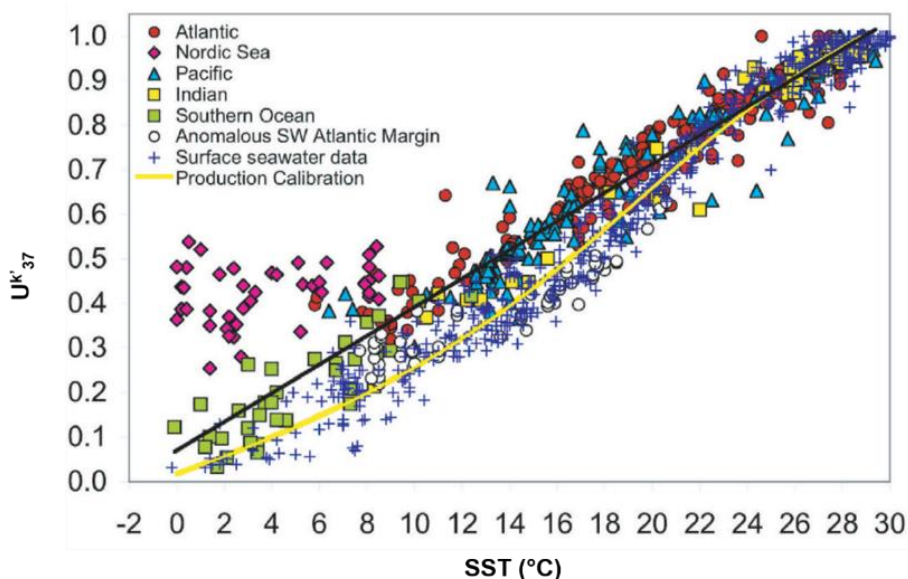


Figure 1.21 Calibration globale entre l'indice $U^{k'}_{37}$ et la température de surfaces marine (SST) (592 échantillons de sédiments de surface répartis au sein de différents océans) proposée par Conte et al. (2006).

Les alcénones ont été détectés dans des sédiments vieux de plusieurs dizaines de millions d'années (Brassell et Dumitrescu, 2004) et ont été utilisés pour reconstruire des paléo SST, fournissant des enregistrements totalement indépendants mais complémentaires de ceux obtenus à l'aide d'autres marqueurs (e.g. isotopes stables de l'oxygène, rapport Mg/Ca). En effet, les enveloppes de carbonate de calcium des foraminifères ne sont pas toujours bien conservées alors que les alcénones sont plus résistants à la dégradation (Eglinton et al., 2001; Sachs et al., 2001; Haug et al., 2005; Zhao et al., 2006; Eglinton and Eglinton, 2008).

Outre l'étude de l'abondance et de la distribution des alcénones, il est possible d'effectuer des mesures isotopiques (e.g. $\Delta^{14}C$, $\delta^{13}C$) sur ces composés et d'en tirer des

informations environnementales. Le $\delta^{13}\text{C}$ des alcénones est ainsi directement corrélé aux concentrations de CO_2 (pCO_2) dans l'océan, en équilibre avec la pCO_2 atmosphérique (Jasper and Hayes, 1990; Pagani, 2002; Pagani et al., 2005; Badger et al., 2019). Néanmoins, l'interprétation de ces mesures est à considérer avec précaution, car d'autres facteurs peuvent influencer l'enregistrement de la pCO_2 par les organismes sources, comme les phases et le taux de croissance des coccolithophoridés (Bidigare et al., 1997; Popp et al., 2006; Badger et al., 2019). Les mesures de $\Delta^{14}\text{C}$ réalisées sur les alcénones peuvent quant à elles servir à affiner les mesures d'âge sur des sédiments marins et, dans certains cas, identifier le transport à longue distance des particules sédimentaires par les courants océaniques (Eglinton and Eglinton, 2008 et références incluses).

Outre le milieu océanique, les alcénones ont également été détectés dans des environnements d'eau saumâtre et d'eau douce (Cranwell, 1985; Castañeda and Schouten, 2011 et références incluses). Néanmoins, les organismes responsables de la production des alcénones en milieu lacustre sont actuellement mal connus. Des études génétiques ont mis en lumière la présence d'ADN de multiples espèces d'haptophytes au sein des lacs (Theroux et al., 2010; Longo et al., 2018). Il semblerait que les principaux producteurs d'alcénones varient d'un lac à l'autre. Contrairement au milieu océanique, où les homologues majoritaires sont les alcénones en $\text{C}_{37:2}$ et $\text{C}_{37:3}$, l'homologue $\text{C}_{37:4}$ semble prédominant en milieu lacustre. La présence en proportions importantes d'homologues en C_{38} a également été observée dans certains lacs, en particulier l'homologue $\text{C}_{38:3}$ (Pearson et al., 2008; Kristen et al., 2010). Il apparaît difficile d'appliquer les calibrations marines de l'indice U^k_{37} aux milieux lacustres, car la prédominance des composés $\text{C}_{37:4}$ par rapport aux homologues $\text{C}_{37:2}$ dans de nombreux lacs se traduit par des estimations de température faussées (Zink et al., 2001). Des tentatives de calibrations en milieu lacustre ont été réalisées, en tenant compte des composés majoritaires dans ces environnements, par exemple en substituant dans l'indice U^k_{37} le composé C_{37} par le composé C_{38} (Pearson et al., 2008). Néanmoins, ces différentes approches restent locales, et spécifiques à chaque lac étudié.

Malgré sa fiabilité et sa robustesse démontrée au cours des 30 dernières années, l'indice U^k_{37} présente certaines incertitudes. En effet, en milieu océanique, cet indicateur peut être influencé par d'autres paramètres que la température de l'eau, comme la lumière (photosynthèse), ou encore la teneur en nutriments (Prahl et al., 2003). De plus, très peu

de connaissances ont été acquises sur les organismes sources des alcénones, du fait de la difficulté de cultiver les coccolithophoridés en milieu contrôlé. Enfin, une autre contrainte est liée à l'influence de la saisonnalité, et plus particulièrement dans le cadre des alcénones marines, de la circulation océanique et du transport latéral (Brassell et al., 1986; Castañeda and Schouten, 2011 et références incluses).

Ces différentes contraintes affectent également l'utilisation de l'indice d'insaturation des alcénones en milieu lacustre, expliquant probablement une part non négligeable de la dispersion observée au sein des régressions linéaires entre les indices issus des alcénones lacustres et les températures de surface de l'eau (Castañeda and Schouten, 2011 et références incluses).

1.3.1.2. Les diols à longue chaîne (indice LDI)

Les diols à longue chaîne constituent d'autres marqueurs de température en milieu aquatique. Ces composés ont été découverts dans des sédiments de la mer Noire (de Leeuw et al., 1981), puis détectés dans des sédiments datant du Quaternaire, à toutes les latitudes (Versteegh et al., 1997; Versteegh et al., 2000 et références incluses). Des cultures de microalgues appartenant à la famille des eustigmatophytes (marines et d'eau douce) produisent des diols à longues chaînes, composés principalement de 1,13 et 1,15-diols en C₂₈-C₃₀ (Fig. 1.22 ; Volkman et al., 1992; Volkman et al., 1999a,b; Villanueva et al., 2014; Rampen et al., 2014a; Rampen et al., 2014b).

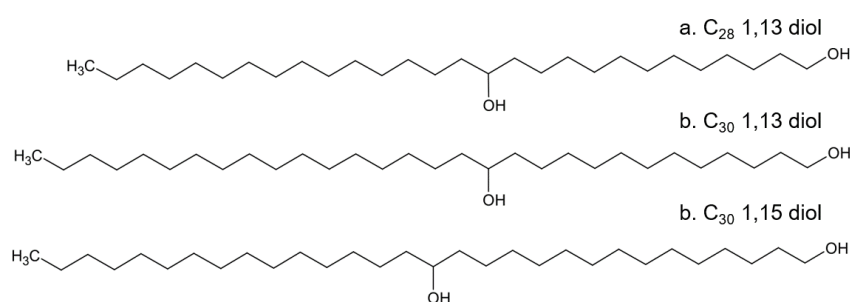


Figure 1.22 Structures des diols à longues chaînes : a) C₂₈ 1,13-diol ; b) C₃₀ 1,13-diol ; c) C₃₀ 1,15-diol.

Cependant, il n'y a pas consensus sur le fait que seuls les eustigmatophytes produisent des diols en milieu marin. En effet, les rares données montrent que

l'abondance relative des diols diffère entre les algues eustigmatophytes marines et les sédiments (Volkman et al., 1992; Versteegh et al., 1997; Rampen et al., 2012).

Un rapport de l'abondance relative des diols C₂₈ à C₃₀, l'indice LDI (Long chain Diol Index), a été défini par Rampen et al. (2012 ; Eq. 1.3) :

$$\text{Long chain Diol Index (LDI)} = \frac{C_{30\ 1,15\text{-diol}}}{C_{28\ 1,13\text{-diol}} + C_{30\ 1,13\text{-diol}} + C_{30\ 1,15\text{-diol}}} \quad (\text{Eq. 1.3})$$

Où C₂₈ 1,13-diol, C₃₀ 1,13-diol, C₃₀ 1,15-diol correspondent aux abondances relatives des diols à longues chaînes représentés en Fig 1.22.

Une relation linéaire statistiquement significative (Eq. 1.4) a été établie entre l'indice LDI et les températures de surface marines (SST) après analyse de ces composés dans 162 échantillons de sédiments marins (Rampen et al., 2012) :

$$\text{LDI} = 0.033 \times \text{SST} + 0.095 \quad (R^2 = 0.969 ; n = 162) \quad (\text{Eq. 1.4})$$

En milieu lacustre, le potentiel de ces composés comme marqueurs de température de surface dans les lacs a été étudié par Rampen et al. (2014a). L'indice LDI révèle une corrélation relativement faible ($R^2 = 0,33$) avec la température annuelle moyenne de l'air utilisée comme approximation de la température moyenne annuelle de la surface des lacs.

Des incertitudes persistent encore sur l'origine et la fonction biologique des diols à longue chaîne et donc sur leur utilisation en routine dans le cadre de reconstitutions paléoclimatiques. Comme précisé plus haut, la comparaison du profil des diols produits en cultures avec ceux retrouvés dans les sédiments marins ou lacustres semble suggérer l'existence d'autres sources que les seuls eustigmatophytes pour ces composés (Villanueva et al., 2014; Rampen et al., 2014a, b). Une étude conjointe des lipides et des gènes ARNr 18S (unité codante spécifique des organismes eucaryotes) a montré que les diols à longue chaîne sont produits, entre autres, par des eustigmatophytes dans les eaux de surface des lacs (Villanueva et al., 2014). Un manque d'information persiste par ailleurs sur le fait que d'autres facteurs environnementaux que la température (salinité, disponibilité en nutriments, saisonnalité...) puissent influencer la distribution de ces composés. Il reste également à évaluer la résistance à la diagenèse et à la dégradation de ces molécules. Rodrigo-Gámiz et al. (2016) ont par exemple suggéré que les diols à

longue chaîne pouvaient être dégradés par oxydation, faussant alors les tentatives de reconstructions de températures réalisées avec le LDI.

1.3.1.3. *Les alkyl tetraéthers de glycérol à chaînes isopréniques*
(indice TEX_{86})

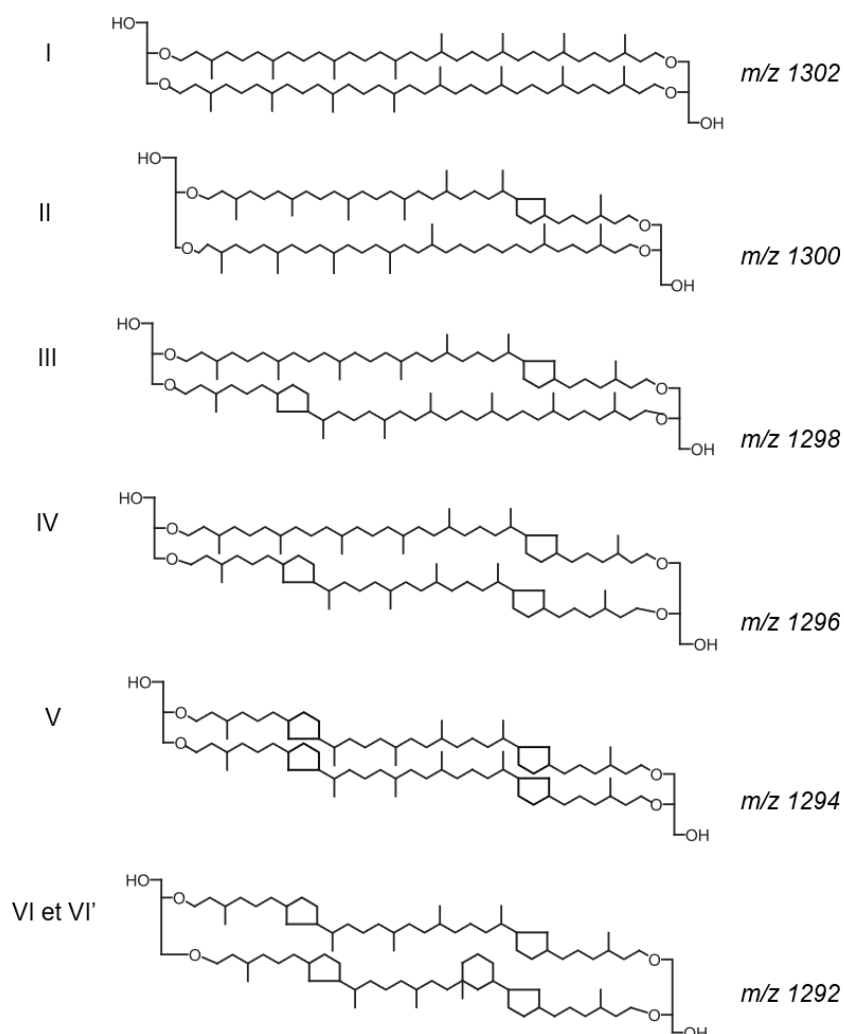


Figure 1.23 Structure des tetraéthers isopréniques (isoGDGTs), avec la nomenclature associée ainsi que les rapports masse/charge (m/z) correspondants.

Les alkyl tetraéthers de glycérol (ou GDGTs, pour glycerol dialkyl glycerol tetraethers) sont une famille de lipides présents dans la membrane des archées et de certaines bactéries. Ces composés possèdent deux chaînes aliphatiques reliées à deux unités de glycérol par des liaisons éther (Fig.1. 23). Ils ont été détectés pour la première fois dans des cultures d'une souche d'archées, *Thermoplasma acidophilum* (Langworthy et al., 1972, 1977, 1983). Leur découverte dans des membranes de microorganismes extrêmophiles a conduit à l'hypothèse que ces molécules étaient biosynthétisées en réponse à des conditions environnementales extrêmes, comme des températures élevées,

des milieux très acides ou hypersalins. Depuis, les tetraéthers archéens ont été détectés dans tous les types d'environnements, terrestres – sols (Wang et al., 2012; Schouten et al., 2013; Zheng et al., 2019), tourbières (Huguet et al., 2010), spéléothèmes (Yang et al., 2011) – et aquatiques – marins (Schouten et al., 2002; Peterse et al., 2015), eaux douces (Powers et al., 2010; Schouten et al., 2012).

Les chaînes aliphatiques des GDGTs identifiés dans des souches d'archées sont caractérisées par la répétition d'unités isoprènes, liées au glycérol par des liaisons éthers. Les tetraéthers archéens sont ainsi qualifiés de tetraéthers de glycérol isopréniques (ou isoGDGTs), ils possèdent 86 carbones et 0 à 8 cycles pentane (Fig 1.23).

Un tetraéther particulier, le crenarchaeol (isoGDGT VI et son régioisomère, l'isoGDGT VI'), contient quatre cycles pentane et un cycle hexane (Fig 1.23). Ce tetraéther est synthétisé spécifiquement par un phylum d'Archaea, les *Thaumarchaeota*, qui sont des archées oxydatrices de l'ammonium, particulièrement abondantes en milieu marin (Schouten et al., 2002), même s'il a été récemment suggéré qu'il pourrait également être produit par le groupe marin II des *Euryarchaeota* (Lincoln et al., 2014). Le caldarcheol (GDGT I) a quant à lui été détecté dans tous les embranchements d'archées et est particulièrement abondant au sein des archées méthanogènes/méthanotrophes (Pearson and Ingalls, 2013). Les autres isoGDGTs contenant de 1 à 4 cycles pentane (GDGTs II-V) sont produits par l'ensemble des Archaea, i.e. *Thaumarchaeota*, *Crenarchaeota* et certains *Euryarchaeota* (Pearson and Ingalls, 2013) (Partie 1.2.2).

Schouten et al. (2002) ont étudié la relation entre le degré de cyclisation des tetraéthers isopréniques et les variations des températures de surface marine (SST) dans un ensemble de 44 échantillons sédiments marins provenant du monde entier. Ils ont défini un indice basé sur l'abondance relative des tetraéthers isopréniques comportant un ou plusieurs cycles, l'indice TEX_{86} (Eq. 1.5 ; Schouten et al., 2002) :

$$TEX_{86} = \frac{[III]+[IV]+[VI]}{[II]+[III]+[IV]+[VI]} \text{ (Eq. 1.5)}$$

Les chiffres romains correspondent à l'abondance relative des structures présentées en Fig. 1.23. Une relation linéaire significative entre le TEX_{86} et les températures de surface marines a été observée (Schouten et al., 2002 ; Eq. 1.6 ; Fig.1.24) :

$$\text{TEX}_{86} = 0.015 \times \text{SST} + 0.28 \quad (R^2 = 0.92, \text{EQM}^1 = 2^\circ\text{C}, n=44; \text{Eq. 1.6})$$

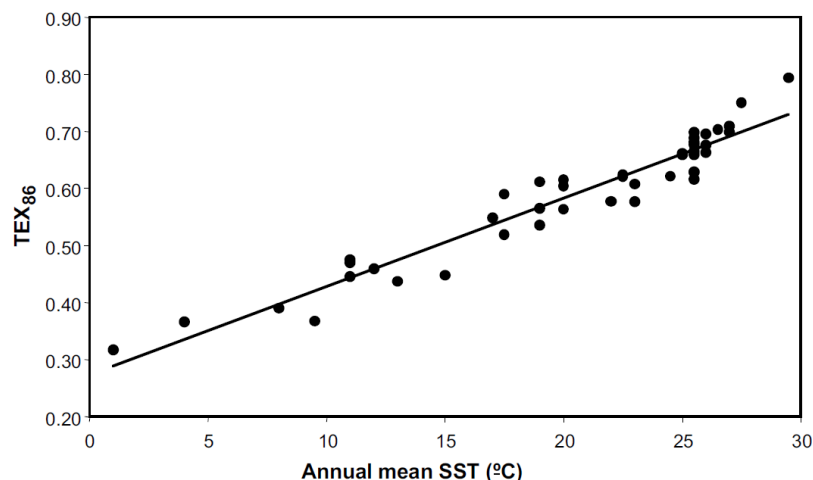


Figure 1.24 Régression linéaire entre l'indice TEX_{86} et la température de surface marine ($n=44$), correspondant aux équations 1.5 et 1.6.

Cette variation de la distribution des isoGDGTs avec la température reflète le phénomène naturel d'adaptation homeovisqueuse des archées (voir Partie 1.2.3).

Le TEX_{86} est aujourd'hui couramment utilisé dans le cadre de reconstitutions paléoenvironnementales en milieu marin.

De nombreuses calibrations ont été proposées ces dernières années à la suite de celle initialement développée par Schouten et al. (2002), améliorant la précision des reconstructions. Ainsi, Kim et al. (2008) ont analysé un ensemble plus grand de sédiments marins ($n=287$), donnant lieu à la calibration globale suivante entre les températures de surface marines et l'indice TEX_{86} (Eq. 1.7) :

$$\text{SST} = 56,2 \times \text{TEX}_{86} - 10,78 \quad (R^2=0.94 ; \text{EQM} = 1.7 \text{ }^\circ\text{C}) \quad (\text{Eq. 1.7})$$

¹ *EQM (Erreur Quadratique Moyenne) étant traduit par RMSE en anglais (Root Mean Squared Error)*

Néanmoins, des incertitudes subsistent sur les reconstitutions de températures dans les milieux de haute latitude. Deux calibrations indépendantes ont été proposées à partir de variants de l'indice TEX_{86} pour remédier à cela (Kim et al., 2010) : le $\text{TEX}_{86}^{\text{L}}$ (Eq. 1.8) pour les régions où les SST sont inférieures à 15 °C (Eq. 1.9 pour la calibration correspondante) et le $\text{TEX}_{86}^{\text{H}}$ (Eq. 1.10) pour les régions où les températures sont supérieures à 15°C (Eq. 1.11 pour la calibration correspondante). L'ensemble de ces équations est présenté ci-dessous :

$$\text{TEX}_{86}^{\text{L}} = \log \left(\frac{[\text{III}]}{[\text{II}]+[\text{III}]+[\text{IV}]} \right) \quad (\text{Eq. 1.8})$$

$$\text{SST} = 67.5 \times \text{TEX}_{86}^{\text{L}} + 46.9 \quad (\text{R}^2= 0,86; \text{EQM} = 4,0 \text{ °C}; n=396) \quad (\text{Eq. 1.9})$$

$$\text{TEX}_{86}^{\text{H}} = \log \left(\frac{[\text{III}]+[\text{IV}]+[\text{VI}']]}{[\text{II}]+[\text{III}]+[\text{IV}]+[\text{VI}']} \right) \quad (\text{Eq. 1.10})$$

$$\text{SST} = 68.4 \times \text{TEX}_{86}^{\text{H}} + 38.6 \quad (\text{R}^2= 0,87 ; \text{EQM} = 2,5 \text{ °C} ; n=255) \quad (\text{Eq. 1.11})$$

Une différence majeure entre les indices $\text{TEX}_{86}^{\text{L}}$ et $\text{TEX}_{86}^{\text{H}}$ est la prise en compte ou non du régioisomère du crénarchaeol (VI' ; Fig 1.23), dont l'abondance relative est fortement corrélée aux températures de surface de l'eau de mer dans les régions chaudes (Kim et al., 2010). Néanmoins, l'erreur quadratique moyenne de ces nouvelles calibrations régionales (Eqs. 1.9 et 1.11) est supérieure à celle de la calibration globale du TEX_{86} .

Le TEX_{86} est un marqueur de température qui a été largement étudié au cours des 20 dernières années et qui s'avère globalement robuste. Néanmoins, il présente, comme les indices U^{k}_{37} ou LDI plusieurs incertitudes. L'écologie des *Thaumarchées*, microorganismes à la base du TEX_{86} , est ainsi mal connue, notamment en ce qui concerne leur profondeur de vie ou leur saisonnalité (Schouten et al., 2013 et références incluses). Par ailleurs, les isoGDGTs sont également produits en milieu terrestre (Blaga et al., 2009; Powers et al., 2010). Il est donc nécessaire de prendre en compte l'apport des isoGDGTs provenant des sols en milieu lacustre ou marin côtier, ces derniers pouvant fausser les reconstructions paléoclimatiques effectuées à partir des composés produits *in situ*.

1.3.1.4. *Conclusion partielle*

Plusieurs biomarqueurs organiques existent pour la reconstruction des températures de surface en milieu marin (par exemple; Castañeda and Schouten, 2011; Lattaud et al., 2018; de Bar et al., 2019) et de manière plus réduite en milieu lacustre (par exemple; Powers et al., 2010; Blaga et al., 2013; Randlett et al., 2014). Ces marqueurs ont été développés et largement utilisés en milieu marin, mais montrent des limites lors de leur application en milieu lacustre, du fait de l'hétérogénéité des environnements continentaux et de la multiplicité des sources et variables environnementales pouvant influencer la production et donc la distribution des lipides.

De plus, il a été mis en évidence que les marqueurs de températures existants présentent tous des contraintes, souvent semblables, pouvant fausser les valeurs estimées : influence de la saisonnalité ou d'autres paramètres environnementaux que la température sur la distribution des lipides, écologie des organismes producteurs de ces derniers, transport latéral des marqueurs organiques en milieu marin... Il apparaît indispensable d'étudier les points susmentionnés afin de pouvoir utiliser de manière fiable les marqueurs correspondants dans le cadre de reconstitutions climatiques.

Contrairement au milieu aquatique, le nombre de marqueurs, notamment de température, en milieu terrestre apparaît très limité du fait de la difficulté au sein de ces environnements d'établir des relations linéaires entre les abondances des lipides et les températures de l'air, de multiples variables environnementales pouvant influencer la structure de ces molécules (pH, humidité, granulométrie, type de sols, couvert végétal...). Dans la partie suivante, nous allons nous intéresser plus particulièrement aux marqueurs de températures et de pH d'origine moléculaire en milieu terrestre.

1.3.2. Marqueurs de température et de pH en milieu terrestre

A l'heure actuelle, les seuls marqueurs moléculaires disponibles pour les reconstructions de température et de pH en milieu terrestre sont les tétraéthers bactériens. Récemment, le potentiel comme marqueur d'environnement d'une autre famille de lipides, les acides gras 3-hydroxylés, a été mis en lumière. Ces deux familles de lipides sont présentés ci-après.

1.3.2.1. *Les tétraéthers de glycérol bactériens*

• Structures et sources des tétraéthers bactériens

Outre les tétraéthers isopréniques produits par les archées et présentés dans la section 1.3.1.3, un autre type de tétraéther de glycérol a été découvert, initialement dans une tourbière, par Sinninghe Damsté et al. (2000). Ces composés (brGDGTs ; Fig. 1.25) se différencient des isoGDGTs produits par les archées par (1) la présence de chaînes alkyles ramifiées non isopréniques et (2) la configuration stéréochimique du glycérol caractéristique des bactéries et opposée à celle des archées (Sinninghe Damsté et al., 2000; Weijers et al., 2006a, b; Schouten et al., 2013). L'origine bactérienne des brGDGTs est confortée par le fait que des tétraéthers à chaînes alkyles ramifiées aient été détectés au sein de souches de bactéries thermophiles (Langworthy et al., 1983; Huber et al., 1992).

Les tétraéthers de glycérol à chaînes alkyles ramifiées contiennent 4 à 6 groupements méthyle et peuvent posséder jusqu'à deux cycles pentane (Fig. 1.25). Initialement, les ramifications méthyle de ces molécules (en plus des 4 ramifications méthyle présents au centre de la chaîne) avaient été détectées sur le 5^{ème} carbone de chaque chaîne alkyle. L'optimisation des techniques de séparation des tétraéthers bactériens a permis de détecter des isomères de position, les groupements méthyle pouvant se trouver en 5^{ème}, 6^{ème} voire en 7^{ème} position (De Jonge et al., 2013, 2014; Ding et al., 2016). Ces composés présentent donc une grande diversité de structures. Ils sont actuellement classés en trois grandes familles (Fig. 1.25) :

- Les brGDGTs tétraméthylés (brGDGTs I),
- Les brGDGTs pentaméthylés (brGDGTs II)
- Les brGDGTs hexaméthylés (brGDGTs III).

Au sein de ces 3 catégories, les tetraéthers ramifiés peuvent être caractérisés par l'absence de cycles pentane (Ia, IIa, IIa', IIIa, IIIa'), la présence d'un cyclopentane (Ib, IIb, IIb', IIIb, IIIb'), ou deux cycles pentane (Ic, IIc, IIc', IIIc, IIIc'). Enfin, les brGDGTs penta- et hexaméthylés peuvent présenter une ramification méthyle en 5^{ème} position sur leurs chaînes aliphatiques (isomères 5-méthyle) ou en 6^{ème} position (isomères 6-méthyle ; notation par un prime sur la Fig 1.25).

Les bactéries sources de ces lipides restent encore inconnues à ce jour (Schouten et al., 2013 et références incluses), bien que certaines d'entre elles puissent appartenir au phylum *Acidobacteria* (Weijers et al., 2009a; Sinninghe Damsté et al., 2011, 2014, 2018). Des quantités importantes de brGDGTs et d'Acidobactéries ont ainsi été observées au sein d'une carotte de tourbe suédoise. D'autre part, des traces du composé I (m/z 1022 ; Fig. 1.24) ont été détectées dans une subdivision de culture d'Acidobactérie (Sinninghe Damsté et al., 2011), de grandes quantités d'acide isodiabolique (acide 13,16-diméthyl octacosanedioïque) lié à un glycérol – un précurseur possible des brGDGTs – étant elles présentes dans la plupart des subdivisions d'Acidobactéries connues (Sinninghe Damsté et al., 2014). Toutefois, des recherches supplémentaires sont encore nécessaires pour mieux contraindre la nature et diversité des microorganismes sources des brGDGTs. Ceci est complexe, car la diversité des bactéries sur Terre reste largement inconnue et leur culture représente un énorme défi, moins d'1% des bactéries étant cultivables (Amann et al., 1995).

Les tetraéthers d'origine bactérienne sont des composés ubiquistes (Schouten et al., 2013), présents dans les sols (Weijers et al., 2006b, 2007; Peterse et al., 2012), tourbières (Huguet et al., 2013; Naafs et al., 2017b), sédiments lacustres (Pearson et al., 2011; Russell et al., 2018), côtiers et marins (Zell et al., 2014; Sinninghe Damsté, 2016). Ils ont fait l'objet d'un intérêt grandissant ces dernières années, car leur distribution varie en fonction des paramètres environnementaux (température et pH principalement) et ils peuvent donc être utilisés en vue de reconstructions paléoenvironnementales, comme discuté ci-dessous.

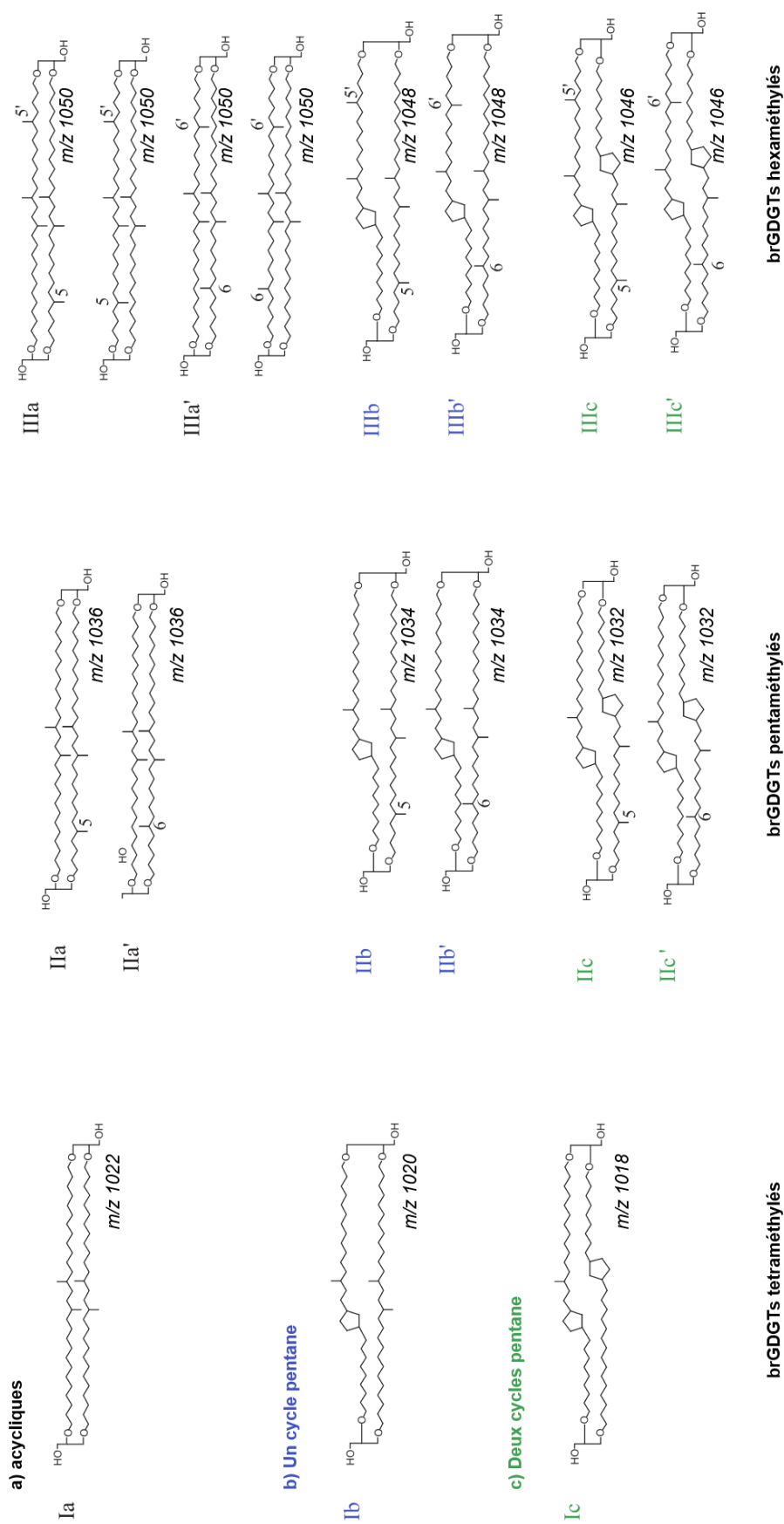


Figure 1.25 . Structure et nomenclature des tétraéthers ramifiés bactériens, avec les isomères de position 5- et 6-méthyles, les nomenclatures associées (en chiffres romain) et les rapports masse/charge (*m/z*) des différents composés (d'après De Jonge et al., 2014)

- Utilisation des tétraéthers bactériens comme marqueurs de température et de pH en milieu terrestre

L'analyse des brGDGTs dans un grand nombre de sols provenant du monde entier ($n=134$) a montré que leur abondance relative était principalement influencée par la température et le pH (Weijers et al., 2007). Ainsi, le nombre moyen de cycles pentane des brGDGTs, reflété par l'indice CBT (Eq. 1.12), est corrélé avec le pH du sol, tandis que le nombre moyen de groupes méthyle de ces mêmes composés, appelé indice MBT (Eq. 1.13), est corrélé avec la température moyenne annuelle de l'air (TMAA) et, dans une moindre mesure, avec le pH du sol (Weijers et al., 2007). Les indices CBT et MBT sont définis comme suit :

$$CBT = -\log \left(\frac{[Ib]+[IIb]}{[Ia]+[IIa]} \right) \quad (\text{Eq. 1.12})$$

$$MBT = \frac{[Ia+Ib+Ic]}{[Ia+Ib+Ic]+[IIa+IIa'+IIb+IIb'+IIc+IIc']+[IIIa+IIIa'+IIIb+IIIb'+IIIc+IIIc']} \quad (\text{Eq.1.13})$$

Les chiffres romains présents au sein des équations 1.12 et 1.13 se réfèrent aux abondances relatives des brGDGTs présentés en Fig. 1.25.

Des régressions linéaires entre les indices MBT/CBT et les paramètres TMAA/pH ont été établies par Weijers et al. (2007 ; Fig 1.26) et correspondent aux Eqs 1.14 et 1.15 :

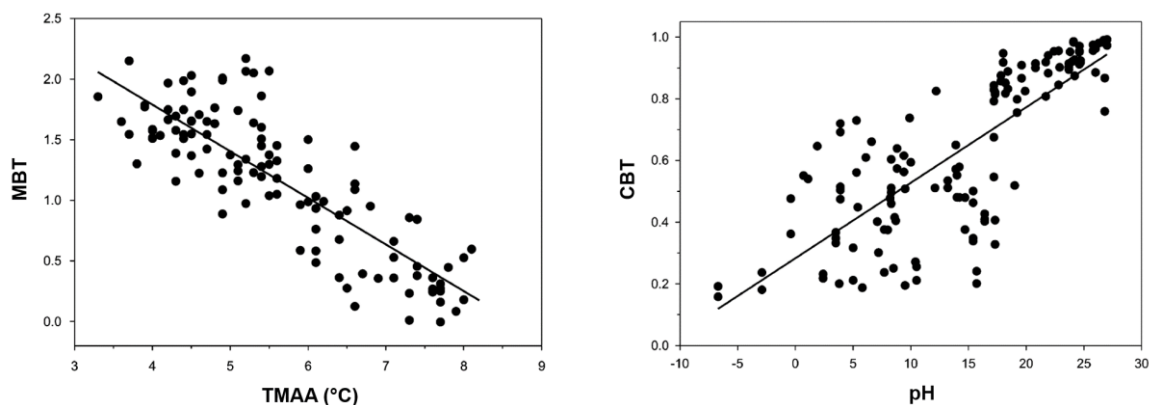


Figure 1.26 Calibrations proposées par Weijers et al. (2007) entre les indices MBT/CBT et la TMAA et le pH respectivement

$$\text{CBT} = 3,33 - \text{pH} \times 0,38 \quad (R^2=0.70 ; n = 134) \quad (\text{Eq. 1.14})$$

$$\text{MBT} = 0,28 + \text{TMAA} \times 0,025 \quad (R^2=0,62 ; n = 134) \quad (\text{Eq. 1.15})$$

Le MBT est également fortement influencé par le pH des sols (Weijers et al., 2007). Une régression multiple entre les valeurs de TMAA, l'indice CBT et l'indice MBT (Eq. 1.16) montre ainsi un meilleur coefficient de corrélation ($R^2 = 0,77$; Fig 1.27) que pour la relation entre la TMAA et l'indice MBT seul (Fig 1.26) :

$$\text{MBT} = 0,122 + 0,187 \times \text{CBT} + 0,020 \times \text{TMAA} \quad (R^2 = 0,77; \text{EQM} = 4,8^\circ\text{C}; n = 134) \quad (\text{Eq. 1.16})$$

Néanmoins, même si cette régression multiple est statistiquement significative ($p < 0.05$), et présente un coefficient de régression satisfaisant ($R^2 = 0,77$), la dispersion associée reste importante ($\text{EQM} = 4,8^\circ\text{C}$).

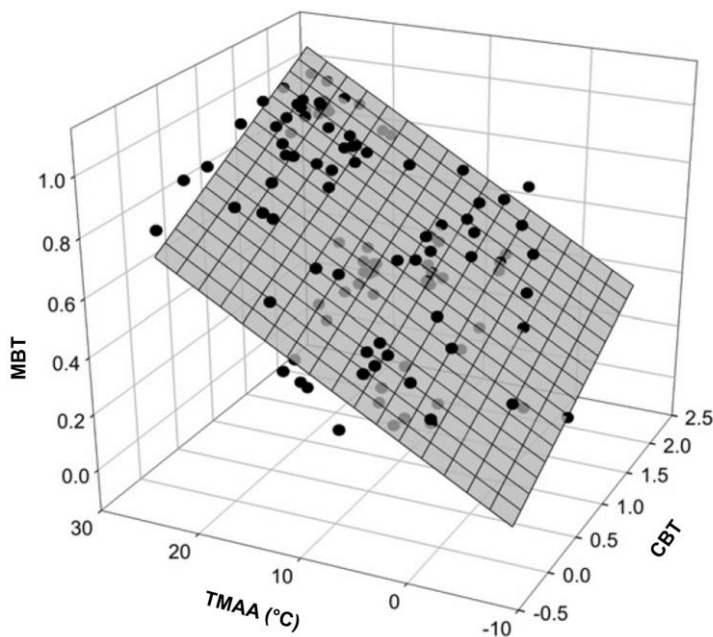


Figure 1.27 Régression linéaire multiple entre les indices MBT, CBT et la TMAA établie pour 134 sols (Eq. 1.16 ; Weijers et al., 2007).

Peterse et al. (2012) ont ensuite proposé une version modifiée du MBT, le MBT' (Eq. 1.17). Il diffère du MBT par la non-prise en compte des composés IIIb et IIIc, en général peu abondants dans les sols.

$$\text{MBT}' = \frac{[Ia+Ib+Ic]}{[Ia+Ib+Ic] + [IIa+IIa'+IIb+IIb'+IIc+IIc'] + [IIIa+IIIa']} \quad (\text{Eq. 1.17})$$

Une nouvelle régression multiple entre le MBT', le CBT et la TMAA a été établie (Peterse et al., 2012 ; Eq. 1.18) en ajoutant des sols supplémentaires aux échantillons de Weijers et al. (2007) :

$$TMAA = 0,81 - 5,67 \times CBT + 31,0 \times MBT' \quad (R^2 = 0,59; EQM = 5,0^\circ C; n = 176)$$

(Eq.1.18)

Par la suite, l'amélioration des méthodes d'analyse des brGDGTs, avec l'utilisation de colonnes de silice en série (De Jonge et al., 2013, 2014), a conduit à la détection et séparation des isomères 5- et 6-méthyles des brGDGTs (Fig. 1.25). Les isomères 6-méthyles apparaissent fortement dépendants du pH du sol et les isomères 5-méthyles de la température (De Jonge et al., 2014). Un nouvel indice, excluant les isomères 6-méthyle et basé uniquement sur les homologues 5-méthyle, l'indice MBT'_{5Me}, a par conséquent été proposé (Eq. 1.18 ; De Jonge et al., 2014) :

$$MBT'_{5Me} = \frac{[Ia+Ib+Ic]}{[Ia+Ib+Ic] + [IIa+IIb+IIc] + [IIIa]} \quad (\text{Eq. 1.18})$$

Les chiffres romains présents au sein de l'équation 1.18 se réfèrent aux abondances relatives des structures présentées en Fig. 1.25.

De nouvelles régressions globales entre le MBT'_{5Me} et la TMAA, ainsi qu'entre l'indice CBT et le pH ont été proposé par De Jonge et al. 2014 (Eqs. 1.19 et 1.20) :

$$TMAA = -8,57 + 31,45 \times MBT'_{5Me} \quad (R^2 = 0,64; EQM = 4,8^\circ C; n = 231; \text{Eq. 1.19})$$

$$pH = 7,15 + 1,59 \times CBT \quad (R^2 = 0,85; EQM = 0,52; n = 231 ; \text{Eq. 1.20})$$

L'indice MBT'_{5Me} permet ainsi de reconstituer les paléotempératures de manière directe, à l'aide d'une régression linéaire simple, non influencée par le pH. Néanmoins, la dispersion associée à cette calibration (EQM = 4,8°C) reste importante.

Par la suite, Dang et al. (2016a, b) ont montré que l'indice MBT' était significativement corrélé avec la TMAA dans des sols chinois quand ceux-ci présentaient

une proportion plus importante de brGDGT méthylés en 5^{ème} position qu'en 6^{ème} position (rapport IR_{6Me} < 0.5 ; Eq. 1.21) :

$$IR_{6Me} = \frac{[IIa+IIb'+IIc'+IIIa+IIIb'+IIIC']}{[IIa+IIb'+IIc'+IIIa+IIIb'+IIIC'+IIa+IIb+IIc+IIIa+IIIb+IIIC]} \text{ (Eq. 1.21)}$$

Sur cette base, Naafs et al., (2017a) ont ainsi proposé une nouvelle calibration globale, ne tenant compte que des sols avec un IR_{6Me} > 0.5 (n=177 ; R²=0.76 ; EQM = 4.1°C). Cette calibration est substantiellement meilleure que celle proposée par De Jonge et al. (2014), même si effectuée sur un nombre plus restreint d'échantillons de sols.

Malgré les améliorations analytiques et celles des calibrations proposées, l'EQM associée à la reconstruction de la TMAA et du pH à l'aide des tétraéthers bactériens reste élevée. Plusieurs hypothèses ont été proposées pour expliquer la dispersion observée sur les différentes calibrations (Peterse et al., 2012; De Jonge et al., 2014; Naafs et al., 2017a) :

(1) Une hétérogénéité importante des différents sols étudiés.

(2) L'influence probable d'autres paramètres environnementaux que la TMAA et le pH sur l'abondance relative des tétraéthers bactériens, comme par exemple l'humidité relative des sols, associée au régime de précipitations (Huguet et al., 2010; Dirghangi et al., 2013; Anderson et al., 2014; Menges et al., 2014; Zell et al., 2014; Yang et al., 2015; Ding et al., 2015; Lei et al., 2016). Il a ainsi été observé une corrélation plus faible entre le degré de méthylation des brGDGTs (indices MBT, MBT'_{5Me}) et la TMAA dans les sols arides (< 500 mm précipitation/an) ou alcalins (Peterse et al., 2012; Dirghangi et al., 2013; Anderson et al., 2014; Zell et al., 2014; Ding et al., 2015). Une surestimation du pH estimé à partir des brGDGTs a également été observée dans les sols arides (Yang et al., 2014; Zell et al., 2014). Outre l'humidité, plusieurs paramètres, tels que la nature du sol (Mueller-Niggemann et al., 2016; Davtian et al., 2016), les communautés végétales (Naeher et al., 2014; Liang et al., 2019) ou la saisonnalité (Weijers et al., 2011a; Huguet et al., 2013; De Jonge et al., 2019) pourraient également influencer la distribution des brGDGTs dans les sols et donc jouer sur les reconstructions de température effectués à partir de ces composés.

(3) Enfin, la différence entre les températures moyennes annuelles de l'air reconstituées et les températures des sols de surfaces prélevés pourrait expliquer la

dispersion des points sur les calibrations. En première approximation, il est souvent considéré que les températures de surface des sols sont liées aux variations de températures de l'air. Néanmoins, il est nécessaire de garder à l'esprit ce biais potentiel lors de construction de calibrations et de l'utilisation de ces marqueurs.

Plusieurs études recommandent l'utilisation de calibrations non pas globales mais locales (par exemple; Tierney et al., 2010; Yang et al., 2015; Ding et al., 2015), qui sont par nature plus précises, car intégrant les spécificités environnementales d'une région donnée. Mais l'utilisation de ces calibrations « locales » soulève plusieurs questions : qu'en est-il de la fiabilité d'une telle calibration dans le cadre de la reconstruction d'événements climatiques dans un passé lointain (i.e. au-delà du Quaternaire) ? Cela nécessite l'hypothèse d'un environnement qui n'aurait pas ou peu varié au cours d'une longue période de temps, ou alors de limiter l'utilisation de ces calibrations aux seules reconstructions lors du Quaternaire. En outre, l'utilisation de calibrations locales pose une question d'échelle spatiale. Est-il nécessaire de produire des calibrations locales pour chaque continent ? Chaque région climatique du globe ? Chaque pays, voire chaque lac ou montagne ? Une calibration globale, certes moins précise, a l'avantage de permettre la comparaison de différentes paléoreconstructions et de faciliter son utilisation par l'ensemble de la communauté scientifique. Outre les modèles linéaires simples, plusieurs méthodes ont été récemment utilisées pour définir des calibrations globales à partir de l'abondance relative des brGDGTs ou de leurs indices associés, telles que les régressions multiples (Russell et al., 2018), la régression de Deming (Naafs et al., 2017a), les modèles de régression linéaire bayésiens (Tierney and Tingley, 2014; Dearing Crampton-Flood et al., 2020) ou les modèles d'apprentissage machine supervisé (Dunkley Jones et al., 2020). Ces approches montrent que le développement conjoint des méthodes analytiques et de traitement de données permettront une meilleure compréhension et utilisation des marqueurs d'environnement.

- **Tetraéthers isopréniques et ramifiés : Indice BIT**

Outre les reconstructions de température et de pH, les tetraéthers peuvent également être utilisés pour estimer les apports de matière organique terrigène en milieu aquatique

(e.g. Weijers et al., 2009b; Zell et al., 2013) via l'indice BIT (Branched and Isoprenoid Tetraethers ; Hopmans et al., 2004) défini comme suit :

$$\text{BIT} = \frac{[Ia+(IIa+IIa')+(IIIa+IIIa')]}{[\text{Crenarchaeol}+Ia+(IIa+IIa')+(IIIa+IIIa')]} \quad (\text{Eq. 1.22})$$

Les chiffres romains présents au sein de l'équation 1.22 se réfèrent aux abondances relatives des structures présentées en Fig. 1.23 et 1.25.

La définition de cet indice est basée sur l'hypothèse que les brGDGTs sont produits principalement au sein des environnements terrestres et le crénarchaeol (GDGT isoprénique, structure VI, partie 1.3.1.3.) uniquement en milieu aquatique (Sinninghe Damsté et al., 2000 ; Schouten et al., 2013, et références incluses). Néanmoins, de nombreuses études ont montré que les tetraéthers archéens et bactériens sont tous deux produits aussi bien en milieu terrestre qu'aquatique (eg; Tierney et al., 2010), faussant potentiellement les valeurs du BIT, qu'il convient donc d'interpréter avec précaution.

1.3.2.2. Validation et application des brGDGTs comme marqueurs de température et de pH en milieu continental

• Validation des brGDGTs comme marqueurs de température et de pH

Plusieurs études « mécanistiques » ont été réalisées sur des échantillons de sols, de tourbières et de lacs actuels afin de mieux comprendre comment les brGDGTs sont influencés par les variables environnementales dans ces écosystèmes. Ainsi, les brGDGTs ont été analysés au sein d'échantillons de sols de surfaces prélevés le long de différents transects altitudinaux dans le monde entier. L'avantage de ces transects altitudinaux est de pouvoir disposer au sein d'une région restreinte de variations naturelles de paramètres environnementaux, en particulier la température, avec l'altitude. En première approximation, il est souvent considéré que seule la température varie et les autres paramètres environnementaux sont considérés comme constants. La distribution des brGDGTs le long de gradients altitudinaux a par exemple été déterminée en Chine – Mts. Meghalaya, (Ernst et al., 2013); Jianfengling, (Yang et al., 2010); Shennongjia, (Yang et al., 2015); Gongga, (Peterse et al., 2009a); Xiangpi, (Liu et al., 2013) – en Afrique – Mt. Kilimanjaro (Sinninghe Damsté et al., 2008), Mt Rungwe (Coffinet et al., 2014; Coffinet,

2015; Huguet et al., 2019) – ou en Amérique du Sud (Andes péruviennes ; Anderson et al., 2014)).

Par ailleurs, l'analyse des brGDGTs dans des échantillons modernes de tourbières, particulièrement en Europe (Weijers et al., 2004, 2006a, 2011; Huguet et al., 2010, 2013), suggère que ces composés seraient produits par des microorganismes préférentiellement anaérobies. Les tourbières étant des sols particuliers, riches en matière organique et très humides, il est apparu nécessaire de développer des calibrations spécifiques entre la distribution des brGDGTs et la température/le pH pour ces milieux (Naafs et al., 2017b).

Enfin, de nombreuses études se sont intéressés aux brGDGTs en milieu lacustre (Blaga et al., 2009; Sinninghe Damsté et al., 2009; Tierney and Russell, 2009; Powers et al., 2010; Loomis et al., 2011). Ces composés peuvent avoir plusieurs origines : production *in situ* dans la colonne d'eau et/ou dans les sédiments, et/ou apports terrigènes via l'érosion des sols (Tierney, 2012; Loomis et al., 2014; Buckles et al., 2014). Dans de nombreux lacs, la production *in situ* de brGDGTs est non négligeable, et les tétraéthers présentent une signature différente de celle observée dans les sols des bassins versants associés. Il est donc nécessaire de développer des calibrations spécifiques aux milieux lacustres entre la distribution des brGDGTs et la température/le pH (Pearson et al., 2011; Loomis et al., 2012; Russell et al., 2018).

• Applications paléoenvironnementales

Les brGDGTs ont été utilisés comme marqueurs de température et pH dans de nombreuses archives sédimentaires terrestres, en particulier des séquences de loess-paléosol. Parmi les sites étudiés, se trouve le « Chinese Loess Plateau » (CLP) également appelé « Yellow Earth Plateau » (Wang et al., 2020 et références incluses), qui présente des séquences de loess-paléosol sur plus de 640 000 km² et permet de reconstituer l'histoire climatique sur un intervalle allant de 2,4 à 2,6 Ma à l'Holocène (Lu et al., 2016; Wang et al., 2020). Grâce à l'application sur ces séquences de loess-paléosols des indices dérivés des brGDGTs (MBT'/CBT), les variations de températures de l'air ont été reconstruites sur les 130 000 dernières années (Peterse et al., 2011, 2014; Gao et al., 2012; Jia et al., 2014; Lu et al., 2016). Les brGDGTs ont également permis de reconstituer les

températures dans des séquences de loess en Europe de l'est (e.g. Schreuder et al., 2016), ou au Tibet, par exemple (e.g. Ding et al., 2015; Feng et al., 2019).

Les brGDGTs ont par ailleurs été utilisés comme marqueurs d'environnement dans des tourbières, e.g. en Arctique (archive datant du début du Pliocène ; Ballantyne et al., 2010), le Jura suisse (Weijers et al., 2011b), en Afrique (Coffinet et al., 2018) et en Asie (Zheng et al., 2015; Wu et al., 2020). Plusieurs applications des brGDGTs comme paléothermomètres concernent par ailleurs les milieux lacustres, par exemple en Asie (Cao et al., 2017; Ning et al., 2019), en Atlantique nord (de Wet, 2017), en Amérique (Miller et al., 2018) et en Europe (Martin et al., 2019).

1.3.2.3. Les acides gras 3-hydroxylés

Outre les tétraéthers, des études récentes (C. Wang et al., 2016; Wang et al., 2018; Huguet et al., 2019; Yang et al., 2020) ont révélé le potentiel d'une autre famille de lipides comme nouveaux marqueurs de température et de pH en milieu terrestre : les acides gras 3-hydroxylés (3-AGH). Les sources, structures et applications potentielles de ces composés comme marqueurs d'environnements sont explicitées dans cette section.

• Structures et sources des acides gras 3-hydroxylés

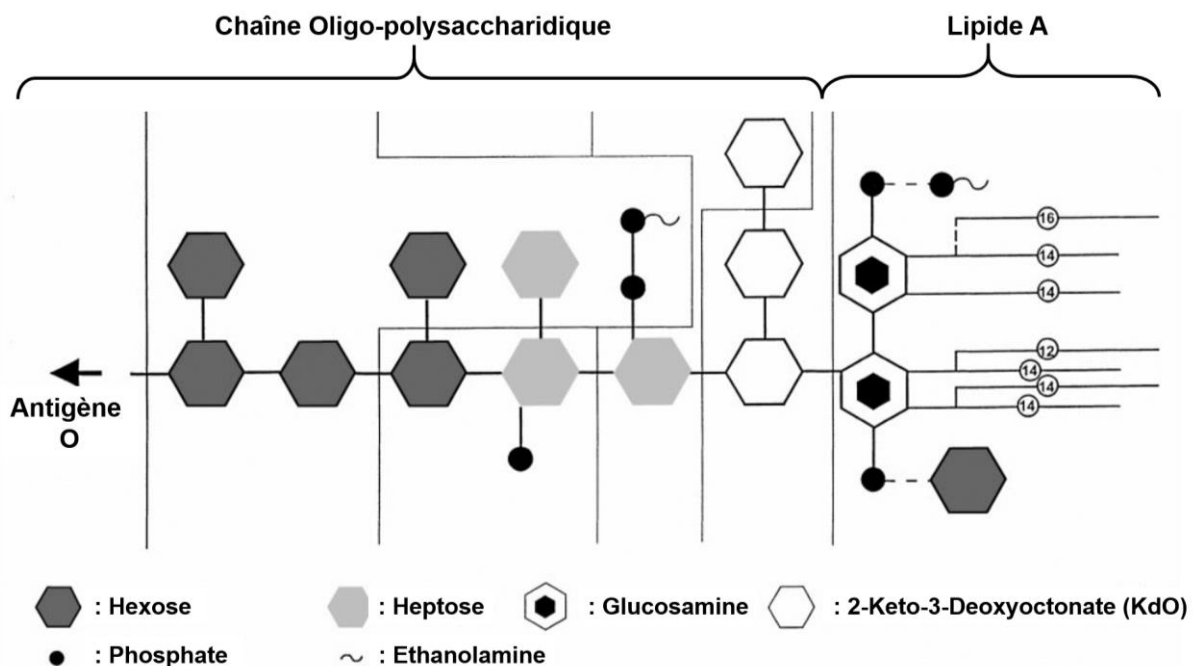


Figure 1.28 Structure schématique du LPS (lipopolysaccharide) ancré dans la membrane par l'intermédiaire du lipide A (d'après Brandenburg et al., 1999). Sur le lipide A sont montrées les chaînes d'acides gras (de différentes longueurs).

Les lipopolysaccharides (LPS) sont le composant principal de la membrane externe des bactéries Gram-négatif (partie 1.2.1.3.1 ; Fig. 1.13). Ce sont des polymères de sucres rattachés à la membrane cellulaire par l'intermédiaire d'une structure appelée lipide A (Brandenburg et al., 1999; Bryant et al., 2010).

Le lipide A est composé d'unités glucosamines phosphorylées liées à des acides gras, en particulier les acides gras 3-hydroxylés (Madigan et Martinko, 2018 ; Fig 1.28, 1.29). Il est responsable du caractère pathogène des bactéries Gram-négatif, mais, étant

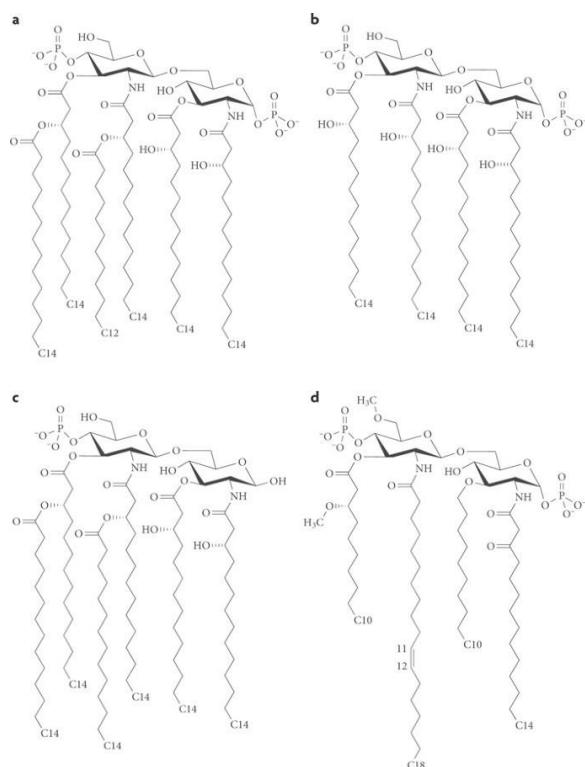


Figure 1.29 Quelques exemples de structures du lipide A, constitué d'unités glucosamines, ancrées dans la membrane par l'intermédiaire d'acides gras (notamment les acides gras 3-hydroxylés). Adapté de Bryant et al. (2010).

ancré dans la membrane, il ne s'active que lors de la mort et donc de la lyse de l'organisme qui permet sa libération (Sonesson et al., 1990; Mielniczuk et al., 1993; Saraf et al., 1997; Szponar et al., 2002, 2003; Keinänen et al., 2003; Wakeham et al., 2003; Lee et al., 2004; Krašnik et al., 2006).

Le lipide A est sujet à une importante variabilité, de différents types :

(1) variabilité inter-espèces—le profil lipidique varie en fonction des espèces de bactéries Gram négatif ;

(2) adaptation « pathogène »— la bactérie source du lipide A est susceptible de modifier la composition de ce dernier pour s'adapter à son hôte et développer une réponse immunitaire ;

(3) changement de structure en réponse à un stress environnemental (Sali, 2014 et références incluses).

Les acides gras 3-hydroxylés (3-AGH) présents au sein du lipide A contiennent 10 à 18 atomes de carbone et un groupement hydroxyle en 3^{ème} position (Fig. 1.30). Même si les 3-AGH bactériens ont majoritairement été détectés chez des bactéries dites à Gram-négatif, une étude a suggéré qu'une souche de bactérie Gram-positif (*Lactobaccillus plantarum*) pouvait produire ces composés (Sjögren et al., 2003). On note par ailleurs l'existence de 3-AGH avec des chaînes carbonées plus longues (> 18 C), produits par d'autres types d'organismes. Ainsi, des 3-AGH présentant C₂₆ à C₃₀ ont été détectés au sein de souches de microalgues de la classe des Eustigmatophytes (Volkman et al., 1999a), et au sein de souches de champignons (Van Dyk et al., 1994 et références incluses). Dans ce manuscrit, nous nous focaliserons sur les 3-AGH d'origine bactérienne (homologues C₁₀ à C₁₈).

On peut distinguer trois types de 3-AGH : à chaîne normale et à chaîne ramifiée, soit *iso* soit *anteiso* (Fig. 1.30). Les 3-AGH *iso* et *anteiso* possèdent une ramification

méthyle en avant-dernière ou en antépénultième position, respectivement (Fig. 1.30). Ces composés ont été largement utilisés (i) pour détecter la présence d'endotoxines libérées par ces microorganismes, notamment dans le cadre d'études cliniques (Wollenweber and Rietschel, 1990; Saraf et al., 1997; Szponar et al., 2002, 2003; Keinänen et al., 2003) et (ii) pour quantifier les bactéries Gram-négatif dans des échantillons environnementaux tels que les aérosols (Lee et al., 2004; Cheng et al., 2012), la matière organique dissoute (Wakeham et al., 2003) ou les sols (Zelles et al., 1995; Zelles, 1999). Néanmoins, jusqu'à récemment, l'effet des paramètres environnementaux sur la structure de ces composés et leur rôle dans l'adaptation homéovisqueuse des bactéries n'avaient été que peu étudiés (C. Wang et al., 2016; Huguet et al., 2019; Yang et al., 2020).

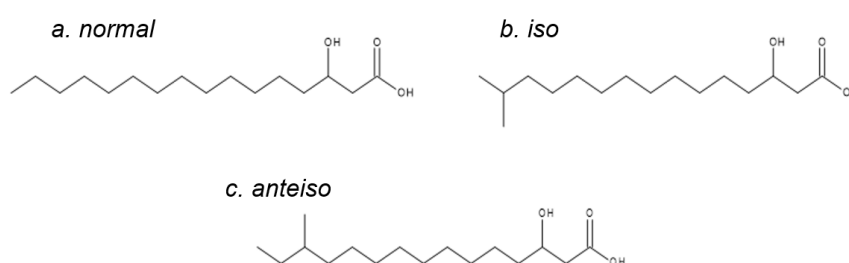


Figure 1.30 Les différents isomères des acides gras 3-hydroxylés.

- **Les 3-AGH comme marqueurs de température et de pH en milieu terrestre**

C. Wang et al. (2016) ont récemment mis en exergue le potentiel des 3-AGH comme marqueurs de température et de pH en milieu terrestre. Ces auteurs ont analysé les 3-AGH dans 26 sols prélevés entre 315 et 2840 m d'altitude le long du Mont Shennongjia (Chine) et ont établi différentes corrélations entre l'abondance relative de ces composés et la température de l'air ou le pH. C. Wang et al. (2016) ont tout d'abord montré que l'indice RIAN (Eq. 1.23), qui reflète l'abondance relative des 3-AGH C₁₀ à C₁₈ ramifiés (*iso* et *anteiso*) par rapport à ceux à chaînes non ramifié (normaux), présentait une corrélation significative avec le pH des sols (C. Wang et al., 2016 ; Fig 1.31) :

$$\text{RIAN} = -\log [3\text{-AGH } iso + 3\text{-AGH } anteiso / 3\text{-AGH } normaux] \text{ (Eq. 1.23)}$$

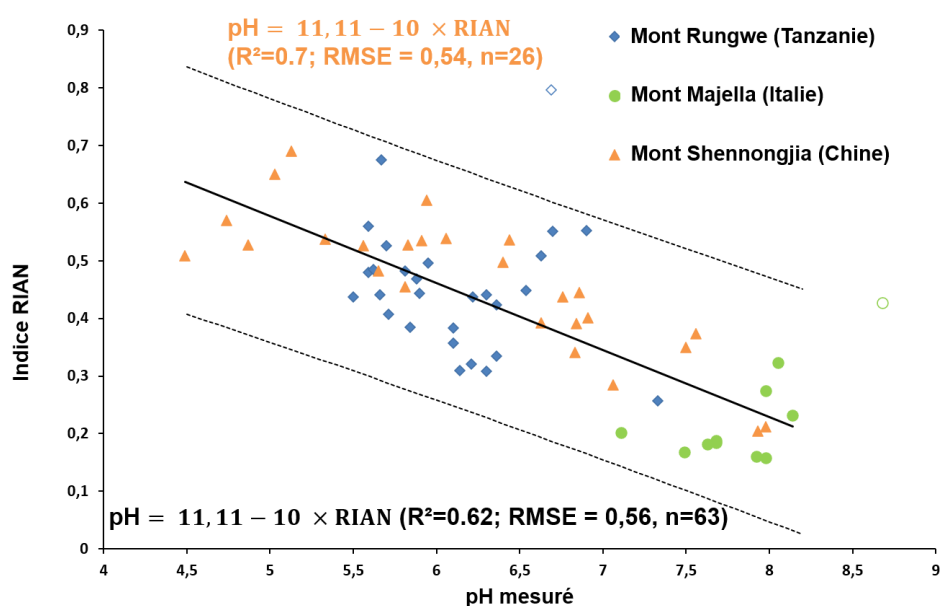


Figure 1.31 Régression linéaire entre l'indice RIAN et les valeurs de pH mesurées au sein des échantillons de sols prélevés le long des transects du Mont Shennongjia (Chine; C.Wang et al., 2016), et des Mont Majella (Italie) et Rungwe (Tanzanie ; Huguet et al., 2019). La droite noire est la régression globale pour l'ensemble des échantillons de sols proposée par Huguet et al. (2019).

Huguet et al. (2019) a confirmé le potentiel des 3-AGH comme marqueurs de pH en milieu terrestre grâce à l'étude d'échantillons de sols prélevés le long de deux autres transects altitudinaux : le Mont Majella (Italie, $n=11$), et le Mont Rungwe (Tanzanie, $n=28$). Lorsque ces deux transects sont considérés individuellement, il y a une absence ou une faible relation linéaire entre les valeurs de pH mesurées et l'indice RIAN, ce qui pourrait être dû à la faible gamme de variation de pH le long de ces deux montagnes. Cependant, lorsque l'ensemble des données disponibles sont combinées, une relation linéaire statistiquement significative entre l'indice RIAN et le pH est obtenue (Fig 1.31 ; Huguet et al., 2019) :

En parallèle, C. Wang et al. (2016) a montré le potentiel des 3-AGH, comme marqueurs de température dans les sols prélevés le long du Mont Shennongjia, au travers de deux nouveaux indices : le RAN_{15} et le RAN_{17} . Ces indices sont définis comme étant le rapport des 3-AGH *anteiso* C_{15}/C_{17} sur les 3-AGH *normaux* C_{15}/C_{17} (Fig. 1.32 ; Eqs. 1.24 et 1.25) :

$$\text{RAN}_{15} = [\text{anteiso } C_{15}] / [\text{normal } C_{15}] \text{ (Eq. 1.24)}$$

$$\text{RAN}_{17} = [\text{anteiso } C_{17}] / [\text{normal } C_{17}] \text{ (Eq. 1.25)}$$

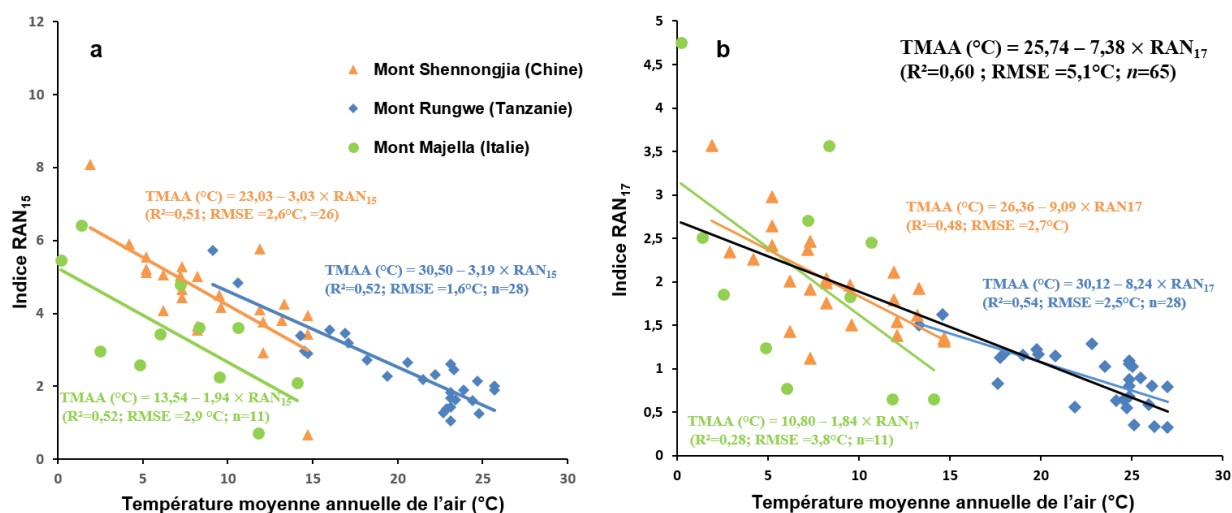


Figure 1.32 Régressions linéaires entre a) l'indice RAN_{15} et b) l'indice RAN_{17} et les valeurs de températures moyennes annuelles de l'air au sein des échantillons de sols prélevés le long des transects du Mont Shennongjia (Chine ; C.Wang et al., 2016), et des Mont Majella (Italie) et Rungwe (Tanzanie ; Huguet et al., 2019). La droite noire sur le graphique b correspond à la régression globale pour l'ensemble des échantillons de sols proposée par Huguet et al. (2019).

Comme pour le pH, Huguet et al. (2019) a confirmé le potentiel des 3-AGH comme marqueurs de température en milieu terrestre grâce à l'étude d'échantillon de sols prélevés le long des transects des Monts Majella (Italie, $n=11$) et Rungwe (Tanzanie, $n=28$). Des relations linéaires statistiquement significatives entre les indices RAN_{15} et RAN_{17} et les valeurs de TMAA ont pu être obtenues (Fig. 1.32).

Les études effectuées le long des trois transects altitudinaux (C. Wang et al., 2016; Huguet et al., 2019) suggèrent que les bactéries Gram-négatif réagissent de la même manière face aux variations de températures avec une augmentation de la proportion des 3-AGH *anteiso*- C_{15}/C_{17} par rapport aux homologues *normaux*- C_{15}/C_{17} lorsque la température diminue, permettant de maintenir une fluidité appropriée de la membrane bactérienne quelle que soit la température. Les relations linéaires entre l'indice RAN_{15} et les températures moyennes annuelles de l'air présentent des pentes similaires mais des ordonnées à l'origine différentes le long des Mts. Shennongjia, Rungwe et Majella (C. Wang et al., 2016 ; Huguet et al., 2019) et semblent donc site-dépendantes. Cela suggère

que des calibrations régionales pourraient être plus adaptées pour appliquer le RAN_{15} comme paleothermomètre. A l'inverse, une régression significative ($R^2 = 0,60$; Fig 1.27b) a pu être établie entre l'indice RAN_{17} et les TMAA estimées le long des Mts. Shennongjia, Rungwe et Majella (C. Wang et al., 2016 ; Huguet et al., 2019).

Les régressions linéaires entre les données de pH/TMAA et les indices basés sur la distribution des 3-AGH, bien que prometteuses, ont été obtenues à partir d'un nombre d'échantillons de sols restreint (environ 70) et montrent un degré de dispersion élevé, conduisant à une erreur quadratique moyenne non négligeable (TMAA-EQM = $5,1^\circ\text{C}$; pH-EQM = $0,56$; Huguet et al., 2019). La distribution des 3-AGH dans les sols pourrait être influencée par des paramètres environnementaux autres que la température et le pH, comme observé pour les brGDGTs. La variabilité de la composition en 3-AGH au sein des différentes espèces de bactéries Gram-négatif (Sali, 2014) doit également être prise en compte.

Les indices calculés à partir des 3-AGH ont toutefois été récemment appliqués pour estimer les changements de température et d'hydrologie au cours des 10 000 dernières années sur un spéléothème en Chine (Wang et al., 2018), montrant le potentiel de ces molécules en tant qu'outils indépendants pour les paléoreconstructions dans les milieux terrestres. Une étude très récente basée sur des échantillons de sédiments marins de l'océan Pacifique Nord a par ailleurs suggéré que la distribution des 3-AGH pourrait également être utilisée pour reconstruire la température de surface en milieu océanique (Yang et al., 2020).

1.3.3. Conclusion partielle

Cet état de l'art, non exhaustif, des biomarqueurs existants utilisés dans le cadre de reconstitutions paléoenvironnementales et paléoclimatiques permet de soulever plusieurs remarques :

(1) **Peu de marqueurs en milieux terrestres** : La grande majorité des biomarqueurs ont été développées en domaine océanique, alors que la surface continentale est un composant majeur du système climatique, de par ses interactions complexes avec le compartiment atmosphérique - l'atmosphère transporte deux fois plus d'énergie, sous forme de chaleur, que l'océan (Ruddiman, 2001).

(2) **Des incertitudes sur les biomarqueurs existants** : Le potentiel des biomolécules préservées dans les archives sédimentaires dans l'objectif de reconstituer des paléoenvironnements est immense, mais ceux-ci présentent de nombreuses contraintes. Certaines de ces limitations sont partagées par beaucoup de marqueurs utilisés en paléoenvironnement, comme l'impact de la saisonnalité, le manque de connaissances sur les organismes sources ou l'utilisation de calibrations locales vs. globales.

(3) **L'influence des variables autres que la température et le pH sur les biomarqueurs organiques microbiens reste mal connue.**

Dans ce contexte, mon travail de thèse a pour objectif premier d'évaluer l'applicabilité des acides gras 3-hydroxylés comme marqueurs de température et de pH en milieu terrestre et leur complémentarité avec les brGDGTs. Pour ceci, je propose de répondre à plusieurs grands objectifs :

(1) *Comprendre l'influence de la température de l'air et du pH du sol, ainsi que d'autres variables environnementales, sur la distribution des 3-AGH et des brGDGTs*

(2) *Obtenir des informations sur la vitesse d'adaptation des bactéries produisant des 3-AGH et des brGDGTs dans les sols.*

(3) *Tester l'applicabilité des 3-AGH comme marqueurs de température de l'air et de pH du sol en milieu terrestre à l'échelle mondiale. Développer une calibration globale reliant l'abondance relative des 3-AGH avec la température de l'air d'une part et le pH des sols d'autre part.*

(4) Proposer une nouvelle calibration globale de température de l'air basée sur les brGDGTs en milieu terrestre.

(5) Valider l'applicabilité des calibrations globales proposées sur les 3-AGH et les brGDGTs en les testant sur des archives sédimentaires connues.

Afin de répondre à ces différents objectifs, une double approche a été initiée, mêlant à la fois des échantillonnages sur le terrain, et des analyses en laboratoire. Les acides gras 3-hydroxylés et les tétraéthers bactériens ont été analysés dans différents sols prélevés le long de différents transects altitudinaux, mais aussi sur des sols choisis placés en incubations. Ces différentes expérimentations et les méthodes associées sont présentées dans la partie suivante.

Chapitre 2 – MATRIELS ET METHODES

2.1. Sites d'étude et échantillonnage

Dans le cadre de cette thèse, une étude géochimique et microbiologique détaillée a été réalisée sur des sols des Alpes françaises provenant d'un transect altitudinal composite (prélèvements de terrain, incubations en laboratoire). Un total de 49 échantillons de sol a été collecté. Après avoir été transportés au laboratoire, ils ont été directement stockés à -20 °C, puis lyophilisés, broyés et tamisés à 2 mm. Une certaine proportion des échantillons a été conservée à 4°C puis placés en incubations (expériences de microcosmes ; voir Partie 2.2.).

Afin de compléter cette étude, d'autres échantillons de sols issus de transects altitudinaux du monde entier ont été analysés. Ces échantillons proviennent de zones climatiques contrastées : Italie (Mont Pollino), Tibet (Mont Shegyla), Chili (Suelos Araucanos) et Pérou (Andes) (Fig 2.1 ; Tableau 2.1).

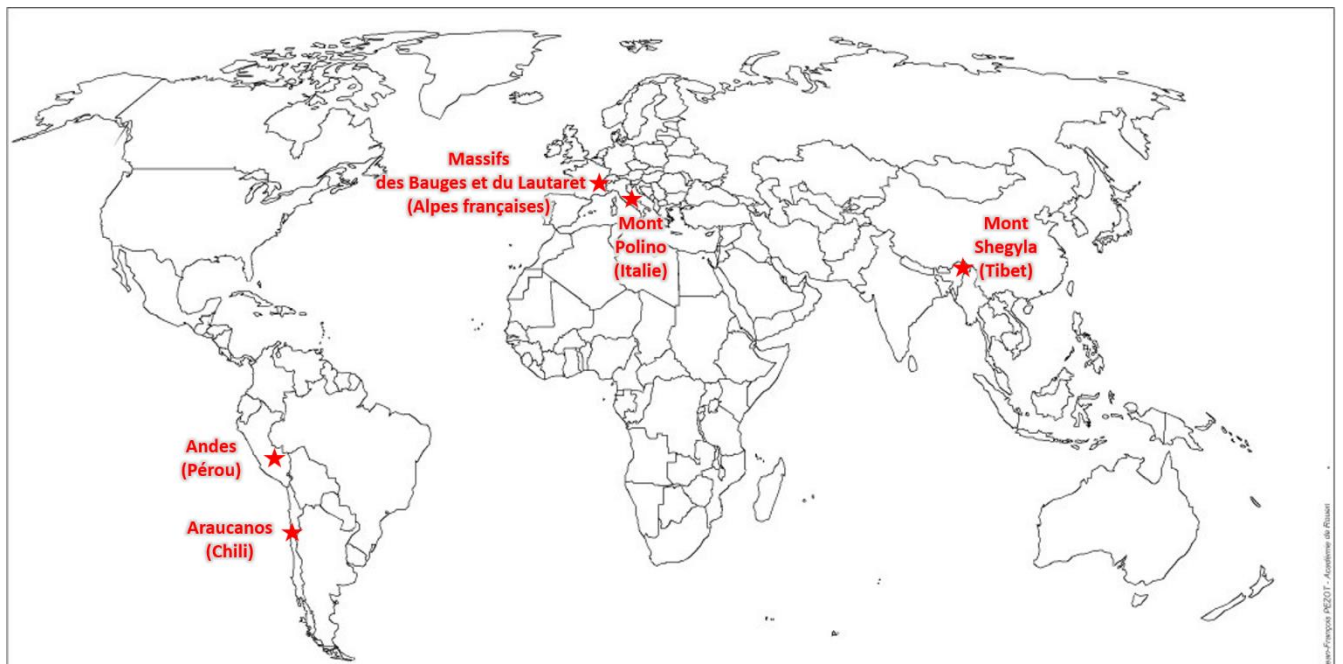


Figure 2.1 Emplacement des différents transects altitudinaux étudiés dans cette thèse.

MATERIELS ET METHODES

Sites d'étude	Echantillons	Lipides bactériens				Analyses globales, isotopiques et élémentaires								
		3-AGH	brGDGTs	TMAA (°C)	pH	Précipitation (mn/an)	%C org	% N	$\delta^{13}\text{C}$	$\delta^{15}\text{N}$	Humidité relative (%)	Granulométrie	CEC (cmol +/-kg)	Diversité microbienne
Alpes françaises	49	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
Incubations 3 mois	32	✓	✓	✓	✓									✓
Incubations 12 mois	32	✓	✓	✓	✓									✓
Mont Polino Italie	13	✓	✓	✓	✓									
Mont Shegyla Tibet	17	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
Andes Pérou	14	✓	✓	✓	✓	✓		✓						
Selos Araucanos Chili	10	✓	✓	✓	✓		✓							

Tableau 2.1 Données disponibles sur les différents transects altitudinaux étudiés. Les marques rouges correspondent à des données obtenues au cours de cette thèse.

2.1.1. Massifs des Bauges et du Lautaret-Galibier



Figure 2.2 Quelques photographies illustrant la diversité et la beauté des paysages rencontrés au sein des Alpes françaises.

J'ai eu l'opportunité de prélever, au côté de Jérôme Poulénard, des échantillons de sol de surface (0-10 cm de profondeur) au sein des Alpes françaises en octobre 2017. Ces échantillons ont été prélevés le long de deux toposéquences climatiques bien documentées : le massif des Bauges (entre 232 et 1475 m d'altitude) et le massif du Lautaret-Galibier (entre 1540 et 2700 m d'altitude). Un observatoire intégré à long terme a été mis en place dans la région dans le cadre de la Zone Atelier Alpes donnant accès à la mesure de la température moyenne annuelle de l'air et à une description précise des types de sol et des communautés végétales (Bounemoura et al., 1998; Carlson et al., 2017; Choler, 2018). Les massifs du Lautaret-Galibier et des Bauges couvrent une gamme d'altitude complémentaire et permettent d'obtenir un transect altitudinal composite représentatif des variations de température, des caractéristiques pédologiques et des communautés végétales au sein des Alpes françaises (Fig 2.2 et 2.3 ; Annexe 1, Supp. Fig 3.1).

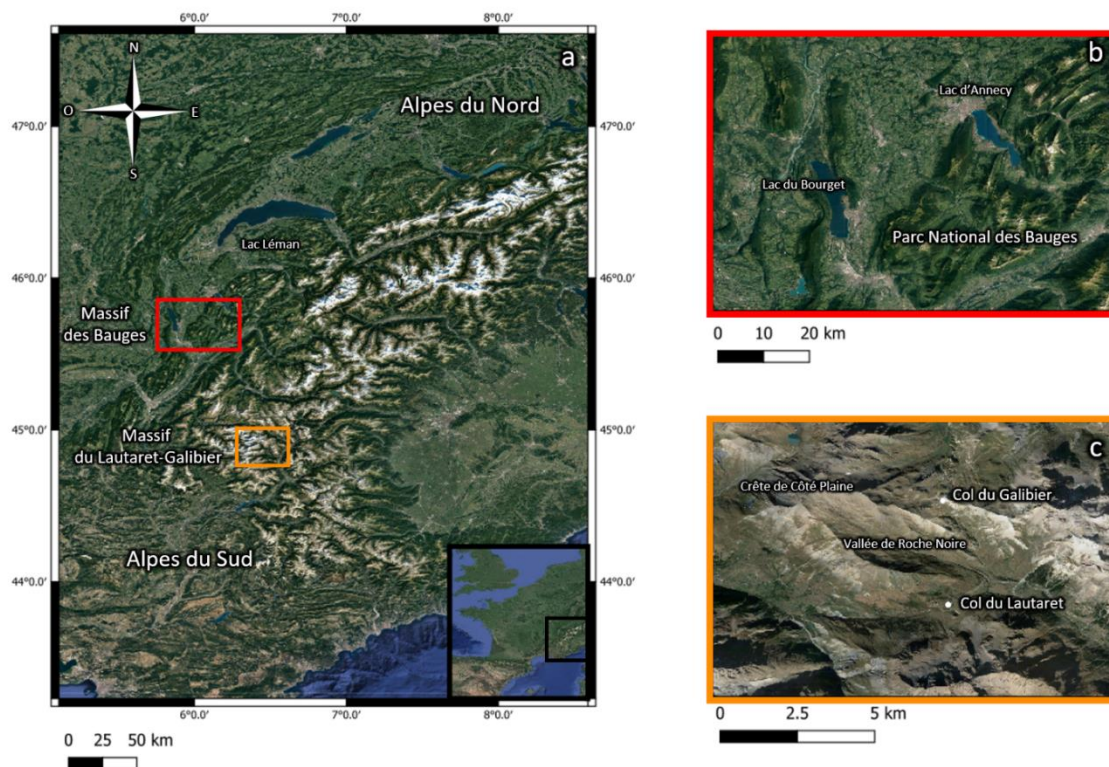


Figure 2.3 Cartes des (a) Alpes françaises, (b) du massif des Bauges et (c) du Lautaret-Galibier. Sources Google Satellites.

2.1.1.1. *Massif des Bauges*

Vingt-quatre échantillons de sol ont été prélevés le long du versant ouest du massif des Bauges. Ce massif, situé dans la partie nord-ouest des Alpes françaises, dans le Parc national français du massif des Bauges, est un site de basse à moyenne altitude (altitude maximale de 2217 m). Au nord-est, ce massif est bordé par un ensemble de montagnes et par le lac d'Annecy. Il est limité au sud-ouest par le lac du Bourget. Il s'agit d'un massif subalpin carbonaté, dominé par des chaînes calcaires orientées nord-nord-est / sud-sud-ouest. Son sommet est principalement composé de calcaires de l'Urgonien. Sur le versant sud, surplombant le lac du Bourget, le massif des Bauges offre un panorama escarpé de montagnes culminant à plus de 2 000 m. Dans la partie basse du massif, à 200 m d'altitude, la pluviométrie annuelle moyenne est de 1 221 mm et la Température Moyenne Annuelle de l'Air (TMAA en °C) de 11,4 °C (pour la période 1981-2010, station météorologique de Chambéry). Dans la partie supérieure du massif des Bauges, à une altitude d'environ 1 300 m, la pluviométrie moyenne annuelle est de 1 704,9 mm et la TMAA de 5,8 °C (pour

la période 1981-2000, station météorologique du Feclaz) (Bounemoura et al., 1998; Carlson et al., 2017; Choler, 2018).

La végétation dans la partie inférieure du massif est caractérisée par des forêts de sorbiers (*Fraxinus excelsior* ; environ 200 m d'altitude) et de hêtres (*Fagus sylvatica* ; entre 900 et 1 000 m d'altitude) ainsi que par des prairies, dominées par l'espèce *Carex foetida*. La partie supérieure du massif est caractérisée par des patchworks de landes (*Vaccinium myrtillus*, entre 1 200 et 1 350 m d'altitude) et de forêts de conifères (*Pinus sylvestris* vers 1350 m ; Fig 2.4).

Les sols de la partie basse du massif des Bauges (environ 200 m) sont des calcosols, influencés par le socle calcaire. Au sein de la hêtraie, la pédologie est dominée par les organosols sur substrat calcaire. Au sein de la forêt de conifères, les sols sont divisés en patchworks d'organosols – résultant de l'accumulation de litière organique sur des substrats calcaires – et de brunisols, résultant principalement de l'accumulation d'argiles dans des creux topographiques.

2.1.1.2. Massif du Lautaret-Galibier

Vingt-cinq échantillons de sol ont également été collectés le long du versant de la vallée de la Roche Noire, entre 1 540 m et 2 700 m d'altitude. La vallée est située dans le massif du Lautaret-Galibier-Galibier. C'est un site de moyenne à haute altitude, situé dans la partie sud-ouest des Alpes françaises, au nord du Parc national français des Ecrins, entre les cols du Lautaret-Galibier (2 058 m d'altitude) et du Galibier (2 645 m d'altitude). Le prélèvement a eu lieu dans la vallée de la Roche Noire, surplombée par la crête de Côte Plaine. Les précipitations annuelles moyennes le long du massif du Lautaret-Galibier-Galibier sont de 1 300 mm. Dans la partie basse du massif, à 1 500 m d'altitude, la TMAA est d'environ 6°C. Dans la partie supérieure du massif, à une altitude de 2 700 m, la TMAA est d'environ 0°C (Bounemoura et al., 1998 ; Choler, 2018). La végétation boisée, dominée par la forêt d'épicéas (*Larix decidua*), se trouve entre 1600 m et 1900 m d'altitude. Des landes (*Vaccinium myrtillus*) sont observées entre 2 200 m et 2 300 m d'altitude ainsi que quelques prairies subalpines (*Festuca paniculata*) autour de 2 000 m d'altitude. La partie supérieure du massif est dominée par des prairies alpines, avec trois espèces dominantes : *Poa alpina* (entre 2 470 m et 2 700 m d'altitude), *Carex foetida* (à

environ 2 500 m d'altitude) et *Kobresia simpliciuscula* (entre 2 600 m et 2 700 m d'altitude).

Les calcosols se trouvent dans la partie la plus basse du massif, entre 1 500 et 2 000 m d'altitude, tandis que les prairies subalpines (environ 2 000 m d'altitude) et alpines (entre 2 400 et 2 700 m d'altitude) sont dominées par les brunisols. Des aloccrisols – se développant sur des roches acides et intermédiaires entre les brunisols et les sols podzoliques – sont également observés entre 2 200 et 2 300 m d'altitude, ainsi que des rendisols – peu évolués et indifférenciés poussant sur des substrats calcaires – à environ 2 700 m d'altitude et des colluviosols – formés par des apports gravitationnels de sols préexistants – à environ 2 500 m d'altitude.

2.1.1.3. Echantillonnage

La végétation le long des deux massifs (Bauges et Lautaret-Galibier) est répartie selon différents biomes en fonction du régime thermique, mais aussi des paramètres de précipitation et de microtopographie le long du gradient altitudinal. Ainsi, des échantillons représentatifs des différents types de sol et/ou de végétation ont été collectés à chaque altitude. Au total, 49 échantillons de sol ont été collectés (Fig 2.4; Tableau 2.2). Pour l'ensemble des sols étudiés, entre 50 et 100g d'échantillon ont été préalablement congelés à -20°C , puis lyophilisés. Dans un deuxième temps, les échantillons ont été broyés et tamisés à 2 mm, afin d'obtenir des sols « homogènes ».

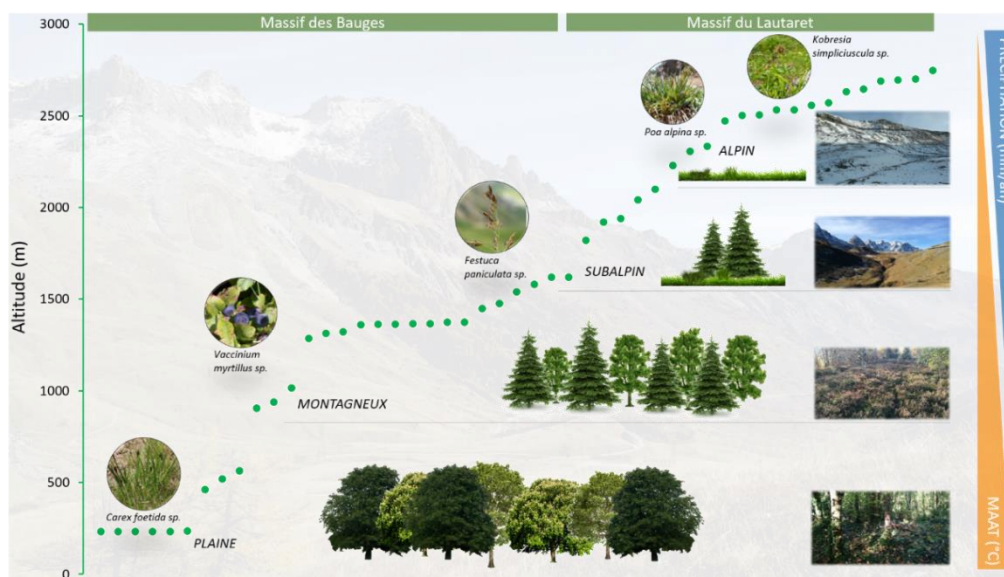


Figure 2.4 Transect altitudinal composite des massifs des Bauges et du Lautaret-Galibier-Galibier.

Echantillon	Nomenclature	Massifs	Coordonnées	Altitude (m)	Types de sol	Végétation
1	LAU2700		45° 3'35.24"N 6°23'30.96"E	2700	Rendisol	Prairie alpine (<i>Kobresia simpliciuscula</i>)
2	KS46		45° 3'35.24"N 6°23'30.96"E	2646	Rendisol	Prairie alpine (<i>Kobresia simpliciuscula</i>)
3	KS47		45° 3'36.04"N 6°21'54.02"E	2748	Rendisol	Prairie alpine (<i>Kobresia simpliciuscula</i>)
4	KS48		45° 3'32.98"N 6°22'4.63"E	2695	Rendisol	Prairie alpine (<i>Kobresia simpliciuscula</i>)
5	LAU 2470 – PA129		45° 3'15.48"N 6°24'3.11"E	2470	Brunisol	Prairie alpine (<i>Poa alpina</i>)
6	PA125		45° 3'29.19"N 6°22'22.86"E	2632	Brunisol	Prairie alpine (<i>Poa alpina</i>)
7	PA126		45° 3'33.77"N 6°22'9.07"E	2688	Brunisol	Prairie alpine (<i>Poa alpina</i>)
8	CF31		45° 3'0.43"N 6°23'6.15"E	2503	Colluviosol	Prairie alpine (<i>Poa alpina</i>)
9	CF32		45° 3'15.75"N 6°22'30.15"E	2571	Colluviosol	Prairie alpine (<i>Poa alpina</i>)
10	CF33		45° 3'13.32"N 6°22'41.51"E	2531	Colluviosol	Prairie alpine (<i>Poa alpina</i>)
11	EN51		45° 2'55.20"N 6°23'14.02"E	2504	Lithosol	Eboulis
12	EN52		45° 3'1.99"N 6°22'54.88"E	2533	Lithosol	Eboulis
13	EN53	Lautaret	45° 3'10.66"N 6°22'39.90"E	2558	Lithosol	Eboulis
14	V26		45° 2'38.94"N 6°24'8.20"E	2230	Allocrisol	Landes subalpine (<i>Vaccinium myrtillus</i>)
15	V27		45° 2'53.79"N 6°23'45.81"E	2305	Allocrisol	Landes subalpine (<i>Vaccinium myrtillus</i>)
16	V28		45° 2'59.32"N 6°23'32.53"E	2333	Allocrisol	Landes subalpine (<i>Vaccinium myrtillus</i>)
17	LAU 2100		45° 2'42.11"N 6°24'25.73"E	2100	Brunisol	Prairie subalpine (<i>Festuca paniculata</i>)
18	LAU1920		45° 2'26.05"N 6°25'8.51"E	1920	Brunisol	Prairie subalpine (<i>Festuca paniculata</i>)
19	FP88		45° 2'25.32"N 6°24'30.39"E	2041	Brunisol	Prairie subalpine (<i>Festuca paniculata</i>)
20	LD113		45° 2'22.35"N 6°26'6.73"E	1940	Calcosol	Forêt subalpine (<i>Larix decidua</i>)
21	LD1820		45° 1'33.34"N 6°27'8.63"E	1820	Calcosol	Forêt subalpine (<i>Larix decidua</i>)
22	LD1620		44°59'47.61"N 6°28'50.40"E	1620	Calcosol	Forêt subalpine (<i>Larix decidua</i>)
23	LECA S1620		45° 0'49.54"N 6°28'0.44"E	1620	Calcosol	Prairie montagnaise
24	LECA S1580		45° 0'4.78"N 6°28'27.34"E	1580	Calcosol	Prairie montagnaise
25	LECA S1540		44°59'39.52"N 6°28'43.30"E	1540	Calcosol	Prairie montagnaise
26	FEC21		45°40'2.47"N 5°58'17.07"E	1475	Brunisol	Prairie montagnaise
27	FEC22		45°40'14.96"N 5°58'25.39"E	1450	Brunisol	Prairie montagnaise
28	FEC23		45°40'29.61"N 5°59'53.47"E	1360	Brunisol	Prairie montagnaise
29	FEC31O		45°40'31.96"N 5°59'46.20"E	1375	Organosol	Forêt montagnaise
30	FEC32O		45°40'33.31"N 5°59'51.52"E	1367	Organosol	Forêt montagnaise
31	FEC33O		45°40'27.77"N 5°59'47.87"E	1362	Organosol	Forêt montagnaise
32	FEC31B		45°40'31.96"N 5°59'46.20"E	1375	Brunisol	Forêt montagnaise
33	FEC32B		45°40'33.31"N 5°59'51.52"E	1367	Brunisol	Forêt montagnaise
34	FEC33B		45°40'27.77"N 5°59'47.87"E	1362	Brunisol	Forêt montagnaise
35	FEC11		45°38'20.27"N 5°59'36.04"E	1286	Podzosol	Landes montagnaise (<i>Vaccinium myrtillus</i>)
36	FEC12		45°38'43.73"N 5°59'26.42"E	1314	Podzosol	Landes montagnaise (<i>Vaccinium myrtillus</i>)
37	FEC13		45°38'47.71"N 5°59'24.09"E	1321	Podzosol	Landes montagnaise (<i>Vaccinium myrtillus</i>)
38	MOU31	Bauges	45°40'22.12"N 5°57'43.67"E	905	Organosol	Forêt montagnaise (<i>Fagus sylvatica</i>)
39	MOU32		45°40'17.69"N 5°57'44.82"E	938	Brunisol	Forêt montagnaise (<i>Fagus sylvatica</i>)
40	MOU33		45°40'6.46"N 5°57'41.22"E	1017	Organosol	Forêt montagnaise (<i>Fagus sylvatica</i>)
41	MOU11		45°40'48.07"N 5°56'30.30"E	462	Brunisol	Prairie collinéaire
42	MOU12		45°40'47.24"N 5°57'1.91"E	565	Brunisol	Prairie collinéaire
43	MOU13		45°40'52.82"N 5°56'52.87"E	520	Brunisol	Prairie collinéaire
44	BOU11		45°38'53.89"N 5°52'3.84"E	232	Calcosol	Forêt collinéaire (<i>Fraxinus excelsior</i>)
45	BOU12		45°38'51.44"N 5°52'3.85"E	234	Calcosol	Forêt collinéaire (<i>Fraxinus excelsior</i>)
46	BOU13		45°38'9.90"N 5°52'9.58"E	237	Calcosol	Forêt collinéaire (<i>Fraxinus excelsior</i>)
47	BOU21		45°38'54.08"N 5°51'58.44"E	232	Calcosol	Prairie collinéaire (<i>Carex foetida</i>)
48	BOU22		45°38'59.87"N 5°52'0.01"E	232	Calcosol	Prairie collinéaire (<i>Carex foetida</i>)
49	BOU23		45°38'22.21"N 5°52'12.53"E	234	Calcosol	Prairie collinéaire (<i>Carex foetida</i>)

Tableau 2.2 Liste des échantillons de sol prélevés le long des massifs des Bauges et du Lautaret-Galibier, associés au type de végétation et aux différents couverts pédologiques.

2.1.1.4. Données de température

Des capteurs de température miniaturisés ont été placés à 5 cm sous la surface du sol dans les 24 sites du massif des Bauges, permettant de mesurer la température du sol en surface. Les températures de surface des sols sont considérées comme équivalentes aux températures de l'air (Choler, 2018) et peuvent être utilisées en première approximation comme estimations de la TMAA (Bartlett et al., 2006; Weijers et al., 2007). Les données correspondantes ont été compilées pour la période 2016-2018 et ont été complétées par les mesures de TMAA à la station météorologique du Feclaz (1 350 m d'altitude). Aucune donnée de température enregistrée n'étant disponible pour la

période 2016-2018 sur le massif du Lautaret-Galibier, les valeurs de TMAA du massif de l'Oisans, sélectionnées pour sa proximité géographique et climatique avec le Lautaret-Galibier, ont été utilisées à la place, avec 8 points répartis tous les 350 m entre 750 et 3550 m d'altitude. Les données ont été fournies le modèle SAFRAN-SURFEX/Crocus-MEPRA (S2 M) développée par Météo France pour les Alpes françaises (Durand et al., 2009; Vionnet et al., 2012).

Une régression linéaire a été établie entre les données de température disponibles (massifs des Bauges et de l'Oisans ainsi que la station du Feclaz) et l'altitude (Fig 2.5) pour obtenir un modèle permettant la reconstruction de la TMAA sur les différents sites d'échantillonnage (c'est-à-dire le transect composite le long des massifs des Bauges et du Lautaret-Galibier). Etant donnée sa fiabilité ($R^2 = 0,97$; EQM (Erreur Quadratique Moyenne) = $0,67$ °C), les valeurs de TMAA rapportées et discutées dans le reste du manuscrit seront celles obtenues à partir de ce modèle.

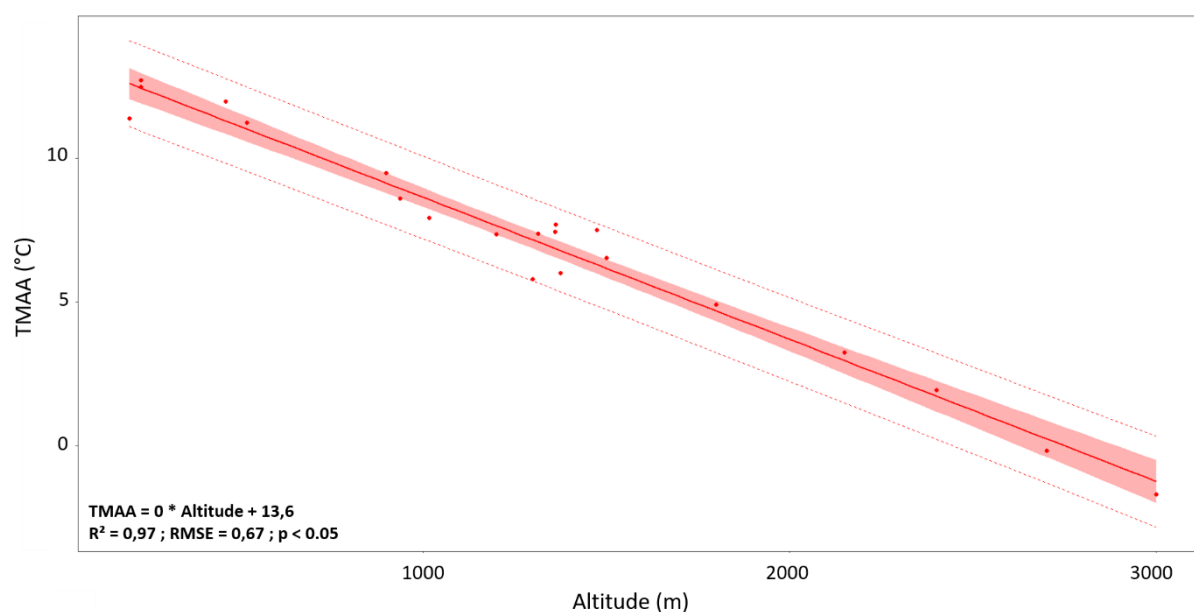


Figure 2.5 Régression linéaire du modèle de TMAA (°C) permettant la reconstruction des températures sur le transect altitudinal composite des Alpes françaises. Les lignes pointillées représentent l'intervalle de prédiction à 95% et la zone colorée représente l'intervalle de confiance à 95%.

2.1.2. Mont Pollino (Italie)

13 échantillons de sols de surface ont été collectés le long du Mont Pollino en Italie. Ces échantillons nous ont été fournis par Giulano Bonanomi (Università di Napoli Federico II).

Le Mont Pollino est situé dans la chaîne calcaire des Apennins et s'élève à 2 248 m d'altitude. Il est encadré au nord-ouest par la Sierra de Prete (2 181 m d'altitude) et au sud par le gouffre du Pollino. La zone alpine à subalpine (au-dessus de 2 100 m d'altitude) est caractérisée par la présence de prairies méditerranéennes (*Festuca bosniaca*, *Carex kitaibeliana*) et la présence de dolines (Todaro et al., 2007; Scalercio et al., 2014). La végétation montagnaise (plus de 1 200 m d'altitude) est dominée par des forêts d'hêtres communs (*Fagus sylvatica*) et par des pins (*Pinus leucodermis*) épars (Bonanomi et al., 2020). Le sol est peu différencié et est majoritairement calcaire. Entre 0 et 1 200 m d'altitude (Scalercio et al., 2014 et références incluses), le Mont Pollino est caractérisé par la présence de forêts de chênes verts (*Quercus ilex*). Le climat le long de cette montagne est méditerranéen humide, avec des températures estivales élevées et une distribution irrégulière des précipitations tout au long de l'année avec une sécheresse estivale prononcée (39,5% en hiver, 23,7% au printemps, 29,2% en automne, 7,6% en été ; précipitations annuelles moyennes : 1 570 mm ; voir Todaro et al., 2007). La TMAA est comprise entre 7 °C (2 200 m au-dessus du niveau de la mer) et 18 °C (0 m au-dessus du niveau de la mer ; Scalercio et al., 2014). Les températures le long du Mont Pollino ont été estimées en utilisant une régression linéaire entre deux températures annuelles moyennes (16 °C à 400 m d'altitude et 10 °C à 1 600 m d'altitude) à partir de données

Echantillons	Altitude (m)	TMAA (°C)	pH
0m	0	18	6,78
200m	200	17	6,19
400m	400	16	6,13
600m	600	15	6,14
800m	800	14	4,53
1000m	1000	13	5,41
1200m	1200	12	6,37
1400m	1400	11	5,62
1600m	1600	10	4,93
1800m	1800	9	4,91
2000m	2000	8	5,52
2100m	2100	7,5	5,91
2200m	2200	7	5,85

Tableau 2.3 Liste des échantillons de sol le long du transect du Mont Pollino et les caractéristiques environnementales et physico-chimiques associées.

météorologiques (station de Castrovillari) enregistrées par Scalercio et al. (2014). Le pH des sols analysés dans la présente étude varie entre 4,5 et 6,8 (tableau 2.3).

2.1.3. Mont Shegyla (Tibet)

17 sols de surface ont été prélevés le long du mont Shegyla par Wang et al. (2015). Le mont Shegyla est situé dans le sud-est du Tibet, en Chine (94°25'-94°45' E, 29°35'-29°57' N). Il s'agit d'un site de haute altitude (2700 à 4500 m). Le long de ce transect altitudinal, on observe différentes zonations climatiques. Ainsi, on retrouve une zone tempérée montagnaise entre 2700 et 3400 m, une zone tempérée froide subalpine entre 3400 et 4300 m et une zone alpine froide au-dessus de 4300 m. Ce gradient altitudinal est caractérisé par des communautés végétales très diverses. Des espèces végétales caractéristiques des étages montagneux à subalpin, comme le chêne brun (*Quercus semecarpifolia*) ou le sapin commun (*Abies alba*) sont abondantes au sein de l'étage montagneux tempéré. On retrouve au sein de l'étage froid subalpin une espèce endémique des régions de Chine occidentale, le sapin forestier (*Abies georgei* var. *smithii*). Enfin, au niveau de la zone alpine froide, on observe la présence d'espèces de conifères (*Sabina saltuaria*) ainsi que des espèces typiques des régions montagneuses comme le *Rhododendron* (Tableau 2.4). Les TMAA ont été estimées en utilisant une régression linéaire entre 7 températures mesurées à partir des données enregistrées par Wang et al. (2015). La TMAA moyenne le long du transect est de 4,6 °C, avec un minimum de 1,1 °C à 4 500 m d'altitude et un maximum de 8,9 °C à env. 3 100 m d'altitude (tableau 2.4).

Echantillons	Altitude (m)	TMAA (°C)	pH	%C org	%N	CEC cmol kg ⁻¹	H ₂ O %	Teneur en argile (g/kg)	Teneur en sable (g/kg)
T1	3106	8,9	5,5	4,57	0,31	21,9	11,1	22,2	60
T2	3117	8,9	6,4	2,63	0,18	13,1	5	18,3	70,9
T3	3132	8,8	6,0	1,83	0,14	14,2	6,1	14,2	72,3
T4	3344	7,6	6,0	2,52	0,19	14,5	20,8	6,9	77,9
T5	3355	7,5	5,9	4,87	0,35	23,9	35,4	15,8	57,9
T6	3356	7,5	5,5	3,98	0,29	18,4	26,4	8,5	75,1
C1	4030	3,7	5,2	4,63	0,25	24,2	35,6	23,8	56,3
C2	4046	3,6	4,7	5,23	0,44	29,7	39,4	28,2	47,9
C3	4050	3,6	4,6	6,37	0,51	31,6	39,6	19,6	60,6
C4	3912	4,3	5,0	6,21	0,32	22,8	19,1	21	69
C5	3918	4,3	4,7	5,67	0,37	28,6	27,3	25,7	58,5
C6	4298	2,1	5,0	5,57	0,31	22,2	35,5	27,9	52,5
C7	4295	2,2	4,9	6,84	0,54	37,1	46,1	22,1	54,6
C8	4304	2,1	5,3	6,07	0,42	27,7	38,4	29,1	51,2
F1	4479	1,1	5,3	5,34	0,33	12,6	30,9	18,6	73,8
F2	4479	1,1	5,1	5,28	0,25	15,3	26	19,8	75,1
F3	4474	1,1	5,2	6	0,45	26,9	42,5	24,9	56,7

Tableau 2.4 Liste des échantillons de sol le long du transect du Mont Shegyla et les caractéristiques environnementales et physico-chimiques associées (Wang et al., 2015)

Le pH du sol varie entre 4,6 et 6,4 (tableau 2.4). Les caractéristiques globales des échantillons analysés dans le cadre de cette étude sont présentées dans le Tableau 2.4.

2.1.4. Andes, transect du Kosñipata (Pérou)

14 sols ont été prélevés dans les Andes péruviennes par Andrew Nottingham (University of Edinburgh) au niveau du transect de Kosñipata (Tableau 4 ; Nottingham et al., 2015). Celui-ci est situé dans le sud-est du Pérou, dans la partie supérieure du bassin versant Madre de Dios/Madeira, à l'est de de la cordillère des Andes. Il s'agit d'un gradient altitudinal complet, (194 m à 3644 m d'altitude) particulièrement bien étudié dans de nombreuses études de biodiversité et bien documenté (Malhi et al., 2010; Nottingham et al., 2015). Le long de ce gradient, on observe une zonation végétale altitudinale depuis la forêt tropicale à basse altitude, jusqu'à une canopée forestière de haute montagne. La limite de la zone forestière se situe entre 3200 et 3700 m, zone à partir de laquelle on découvre une végétation prairiale de haute altitude. Contrairement au transect composite des Alpes françaises, les précipitations ne varient pas de manière notable avec l'altitude (moyenne=2448 mm/an, écart-type = 503 mm/an ; Rapp and Silman, 2012; Nottingham et al., 2015). Les TMAA sont comprises entre 26,4 °C à 194 m d'altitude et 6,5 °C à 3644 m d'altitude (tableau 2.5). Le pH est caractéristique de sols acides (3,4 - 4,7 ; tableau 2.5). De plus amples informations sur ces sites et ces sols sont disponibles dans Nottingham et al. (2015).

Echantillons	Altitude (m)	TMAA (°C)	pH	Précipitation (mm/an)	Types de végétation
TAM4	194	26,4	3,7	2730	Forêt tropicale collinéaire
TAM3	210	26,4	4	3199	
VC	1063	20,7	4,7	3087	Forêt tropicale prémontagneuse
SP2	1500	17,4	3,5	2631	Forêt subalpine
SP1	1750	15,8	3,6	2631	
TU8	1850	16	3,5	2472	Forêt alpine
TU7	2020	14,9	3,4	1827	
TU5	2520	12,1	3,7	n.d.	
TU4	2720	11,1	3,6	2318	
TU3	3020	9,5	3,4	1776	Prairie alpine
TU2	3200	8,9	3,5	n.d.	
WAY	3025	11,1	3,5	1706	
TU1	3400	7,7	3,4	2555	
TC	3644	6,5	3,4	n.d.	

Tableau 2.5 Liste des échantillons de sols prélevés au niveau du transect du Kosñipata et caractéristiques environnementales et physico-chimiques associées. (Nottingham et al., 2017). (n.d. : non disponible)

2.1.5. Araucanos (Chili)

10 sols d'un transect altitudinal chilien ont été collectés dans le cadre d'une collaboration avec le Dr Sergio Contreras (Département de chimie environnementale, Faculté des sciences et Centre de recherche sur la biodiversité et les environnements durables (CIBAS), Universidad Católica de la Santísima Concepción). Ces échantillons de sols de surface ont été prélevés entre 690 m et 1385 m d'altitude autour de 10 lacs andins. De plus, des mesures à haute fréquence (toutes les heures) de la température *in situ* sur une période d'un an sont disponibles pour les différents sites échantillonnés. Les TMAA sont comprises entre 5,75°C et 9,2°C. Le pH des sols sont compris entre 4,4 et 6,8 (Tableau 2.6).

Echantillons	Altitude (m)	TMAA (°C)	pH	Précipitation (mm/an)	%C org
Laguna Negra	690	9,2	5,4	2375	18,2
Quepe	870	8,9	5,6	2282	12,3
Malleco	891	7,9	4,9	2881	15,0
San Pedro	915	n.d.	6,8	n.d.	n.d.
Verde conguillio	980	8,5	5,6	2335	7,2
Captren	985	5,8	4,7	2644	17,8
Conguillio	1125	6,0	5,0	2480	15,2
Galletue	1151	6,0	5,9	2133	14,3
El Toro	1196	7,1	5,8	1970	8,1
Verde Tolhuaca	1385	n.d.	4,4	2700	28,7

Tableau 2.6 Liste des échantillons de sols prélevés au niveau du transect Araucanos et les caractéristiques environnementales et physico-chimiques associée. (n.d. : non disponible)

2.1.6. Jeu de données global (brGDGTs)

Une partie de mon travail de thèse a consisté à développer une nouvelle calibration globale terrestre des brGDGTs comme marqueurs de température (Chapitre 5). Pour cela, les échantillons décrits précédemment ont été ajoutés à un ensemble d'échantillons de sols de surface et de tourbières répartis à l'échelle mondiale, précédemment analysés pour les brGDGTs et publiés au sein de précédentes études (Weijers et al., 2007; Peterse et al., 2012; De Jonge et al., 2014; Yang et al., 2015; Ding et al., 2015; Xiao et al., 2015; Lei et al., 2016; H. Wang et al., 2016; Naafs et al., 2017b; Dearing Crampton-Flood et al., 2020) (Tableau 2.7).

Site	N (échantillons)	Références	Paramètres disponibles
Italie	13	Véquaud et al., 2020	TMAA, pH
Italie	11	Huguet et al., 2019, Véquaud et al., 2020	TMAA, pH
Tibet	17	Véquaud et al., 2020	TMAA, pH, PMA
Perou	14	Véquaud et al., 2020	TMAA, pH, PMA
France	49	Véquaud et al., 2021	TMAA, pH
Distribué globalement	229	Weijers et al. (2007a), Peterse et al. (2012), De Jonge et al. (2014b)	TMAA, pH, PMA, FRS
Inde	46	Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Russie	4	Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Nouvelle-Zélande	1	Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Chine	15	Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Chine	27	Xiao et al. (2015), Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Chine	26	Yang et al. (2015), Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Chine	27	Ding et al. (2015), Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Chine	44	Lei et al. (2016), Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Chine	148	Wang et al. (2016), Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
Distribué globalement	96	Naafs et al. (2017b), Dearing Crampton-Flood et al., (2020)	TMAA, pH, PMA, FRS
TOTAL	767		

Tableau 2.7 Emplacement, nombre et références des échantillons de sols et de tourbières dans le jeu de données global de brGDGT. Les paramètres disponibles pour les différents sites d'échantillonnage sont indiqués : TMAA (Température moyenne annuelle de l'air (°C), pH, PMA (Précipitation moyenne annuelle (mm/an), FRS (Nombre de jours de gel par an).

Les valeurs TMAA et pH sont disponibles pour la plupart des échantillons de ce jeu de données. Néanmoins, pour les autres échantillons, les paramètres MAAT et les valeurs de pH ont été extraits du jeu de données CRU TS v. 3.26 (Harris et al., 2014). En outre, le nombre de jours de gel pendant l'année (FRS) a également été acquis pour la plupart des échantillons (pour lesquels les coordonnées des sites d'échantillonnages étaient disponibles).

Le FRS a été utilisé à la place de l'indice des degrés-jours de croissance au-dessus du point de congélation (GDD), calculé en additionnant les températures quotidiennes au-dessus de 0°C sur une année dans un sol (Choler 2018), et interprété comme un indicateur de l'accumulation de chaleur dans un sol (McMaster and Wilhelm, 1997; Choler, 2018). Le GDD reflète mieux les températures de croissance rencontrées par les communautés bactériennes dans les sols que le MAAT et a notamment été utilisé par Naafs et al. (2017a) pour contraindre l'applicabilité de l'indice MBT'_{5Me}. Malheureusement, comme les températures quotidiennes n'étaient pas disponibles pour l'ensemble des données, l'indice GDD n'a pas pu être calculé et le FRS a été utilisé comme indicateur de saisonnalité.

2.2. Etudes de sols en environnements contrôlés (microcosmes)

2.2.1. Choix des sols

En milieu naturel, les microorganismes des sols sont influencés par une multitude de paramètres environnementaux. Des expériences en conditions contrôlées (microcosmes) ont été réalisées pour étudier les variations des abondances des lipides bactériens (acides gras 3-hydroxylés (3-AGH) et tétraéthers bactériens (brGDGTs)) et de la diversité microbienne en fonction de la température comme unique variable. Afin de disposer de sols contrastés d'un point de vue microbiologique, il a été décidé dans un premier temps de déterminer la réponse physiologique des communautés microbiennes de plusieurs sols alpins afin de comparer les diversités métaboliques de ces dernières. Les incubations ont été réalisées sur un nombre de sols limités, prélevés en quantité suffisante.

La diversité métabolique des communautés microbienne de 10 sols choisis le long du transect a été caractérisée à l'aide de tests de respiration sur des plaques Biolog® EcoPlate™. Le principe de ce test est d'offrir à des communautés microbiennes présentes dans des échantillons environnementaux (dans ce cas, des sols) différentes sources de carbones organiques (substrats) et d'observer la réponse physiologique des communautés microbiennes en fonction des différents substrats proposés. Les plaques contiennent 31 substrats de carbone différents, répliqués 3 fois, ainsi que trois puits de contrôles (sans substrats). Dans chaque puits est déposé un réactif, le tétrazolium. Ce réactif est réduit lors de la consommation hétérotrophe des substrats de carbone organique par les bactéries, formant un chromophore violet, du formazan. L'intensité de la réduction du tétrazolium permet d'observer l'efficacité de la consommation des sources de carbone présentes dans les puits par les communautés microbiennes (Fig 2.6).

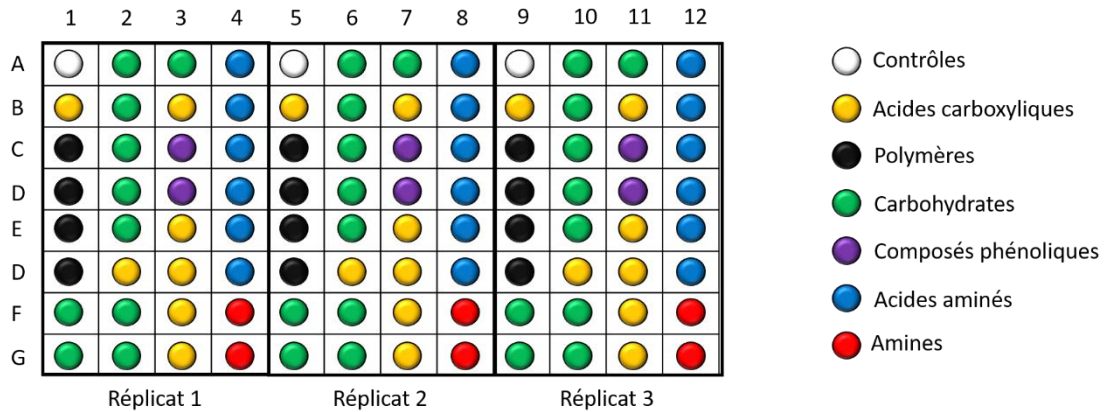


Figure 2.6 Schéma d'une plaque de respiration et répartitions des différents substrats carbonés.

Pour l'étude de la diversité métabolique, 2 g de sol frais ont été ajoutés à 200 ml de solution saline (0,9 % NaCl) et agités à 200 tr/min pendant 20 minutes à l'aide d'un agitateur rotatif. Le surnageant a ensuite été pipeté et dilué 10 fois avec la solution saline, puis inoculé dans les puits des microplaques Biolog® (150 µl), à raison d'un échantillon par plaque. Les plaques ont été incubées à 30°C. L'utilisation des substrats est révélée par un changement de couleur des différents puits de la plaque. L'intensité de ce changement de couleur a été déterminée par spectrophotométrie (mesure quotidienne de l'absorbance à 590 nm à l'aide d'un lecteur de microplaques) et permet de déterminer 2 paramètres reliés à l'activité des communautés microbiennes :

-La réponse métabolique moyenne (Average Metabolic Response ou AMR) décrit avec une seule valeur (possibilité de comparaison inter-échantillons) l'utilisation moyenne par les communautés microbiennes des différentes sources de carbone disponible. Cette valeur est déterminée en calculant la moyenne des différences entre l'intensité de l'absorbance mesurée sur les puits de sources de carbone (Abs échantillon) et les puits des « contrôles » (Abs contrôle) (Equation 2.1)

$$AMR = [\sum (Abs \text{ échantillon} - Abs \text{ contrôle})] / n \quad (Eq . 2.1)$$

-La diversité métabolique (Community Metabolic Diversity ou CMD) est représentée par le nombre de substrats différents utilisé par la communauté microbienne d'un échantillon. Il s'agit de la somme de réponse positive sur la plaque (puits présentant un changement de couleur).

Parmi les 49 sols des Alpes françaises, quatre d'entre eux - BOU 11, MOU 32, FP 88, et KS48 (Fig 2.4 ; Tableau 2.2) - ont été choisis pour les expériences en microcosme. Ces échantillons ont été sélectionnés étant donné leur représentativité des environnements contrastés rencontrés au sein des Alpes françaises, et de la diversité des communautés microbiennes associées à ces environnements. Ainsi, ces 4 échantillons sont répartis de manière régulière le long du transect altitudinal (232 m, 938 m, 2041m et 2695 m), et reflètent des environnements et des types de sols variés, avec des valeurs de pH s'échelonnant entre 5,3 et 7,3 (Tableau 2.2 ; Tableau 3.1 ; Fig 2.4).

KS48 est un échantillon de sol prélevé à haute altitude (2695 m), de type rendisol. La végétation du lieu de prélèvement de l'échantillon est caractérisée par une prairie alpine, dominée par l'espèce *Kobresia simpliciuscula*. Le pH de cet échantillon est de 6,5, tandis que sa teneur en carbone organique est de 12,7%. A cette altitude, la température moyenne annuelle est de 0,04°C.

FP88 est un échantillon de sol prélevé en milieu subalpin (2041 m), caractérisé par une végétation prairiale dominée par l'espèce *Festuca paniculata*. La température moyenne annuelle à cette altitude est de 3,6°C. FP88 est un échantillon de brunisol, caractérisé par un pH légèrement acide (5,2) et une teneur en carbone organique de 6,6%.

MOU32 a été prélevé au sein de l'étage montagneux, à 938 m d'altitude (TMAA = 8.6°C), dans une forêt dominée par l'espèce *Fagus sylvatica*. Il s'agit d'un échantillon de brunisol, caractérisé par un pH neutre (7,2) et une teneur en carbone organique de 11,6%.

Enfin, l'échantillon BOU11 a été prélevé en plaine (232 m) au sein d'une forêt dominée par l'espèce *Fraxinus excelsior*. La température moyenne annuelle à cette altitude est de 12,5°C. BOU11 est un calcosol, caractérisé par un pH de 7,3 et une teneur en carbone organique de 5,2°C.

2.2.2. Protocole d'incubation

Chacun des 4 sols a été incubé pendant un an à 4 températures (5°C, 15°C, 25°C, 35°C). Pour chaque point de la cinétique (1, 3, 6, 9 et 12 mois d'incubation), 30 g de sol ont été incubés en duplicat (flacons A et B) dans une bouteille Duran de 250 mL fermée et ouverte une fois par semaine pour oxygéner l'échantillon. Ceci évite une perte de l'humidité au sein des flacons. Au cours de la thèse, les échantillons en microcosmes ont

été analysés pour les pas de temps 3 mois et 12 mois (duplicat A et B) dans le cas des 3-AGH et pour le pas de temps 12 mois (duplicat A et B) pour les brGDGTs (Fig 2.7).

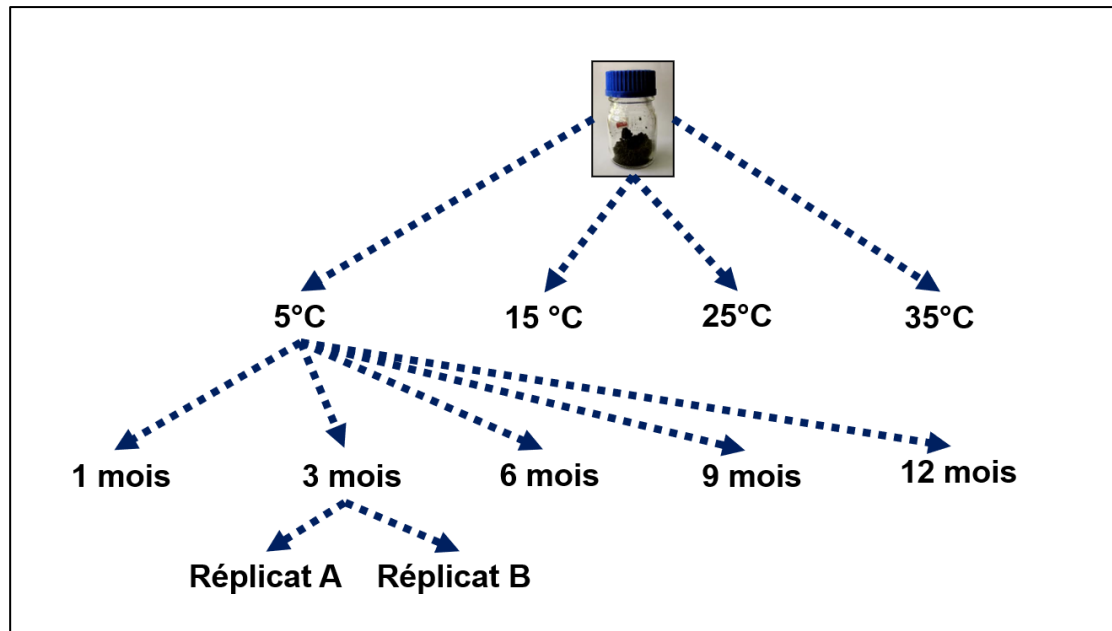


Figure 2.7 Schéma illustrant le protocole d'incubation des 4 échantillons de sols en microcosme

A chaque pas de temps, la viabilité de l'échantillon (présence de microorganismes vivants) a été testée par numération des communautés microbiennes. Pour ce faire, les bactéries ont dans un premier temps été extraites des 4 sols par remise en suspension de 2 g d'échantillon dans une solution saline (NaCl 0,9g/L), puis diluées et étalées sur un milieu gélosé R2A, additionné d'un antifongique (cycloheximide) et de trois antibiotiques (vancomycine, teicoplanine, daptomycine) inhibant de façon spécifique la croissance des bactéries Gram-positif.

La numérotation des communautés bactériennes Gram-négatif cultivables est réalisée par comptage des Unités Formant Colonies (UFC, ou Colony-Forming Unit CFU). Ce dénombrement a pour but d'estimer le nombre de cellules présentes en fonction de leur capacité à donner naissance à des colonies dans des conditions spécifiques de milieu nutritif, de température et de temps. En théorie, une cellule viable (vivante) et cultivable peut donner naissance à une colonie par réplication. Le comptage des UFC suppose que chaque colonie soit séparée et issue d'une seule cellule microbienne viable (Goldman and Green, 2015). Néanmoins, cette méthode de dénombrement peut dans

certain cas sous-estimer le nombre des cellules vivantes présente dans un échantillon. En effet, seul 1% des bactéries sont cultivables (Amann et al., 1995), certaines colonies pouvant être formées par plusieurs cellules physiquement proches, tandis que d'autres cellules vivantes peuvent être en « dormance » et ne pas former de colonies viables (Goldman and Green, 2015; communication personnelle Sylvie Collin, 2020).

Dans l'objectif d'étudier la diversité bactérienne, en particulier celle des bactéries Gram-négatifs produisant les 3-AGH, une aliquote de chaque échantillon de sol a été placée dans 8 vials de 1,5 ml (réplicats) et conservée à -20°C pour les analyses de l'ADN génomique. Une fois ces prélèvements effectués, les flacons contenant le reste des échantillons de sols ont été congelés à -20°C puis lyophilisés dans l'attente des analyses lipidiques (acides gras 3-hydroxylés et tétraéthers bactériens). Les résultats de diversité microbienne sont en cours d'analyse au sein de l'UMR METIS et ne seront pas présentés dans la suite de ce manuscrit.

2.3. Analyses élémentaires, pH et humidité

Différentes analyses ont été effectuées sur les échantillons de sols : humidité, analyses élémentaires et isotopiques, mesures de pH, granulométrie et mesure de la capacité d'échange cationique (CEC)

2.3.1. Humidité relative

Les échantillons frais ont été préalablement pesés, congelés puis lyophilisés. Une mesure de la masse des échantillons avant et après lyophilisation a permis de calculer l'humidité relative au poids initial de sol (en %) de chaque échantillon.

2.3.2. Analyses élémentaires et isotopiques

Les échantillons préalablement lyophilisés ont été décarbonatés avant les analyses élémentaires (teneur en carbone organique total et azote) et isotopiques ($\delta^{13}\text{C}$ et $\delta^{15}\text{N}$). Pour ce faire, on ajoute 20 mL d'acide chlorhydrique 3M à environ 1,5 g de sol dans des tubes à centrifuger. Cette suspension est agitée pendant 3 heures. L'échantillon est ensuite centrifugé (3000 tours/minutes pendant 10 minutes à 15°C). Le culot ainsi obtenu est rincé avec de l'eau distillée, et recentrifugé dans les conditions décrites précédemment.

Cette étape est généralement répétée deux fois, jusqu'à neutralité de l'eau de rinçage. L'échantillon décarbonaté est ensuite conservé à -20 °C puis lyophilisé.

Les analyses élémentaires et isotopiques ont été réalisées sur la plateforme ALYSES, (Université de la Sorbonne / IRD) à Bondy, France. Environ 10 µg de sols décarbonatés ont été typiquement pesés dans des capsules en étain avant d'être analysés à l'aide d'un analyseur Elementar Vario EL III (Elementar Analysensysteme GmbH, Allemagne).

2.3.3. Mesures de pH

Les mesures de pH sont réalisées sur l'échantillon préalablement lyophilisé. Pour ce faire, une solution de sol artificielle est préparée en ajoutant de l'eau ultrapure (Elga) à l'échantillon sec, dans des proportions connues (sol/eau dans des proportions 1 :2,5). Il est typiquement prélevé 4 g d'échantillon sec, auquel on ajoute 10 ml d'eau distillée. La solution de sol ainsi obtenue est agitée pendant 30 minutes, avant décantation pendant 1 heure (Carter et al., 2007). La mesure du pH est ensuite réalisée avec une sonde pH préalablement calibrée à l'aide de solutions tampons (pH 4 et pH 7).

2.3.4. Granulométrie et capacité d'échange cationique

La répartition de la taille des particules en cinq fractions (argiles (< 2µm), limons fins (entre 2 et 20 µm), limons grossiers (entre 20 et 50 µm), sables fins (entre 50 et 200 µm) et sables grossiers (entre 200 et 2000 µm)) de sol tamisés et broyés a été déterminée au laboratoire d'analyse des sols INRA d'Arras pour l'ensemble des 49 sols des Alpes françaises. La capacité d'échange cationique (CEC) a été mesurée dans le laboratoire d'analyse des sols d'Arras à l'aide de la méthode METSON (Metson, 1957).

2.4. Analyse des lipides bactériens : brGDGT et 3-AGH

Différentes méthodes d'extraction et séparation ont été utilisées en vue de l'analyse des deux familles de lipides bactériens étudiés au cours de cette thèse : les tétraéthers bactériens (brGDGTs) et les acides gras 3-hydroxylés (3-AGH).

2.4.1. Extraction, séparation et analyse des brGDGTs

2.4.1.1. *Protocole d'extraction et de séparation des tetraéthers bactériens*

Les tetraéthers bactériens (brGDGTs) ont été extraits des sols à l'aide d'un extracteur automatisé Accelerated Solvant Extractor 100 (ASE 100 ; Thermo Fisher-Dionex), comme préalablement décrit par Coffinet et al. (2014). 10 g de sols lyophilisés sont typiquement placés dans une cellule puis extraits grâce à un mélange dichlorométhane (DCM)/méthanol (9 :1) pendant 3×5 minutes à 100°C et 10×10^6 Pa. L'extrait lipidique total ainsi obtenu est séparé en fractions de polarité croissante sur une colonne d'alumine préalablement activée à 150 °C pendant 1 nuit. Les fractions sont éluées respectivement avec un mélange de 30 mL d'heptane : DCM 9 :1 (v :v) puis 30 ml de DCM : MeOH 1 :1 (v :v). Les GDGTs sont contenus dans la seconde fraction.

Les échantillons sont ensuite préparés pour l'analyse. 300 μ L de la fraction 2 (initialement dans un mélange DCM/MeOH 5/1) sont placés dans un tube Eppendorf d'1,5 mL puis séchés sous flux d'azote. Ensuite 300 μ L d'heptane sont ajoutés dans le tube eppendorf. La solution ainsi obtenue est homogénéisée (vortex pendant 10 secondes) et centrifugée pendant 60 secondes à 7000 tours/minute. La solution surnageante, contenant les brGDGTs, est prélevée puis conservée dans un flacon de 2 ml. 45 μ L de cette solution ont été analysés après ajout de 5 μ L d'un standard (tetraéther synthétique en C₄₆) à 0,01 mg/mL (Fig 2.8 ; Huguet et al., 2006).

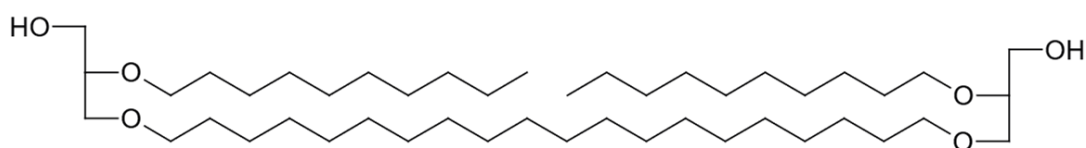


Figure 2.8 Structure interne du tetraéther synthétique en C₄₆ utilisé pour la semi-quantification des tetraéthers.

2.4.1.2. *Protocole d'analyse des tetraéthers bactériens*

L'analyse est réalisée par chromatographie liquide couplée à un spectromètre de masse (Shimadzu LCMS 2020) avec une source APCI (Ionisation Chimique à Pression Atmosphérique). La séparation des molécules se fait sur deux colonnes de silice en

tandem Hypersil Gold (150 mm × 2,1 mm, 1,9 µm ; Thermo Finnigan, USA) thermostatées à 40 °C, comme décrit par Huguet et al. (2019) et utilise deux solvants A et B (A=Hexane et B=isopropanol) à 0,2 ml/min. L'élution débute à 82% de A et 18% de B pendant 25 minutes, puis augmente suivant un gradient linéaire avec 65% de A et 35% de B en 25 minutes. Un second gradient linéaire conduit à un mélange composé de 20% de A et 80% de B en 30 minutes. Ce gradient est maintenu pendant 10 minutes avant de revenir à l'état initial (82% de A et 18% de B) en 10 minutes. La colonne est équilibrée dans ces conditions pendant 30 minutes. Le volume d'injection est de 10 µL. Cette méthodologie permet de séparer les isomères 5- et 6-méthyle. La détection des tétraéthers est réalisée en mode SIM (Single Ion Monitoring), qui permet d'analyser spécifiquement les composés d'intérêt. La semi-quantification des brGDGTs a été réalisée en comparant le signal intégré du composé considéré avec celui du standard interne en C₄₆ (Huguet et al., 2006) en supposant que leurs facteurs de réponse sont identiques.

2.4.2. Extraction, séparation et analyse des acides gras 3-hydroxylés

2.4.2.1. *Protocole d'extraction et de séparation des acides gras 3-hydroxylés*

Le protocole de préparation et d'analyse des acides gras 3-hydroxylés a été modifié d'après C. Wang et al. (2016). Entre 5 et 10 g de sol préalablement lyophilisés sont soumis à une hydrolyse acide (reflux pendant 3 heures dans de l'acide chlorhydrique (HCl) 3M avec 30 ml pour 10g d'échantillon ; bain d'huile à 130 °C) afin de libérer les 3-AGH liés aux unités glucosamine du lipopolysaccharide de la membrane des bactéries Gram-négatif. L'échantillon est ensuite rincé à l'eau ultrapure (20 ml), puis centrifugé pendant 10 minutes à 15°C et 3000 tours/minute. Le surnageant est récupéré dans un erlenmeyer. Le culot est ensuite extrait à l'aide de différents mélanges de solvants : DCM /Méthanol 1 :1 (v:v ; 2 fois), puis DCM (2 fois). Chaque extraction est réalisée à température ambiante par ultrasonication, et le surnageant obtenu après centrifugation (10 min à 15°C et 3000 tours/minute) est cumulé avec les précédents.

La phase organique obtenue après les différentes extractions est séparée de la phase non organique (acide chlorhydrique dans l'eau) à l'aide d'une ampoule à décanter, l'échantillon étant rincé 3 fois avec 20 ml de DCM. La fraction organique ainsi obtenue

est séchée sur sulfate de sodium, filtrée sur coton puis concentrée à sec grâce à un évaporateur rotatif.

L'extrait lipidique après hydrolyse est ensuite méthylé dans une solution HCl-Méthanol 1M pendant 1 heure à 80°C. Il est ensuite concentré à sec, puis rincé 3 fois avec du méthanol pur et évaporé pour éliminer l'excès d'HCl. Les acides gras méthylés hydroxylés et non-hydroxylés sont ensuite séparés sur une colonne de silice activée à 150 °C pendant 1 nuit. Trois fractions sont obtenues respectivement après élution avec (1) 30 mL d'heptane/acétate d'éthyle 98 :2, (2) 30 mL d'acétate d'éthyle et (3) 30 mL de méthanol pur. Les 3-AGH sont contenus dans la deuxième fraction.

2.4.2.2. Protocole initial d'analyse des acides gras 3-hydroxylés

Les conditions initiales de préparation des échantillons pour l'analyse des 3-AGH sont identiques à celle rapportée par Huguet et al. (2019) et proposé par C. Wang et al., 2016. Les 3-AGH sont des molécules de petites tailles, analysables en chromatographie gazeuse et spectrométrie de masse (GC-MS). Ce sont des molécules polaires qu'il est nécessaire de dériver avant analyse. Ainsi, un agent silylant N,O, bis(triméthylsilyl)trifluoroacétamide (BSTFA) - Trimethylchlorosilane (TMCS) 99:1 (Grace Davison Discovery Science, USA) est ajouté à l'échantillon comme suit. Une aliquote de la fraction 2, obtenu après séparation sur colonne de silice, est inséré dans un flacon conique puis séché pour enlever toute trace de solvant protique (dans ce cas, le méthanol). Ensuite l'échantillon est redissous dans du DCM. A cette solution est ajouté un standard synthétique – acide gras 3-hydroxylé C₁₄ deutéré (acide gras 3-hydroxytetradecanoïque, 2,2,3,4,4-d₅ ; Sigma Aldrich) (Figure 2.9) – en quantité connue, pour déterminer l'abondance des 3-AGH naturels, et un agent dérivant, du BSTFA (5%), dans le but de silyler l'échantillon et le standard. La silylation est effectuée en étuve pendant 30 min à 70°C.

L'analyse est réalisée à l'aide d'un chromatographe en phase gazeuse Agilent Network 6980 couplé à un spectromètre de masse. La colonne utilisée pour la séparation des composés est une colonne Restek RXI-5 Sil MS silica (30m × 0,25 mm). La phase porteuse est constituée d'hélium avec un débit de 2 ml/min. Le four du chromatographe est programmé suivant le protocole suivant : 70°C à 200 °C, 10°C/mn ; puis 200°C à

310°C, 2°C/mn. Le four est ensuite stabilisé à 310 °C pendant 47 mn. Les échantillons sont injectés en mode splitless à 280°C.

Les 3-AGH ont été semi-quantifiés en intégrant le pic approprié et en comparant l'aire de l'étalon interne (3-AGH C₁₄ deutéré et méthylé ; Fig 2.9). Ce standard interne a pour objectif principal de vérifier la répétabilité des analyses au cours de l'injection des échantillons. Les différents 3-AGH ont été identifiés sur la base de leur temps de rétention, après extraction du fragment m/z 175 caractéristique (m/z 178 pour l'étalon interne deutéré ; cf. Huguet et al., 2019).

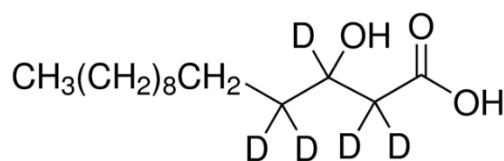


Figure 2.9 Structure chimique du standard C₁₄ deutéré, utilisé pour la quantification des acides gras 3-hydroxylés.

2.4.2.3. Développement analytique

Dans un premier temps, l'analyse des acides gras 3-hydroxylés dans les échantillons de sols a été réalisée suivant le protocole proposé par C. Wang et al. (2016) et décrit ci-dessus. Néanmoins, l'étude approfondie de ce protocole a montré plusieurs artefacts qu'il a fallu déterminer puis élucider.

• Effet du séchage sous azote

Ainsi, l'analyse préliminaire en GC-MS de 17 échantillons de sols sélectionnés parmi les 49 prélevés le long du transect altitudinal des Alpes françaises a permis de révéler un effet du séchage sous flux d'azote des échantillons sur la distribution et l'abondance des 3-AGH (Figure 2.10). Ces étapes de séchage étaient initialement au nombre de deux. Un premier séchage de l'échantillon en solution dans du DCM/Méthanol 5 :1 avait pour but d'éliminer toute trace du solvant protique, dans notre cas le méthanol, avant de resolubiliser l'extrait lipidique dans du DCM. Cette étape est indispensable afin de pouvoir ensuite silyler l'échantillon. Une deuxième étape de séchage, post-silylation,

avait pour but de limiter l'usure de la colonne en présence du réactif de silylation (BSTFA-TMCS).

Il a été montré que ces étapes de séchages induisaient une perte des composés les plus volatils. La première étape de séchage du protocole, indispensable afin d'éliminer le solvant protique a toutefois été conservée mais en diminuant le flux d'azote (séchage doux).

La deuxième étape de séchage, post-silylation, a été supprimé afin de limiter la perte des composés les plus volatils. Dans l'objectif de limiter l'usure de la colonne en présence du réactif de silylation, un blanc de DCM est injecté entre chaque échantillon afin d'éliminer les traces potentielles de BSTFA.

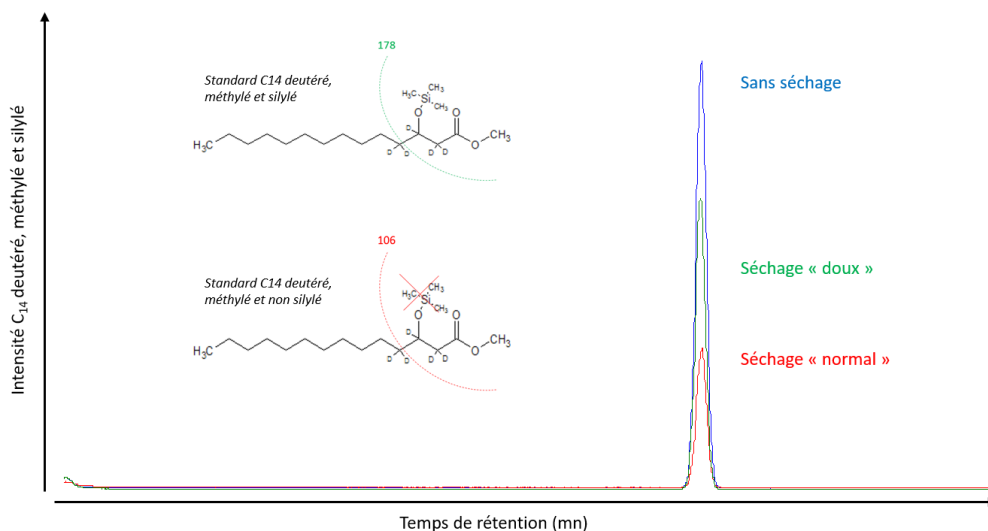


Figure 2.10 Illustration de l'effet du séchage sous flux d'azote des échantillons sur un standard en C₁₄ deutéré, méthylé et silylé, avant analyse en GC-MS

• Répétabilité et reproductibilité des analyses

Par la suite, avant de pouvoir appliquer le protocole en routine, la répétabilité et la reproductibilité des analyses en GC-MS a été testée sur un échantillon de sol des Alpes (BOU11, 232 m, Tableau 2.2). Cet échantillon a été analysé en triplicat, dans 3 vials différents (test de la reproductibilité). Chaque flacon a été injecté 3 fois, en série (test de la répétabilité). Ces analyses n'ont pas montré de différences substantielles au niveau de l'abondance relative des 3-AGH entre les différentes injections. Il a été conclu que les conditions de répétabilité et reproductibilité étaient respectées (Fig 2.11).

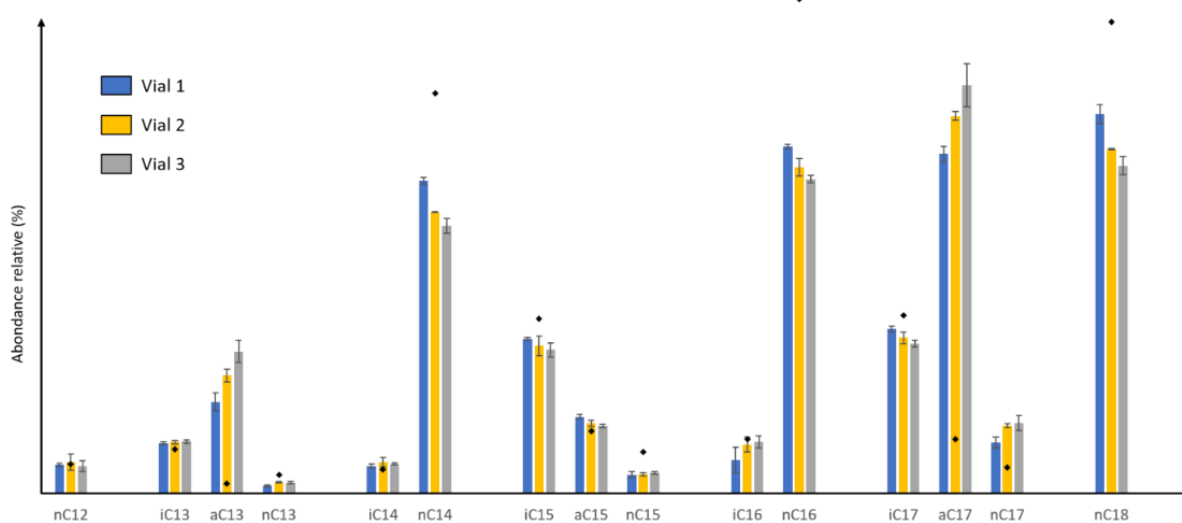


Figure 2.11 Test de répétabilité et de reproductibilité sur un triplicat du sol des Alpes françaises BOU11 (232 m d'altitude) analysé en GC-MS. Des valeurs aberrantes ont été observées sur un des triplicats d'injection, dû à une erreur analytique. Ces erreurs sont représentées sous la forme de points sur le graphique.

• Désilylation des échantillons au cours de l'analyse

Néanmoins, ces tests ont montré une désilylation et/ou une silylation partielle des 3-AGH, par la présence d'un pic de masse m/z 106 dans les échantillons, correspondant au standard C_{14} deutéré, méthylé et non silylé (Fig 2.10). Le rapport de l'aire du standard 3-AGH en C_{14} deutéré, méthylé et silylé (masse m/z 178) avec le pic de standard de 3-AGH en C_{14} deutéré, méthylé et non silylé (masse m/z 106 ; Fig 2.12) présent dans les échantillons a permis de montrer une évolution à la hausse au cours du temps, suggérant une désilylation des échantillons lors des analyses.

Ce problème a été résolu en augmentant la proportion (de 5 à 20%) du réactif de silylation (BSTFA-TMCS) dans les flacons d'injection et d'effectuer la réaction de silylation par le passeur sur chaque échantillon juste avant l'injection.

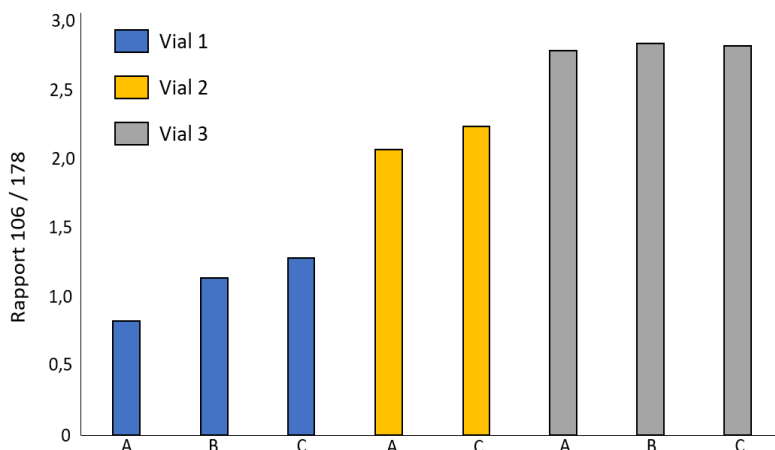


Figure 2.12 Rapport des fragments m/z 106 (standard C_{14} deutéré non silylé) et 178 (standard C_{14} deutéré silylé) dans l'échantillon BOU11 analysé en triplicat (3 vials), chaque vial étant injecté 3 fois.

- **Présence de coélutions**

En outre, l'analyse d'échantillons des Alpes selon le protocole initial (Partie 2.4.2.2) a révélé la présence de coélutions importantes (notamment sur le 3-AGH C_{17}) d'acide gras non hydroxylés avec les acides gras 3-hydroxylés (Fig 2.13). Afin de résoudre ces artefacts, il a été décidé dans un premier temps d'améliorer les séparations préparatives sur colonne de silice en augmentant la quantité de solvant éluant sur les différentes fractions (60 ml au lieu de 30 ml) et en réalisant une seconde séparation sur la fraction 2, contenant les 3-AGH. Néanmoins, ces nouvelles séparations n'ont pas permis de résoudre la présence de coélutions.

Nous avons dans un second temps remplacé la colonne de chromatographie du GC-MS, initialement d'une longueur de 30 m, par une colonne de 60 m, et avons augmenté le débit du gaz vecteur (hélium), passant de 1 ml/min à 2ml/min. Ces nouveaux paramètres analytiques ont permis de limiter la présence de coélutions (Fig 2.13).

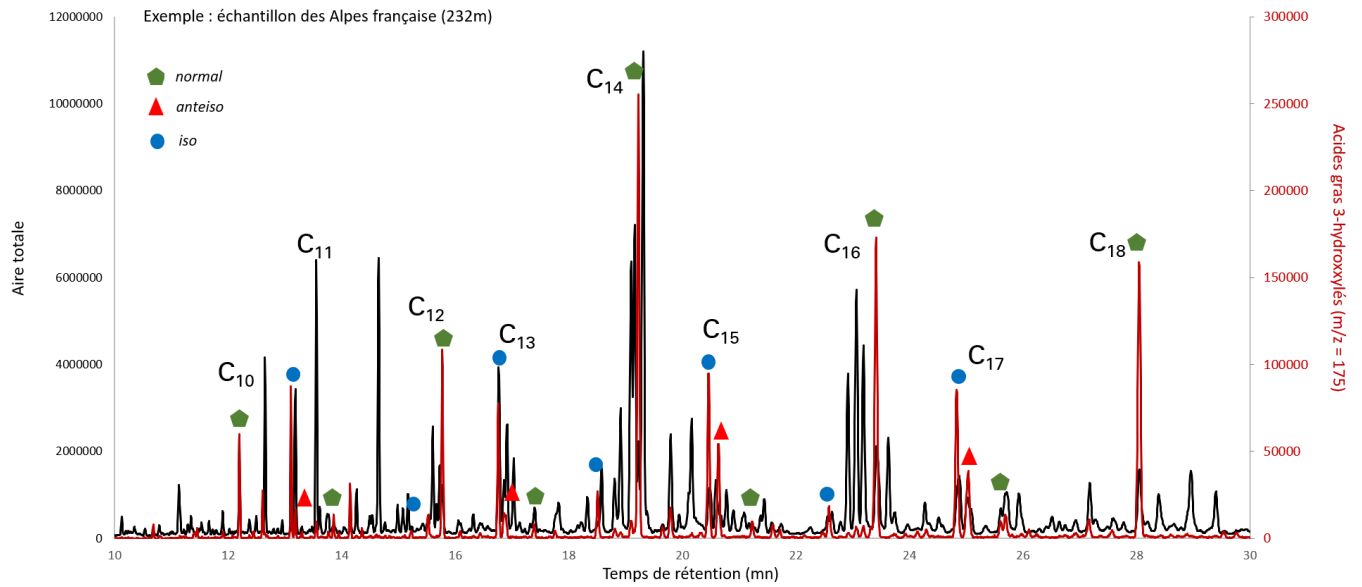


Figure 2.13 Illustration de la présence de coélutions d'acides gras non hydroxylés sur les acides gras 3-hydroxylés

2.4.2.4. *Nouvelles conditions d'analyse des acides gras 3-hydroxylés*

Le développement analytique réalisé au cours de cette thèse a permis d'améliorer les conditions d'analyse des 3-AGH extraits des sols. Ainsi, il a été décidé de supprimer la deuxième étape de séchage et de réaliser la silylation sur le passeur juste avant l'analyse en GC-MS, avec du BSTFA-TMCS en proportion 20%. Afin d'améliorer la séparation des composés, une colonne de 60 m a été utilisée à la place d'une colonne de 30 m. Le tableau 2.6 présente le protocole d'analyse des 3-AGH initialement proposé par C. Wang et al. (2016) et celui utilisé au cours de cette thèse après optimisation (Tableau 2.8).

	Protocole original (Wang et al., 2016)	Protocole modifié (cette étude)
Extraction	Hydrolyse acide HCl 3M (3h ; 130°C)	Hydrolyse acide HCl 3M (3h ; 130°C)
	Extraction aux solvants organiques	Extraction aux solvants organiques
	Séparation phases organique et aqueuse	Séparation phases organique et aqueuse
Séparation	Méthylation (1h ; 80°C)	Méthylation (1h ; 80°C)
	Séparation en 3 fractions	Séparation en 3 fractions
	Fractions 2 collectée dans un vial (DCM/MeOH) 5:1	Fractions 2 collectée dans un vial (DCM/MeOH) 5:1
	Séchage du solvant protique	Séchage doux du solvant protique
	Ajout de DCM, standard C ₁₄ deutéré et méthylé, et de BSTFA-TMCS (5%)	Ajout de DCM, standard C ₁₄ deutéré et méthylé, et de BSTFA-TMCS (20%)
Silylation	Silylation dans une étuve (30 min ; 70°C)	Silylation sur le passeur de la GC-MS (30min ; 70°C ; agitation)
	Séchage du BSTFA-TMCS	
	Ajout de DCM et vortex (10s)	
Analyse	Injection en GC-MS (Splitless, 280°C, débit 1 ml/min)	Injection en GC-MS (Splitless, 280°C, débit 2 ml/min)
	Colonne de chromatographie Restek RXI-5 Sil MS silica (30m x 0,25 mm).	Colonne de chromatographie Restek RXI-5 Sil MS silica (60m x 0,25 mm).
	Température du four : 70 °C à 200°C par pas de 10°C/min puis 200°C à 310°C par pas de 2°C/min. Stabilisation à 310°C pendant 47 min	Température du four : 70 °C à 200°C par pas de 10°C/min puis 200°C à 310°C par pas de 2°C/min. Stabilisation à 310°C pendant 20 min

Tableau 2.8 Comparaison des protocole d'extraction, séparation et d'analyse des 3-AGH. En bleu sont indiquées les modifications apportées au protocole original (C. Wang et al., 2016).

2.5. Analyses statistiques

Les analyses statistiques utilisées lors de cette thèse peuvent être classées en trois grandes catégories en fonction du nombre et du type de variables étudiées, mais aussi et surtout des objectifs et résultats attendus.

2.5.1. Analyses unies et bivariées

Afin d'observer la différence entre des groupes d'échantillons selon différentes modalités (types de sols, communautés de végétations), des moyennes et écart types ont été calculés. Afin de déterminer si ces valeurs associées aux différents groupes étaient significativement différentes, le test de Kruskal-Wallis, associé au test post hoc de Dunn, a été utilisé.

Afin d'étudier les relations entre les différentes variables environnementales, des tests de corrélations ont été effectués. Il a été choisi lors de cette thèse d'utiliser le test de corrélation non paramétrique de Spearman afin de disposer d'une analyse robuste face au faible nombre d'échantillons à notre disposition. Ce test permet de calculer un coefficient de corrélation pour des variables prises deux à deux et de tester la significativité de cette corrélation. Ainsi, une corrélation apparaît significative lorsque la p -value est $<0,05$.

Des régressions linéaires ont été utilisées pour étudier les relations entre la distribution des lipides bactériens et différents paramètres environnementaux. Ces différents tests statistiques ont été réalisés à l'aide du logiciel R (Version 3.6.0)

2.5.2. Analyses multifactorielles

Des analyses en composantes principales (ACP) différents échantillons pour les 3-AGH d'une part et les brGDGTs d'autre part. Ces analyses permettent de réduire un ensemble de n variables dans un espace bidimensionnel. Ces dimensions, également appelées composantes principales, synthétisent au mieux toutes les variables environnementales de l'ensemble des échantillons, ce qui permet de les simplifier. Les relations entre les variables sont décrites par rapport à ces composantes principales. Ainsi, l'ACP permet de décrire sur un espace géométrique simple les relations générales entre toutes nos variables. Il est possible dans un deuxième temps d'estimer quelles sont les variables qui expliquent le mieux la distribution ou les différences entre tous nos

échantillons. Cette analyse, largement utilisée dans les sciences de l'environnement, est intéressante lorsque le nombre de variables est important. Néanmoins, l'ACP, si elle permet de décrire les relations générales entre les variables, ne permet pas de mettre en évidence l'influence des variables sur d'autres variables, par exemple dans le cas de cette étude, l'influence des paramètres environnementaux sur l'abondance des lipides membranaires. A cette fin, des analyses de redondance (RDA) ont été effectuées. La RDA est une analyse "sous contraintes", contrairement à l'ACP qui est "sans contraintes". Avant l'analyse de redondance, deux groupes de variables sont définis : les variables explicatives (c'est-à-dire les paramètres physico-chimiques et environnementaux des sols) et les variables à expliquer (c'est-à-dire les abondances relatives des lipides bactériens, brGDGTs et 3-AGH). La RDA permet ainsi de visualiser directement la réponse des lipides membranaires en fonction des variations des paramètres physico-chimiques et environnementaux (Fig 2.14). Les ACP (centrées et réduites) sont effectués sur le logiciel R (Version 3.6.0), à l'aide du package FactoMineR. Les analyses de redondance (centrées et réduites) sont effectuées à l'aide du logiciel CANOCO version 5 (Braak and Smilauer, 2002).

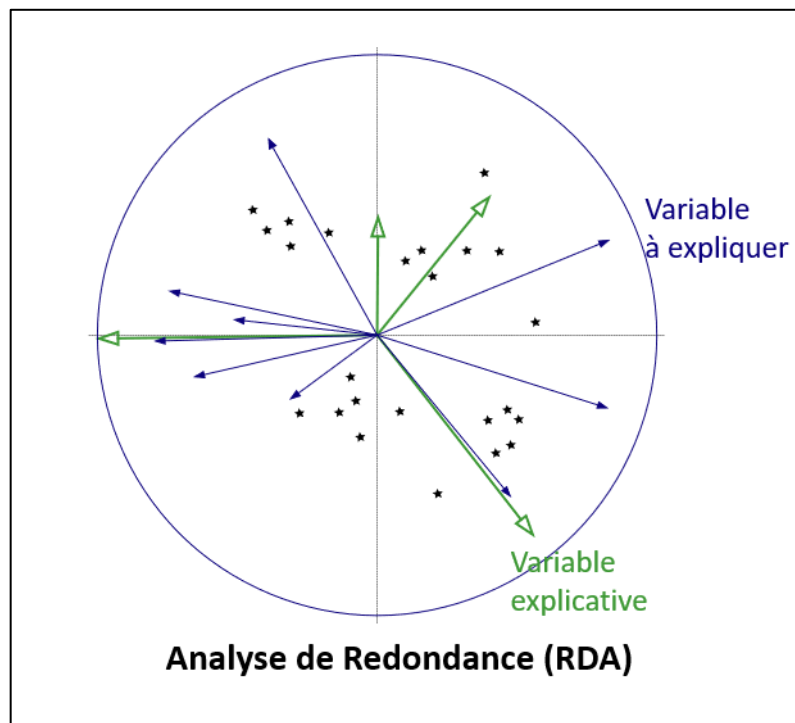


Figure 2.14 Principe de l'analyse de redondance ou RDA

2.5.3. Machine learning

Afin d'établir des calibrations de marqueurs (de pH ou de température par exemple), les méthodes de régression linéaire sont le plus souvent utilisées mais celles-ci souffrent du phénomène de dilution de la régression (biais de la pente de régression vers zéro causé par des erreurs dans la variable indépendante (axe des x)), comme cela a déjà été souligné (Naafs et al., 2017a; Dearing Crampton-Flood et al., 2020). En outre, ces modèles sont paramétriques et ne prennent pas en compte les potentielles influences non linéaires. C'est pourquoi d'autres modèles que les régressions simples/multiples ont également été proposés dans cette thèse : le modèle des k-plus proche voisin (k-NN) et le modèle Random Forest (non paramétriques et tenant compte des influences non linéaires). Ces modèles sont basés sur des algorithmes d'apprentissage automatique (ou machine learning), qui sont construits sur une proportion du jeu de données total (défini de manière aléatoire, il s'agit du jeu de données d'entraînement), puis testés sur le reste du jeu de données, considéré comme indépendant (jeu de données de test).

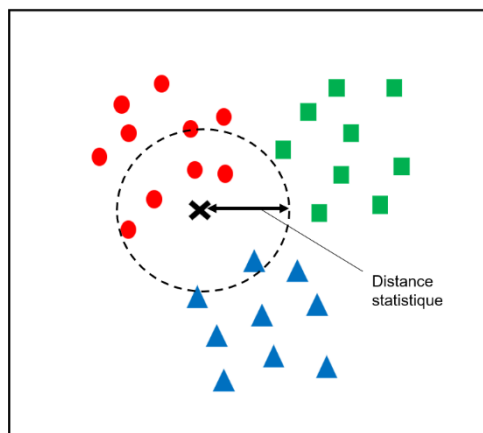


Figure 2.15 Principe de la méthode k-NN. L'échantillon inconnu x appartient vraisemblablement au groupe rouge selon cette méthode.

Le modèle k-NN est basé sur l'estimation des distances moyennes entre les différents échantillons. Il s'agit d'une méthode d'apprentissage supervisée (par exemple, Gangopadhyay et al., 2009). Une base de données d'apprentissage composée de N couples "entrée-sortie" est initialement constituée, afin de pouvoir estimer un résultat associé à une nouvelle

entrée x . La méthode des k-voisins prend en compte les k échantillons d'apprentissage dont l'entrée est la plus proche de la nouvelle entrée x , selon une distance à définir. Cette méthode est non-paramétrique et est utilisée pour la classification et la régression. Dans la régression par méthode k-NN, le résultat est la valeur de l'objet x , qui est la moyenne des valeurs des k plus proches voisins (Fig 2.15). Ses contraintes résident dans le fait que, par définition, si une plage de valeurs est plus fréquente que les autres, alors elle sera statistiquement

prédominante parmi les k plus proches voisins. Mais, dans le cas présent, la sélection des données a été effectuée de manière aléatoire sur le jeu de données avec une modalité de stratification selon le TMAA ou le pH pour limiter l'impact des valeurs extrêmes comme détaillé ci-dessous. Ceci permet de pallier cette limitation de la méthode k -NN.

L'algorithme Random Forest est également une méthode d'apprentissage supervisé utilisée, entre autres, pour les régressions (par exemple, Ho, 1995; Breiman, 2001; Denisko and Hoffman, 2018). Ce modèle fonctionne en construisant une multitude d'arbres de décision et en calculant la prédiction moyenne de ces différents arbres (Fig. 2.16). La robustesse de ce modèle réside dans le fait qu'il est (1) intuitif, (2) et que les résultats prédits sont les résultats d'une multitude de modèles indépendants (arbres de décisions)

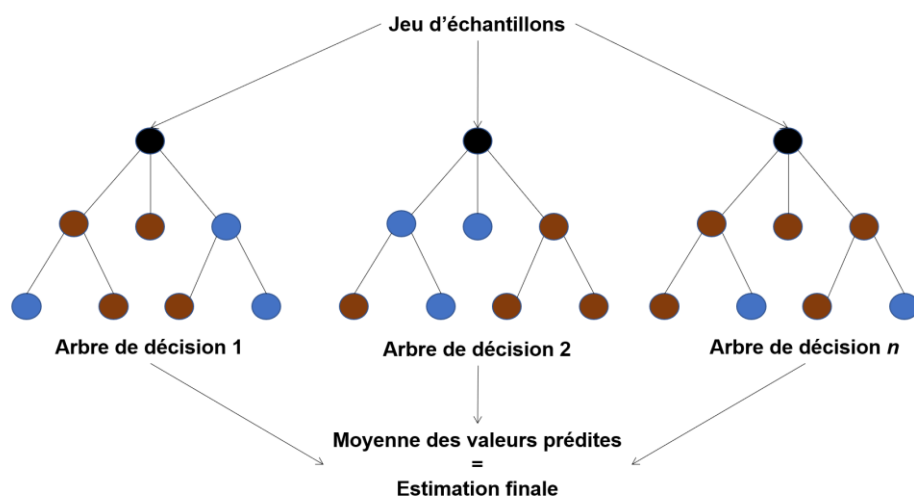


Figure 2.16 Principe de la méthode Random Forest pour les régressions. D'après Breiman (2001)

La phase d'apprentissage requise pour les différents modèles a été effectuée sur 75% de l'ensemble de l'échantillon avec une itération de dix validations croisées par modèle. La sélection des données a été effectuée de manière aléatoire sur l'ensemble des données (sans prétraitement) mais avec une modalité de stratification en fonction de la TMAA ou du pH pour limiter l'impact des valeurs extrêmes. Ensuite, la robustesse et la précision des différents modèles ont été testées sur les 25 % d'échantillons restants, considérés comme un ensemble de données indépendant. Les modèles k -NN et la random forest ont été réalisés avec le logiciel R, version 3.6.1 et version 4.0.3 (R Core Team, 2014) en utilisant les packages - tidymodels (version 0.1.0) - kknn (version 1.3.1), ranger (version 0.12.1).

Une application web est disponible en ligne (<https://athibault.shinyapps.io/paleotools>) pour la reconstruction des TMAA dérivées des 3-AGH en utilisant les modèles d'apprentissage automatique proposés dans la présente étude (Chapitre 4).

Un package R avec une application web est disponible sur un répertoire GITHUB ([paleoFROG](#)) pour la reconstruction des TMAA à partir des brGDGT en utilisant les modèles FROG proposés dans la présente étude (Chapitre 5).

Les performances du modèle FROG ont été comparées à celles des modèles bayésiens BayMBT et BayMBT₀ proposés par Dearing Crampton-Flood et al. (2020). Ces derniers ont été réalisés avec le code MATLAB disponible sur le dépôt GITHUB de Jessica Tierney (<https://github.com/jesstierney> ; Dearing Crampton-Flood et al., 2020), en utilisant MATLAB, version 9.8. La moyenne *a priori* a été fixée à la moyenne des TMAA pour tous les échantillons de sol (10°C), avec un écart type *a priori* de 30°C.

**Chapitre 3 – ETUDE DE MARQUEURS
LIPIDIQUES BACTERIENS AU SEIN DE
SOLS DES ALPES FRANÇAISES**

Dans ce chapitre sont présentées les analyses de marqueurs moléculaires de températures de l'air et de pH des sols, les 3-AGH et les brGDGT, au sein d'échantillons de sols des Alpes françaises. Ce chapitre est divisé en deux sous parties. Dans une première partie publiée dans *Organic Geochemistry*, l'influence des paramètres environnementaux (température, pH, végétation, granulométrie...) sur la distribution des 3-AGH et des brGDGTs a été étudiée dans 49 sols prélevés le long de transects altitudinaux dans les Alpes françaises (Massif du Galibier-Lautaret et des Bauges). Dans la deuxième partie (en français), sont présentées des expériences d'incubations en microcosmes réalisées sur quatre sols des Alpes, pendant un an à différentes températures.

3.1. Influence of environmental parameters on the distribution of bacterial lipids in soils from the French Alps: Implications for paleo-reconstructions

3.1.1. Introduction

Direct measurement of environmental data, such as temperature and precipitation, has been possible for the last two centuries, the so-called "instrumental" period. Beyond this period, it is necessary to use indirect approaches to obtain information on the different environmental parameters. "Indirect" indicators of these parameters – so-called proxies – have thus been developed and used regularly for last decades.

Organic biomarkers have been of great interest to the scientific community for the reconstruction of past environments over the last decades (Eglinton and Eglinton, 2008), especially microbial compounds. Microorganisms modify the lipid composition of their membranes (e.g. carbon chain length, number of saturations, branching level) in response to variations of environmental parameters (pH, temperature, osmotic pressure) (Ernst et al., 2016). Such modifications are considered to maintain a functional fluidity and permeability of the microbial membrane (Singer and Nicolson, 1972; Sinensky, 1974; Hazel and Williams, 1990; Denich et al., 2003). These microbial lipids can be preserved in soils and sediments and be used as proxies of past environmental conditions.

In this context, glycerol dialkyl glycerol tetraethers (GDGTs) are a family of microbial lipids widely used for paleoenvironmental reconstructions. These compounds are ubiquitous in terrestrial (Weijers et al., 2007; Peterse et al., 2012; De Jonge et al., 2014; Naafs et al., 2017a) and aquatic environments (Schouten et al., 2002, 2012; Powers

et al., 2010; Peterse et al., 2015; Weber et al., 2015). They are characterized by aliphatic chains connected to two glycerol units via ether bonds. Two groups of GDGTs – isoprenoid and branched – can be distinguished (Schouten et al., 2013 and references therein). Isoprenoid GDGTs (iGDGTs) are produced by archaea. In contrast, GDGTs with branched chains but no isoprenoid alkyl chains – so-called branched GDGTs (brGDGTs) – are produced by still unidentified bacteria, although some of them may belong to the phylum *Acidobacteria* (Sinninghe Damsté et al., 2011, 2014, 2018). They were discovered in peat (Sinninghe Damsté et al., 2000). The analysis of brGDGTs in a large number of soils distributed worldwide then showed that the relative distribution of these compounds is dependent on environmental parameters, mainly temperature and pH (Weijers et al., 2007; De Jonge et al., 2014, 2019).

Despite the wide application of brGDGT indices for paleoenvironmental reconstructions in terrestrial settings (Weijers et al., 2011b; Wang et al., 2017; Coffinet et al., 2018), the derived results must be interpreted with caution, as parameters other than temperature or pH, such as humidity (Huguet et al., 2010; Menges et al., 2014), soil type (Mueller-Niggemann et al., 2016; Davtian et al., 2016), vegetation composition (Weijers et al., 2011b; Naeher et al., 2014; Liang et al., 2019) or seasonality (Huguet et al., 2013) can also have an influence on the relative abundances of brGDGTs. These lipids are the only molecular proxies of temperature and pH available in the terrestrial environment to date, as most of the available proxies have been developed in oceanic environments. Despite improvements in brGDGT analytical methods and use of refined calibration models (De Jonge et al., 2014; Hopmans et al., 2016; Dearing Crampton-Flood et al., 2020), the Root Mean Square Error (RMSE) associated with Mean Annual Air Temperature (MAAT) reconstruction using the global brGDGT calibrations in soils remains high ($>4^{\circ}\text{C}$), possibly due to the lack of knowledge on all influencing environmental parameters. The development of new molecular proxies, independent of and complementary to brGDGTs, is essential to improve the reliability of environmental reconstructions in terrestrial settings.

Recent studies (C. Wang et al., 2016; Wang et al., 2018; Huguet et al., 2019; Yang et al., 2020) have unveiled the potential of another family of lipids – 3-hydroxy fatty acids (3-OH FAs) – for temperature and pH reconstructions. 3-OH FAs with 10 to 18 carbon atoms are specifically produced by all Gram-negative bacteria and are bound to the

lipopolysaccharide (LPS; main component of the outer membrane) by ester or amide bonds. 3-OH FAs are part of the lipid A, which anchors the LPS in the outer membrane of Gram-negative bacteria. Three types of 3-OH FAs can be distinguished: with *normal* (i.e. straight and unbranched) or branched carbon chains, either *iso* or *anteiso*. These compounds were widely used (i) to quantify Gram-negative bacteria in clinical studies and detect the presence of endotoxins they release (Wollenweber and Rietschel, 1990; Saraf et al., 1997; Szponar et al., 2002, 2003; Keinänen et al., 2003) and (ii) to characterize and quantify Gram-negative bacteria communities in environmental samples such as aerosols (Lee et al., 2004; Cheng et al., 2012), dissolved organic matter (Wakeham et al., 2003), or soils (Zelles et al., 1995; Zelles, 1999).

3-OH FAs have been recently proposed as proxies of temperature and pH in terrestrial environments after analysis of these compounds in 26 soils along Mount Shennongjia, China (C. Wang et al., 2016). New indices were developed, with RAN_{15} and RAN_{17} (defined as the ratio of C_{15} or C_{17} *anteiso* 3-OH FAs to *normal* C_{15} or C_{17} 3-OH FAs) being correlated with MAAT and RIAN (defined as $-\log ([normal\ 3-OH\ FAs]/[anteiso+iso\ 3-OH\ FAs])$) being dependent on soil pH.

Significant relationships between 3-OH FA distribution and temperature as well as pH were similarly observed along two additional altitudinal transects: Mount Majella in Italy and Mount Rungwe in Tanzania (Huguet et al., 2019). The RAN_{15}/RAN_{17} indices were negatively correlated to air temperature along the three mountains investigated so far (C. Wang et al., 2016; Huguet et al., 2019). This suggests that Gram-negative bacteria respond to colder temperatures with an increase in *anteiso*- C_{15}/C_{17} vs. *n*- C_{15}/C_{17} 3-OH FAs in order to maintain a proper fluidity and optimal state of the bacterial membrane, the so-called homeoviscous adaptation mechanism (Sinensky, 1974; Hazel and Williams, 1990). Nevertheless, the relationships between RAN_{15} and MAAT along Mts. Shennongjia, Rungwe and Majella showed the same slopes but different intercepts (C. Wang et al., 2016; Huguet et al., 2019), suggesting that regional calibrations may be more adapted to apply RAN_{15} as a temperature proxy in soils. In contrast, a significant ($R^2 = 0.60$) combined calibration between RAN_{17} and MAAT could be established using data from Mts. Shennongjia, Rungwe and Majella (C. Wang et al., 2016; Huguet et al., 2019). Similarly, RIAN was shown to be strongly negatively correlated with soil pH along the three aforementioned mountains (C. Wang et al., 2016; Huguet et al., 2019), reflecting a

general relative increase in normal homologues compared to branched (*iso* and *anteiso*) ones with increasing pH. This mechanism was suggested to reduce the permeability and fluidity of the membrane for the cell to cope with lower pH (Watanabe and Takakuwa, 1984; Lepage et al., 1987; Russell, 1989; Russell et al., 1995; Denich et al., 2003; Beales, 2004). 3-OH FA indices were recently applied to estimate temperature and hydrological changes over the last 10,000 years in a speleothem from China (Wang et al., 2018), showing the potential of 3-OH FAs as independent tools for environmental reconstruction in terrestrial settings. A very recent study based on marine sediments from the North Pacific Ocean suggested that the distribution of 3-OH FAs could also be used to reconstruct sea surface temperature (Yang et al., 2020).

Even though these results are promising, the linear regressions between pH/MAAT data and 3-OH FA indices in terrestrial environments are still based on a rather small dataset (ca. 70 soil samples) and show a large degree of scatter, leading to substantial errors in 3-OH FA-based MAAT and pH reconstitutions (MAAT-RMSE = 5.1°C; pH-RMSE = 0.56; Huguet et al., 2019). One can anticipate that the distribution of 3-OH FAs in soils is impacted by environmental parameters other than the temperature and pH, as observed for brGDGTs. A full understanding of the environmental parameters (e.g. MAAT, pH, soil moisture, organic carbon and nitrogen content, soil types and vegetation communities) influencing bacterial lipid distribution (3-OH FAs and brGDGTs) in soils is required to increase the reliability of the temperature and pH proxies. Nevertheless, to date, there is no report of a detailed investigation of the environmental controls on 3-OH FA distribution in soils and only a limited number of studies on brGDGTs address this question. The aim of this work is thus to determine and quantify the cumulative effect of environmental parameters on the distribution of both brGDGTs and 3-OH FAs along a composite altitudinal transect in a well-constrained sampling site in the French Alps (Annex 1; Supp. Fig 3.1).

3.1.2. Results

3.1.2.1. *Bulk soil properties*

The organic carbon content (C_{org}) shows a wide range of variation— from 1.4% for poorly evolved soils (lithosol) to 49.10% for forest organosol (litter soil) – with an average of 9.7% and a standard deviation of 10.3% (Table 3.1). Only a minority of soils ($n = 4$) display high C_{org} values ($> 30\%$). The $\delta^{13}C$ values show a unimodal distribution ($-26.7\text{‰} \pm 2.3\text{‰}$) and range between -29.9‰ and -24.4‰ (Table 3.1). The N content of the soil samples is normally distributed ($0.7\% \pm 0.5\%$; Table 3.1). Relative soil moisture is unimodally distributed and largely varies between soils, ranging between 3.4 and 68.2% (mean $30.3\% \pm 15.6\%$; Table 3.1). With regard to the texture of the sampled soils, the clay contents are on average 313.8 ± 126.7 g/kg of soil and largely vary between samples (minimum and maximum values of 33 and 546 g/kg of soil, respectively). This large variability of the soil characteristics is also reflected in the silt (406.4 ± 117.0 g/kg soil) and sand (279.8 ± 194.1 g/kg soil) contents, as well as in the CEC values (24.2 ± 17.0 cmol +/kg; Table 3.1). Finally, the pH values show a wide range of variation (3.62–7.49), with an average of 5.83 and a standard deviation of 1.08.

ETUDE DE MARQUEURS LIPIDIQUES BACTERIENS AU SEIN DE SOLS DES ALPES
FRANÇAISES

Samples	MAAT (°C)	pH	C org (%)	N (%)	$\delta^{13}\text{C}$	$\delta^{15}\text{N}$	Moisture (%)	Clay (g/kg)	Silt (g/kg)	Sand (g/kg)	CEC (cmol +/kg)
1	0.4	5.1	3.8	0.31	-25.9	2.8	12.4	112	207	681	49.1
2	0.6	6.3	6.2	0.59	-25.4	3.0	27.5	273	411	316	22.6
3	0.1	6.0	5.9	0.57	-24.4	6.7	22.7	312	438	250	21.4
4	0.4	6.5	12.7	1.22	-24.9	2.8	39.4	323	407	270	38.3
5	1.5	4.7	7.0	0.62	-25.4	3.6	29.5	380	457	163	20.8
6	0.7	4.9	5.0	0.46	-25.1	4.3	30.3	290	579	131	16.1
7	0.4	4.9	7.7	0.59	-25.6	2.1	28.9	385	467	148	21
8	1.3	4.8	2.0	0.21	-25.0	4.0	26.0	248	623	129	10.3
9	1.0	4.5	4.1	0.37	-25.2	3.7	31.4	321	557	122	13.4
10	1.2	4.8	7.3	0.56	-25.7	8.5	34.1	378	461	161	19.2
11	1.3	5.7	2.0	0.18	-26.3	4.2	8.6	197	423	380	9.48
12	1.2	5.9	1.4	0.15	-25.7	4.2	9.1	136	381	483	8.27
13	1.1	7.1	1.5	0.16	-25.4	1.9	3.4	115	310	575	5.98
14	2.7	4.7	17.8	1.10	-25.8	-0.2	45.8	448	406	146	31.9
15	2.3	5.1	13.6	0.81	-25.3	0.0	44.8	368	405	227	28.6
16	2.2	5.5	17.7	1.22	-25.0	2.0	52.9	432	444	124	47.5
17	3.3	4.8	9.1	0.71	-25.5	2.9	26.3	416	424	160	23.5
18	4.2	5.2	5.9	0.50	-26.1	2.5	16.6	377	420	203	20.4
19	3.6	5.2	6.6	0.58	-25.4	3.3	22.9	412	428	160	23.1
20	4.1	6.2	6.5	0.57	-25.7	1.7	16.0	395	392	213	24.6
21	4.7	7.3	6.3	0.54	-25.7	2.6	12.2	385	399	216	19.4
22	5.7	6.3	5.1	0.34	-27.4	1.7	5.8	194	232	574	13.2
23	5.7	6.2	9.4	0.78	-27.0	1.2	23.9	254	299	447	24.6
24	5.9	6.6	7.5	0.64	-26.9	3.2	11.7	277	341	382	23.3
25	6.1	6.7	5.8	0.56	-26.4	4.2	13.6	350	381	269	20.2
26*	7.5	4.7	3.1	0.28	-26.7	1.1	31.8	321	443	236	15.8
27	6.5	6.0	3.8	0.40	-26.9	5.1	33.1	344	457	199	21
28*	7.4	5.1	5.9	0.53	-26.2	1.0	35.6	396	434	170	23.1
29*	6.0	4.9	32.9	1.89	-26.7	-5.0	61.7	513	458	29	74.7
30	6.9	4.8	49.1	2.10	-25.1	-2.50	64.7	540	428	32	64.3
31*	7.7	5.2	41.0	2.03	-39.9	-3.8	68.2	546	452	2	96
32	6.9	4.8	4.7	0.33	-25.6	-1.0	33.7	305	525	170	21.5
33	6.9	4.9	6.4	0.49	-26.5	0.0	36.4	398	458	144	26.9
34	6.9	5.5	6.8	0.42	-25.2	-0.8	39.9	345	541	114	30.5
35	7.3	3.8	29.0	1.37	-27.7	-1.8	62.7	114	191	695	11.5
36*	7.4	3.6	31.0	1.74	-28.8	-1.5	53.4	98	100	802	13.2
37	7.1	5.5	2.4	0.12	-27.9	-0.3	21.0	33	98	869	8.05
38	9.2	6.4	19.8	1.05	-27.9	-2.7	46.4	411	285	304	32.9
39*	8.6	7.2	11.6	0.88	-26.5	-1.0	28.0	432	304	264	34.7
40*	7.9	6.7	5.8	0.39	-27.1	-0.6	30.4	516	294	190	35.8
41*	12.0	7.5	8.7	0.86	-27.1	2.3	23.3	357	329	314	24.7
42	10.8	6.7	4.3	0.35	-28.4	-0.2	26.1	477	314	209	28.4
43*	11.2	7.1	1.7	0.18	-25.9	7.0	17.5	276	330	394	17.8
44*	12.5	7.3	5.2	0.53	-27.9	2.9	26.8	227	577	196	11.8
45	12.5	7.5	7.3	0.69	-28.0	1.9	26.5	181	510	309	10.8
46	12.4	7.3	2.7	0.24	-29.4	1.3	31.6	95	501	404	9.65
47*	12.7	7.3	2.2	0.23	-28.3	5.0	27.2	171	559	270	9.02
48	12.5	7.5	9.8	0.90	-28.7	2.2	47.2	327	599	74	26.9
49	12.5	7.6	2.4	0.21	-27.7	6.2	16.0	175	434	391	7.32

Table 3.1 Physicochemical characteristics of the 49 samples collected along the Bauges and Lautaret-Galibier massifs. The numbering of the samples corresponds to that used in Table 2.2. Samples with stars represent sites instrumented with temperature sensors.

3.1.2.2. *BrGDGT* and 3-OH FA distribution

3-OH FAs and brGDGTs were detected in the 49 studied soil samples (Annex 1; Supp. Tables 3.1 and 3.2). BrGDGT distribution was dominated by acyclic compounds (Ia, IIa, IIa', IIIa, IIIa') which present in average ca. 80% of total brGDGTs (Fig. 3a). BrGDGTs Ia and IIa were the most abundant homologues in all samples. The pentamethylated brGDGTs (IIa-c; mean 50%, standard deviation (SD) of 6.9%) were predominant over the tetramethylated (Ia-c) and hexamethylated (IIIa-c; Fig. 3.1a). The 5-methyl isomers were present in higher proportion (mean 74%, SD 15.6%) than the 6-methyl compounds (mean 26%, SD 15%) (Fig. 3.1a).

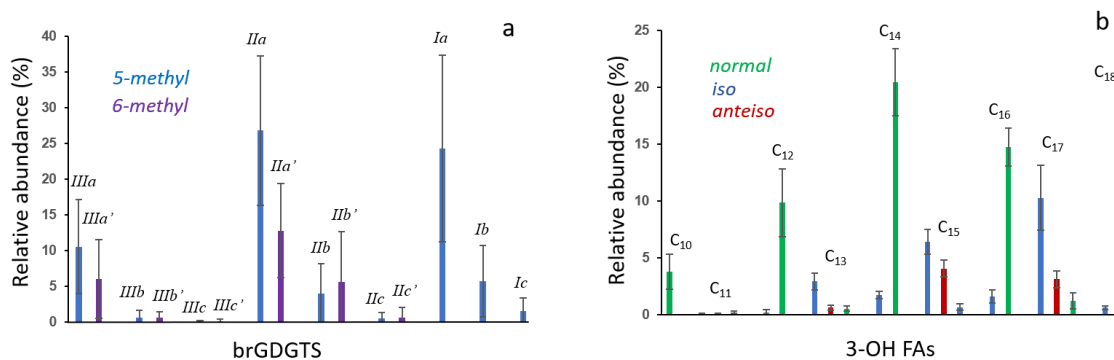


Figure 3.1 Average distribution of (a) brGDGTs and (b) 3-OH FAs in the 49 soil samples of the French Alps.

3-OH FAs with chain lengths between 8 and 26 C were identified. These compounds may have various origins depending on their chain length (bacteria, plants and fungi; Zelles, 1999; C. Wang et al., 2016; and reference therein). Nevertheless, most of the 3-OH FAs present in our samples are homologues with 10 to 18 C atoms, generally associated with Gram-negative bacteria (Saraf et al., 1997; Szponar et al., 2002, 2003; Keinänen et al., 2003) and only these homologues will be considered in the following. 3-OH FAs with even carbon number and *normal* chains were the most abundant compounds in all samples (about 65% of the 3-OH FAs, SD 12%), with a predominance of the *n*-C₁₄ homologue (Fig. 3.1b). *Iso* (mean 24%, SD 4.5%) and *anteiso* (mean 8%, SD 1%) isomers were also present. It must be noted that *anteiso* isomers were only detected for odd 3-OH FAs (Fig. 3.1b).

3.1.3. Discussion

The distribution of brGDGTs (Weijers et al., 2007; Peterse et al., 2012; De Jonge et al., 2014) and 3-OH FAs (C. Wang et al., 2016; Huguet et al., 2019) in soils was reported to be influenced mainly by MAAT and soil pH. Therefore, the effect of these two environmental parameters on the relative abundance of 3-OH FAs and brGDGTs in the 49 soils of the French Alps was first investigated.

3.1.3.1. *BrGDGTs and 3-OH FAs as proxies of pH and temperature*

- Relationship between pH and bacterial lipid distribution

BrGDGT-derived CBT index has been shown to be strongly correlated with pH in a large set of globally distributed soils (De Jonge et al., 2014; Naafs et al., 2017a). Consistently, in the French Alpine soils, CBT displayed a significant ($p < 0.05$) and strong correlation ($R^2 = 0.85$; RMSE = 0.42) with soil pH (Fig. 3.2a), in accordance with the global CBT-pH calibration proposed by De Jonge et al. (2014). This confirms the robustness of the CBT as a soil pH proxy taking into account the highly contrasting environmental characteristics of these French Alps soils.

The soils of the French Alps (Table 3.1) and those of Mts. Rungwe, Majella and Shennongjia, previously investigated for 3-OH FA distribution (C. Wang et al., 2016; Huguet et al., 2019), cover a similar range of pH (ca. 3–8) and of RIAN index (0.13–0.55 in the French Alps, Table 3.1; 0.15–0.70 for the other soils, C. Wang et al., 2016; Huguet et al., 2019). In the French Alps, the RIAN index was observed to be moderately ($R^2 = 0.43$; RMSE = 0.83) but significantly negatively correlated ($p < 0.001$) with soil pH (Fig. 3.2b), consistent with the significant correlation observed along the previously investigated altitudinal transects (C. Wang et al., 2016; Huguet et al., 2019). This confirms the general influence of pH on the relative abundance of 3-OH FAs in soils.

Nevertheless, a stronger correlation between RIAN and pH ($R^2 = 0.62$) was obtained for the dataset combining soils from Mts. Rungwe, Majella and Shennongjia (Huguet et al., 2019) compared to the one from the French Alps. In addition, the RIAN-pH correlations along the French Alps (Fig. 3.2b) and along Mts. Rungwe, Majella and Shennongjia (Huguet et al., 2019) displayed significantly different slopes and intercepts

(p -value = 0.049), reflecting local trends between 3-OH FA distribution and pH. Finally, the scatter in the RIAN-soil pH relationship along the French Alps (Fig. 3.2b), reflected in the average value of the prediction interval (0.35), suggests that parameters other than pH also affect the distribution of 3-OH FAs along this transect.

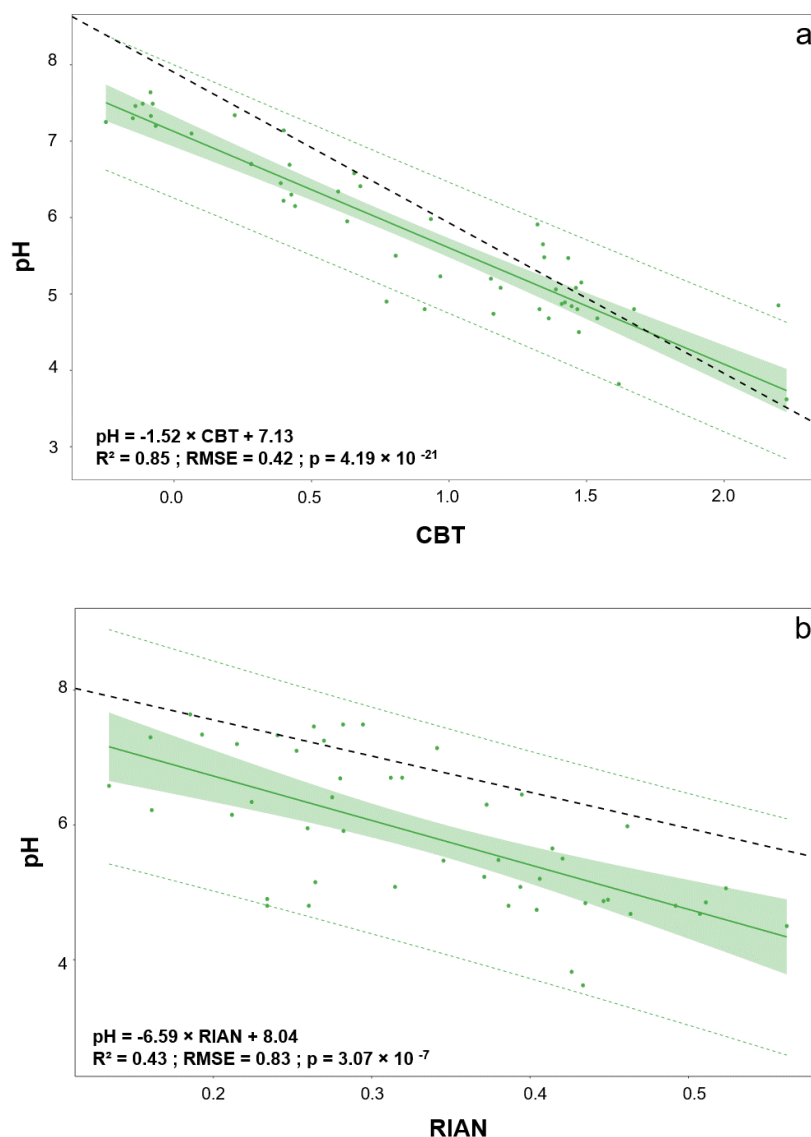


Figure 3.2 (a) Linear regression between CBT and pH along the two massifs in the French Alps. The black dotted line corresponds to the global linear regression between CBT and pH from De Jonge et al. (2014). (b) Linear regression between RIAN and pH along the two massifs in the French Alps. The black dotted line corresponds to the linear regression between RIAN and pH from Huguet et al. (2019). Dotted and colored lines represent the 95% prediction interval for each regression and colored areas represent the 95% confidence interval for each regression.

- Relationship between MAAT and bacterial lipid distribution

In addition to pH, the relationship between bacterial membrane lipid distribution and MAAT was also investigated. BrGDGT-based MBT'_{5Me} was significantly ($p < 0.001$) negatively correlated with MAAT in the French Alps (Fig. 3.3a). Nevertheless, this correlation was substantially weaker ($R^2 = 0.35$) than the one between MAAT and MBT' or MBT'_{5Me} for (i) soils distributed worldwide ($R^2 = 0.66$; De Jonge et al., 2014) and (ii) along other altitudinal transects: Mount Kilimanjaro, Tanzania ($R^2 = 0.75$; Sinninghe Damsté et al., 2008); Mount Xiangpi, China ($R^2 = 0.64$; Liu et al., 2013); Mount Rungwe, Tanzania ($R^2 = 0.74$; Coffinet et al., 2014); Andean transect, Colombia ($R^2 = 0.67$; Anderson et al., 2014), East Coast of the USA ($R^2 = 0.47$; Dirghangi et al., 2013). The lower RMSE in this study (3.3 °C), in comparison with the RMSE of the global calibration (4.8 °C) could be partly explained by the relatively small number of samples ($n = 49$) analyzed within the French Alps, compared to the global dataset ($n = 231$) presented by De Jonge et al. (2014). Even though the relationship between brGDGT distribution and MAAT is generally strong, there is a substantial scatter in the MBT'_{5Me}-MAAT global calibration, leading to the high RMSE mentioned above (De Jonge et al., 2014), which reflects the heterogeneous and local response of brGDGT distribution to temperature. Such heterogeneity is also visible at the regional level, with e.g. significant differences in the relationship between brGDGT distribution and MAAT/soil pH along several altitudinal transects subjected to different precipitation regimes (Anderson et al., 2014). The weak relationship between MBT'_{5Me} and MAAT in the French Alps, associated with an uncertainty (RMSE) of 3.3 °C, reflects the fact that local parameters other than MAAT should be the primary controls of 5-methyl brGDGT relative abundance along this specific transect.

Regarding 3-OH FAs, the temperature proxies based on these compounds, namely RAN₁₅ and RAN₁₇, did not show any correlation with MAAT in the French Alps (Fig. 3.3b, c). In contrast, significant negative correlations between RAN₁₅ and MAAT were previously observed along Mt. Shennongjia ($R^2 = 0.51$; C. Wang et al. 2016), Mt. Rungwe and Mt. Majella ($R^2 = 0.80$; $p < 0.001$ and $R^2 = 0.54$; $p = 0.01$, respectively; Huguet et al., 2019) with similar slopes but different intercepts. This was interpreted as the result of local effects on RAN₁₅, despite a similar general response to temperature changes

(Huguet et al., 2019). As observed for RAN₁₅, RAN₁₇ was significantly negatively correlated with MAAT along Mt. Shennongjia ($R^2 = 0.48$; C. Wang et al. 2016), Rungwe and Majella ($R^2 = 0.54$ and 0.28 , respectively; Huguet et al., 2019) with similar slopes and ordinates, showing a comparable response of RAN₁₇ to temperature variations in these different locations. Nevertheless, there was only a weak correlation between RAN₁₇ and MAAT along Mt. Majella (Huguet et al., 2019), characterized by the closest altitudinal, geographical and climatic conditions to those of the French Alps in comparison with other investigated mountains. This result, combined with the absence of correlation between either RAN₁₅ or RAN₁₇ and the MAAT in the French Alps (Fig. 3.3b), suggests that locally the relative abundance of C₁₅ and C₁₇ 3-OH FAs may be mainly influenced by environmental parameters other than MAAT.

The Bauges and Lautaret-Galibier massifs, investigated in the present study, are a combination of highly diverse microecosystems resulting from microtopographical and geological (nature of the bedrock) differences and characterized by specific vegetation and soil types, precipitation or snow cover. This multiplicity of environmental variables may therefore affect the distribution of bacterial lipids in soils, as previously shown for brGDGTs (e.g. soil moisture, Dirghangi et al., 2013; granulometry, Peterse and Eglinton, 2017; vegetation or soil type, Davtian et al., 2016, Liang et al., 2019). This might reflect either changes in bacterial communities as previously reported (Siles and Margesin, 2016; Hofmann et al., 2016; Shen et al., 2019) and/or adaptation mechanisms of a constant community. Statistical analyses were performed to disentangle and quantify the main drivers of changes in bacterial lipid distribution along the Bauges and Lautaret-Galibier massifs and thus explain the weakness or lack of correlations of the 3-OH FA-/brGDGT-derived proxies with MAAT.

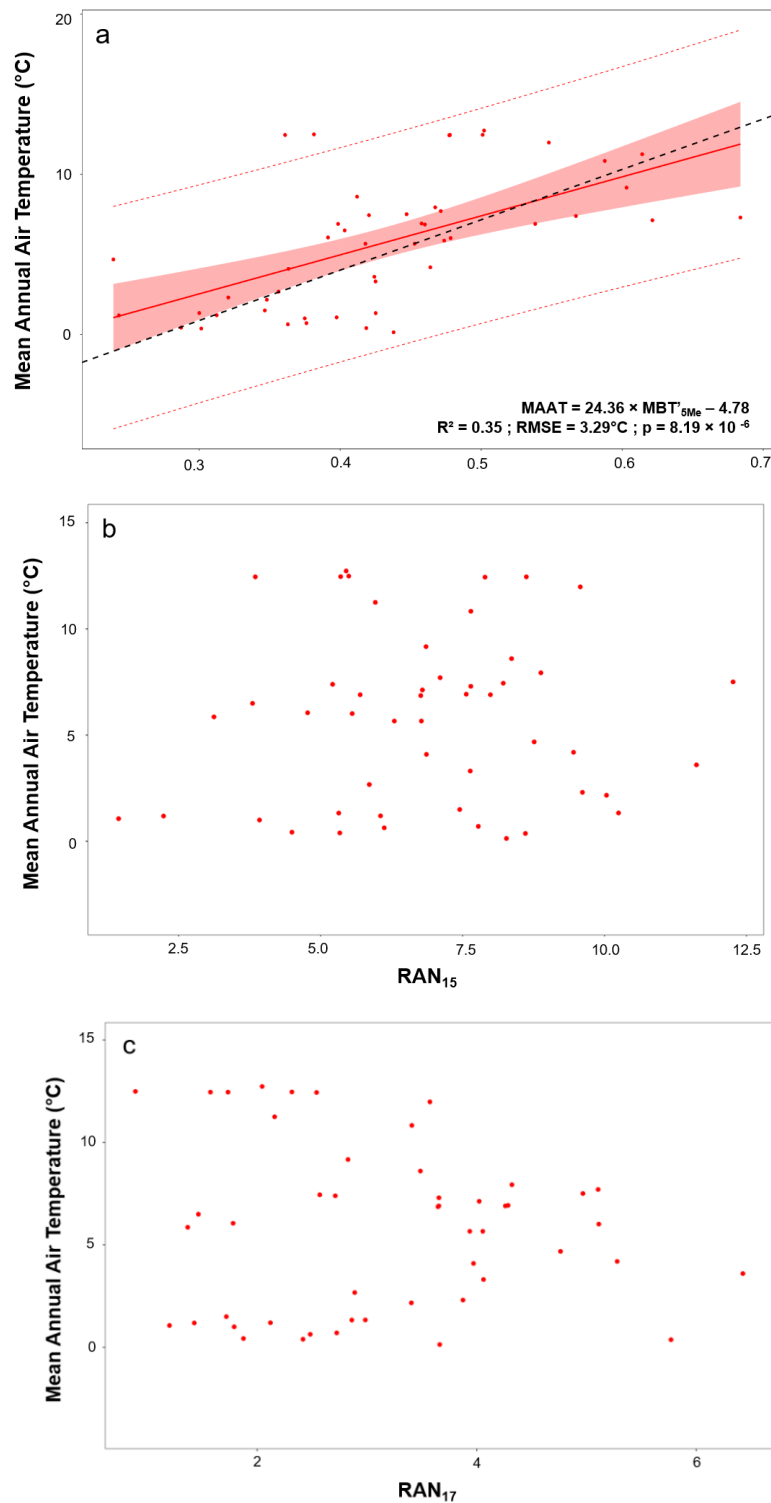


Figure 3.3 (a) Linear regression between MAAT and MBT'_{5Me} along the two massifs in the French Alps. The black dotted line corresponds to global the linear regression between MBT'_{5Me} and MAAT from De Jonge et al. (2014). Linear regressions between (b) RAN_{15} and (c) RAN_{17} along the two massifs in the French Alps. Dotted and colored lines represent the 95% prediction interval for each linear regression and colored areas represent the 95% confidence interval for each regression.

3.1.3.2. *Impact of other environmental variables on bacterial lipid distribution*

- Selection of non-redundant environmental variables for statistical analyses

Before constraining the influence of the different environmental factors on 3-OH FA and brGDGT distribution using redundancy analyses (RDA), non-redundant environmental variables, which carry significantly different information, had to be selected. To this aim, a Spearman correlation matrix was first computed (Fig. 3.4), allowing the determination of the correlations between the different environmental variables. Highly correlated variables can be considered as redundant. Only correlation coefficients with red (negative correlation) or blue (positive correlation) colors are statistically significant ($p < 0.05$) (Fig. 3.4).

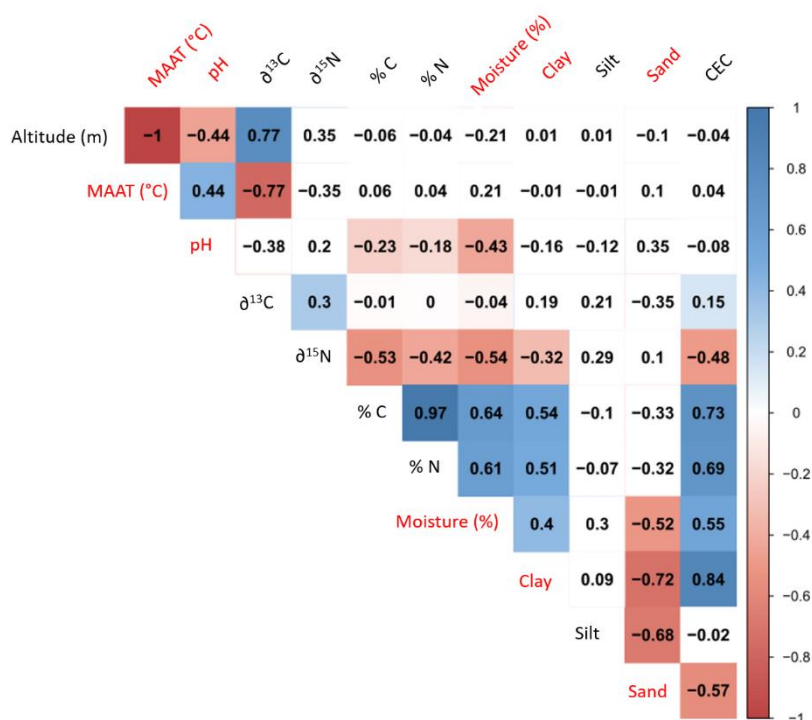


Figure 3.4 Correlation matrix between the different environmental variables and soil properties. Correlation coefficients in red (negative correlation) or blue (positive correlation) are statistically significant. Red variables have been chosen as explanatory variables for statistical tests.

As MAAT values were reconstructed using a linear model with altitude (cf. section 2.1.1.4.), MAAT and altitude were logically anticorrelated along the Bauges and Lautaret-Galibier, reflecting the natural cooling of air with elevation (Fig. 3.4). Altitude (and thus MAAT) was moderately correlated with soil pH ($r = -0.44$) and strongly correlated with $\delta^{13}\text{C}$ ($r = 0.77$). A relationship between soil pH and altitude was similarly observed along other altitudinal gradients but the sign of the correlation was shown to be strongly site-dependent (Cheng-Jim et al., 2014; Bhandari and Zhang, 2019). It can be influenced by multiple factors such as soil types, source rock (e.g. calcareous or granitic) or vegetation composition (e.g. presence or absence of conifers; Smith et al., 2002; Djukic et al., 2010; Gutiérrez-Girón et al., 2015). Similarly, the relationship between $\delta^{13}\text{C}$ and altitude could be due to different vegetation, litter input and degradation of OM with thermal regimes, as suggested along other mountains (Wei and Jia, 2009; Du et al., 2014). Indeed at higher altitudes, the degradation of OM may be reduced, in contrast with the lower altitudes and higher temperatures, where plant accumulation is higher than decomposition (Davidson and Janssen, 2006; Conant et al., 2011).

Altitude is correlated to a set of environmental parameters, such as MAAT, carbon content, pH and can be considered as an integrative variable of the changes in vegetation, soil and relief types. In order to observe the sole effect of MAAT variation on brGDGT and 3-OH FA distribution, MAAT rather than altitude was kept as a non-redundant parameter. The $\delta^{15}\text{N}$ of the different soil samples showed moderate correlations ($r < 0.7$) with six of the eleven parameters investigated (Fig. 3.4) and was therefore considered as a confounding parameter. Organic carbon content (C_{org}) of the soils was observed to be strongly correlated with the nitrogen content ($r = 0.97$) and $\delta^{15}\text{N}$ ($r = -0.53$) as well as with the cation exchange capacity ($r = 0.73$; Fig. 3.4). C_{org} was also moderately correlated with soil water content ($r = 0.64$) and soil clay content ($r = 0.54$; Fig. 3.4) and was therefore considered as a redundant variable in the present study. Similarly, based on collinearity relationships with other variables, N content, $\delta^{15}\text{N}$ and CEC will not be considered in the following statistical tests. In order to confirm the choice of the non-redundant variables for RDA analyses made using the Spearman correlation matrix (Fig. 3.4) and determine the best set of variables explaining the distribution of brGDGTs/3-OH FAs, a forward selection analysis was performed through the Canoco software. This highlighted MAAT,

pH, soil moisture, sand and clay contents (reflecting the grain size gradient of the soil) as explanatory variables, in line with Spearman correlation tests (Fig. 3.4).

The choice of these variables for subsequent RDA analysis is not unexpected, as MAAT and pH were shown to have a major influence on bacterial communities, which develop adaptation mechanisms to cope with variations of such parameters (Denich et al., 2003; Beales, 2004; Fierer and Jackson, 2006). This is especially the case for brGDGT- and 3-OH FA-producing bacteria (Huguet et al., 2019; De Jonge et al., 2019). Similarly, soil moisture may affect the relative abundance of bacterial membrane lipids, as shown for brGDGTs (e.g. Peterse et al., 2012; Dirghangi et al., 2013; Ding et al., 2015). Soil moisture varies, as a first approximation, according to the precipitation regime, but it can also be related to other factors such as evapotranspiration, grain size or vegetation cover (Crave and Gascuel-Odoux, 1997; Gómez-Plaza et al., 2001). The grain size gradient is representative of the edaphic parameters which may influence bacterial diversity in soils (Hemkemeyer et al., 2018) in relationship with changes in soil (micro)structure and organic matter availability and which may also in turn affect the distribution of 3-OH FA and brGDGT source microorganisms in soils, as proposed below.

Selecting only non-redundant variables for RDA simplifies the analysis and makes it more robust. Nevertheless, RDA analyses carried out with all the variables, confirmed that MAAT, pH and soil moisture were the parameters mainly influencing 3-OH FA and lipid distribution (data not shown). Therefore, in the following, we will only discuss the results of the RDA analyses performed with the aforementioned non-redundant variables.

- **RDA analysis of brGDGT distribution and environmental variables**

Regarding the RDA related to brGDGTs (Fig. 3.5a, Table 3.2), the first two axes explain 50.2 % of the total inertia of the dataset. The selected environmental variables are sufficient to explain 89.4 % of the variance in the brGDGT relative abundance (referred to as explained fitted variation in Table 3.2). On the first axis, the RDA analysis shows a predominant effect of pH ($r = 0.96$; Table 3.2) and a weaker contribution of MAAT ($r = 0.71$) and soil moisture ($r = -0.26$) on the distribution of brGDGTs. Axis 2, on the other hand, is influenced by MAAT ($r = 0.62$), relative soil moisture ($r = 0.53$) and to a lesser extent by pH ($r = -0.26$) and soil texture parameters, with a positive correlation with the

ÉTUDE DE MARQUEURS LIPIDIQUES BACTÉRIENS AU SEIN DE SOLS DES ALPES FRANÇAISES

sand content ($r = 0.08$) and a negative one with the clay content ($r = -0.20$; Fig. 3.5a and Table 3.2).

To go further, the influence of each environmental variable on the relative distribution of brGDGTs can be quantified through RDA analysis (Table 3.3). When the effect of each variable is considered individually, the influence of pH on brGDGT distribution appears predominant (41.1 %), followed by MAAT (25.2 %) and to a lesser extent soil moisture (5.3 %; Table 3.3). This is in line with observations made from the global soil dataset by De Jonge et al. (2014), where the variance of the relative abundances of the 15 individual brGDGTs in 231 soils was investigated using a Principal Component Analysis (PCA). When the first two components of this PCA are considered, it appears that pH was positively correlated with the largest component of the variance (40 %), while MAAT and MAP were correlated together with the other component (17 %).

Variables	brGDGTs		3-OH FAs	
	Axis 1	Axis 2	Axis 1	Axis 2
pH	0.96	-0.26	0.95	0.23
MAAT (°C)	0.71	0.62	0.29	-0.93
Moisture (%)	-0.26	0.53	-0.56	-0.41
Clay (g/kg)	-0.14	-0.2	-0.32	0.006
Sand (g/kg)	0.02	0.08	-0.21	-0.06
Expl. variation (%)	44.09	6.13	21.27	10.51
Expl. fitted variation (%)	78.47	10.91	54.39	26.84

Table 3.2 RDA correlation coefficients of the environmental variables selected along axes 1 and 2.

In the natural environment, environmental parameters do not influence bacterial lipid distribution separately, but are interrelated and act concomitantly, with e.g. soil moisture impacting the capacity of a soil to retain heat, thus indirectly influencing the temperature (Idso et al., 1975; Davidson et al., 1998). Therefore, the combined effect of the different environmental variables on brGDGT distribution was also investigated statistically. The conditional influence of these variables (Table 3.3) on the brGDGT distribution shows once again a clear predominance of the effect of pH (41.1 %), with a reduced influence of temperature (8.1 %), particle size (5.5 %) and soil moisture (1.5 %; Table 3.3). pH, grain size (clay/sand) content and soil moisture can be all considered as variables related to the local environment, i.e. microtopography, vegetation cover and soil type. When combined together, the influence of these local parameters on brGDGT

distribution amounts to 48.1 % against only 8.1 % for MAAT (Table 4), which can explain the weakness of the MBT'_{5Me}-MAAT relationship in soils from the French Alps and its large degree of scatter (Fig. 3.5a).

Variables	brGDGTs		3-OH FAs	
	Simple effets (%)	Conditional effects (%)	Simple effets (%)	Conditional effects (%)
pH	41.1 ***	41.1***	20.0 ***	20.0 ***
MAAT (°C)	25.2 ***	8.1 ***	11.2 **	10.9 ***
Moisture (%)	5.3 *	1.5 °	9.9 ***	3.6*
Clay (g/kg)	2.8	2.0 *	3.8°	2.5 °
Sand (g/kg)	1.4	3.5 **	2.1	2.2

Table 3.3 Quantification of the influence of the different environmental variables on brGDGTs and 3-OH FAs. Statistical significancies are shown as follows: * (0<p<0.001); ** (0.001<p<0.01); * (0.01<p<0.05); ° (0.05 <p<0.1).**

The RDA triplot also illustrates the influence of the different environmental variables on the relative abundance of the individual brGDGTs (Fig. 3.5a). BrGDGTs comprising cyclic moieties (i.e with subscript b or c) are strongly and mainly positively correlated with the first axis of the RDA, and thus positively correlated with pH. As most of these cyclic compounds are used in the calculation of the CBT index, this explains the strong correlation between CBT and pH observed in the soils of the French Alps ($R^2 = 0.85$; Fig. 3.2a), consistent with the trends observed in soils distributed worldwide (De Jonge et al., 2014). In contrast with cyclic brGDGTs, acyclic 5-methyl brGDGTs (with subscript a) are located in the left half of the RDA (Fig. 3.5a) and show a moderate to strong correlation with the second axis of the RDA, especially brGDGT IIIa which is highly negatively influenced by MAAT. However, except compound IIIa, all the brGDGTs involved in the calculation of the MBT'_{5Me} are more influenced by pH than by temperature (Fig. 3.5a), hence a rather moderate correlation between this index and MAAT ($R^2=0.35$; Fig. 5a) is observed. Therefore, in contrast with previous studies where 5-methyl brGDGT distribution in soils was suggested to be mainly affected by MAAT and to a lesser extent by soil pH (De Jonge et al., 2014; Naafs et al., 2017a), the relative abundance of the 5-methyl brGDGTs in the soils from the French Alps seems to be mainly driven by soil pH, with only compound IIIa showing a strong correlation with MAAT (Fig. 3.5a). This may be related to a recent study by De Jonge et al., (2019), which

highlighted the strong effect of pH on brGDGT distribution in addition to MAAT, with a different pH-dependency in soils from warm and cold environments. As for the 6-methyl brGDGTs, located in the right part of the RDA (Fig. 3.5a), their relative abundance in the soils of the French Alps appear well-correlated with pH, as observed in a large set of globally distributed soils ($n = 350$) by Naafs et al. (2017a). BrGDGT distribution is only weakly influenced by soil moisture (especially compound Ia) and soil texture in soils from the French Alps (Fig. 3.5a).

- **RDA analysis of 3-OH FA distribution and environmental variables**

Similarly to brGDGTs, RDA analysis was performed to investigate the effect of the different environmental parameters on 3-OH FA distribution (Fig. 3.5b). The first two axes of the RDA explain 31.8 % of the total inertia of the dataset. The explained fitted variation represents 81.2 % of the total inertia of the relationship between 3-OH FAs and the environmental variables (Table 3.2). The RDA analysis shows a predominant contribution of pH ($r = 0.95$), positively correlated with axis 1, on the distribution of 3-OH FAs (Fig. 3.5b, Table 3.2). Soil moisture ($r = -0.56$) MAAT ($r = 0.29$), sand content ($r = -0.21$) and clay content ($r = -0.32$) show a weaker correlation with the axis 1. Axis 2 is mainly related to the MAAT ($r = -0.93$), to the relative soil moisture ($r = -0.41$; Fig. 3.5b, Table 3) and to a lesser extent with pH ($r = 0.23$). Soil structure parameters (sand and clay contents) are poorly represented on the axis 2 of the RDA. (Fig. 3.5b; Table 3.2).

These qualitative trends are confirmed after quantification of the influence of each environmental variable on 3-OH FA distribution (Table 3.3). The study of simple effects also shows a predominant influence of pH on 3-OH FA distribution (20.0 %) followed by relative soil moisture (9.9 %), temperature (11.2 %) and particle size (3.8%). The combined influence of the different variables on the 3-OH FA relative abundance shows similar results, with pH (20.0 %) and to a lesser extent temperature (10.9 %) and soil moisture (3.6 %) being the main parameters affecting the lipid distribution (Table 3.3). Like brGDGTs, the influence of local parameters related to soil type and vegetation (i.e. soil pH, grain size content and soil moisture) on 3-OH FA distribution is predominant (26.1 %; Table 3.3) over MAAT, which could explain the absence of correlation between 3-OH FA distribution, especially RAN₁₅, RAN₁₇ indices and MAAT (Fig. 3.3b, c).

When considering the relative abundances of the individual 3-OH FAs in the RDA, it appears that most 3-OH FAs are grouped in a narrow sector of the circle (Fig. 3.5b). They encompass compounds with branched chains (*anteiso*-C₁₁, *iso*-C₁₂, *iso*-C₁₃, *anteiso*-C₁₃, *iso*-C₁₄, *anteiso*-C₁₅, *iso*-C₁₆, *iso*-C₁₇, *anteiso*-C₁₇), but also *normal* ones (C₁₅, C₁₇). This sector is close to the positive part of the first axis and thus these compounds are mostly influenced by pH. In contrast, this sector appears orthogonal to MAAT and opposite to soil moisture. Outside this sector, the *normal* C₁₄ 3-OH FA is negatively correlated with the first axis and the *iso* C₁₅ on the one hand and the *normal* C₁₀ and C₁₁ on the other hand are respectively positively and negatively correlated with the MAAT (Fig. 3.5b).

The predominant influence of pH on the relative abundance of all *iso*, *anteiso* and *normal* 3-OH FAs except the *iso* C₁₅, *normal* C₁₈ and *normal* C₁₀ homologues (Fig. 3.5b) explains the existence of a relationship between the RIAN index and pH (Fig. 3.2b). Nevertheless, this correlation remains moderate due to the rather strong and opposite influence of soil moisture (Table 3.3) and to the (positive or negative) correlation of some 3-OH FA with MAAT. The RDA analysis also shows that the compounds used for the calculation of the RAN₁₅ and RAN₁₇ indices (*anteiso* and *normal* C₁₅/C₁₇ homologues) are located in the pH sector, orthogonal to MAAT (Fig. 3.5b). This explains why there are no linear relations between these indices and MAAT in soils from the French Alps (Fig. 3.3b, c).

• **Influence of vegetation and soil type on 3-OH FA and brGDGT distribution**

As discussed above, the influence of local edaphic parameters (i.e. grain size, soil moisture, pH) related to soil type and vegetation on both brGDGT and 3-OH FA distribution was shown to be higher than that of MAAT. This is further confirmed by the distribution of the samples in the RDAs (Fig. 3.5 a, b), with the samples being distributed according to their affiliation to certain types of soils or vegetation communities. The fact that samples are grouped in clusters may partly be due to the fact that some of them are geographically proximal. But some of the samples, spatially distant, are also clustered in the RDA, for example, according to soil types : samples 20 to 25 (calcosols, between

1,540 m and 1,940 m a.s.l.), samples 5 to 7 and 17 to 19 (brunisol, between 1,920 m and 2,688 m a.s.l.), or vegetation compositions : samples 1 to 10 which are alpine meadow (between 2,531 and 2,700 m a.s.l.) (Fig. 3.5; Chapter 2, Table 2.2). The calcosols are distributed over the entire right-hand side of the two RDAs, as they are essentially differentiated based on their alkaline pH. The calcosols are also distributed along axis 2, driven by altitude and relative soil moisture. Thus, lowland calcosols appear in the bottom or upper right quadrants of the two RDAs respectively (Fig. 3.5a, b), whereas subgroups of high-altitude calcosols, reflecting mountainous and subalpine vegetation covers, can be observed at the bottom or upper right, respectively. This shows that calcosols can also be differentiated based on the plant cover patterns and altitude. In contrast with the calcosols, brunisols are distributed throughout the whole correlation circle. However, they can be subdivided into 3 distinct subgroups based on the vegetation cover and altitude: a first subgroup with intermediate altitude samples and mountainous vegetation, positively correlated with soil moisture; a second subgroup, with high altitude samples and meadow vegetation, negatively correlated with MAAT; a third subgroup of individuals, characterized by low altitude soils and lowland vegetation, positively correlated with MAAT. Altogether, the RDA results show that soil groups are discriminated according to their bacterial lipid distributions, which are strongly dependent on vegetation cover and soil types.

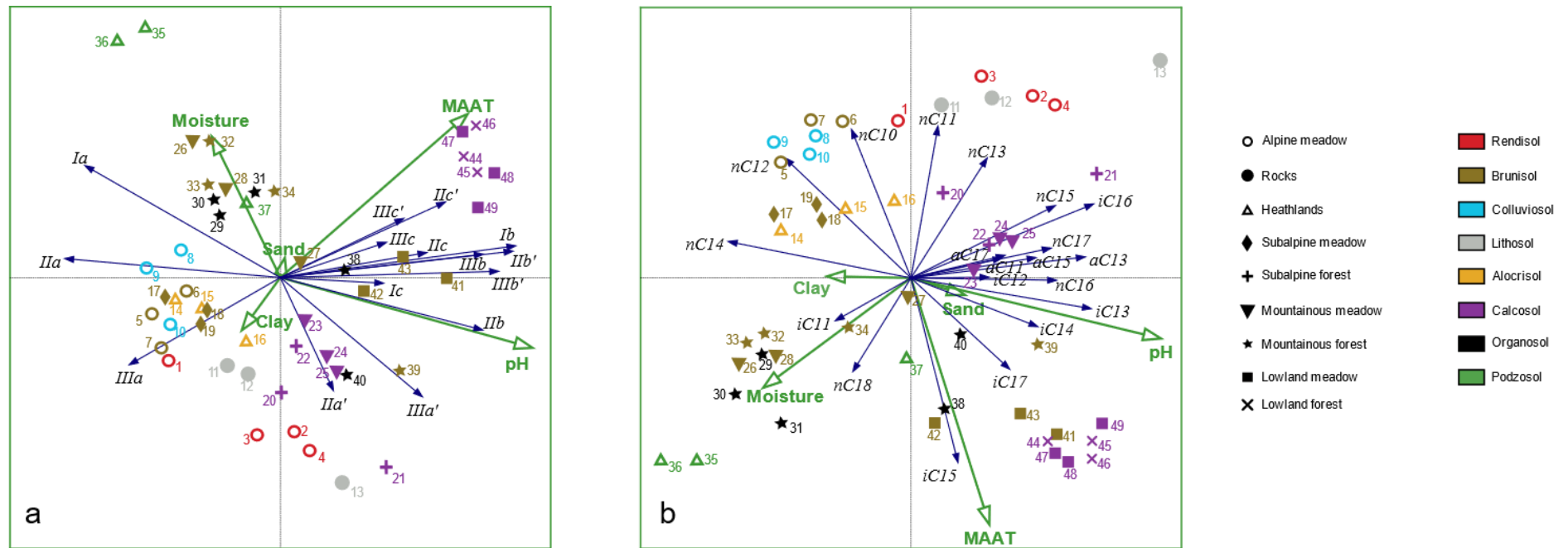


Figure 3.5 Triplot presenting the results of the RDA carried out on the relative abundances of (a) brGDGTs and (b) 3-OH FAs and the non-redundant physicochemical parameters: pH, MAAT, soil moisture, sand and clay contents.

3.1.3.3. *Development of local calibrations between the relative abundance of bacterial lipids and MAAT/pH*

The wide diversity of microenvironments encountered the French Alps could explain the absence or weakness of the relationships between MAAT and brGDGT-/3-OH FA-based indices, as well as between pH and RIAN index. This stresses the need for identifying the individual lipids whose fractional abundance is mainly correlated with MAAT or pH, which is possible through RDA analyses (Fig. 3.5). The identification of these individual lipids then allow the development of local relationships between environmental parameters (MAAT/pH) and the relative abundances (%) of the selected compounds (Figs. 3.6 and 3.7), as detailed below.

• BrGDGTs

Even though the MBT'_{5Me} index is not suitable for MAAT reconstruction in the French Alps ($R^2 = 0.35$; Fig. 3.3), the RDA analysis showed that the fractional abundance of some of the brGDGTs (mainly compound *IIIa* and to a lesser extent compounds *Iib'* and *Ib*) are moderately to strongly correlated with MAAT. An alternate transfer function with MAAT was defined based on a linear combination of the fractional abundance of these three compounds (Eq. 3.1; Fig.3.6):

$$\text{MAAT} = -0.255 \times [\text{IIIa}] + 0.195 \times [\text{Ib}] + 0.171 \times [\text{Iib}'] + 6.34 \quad (3.1)$$

$$(n = 49; R^2 = 0.69; \text{RMSE} = 2.26 \text{ } ^\circ\text{C})$$

This new calibration (Eq. 3.1; Fig. 3.6) shows a much higher determination coefficient and lower RMSE than the MBT'_{5Me}-MAAT relationship (Fig. 3.3a). The same approach was used to propose a new local calibration for the brGDGTs to reconstruct pH values, but this did not improve the linear relationship demonstrated between the CBT index and pH (Fig. 3.2a). Altogether, these results suggest that brGDGTs can be used for both pH and temperature reconstructions even in highly diverse and contrasted

environments such as the French Alps. Nevertheless, the present study shows that a detailed investigation of the environmental controls of brGDGT distribution and the development of local calibrations may be required to confidently use these compounds as paleoproxies.

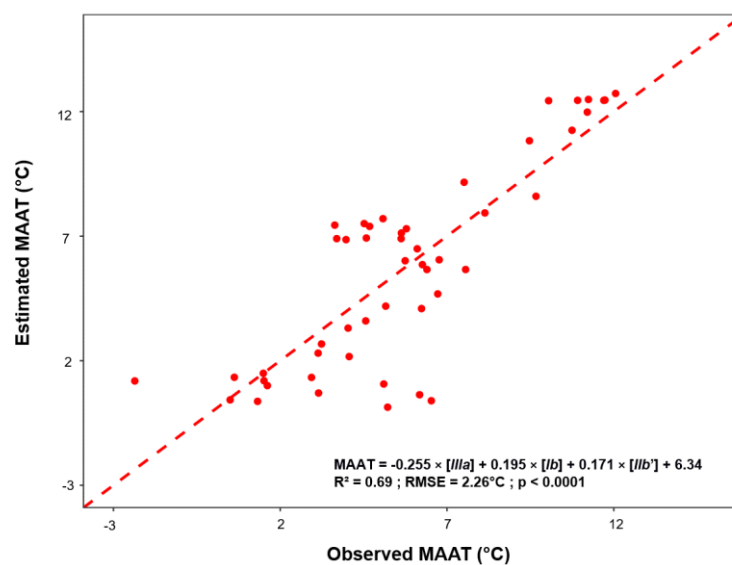


Figure 3.6 MAAT estimated using the new calibration based on a linear combination of brGDGT relative abundances (Eq. 3.1) vs. MAAT observed along the the Bauges and Lautaret-Galibier massifs. The roman numerals correspond to the different GDGT structures presented in De Jonge et al. (2014).

- 3-OH FAs

The RDA results allowed the identification of the individual 3-OH FAs whose relative abundance is mainly correlated with MAAT or pH (Fig. 3.5b). Two alternate transfer functions were proposed based on a linear combination of the relative abundance of 3-OH FAs (Eqs. 3.2 and 3.3), using a similar approach as for brGDGTs (Eq. 3.1):

$$pH = -0.144 \times [nC_{12}] + 0.177 \times [iC_{13}] - 0.171 \times [nC_{14}] + 0.148 \times [iC_{16}] + 10.01$$

($n = 49$; $R^2 = 0.70$; $RMSE = 0.6$) (3.2)

$$\text{MAAT} = -0.629 \times [nC_{10}] + 0.883 \times [iC_{15}] + 0.496 \times [iC_{17}] + 0.629 \times [nC_{18}] - 13.16$$

$$(n = 49; R^2 = 0.63; \text{RMSE} = 2.43 \text{ } ^\circ\text{C}) \quad (3.3)$$

The proposed relationship between pH and 3-OH FA relative abundances (Eq. 3.2; Fig 3.7a) presents a higher determination coefficient ($R^2 = 0.70$ vs. 0.43) and accuracy (RMSE = 0.60 vs. 0.83) than the one with the RIAN index. Similarly, the identification of the individual 3-OH FAs strongly influenced by MAAT allowed obtaining a new relationship with MAAT (Eq. 3.3; Fig. 3.7b) which was not visible through the previously defined RAN_{15} and RAN_{17} indices. This shows once again the interest and importance of constraining the environmental factors affecting the relative abundance of microbial lipids in complex and highly variable environments such as the French Alps, where conventional indices are not reliable for temperature and pH reconstructions. Specific calibrations have to be developed instead to take into account the effect of local confounding parameters on lipid relative distribution.

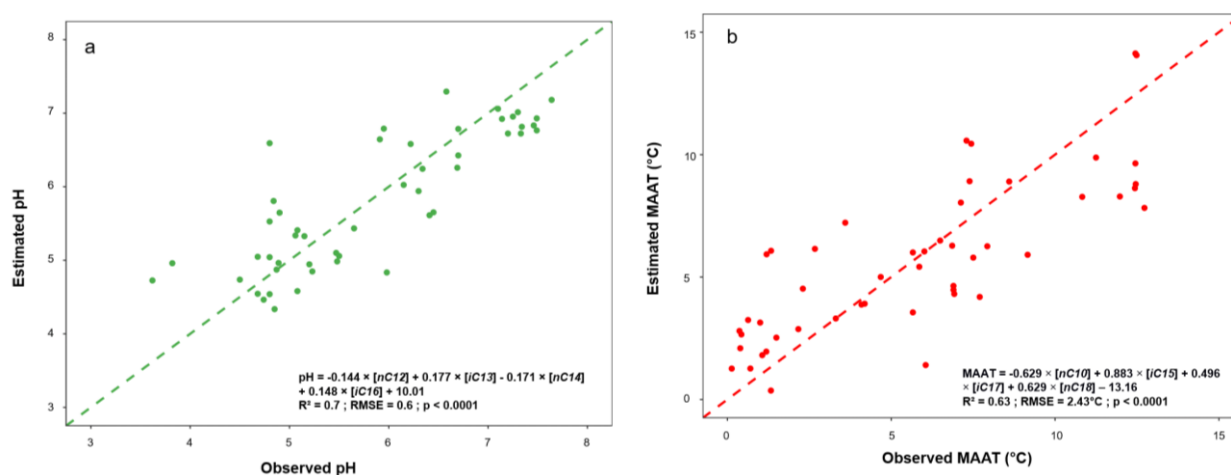


Figure 3.7 (a) pH estimated using the new calibration based on a linear combination of 3-OH FA relative abundances (Eq. 7) vs. pH observed along the Bauges and Lautaret-Galibier massifs. (b) MAAT estimated using the new calibration based on a linear combination of 3-OH relative abundances (Eq. 8) vs. MAAT observed along the Bauges and Lautaret-Galibier massifs. The roman numerals correspond to the different GDGT structures presented in De Jonge et al. (2014).

*3.1.3.4. Constraints on the applicability of brGDGTs and 3-OH FAs
as environmental proxies*

3-OH FAs were proposed only recently as potential temperature and pH proxies in soils and were to date investigated in a limited amount of samples – ca. 70 soils, excluding the presently studied ones (C. Wang et al., 2016; Huguet et al., 2019). This contrasts with brGDGTs, whose applicability as environmental proxies in terrestrial (and aquatic) environments has been extensively investigated over the last 15 years (e.g. Weijers et al., 2007; Peterse et al., 2012; Coffinet et al., 2014; De Jonge et al., 2014; Huguet et al., 2014; Dang et al., 2016; Lei et al., 2016; Naafs et al., 2017a; Dearing Crampton-Flood et al., 2020). Studies investigating the constraints on the applicability of 3-OH FAs as environmental proxies in terrestrial settings are thus needed. pH was shown to be the main factor influencing both brGDGT and 3-OH FA distribution in soils of the French Alps, even though RIAN shows a much more moderate correlation with a lower prediction interval (Fig. 3.2b; $R^2=0.43$, RMSE = 0.83) than CBT (Fig. 3.2a; $R^2 = 0.85$, RMSE = 0.42). The calibration between the linear combination of 3-OH FA relative abundances and pH proposed in this study (Eq. 3.2; Fig. 3.7a) is also slightly weaker ($R^2 = 0.70$) and less accurate (RMSE = 0.60) than the one with the CBT (Fig. 3.2a). Altogether, these results show the higher reliability of the latter index as a pH proxy in the French Alps. In contrast with pH, the MAAT was only moderately influencing the bacterial lipid distribution in soils reflected by the weak relationship between MBT'_{5Me} and MAAT and the absence of relationship of RAN₁₅ and RAN₁₇ with MAAT (Fig. 3.3). Nevertheless, the environmental variables investigated in the present study (including MAAT and pH) were not able to explain the total variability of the lipid distribution, the unexplained variation remaining higher for 3-OH FAs than for brGDGTs (Table 3.2). Seasonality might be one of the factors influencing the production of bacterial lipids in the French Alps soils, as previously suggested for brGDGTs in the French Jura Mountains (Huguet et al., 2013). Indeed, during part of the year, some of the French Alps soils are covered with snow, protecting them from extreme temperatures. Parameters such as microtopography or snowfall are known to have an important role on the thermal regime of soils and thus on the microorganisms living in them (Margesin et al., 2009). The effect

of the thermal regime could therefore explain part of the dispersion remaining in the local calibrations between MAAT and bacteria lipid distribution.

The lack of information regarding the ecology of the 3-OH FA and brGDGT producers should indeed be taken into account while discussing the high uncertainty related to the lipid-based transfer functions. C₁₀ to C₁₈ 3-OH FAs, used for the calculation of the RAN₁₅ and RAN₁₇ pH/temperature proxies, are known to be produced by Gram-negative bacteria, which are a highly diverse group of microorganisms represented by numerous genera. This non-monophyletic group of bacteria does not share a common ancestor (Lecointre and Guyader, 2006), which explains the large genetic and biochemical differences between the various Gram-negative bacteria. The numerous genera and subgenera of Gram-negative bacteria are also characterized by highly diverse lipid profiles, with different relative abundances of 3-OH FAs, and with the whole suite of C₁₀-C₁₈ 3-OH FAs homologues not being present in all the strains (Wilkinson et al., 1988). Therefore, the 3-OH FA lipid distribution is highly dependent upon what Gram-negative bacterial species are present (e.g. Parker et al., 1982; Bhat and Carlson, 1992; Zelles, 1999). Furthermore, the diversity of Gram-negative bacteria varies with altitude (Margesin et al., 2009; Siles and Margesin, 2016). Taken together, this may partly explain the lack of relationship or poor to moderate correlations observed between 3-OH FA indices (RAN_{15/17}, RIAN) and MAAT/soil pH (Figs. 3.2b, 3.3b, c). Additionally, it might also not be excluded that the various soil Gram-negative bacteria genera/subgenera respond differently to variations of environmental parameters. In brief, the 3-OH FA profile of the various Gram-negative bacteria species likely reflects the diversity of microenvironments present in the French Alps, which are subjected concomitantly to large variations in several environmental parameters (MAAT, soil pH, soil moisture, microtopography, vegetation cover, soil type etc...).

In contrast, although they appear to be ubiquitous, brGDGTs might be produced by a more restricted number of bacterial species than 3-OH FAs. Indeed, so far only soil *Acidobacteria* have been found to contain the building blocks of brGDGT biosynthesis and are thus considered as potential brGDGT sources (Sinninghe Damsté et al., 2018). If the diversity of brGDGT producers is indeed lower than that of 3-OH FA-producing microorganisms, it may explain the more homogenous response and lower scatter of the relationships between MAAT/pH and brGDGT-derived indices than those with 3-OH FA

indices (Figs. 3.2 and 3.3). Such relationships may also be affected by the different physiological responses of brGDGT and 3-OH FA producers to environmental parameters. These hypotheses remain purely speculative, as to date the information on brGDGT source microorganisms remains limited. This calls for further studies assessing and comparing the influence of environmental parameters on brGDGT and 3-OH FA producing bacteria in terrestrial environments. Such work is essential to improve the reliability and accuracy of the complementary temperature and pH proxies based on the two lipid families.

3.1.4. Conclusions

This study thoroughly investigated the environmental factors controlling the distribution of brGDGTs and 3-OH FAs in soils collected along well-documented altitudinal transects in the French Alps. The influence of local parameters (pH and to a lesser extent soil moisture and grain size, related to vegetation and soil types) on brGDGT and 3-OH FA was more important than MAAT. This likely explains the absence or weak relationships between MAAT and brGDGT/3-OH FA-based indices and stresses the need for identifying the individual lipids whose fractional abundance is correlated with MAAT or pH. Such work at the molecular level led to the development of strong local calibrations between the individual lipids identified through statistical analyses and MAAT/pH. They can be applied to soils from the French Alps, representative of highly contrasted microenvironments. The present study highlights the importance of constraining the environmental factors affecting the distribution of 3-OH FAs and brGDGTs in terrestrial settings prior to any environmental reconstruction using these lipid biomarkers. Such an approach should be reproduced in other sites, where local factors would also strongly influence the bacterial lipid distribution. Local effects are expected to explain a large part of the scatter observed in the global calibrations between MAAT/pH and 3-OH FA as well as brGDGT distribution. These local confounding parameters may notably explain why the calibration between 3-OH FA indices (RAN_{15/17}) and MAAT in the French Alps differ from those previously obtained in other sites. As 3-OH FAs and brGDGTs may be produced by different microorganisms, further work is needed to assess and compare the impact of environmental parameters, especially temperature and pH, on the microbial

diversity in contrasted soils and the associated lipid profiles. Even though the lack of knowledge on brGDGT-producing bacteria complicates such studies, the latter are essential to improve the reliability and accuracy of the complementary temperature and pH proxies based on 3-OH FAs and brGDGTs.

3.2. Etude de l'effet de la température sur la distribution des acides gras 3-hydroxylés et des tétraéthers bactériens au sein de sols des Alpes françaises en conditions contrôlées.

3.2.1. Introduction

Dans la précédente partie de ce troisième chapitre, présentée sous forme d'article, une des conclusions était qu'au sein des sols des Alpes françaises la température n'est pas le seul facteur influençant la distribution des lipides membranaires bactériens (acides gras 3-hydroxylés (3-AGH) et tétraéthers bactériens (brGDGT)). Une des hypothèses pour expliquer ceci est la grande diversité d'environnements présents au sein des Alpes françaises, et l'influence d'autres paramètres que la température sur la distribution des lipides, comme le pH, l'humidité, ou encore la granulométrie.

Afin de tester directement l'effet d'une variation de température sur l'abondance relative des acides gras 3-hydroxylés bactériens (C₁₀ à C₁₈) et des tétraéthers bactériens dans les sols, des incubations en conditions contrôlées ont été réalisées à quatre températures différentes (5°C, 15°C, 25°C, 35°C) pendant 1 an. De précédentes études ont mis en exergue le potentiel de telles expérimentations pour observer la vitesse d'adaptation, en fonction de la température, des organismes sources des tétraéthers bactériens et du renouvellement des lipides bactériens dans des échantillons de sols et de tourbières (Huguet et al., 2014, 2017; Peterse et al., 2015; Chen et al., 2018; Martínez-Sosa and Tierney, 2019).

Dans le cadre de notre étude, quatre sols contrastés en termes de types de végétation et de sols ont été prélevés pour réaliser ces expériences, comme décrit dans la partie matériel et méthodes (section 2.2). Ces sols sont BOU11, MOU32, FP88 (respectivement les échantillons 44, 39, 19 et 4 ; Tableaux 2.2. et 3.1.).

3.2.2. Résultats et discussion

3.2.2.1. *Viabilité des organismes au sein des microcosmes*

A cours de l'année d'incubation des quatre sols des Alpes françaises, la viabilité des bactéries Gram-négatif au sein des différents échantillons a été testée par numération de ces microorganismes. Pour ce faire, les bactéries Gram-négatif cultivables ont été

isolées à partir des quatre échantillons de sols choisis précédemment. L'intérêt de ces analyses est de vérifier qu'au cours des 12 mois d'incubation, il y a toujours des organismes vivants, notamment ceux produisant des acides gras 3-hydroxylés. En outre, de précédentes études ont proposés que les brGDGTs sont synthétisés par des Acidobactéries, qui sont des bactéries Gram-négatifs (Weijers et al., 2009a; Sinninghe Damsté et al., 2018 et références incluses). Pour les 4 échantillons de sols, la viabilité des microcosmes a été vérifiée juste après le prélèvement des sols, ainsi qu'après 1 mois, 6 mois, 9 mois et 12 mois d'incubation, aux différentes températures (5°C, 15°C, 25°C, 35°C ; section 2.2).

Le comptage des unités formant colonies (UFC) pour les différents échantillons (Fig. 3.8) ne montre globalement pas de tendance nette à la diminution des cellules de bactéries Gram-négatif cultivables au cours des 12 mois d'incubation, quelle que soit la température. Cette observation révèle la présence de bactéries Gram-négatif viables, même après un an d'incubation en conditions contrôlées. Cette première conclusion confirme la présence d'organismes vivants, et nous permet dans la suite de cette étude de suivre et comparer l'évolution des concentrations et abondances relatives des acides gras 3-hydroxylés et des tetraéthers bactériens au cours de l'expérimentation en fonction des différentes températures d'incubation.

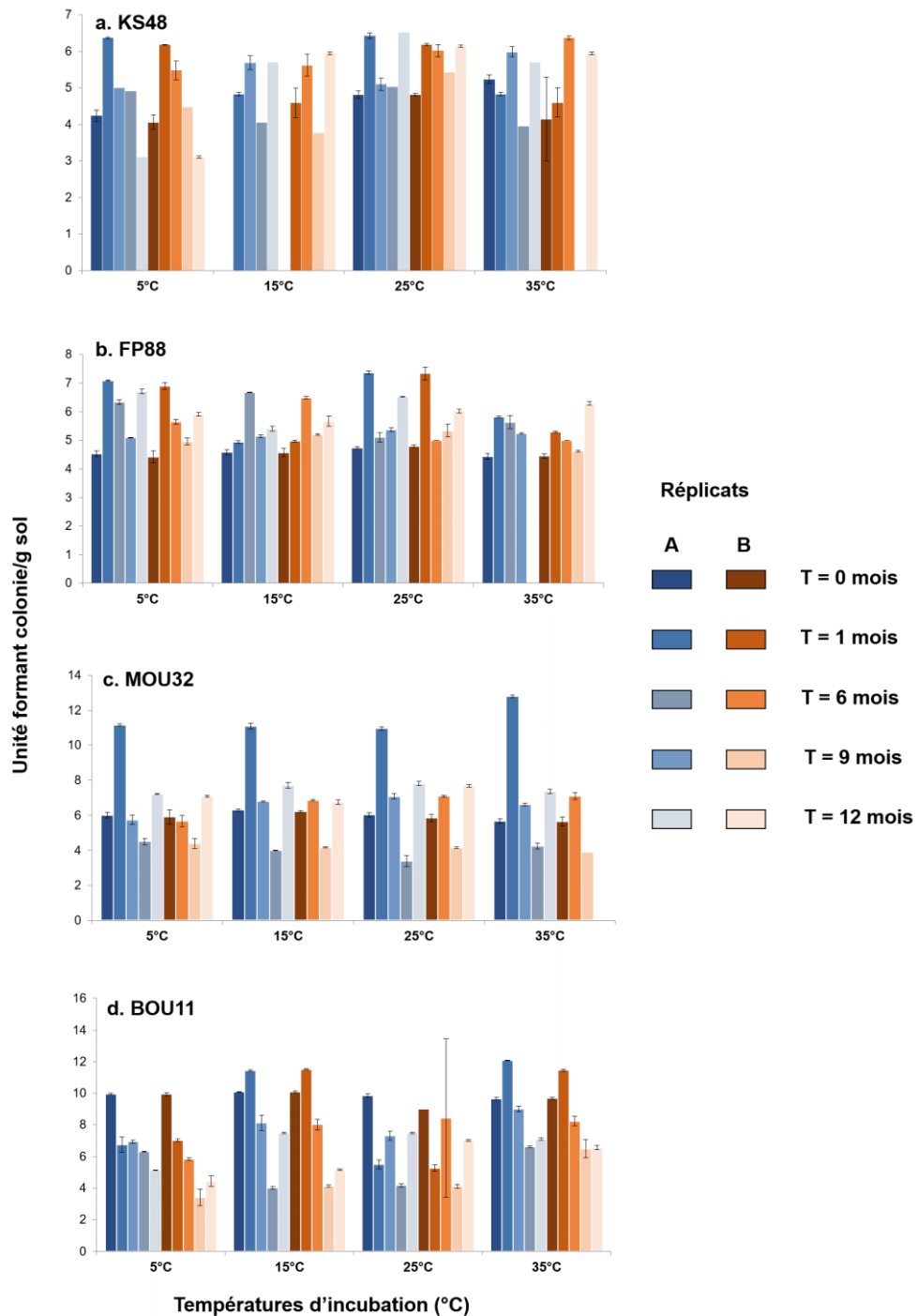


Figure 3.8 Données de viabilité des bactéries Gram-négatif dans les 4 échantillons incubés aux différentes températures et pour les différents pas de temps (0, 1, 6, 9 et 12 mois). Les concentrations d'unités formant des colonies sont exprimées en utilisant une notation logarithmique, où la valeur indiquée est le logarithme en base 10 de la concentration.

3.2.2.2. Concentrations et abondances relatives des lipides bactériens

• Les acides gras 3-hydroxylés (3-AGH)

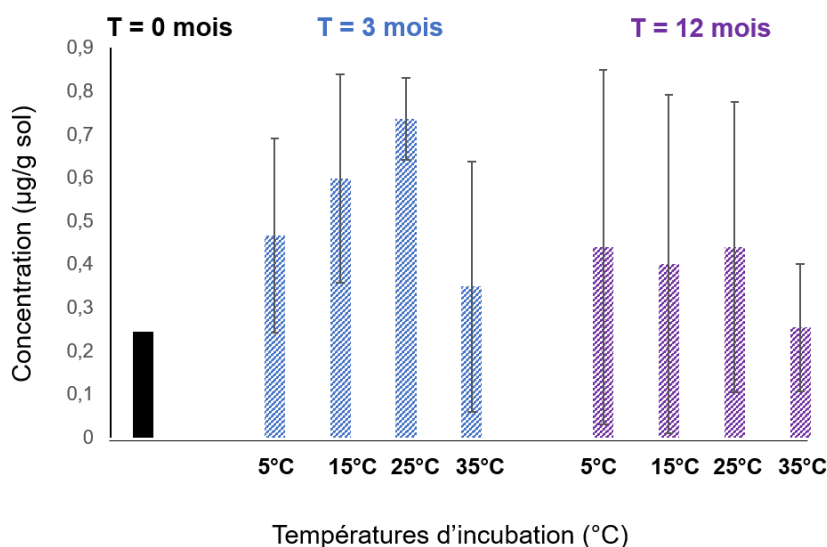


Figure 3.9 Concentrations des 3-AGH (µg/g sol sec) au sein de l'échantillon FP88 aux différents pas de temps et températures d'incubation considérées. Après 3 et 12 mois d'incubation les moyennes de deux réplicats avec les écarts types associés sont montrées.

Les 3-AGH ont été analysés dans les quatre échantillons de sols précédemment cités au temps initial puis après 3 et 12 mois d'incubation, pour les différentes températures. Il n'a pas été observé une évolution notable de la concentration totale en 3-AGH au cours de l'année d'incubation, quelle que soit la température d'incubation et l'échantillon considéré (annexe 1 ; Tableau supplémentaire 3.3), comme montré à titre d'exemple pour le sol FP88 (Fig. 3.9). Cela suggère que le changement de température n'a pas eu d'effet majeur (i.e. production ou dégradation) sur la quantité totale des 3-AGH dans les échantillons de sols au cours de cette année d'incubation (annexe 1 ; Tableau supplémentaire 3.3).

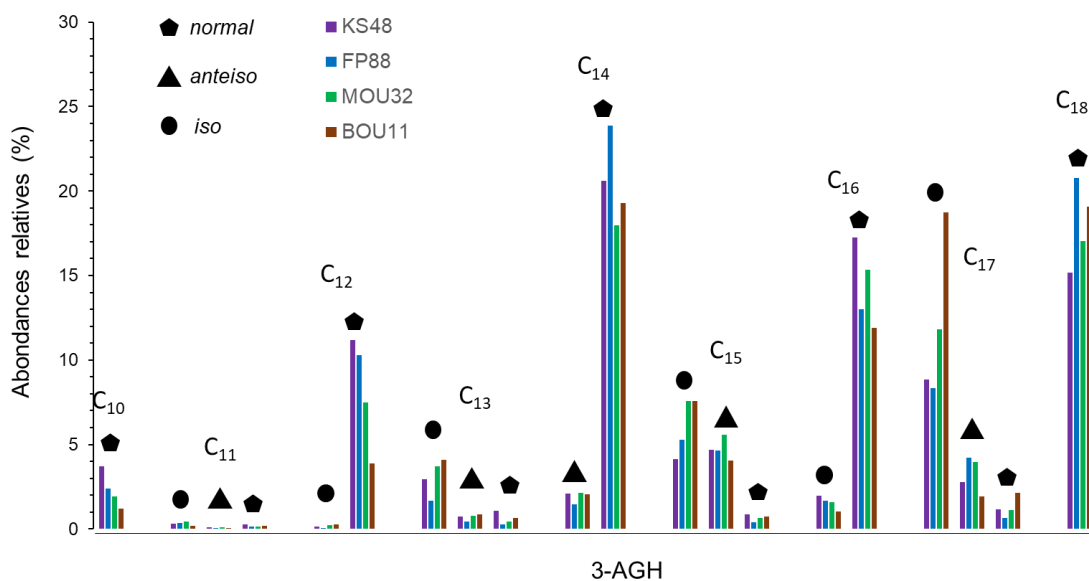


Figure 3.10 Abondances relatives des 3-AGH pour les échantillons initiaux ($t = 0$ mois) KS48, FP88, MOU32 et BOU11.

L'abondance relative des 3-AGH individuels C_{10} à C_{18} a été déterminée dans les 4 échantillons au temps initial (Fig 3.10), ainsi qu'après 3 et 12 mois d'incubation et aux différentes températures. Les échantillons initiaux présentent une prédominance des homologues à chaînes normales ($66,1 \pm 6,4\%$) par rapport aux 3-AGH *iso* ($25,2 \pm 7,0\%$) et *anteiso* ($11,1 \pm 2,0\%$) (Tableau 3.4). Ces résultats sont en bon accord avec ceux obtenus pour les 49 sols alpins (voir section 3.1.). On remarque cependant que le degré de ramification des 3-AGH, comme défini dans l'équation 3.4, diffère en fonction de l'échantillon considéré :

$$\text{Degré de ramification} = [Iso + Anteiso] / [Normal] \quad (\text{Eq.3.4})$$

Celui-ci est plus élevé pour les échantillons prélevés à faible altitude (MOU32 et BOU11, avec des valeurs de 0,60 et 0,69 respectivement) par rapport à ceux provenant d'altitudes élevées (KS48 et FP88, avec un rapport de 0,40 et 0,39 respectivement). Cette observation suggère une augmentation de la proportion des homologues à chaînes normales par rapport aux composés ramifiés lorsque l'altitude augmente. Cette tendance est également visible au sein des 49 échantillons de sols prélevés au sein des Alpes

françaises (Fig. 3.11) et pourrait représenter un mécanisme d'adaptation à la température des microorganismes produisant des 3-AGH (Ernst et al., 2016).

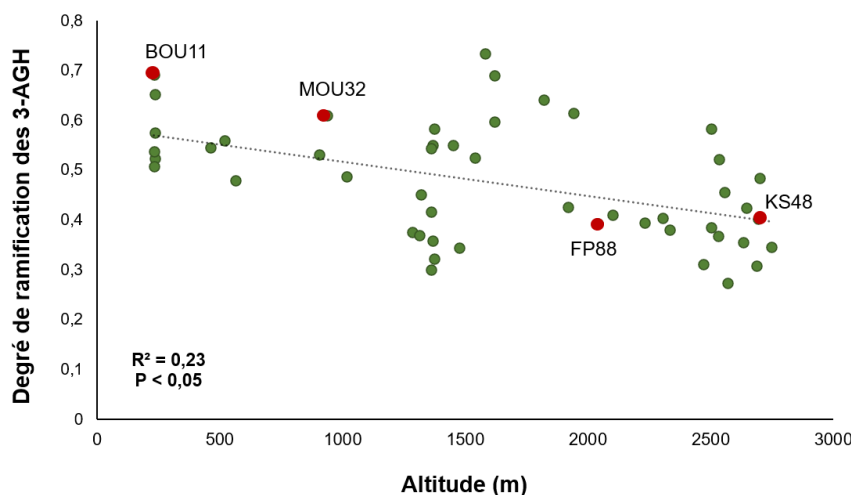


Figure 3.11 Degré de ramification des 3-AGH mesurés au sein des 49 échantillons de sols des Alpes françaises en fonction de l'altitude. L'écart type est compris dans la taille du point sur cette figure.

L'abondance relative des 3-AGH n'a pas été modifiée de manière notable au cours des incubations, quels que soient la température et l'échantillon de sol considérés (Fig. 3.13, Tableau 3.4). Néanmoins, pour tous les échantillons sauf KS48, au bout de 12 mois, le degré de ramification est plus faible à 35 °C qu'aux 3 autres températures. Ceci est particulièrement visible pour les sols MOU32 et BOU11 (Tableau 3.4 ; Fig 3.12). Ainsi, dans le cas de l'échantillon BOU11, la diminution du degré de ramification en fonction de l'augmentation de la température est visible dès le troisième mois d'expérience en microcosme. La même tendance est observable au bout de 12 mois d'incubation, avec un degré de ramification plus faible à 35 °C qu'aux 3 autres températures d'incubation (Tableau 3.4).

Ces tendances sont en opposition avec la tendance observée sur les 49 échantillons de sols au pas de temps initial (diminution du degré de ramification quand la température diminue). Néanmoins, la diminution du degré de ramification observée dans les microcosmes lors de l'augmentation de la température au cours du temps est en accord

avec la littérature. En effet, différentes études ont montré qu'une diminution de la température du milieu de vie des microorganismes entraîne en réaction la biosynthèse de lipides membranaires à chaînes plus courtes, non saturées et/ou ramifiées permettant d'augmenter la fluidité de la membrane (Denich et al., 2003; Haque and Russell, 2004; Ernst et al., 2016). La tendance inverse observée dans les 49 échantillons de sols des Alpes françaises pourrait être liée au fait que la diversité des bactéries Gram-négatif varie avec l'altitude, comme suggéré par de précédentes études (Margesin et al., 2009; Siles and Margesin, 2016), sans qu'on puisse exclure que ces différentes communautés bactériennes ne présentent pas le même mécanisme d'adaptation à la température.

Ainsi, au bout de 12 mois d'incubations à 35°C, il semble y avoir une évolution du degré de ramification des 3-AGH, vraisemblablement en accord avec un mécanisme d'adaptation homéo visqueux. Afin d'infirmier ou de confirmer cette tendance, il serait nécessaire d'analyser les pas de temps restants, voire de refaire ces expériences d'incubations avec des triplicats afin de pouvoir réaliser des tests statistiques.

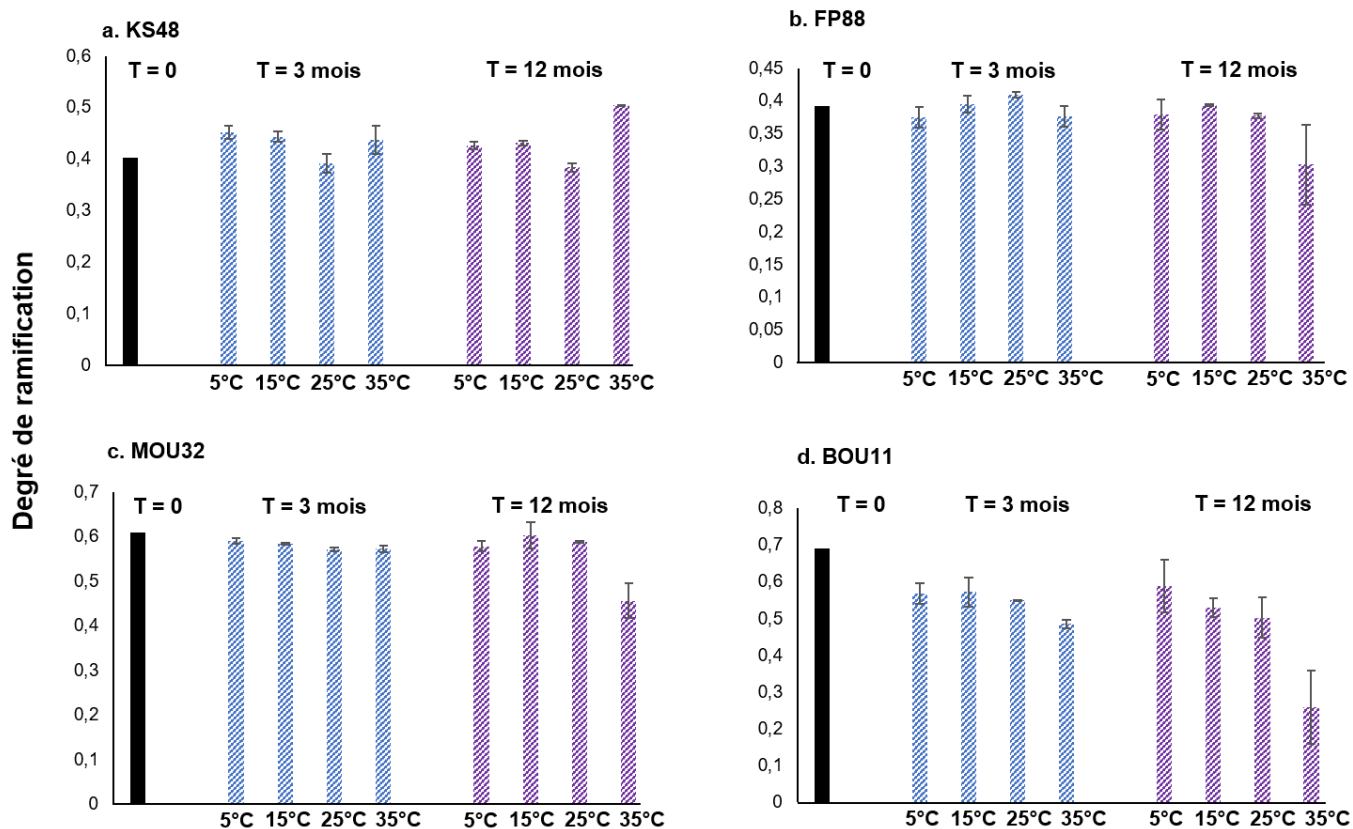


Figure 3.12 Degré de ramification des acides gras 3-hydroxylés (Eq 1) aux différentes températures d'incubation et aux différents pas de temps pour les échantillons a) KS48, b) FP88, c) MOU32 et d) BOU11. Les résultats présentés correspondent à la moyenne sur les deux réplicats.

ÉTUDE DE MARQUEURS LIPIDIQUES BACTÉRIENS AU SEIN DE SOLS DES ALPES FRANÇAISES

	KS 48				FP88				MOU32				BOU11			
	<i>Iso</i>	<i>Anteiso</i>	<i>Normal</i>	Degré de ramification	<i>Iso</i>	<i>Anteiso</i>	<i>Normal</i>	Degré de ramification	<i>Iso</i>	<i>Anteiso</i>	<i>Normal</i>	Degré de ramification	<i>Iso</i>	<i>Anteiso</i>	<i>Normal</i>	Degré de ramification
T= 0 mois	20.41	12.02	71.28	0.40	18.85	11.74	71.81	0.39	27.47	12.31	62.12	0.61	33.97	8.14	59.12	0.69
5°C	22.37 <i>1.40</i>	8.74 <i>0.76</i>	68.89 <i>0.64</i>	0.45 <i>0.01</i>	18.14 <i>1.43</i>	9.11 <i>0.58</i>	72.75 <i>0.85</i>	0.37 <i>0.02</i>	28.04 <i>0.66</i>	9.09 <i>0.91</i>	62.87 <i>0.25</i>	0.59 <i>0.01</i>	27.39 <i>0.72</i>	8.84 <i>1.86</i>	63.77 <i>1.14</i>	0.57 <i>0.03</i>
15°C	22.58 <i>1.00</i>	8.12 <i>0.49</i>	69.30 <i>0.51</i>	0.44 <i>0.01</i>	18.76 <i>0.99</i>	9.54 <i>0.30</i>	71.70 <i>0.69</i>	0.39 <i>0.01</i>	27.60 <i>0.55</i>	9.28 <i>0.49</i>	63.13 <i>0.07</i>	0.58 <i>0.00</i>	28.97 <i>3.04</i>	7.40 <i>1.46</i>	63.63 <i>1.58</i>	0.57 <i>0.04</i>
25°C	19.47 <i>0.91</i>	8.66 <i>0.05</i>	71.87 <i>0.96</i>	0.39 <i>0.02</i>	19.71 <i>0.10</i>	9.33 <i>0.31</i>	70.97 <i>0.21</i>	0.41 <i>0.00</i>	27.76 <i>0.45</i>	8.58 <i>0.26</i>	63.66 <i>0.19</i>	0.57 <i>0.00</i>	25.92 <i>0.95</i>	9.58 <i>0.98</i>	64.50 <i>0.03</i>	0.55 <i>0.00</i>
35°C	21.59 <i>1.90</i>	8.81 <i>0.55</i>	69.60 <i>1.36</i>	0.44 <i>0.03</i>	18.81 <i>0.03</i>	8.51 <i>0.80</i>	72.67 <i>0.83</i>	0.38 <i>0.02</i>	27.68 <i>0.25</i>	8.74 <i>0.54</i>	63.58 <i>0.30</i>	0.57 <i>0.01</i>	25.69 <i>0.30</i>	6.96 <i>0.21</i>	67.35 <i>0.51</i>	0.48 <i>0.01</i>
Moyenne	21.50 <i>1.42</i>	8.59 <i>0.31</i>	69.91 <i>1.33</i>	0.43 <i>0.03</i>	18.86 <i>0.65</i>	9.12 <i>0.44</i>	72.02 <i>0.85</i>	0.39 <i>0.02</i>	27.77 <i>0.19</i>	8.92 <i>0.32</i>	63.31 <i>0.38</i>	0.58 <i>0.01</i>	26.99 <i>1.52</i>	8.19 <i>1.22</i>	64.81 <i>1.73</i>	0.54 <i>0.04</i>
5°C	20.96 <i>0.97</i>	8.89 <i>0.60</i>	70.15 <i>0.37</i>	0.43 <i>0.01</i>	18.59 <i>1.68</i>	8.86 <i>0.48</i>	72.55 <i>1.20</i>	0.38 <i>0.02</i>	28.30 <i>0.25</i>	8.35 <i>0.18</i>	63.36 <i>0.44</i>	0.58 <i>0.01</i>	30.14 <i>3.65</i>	6.82 <i>0.77</i>	63.03 <i>2.88</i>	0.59 <i>0.07</i>
15°C	21.57 <i>0.36</i>	8.55 <i>0.56</i>	69.88 <i>0.20</i>	0.43 <i>0.00</i>	19.39 <i>0.66</i>	8.85 <i>0.59</i>	71.75 <i>0.07</i>	0.39 <i>0.00</i>	27.87 <i>1.06</i>	9.76 <i>0.09</i>	62.37 <i>1.16</i>	0.60 <i>0.03</i>	26.17 <i>1.66</i>	8.41 <i>0.55</i>	65.41 <i>1.11</i>	0.53 <i>0.03</i>
25°C	19.18 <i>0.49</i>	8.56 <i>0.06</i>	72.26 <i>0.43</i>	0.38 <i>0.01</i>	18.96 <i>0.55</i>	8.44 <i>0.38</i>	72.60 <i>0.17</i>	0.38 <i>0.00</i>	27.45 <i>1.03</i>	9.59 <i>1.09</i>	62.96 <i>0.06</i>	0.59 <i>0.00</i>	25.77 <i>0.12</i>	7.63 <i>2.55</i>	66.60 <i>2.43</i>	0.50 <i>0.05</i>
35°C	24.10 <i>0.04</i>	9.43 <i>0.07</i>	66.47 <i>0.03</i>	0.50 <i>0.00</i>	15.09 <i>2.92</i>	8.05 <i>0.70</i>	76.86 <i>3.61</i>	0.30 <i>0.06</i>	22.60 <i>1.23</i>	8.71 <i>0.62</i>	68.69 <i>1.84</i>	0.46 <i>0.04</i>	15.85 <i>4.06</i>	4.52 <i>2.22</i>	79.62 <i>6.28</i>	0.26 <i>0.10</i>
Moyenne	21.45 <i>2.04</i>	8.86 <i>0.41</i>	69.69 <i>2.39</i>	0.44 <i>0.05</i>	18.01 <i>1.97</i>	8.55 <i>0.39</i>	73.44 <i>2.31</i>	0.36 <i>0.04</i>	26.55 <i>2.66</i>	9.10 <i>0.68</i>	64.35 <i>2.93</i>	0.56 <i>0.07</i>	24.48 <i>6.08</i>	6.85 <i>1.68</i>	68.67 <i>7.45</i>	0.47 <i>0.14</i>

Tableau 3.4 Proportion des isomères d'acides gras 3-hydroxylés (en %) et degré de ramification pour les échantillons BOU11, FP88, MOU32 et KS48 aux différents pas de temps et températures d'incubation considérés. Les écart-types entre réplicats sont indiqués en italique.

ETUDE DE MARQUEURS LIPIDIQUES BACTERIENS AU SEIN DE SOLS DES ALPES FRANÇAISES

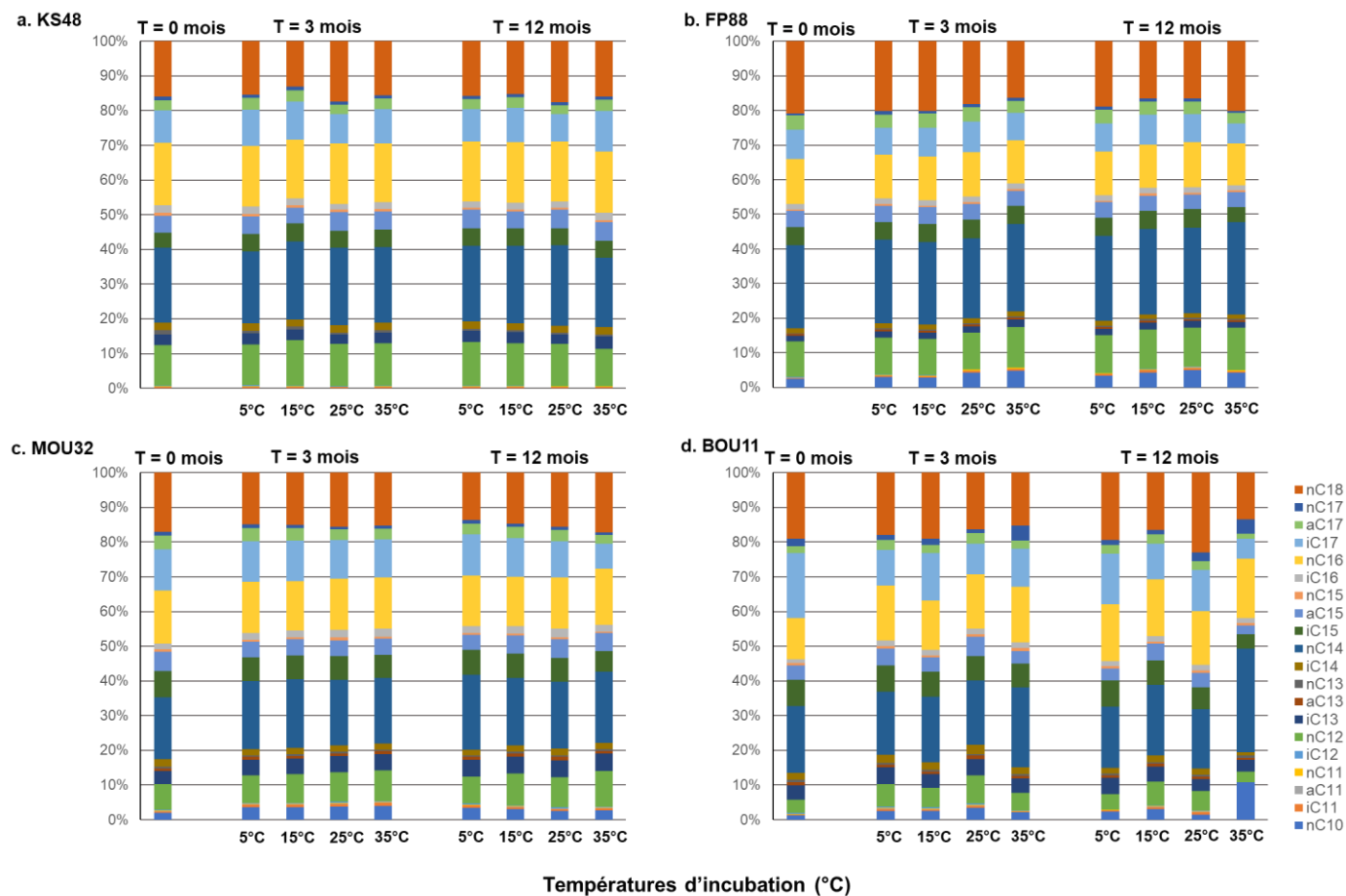


Figure 3.13 Abondances relatives des acides gras 3-hydroxylés aux différentes températures d'incubation et aux différents pas de temps pour les échantillons a) KS48, b) FP88, c) MOU32 et d) BOU11. Les résultats présentés correspondent à la moyenne des abondances relatives sur les deux répliqués.

• *Les tétraéthers bactériens (brGDGT)*

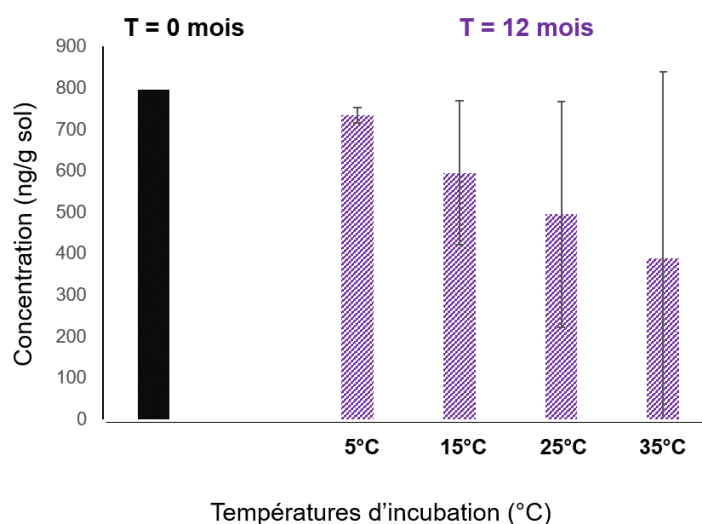


Figure 3.14 Concentrations des tétraéthers bactériens (ng/g sol sec) dans l'échantillon FP88 aux différents pas de temps et températures d'incubation considérés. A 3 mois sont montrées les moyennes des deux réplicats avec les écarts types associés.

Les concentrations totales en tétraéthers bactériens ont été calculées au sein des quatre échantillons de sols, pour les différentes modalités d'incubation (Fig. 3.14). Ces dernières ne montrent pas d'évolution majeure au cours du temps, quels que soient la température et l'échantillon considérés (annexe 1 ; Tableau supplémentaire 3.4), comme décrit pour l'échantillon FP88 présenté en exemple (Fig. 3.14). Cela suggère qu'il n'y a pas eu d'effet majeur (i.e. production ou dégradation) d'un changement de température sur la quantité totale de tétraéthers bactériens dans les échantillons de sols au cours de cette année d'incubation, comme également observé pour les 3-AGH (annexe 1 ; Tableau supplémentaire 3.3).

En parallèle des concentrations totales, les abondances relatives des différents tétraéthers bactériens ont été déterminées au sein des différents échantillons de sols (Fig. 3.15 ; Tableau 3.5 ; annexe 1 ; Tableau supplémentaire 3.4). Les échantillons KS48 et FP88 montrent une prédominance des homologues acycliques tetraméthylés (*Ia*) et pentaméthylés (*Ila*), qui représentent plus de 75 % des tétraéthers bactériens (Fig. 3.15 ; Tableau 3.5 ; annexe 1 ; Tableau supplémentaire 3.4). Par ailleurs, dans ces échantillons,

ÉTUDE DE MARQUEURS LIPIDIQUES BACTÉRIENS AU SEIN DE SOLS DES ALPES FRANÇAISES

les isomères 5-méthyle sont plus abondants que les composés 6-méthyle (63,1% et 84,1% respectivement). En revanche, dans les échantillons MOU32 et BOU11, les homologues *Ila'*, *Ilb'* et *IIla'* sont les plus abondants, avec une prédominance des composés acycliques (ca. 60% ; Fig. 3.15 ; Tableau 3.5; annexe 1 ; Tableau supplémentaire 3.4). Les isomères 5-méthyle sont présents en proportion plus faible au sein de ces deux échantillons (38, 2 et 55,9 %, respectivement) par rapport aux sols KS48 et FP88 (Fig. 3.15; annexe 1 ; Tableau supplémentaire 3.4). Les différences d'abondances relatives au sein de ces 2 sous-groupes d'échantillons, prélevés à haute (KS48 et FP88 - 2695 m et 2041 m) et basse altitude (MOU32 et BOU11 - 938 et 232 m) peuvent être expliquées par les conditions environnementales contrastées (en particulier la température) dans lesquelles se trouvent ces sols, influençant la composition lipidique des microorganismes sources des tétraéthers et/ou la diversité bactérienne.

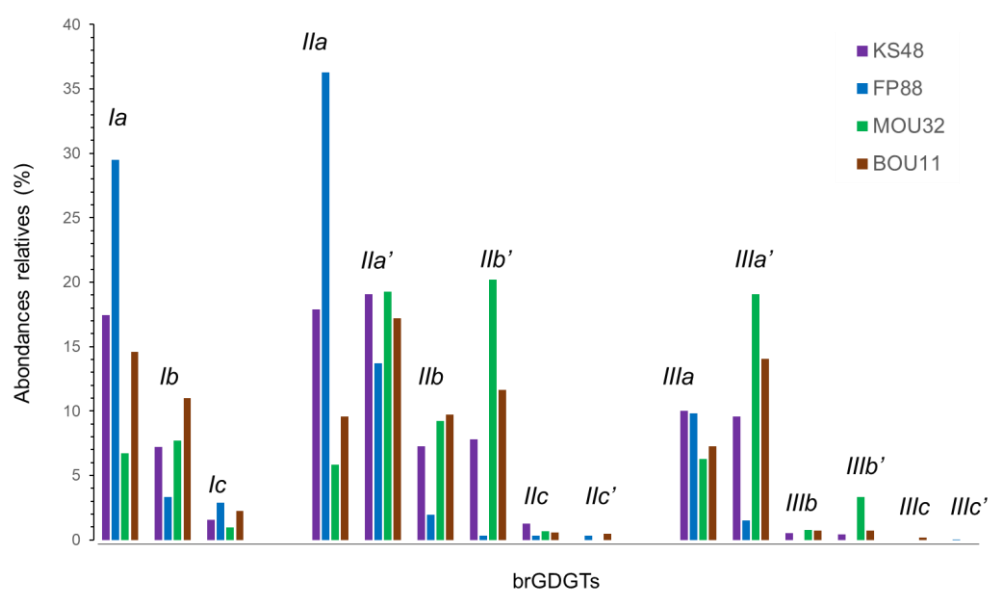


Figure 3.15 Abondances relatives des brGDGTs au pas de temps initial pour les échantillons KS48, FP88, MOU32 et BOU11.

Après 12 mois d'incubation, il est observé une augmentation de la proportion relative des composés acycliques tétraméthylés *Ia* au sein des 4 échantillons incubés à 35°C (Fig. 3.16; Tableau 3.5). Il est aussi noté une augmentation relative du composé *Ib* (un cycle pentane) dans le cas de l'échantillon BOU11 placé à 35°C pendant 12 mois. Ces changements de distribution se reflètent au niveau de l'indice de méthylation

MBT'_{5Me}, qui augmente significativement (p-value < 0,05 sur le test de Student pour échantillons appariés) après 12 mois d'incubation à 35°C pour 3 des 4 (sauf MOU32) échantillons considérés (Fig. 3.17). Néanmoins ces résultats sont à interprétés avec précaution, car il n'y a que deux réplicats considérés dans le test de student. Un minimum de 3 échantillons (triplicat) est requis pour que ce test soit fiable. Ces résultats sont plus précisément discutés dans la suite de ce chapitre.

ÉTUDE DE MARQUEURS LIPIDIQUES BACTÉRIENS AU SEIN DE SOLS DES ALPES FRANÇAISES

		Acyclique	Tetraméthylés	Pentaméthylés	Hexaméthylés	5-méthyle	6-méthyle
KS48	T= 0 mois	74.01	26.22	53.28	20.51	63.15	36.85
	5°C	85.10 <i>1.47</i>	28.75 <i>1.91</i>	54.15 <i>0.45</i>	17.10 <i>2.35</i>	69.30 <i>2.95</i>	30.70 <i>2.95</i>
		15°C	84.59 <i>0.64</i>	28.34 <i>0.79</i>	55.07 <i>0.87</i>	16.60 <i>1.66</i>	70.93 <i>0.86</i>
	T= 3 mois	86.98 <i>0.94</i>	33.92 <i>1.73</i>	51.72 <i>0.41</i>	14.36 <i>1.32</i>	76.65 <i>0.97</i>	23.35 <i>0.97</i>
	35°C	86.68 <i>0.78</i>	45.38 <i>4.82</i>	41.70 <i>3.36</i>	12.93 <i>1.46</i>	77.30 <i>3.90</i>	22.70 <i>3.90</i>
	Moyenne	85.83 <i>1.17</i>	34.10 <i>7.94</i>	50.66 <i>6.14</i>	15.24 <i>1.95</i>	73.54 <i>4.02</i>	26.46 <i>4.02</i>
FP88	T= 0 mois	90.83	35.68	52.98	11.34	84.10	15.90
	5°C	88.42 <i>1.86</i>	38.21 <i>0.61</i>	49.33 <i>0.67</i>	12.46 <i>0.06</i>	84.42 <i>2.09</i>	15.58 <i>2.09</i>
		15°C	89.00 <i>0.11</i>	40.02 <i>0.08</i>	48.59 <i>0.42</i>	11.39 <i>0.34</i>	87.08 <i>1.06</i>
	T= 3 mois	88.60 <i>0.35</i>	40.68 <i>1.56</i>	48.47 <i>1.54</i>	10.85 <i>0.03</i>	87.51 <i>1.25</i>	12.49 <i>1.25</i>
	35°C	88.95 <i>0.32</i>	45.93 <i>2.39</i>	44.68 <i>1.28</i>	9.39 <i>1.11</i>	88.07 <i>0.96</i>	11.93 <i>0.96</i>
	Moyenne	88.74 <i>0.28</i>	41.21 <i>3.32</i>	47.77 <i>2.09</i>	11.02 <i>1.28</i>	86.77 <i>1.62</i>	13.23 <i>1.62</i>
MOU32	T= 0 mois	57.17	15.40	55.17	29.43	38.15	61.85
	5°C	60.54 <i>0.58</i>	16.78 <i>0.23</i>	53.49 <i>0.52</i>	29.73 <i>0.75</i>	35.86 <i>1.91</i>	64.14 <i>1.91</i>
		15°C	51.68 <i>14.42</i>	24.00 <i>9.46</i>	53.44 <i>0.98</i>	22.56 <i>10.43</i>	45.51 <i>13.35</i>
	T= 3 mois	48.29 <i>14.07</i>	24.87 <i>9.90</i>	53.70 <i>0.06</i>	21.43 <i>9.84</i>	45.57 <i>11.61</i>	54.43 <i>11.61</i>
	35°C	59.53 <i>2.05</i>	24.83 <i>0.20</i>	50.04 <i>0.71</i>	25.13 <i>0.91</i>	42.95 <i>0.95</i>	57.05 <i>0.95</i>
	Moyenne	55.01 <i>5.98</i>	22.62 <i>3.91</i>	52.67 <i>1.76</i>	24.71 <i>3.69</i>	42.47 <i>4.57</i>	57.53 <i>4.57</i>
BOU11	T= 0 mois	62.65	27.82	49.22	22.96	55.91	44.09
	5°C	40.91 <i>0.23</i>	29.75 <i>0.49</i>	54.89 <i>0.10</i>	15.36 <i>0.39</i>	51.03 <i>0.11</i>	48.97 <i>0.11</i>
		15°C	46.63 <i>8.30</i>	26.69 <i>4.56</i>	53.92 <i>1.07</i>	19.40 <i>5.63</i>	48.03 <i>5.45</i>
	T= 3 mois	46.91 <i>13.47</i>	23.95 <i>8.26</i>	55.31 <i>1.03</i>	20.74 <i>9.29</i>	45.25 <i>8.34</i>	54.75 <i>8.34</i>
	35°C	42.28 <i>1.44</i>	43.28 <i>6.74</i>	44.70 <i>5.68</i>	12.03 <i>1.05</i>	60.28 <i>2.98</i>	39.72 <i>2.98</i>
	Moyenne	44.18 <i>3.04</i>	30.92 <i>8.58</i>	52.20 <i>5.04</i>	16.88 <i>3.96</i>	51.15 <i>6.53</i>	48.85 <i>6.53</i>

Tableau 3.5 Abondances relatives des tetraéthers bactériens individuels (en %) pour les échantillons BOU11, FP88, MOU32 et KS48 aux différents pas de temps et températures d'incubation considérés. Les écart-types entre réplicats sont indiqués en italiques.

ÉTUDE DE MARQUEURS LIPIDIQUES BACTÉRIENS AU SEIN DE SOLS DES ALPES FRANÇAISES

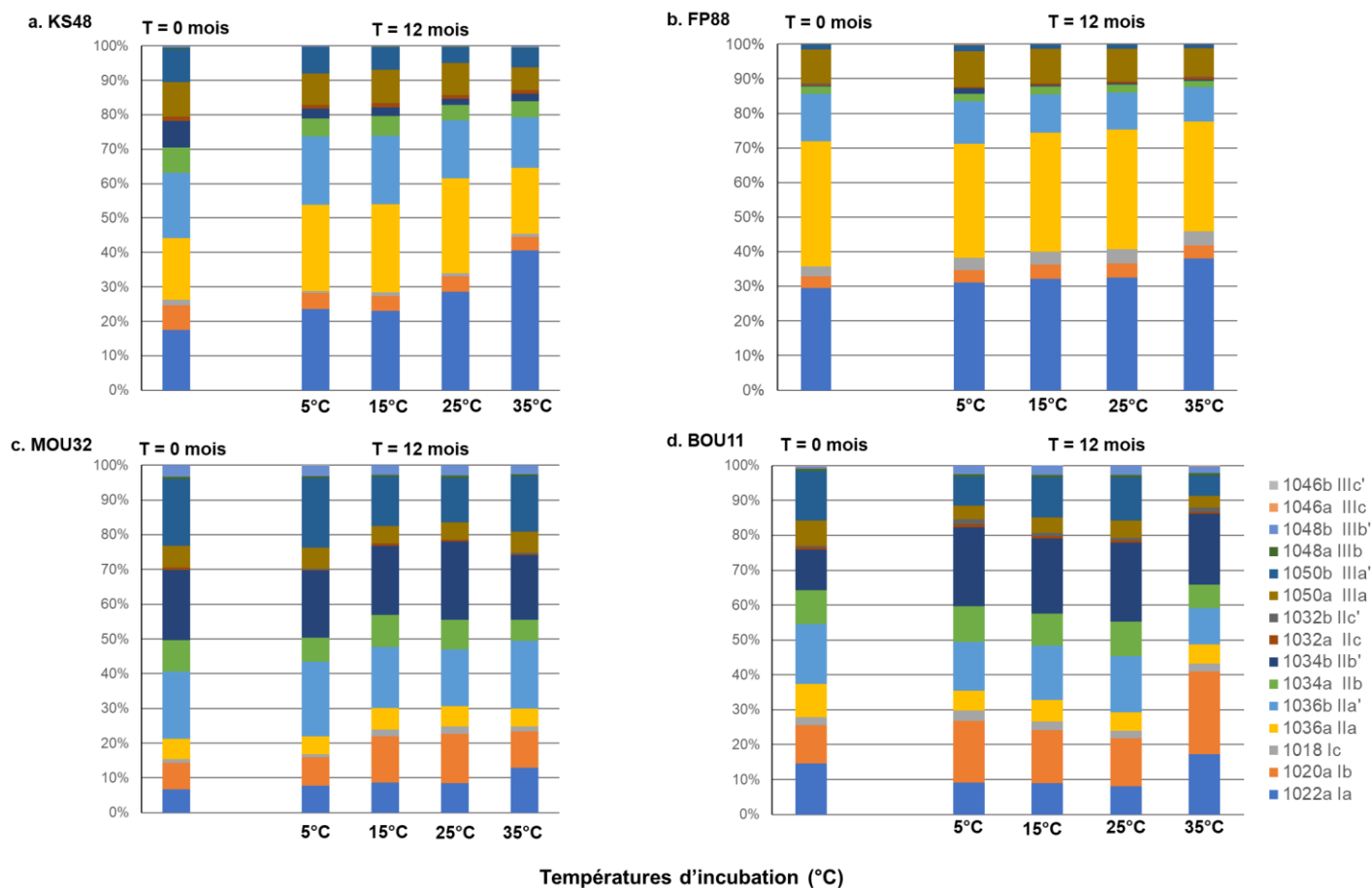


Figure 3.16 Abondances relatives des tétraéthères bactériens aux différentes températures d'incubations et aux différents pas de temps pour les échantillons a) KS48, b) FP88, c) MOU32 et d) BOU11. Les résultats présentés correspondent à la moyenne des abondances relatives sur les deux réplicats.

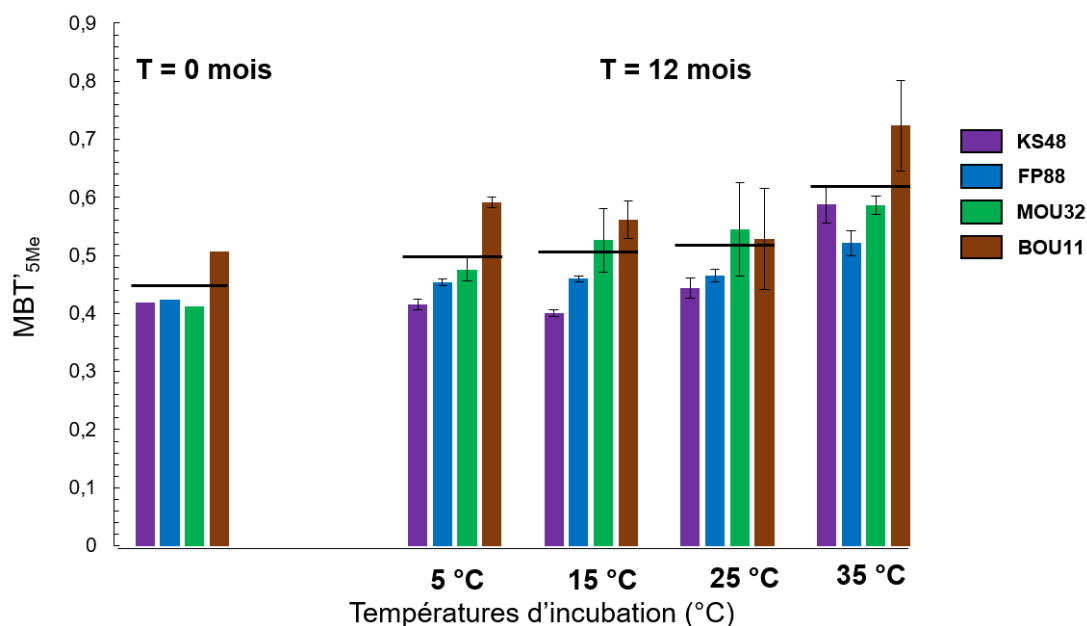


Figure 3.17 Indice MBT'_{5Me} pour les 4 échantillons des sols, aux différentes températures d'incubation et pour les différents pas de temps considérés. Les traits noirs représentent les moyennes de l'indice MBT'_{5Me} des 4 échantillons de sols pour les différentes modalités d'incubations.

3.2.2.3. Adaptation des microorganismes et production des lipides membranaires en réponse aux variations de températures

Les expériences en microcosmes sont particulièrement utiles pour contraindre l'influence d'une variable unique sur les microorganismes et la distribution de leurs lipides membranaires (par exemple pour les tétraéthers bactériens: Huguet et al., 2014; Huguet et al., 2017; Martínez-Sosa and Tierney, 2019).

Les analyses des acides gras 3-hydroxylés dans les sols des Alpes incubés en microcosmes pendant 12 mois n'ont pas révélé de différences notables en termes de concentrations et d'abondances relatives au cours du temps ($t = 0$ mois, 3 mois, 12 mois) et quelle que soit la température (5 à 35 °C), malgré la présence de biomasse viable en fin d'incubation. Ceci indique qu'une durée supérieure à douze mois est nécessaire pour que le stock de lipides présents initialement dans les sols soit profondément renouvelé et reflète les nouvelles conditions de température. Néanmoins, la diminution du degré de ramification des 3-AGH avec l'augmentation de température après 12 mois d'incubation

pour la plupart des échantillons placés en microcosmes suggère une influence de la température sur le profil lipidique des bactéries sources des 3-AGH. Cette étude est la première à proposer des résultats sur un effet de la température sur la distribution des acides gras 3-hydroxylés, et de manière indirecte sur la vitesse de renouvellement de ces composés au sein des sols.

A l'inverse, pour les tétraéthers bactériens, les expériences en microcosmes montrent une augmentation de la proportion des tétraéthers bactériens tetraméthylés *Ia* au sein de tous les échantillons et *Ib* dans le cas de BOU11. Cette augmentation se reflète également dans celle de l'indice de méthylation MBT'_{5Me} au bout de 12 mois d'incubations à 35°C pour les 4 échantillons considérés (Figure 8). Les températures moyennes annuelles de l'air observées sur site s'échelonnent entre 0°C (KS48) et 12,4°C (BOU11). Il est supposé que l'augmentation brutale des températures, par rapport aux conditions de températures naturelles, lors de l'incubation à 35°C, a provoqué une adaptation des organismes sources des tétraéthers, résultant en la synthèse préférentielle et rapide de tétraéthers acycliques (notamment le composé *Ia*) pour rigidifier la membrane (Sinensky, 1974; Ernst et al., 2016). L'augmentation de l'indice MBT'_{5Me} dans les sols avec la température est classiquement observée dans la littérature (Huguet et al., 2013, 2014).

Différentes études ont montré que le taux de renouvellement des tétraéthers bactériens (notamment des lipides de cœur, i.e. le stock fossile) était de l'ordre de quelques années dans les sols (Weijers et al., 2010; Peterse et al., 2010; Huguet et al., 2014). Ainsi, Peterse et al., (2010) ont estimé que celui-ci était inférieur à 45 ans et Weijers et al. (2010) ont estimé que le temps de renouvellement du stock de brGDGTs était d'environ 18 ans. De la même façon, Huguet et al. (2014), grâce à des expériences *in situ* et en laboratoire ont montré qu'entre trois et douze mois étaient nécessaires pour renouveler le contenu en lipides intacts (Intact Polar Lipids ou IPL, indicateur de biomasse vivante) au sein d'échantillons de tourbières, et supérieur à douze mois pour renouveler la distribution des lipides de cœur (lipides membranaires intacts ayant perdu leurs têtes polaires, considéré comme un « stock fossile) (Huguet et al., 2013, 2014). Ces différentes études pourraient expliquer l'absence ou la faible évolution des abondances relatives et des concentrations en acides gras 3-hydroxylés et en tétraéthers bactériens au sein des échantillons placés en microcosmes pendant 12 mois, le temps de renouvellement

du stock de lipides « de cœur » étant vraisemblablement plus long que celui des lipides intacts.

3.2.3. Conclusions

Les expériences de microcosmes menées sur quatre échantillons de sols contrastés, représentatifs des paramètres édapho-climatiques des Alpes françaises avaient pour objectif de contraindre une adaptation membranaire des microorganismes sources des 3-AGH et des tétraéthers bactériens en fonction de la température. Les expériences menées au cours de cette thèse n'ont pas permis de mettre en lumière une évolution notable des concentrations et des abondances relatives des 3-AGH. Il n'est pas totalement exclu que les bactéries sources de 3-AGH aient modifié leur profil lipidique en réponse aux changements de température. Notre expérience suggère cependant qu'il faut plus d'un an pour que le stock « fossile » de 3-AGH présent initialement dans les sols soit renouvelé et reflète les nouvelles conditions de température. Concernant les tétraéthers bactériens, au bout de 12 mois d'incubation, une légère augmentation de la proportion de tétraéthers tétraméthylés acycliques et de l'indice MBT'_{5Me} a pu être constatée en accord avec la littérature.

Afin de compléter ces expériences, il faudrait analyser les autres pas de temps des microcosmes (1, 3, 6 et 9 mois). Pour aller plus loin, il serait nécessaire de distinguer clairement au sein des échantillons le stock de lipides intacts (fraichement produits) et le stock de lipides de cœur (lipides fossiles), pour les tétraéthers bactériens et pour les 3-AGH. Cette approche permettrait d'observer l'évolution de la distribution et de la concentration des lipides intacts au sein des 4 sols pour les différents pas de temps et aux différentes températures. Néanmoins, si la méthode de séparation des lipides intacts et des lipides de cœurs existe déjà pour les brGDGTs (par exemple, Huguet et al., 2014), elle reste à être développer pour les acides gras 3-hydroxylés. Il serait aussi possible d'évaluer l'activité, la croissance et le temps de renouvellement des communautés microbiennes et des lipides membranaires associés (3-AGH et tétraéthers bactériens) dans ces échantillons de sols par marquage isotopique (Stable Isotope Probing, SIP). Cette méthode permettrait de suivre l'incorporation de substrats marqués au sein des lipides microbiens, et présente l'avantage de ne pas nécessiter de cultures de souches pures (par

exemple Boschker et al., (1998), Kopf et al., (2016)). Une méthode possible, utilisée par exemple par Huguet et al., (2017), propose une approche double SIP, avec des ajouts d'eau deutérée (D_2O) et de carbone inorganique dissous marqué au ^{13}C ($^{13}C_{DIC}$). Cette approche « double SIP » avec l'ajout simultané d'eau deutérée et le carbone inorganique marqué au ^{13}C ($^{13}C_{DIC}$) permettrait dans un premier temps de quantifier la proportion et la vitesse de production des lipides nouvellement produit au cours de l'expérience. En effet, il a été montré que la teneur en deutérium des lipides nouvellement synthétisés au cours de l'expérience est en équilibre avec la teneur en deutérium de l'eau ajouté comme substrat. De plus, l'incorporation de carbone inorganique dissous permet l'obtention d'information sur les voies de biosynthèse des lipides (autotrophie vs hétérotrophie). Le rapport entre le $^{13}C_{DIC}$ et le D_2O -incorporé dans les lipides nouvellement produits permet de distinguer ces deux voies de biosynthèse au sein des échantillons de sols placés en incubation. Ainsi cette expérience permettrait de fournir des informations complémentaires sur les organismes sources des 3-AGH et des tétraéthers bactériens, notamment leur métabolisme. Dans le même ordre d'idée, des expériences effectuées sur des souches pures des organismes sources des 3-AGHs, les bactéries Gram-négatif, pourraient être réalisées. Concernant les tétraéthers bactériens, leurs organismes sources sont encore largement inconnus (Weijers et al., 2009a; Sinninghe Damsté et al., 2018). Cette lacune empêche à l'heure actuelle l'utilisation de cultures de souches pures, excluant l'étude approfondie de la réponse des brGDGTs en fonction de paramètres environnementaux spécifiques en milieu contrôlé.

**Chapitre 4 – DEVELOPMENT OF
GLOBAL TEMPERATURE AND PH
CALIBRATIONS BASED ON BACTERIAL 3-
HYDROXY FATTY ACIDS IN SOILS**

Le chapitre 4 correspond à un manuscrit accepté pour publication dans *Biogeosciences*. Dans ce chapitre, nous allons tester l'applicabilité des 3-AGH comme marqueurs de température et de pH dans des sols du monde entier ($n=168$). La réponse des 3-AGH aux changements de températures et de pH dans ces sols a été comparée à celle des brGDGTs, utilisés comme marqueurs «de référence».

4.1. Introduction

Investigating past climate variations is essential to understand and predict future environmental changes, especially in the context of global anthropogenic change. Direct records of environmental parameters are available for the last decades, the so-called "instrumental" period. Beyond this period, proxies can be used to obtain indirect information on environmental parameters. A major challenge is to develop reliable proxies which can be applied to continental environments in addition to marine ones. Indeed, available proxies have been mainly developed and used in marine settings, as the composition and mechanism of formation of marine sedimentary cores are less complex than in continental settings, which are highly heterogeneous. Several environmental proxies based on organic (e.g. the alkenone unsaturation index (U^k_{37} ; Brassell et al., 1986) and inorganic (Mg/Ca ratio and $^{18}O/^{16}O$ ratio of foraminifera; Emiliani, 1955; Erez and Luz, 1983) fossil remains were notably developed for the reconstruction of sea surface temperatures.

Some of the existing proxies are based on membrane lipids synthesized by certain microorganisms (Eglinton and Eglinton, 2008; Schouten et al., 2013). These microorganisms are able to adjust the composition of their membrane lipids in response to the prevailing environmental conditions in order to maintain an appropriate fluidity and to ensure the optimal state of the cellular membrane (Singer and Nicolson, 1972; Sinensky, 1974; Hazel and Williams, 1990; Denich et al., 2003). The structure of glycerol dialkyl glycerol tetraethers (GDGTs), which are membrane lipids biosynthesized by archaea and some bacteria, is especially known to be related to environmental conditions. Archaeal GDGTs are constituted of isoprenoid alkyl chains ether-linked to glycerol, whereas bacterial GDGTs are characterized by branched alkyl chains instead of isoprenoid ones. The latter compounds are ubiquitous in terrestrial (Weijers et al., 2007;

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

Peterse et al., 2012; De Jonge et al., 2014; Naafs et al., 2017a) and aquatic environments (Peterse et al., 2009b; Tierney and Russell, 2009; Sinninghe Damsté et al., 2009; Loomis et al., 2012; Peterse et al., 2015; Weber et al., 2015). These branched GDGTs (brGDGTs) are produced by still unidentified bacteria, although some of them may belong to the phylum Acidobacteria (Sinninghe Damsté et al., 2011, 2014, 2018). The analysis of brGDGTs in a large number of soils distributed worldwide showed that the relative distribution of these compounds is mainly related to mean annual air temperature (MAAT) and soil pH (Weijers et al., 2007; Peterse et al., 2012; De Jonge et al., 2014). Even though brGDGT proxies were largely investigated over the last 10 years (De Jonge et al., 2014; Dearing Crampton-Flood et al., 2020) and were applied to various paleorecords (e.g. Coffinet et al., 2018; Wang et al., 2020), new molecular proxies, independent of and complementary to brGDGTs, are needed to improve the reliability of temperature reconstructions in terrestrial settings.

Recent studies have unveiled the potential of another family of bacterial lipids – 3-hydroxy fatty acids (3-OH FAs) – for temperature and pH reconstructions in terrestrial (Wang et al., 2016, 2018; Huguet et al., 2019) and marine (Yang et al., 2020) settings. 3-OH FAs with 10 to 18 carbon atoms are specifically produced by Gram-negative bacteria and are bound to the lipopolysaccharide (LPS) by ester or amide bonds (Wollenweber et al., 1982; Wollenweber and Rietschel, 1990). Three types of 3-OH FAs can be distinguished, with either *normal* chains or branched chains, *iso* or *anteiso*.

The analysis of 3-OH FAs in soils showed that the ratio of C₁₅ or C₁₇ *anteiso* 3-OH FA to *normal* C₁₅ or C₁₇ 3-OH FA (RAN₁₅ and RAN₁₇ indices, respectively) were negatively correlated with MAAT along the three mountains investigated so far: Mts. Shennongjia (China; Wang et al., 2016), Rungwe and Majella (Tanzania and Italy, respectively; Huguet et al., 2019). This suggests that Gram-negative bacteria producing these fatty acids respond to colder temperatures with an increase in *anteiso*-C₁₅/C₁₇ vs. *n*-C₁₅/C₁₇ 3-OH FAs, in order to maintain a proper fluidity and optimal state of the bacterial membrane, the so-called homeoviscous adaptation mechanism (Sinensky, 1974; Hazel and Eugene Williams, 1990). Nevertheless, the relationships between RAN₁₅ and MAAT along the three mountain transects showed the same slopes but different intercepts (Wang et al., 2016; Huguet et al., 2019), suggesting that regional or local RAN₁₅ relations may be more appropriate to apply for temperature reconstructions in terrestrial settings. In

contrast, a significant calibration between RAN₁₇ and MAAT could be established using combined data from the three mountain regions (Wang et al., 2016; Huguet et al., 2019).

Another index, defined as the cologarithm of the sum of *anteiso* and *iso* 3-OH FAs divided by the sum of *normal* homologues (RIAN index), was shown to be strongly negatively correlated with soil pH along the three aforementioned mountains (Wang et al., 2016; Huguet et al., 2020), reflecting a general relative increase in normal homologues compared to branched (*iso* and *anteiso*) ones with increasing pH. This mechanism was suggested to reduce the permeability and fluidity of the membrane for the cell to cope with lower pH (Russell et al., 1995; Denich et al., 2003; Beales, 2004).

For the first time, 3-OH FA indices were recently applied to the reconstruction of the temperature and hydrological changes over the last 10,000 years in a speleothem from China (Wang et al., 2018), showing the potential of 3-OH FAs as independent tools for environmental reconstruction in terrestrial settings. A very recent study based on marine sediments from the North Pacific Ocean suggested that the distribution of 3-OH FAs could also be used to reconstruct sea surface temperature (Yang et al., 2020).

Even though these results are promising, the linear regressions between pH or MAAT and 3-OH FA indices in terrestrial environments are still based on a rather small dataset (ca. 70 soil samples; Wang et al., 2016; Huguet et al., 2019). The aim of this study was to investigate the applicability of 3-OH FAs as MAAT and pH proxies at the global level using an extended soil dataset and refined statistical tools. The 3-OH FA distribution from 54 soils was determined in four globally distributed altitudinal transects (Tibet, Italy, the Peruvian Andes and Chile) and was combined with data previously published by Wang et al. (2016; Mt Shennongjia, China), Huguet et al. (2019; Mt. Rungwe, Tanzania, and Mt. Majella, Italy) and Véquaud et al. (2021; Mt. Bauges and Mt. Lautaret-Galibier, France), leading to a total of 168 samples. In addition to linear regressions, non-parametric, machine learning models were used to improve the global relationships between 3-OH FA distribution and MAAT/pH. These models present the advantage of taking into account non-linear environmental influences, in line with the intrinsic complexity of the environmental settings. Finally, these new models were tested and compared by applying them to a speleothem archive (Wang et al., 2018) representing to date the only available MAAT record derived from 3-OH FA proxies in a continental setting. As brGDGTs are the only microbial organic proxies which can be used for

temperature and pH reconstructions in terrestrial settings so far, they can serve as a reference proxy to understand the temperature and pH dependency of 3-OH FAs analysed in the same dataset. 3-OH FAs and brGDGTs have thus been concomitantly analyzed to assess their reliability and complementarity as independent temperature and pH proxies.

4.2. Results

4.2.1. Distribution of bacterial lipids

4.2.1.1. 3-OH-FAs

Representing eight elevation transects and 168 samples, 3-OH FAs were identified in the whole dataset, (Annex 2; Supp. table 4.1; Yang et al., 2015; C. Wang et al., 2016; Coffinet et al., 2017; Huguet et al., 2019; Véquaud et al., 2021). Their chain lengths range between 8 and 26 C atoms, indicating that these compounds have various origins (bacteria, plants, and fungi; Zelles, 1999; Wang et al., 2016 and reference therein). The homologues of 3-OH FAs with 10 to 18 C atoms are considered to be produced exclusively by Gram-negative bacteria (Wollenweber and Rietschel, 1990; Szponar et al., 2003) and will be the only ones considered in the following. Compounds with an even carbon number and *normal* chains were the most abundant 3-OH FAs in all samples (mean = 67.9 % of the total 3-OH FAs, standard deviation (SD) = 6.8%), with a predominance of the *n*-C₁₄ homologue (21.9%, SD = 3.23%; Fig. 1). *Iso* (mean = 22.9%, SD = 5.01%) and *anteiso* (mean = 6.33 %, SD = 1.79%) isomers were also present. It must be noted that *anteiso* isomers were only detected for odd carbon-numbered 3-OH FAs (Yang et al., 2015; C. Wang et al., 2016; Coffinet et al., 2017; Huguet et al., 2019).

The distribution of 3-OH FAs in the soils of the different altitudinal transects did not show a large variability (Fig. 4.1). Thus, there was no major difference in the relative abundances of most of the 3-OH FAs (*i*-C₁₁, *a*-C₁₁, *n*-C₁₁, *i*-C₁₂, *a*-C₁₃, *n*-C₁₃, *i*-C₁₄, *n*-C₁₅, *i*-C₁₆, *a*-C₁₇ and *n*-C₁₇) between the eight study sites, even though slight differences could be observed for some compounds as detailed below. For example, the Peruvian samples were characterized by higher average proportions of *n*-C₁₈ 3-OH FA and lower contribution of the *n*-C₁₀ and *n*-C₁₂ homologues than those from the other transects. Soils

DEVELOPMENT OF GLOBAL TEMPERATURE AND PH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

from Mt. Shegyla were characterized by lower average proportions of *n*-C₁₄ 3-OH FAs and higher abundances of *i*-C₁₇ compounds compared to the other transects (Fig. 4.1).

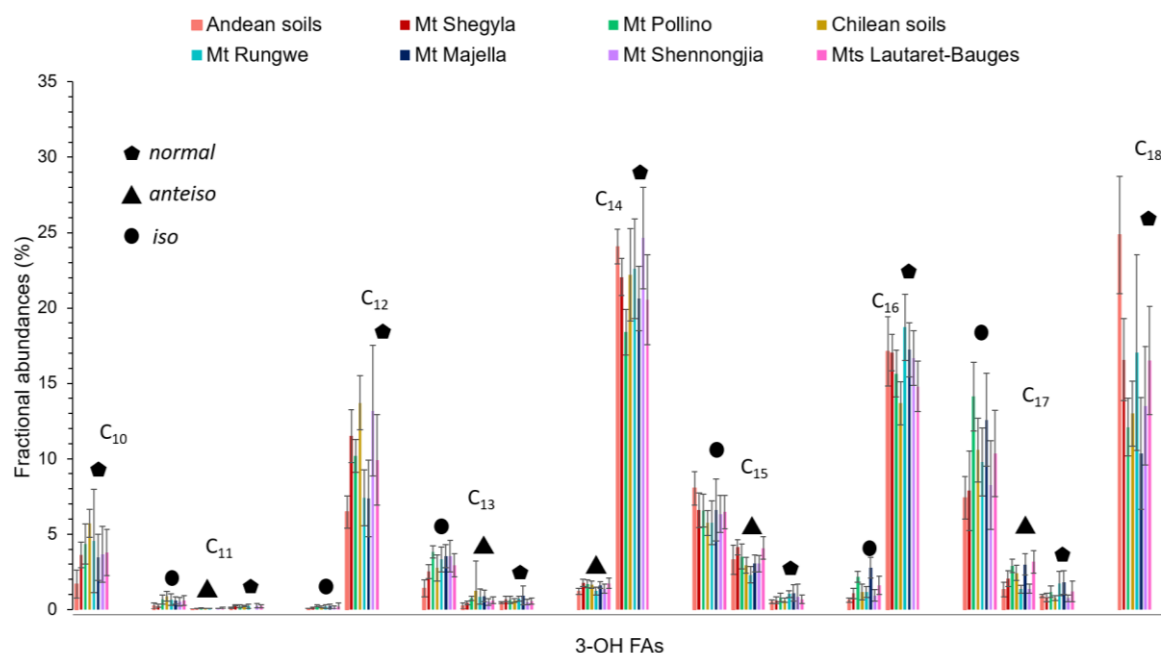


Figure 4.1 Average distribution of 3-OH FAs along the eight altitudinal transects investigated in this study. Data from Mt. Majella and Mt. Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Wang et al. (2016). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021).

4.2.1.2. *brGDGTs*

The relative abundances of *brGDGTs* were compared between the same transects as for 3-OH FAs, representing a total of 168 samples. The 5- and 6-methyl isomers were separated in most of the samples (Fig. 4.2; Annex 2; Supp. table 4.2), except in older dataset, i.e. soils from Mt. Rungwe (Coffinet et al., 2014, 2017). The *brGDGT* data from Mt. Rungwe will not be further considered in this study.

The *brGDGT* distribution was dominated by acyclic compounds (Ia, IIa, IIa', IIIa, IIIa') which represent on average ca. 83.4% of total *brGDGTs* (SD = 14.5%; Fig. 4.2). The tetramethylated (Ia-c; mean = 39.3%, SD of 20.5%) and the pentamethylated (IIa-c; 44.8%, SD = 12.8%) *brGDGTs* were predominant over the hexamethylated ones (IIIa-c; Fig. 4.2). The 5-methyl isomers were on average present in a higher proportion (mean = 71.9%, SD = 23.4%) than the 6-methyl compounds (Fig. 4.2).

DEVELOPMENT OF GLOBAL TEMPERATURE AND PH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

High variability in the brGDGT distribution was observed among the different transects. The relative abundance of brGDGT Ia was much higher in the Peruvian soils (mean = 83%, SD = 12.6%) than in the other transects (mean between 17.3% and 61.7%; Fig. 4.2). The 5-methyl isomers were more abundant than the 6-methyl isomers for all sites except for Mt. Pollino (mean 5-methyl = 44%, SD=11.7%) and Mt. Majella (mean 5-methyl = 33.7 %, SD = 5.5%; Fig. 4.2).

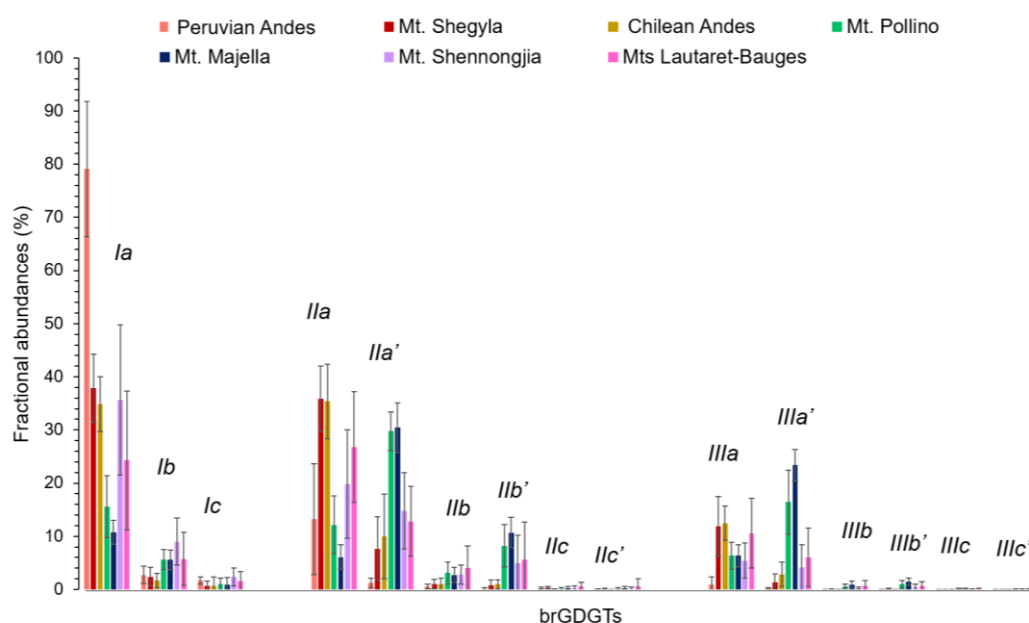


Figure 4.2 Average distribution of 5- and 6-methyl brGDGTs, along Mt. Shegyla, Mt. Pollino Majella, Lautaret-Bauges, the Peruvian Andes and the Chilean Andes. Data from Mt. Majella were taken from Hugué et al. (2019). Data from Mt. Shennongjia were taken from Yang et al. (2015). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021).

4.2.2. 3-OH FA- and brGDGT-derived indices

4.2.2.1. 3-OH FAs

The RIAN index varied between 0.1 and 0.8 among the eight elevation transects (Table 4.1). The RIAN index ranged from 0.37 to 0.67 for the Peruvian Andes, 0.23 to 0.56 for Mt. Shegyla, 0.15 to 0.34 for Mt. Pollino, 0.21 to 0.53 for the Chilean Andes, 0.26 to 0.80 for Mt. Rungwe (Hugué et al., 2019), 0.16 to 0.46 for Mt. Majella (Hugué

et al., 2019), 0.20 to 0.69 for Mt. Shennongjia (C. Wang et al., 2016) and 0.13 to 0.56 for the French Alps (Véquaud et al., 2021).

The RAN₁₅ varied greatly among the different sites (Table 4.1). It was in the same range along Mt. Rungwe (1.04-5.73) and Mt. Majella (0.68-6.43; Huguet et al., 2019). In contrast, its upper limit was higher for Mt. Shennongjia (0.67-10.77; Wang et al., 2016), Mt. Shegyla (4.07-12.17), Mt. Pollino (2.41-10.26), the Peruvian Andes (2.45-13.77) and the French Alps (1.44-12.26). The range of variation in RAN₁₅ was narrower for the Chilean Andes (3.82-6.40).

The RAN₁₇ values were similar among the different altitudinal transects (Table 4.1), ranging from 1.72 to 3.90 along Mt. Shegyla, 0.68 to 6.43 along Mt. Majella (Huguet et al., 2019), 1.19 to 4.54 along Mt. Pollino, 1.91 to 4.25 for the Chilean Andes and 1.12 to 3.57 along Mt. Shennongjia (C. Wang et al., 2016). The range of RAN₁₇ values was narrower for Mt. Rungwe (0.33-1.62; Huguet et al., 2019) and the Peruvian Andes (0.61-2.39) and wider for the French Alps (0.89-6.42; Véquaud et al., 2021) compared to the other sites.

4.2.2.2. *brGDGTs*

The range of variation in the MBT'_{5Me} index was homogeneous along most transects (0.32-0.63; Table 4.1), except the Peruvian Andes, with higher values (0.58-0.98; Table 4.1). Regarding the CBT' index, it showed similar ranges along the Chilean Andes (-2.28 to -0.32) and Mt. Shegyla (-2.39 to -0.35; Table 4.1). This index showed different ranges of variations along the other altitudinal transects : Mt. Shennongjia (-1.18 to 0.50; Yang et al., 2015), Mt. Pollino (-0.24 to 0.43) and the Peruvian Andes (-1.91 to -1.09). Finally, the CBT' values varied within a narrow range along Mt. Majella (0.23-0.59; Huguet et al., 2019) and within a wide range along the French Alps (-2.29 to 0.52 ; Véquaud et al., 2021).

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

ID	Location	Altitude (m)	MAAT (°C)	pH	RAN ₁₅	RAN ₁₇	RIAN	MBT' _{SMt}	CBT'
1	Peruvian Andes	194	26.4	3.7	2.45	0.96	0.47	0.96	-1.09
2	Peruvian Andes	210	26.4	4	2.56	0.61	0.60	0.97	-1.92
3	Peruvian Andes	1063	20.7	4.7	3.46	0.70	0.54	0.98	-1.76
4	Peruvian Andes	1500	17.4	3.5	4.15	0.93	0.51	0.91	-1.55
5	Peruvian Andes	1750	15.8	3.6	5.30	1.32	0.51	0.92	-1.62
6	Peruvian Andes	1850	16	3.5	6.81	1.23	0.54	0.96	-1.76
7	Peruvian Andes	2020	14.9	3.4	7.00	1.19	0.54	0.95	-1.68
8	Peruvian Andes	2520	12.1	3.7	8.40	1.59	0.53	0.74	-1.42
9	Peruvian Andes	2720	11.1	3.6	8.42	1.73	0.48	0.83	-1.45
10	Peruvian Andes	3020	9.5	3.4	13.78	2.21	0.44	0.83	-1.21
11	Peruvian Andes	3200	8.9	3.5	6.91	2.35	0.37	0.71	-1.48
12	Peruvian Andes	3025	11.1	3.5	8.86	1.74	0.52	0.82	-1.66
13	Peruvian Andes	3400	7.7	3.4	9.10	2.39	0.40	0.71	-1.39
14	Peruvian Andes	3644	6.5	3.4	8.93	2.03	0.67	0.58	-1.21
15	Mt. Shegyla, Tibet	3106	8.9	5.53	6.22	2.02	0.51	0.59	-0.83
16	Mt. Shegyla, Tibet	3117	8.9	6.43	4.47	1.86	0.36	0.57	-0.35
17	Mt. Shegyla, Tibet	3132	8.8	6.01	4.07	1.72	0.43	0.61	-0.47
18	Mt. Shegyla, Tibet	3344	7.6	6.03	5.40	2.80	0.34	0.51	-0.67
19	Mt. Shegyla, Tibet	3355	7.5	5.87	4.09	2.71	0.23	0.44	-0.39
20	Mt. Shegyla, Tibet	3356	7.5	5.52	3.87	2.14	0.25	0.42	-0.70
21	Mt. Shegyla, Tibet	4030	3.7	5.21	8.21	3.64	0.43	0.49	-1.10
22	Mt. Shegyla, Tibet	4046	3.6	4.68	8.37	3.00	0.49	0.52	-1.17
23	Mt. Shegyla, Tibet	4050	3.6	4.61	8.94	2.47	0.50	0.44	-1.33
24	Mt. Shegyla, Tibet	3912	4.3	5.04	9.74	2.30	0.48	0.40	-2.39
25	Mt. Shegyla, Tibet	3918	4.3	4.68	8.67	1.80	0.56	0.45	-2.23
26	Mt. Shegyla, Tibet	4298	2.1	5.04	10.00	2.78	0.50	0.45	-2.04
27	Mt. Shegyla, Tibet	4295	2.2	4.87	12.17	3.90	0.50	0.42	-1.07
28	Mt. Shegyla, Tibet	4304	2.1	5.26	10.10	3.20	0.46	0.46	-1.14
29	Mt. Shegyla, Tibet	4479	1.1	5.26	10.11	3.42	0.52	0.35	-1.27
30	Mt. Shegyla, Tibet	4479	1.1	5.07	5.71	3.00	0.50	0.35	-0.84
31	Mt. Shegyla, Tibet	4474	1.1	5.24	7.88	3.65	0.42	0.32	-1.15
32	Mt. Pollino, Italy	0	18	6.78	2.71	1.19	0.15	0.50	0.31
33	Mt. Pollino, Italy	200	17	6.19	2.41	1.28	0.30	0.63	0.34
34	Mt. Pollino, Italy	400	16	6.13	4.26	2.29	0.22	0.58	0.35
35	Mt. Pollino, Italy	600	15	6.14	4.15	2.36	0.22	0.55	0.43
36	Mt. Pollino, Italy	800	14	4.53	3.34	2.77	0.34	0.51	-0.24
37	Mt. Pollino, Italy	1000	13	5.41	3.06	1.83	0.28	0.48	0.10
38	Mt. Pollino, Italy	1200	12	6.37	4.21	1.91	0.24	0.55	0.43
39	Mt. Pollino, Italy	1400	11	5.62	5.77	4.16	0.18	0.52	0.40
40	Mt. Pollino, Italy	1600	10	4.93	7.64	4.54	0.27	0.44	-0.13
41	Mt. Pollino, Italy	1800	9	4.91	3.45	3.17	0.25	0.45	-0.07
42	Mt. Pollino, Italy	2000	8	5.52	6.35	4.52	0.19	0.56	0.40
43	Mt. Pollino, Italy	2100	7.5	5.91	10.26	3.62	0.19	0.42	0.38
44	Mt. Pollino, Italy	2200	7	5.85	6.21	2.82	0.31	0.47	0.34
45	Chilean Andes	690	9.2	5.38	5.01	3.51	0.42	0.41	-0.80
46	Chilean Andes	870	8.9	5.62	5.21	2.43	0.39	0.49	-0.52
47	Chilean Andes	891	7.9	4.94	5.18	2.69	0.53	0.44	-0.94
48	Chilean Andes	915	NA	6.75	4.67	4.25	0.21	NA	NA
49	Chilean Andes	980	8.5	5.63	3.87	3.83	0.28	0.46	-0.66
50	Chilean Andes	985	5.8	4.67	6.41	3.12	0.48	0.41	-1.83
51	Chilean Andes	1125	6.0	5.00	3.83	4.18	0.46	0.42	-1.02
52	Chilean Andes	1151	6.0	5.89	4.74	2.89	0.33	0.43	-0.32
53	Chilean Andes	1196	7.1	5.79	5.70	4.07	0.34	0.43	-0.40
54	Chilean Andes	1385	NA	4.43	4.85	1.91	0.39	0.41	-2.28

Table 4.1 List of the soil samples collected along Mts. Shegyla, Pollino, Peruvian Andes and Chilean Andes, with corresponding altitude (m), MAAT (°C), pH and 3-OH FA/brGDGT-derived indices.

4.2.3. Principal component analysis and clustering of 3-OH FA and brGDGT distribution

Principal component analyses were performed to refine the comparison of bacterial lipid distribution (3-OH FAs and brGDGTs) among the different altitudinal transects.

4.2.3.1. 3-OH FAs

The first two axes of the 3-OH FA PCA explained 39.1% of the total variance in the dataset (Fig. 4.3a). Dimension 1 (23.9%) opposed samples from Mt. Pollino in the right quadrant to Peruvian soils and samples from Mt. Shennongjia. Dimension 2 (15.2%) especially separated individuals from Chile and Mt. Rungwe. The Wilks' test showed that the location of the sampling sites was the best variable discriminating the distribution of the individuals in the PCA.

Principal component analysis performed on the temperature (RAN₁₅, RAN₁₇) and pH (RIAN) indices derived from 3-OH FAs showed that most of the variance was carried by the first two axes of the PCA (Axis 1 = 56.09%; Axis 2 = 35.29%; Annex 2; Supp. Fig 4.1). The first axis was highly correlated with the RAN₁₅ ($r = 0.87$) and RAN₁₇ ($r = 0.93$) as well as with MAAT ($r = -0.67$), while Axis 2 showed strong correlations with the RIAN ($r = 0.96$) and pH ($r = -0.61$). The PCA allowed the visualization of relationships at the scale of the whole dataset between MAAT and RAN₁₅ and RAN₁₇ ($r = -0.61$; $r = -0.64$ respectively) and between pH and RIAN ($r = -0.53$).

4.2.3.2. brGDGTs

The first two axes of the brGDGT PCA explained 57.7% of the total variance in the dataset (Fig. 4.3b). Dimension 1 (42.6%) strongly discriminated soils from Mt. Majella and, to a lesser extent, Mt. Pollino in the right quadrant from those from Mt. Shegyla, the Peruvian Andes and the Chilean Andes in the left quadrant. Mt. Majella and Mt. Pollino were also discriminated negatively along dimension 2 (15.1%). Samples from Mts. Shennongjia and Lautaret-Galibier were distributed over the entire PCA. As for the 3-OH FAs, Wilks' test showed that the location of the sampling sites was the best variable discriminating the distribution of the brGDGTs in the PCA.

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

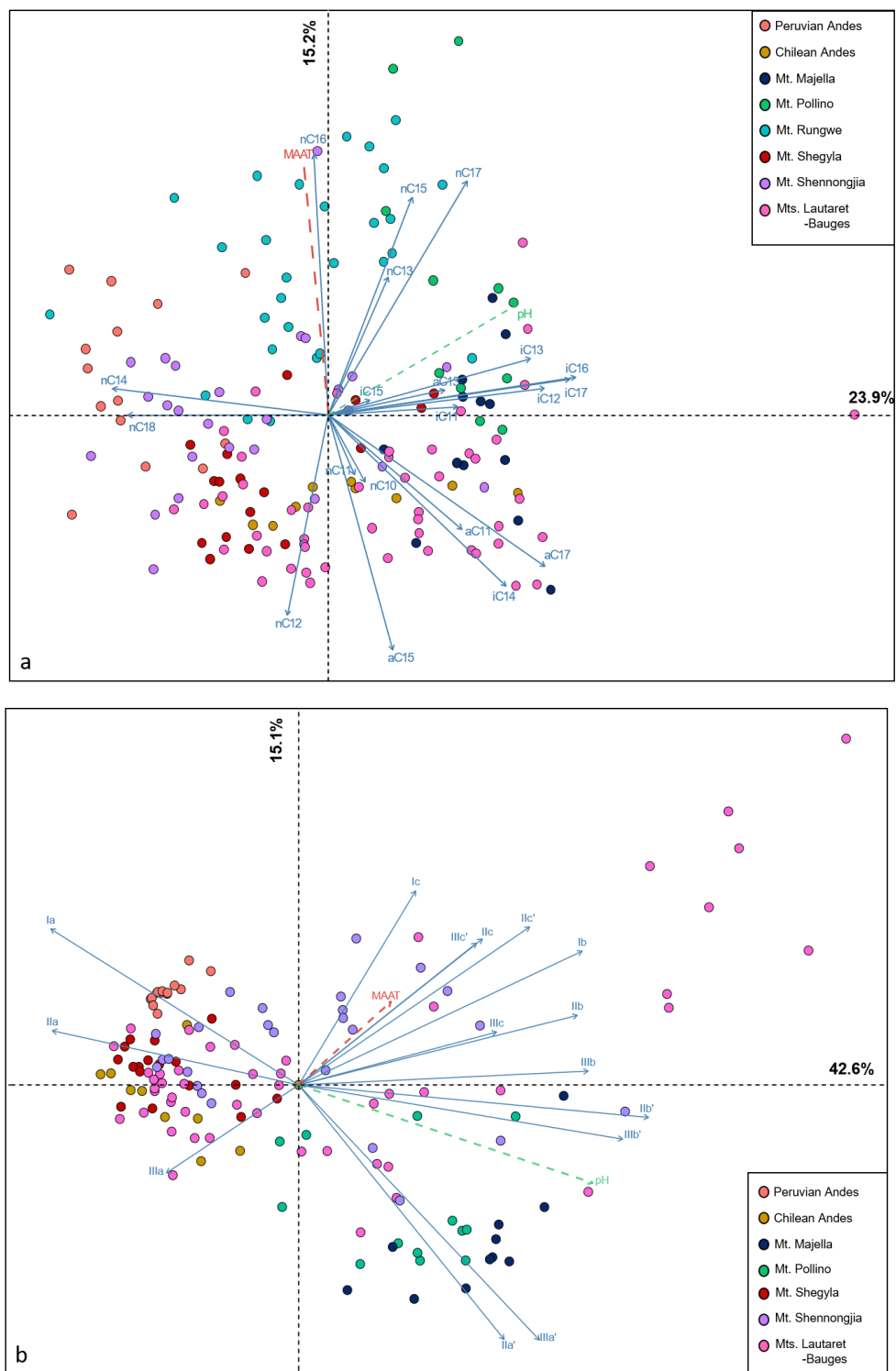


Figure 4.3 PCA biplot of (a) 3-OH FA fractional abundances in soil samples from the eight altitudinal transects and (b) brGDGT fractional abundances in soil samples from seven of the eight altitudinal transects. The brGDGT data from Mt. Rungwe, for which 5- and 6-methyl isomers were not separated, were not included in the PCA.

4.3. Discussion

4.3.1. 3-OH FA- and brGDGT-derived proxies

Previous studies conducted on soils from individual altitudinal transects revealed (1) local linear relationships between MAAT/pH and 3-OH FA indices and (2) the potential for combined calibrations using simple linear regressions (C. Wang et al., 2016; Huguet et al., 2019; Véquaud et al., 2021). In the present study, the existence of linear relationships between 3-OH-FA-derived indices and environmental variables was further investigated using an extended soil dataset and the corresponding results were compared with those derived from the brGDGTs, used as an established reference proxy.

4.3.1.1. Relationships between pH- and bacterial lipid-derived proxies

The relationship between RIAN and pH was investigated along each of the altitudinal transects (Fig. 4.4a; Annex 2; Supp. table 4.3). No significant linear relationship was obtained for the Peruvian Andes, Mt. Rungwe, Mt. Pollino and Mt. Majella (Huguet et al., 2019) and weak to moderate correlations were observed along Mt. Shegyla and Lautaret-Bauges ($R^2 = 0.29-0.46$; Annex 2; Supp. table 4.3). In contrast, strong regressions between RIAN and pH were observed along Mt. Shennongjia ($R^2 = 0.71$) and in Chilean Andes ($R^2 = 0.66$). A weak linear relationship between RIAN and pH ($R^2=0.34$, $RMSE = 0.99$, $p = 7.39 \times 10^{-17}$) was also obtained when considering the 168 samples for the eight elevation transects altogether. Therefore, our results confirm the general influence of pH on the relative abundance of 3-OH FAs (Huguet et al., 2019) but suggest that strong linear correlations between RIAN and pH can only be obtained (i) at a local level and (ii) only for some of the sites.

As previously suggested (Huguet et al., 2019), the absence or weakness of linear correlations between RIAN and pH may be at least partly due to the small range of variation in pH (<2 units) along some mountains, such as Mt. Rungwe, Mt. Majella, and the Peruvian Andes (Fig. 4.4a; Table 4.1, Huguet et al., 2019). Transects for the Peruvian Andes and Mt. Majella were also characterized by the absence of relationships between pH and the brGDGT-derived CBT' index, supporting the hypothesis that low pH ranges

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

limit the potential of obtaining linear relationships between indices based on bacterial lipids and pH. Nevertheless, the existence of a narrow pH range was not the only limiting factor in obtaining a strong linear regression between RIAN and pH. Indeed, MAAT rather than soil pH was the dominant driver of soil bacterial diversity and community composition for the Peruvian transect (using 16S rRNA sequencing (Nottingham et al., 2018); and using phospholipid fatty acids (Whitaker et al., 2014)), consistent with the weak correlation between soil pH and bacterial lipids. The weakness of the RIAN-pH relationship may also be partly due to the heterogeneity of soils encountered along a given altitudinal transect, representing specific microenvironments, and to the large diversity of bacterial communities in soils from different elevations (Siles and Margesin, 2016). The distribution of 3-OH FAs varies greatly among Gram-negative bacterial species (Bhat and Carlson, 1992), which may account for the significant variability in RIAN values observed in soils from a given transect. Altogether, these results suggest that linear models are not the most suitable for establishing a global calibration between RIAN and pH in soils.

Concerning GDGTs, moderate to strong relationships between brGDGT-derived CBT' index and pH were observed along five of the seven altitudinal transects investigated (Fig. 4.4b; Annex 2; Supp. table 4.3). All the individual linear relationships between CBT' and pH, where present, had similar slopes and ordinates and share (for most of the samples) the same 95% confidence intervals (pvalue <0.5). This resulted in a strong linear relationship between CBT' index and pH values for the dataset ($R^2 = 0.68$, RMSE = 0.71, $n = 140$), which is weaker than the global calibration ($R^2 = 0.85$, RMSE = 0.52, $n = 221$) proposed by De Jonge et al. (2014).

The discrepancy in relationships between temperature and brGDGTs and 3-OH FAs might partly be due to differences in the relative abundance of these lipids among bacterial communities. The brGDGTs are produced by a more restricted and less diverse number of bacterial species than 3-OH FAs, which are arguably biosynthesized by a large diversity of Gram-negative bacteria species (e.g. Wakeham et al., 2003, Zelles et al., 1995; Zelles, 1999). So far, only bacteria from the Acidobacteria phylum were identified as putative brGDGT producers in soils (Sinninghe Damsté et al., 2018). The hypothetical lower diversity of brGDGT producers, in contrast with 3-OH FAs might explain the more homogenous response and lower scatter of the relationships between pH and CBT' index.

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

Moreover, the CBT' index is a ratio based on a restricted number of compounds, representing the direct dependence of the degree of cyclization of bacterial GDGTs on pH. Conversely, the RIAN index is calculated from the relative abundances of all the individual 3-OH FAs between C₁₀ and C₁₈ (C. Wang et al., 2016). It cannot be ruled out that some of the compounds used to calculate the RIAN index are preferentially synthesized, as part of the homeoviscous mechanism, in response to environmental variables other than pH. This calls for a better understanding of the ecology of 3-OH FA-producing bacteria and their adaptation mechanisms.

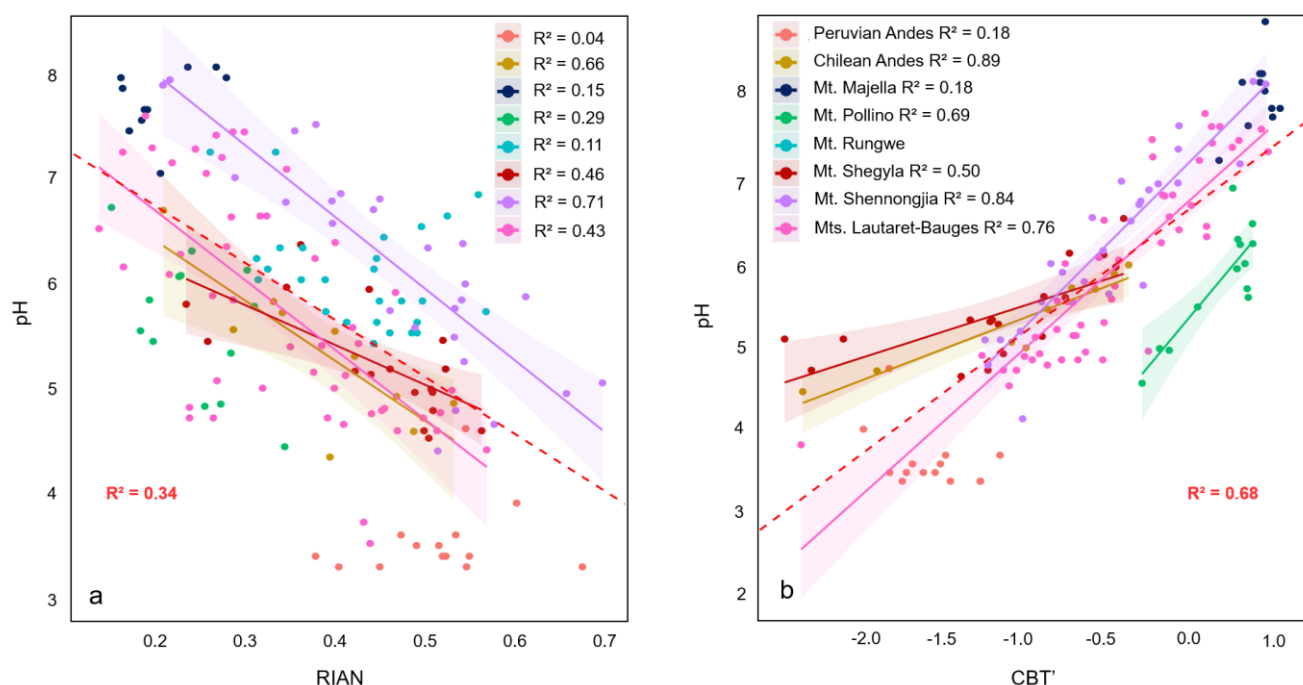


Figure 4.4 Linear regressions between (a) pH and RIAN and (b) pH and CBT' along the eight altitudinal transects investigated. Dotted lines represent the 95% confidence interval for each regression and coloured areas represent the 95% confidence interval for each regression. Data for Mt. Majella and Mt. Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Yang et al. (2015) and C. Wang et al. (2016). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021). Only significant regressions ($p < 0.05$) are shown.

4.3.1.2. Relationships between MAAT- and bacterial lipid-derived proxies

RAN₁₅ was previously shown to be correlated with MAAT along Mt. Rungwe, Mt. Majella and Mt. Shennongjia (C. Wang et al., 2016; Huguet et al., 2019). Moderate to strong linear correlations ($R^2 = 0.49-0.79$) between RAN₁₅ and MAAT were also observed along most of the individual transects investigated (Fig. 4.5a; Annex 2; Supp. table 4.3), except along the Chilean and Lautaret-Bauges transects. The individual correlations do not share the same 95% confidence intervals and even when some of them present similar slopes, the regression lines display significantly different intercepts (p -value > 0.05) (Fig. 4.5a). This supports the hypothesis of a site-dependent effect of the linear RAN₁₅-MAAT relationship previously made by Huguet et al. (2019).

Similarly to RAN₁₅, RAN₁₇ was moderately to strongly correlated ($R^2 = 0.53-0.81$) with MAAT along 5 out of 8 individual transects (Fig. 4.5b; Annex 2; Supp. table 4.3). The small range of variation in MAAT along the Chilean transect (6.0-9.2 °C) (Table 4.1), associated with that of the RAN₁₅ /RAN₁₇, could explain the lack of a linear relationship between the MAAT and these indices. As for the French Alps (Mts Lautaret-Bauges), the influence of local environmental parameters (pH and to a lesser extent soil moisture and grain size, related to vegetation and soil types, or thermal regimes associated with the snow cover) on 3-OH FA distribution was shown to be predominant over that of MAAT (Véquaud et al., 2021). In contrast with RAN₁₅, the linear regressions between RAN₁₇ and MAAT along Mts. Shegyla, Shennongjia, Rungwe and the Peruvian Andes transects share confidence intervals at 95% and have similar slope and intercept values (p -value < 0.05 ; Fig. 4.5b; Annex 2; Supp. table 4.3), suggesting that RAN₁₇ could be a more effective global proxy for MAAT reconstructions than RAN₁₅.

In order to test the hypothesis that RAN₁₇, rather than RAN₁₅, is a more effective global proxy for MAAT, the global calibrations between RAN₁₅/RAN₁₇ and MAAT based on the entire soil dataset ($n = 168$) were compared. The two linear regressions had similar moderate determination coefficients ($R^2 = 0.37$ and 0.41 for RAN₁₅ and RAN₁₇, respectively) and similar high RMSE (RMSE = 5.46°C and 5.28°C for RAN₁₅ and RAN₁₇, respectively; Annex 2; Supp. table 4.3). For all transects (except for the Mt Majella RAN₁₇/MAAT relationship), the individual local regressions between

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

RAN₁₅/RAN₁₇ and MAAT outperformed the proposed global linear calibrations in terms of determination coefficients (0.49-0.81) and RMSE (1.98-3.57 °C; Annex 2; Supp. table 4.3), suggesting that local rather than global linear transfer functions based on RAN₁₅ or RAN₁₇ may be more appropriate for paleotemperature reconstructions in soils.

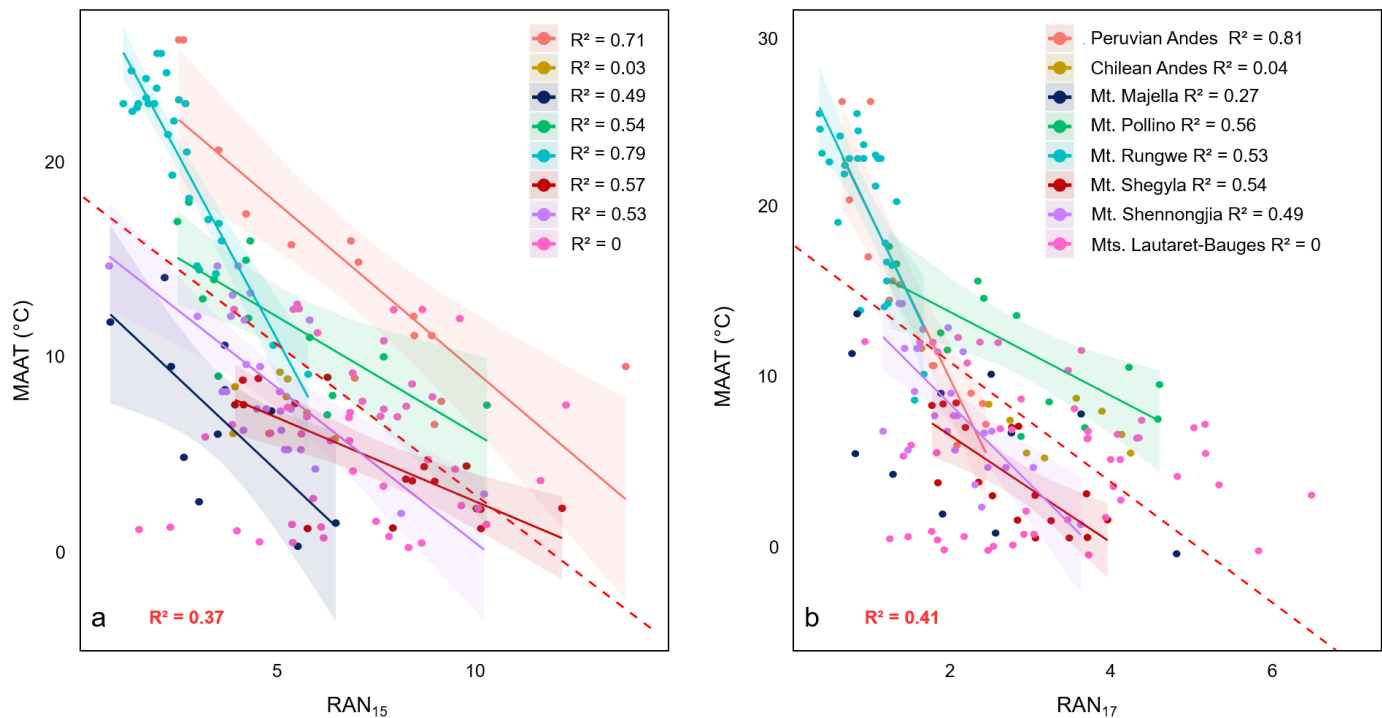


Figure 4.5 Linear regressions between (a) MAAT and RAN₁₅ and (b) MAAT and RAN₁₇ along the eight altitudinal transects investigated. Dotted lines represent the 95% confidence interval for each regression and coloured areas represent the 95% confidence interval for each regression. Data from Mt. Majella and Mt. Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from C. Wang et al. (2016). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021). Only significant regressions ($p < 0.05$) are shown.

The difficulties in establishing global linear RAN₁₅/RAN₁₇-MAAT calibrations may partly be due to the fact that microbial diversity, especially for 3-OH-FA-producing Gram-negative bacteria (Margesin et al., 2009; Siles and Margesin, 2016), can vary greatly from one soil to another, resulting in variation in the RAN₁₅/RAN₁₇ indices, as also assumed for the RIAN. The strong regional dependence of the 3-OH-FA distribution may thus explain the weak correlation between 3-OH FA-derived indices (RAN₁₅, RAN₁₇ and RIAN) and environmental variables (MAAT/pH) at a global level. This regional dependency was further supported by the PCA of the relative abundance of 3-OH FAs

DEVELOPMENT OF GLOBAL TEMPERATURE AND PH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

across the global dataset, which showed that the individuals were grouped based on the sampling location (Fig. 4.3a).

In addition to 3-OH FAs, the relationships between brGDGT distribution and MAAT were investigated along the seven transects for which the 5- and 6-methyl brGDGT isomers were separated (Mt. Shegyla, Mt. Pollino Majella, Lautaret-Bauges, Mt. Shennongjia, the Peruvian Andes and the Chilean Andes). These individual transects showed moderate to strong relationships between MAAT and MBT'_{5Me} (R^2 0.35-0.89; Fig. 4.6 and Annex 2; Supp. table 4.3), with similar slopes and ordinates (except for the Peruvian Andes) and shared 95% confidence intervals for most of the samples. A distinct relationship between MBT'_{5Me} and MAAT was observed along the Peruvian Andes and Mt Majella transects (Fig. 4.6), as also observed for the RIAN and RAN_{15} indices (Figs 4.4a and 4.5a). The singularity of the Peruvian soils is also visible on the PCA performed in the brGDGT distribution (Fig. 4.3b), where the samples from this region are pooled very narrowly from the rest of the dataset. This specific trend is difficult to explain, even though the Peruvian Andes are subjected to warmer climatic conditions (Table 4.1) than the other temperate transects, which may in turn affect the nature of the microbial communities encountered in the soils and the bacteria lipid distribution (Siles and Margesin, 2016; Hofmann et al., 2016; De Jonge et al., 2019).

A moderate linear relationship between MAAT and MBT'_{5Me} ($MAAT = 24.3 \times MBT'_{5Me} - 5.03$; $R^2 = 0.57$, $RMSE = 3.39$ °C, $n = 140$; Annex 2; Supp. table 4.3) was observed after combining the data for the seven aforementioned altitudinal transects. This global relationship follows a similar trend as the calibration proposed by De Jonge et al. 2014 ($MAAT = 31.45 \times MBT'_{5Me} - 8.57$) and is more robust and accurate than those obtained between the RAN_{15}/RAN_{17} and MAAT (Annex 2; Supp. table 4.3). This confirms that the MBT'_{5Me} index can be applied at a global scale using a simple linear regression model as previously shown (De Jonge et al., 2014; Naafs et al., 2017a), in contrast with the RAN_{15} and RAN_{17} proxies, for which only strong local calibrations with MAAT were found.

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

As a similar conclusion was obtained for the RIAN-pH proxy, it appears necessary to use more complex models to develop global calibrations between 3-OH-FA-derived proxies and MAAT/pH. This novel method allows the complexity and specificity of each environmental site to be taken into account.

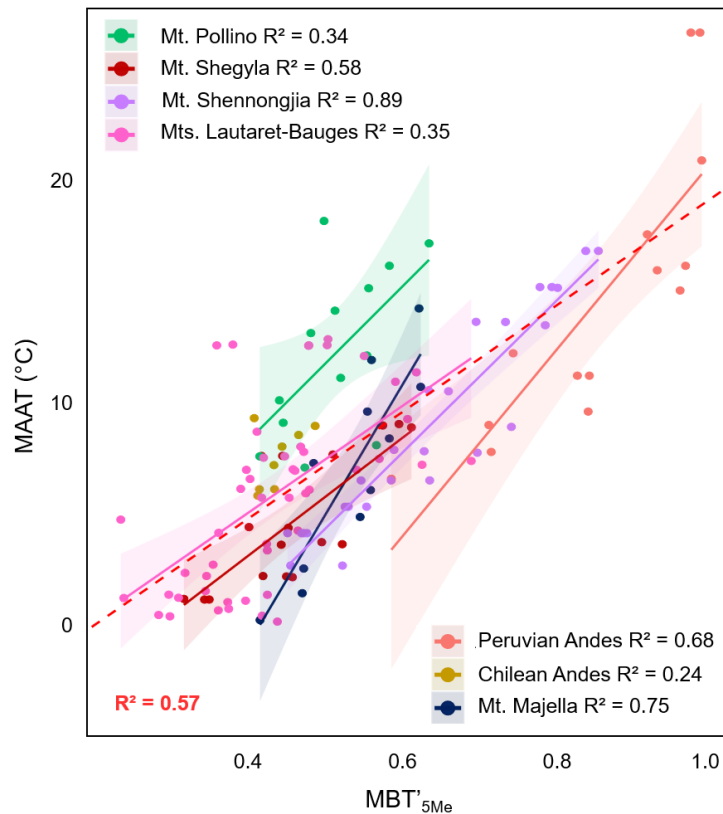


Figure 4.6 Linear regressions between (a) MAAT and MBT'_{5Me} along seven of the eight altitudinal transects investigated. Data from Mt. Rungwe (Coffinet et al., 2014), for which 5- and 6-methyl brGDGTs were not separated, were not included in this graph. Dotted lines represent the 95% confidence interval for each regression and colored areas represent the 95% confidence interval for each regression. Data from Mt. Majella were taken from Huguet et al. (2019). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021). Data from Mt. Shennongjia were taken from Yang et al. (2015). The global soil calibration by De Jonge et al. (2014) was applied to all these transects. Only significant regressions ($p < 0.05$) are shown.

4.3.2. Development of new models for the reconstruction of MAAT and pH from 3-OH FAs

Several complementary methods were recently used to derive calibrations with environmental parameters from organic proxies. Most calibrations between lipid distribution and environmental variables were based on simple linear regression models, most often the ordinary least square regression (e.g. for brGDGTs: De Jonge et al., 2014; C. Wang et al., 2016), as it is simple and easy to implement and understand. Other linear models, such as Deming regression (Naafs et al., 2017a) or Bayesian regression (Tierney and Tingley, 2014; Dearing Crampton-Flood et al., 2020), were also used. Nevertheless, these single linear regression methods rely on a given index (e.g. MBT'_{5Me} or CBT' for brGDGTs) which is correlated with environmental parameters. This represents a limitation, as the relative distribution of bacterial lipids can be concomitantly influenced by several environmental parameters (e.g. Véquaud et al., 2021) and can also depend on the microbial diversity and sources of these compounds (Parker et al., 1982; Bhat and Carlson, 1992; Zelles, 1999). In contrast, using bacterial relative abundances rather than a single index in the relationships with environmental variables appears less restrictive and more representative of the environmental complexity. Other models can be used in this way, such as those based on multiple regressions (e.g. Peterse et al., 2012; De Jonge et al., 2014; Russell et al., 2018), describing the relationships between one or several explained variables (e.g. bacterial lipid abundances) and one or several explanatory variables (e.g. MAAT, pH). Multiple regressions can reveal the presence of linear relationships among several known variables but cannot take into account non-linear influences, which may occur in complex environmental settings. This limitation, common to all linear models, can be overcome using non-parametric methods such as some of the machine learning algorithms (e.g. nearest neighbours or random forest; Dunkley Jones et al., 2020). The reliability of the latter models lies in the fact that they are non-linear, which helps capture the intrinsic complexity of the environmental setting, and that they avoid the regression dilution phenomenon observed in most linear models. Moreover, their robustness is improved by the fact that they are built on a randomly defined proportion of the total dataset and then tested on the rest of the dataset, considered to be

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

independent. Last, these machine learning algorithms are flexible and are continuously evolving when adding new samples.

As shown in section 4.3.1., robust global calibrations between 3-OH-FA-derived indices (RIAN, RAN₁₅ and RAN₁₇) and MAAT/pH could not be established using a simple linear regression model, contrary to what was observed with brGDGT-derived indices. Therefore, three different independent and complementary models were tested to potentially establish stronger statistical relationships between 3-OH FA distributions and pH/MAAT at the global level : (i) a parametric model – multiple linear regression; (ii) two non-parametric models – random forest (e.g. Ho, 1995; Denisko and Hoffman, 2018) and k-NN algorithms (e.g. Gangopadhyay et al., 2009). As discussed above, the multiple linear regression model allows the determination of linear relationships between MAAT/pH and the individual relative abundances of 3-OH-FAs, instead of indices derived from the latter. As for the two non-parametric models, they present among other things the advantage of taking into account non-linear environmental influences.

The three models, based on a supervised machine learning approach, were applied to the total soil dataset ($n=168$). All the 3-OH FA homologues, whatever their abundance, were included in the models to keep the maximum variability and take into account the specificity and complexity of each altitudinal transect. Indeed, the nature of the individual 3-OH FAs whose fractional abundance is mainly influenced by MAAT/pH may be site-dependent, as previously observed (Véquaud et al., 2021). The performances of these three models were compared with those of the linear calibrations between 3-OH FA-derived indices (RAN₁₅, RAN₁₇, RIAN) and MAAT/pH (Table 4.2).

	Model	n (training)	n (test)	R ²	RMSE	Variance in residuals	Mean absolute error	Lower estimation limit	Upper estimation limit
MAAT (°C)	RAN ₁₅	-	168	0.37	5.5	29.8	4.0	-3.1	17.2
	RAN ₁₇	-	168	0.41	5.3	27.9	3.9	-4.3	17.0
	k-NN	128	40	0.77	3.1	9.4	2.3	0.5	25.0
	Multiple linear regression	128	40	0.79	3.0	9.2	2.3	-1.2	25.8
	Random forest	128	40	0.83	2.8	8.0	2.2	0.8	24.9
pH	RIAN	-	168	0.34	1.0	1.0	0.8	4.1	7.9
	k-NN	128	40	0.70	0.7	0.5	0.5	3.4	8.7
	Multiple linear regression	128	40	0.64	0.8	0.6	0.6	4.0	8.3
	Random forest	128	40	0.68	0.7	0.5	0.5	3.5	7.8

Table 4.2 Characteristics of the different models proposed in this study to estimate MAAT and pH: R², RMSE, variance of the residuals, mean absolute error (MAE) and the upper and lower limits of estimation. The "training" samples were used to develop the different machine learning models, which were then tested on a "test" sample set.

4.3.2.1. Temperature calibrations

The multiple linear regression model yielded a strong relationship between 3-OH FA relative abundances and MAAT (Fig. 4.7a, Eq.4.1):

$$\begin{aligned} \text{MAAT } (^{\circ}\text{C}) = & -59.02 \times [nC_{10}] + 102.1 \times [iC_{11}] + 2628.49 \times [aC_{11}] - 165.58 \times [nC_{11}] - \\ & 79.799 \times [nC_{12}] + 89.93 \times [iC_{13}] + 205.06 \times [aC_{13}] - 136.25 \times [nC_{13}] - 309.71 \times [iC_{14}] \\ & - 43.16 \times [nC_{14}] - 9.27 \times [iC_{15}] - 308.53 \times [aC_{15}] + 66.06 \times [nC_{15}] - 60.57 \times [iC_{16}] + \\ & 15.53 \times [nC_{16}] + 13.52 \times [iC_{17}] - 228.76 \times [aC_{17}] - 91.12 \times [nC_{17}] + 43.7 \quad (n = 168; R^2 \\ & = 0.79; \text{RMSE} = 3.0 \text{ } ^{\circ}\text{C}) \end{aligned} \quad (4.1)$$

This model, which takes into account the whole suite of 3-OH-FAs (C₁₀-C₁₈), presents a higher strength than the global linear relationships between 3-OH FA derived indices and MAAT (R²=0.37 and 0.41, RMSE =5.5°C and 5.3°C for RAN₁₅ and RAN₁₇, respectively; Table 4.2). The multiple linear regression also improves the accuracy and robustness of MAAT prediction in comparison with single linear relationships, with lower RMSE (3.0 °C), variance in the residuals (9.2 °C; Fig. 4.7d) and mean absolute error (MAE; 2.3 °C) than with the RAN₁₅ and RAN₁₇ calibrations (RMSE of 5.5 and 5.3 °C; variance of 29.8 and 27.9 °C; MAE of 4.0 and 3.9 °C for RAN₁₅ and RAN₁₇, respectively; Table 4.2).

Similarly to the multiple linear regression model (Fig. 4.7a), the random forest (Fig. 4.7b) and k-NN (Fig. 4.7c) calibrations are characterized by strong determination coefficients (R² 0.83 and 0.77, respectively). The variance in residuals, MAE and RMSE of the random forest calibration are slightly lower than those of the multiple linear regression and k-NN models (Table 4.2). An advantage of the random forest algorithm lies in the fact that the weight of the different variables used to define the model can be quantified using the permutation importance method (Breiman, 2001). The *a*-C₁₅, *i*-C₁₄, *a*-C₁₇, *n*-C₁₂ and *n*-C₁₅, as well as to a lesser extent *n*-C₁₇, *n*-C₁₆ and *i*-C₁₃ 3-OH FAs were observed to be the homologues predominantly used by the model to estimate MAAT values (Fig. 4.9a). They include all the 3-OH FAs involved in the calculation of the RAN₁₅ and RAN₁₇ indices, especially the *a*-C₁₅ homologue. This may explain why linear relationships between the RAN₁₅/RAN₁₇ and MAAT could be established along some of

DEVELOPMENT OF GLOBAL TEMPERATURE AND PH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

the altitudinal transects investigated until now (Wang et al., 2016; Huguet al., 2019; Véquaud et al., 2021; this study). Nevertheless, other individual 3-OH FAs than those appearing in the calculation of the RAN_{15} and RAN_{17} also have a major weight in the random forest model and seem to be influenced by temperature changes, explaining the moderate determination coefficients of the global RAN_{15}/RAN_{17} -MAAT linear relationships observed in this study.

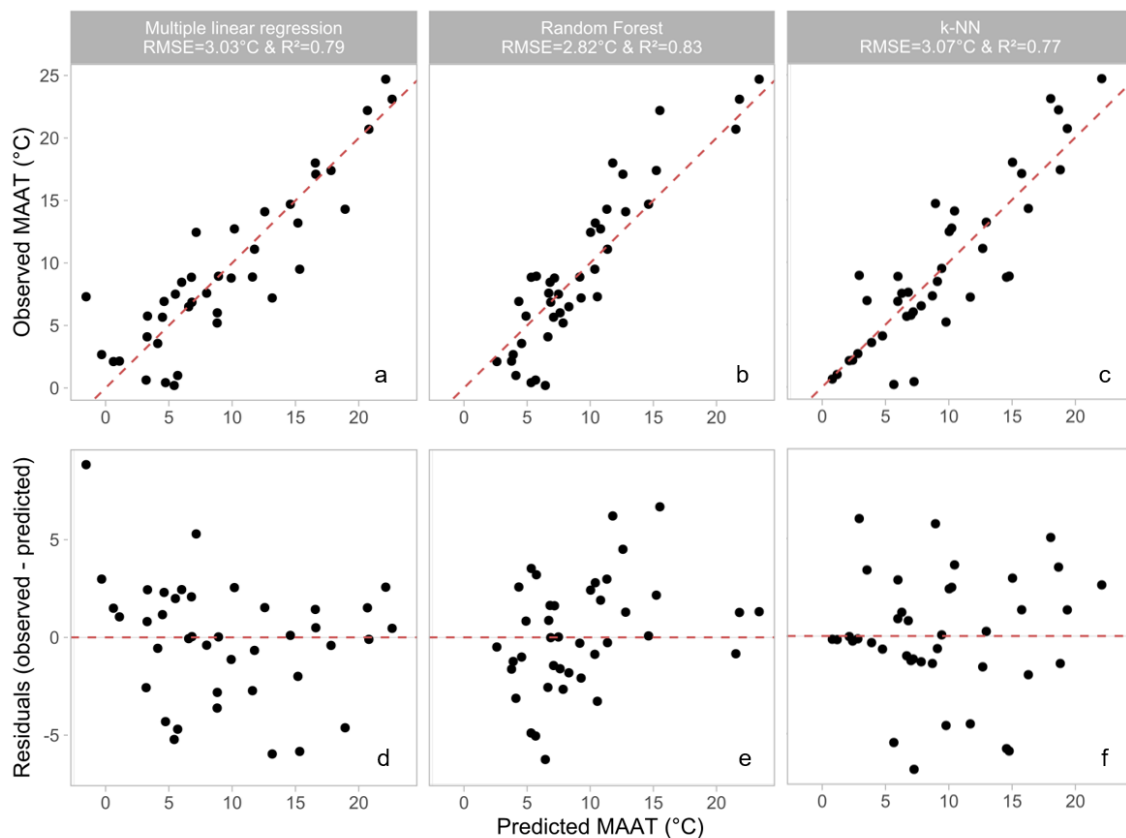


Figure 4.7 Results of the three different models tested to reconstruct the MAAT from 3-OH FA distribution: observed MAAT (°C) vs Predicted MAAT (°C) for (a) the multiple linear regression model, (b) the random forest model and (c) the k-NN method. MAAT residuals plotted against the predicted MAAT for (d) the multiple linear regression model, (e) the random forest model and (f) the k-NN method.

On the whole, the strength and accuracy of the multiple linear regression, k-NN and random forest models are much higher than those based on the RAN₁₅ and RAN₁₇ indices (Table 4.2). This is likely related to the fact that the three aforementioned models integrate the whole suite of 3-OH-FAs homologues (C₁₀ to C₁₈) and thus better capture the complexity of the response of soil Gram-negative bacteria and their lipid distribution to temperature changes than the RAN₁₅ and RAN₁₇ indices. They also present the advantage of increasing the range of temperature which may be predicted by more than 4 °C in comparison with the RAN₁₅ and RAN₁₇ calibrations (Table 4.2). Indeed, even though the lower limit of MAAT estimates for the three models tested in the present study is slightly higher than those based on the RAN₁₅ and RAN₁₇ indices, the upper limit of the MAAT which can be estimated using the multiple linear regression, random forest and k-NN models is substantially higher (ca. 25 °C) than that based on the RAN₁₅ or RAN₁₇ indices (ca. 17 °C; Table 4.2).

The three proposed models show the potential of 3-OH-FAs as MAAT proxies at the global level, which was not visible using RAN₁₅ and RAN₁₇ indices. The non-parametric models (random forest and k-NN) may benefit from the fact that they take into account the complex, non-linear relationships between environmental parameters and bacterial lipid abundance. This is highlighted when comparing the independent variations of the individual 3-OH FA relative abundances with estimated MAAT for the three proposed models, with non-linear trends for the k-NN and random forest models, in contrast with the multiple linear regression (Annex 2; Supp. Fig. 4.2).

4.3.2.2. *pH calibrations*

A robust linear relationship between the RIAN and pH could not be obtained from the whole soil dataset (Fig. 4.4a; Table 4.2). In contrast, the multiple regression model provided a strong correlation between the 3-OH FA fractional abundances and pH (Fig. 4.8a; Eq. 4.2):

$$pH = -1.45 \times [nC_{10}] - 31.70 \times [iC_{11}] - 162.09 \times [aC_{11}] - 53.22 \times [nC_{11}] - 6.21 \times [nC_{12}] + 56.24 \times [iC_{13}] - 2.02 \times [aC_{13}] + 15.10 \times [nC_{13}] + 23.99 \times [iC_{14}] - 4.54 \times [nC_{14}] -$$

**DEVELOPMENT OF GLOBAL TEMPERATURE AND PH CALIBRATIONS BASED ON
BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS**

$$13.79 \times [iC_{15}] - 15.74 \times [aC_{15}] + 1.93 \times [nC_{15}] - 46.29 \times [iC_{16}] - 3.20 \times [nC_{16}] - 1.80 \times [iC_{17}] - 8.90 \times [aC_{17}] + 11.46 \times [nC_{17}] - 3.63 \times [nC_{18}] + 7.84 \quad (n = 168; R^2 = 0.64; RMSE = 0.8) \quad (4.2)$$

The random forest (Fig. 4.8b) and k-NN pH models (Fig. 4.8c) appeared to be slightly more robust and accurate than the multiple linear regression (Fig. 4.8a), as the former two models presented slightly higher determination coefficients ($R^2 = 0.68$ and 0.70 for k-NN and random forest, respectively) and slightly lower RMSE (0.7), variance in residuals (0.5) and MAE (0.5) than the multiple linear regression (Table 4.2).

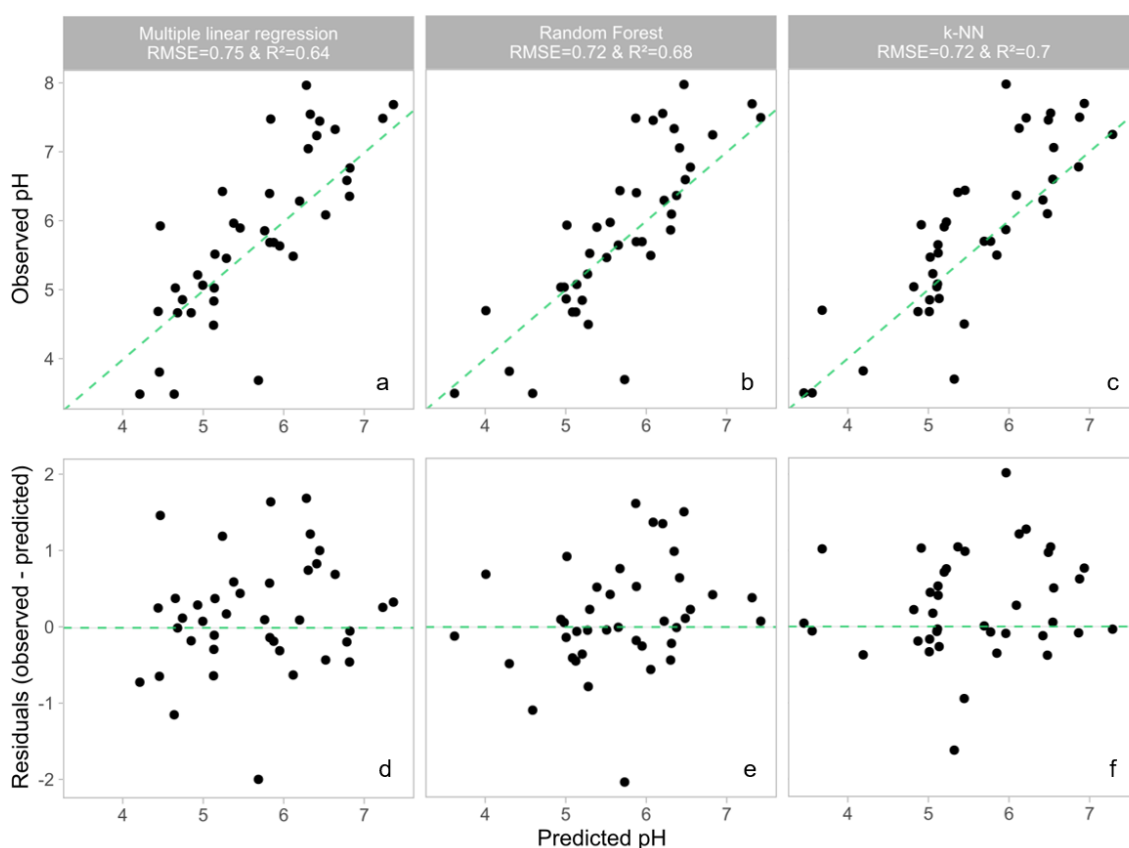


Figure 4.8 Results of the three different models tested to reconstruct the pH from 3-OH FA distribution: observed pH vs predicted pH for (a) the multiple linear regression model, (b) the random forest model and (c) the k-NN method. The pH residuals plotted against the predicted pH for (d) the multiple linear regression model, (e) the random forest model and (f) the k-NN method.

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

As for the MAAT random forest model, the weight of the individual 3-OH FAs in the pH random forest calibration was determined (Fig. 4.9b). Three homologues— *i-C*₁₃, *n-C*₁₅, *i-C*₁₆— had a larger weight in the global pH model than the others (Fig. 4.9b). This is consistent with a detailed study of 3-OH FA distribution in soils from the French Alps (Véquaude et al., 2021), where the *i-C*₁₃ and *i-C*₁₆ 3-OH FAs were observed to be predominantly influenced by pH. Nevertheless, in addition to the three aforementioned homologues, most of the C₁₀ to C₁₈ 3-OH FAs have a non-negligible influence in the random forest pH model, except the *a-C*₁₅ and *i-C*₁₄ compounds (Fig. 4.9b). This is in line with the definition of the 3-OH FA-based pH index (RIAN) defined by C. Wang et al. (2016) which includes the whole suite of 3-OH FAs. These results suggest that soil Gram-negative bacteria may respond to pH variations by modifying the whole distribution of 3-OH-Fas (C₁₀-C₁₈). This would need to be further confirmed by, for example, investigating the influence of pH variations on pure strains of Gram-negative bacteria isolated from soils.

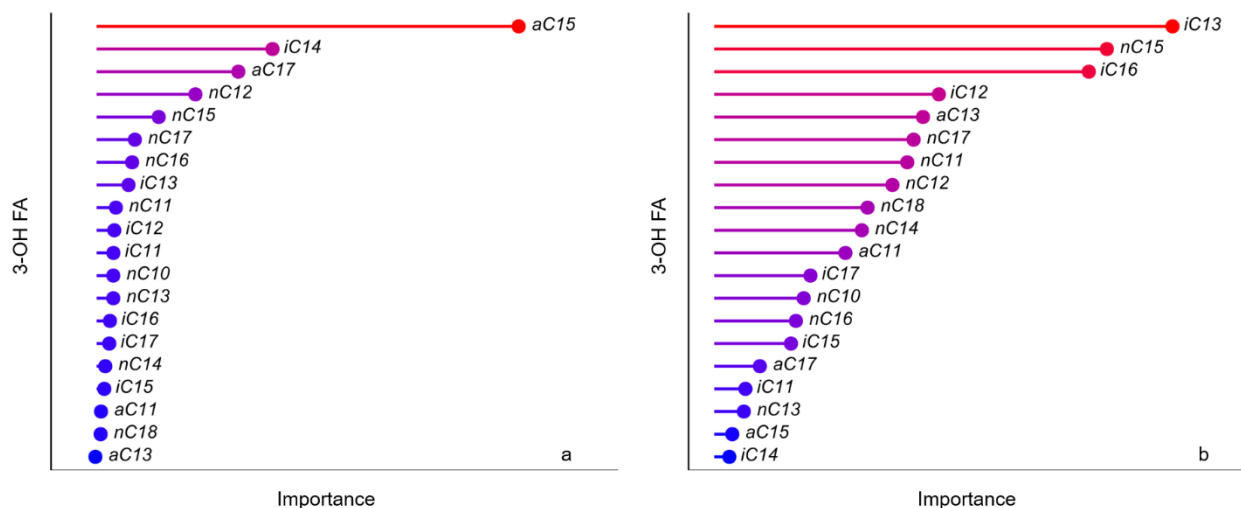


Figure 4.9 Importance (arbitrary unit) of the 3-OH FAs used to estimate (a) MAAT and (b) pH in the random forest models proposed in this study according to the permutation importance method (Breiman, 2001).

In any case, in contrast with the RIAN index, the multiple linear regression, k-NN and random forest models provided strong global calibrations with pH (Fig. 4.8), as robust as the global CBT'-pH relationship (Fig. 4.4b). The three proposed models also increase the range of pH which can be estimated (~ 4 pH units) in comparison with the RIAN global calibration (~ 3 pH units), further strengthening the potential of these models for soil pH reconstruction. As MAAT models, the independent variations in the individual 3-OH-FA relative abundances with estimated pH highlight non-linear trends for the k-NN and random forest models, in contrast with the multiple linear regression (Annex 2; Supp. Fig. 4.3), which might favour the use of the two non-parametric models in order to take into account such non-linear influences. The machine learning MAAT and pH models proposed in this paper are flexible and could be further improved by increasing the number of soil samples analysed and the representativeness of the different MAAT and pH values within the dataset.

4.3.3. Palaeoclimate application of the new 3-OH FA and MAAT models

The multiple regression, random forest and k-NN models developed for MAAT reconstruction using 3-OH FAs were similar in terms of robustness and precision (Figs. 4.7a, b, c; Table 4.2). The performance and validity of these global terrestrial calibrations for palaeotemperature reconstructions was thus tested and compared with the MAAT record from a Chinese speleothem (HS4 stalagmite) covering the last 9,000 years BP (Wang et al., 2018). This terrestrial archive was the object of previous palaeostudies, thus providing a context for the interpretation of the MAAT data and, to the best of our knowledge, represents the only published application of 3-OH FAs as a palaeotemperature proxy in terrestrial settings (Wang et al., 2018). The local comparison of 3-OH FA distributions in the overlying soils and stalagmites and the analyses of bacterial diversity and transport pathways suggested that the 3-OH FAs in the HS4 speleothem were mainly soil-derived (Wang et al., 2018), supporting the application of soil calibrations for MAAT reconstruction from this archive, although not being a palaeosoil itself. The first palaeoapplication of 3-OH FAs (Wang et al., 2018) on this speleothem relied on a local calibration between the RAN₁₅ index and MAAT proposed

by C. Wang et al. (2016) using soils from Mt. Shennongjia. The MAAT estimates derived from our global soil calibrations were compared with those obtained from this local soil calibration (C. Wang et al., 2016).

4.3.3.1. Comparison of the multiple linear regression, k-NN and random forest global MAAT calibrations

The multiple regression model (Eq. 4.1; Fig. 4.7a) yielded MAAT estimates ranging between -35 and 22.8 °C over the last 9,000 years (Annex 2; Supp. Fig. 4.4). The temperature minimum (-35°C) observed at 560 yrs BP can be considered to be an outlier, with a significantly lower MAAT estimate than those provided by the other samples. After having ignored this apparent outlier, the MAAT range over the last 9,000 years was comprised between 3.2°C and 22.8°C, with temperature shifts of up to 15 °C within very short periods of time. The observed range of MAAT and large variations in temperature over such short periods appear far too excessive, as the expected amplitude of MAAT during the Holocene is expected to be up to ca. 2-3 °C (Liu et al., 2014). This highly questions the reliability of the multiple linear regression model for MAAT reconstruction from this archive.

MAAT estimates derived from the k-NN calibration ranged between 6.5 and 19.7 °C over the last 9,000 years (Annex 2; Supp. Fig. 4.4). Abrupt shifts in MAAT of more than 10 °C were observed between 2,000 and 4,000 yrs BP. Such variations, higher than the RMSE of the calibration, appear excessive for the Holocene period, as previously discussed for the multiple regression model. The bias in MAAT estimates may be due to the intrinsic definition of the k-NN model, which is better suited for uniformly distributed datasets. This is not the case here, as the individual transects heterogeneously cover a wide range of temperatures. The application of a global calibration at the local scale – that of the HS4 stalagmite – using the k-NN method and based on the similarities among samples, thus does not appear appropriate. Such a calibration might be improved by extending the dataset with samples more equally distributed across a wider range of global climatic gradients.

Finally, the random forest model yielded MAAT estimates between 10.6 and 19.3°C, i.e. a much smaller range than the k-NN algorithm and multiple regression model (Annex 2; Supp. Fig. 4.4). The amplitude of the shifts observed between 2,000 and 4,000 yrs BP was ca. 4°C, which is climatically more consistent than the variations obtained with the k-NN method and multiple regression model, even though these large variations in MAAT over such short periods of time still appear too excessive. Furthermore, the application of the global random forest calibration roughly provided similar temperature trends as those derived from the local RAN₁₅ calibration by Wang et al. (2018; Fig. 4.10), despite some of the largest oscillations for the global model. These results suggest that the random forest calibration is more reliable than the multiple regression and k-NN ones. This can be explained by the intrinsic definition of the random forest algorithm, which averages the results of several independent models (so-called decision trees), thus reducing the variance and thus the forecast error in the final model. This is also in line with the slightly higher accuracy of the random forest calibration compared with the other two models (Table 4.2), as previously discussed. In contrast, the multiple regression calibration was the less performant of the three models in the investigated archive. This may be related to its parametric nature and the fact that it does not take into account the natural non-linear variations on 3-OH FA fractional abundances highlighted by the random forest and k-NN models (Annex 2; Supp. Figs. 4.2 and 4.3).

In conclusion, the three models proposed in this study, especially the random forest, have potential for MAAT reconstruction, even though the application to a well-known palaeoclimate archive showed their limitations. This highlights the importance of testing new calibrations on well-characterized archives to investigate their reliability.

4.3.3.2. Comparison of the global random forest and local RAN₁₅ calibrations for MAAT reconstruction

The random forest model was observed to be the most reliable of the three proposed global MAAT calibrations (Fig. 4.7). To go further, we compared the temperature record derived from our global random forest calibration with that derived from the local MAAT/RAN₁₅ transfer function proposed by C. Wang et al. (2016; Fig. 4.10). The application of the local RAN₁₅ calibration to the HS4 stalagmite yielded an average

DEVELOPMENT OF GLOBAL TEMPERATURE AND PH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

MAAT of ca. 18.4 °C over the most recent part of the record (last 800 yrs; Fig. 4.10), consistent with the MAAT of 18 °C recorded *in situ* by a temperature logger (Hu et al., 2008; Wang et al., 2018). In contrast, absolute MAAT estimates derived from the random forest model were on average 14.2 °C over the last 800 yrs and were generally lower than those obtained from the local RAN₁₅ calibration over the whole record. Altogether, these results suggest that the random forest model tends to underestimate absolute MAAT, in contrast with the RAN₁₅ calibration proposed by C. Wang et al. (2016). This discrepancy may be due to the fact that the calibration proposed in the present study is based on a global dataset, with samples subject to a large variety of environmental and climatic conditions, whereas the RAN₁₅-MAAT transfer function by C. Wang et al. (2016) was constructed using soil samples from a regional altitudinal transect, located at only 120 km distance from the stalagmite site (Wang et al., 2018).

Even though the local calibration by C. Wang et al. (2016) provides more accurate absolute MAAT values than the present global random forest model, as it could be expected, both calibrations roughly generate similar qualitative MAAT trends over time. A regular slight decrease in temperature of ca. 1 °C was observed between 9,000 and ca. 1,000 yrs BP based on the local RAN₁₅ calibration (Fig. 4.10a; Wang et al., 2018). This general decreasing trend was also visible when using the random forest model, but with larger oscillations and mainly between 9,000 and 4,000 yrs BP, in agreement with the general trend recorded by the $\delta^{18}\text{O}$ record (mixture of temperature and hydrological signals, Wang et al., 2018) of the HS4 stalagmite (Fig. 4.10c,d; Hu et al., 2008). In addition, both the global random forest, local RAN₁₅ calibrations and the $\delta^{18}\text{O}$ record allowed the identification of several climatic events in the Northern Hemisphere, in agreement with the reconstructed total solar irradiance (TSI, Steinhilber et al., 2009, Fig. 4.10e). Thus, both models highlighted, with slightly different amplitudes, the Medieval Warm Period (MWP, 800-1000 years BP) and Little Ice Age (LIA; 200-500 years BP) periods (Mann et al., 2008; Ljungqvist, 2010; Wang et al., 2018). The LIA event is particularly well represented by the global random forest calibration (Fig. 4.10b), in line with the decrease in the TSI (Fig. 4.10c, e) associated with a relative increase in the $\delta^{18}\text{O}$ of HS4 carbonates (dry-cool event, Wang et al., 2018). Before the MWP, the global random forest calibration shows slight oscillations, which can be assumed to be representative of TSI variations between 500 and 1,300 yrs BP. Similarly, an important

cooling event, well correlated with a significant decrease in the TSI (Fig. 4.10a, c, e), was recorded by the two calibrations at 1300 yr BP.

The global random forest calibration also highlighted two cooling events, poorly represented by the local RAN₁₅ calibration: one at ca. 4,200 yrs BP ago and, to a lesser extent, another one between 2,800 and 3,000 yrs BP (Bond et al., 2001; Mayewski et al., 2004). The event at 4,200 yrs BP is consistent with the $\delta^{18}\text{O}$ and solar irradiance records and is referenced in the literature as the "4.2 kiloyear event" (deMenocal, 2001). This intense drought event was suggested to have had a major impact on different civilizations (collapses, migrations; (Gibbons, 1993; Staubwasser et al., 2003; Li et al., 2018; Bini et al., 2019). Thus, in some parts of China, the production of rice fields sharply decreased during this period, leading to a decrease in population (Gao et al., 2007).

Both calibrations additionally show a cooling period between 4,000 yrs and 3,200 yrs BP, more pronounced based on the global random forest model, followed by another cooling between 3,200 years BP and 3,000 yrs BP. This cooling period is consistent with the trends derived from $\delta^{18}\text{O}$ and solar irradiance records. It culminates with a cold episode at 3000 yrs BP, also known as Late Bronze Age Collapse (Kaniewski et al., 2013). Indeed, this cold episode, combined with droughts, may have led to a decrease in agricultural production in China, contributing to the degradation of trade routes and ultimately to the collapse of Bronze Age civilizations (Weiss, 1982; Knapp and Manning, 2016). Last, the global random forest calibration also highlights two additional cold events, between 5,600 and 5,900 yrs BP, as well as around 7,100 yrs BP, corresponding to solar irradiance minima (Bond et al., 2001; Mayewski et al., 2004) and which are not as clearly visible with the local RAN₁₅ calibration by Wang et al. (2016).

The first application of the random forest calibration to a natural archive shows the potential of 3-OH FAs as palaeotemperature proxies at a global scale, as known and documented climatic events were recorded, with a similar RMSE (2.8 °C; Table 4.2) as that of the local calibration by Wang et al. (2.6 °C; 2016). This RMSE is also much lower than the one related to the latest global MAAT-brGDGT calibrations (> 4 °C; De Jonge et al., 2014; Naafs et al., 2017a; Dearing Crampton-Flood et al., 2020), even though the latter are based on a larger number of soil samples than the global 3-OH FA model proposed in the present study. In summary, we demonstrate that 3-OH FAs are promising and effective temperature proxies for terrestrial settings, complementary to, and

DEVELOPMENT OF GLOBAL TEMPERATURE AND pH CALIBRATIONS BASED ON BACTERIAL 3-HYDROXY FATTY ACIDS IN SOILS

independent of, the brGDGTs, and also highlight the usefulness of non-parametric models using machine learning, especially the random forest algorithm, to establish global MAAT calibrations. We expect that analyses of 3-OH FAs in a larger number of globally distributed soils will further improve the accuracy and robustness of the global random forest calibration for palaeotemperature reconstruction. Additional palaeoapplications are also required to further test and validate the applicability of the global MAAT and pH calibrations based on 3-OH FAs presented in this study.

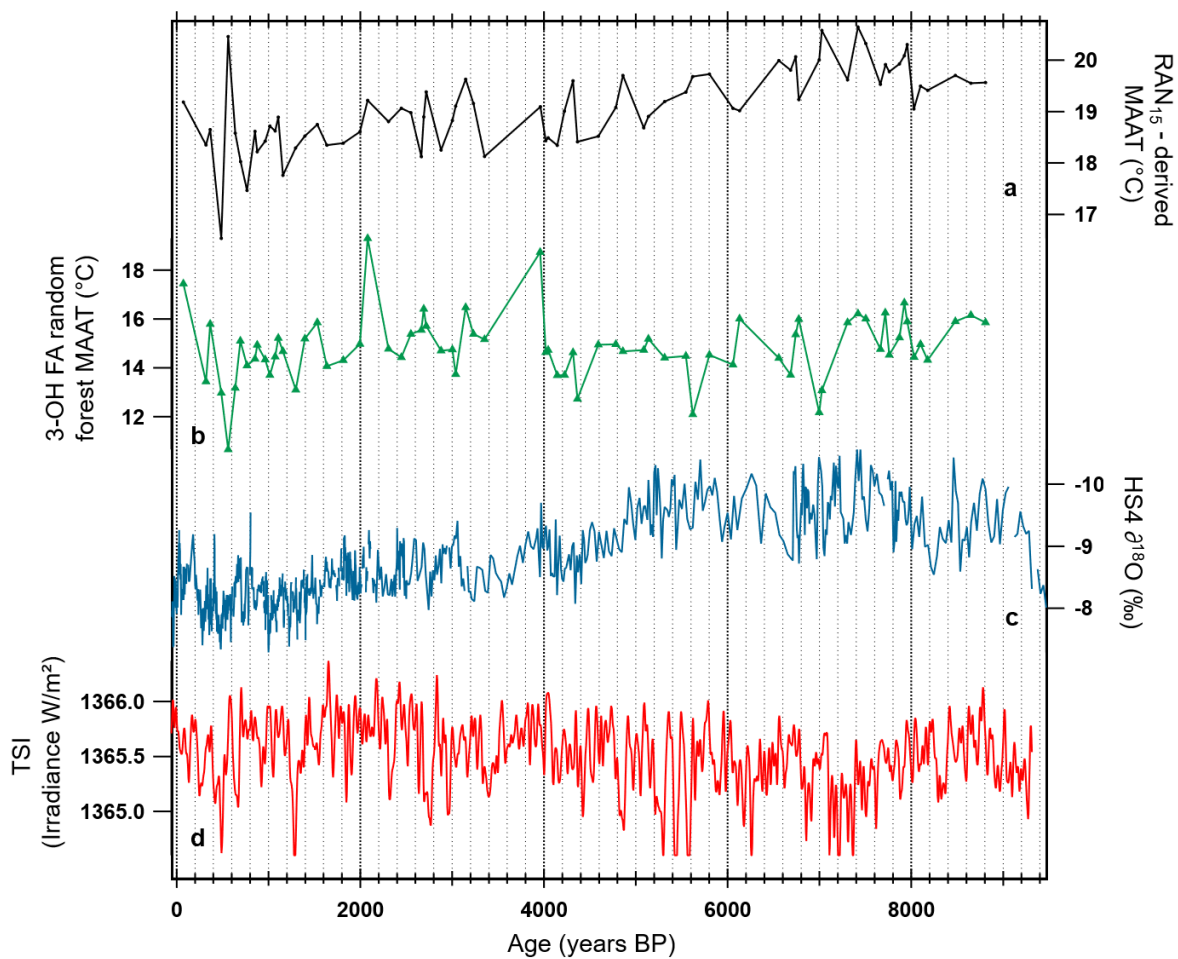


Figure 4.10 Comparison of the 3-OH FA model-MAAT record with other time series and proxy records for the HS4 speleothem (Wang et al., 2018). (a) RAN₁₅-MAAT record reconstructed using a local Chinese calibration (C. Wang et al., 2016; Wang et al., 2018). (b) 3-OH FA random forest model-MAAT. (d) The CaCO₃ oxygen isotope record (Hu et al., 2008b). (e) Total solar irradiance (TSI; W/m²) during the Holocene (past 9300 years) based on a composite described in Steinhilber et al. (2009).

4.4. Conclusions

The 3-OH FAs have been recently proposed as environmental proxies in terrestrial settings, based on local studies. This study investigated for the first time the applicability of these compounds as MAAT and pH proxies at the global scale using an extended soil dataset across a series of globally distributed elevation transects ($n = 168$). Strong linear relationships between 3-OH FA-derived indices (RAN_{15} , RAN_{17} and RIAN) and MAAT/pH could only be obtained locally, for some individual transects, suggesting that these indices cannot be used as palaeoproxies at the global scale through this kind of model. Other algorithms (multiple linear regression, k-NN and random forest models) were tested and, in contrast with simple linear regressions, provided strong global correlations between MAAT/pH and 3-OH FA relative abundances. The applicability of these three models for palaeotemperature reconstruction was tested and compared with the MAAT record from the unique available record: a Chinese speleothem. The calibration based on the random forest model appeared to be the most robust and showed similar trends to previous reconstructions and known Holocene climate variations. Furthermore, the global random forest model highlighted documented climatic events poorly represented by the local RAN_{15} calibration. This new global model is promising for palaeotemperature reconstructions in terrestrial settings and could be further improved by analysing 3-OH FAs in a larger number of globally distributed soils. This study demonstrates the major potential of 3-OH FAs as MAAT/pH proxies in terrestrial environments through the different models presented and their application for palaeoreconstruction.

**Chapitre 5 – FROG: IMPROVEMENT
OF THE GLOBAL BRANCHED TETRAETHER
TEMPERATURE CALIBRATION IN
TERRESTRIAL SETTINGS USING MACHINE
LEARNING**

Le chapitre 5 est présenté sous forme d'un article en préparation pour *Geochimica et Cosmochimica Acta*. Il est proposé de développer une nouvelle calibration globale permettant d'améliorer la précision des reconstructions de température en milieu terrestre, à partir des données de brGDGTs disponibles dans la littérature pour l'ensemble des échantillons de sols et de tourbières ($n=767$). Les outils statistiques développés dans les chapitres précédents ont été utilisés afin de mieux comprendre les paramètres environnementaux influençant la distribution des brGDGTs en milieu terrestre.

5.1. Introduction

Investigating past climate variations is essential to understand and predict future environmental changes, especially in the context of global anthropogenic changes. To this aim, "indirect" indicators of past climates – so-called proxies – have been developed and used regularly since the last century, including those based on microbial lipids. Microorganisms are capable to modify the lipid composition of their membranes to maintain a functional fluidity and permeability of the latter. The temperature and pH of the microorganism's living environment are considered to be the predominant factors influencing the membrane lipid distribution (Lauber et al., 2009; Siles and Margesin, 2016; Hofmann et al., 2016; Shen et al., 2019).

Among microbial lipids, branched glycerol dialkyl glycerol tetraethers (brGDGTs) have been increasingly used as temperature and pH proxies in continental settings over the last 15 years. These membrane lipids are produced by still unidentified bacteria, although some of them may belong to the phylum *Acidobacteria* (Sinninghe Damsté et al., 2011, 2014, 2018). They are ubiquitous in terrestrial (Weijers et al., 2007; Peterse et al., 2012; De Jonge et al., 2014; Naafs et al., 2017a) and aquatic environments (Blaga et al., 2009; Peterse et al., 2009, 2015; Damsté et al., 2009; Tierney and Russell, 2009; Loomis et al., 2012; Weber et al., 2015). The analysis of brGDGTs, based on a large number of soils distributed worldwide, showed that the relative distribution of these compounds is mainly related to mean annual air temperature (MAAT) and soil pH (Weijers et al., 2007; Peterse et al., 2012; De Jonge et al., 2014). The average number of pentane cycles, reflected in the cyclisation of branched tetraethers (CBT) index (Weijers et al., 2007; Peterse et al., 2012), has been correlated with soil pH, while the average number of methyl groups, referred to as the MBT (Methylation Index of Branched

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

Tetraethers) index, has been initially correlated with mean annual mean air temperature (MAAT) and, to a lesser extent, soil pH (Weijers et al., 2007 ($r^2=0.77$; RMSE =4.8°C, $n=134$); Peterse et al., 2012 ($r^2=0.59$; RMSE=5.0°C; $n=176$). More recently, new brGDGT isomers have been detected through improved analytical methods, with methyl groups being present in either 5th, 6th, 7th or 8th position (De Jonge et al., 2013, 2014; Hopmans et al., 2016; Ding et al., 2016). It was observed that 6-methyl isomers were strongly and predominantly dependent on soil pH and 5-methyl brGDGT on temperature (De Jonge et al., 2014). This led to the development of a new MBT index excluding 6-methyl isomers - the MBT'_{5Me} index - which correlates preferentially with MAAT ($r^2 = 0.64$; RMSE= 4.8°C; $n = 231$).

Hence, brGDGTs have emerged over time as robust temperature and pH paleoproxies in multiple types of settings – lakes (Blaga et al., 2009; Powers et al., 2010; Huguet et al., 2010; Fawcett et al., 2011; Lu et al., 2016; Coffinet et al., 2018; Wang et al., 2020), peatlands (Weijers et al., 2011b; Coffinet et al., 2018; Wu et al., 2020), soil/paleosols (Ding et al., 2015 ; Lu et al., 2016 ; Feng et al., 2019 ; Wang et al., 2020) and speleothems (Baker et al., 2019). Nevertheless, numerous studies showed that additional parameters other than temperature and pH, such as soil moisture, thermal regime or vegetation cover (Dirghangi et al., 2013; Menges et al., 2014; Huguet et al., 2014; Liang et al., 2019; De Jonge et al., 2019), can also influence the relative distribution of brGDGTs in terrestrial settings (e.g. Weijers et al., 2011b; Huguet et al., 2013; Davtian et al., 2016; Mueller-Niggemann et al., 2016; Dang et al., 2016; Liang et al., 2019). This can explain a large part of the residual scatter in the global brGDGT terrestrial calibrations with MAAT and pH (De Jonge et al., 2014; Naafs et al., 2017a, b; Dearing Crampton-Flood et al., 2020).

To overcome these limitations, refinements in the global brGDGT calibrations were proposed over the years. De Jonge et al (2014) proposed a global temperature calibration ($R^2 = 0.64$; RMSE = 4.8°C; $n = 231$) excluding the 6-methyl brGDGTs isomers. Naafs et al. (2017a) showed that a stronger correlation between MBT'_{5Me} and MAAT in soils could be obtained by excluding samples dominated by 6-methyl brGDGTs ($n = 177$; $R^2 = 0.76$; RMSE = 4.1°C). More recently, Dearing Crampton-Flood et al. (2020) used Bayesian statistics instead of more classically applied single linear regressions to investigate the relationship between MBT'_{5Me} and MAAT in soils ($n = 353$; $R^2 = 0.64$; RMSE = 6 °C).

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

The robustness of the Bayesian approach relies on the fact that (i) it considers a given index (e.g., the MBT'_{5Me} for brGDGTs) as the variable dependent on environmental parameters, consistent with the fact that bacterial lipids are produced in response to the variations of environmental parameters and that (ii) it avoids regression dilution phenomena, in contrast with most of the models based on linear regressions.

Despite improvements in brGDGT analytical methods and development of refined models, the RMSE associated with global calibrations between brGDGT distribution and MAAT in terrestrial settings remains high (> 4 °C). It should be noted that most of the previous global brGDGT calibrations in soils were based on a correlation between MAAT and a single index (i.e. MBT'_{5Me}; De Jonge et al., 2014; Naafs et al., 2017a, b; Dearing Crampton-Flood et al., 2020), even though the relative distribution of brGDGTs can be concomitantly influenced by several environmental parameters. In contrast, using bacterial relative abundances rather than a single index in models appears more representative of the environmental complexity (Wang et al., 2020; Véquaud et al., 2020; Dunkley Jones et al., 2020). Moreover, by definition, linear models cannot take into account non-linear influences, which may occur in complex environmental settings. Such a limitation can be overcome using non-parametric models such as machine-learning algorithms. Machine-learning models were very recently used to develop global calibrations between the relative abundance of isoprenoid GDGTs and sea surface temperature (SST) in marine settings (Dunkley Jones et al., 2020) and between the relative abundance of bacterial 3-hydroxy fatty acids and MAAT in soils (Véquaud et al., 2020; Wang et al., 2021). These models allow overcoming the use of a single index as they are based on the whole suite of microbial lipids. They are built on a proportion of the total dataset (randomly defined) and then tested on the rest of the dataset, considered as independent. Such an approach improves the robustness of the model and avoids the phenomenon of regression dilution. As they are non-parametric, they also capture non-linear environmental influences, in line with the intrinsic complexity of the environmental settings

This study developed a machine-learning algorithm (random forest) with the aim of proposing a new global brGDGT calibration for MAAT reconstruction in terrestrial settings with a reduced RMSE. It was based on an extended global dataset comprising 775 peat and soil samples (with 112 samples added to the global brGDGTs calibration),

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

first statistically separated into clusters, to better understand the parameters affecting brGDGT distribution in soils at the global scale and highlight potential confounding factors in addition to MAAT. The latter differ by the influence of environmental parameters – MAAT, MAP mean annual precipitation (MAP), soil pH and the number of frost days during the year (FRS) on the relative abundance of brGDGTs. This mechanistic approach then led to the development of a refined brGDGT temperature calibration (so-called FROG model) based on a random forest machine-learning algorithm and the whole suite of individual brGDGTs. Alternative models were also proposed to test the influence of confounding variables on the FROG calibration and potentially further improve its accuracy.

5.2. Results

5.2.1. Principal component analysis and clustering on the global dataset

A Principal Component Analysis (PCA) was performed on the entire brGDGT dataset (Fig. 5.1). The first 3 axes carry most the majority of the variance (70.3%; Fig. 5.1a, b, c). Consequently, the description of the analysis will be restricted to these axes. A cluster classification of the samples based on the k-means method was performed, yielding 5 clusters (Fig. 5.1d). Even though these different clusters may seem to overlap, they are well separated in an n-dimensional space which would explain 100% of the total inertia of the dataset (Fig. 5.1 a,b)

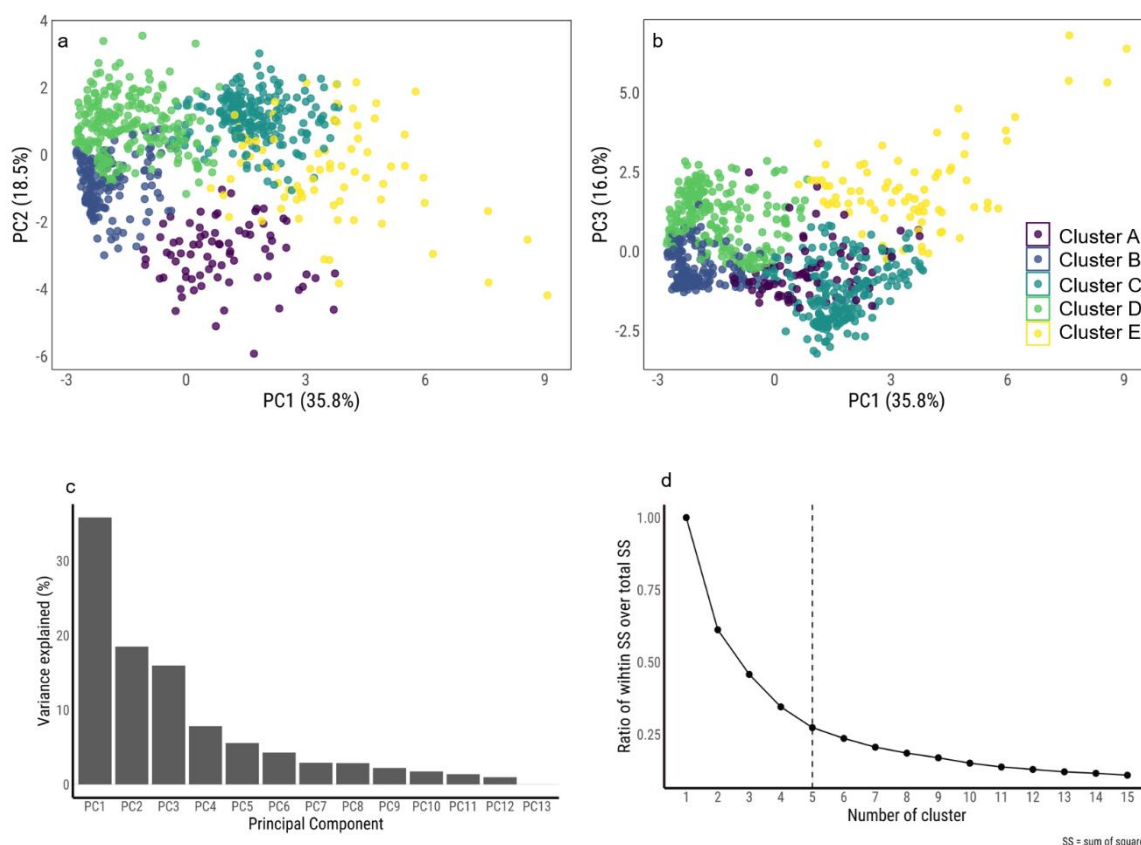


Figure 5.1 Principal Component Analysis performed on the global dataset (n=767) with the k-means clustering. (a) PC2 vs PC1, (b) PC3 vs PC1, (c) Variance explained (%) for each component, (d) Optimal number of clusters according to the k-means method.

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

The distribution of the samples between the different clusters is heterogeneous (between 76 and 230 samples). Each cluster covers a different range of pH, MAAT, MAP and FRS, with various proportions of soil and peat samples (Table 5.1).

	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E	All
Soils	73	132	207	184	75	671
Peats	3	43	2	46	2	96
Samples	76	175	209	230	77	767
Min pH	5.6	3.0	4.9	3.4	4.0	3.0
Max pH	8.1	7.9	8.7	7.5	9.3	9.3
Mean pH	7.1	5.2	7.5	5.4	7.2	6.3
Min MAAT (°C)	10.8	0.4	-6.0	-17.7	-6.7	-17.7
Max MAAT (°C)	28.0	27.4	18.0	17.4	18.5	28.0
Mean MAAT (°C)	22.4	16.0	6.7	3.9	6.5	9.5
Min MAP (mm/yr)	106	383	168	128	156	106
Max MAP (mm/yr)	2770	3584	1055	2177	1191	3584
Mean MAP (mm/yr)	1069	1311	453	786	756	855
Min Frost days	0	0	3	0	2	0
Max Frost days	11	19	29	26	27	29
Mean Frost days	2	5	15	14	13	11

Table 5.1 Quantitative description of the 5 clusters obtained after the k-means on PCA on the brGDGT distribution of the total dataset.

The clusters are well-differentiated, with different means and amplitudes for MAAT, FRS and pH (Fig. 5.2) but also on the basis of their geographical locations (Fig. 5.3). Clusters B and D contain a larger proportion of peat samples representative of acidic environments, which can explain the lower pH range compared to the other clusters (Gorham, 1991; Killops, 2005; Dedysh et al., 2006; Comont et al., 2006) (Table 5.1, Fig. 5.2). Cluster A shows samples mainly distributed over tropical and subtropical latitudes (Fig. 5.3) associated with high MAP and MAAT (Table 5.1). Cluster B samples are distributed over temperate to subtropical latitudes, with similar precipitation amounts as for samples from cluster A but lower MAAT (Fig. 5.3; Table 5.1). Samples of cluster C are mostly distributed in China and correspond to loess-paleosol samples (Fig. 5.3). Within this cluster, MAP and MAAT are on average lower than in clusters A and B, associated with a higher number of frost days during the year (Table 5.1; Fig. 5.2). The samples from clusters D and E show similar geographical distributions, mostly in the

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

northern hemisphere, at temperate latitudes, and even polar latitudes (Fig. 5.3). This results in lower MAP, colder MAAT especially for cluster D, and a high FRS, similar to cluster C. Thus, the statistical differentiation of the brGDGT dataset into different clusters reflects an environmental reality that can be described.

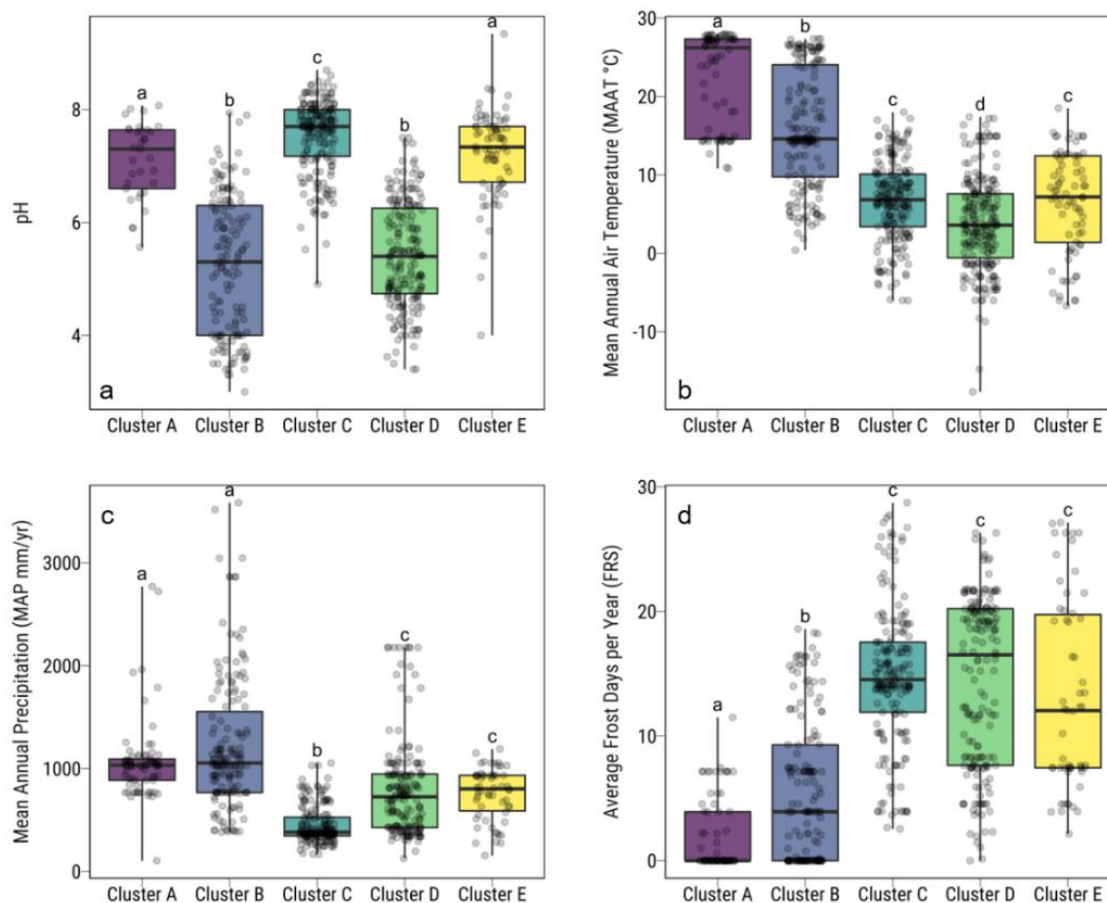


Figure 5.2 Boxplot showing the distribution of the 4 environmental variables considered: (a) pH, (b) MAAT(°C), (c) MAP (mm/year), (d) Average frost days per year (FRS). Interquartile range (IQR) = $Q_3 - Q_1$ where Q_3 is the 75th percentile and Q_1 is the 25th percentile. Outliers are defined with a 1.5 coefficient on the IQR. Letters on the panel show the differences between each cluster according to Kruskal-Wallis and Dunn post-hoc tests.

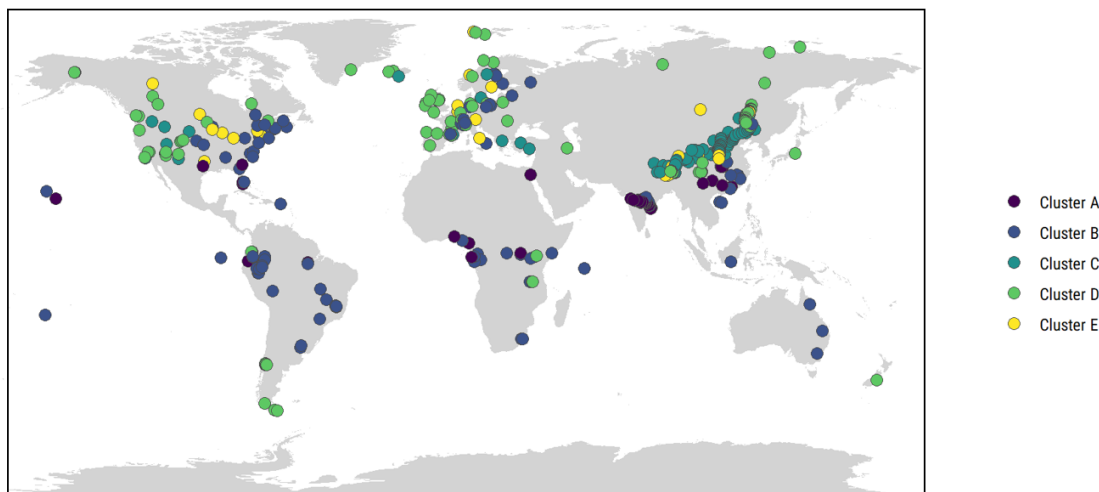


Figure 5.3 Spatial distribution of samples in the global dataset. The colors correspond to the different clusters.

5.2.2. Relationships between MBT'_{5Me} and MAAT

A strong and significant calibration between MAAT and MBT'_{5Me} is observed when considering the total soil dataset (Supp. Fig. 5.1; Eq. 5.1):

$$MAAT (^{\circ}C) = 35.98 \times MBT'_{5Me} - 12.74 \quad (n=767; R^2= 0.65, RMSE= 5.2^{\circ}C) \quad (5.1)$$

Nevertheless, this global calibration shows an important scattering. The linear regressions between the MBT'_{5Me} and MAAT were further explored for each cluster derived from the PCA analysis (Fig. 5.1). Clusters A and B show strong linear relationships ($R^2 = 0.61$ and 0.77 , respectively) between MAAT and MBT'_{5Me} (Fig. 5.4a, b) associated with improved RMSE ($3.8^{\circ}C$ and $3.6^{\circ}C$, respectively) compared to the global calibration (Supp. Fig. 5.1; Eq. 5.1). In contrast, for clusters C-D-E (Figs. 5.4c, d, e), weak or absent relationships between MBT'_{5Me} and MAAT were observed (Fig. 5.4).

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

To further investigate the influence of the proposed environmental variables (MAAT, MAP, pH, frost days per year) on the brGDGT relative abundance in the different clusters, redundancy analyses (RDA) were performed.

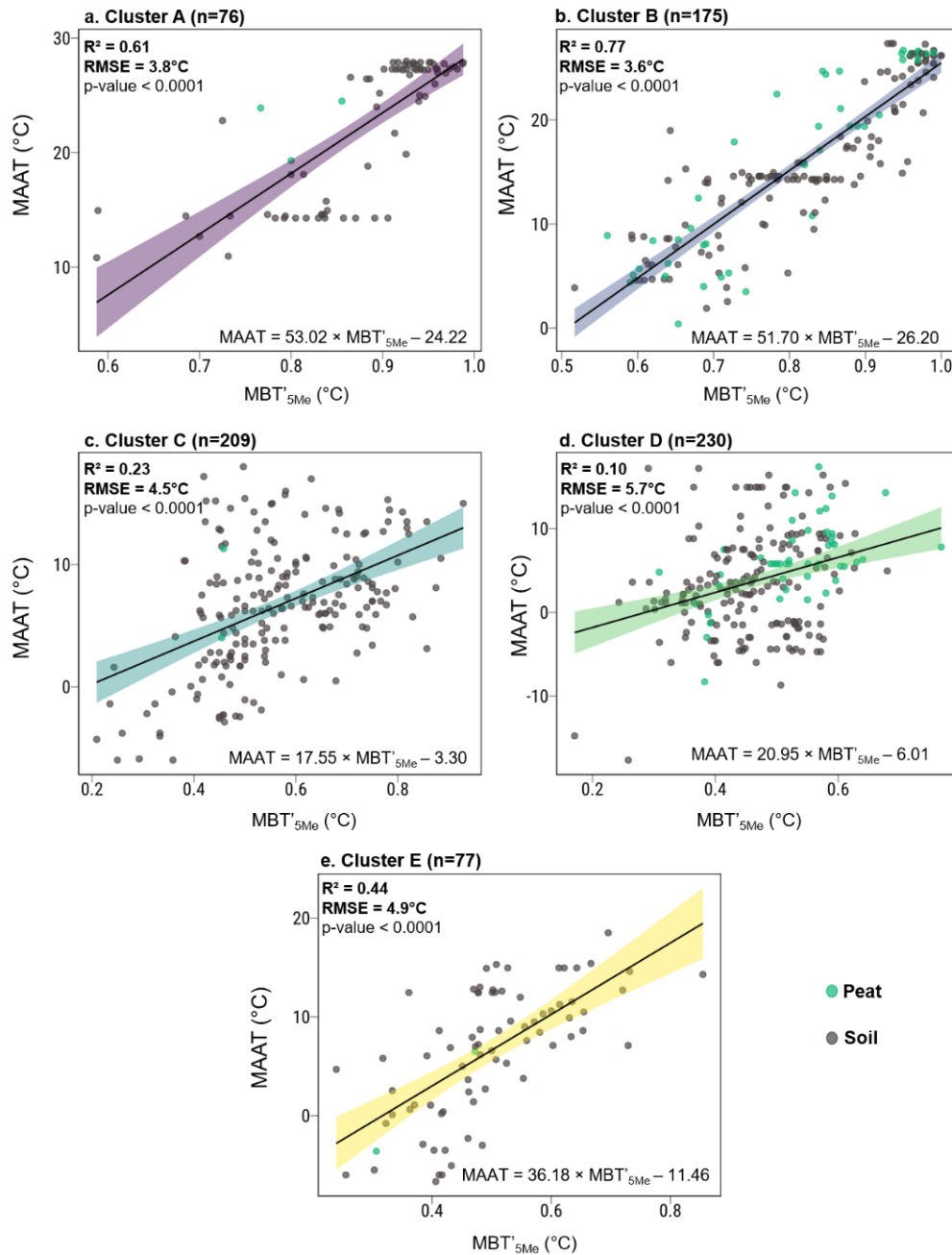


Figure 5.4 Linear regressions between MAAT (°C) and MBT'_{5Me} for (a) cluster A, (b) cluster B, (c) cluster C, (d) cluster D and (e) cluster E. Grey dots correspond to soils and green dots to peats samples.

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

Regarding clusters A and B, the first two axes explain 63.9% and 74.6% respectively of the total variance of the dataset (Fig. 5.5 a, b; Table 5.2). The selected environmental variables explain 97.3 % (cluster A) and 99.9% (cluster B) of the variance of the brGDGT relative abundances (Table 5.2). The first axis of the RDA for clusters A and B is well correlated with MAAT ($r=0.80$; $r=0.90$ respectively), FRS ($r=-0.80$; $r=-0.81$), pH ($r=-0.75$; $r=-0.60$) and MAP ($r=0.77$; $r=0.78$) (Fig. 5.5a, b; Table 3). Axis 2 of cluster A is mainly correlated with the FRS ($r=0.56$) and to a lesser extent MAAT ($r=0.26$) and pH ($r=-0.34$) and MAAT ($r=-0.26$), while axis 2 cluster B is predominantly correlated with pH ($r=-0.76$) and to a lesser extent with MAAT ($r=-0.41$; Fig. 5.5 a, b; Table 5.2). The quantification of the combined influence of the different environmental variables on the brGDGT distribution shows a predominant effect of the FRS (39.6%) and to a lesser extent pH (18.6%) for cluster A and MAAT (57.3%) and to a lesser extent pH for cluster B (16.7%; Table 5.2). The predominant influence of the thermal regime (MAAT and FRS) on the relative distribution of brGDGTs in these two clusters explains why the linear regressions between MBT'_{5Me} and MAAT are stronger (Fig. 5.4) than that observed for the global dataset (Supp. Fig. 5.1; Eq.1).

The first two axes of the RDAs for clusters C and D explain 34.9% and 53.4% of the total inertia of the dataset, with an explained fitted variation (relationship between the fractional abundances of brGDGTs and the selected environmental variables, calculated as the sum of all constrained eigenvalues) of 95.7% and 99.4%, respectively (Fig. 5.5 c, d; Table 5.2). For these two clusters, axis 1 is positively correlated with MAP values ($r=0.59$ and 0.48 , respectively) and strongly negatively correlated with pH ($r=-0.91$ and -0.96 , respectively). Axis 2 is controlled by the thermal regime, being mainly correlated with FRS ($r=0.74$ and -0.68 for clusters C and D, respectively) and MAAT ($r=-0.78$ and 0.39 , respectively). The quantification of the combined influence of the environmental variables on brGDGT distribution in clusters C and D shows a predominant effect of soil pH (25.6% and 47.4%, respectively) and only to a much lower extent MAP and MAAT (<6%) (Table 5.2). This is consistent with the weak correlation between MBT'_{5Me} and MAAT for cluster C, and absence of correlation for cluster D as well as the high scattering of the corresponding values (Fig. 5.4 c, d).

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

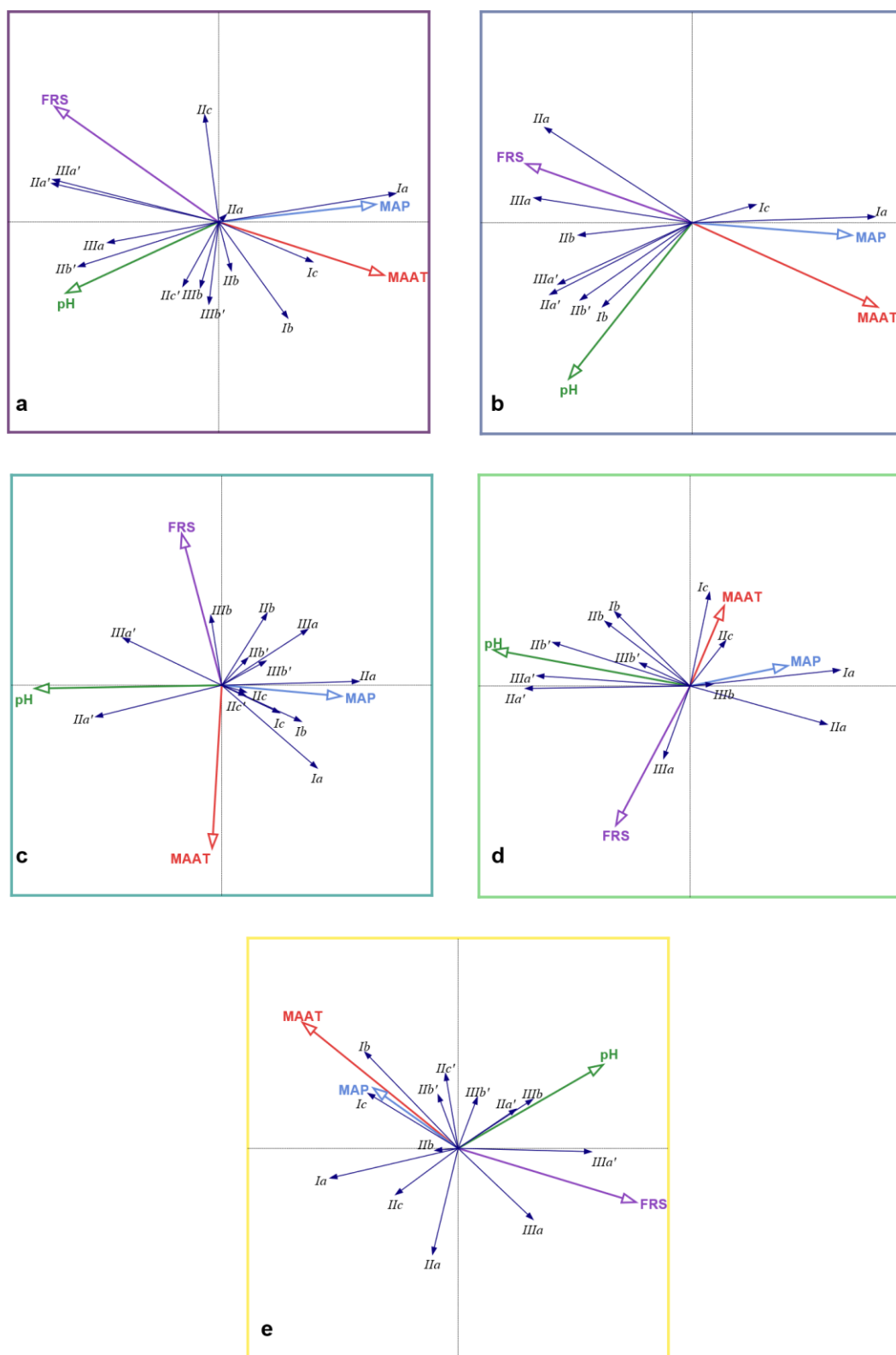


Figure 5.5 Redundancy analysis between brGDGT distribution and environmental variables for (a) cluster A, (b) cluster B, (c) cluster C, (d) cluster D and (e) cluster E.

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

Variables	RDA correlation coefficients		Conditional effects (%)	
	Axis 1	Axis 2		
Cluster A	pH	-0.75	-0.34	18.6*
	MAAT (°C)	0.8	-0.26	5.3*
	MAP (mm/year)	0.77	0.08	2.2
	Number of frost days (FRS)	-0.8	0.56	39.6*
	Expl. variation (%)	59.10	4.82	
	Expl. fitted variation (%)	89.97	7.34	
Cluster B	pH	-0.6	-0.76	16.7*
	MAAT (°C)	0.9	-0.41	57.3*
	MAP (mm/year)	0.78	-0.06	0.2
	Number of frost days (FRS)	-0.81	0.29	0.5
	Expl. variation (%)	68.99	5.62	
	Expl. fitted variation (%)	92.42	7.52	
Cluster C	pH	-0.91	-0.02	25.6*
	MAAT (°C)	-0.05	-0.78	4.5*
	MAP (mm/year)	0.59	-0.05	5.7*
	Number of frost days (FRS)	-0.2	0.74	0.7
	Expl. variation (%)	30.41	4.51	
	Expl. fitted variation (%)	83.36	12.37	
Cluster D	pH	-0.96	0.18	47.4*
	MAAT (°C)	0.17	0.39	1.5*
	MAP (mm/year)	0.48	0.1	3.4*
	Number of frost days (FRS)	-0.36	-0.68	1.5*
	Expl. variation (%)	51.58	1.88	
	Expl. fitted variation (%)	95.9	3.5	
Cluster E	pH	0.71	0.41	7.7*
	MAAT (°C)	-0.76	0.62	6.6*
	MAP (mm/year)	-0.41	0.29	0.6
	Number of frost days (FRS)	0.87	-0.26	15.2*
	Expl. variation (%)	18.19	9.6	
	Expl. fitted variation (%)	62.5	31.9	

Table 5.2 RDA correlation coefficients of the selected environmental variables along axes 1 and 2 for each cluster, and quantification of the influence of the different environmental variables on brGDGT relative abundances. Statistical significance ($p < 0.05$) is shown with an asterisk.

The first two axes of the RDA for cluster E explain 28.4% of the total inertia of the dataset and the explained fitted variation is 94.40% (Fig. 5.5e; Table 5.2). Axis 1 is mainly correlated with MAAT ($r=-0.76$), FRS ($r=0.87$), pH ($r=0.71$) and to lower extent MAP ($r=-0.41$; Table 5.2). Axis 2, on the other hand, is mainly influenced by MAAT ($r=0.62$), and to a lesser extent pH ($r=0.41$; Table 5.2). When examining the combined effect of the different environmental variables in this cluster, it appears that brGDGT distribution is mainly and significantly controlled by FRS (15.2%) and to a lower extent by MAAT (6.6%) and pH (7.7%; Table 5.2). The major influence of the thermal regime (FRS, MAAT) on brGDGT distribution in cluster E is consistent with the relationship ($R^2=0.44$) observed between MBT'_{5Me} and MAAT (Fig. 5.4e). Nevertheless, in contrast with cluster A, the additional influence of pH may explain the moderate determination coefficient of this correlation.

5.3. Discussion

5.3.1. Constraints on the MBT'_{5Me} -MAAT relationship in soils

The MBT'_{5Me} was shown to be linearly and strongly correlated with MAAT in the present extended soil dataset (Supp. Fig. 5.1; Eq. 5.1), as previously observed at the global level (e.g. De Jonge et al., 2014; Dearing Crampton-Flood et al., 2020). Nevertheless, in line with these previous studies, the RMSE remains high ($> 5^\circ\text{C}$), notably due to the influence of other environmental parameters than MAAT on brGDGT distribution.

Soil moisture has been suggested to have an effect on the relative abundance of brGDGT distribution (e.g. Dirghangi et al., 2013; Menges et al., 2014; Dang et al., 2016a; Dang et al., 2016b; Naafs et al., 2017a), with weak or no linear relationships between MBT/MBT'_{5Me} and MAAT in arid and/or alkaline soils. It was proposed that the relative soil moisture, correlated with pH variations (as shown on the RDA, Fig.5.5a, c, d), may have a role on the diversity of bacterial communities (Lauber et al., 2009; Shen et al., 2019) and/or that brGDGT producers may change their membrane composition in response to soil moisture changes (Dang et al., 2016b). The relative soil moisture may also impact the capacity of a soil to retain heat (Idso et al., 1975; Davidson et al., 1998;

Balleza et al., 2014; Dang et al., 2016b and reference therein), indirectly influencing the methylation degree of brGDGTs. Dang et al (2016b) and Naafs et al. (2017a) especially showed that the MBT'_{5Me} was only significantly correlated with MAAT when the ratio of 5- vs. 6-methyl brGDGTs (IR_{6Me}) was lower than 0.5. Therefore, to potentially improve the accuracy of our global MBT'_{5Me}-MAAT calibration (Supp. Fig. 5.1), we divided the total dataset in two sub-groups based on a threshold value of 0.5 for the IR_{6Me} ratio as proposed by Dang et al. (2016b) and Naafs et al. (2017a). Two subgroups with similar number of samples ($n=389$ for IR_{6Me}>0.5; $n=376$ for IR_{6Me}<0.5) were obtained. The linear regressions between MBT'_{5Me} and MAAT in the two subgroups were statistically similar (Supp. Fig. 5.2), even though a slightly higher determination coefficient and lower RMSE were observed when IR_{6Me}<0.5. Moreover, the regressions obtained for the two subgroups did not show obvious improvements with the one derived from the total dataset. Therefore, a separation of the dataset based on the IR_{6Me} values does not appear necessary for this large dataset.

More recently, De Jonge et al. (2019) suggested that the response of brGDGTs to temperature changes is strongly dependent on the nature of bacterial communities present in soils. These authors initially showed that the distribution of brGDGTs in a set of geothermally warmed soils from Iceland changes when the average annual soil temperature is above 14°C. This sudden change in brGDGT distribution coincided with an abrupt shift in the bacterial community composition. This observation was extended to the global scale, as a PCA of the brGDGT distribution over the entire peat and soil dataset compiled by De Jonge et al. (2019) revealed two distinct clusters – one considered as a “cold” subgroup ($n=251$ soils and peats; MAAT between -8.3°C and 18.2°C) and another one as a “warm” subgroup ($n= 195$ soils and peats; MAAT between 0.4°C and 27.1°C) – with different responses to temperature and pH. A relative increase in brGDGT *Ia* vs. homologues *Ila* and *IIla* was observed in the warm subgroup and was reflected by the community index (CI) proposed by De Jonge et al. (2019; Eq. 5.2):

$$CI = [Ia]/([Ia] + [Ila] + [IIla]) \quad (5.2)$$

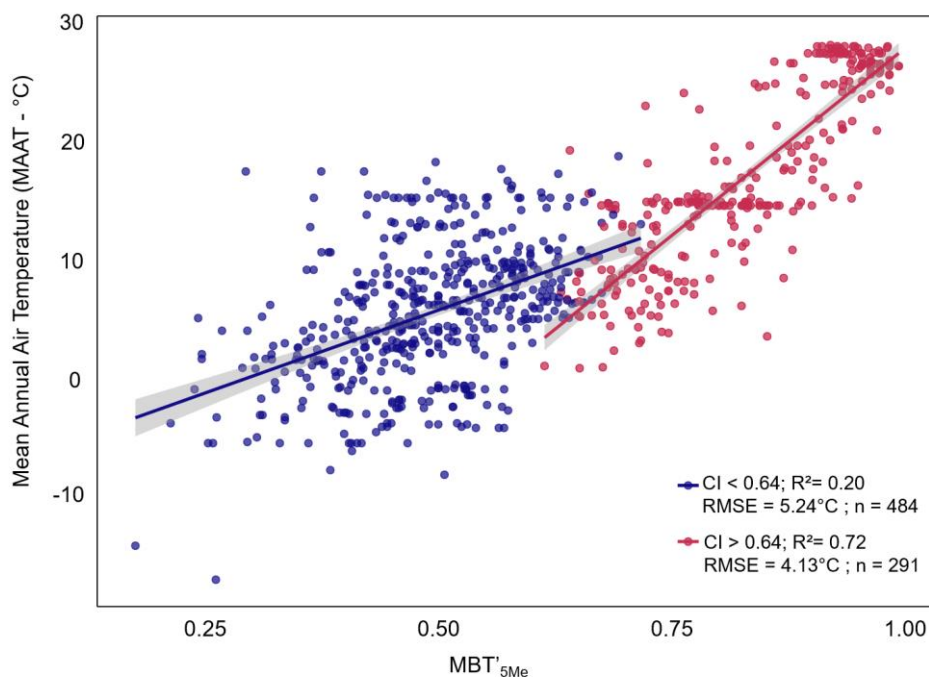


Figure 5.6 Comparison of the linear regressions between MAAT (°C) and MBT'_{5Me} in the two subgroups derived from the extended dataset: “cold” cluster (CI < 0.64) and “warm” cluster (CI > 0.64). The CI index (Eq. 5.2) was defined by De Jonge et al. (2019).

The CI index is similar to the MBT'_{5Me} (Eq. 5.2), except that it excludes the compounds containing cyclopentyl moieties, i.e., those suspected to be pH-sensitive compounds. De Jonge et al. (2019) proposed a CI threshold of 0.64 to separate the warm (>0.64) and cold (<0.64) brGDGT clusters and improved the MBT'_{5Me}-MAAT relationship when considering these two subgroups separately. The extended dataset proposed in the present study ($n=775$) was thus divided into two subgroups based on the CI threshold of 0.64 proposed by De Jonge et al. (2019). A strong linear relationship between MBT'_{5Me} and MAAT was observed for the warm cluster ($R^2=0.72$), while it was much weaker for the cold subgroup ($R^2=0.20$; Fig. 5.6). The discrimination of samples in clusters based on their different brGDGT signature and response to environmental changes allows improving the MBT'_{5Me}-MAAT relationship, even though this requires utilization of subdatasets of reduced size.

Following the approach of De Jonge et al. (2019), we performed a statistical clustering of the extended peat and soil dataset without any *a priori* on the basis of their brGDGT fractional abundances. Five clusters were obtained (Fig. 5.1). The brGDGT

distribution was differently impacted by environmental variables in each of these clusters (Fig. 5.5; Table 5.2). In clusters A and B, the relative distribution of brGDGTs was shown to be mainly impacted by the thermal regime (Table 5.2), as reflected by strong correlations between MAAT and MBT'_{5Me}. In contrast brGDGT distribution was largely influenced by other factors than MAAT, especially pH, in clusters C and D, and to a lesser extent cluster E (Table 3), leading to weak relationships between MAAT and MBT'_{5Me} (Fig. 4). This shows a different dependency of brGDGTs on temperature changes at the global scale, consistent with previous observations by De Jonge et al. (2019), all the more as clusters A and B were mainly characterized by CI > 0.64 (Fig. 5.7) and can be related to “warm” groups as defined by De Jonge et al. (2019), whereas the other clusters, with CI < 0.64, mainly correspond to “cold” groups. The difference in MBT'_{5Me} efficiency between the clusters proposed in this study can be explained by a potential influence of the thermal regime on brGDGT synthesis by microorganisms as already assessed by De Jonge et al. (2019) and by several other studies which suggested that brGDGTs may be preferentially produced during warm seasons (Weijers et al., 2011b; Huguet et al., 2013; Deng et al., 2016)

The CI threshold (0.64) defined by De Jonge et al. (2019) was based on a smaller number of soils ($n=446$) than available in the present study. To refine this value based on the present extended sample set ($n=775$), all linear regressions between MAAT and MBT'_{5Me} were tested by successive iteration for CI values from 0 to 1 (0.001 step) (Supp. Fig. 5.3). The refined threshold is 0.69 (corresponding to the best adjusted R²), in agreement with the clusters presented in this study (Fig. 5.7) and those of De Jonge et al. (2019) and might be considered first and foremost in any use of the MBT'_{5Me} index to reconstruct paleo-MAATs as proposed by De Jonge et al., (2019).

The fact that a strong relationship between MBT'_{5Me} and MAAT is only observed for the warmer clusters (Fig. 5.4) highlights the limitations of using a single index and a simple linear regression model to capture the response of brGDGTs to temperature changes. This response is necessarily complex, as the brGDGT distribution can be concomitantly and non-linearly influenced by several environmental parameters. Other models should be complementarily developed to better reflect such a complex response. They should be based on bacterial relative abundances rather than a single index to be more realistic and representative of the environmental complexity, as the whole suite of

brGDGTs can be influenced by environmental parameters. They should also consider non-linear influences, which may occur in environmental settings. Machine-learning models can be used to this aim and will be tested in the following.

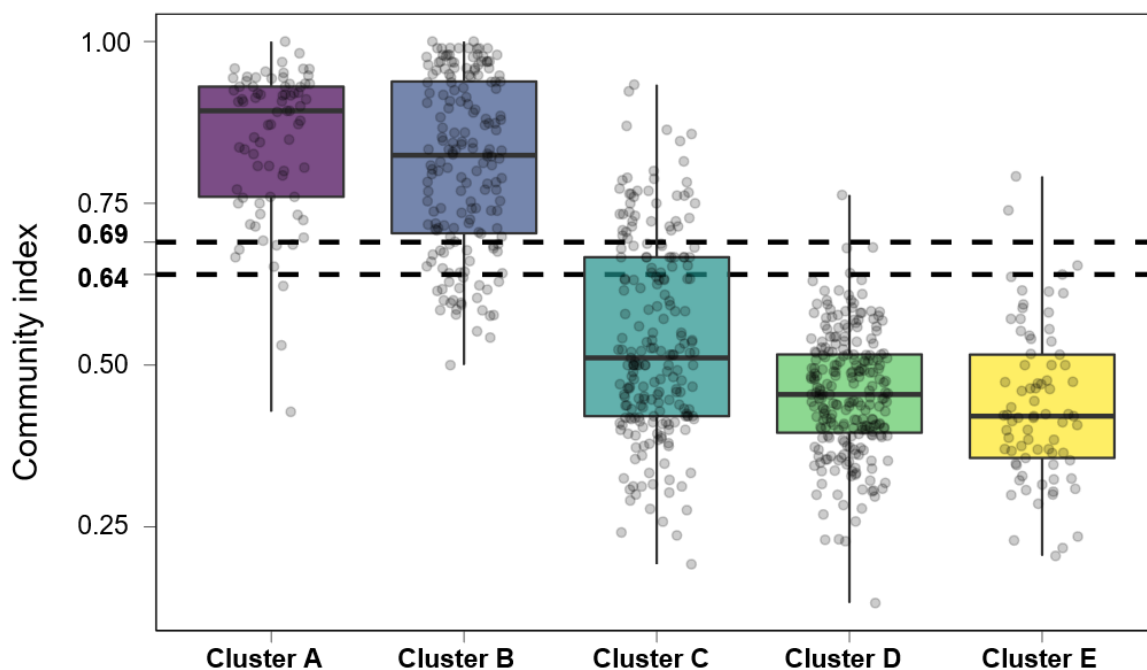


Figure 5.7 Average values of the Community Index (CI; Eq. 3) for each cluster of the global soil dataset defined after PCA analysis (Fig. 2). The threshold of 0.64 for the CI separating “warm” and “cold” groups (De Jonge et al., 2019) is represented.

5.3.2. A novel global calibration between MAAT and brGDGTs using random forest algorithm

5.3.2.1. Model development

A machine-learning model – the random forest algorithm – was tested to potentially derive stronger and more accurate global MAAT calibration from brGDGT distributions than single linear regressions based on the MBT’_{5Me} index.

The random forest algorithm was first “trained” during the learning phase to estimate MAAT from brGDGT relative abundances. During this phase, the model produces decision trees that will automatically discriminate the compounds whose relative abundances are not influenced by temperature, and thus selects those to be used to estimate MAAT. The developed model was called “random Forest Regression for PaleOMAAT using brGDGTs” (FROG).

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

As the 5-methyl brGDGTs are considered to be mainly correlated with MAAT in contrast with 6-methyl isomers (De Jonge et al., 2014), the random forest model was initially applied to the fractional abundances of the 5-methyl brGDGTs only (FROG_{5Me}). After training on the sampling dataset (75% of the global dataset; n=583), a strong ($R^2 = 0.80$) global calibration between 5-methyl brGDGT relative abundance and MAAT was obtained (Fig. 5.8a), with a RMSE of 4.12 °C. The random forest algorithm was then tested using the relative abundances of all brGDGTs including the 6-methyl isomers (FROG). This "alternative" global calibration ($R^2=0.81$ and RMSE =4.09 °C; Fig. 5.8b) appeared similar to the FROG_{5Me} model (Fig. 5.8a). The variance in residuals and RMSE of both random forest calibrations (FROG_{5Me} and FROG) are lower than those of the simple linear regression between MBT'_{5Me} and MAAT (Fig. 5.8c, d; Table 5.3). The residuals of the random forest models do not show clear trends which could reflect the influence of other variables than the MAAT on brGDGT distribution (Fig. 5.8c, d). However, as observed by Naafs et al. (2017a) and Dearing Crampton-Flood et al. (2020) in their global MBT'_{5Me}-MAAT calibration, a major part of the uncertainty in the FROG model is due to a wide dispersion of predicted temperatures for MAAT < 10°C (Fig. 5.8c, d), corresponding to samples from high latitudes and/or high elevations.

	n (samples)	R ²	RMSE (°C)	Variance in residuals (°C)	Lower limit (°C)	Upper limit (°C)
Global Calibration						
FROG _{5Me}	192 (775)	0.8 (0.78)	4.12 (4.14)	17.0 (17.1)	-2.06 (-3.50)	27.33 (27.47)
FROG	192 (775)	0.81 (0.79)	4.09 (4.01)	16.7 (16.1)	-2.38 (-4.23)	27.47 (27.58)
Naafs calibration*	192 (775)	0.70 (0.64)	5.01 (5.23)	25.1 (27.4)	-7.81 (-7.81)	24.59 (24.59)
De Jonge calibration*	192 (775)	0.66 (0.61)	5.31 (5.45)	28.2 (29.7)	-3.19 (-3.19)	22.88 (22.88)
BayMBT*	192 (775)	0.52 (0.43)	6.33 (6.58)	40.1 (43.3)	-15.89 (-15.89)	29.83 (29.83)
MBT' _{5Me} (this study)	192 (775)	0.70 (0.65)	5.01 (5.20)	25.1 (27.0)	-6.73 (-6.73)	23.29 (23.29)
Alternative Calibration						
FROG ₀	164 (661)	0.83 (0.85)	2.53 (2.34)	6.4 (5.5)	5.39 (4.58)	27.02 (27.72)
BayMBT ₀ *	164 (661)	0.56 (0.54)	4.07 (4.07)	16.6 (16.6)	2.68 (1.43)	27.02 (27.68)
FROG ₅₀₀	108 (442)	0.85 (0.82)	3.56 (3.66)	12.7 (13.4)	-1.27 (-1.82)	27.48 (27.79)

Table 5.3 Characteristics of the different models proposed in this study to estimate MAAT: R², RMSE, variance of the residuals and the upper and lower limits of estimation. The "training" samples were used to develop the different machine learning models, which were then tested on the remaining sample set.

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

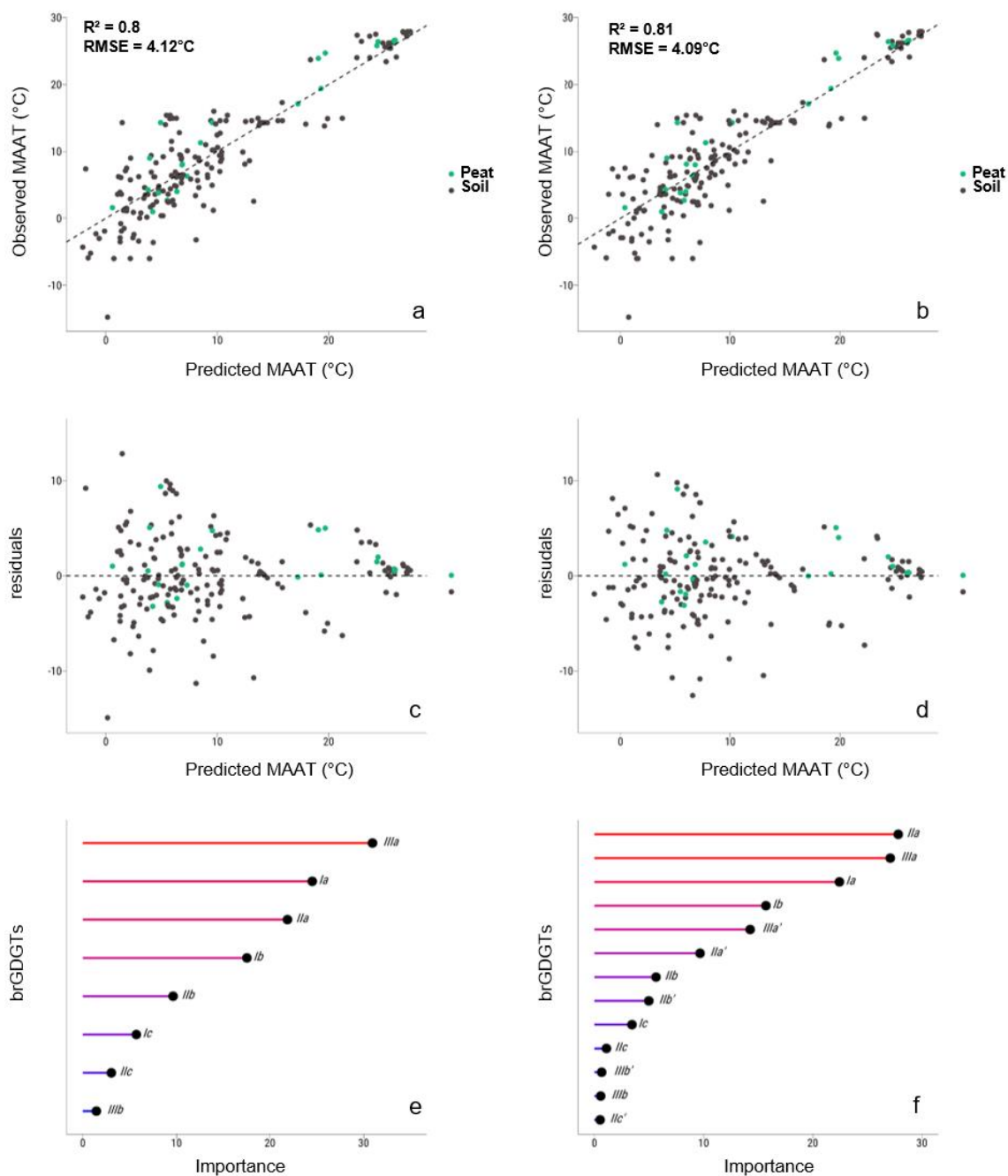


Figure 5.8 MAAT predicted by the random forest models using the relative abundances of (a) 5-methyl brGDGTs (FROG_{5Me}) and (b) all brGDGTs. Residuals of the (c) FROG_{5Me} model and (d) FROG model with all brGDGTs. Importance of the individual brGDGTs in (e) the FROG_{5Me} model and (f) the FROG model with all brGDGTs, according to the permutation importance method (Breiman, 2001). Grey samples correspond to soils and green samples to peats.

An advantage of the random forest algorithm lies in the fact that the weight of the different variables used to define the model can be quantified using the permutation importance method (Fig 8e, f; Breiman, 2001). The brGDGTs *Ia*, *Ila*, *IIla* were the homologues predominantly used by the FROG_{5Me} model (Fig. 5.8e). This is consistent with the linear correlation between the MBT'_{5Me} and MAAT (Supp. Fig. 1; Eq.5.1), as these acyclic brGDGTs have a major weight in the MBT'_{5Me}. Nevertheless, other brGDGT homologues are also taken into account by the FROG_{5Me} model, especially compounds *Ib* and *Ilb*, which contain one cyclopentyl moiety.

Regarding the random forest model established with all brGDGTs (FROG model), the predominant homologues are also the acyclic brGDGTs *Ia*, *Ila*, and *IIla*, and to a lesser extent compound *Ib* as well as compounds *IIla'*, *Ila'*, i.e., 6-methyl isomers, pointing to their contribution to MAAT reconstruction. The FROG model thus presents the advantage of integrating both types of brGDGT isomers, including 6-methyl ones, even though they are usually considered as being predominantly influenced by pH variations (De Jonge et al., 2014b). Taking into account both 5- and 6-methyl brGDGTs in the model reflects the environmental reality and the concomitant influence of other environmental parameters than MAAT such as pH and MAP (Table 5.2) on the distribution of brGDGTs. Therefore, it seems more relevant to use the FROG model for MAAT reconstruction, all the more as the FROG_{5Me} and FROG models have similar determination coefficients, RMSE and estimation ranges (Table 5.3).

5.3.2.2. Comparison of the FROG model with previous global soil calibrations

The FROG model was compared with previously published global soil/peat linear calibrations based on Ordinary Least Square (De Jonge et al., 2014), Deming (Naafs et al., 2017a) and Bayesian (Dearing Crampton-Flood et al., 2020) regressions. Our model is characterized by a higher determination coefficient and lower RMSE (Fig. 5.8; R²=0.81 and RMSE=4.09 °C) than available calibrations (De Jonge et al., 2014: R²=0.61, RMSE=4.8°C; *n* =231; Naafs et al., 2017a: R²= 0.71, RMSE= 4.1°C, *n*=177; Dearing

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

Crampton-Flood et al., 2020: $R^2=0.64$, $RMSE=6.0$ °C, $n=343$; Table 5.3), while being obtained from a larger dataset ($n=775$).

In order to make a direct comparison of the different models, the calibrations by De Jonge et al. (2014), Naafs et al. (2017a) and Dearing Crampton-Flood et al. (2020; BayMBT model) were applied to the same dataset as the one used for the random forest model (Fig. 5.9; Table 4). This confirms that the FROG model performs better in terms of robustness and accuracy than the previously published global terrestrial calibrations (De Jonge et al., 2014; Naafs et al., 2017a; Dearing-Crampton Flood et al., 2020). The FROG model also increases the upper limit of MAAT estimation by > 3 °C in comparison with the calibrations by De Jonge et al. (2014) and Naafs et al. (2017a; Table 5.3). In contrast, the lower limit of MAAT estimates for the random forest model is higher than that of the other calibrations by De Jonge et al. (2014), Naafs et al. (2017a) and especially Dearing Crampton-Flood et al. (2020).

The higher strength and accuracy of the FROG model compared to those based on the MBT'_{5Me} (Table 5.3) could be explained by the fact that the non-parametric random forest algorithm selects the most relevant brGDGTs among the whole suite of homologues to estimate MAAT, while taking into account the concomitant and non-linear influences of other environmental parameters on these compounds, unlike the linear models (De Jonge et al., 2014; Naafs et al., 2017a; Dearing Crampton-Flood et al., 2020). This is highlighted when comparing the independent variations of the brGDGT relative abundances with estimated MAAT for the FROG model, with non-linear trends for most of the individual brGDGTs (Supp. Fig. 4). In addition, one of the prerequisites of the Bayesian model, as recently proposed by Dearing Crampton-Flood et al. (2020), is to determine a *prior*, before analysis. Nevertheless, as shown in this study, the influence of environmental parameters on brGDGT distribution in soils/peat is sample-dependent (e.g., Fig. 5.4; Table 5.2). Therefore, it seems difficult to determine a *prior* adapted to all the samples in such a large dataset. Bayesian models could be more efficient when applied to local/regional calibrations, where the *prior* can be determined more precisely. In any case, the BayMBT model by Dearing-Crampton Flood et al. (2020) is characterized by a higher range of estimated temperatures compared to the random forest FROG model (Table 4), reflecting the fact that each model presents its advantages and limitations. The

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

FROG and BayMBT models should be seen as complementary and show that efficient statistical approaches are needed to improve brGDGT calibrations.

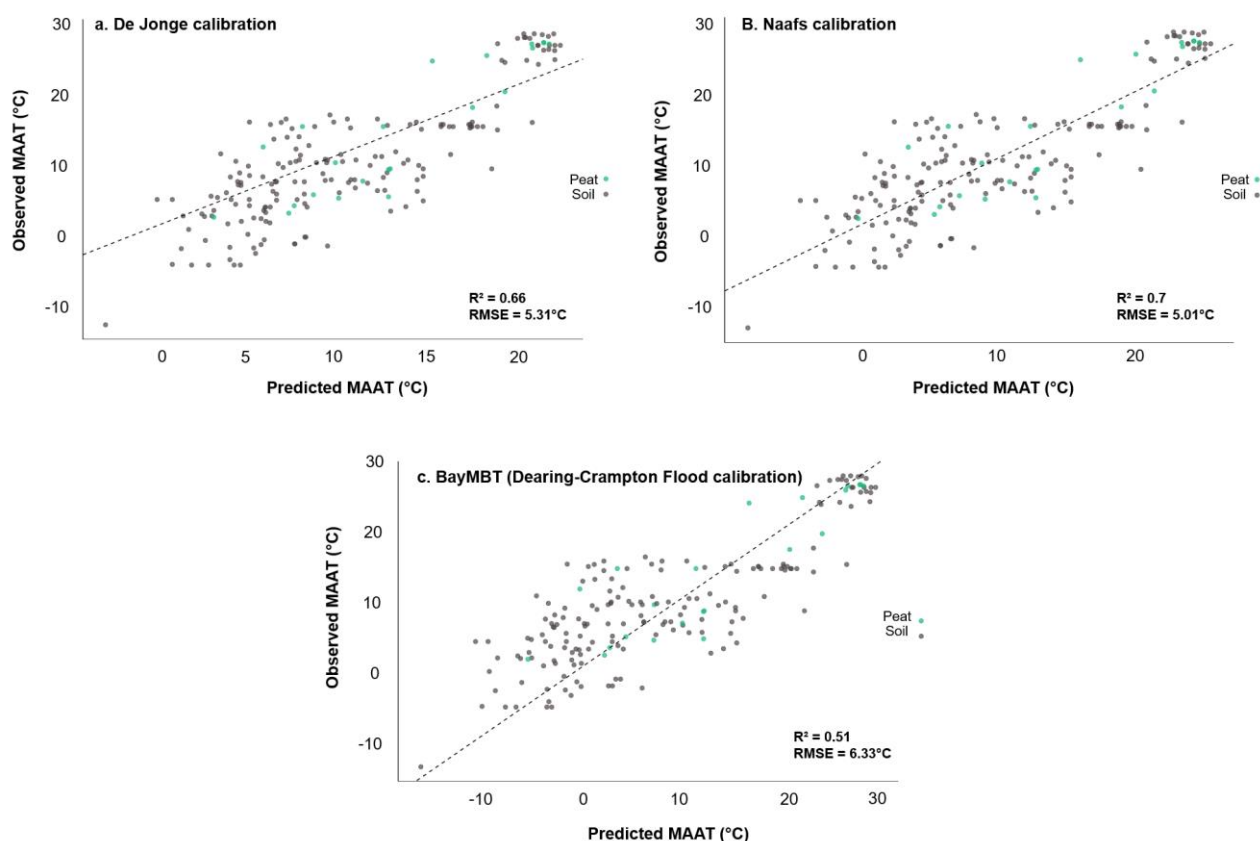


Figure 5.9 Comparison of the global MBT'_{5Me} calibration proposed by (a) De Jonge et al. (2014) and (b) Naafs et al. (2017a) using the same dataset than the FROG model.

5.3.3. Development of alternative MAAT calibrations

Although the FROG MAAT calibration presents a strong determination coefficient, the associated RMSE is still ca. 4 °C (Table 5.3), likely due to potential confounding factors influencing brGDGT distribution in addition to MAAT. To test the influence of the thermal regime and MAP on the FROG model and potentially improve its accuracy, alternative submodels based on different subsets of the extended dataset were developed.

5.3.3.1. *Alternative models considering the influence of thermal regime*

The number of frost days (FRS) was shown to have a significant impact on the brGDGT distribution of the extended soil dataset based on RDA analyses, especially on the 5-methyl isomers (Fig. 5.5; Table 5.2). The FRS can be considered as an indirect indicator of temperature thermal regime which may influence the growth of brGDGT source microorganisms and the production of these lipids. Nevertheless, the impact of temperature seasonality on brGDGT distribution is the object of contrasting observations. Indeed, brGDGT distribution was not observed to be affected by seasonal temperature variations in mid-latitude soils (Weijers et al., 2011; Lei et al., 2016), which may be related to the slow turnover of these compounds in soils (ca. 20 years; Weijers et al., 2010). In addition, it was suggested that microbial communities can continue to grow and biosynthesize membrane lipids in frozen soils (Drotz et al., 2010). In contrast, several studies suggested that brGDGTs may be preferentially produced during summer in peats (Weijers et al., 2011b; Huguet et al., 2013) or soils (e.g., Deng et al., 2016). An increase in microbial biomass production in unfrozen soils was also reported (Schimel and Clein, 1996; Nedwell, 1999; Schimel et al., 2007), reflecting an effect of the temperature seasonality.

To take into account the potential influence of the thermal regime on brGDGT distribution, an alternative Bayesian calibration was proposed by Dearing Crampton-Flood et al. (2020). This model (BayMBT₀) estimates the average temperature of all months that have an average temperature above 0 °C using the MBT'_{5Me} index. The BayMBT₀ model (R²= 0.70; RMSE = 3.8°C; Lower limit = 0.9°C, upper limit = 27.1°C; Dearing Crampton-Flood et al., 2020) improves the strength and accuracy of the BayMBT model but does not allow reconstructing negative temperatures. To evaluate the influence of the thermal regime on the FROG model, the same approach as that proposed by Dearing Crampton-Flood et al. (2020) was applied.

In the present extended dataset, monthly temperatures are available for 663 out of the 775 soil samples. This excludes some of the samples, collected in the French Alps, Peruvian Andes, Mts Pollino, Shegyla and Italy (Véquaud et al., 2020). The alternative calibration based on the average temperature of all months that have a temperature above

0 °C (FROG₀) is as strong as the global random forest calibration (FROG; Fig. 5.8), with an $R^2 = 0.84$, and is more accurate (RMSE = 2.5 °C) than the latter (Table 5.3), even though it is based on a slightly reduced dataset ($n=663$). The FROG₀ calibration also performs better than the BayMBT₀ model ($R^2 = 0.56$; RMSE = 4.1 °C), with a higher R^2 , lower RMSE but with a slightly lower range of MAAT estimation (Table 4). Overall, the present study, through RDA analyses (Fig. 5.5) and the FROG₀ model, highlights the effect of the thermal regime on brGDGT distribution. In particular, it can be assumed that within frozen soils the activity of brGDGT source organisms will be reduced. In addition, snowfall can isolate the soil from the atmospheric compartment, decoupling atmospheric temperatures from soil temperatures during the cold season. This confirms the interest of using alternative models including this effect to improve the accuracy of MAAT reconstruction, as also suggested by Dearing Crampton-Flood et al. (2020).

5.3.3.2. *Alternative model considering the influence of MAP*

As previously discussed, the brGDGT distribution may be largely impacted by relative soil moisture (e.g., Dirghangi et al., 2013; Naafs et al., 2017a). In the present study, MAP was used as a proxy of the soil water content (SWC), which varies as a first approximation according to the precipitation regime, although other factors can play a role, such as relief (topography), evapotranspiration, grain size or vegetation cover (Crave and Gascuel-Oudou, 1997; Gómez-Plaza et al., 2001).

The influence of MAP was tested with the FROG₅₀₀ random forest model. This alternative to the FROG model excludes all the samples with MAP < 500 mm/year, corresponding to soils previously defined as dry, and generally described as alkaline soils from arid regions, low in organic matter (Peterse et al., 2012; Naafs et al., 2017a; Dearing Crampton-Flood et al., 2020). The FROG₅₀₀ model contains 442 samples and only shows a slight improvement in MAAT reconstruction in comparison with the FROG calibration, with a slightly higher determination coefficient ($R^2 = 0.85$) and lower RMSE (3.5 °C; Table 5.3). The slight decrease in the RMSE of the FROG₅₀₀ vs. the FROG model (by ca. 0.5 °C) is likely related to the large reduction of the dataset size by more than 300 samples. Therefore, the use of the FROG₅₀₀ model, as the BayMBT₅₀₀ proposed by Dearing Crampton-Flood et al. (2020), does not seem particularly recommended, as its

performance is only slightly better than the original FROG calibration, even though it contains a much lower number of samples.

5.3.4. Paleo application of the FROG global calibration

The performance and validity of the FROG model was tested and compared with the MAAT record from a Chinese peatland (Wu et al., 2020) and from Pliocene sediments from the North Sea basin (Dearing Crampton-Flood et al., 2018). These archives were the object of previous paleostudies, providing a context for the interpretation of the MAAT data from the FROG model.

5.3.4.1. Application of the FROG model to a peatland covering the Holocene

The FROG calibration was applied to a core from a Chinese peatland (Sahara Sand Peatland; Fig. 10 in Wu et al., 2020), located in the Altai mountains and covering the last 11,000 years (Wu et al., 2020). The FROG model was developed from both soil and peatland samples, making its application possible to a peat archive (Fig. 5.10). This peatland has been the object of previous paleostudies and is ideal for paleoenvironmental and paleoclimate reconstruction (Rao et al., 2019; Wu et al., 2020), as it is not subjected to anthropogenic influences due to its specific position in the northern Arid Central Asia (ACA) at an elevation of 2446 m a.s.l.

Although not exclusively based on peat samples, the MAAT record derived from the FROG model was compared (Fig. 5.10) with the one obtained from the global MBT'_{5Me} peat calibration by Naafs et al. (2017b) and reported by Wu et al. (2020). The global MBT'_{5Me} peat calibration reconstructed a temperature of -6.1 °C for the modern period, consistent with the expected MAAT (-5.4°C; Wu et al., 2020). In contrast, the MAAT estimate derived from the FROG model for the present period is ca. 1.5°C and thus higher than expected (Fig. 5.10). A large amplitude of variation for the reconstructed MAAT was observed during the Holocene (Fig. 5.10) from both the FROG model (~ 8 °C) and the global MBT'_{5Me} peat calibration (~ 10 °C). Such a range of MAAT variation over the Holocene is consistent with previous studies in the area (Wu et al., 2020), with amplitudes of 6-7°C and even 7-8 °C based on pollen-based MAAT reconstructions from

a lake (Jiang et al., 2006) and a peatland (Feng et al., 2017), respectively. Even though the reconstructions based on the peat calibration (Wu et al., 2020) and on the FROG model both showed a warming over the Holocene, different trends were observed from the two models. Thus, the MBT'_{5Me} peat calibration shows a warming trend during the Holocene, with a cold period from 11,000 yr BP to 6,000 yr BP, a warming from 6,000 yr to 3,000 yr BP, and then a plateau between 3,000 yr BP and the present period (Wu et al., 2020). Conversely, the FROG model reconstruction shows a "cold" period, with large oscillations, between 11,000 and 9,000 yr BP (mean = -0.3°C) and a warmer period between 9000 and 0 yr BP (mean = 2.1°C; Fig. 5.10).

The MAAT records derived from the two brGDGT models were compared with the α -cellulose $\delta^{13}\text{C}$ record (Fig. 10; Rao et al., 2019; Wu et al., 2020). The α -cellulose $\delta^{13}\text{C}$ was suggested to be a proxy for summer temperatures. Indeed summer temperature has been presented as the most important climatic factor controlling the α -cellulose $\delta^{13}\text{C}$ on this study site (Rao et al., 2019) based on the study of modern plants, tree-ring samples, and instrumental climate data. The α -cellulose $\delta^{13}\text{C}$ proxy showed a warming trend over the Holocene in the same way as the two brGDGT models. Nevertheless, the α -cellulose $\delta^{13}\text{C}$ summer temperature record presents more similarities with the FROG MAAT record than with the MBT'_{5Me} peat model, with especially an obvious warm period from ca. 8,000 to 4,000 yr B.P. visible in both the α -cellulose $\delta^{13}\text{C}$ and FROG records. In contrast, this warming trend was not observed in the MBT'_{5Me} peat record (Fig. 10), which might be due to the fact that the α -cellulose $\delta^{13}\text{C}$ proxy reflects summer temperature, whereas the MBT'_{5Me} is related to MAAT, as suggested by Wu et al. (2020).

More generally, the MBT'_{5Me} peat model does not record any obvious temperature variation over the 11,000-5,500 yr B.P. period in contrast with the other two records, which may be explained by a possible "temperature threshold" effect on the brGDGT distribution. Indeed, the MBT'_{5Me} was suggested to be insensitive to MAAT variations in extreme conditions (i.e., too low or too high temperature; De Jonge et al., 2014; Naafs et al., 2017a). This bias is avoided by the FROG model, which is based on the individual abundances of the brGDGTs and including the complex and non-linear responses of these compounds to the variations of the environmental parameters including MAAT. In addition, the thermal regime was observed to have an influence on the relative distribution

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

of brGDGTs (Fig. 5.6). These compounds might be more representative of the summer than annual environmental conditions, as suggested in some previous studies (Huguet et al., 2013). This implies that in the present case the FROG model may reflect summer temperatures rather than MAAT, which could explain the similarities between the FROG and α -cellulose $\delta^{13}\text{C}$ records, in contrast with the MBT'_{5Me} peat record. This also could be an explanation of the colder modern MAAT estimate for the MBT'_{5Me} than for the FROG model. It should be noted that the MAAT record derived from the FROG model also shows similar patterns of variation as those derived from other proxies (pollen, alkenones, $\delta^{18}\text{O}$; Fig. 6 in Wu et al., 2020). To conclude, the FROG model shows documented climatic variations, with a reduced error (4°C) compared to previous global calibrations. The use of this model also highlights complementary information to the MBT'_{5Me} index and these two methods of past temperatures estimations should be done concomitantly.

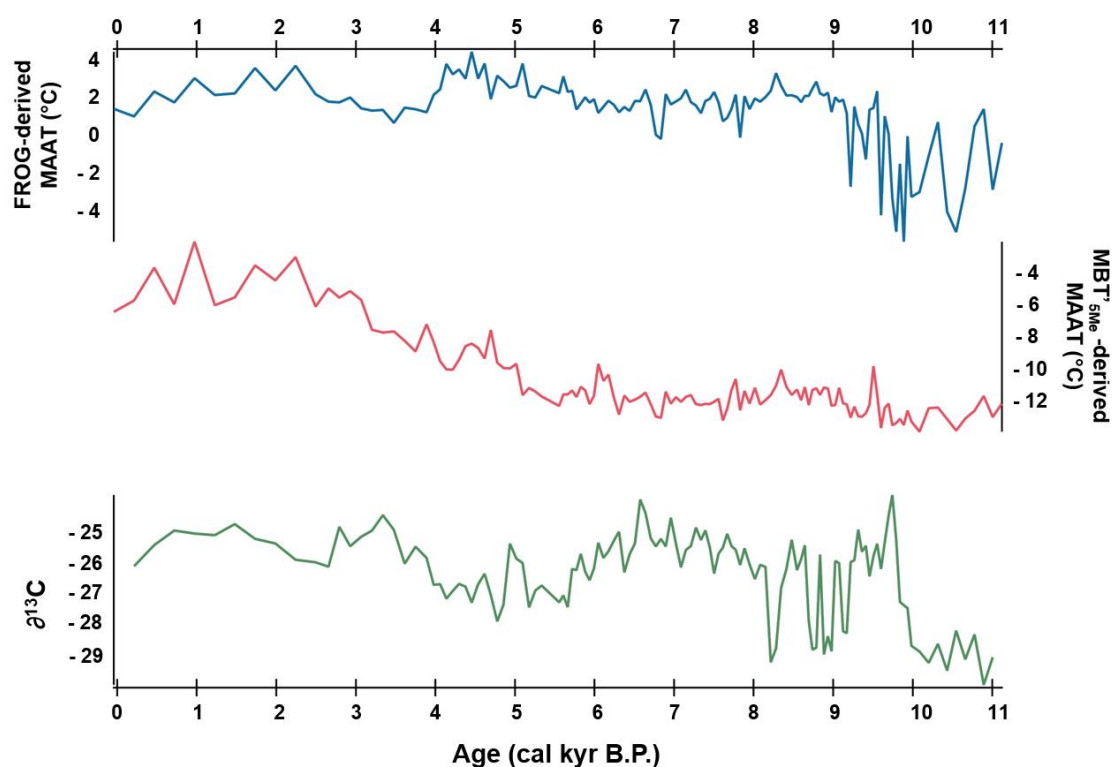


Figure 5.10 Comparison of the MAAT estimates derived from the MBT'_{5Me} peat calibration by Naafs et al. (2017; red) and the FROG random forest model (blue) with the peat α -cellulose $\delta^{13}\text{C}$ record (green) from core ATM10-C7 of the SSP (Rao et al., 2019; Wu et al., 2020).

5.3.4.2. *Application of the FROG model to sediments from the Pliocene*

The performance and reliability of the FROG model were also tested using a paleorecord based on Pliocene sediments from the North Sea basin (Dearing Crampton-Flood et al., 2018, 2020). BrGDGTs were previously analyzed in this archive, with subsequent reconstruction of past temperature variations during the Pliocene using the MBT'_{5Me} (Dearing Crampton-Flood et al., 2018). More recently, the validity of the Bayesian model (BayMBT₀) proposed by Dearing Crampton-Flood et al. (2020) was tested on this archive and compared with the MAAT records derived from the terrestrial brGDGT calibrations by De Jonge et al. (2014) and Naafs et al. (2017a).

The aforementioned calibrations, as well as the FROG model (Fig.5.11a), showed an overall decrease in MAAT during the Pliocene, consistent with the global cooling that is documented in the literature (Lisiecki and Raymo, 2005; Dearing Crampton-Flood et al., 2018, 2020). The MAAT records derived from the different models showed similar qualitative trends over the reconstructed period, even though the range of variation appeared slightly smaller (3-4°C) for the FROG model than for the other calibrations (Fig. 5.11a).

The absolute temperatures reconstructed by the different brGDGT models were generally lower than those expected in NW Europe (13-14 °C) over the Pliocene and derived from pollen assemblages (Dearing Crampton-Flood et al., 2020). The BayMBT record showed the lowest MAAT estimates, with very large oscillations and negative temperatures during the late Pliocene, in contrast with the other global calibrations (Fig.11a). Nevertheless, it is unlikely that temperatures went below 0°C during the Pliocene based on the presence of pollen assemblage of warm-adapted species found in Dutch sediments of this period, as specified by Dearing Crampton-Flood et al. (2020). That is why Dearing Crampton-Flood et al. (2020) favored the use of the BayMBT₀ model, based on the mean temperature of all months above 0 °C, for MAAT reconstruction, as it is more accurate (RMSE = 3.8°C) than the BayMBT model (RMSE = 6°C). The BayMBT₀ model indeed provided higher absolute MAAT estimates than the BayMBT and much more consistent with temperatures estimated from other proxies (i.e. 13-14°C; Fig. 5.11b).

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

Although the FROG model did not result in such negative reconstructed MAAT, the FROG₀ model was also applied to the Pliocene archive to compare with the BayMBT₀ record (Fig. 5.11b). Both models showed a decreasing trend in temperature over the Pliocene, although the BayMBT₀ model displayed larger oscillations (between 3.5 °C and 16.2°C; mean 10.7°C) than the FROG₀ one (between 10 °C and 13.1°C; mean 11.7°C). As the MAAT estimates from the FROG₀ model are associated with a smaller error (2.5 °C; Table 5.3) than the other global brGDGT calibrations including the BayMBT₀ one (3.8 °C; Dearing Crampton-Flood et al., 2020), an improvement in the accuracy of the paleoreconstruction over this period can be anticipated when using the FROG₀ model. This model, unlike the FROG model, estimates the mean temperature of all months above 0°C, which can be considered as more reflecting warm season temperature. Unlike the FROG model, the FROG₀ model can neglect the thermal and nival influences on the samples, and so on the source organisms of brGDGTs. We suggest that the FROG₀ model should be used in addition to the FROG model rather than alone, in order to obtain complementary information on annual and seasonal temperatures dynamics in paleoclimate studies.

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

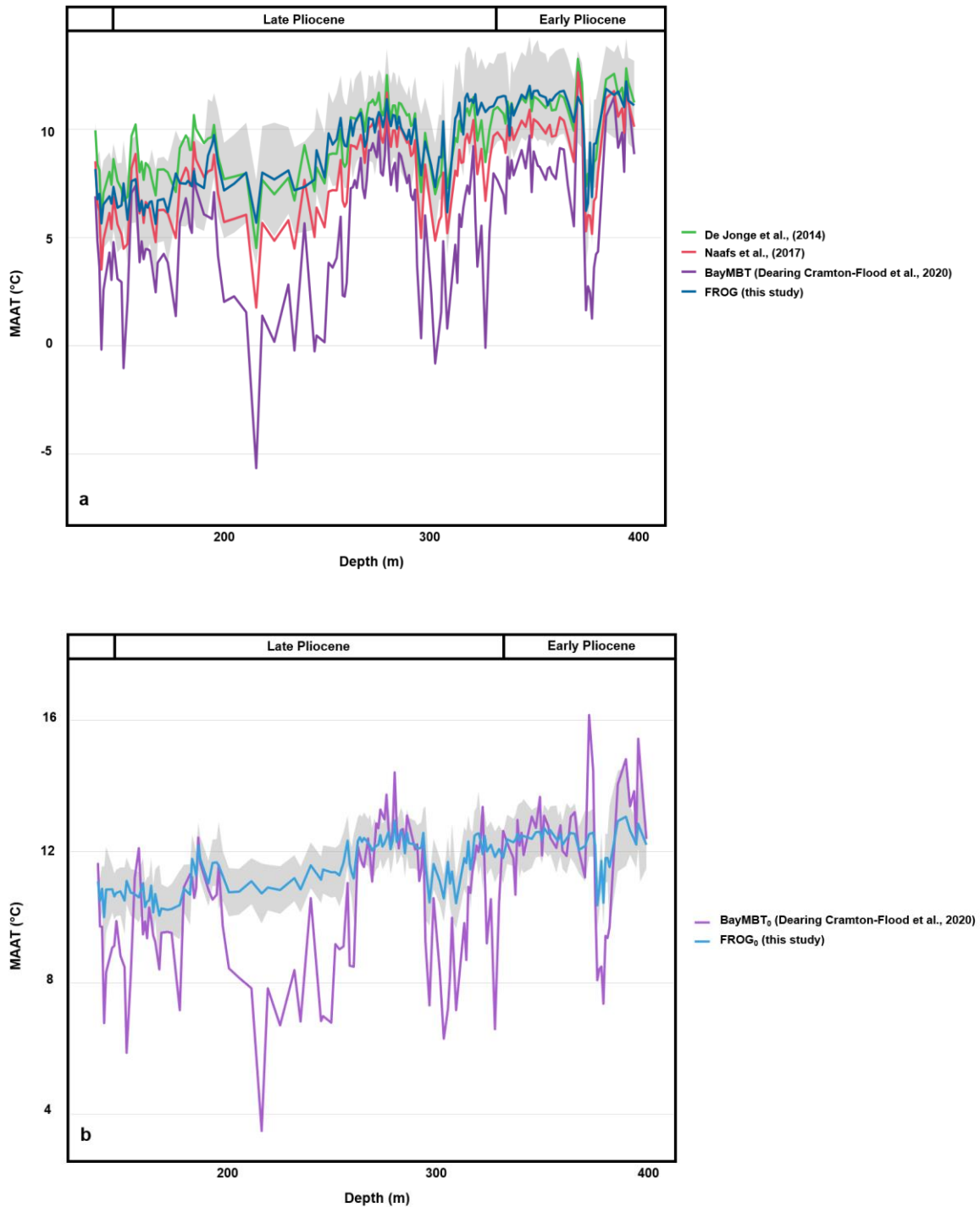


Figure 5.11 Reconstructed MAAT for the Pliocene marine sediment sequence from the Hank core located in the Netherlands (Dearing Crampton-Flood et al., 2018, 2020) (a) derived from the calibrations by De Jonge et al. (2014), Naafs et al. (2017a), Dearing Crampton-Flood et al. (2020; BayMBT) and from the FROG model and (b) derived from the calibrations by Dearing Crampton-Flood et al., (2020; BayMBT₀) and from the FROG₀ model. Grey zones are the 95% intervals for the FROG and FROG₀ models

5.4. Conclusion

Several global brGDGT calibrations for MAAT reconstruction in terrestrial settings have been proposed over the last years. Nevertheless, the uncertainty in brGDGT-based temperature estimates is still substantial, largely due to the influence of the various environmental variables in addition to MAAT on brGDGT distribution. A statistical clustering and analysis of the globally distributed brGDGT dataset allowed a refined understanding of the parameters affecting brGDGT distribution in soils at the global scale and highlighted potential confounding factors in addition to MAAT on the relative abundances of brGDGTs. Furthermore, the community index threshold proposed in a previous study (De Jonge et al., 2019), allowing to discriminate samples in which brGDGTs cope differently with temperatures, was refined (0.69) and will be of a major interest in the use of the MBT'_{5Me} index to reconstruct paleo-MAATs. In addition, thanks to this clustering, the present study confirmed that other factors than MAAT - pH, precipitation and thermal regime - could have a major influence on the distribution of these compounds. This shows the limitations of using a single index and a simple linear regression model to capture the response of brGDGTs to temperature changes.

A new improved calibration based on random forest algorithm was then proposed, the so-called Random **F**orest **R**egression for PaleO**M**AAT using brGDGTs (**FROG**). The FROG model, which is multi-factorial and non-parametric, allows overcoming the use of a single index and appears more representative of the environmental complexity. This new model appears to be more robust and accurate than previous global calibrations while being proposed on an extended soil and peat dataset. This is related to the fact that it is built on the whole suite of brGDGTs instead of only the brGDGTs included in the MBT'_{5Me} index, and that it takes into account the influence of other environmental parameters than MAAT on the brGDGT distribution, such as pH, thermal regime and MAP, as shown through RDA analyses. Finally, the FROG model was applied to two existing paleorecords and compared with available calibrations, showing its suitability for paleoreconstructions. Application of this new model should improve the accuracy of brGDGT-based MAAT reconstructions in terrestrial settings and could be applicable in environments where the MBT'_{5Me} shows some limitations because of potential confounding factors. As the random forest algorithm is adaptive, the FROG

FROG: IMPROVEMENT OF THE GLOBAL BRANCHED TETRAETHER TEMPERATURE CALIBRATION IN TERRESTRIAL SETTINGS USING MACHINE LEARNING

model could be further improved by the implementation of additional samples in the dataset and are freely available to the community through an R package ([paleoFROG](#)).

CONCLUSIONS & PERSPECTIVES

Conclusion

Comprendre la variabilité climatique et environnementale passée est capital pour pouvoir appréhender les changements futurs. Dans ce but, il est nécessaire d'avoir accès aux données environnementales passées. Celle-ci sont disponibles pour les deux derniers siècles (période instrumentale) Au-delà de cette période, il est nécessaire d'utiliser des indicateurs indirects de ces données.

A ce jour, les seuls marqueurs moléculaires terrestres utilisés en routine dans le cadre de paléo reconstructions, notamment de température et de pH, sont les tetraéthers d'origine bactérienne (brGDGTs). Cependant, les bactéries sources des brGDGTs ne sont pas encore identifiées, d'où une plus grande incertitude sur l'utilisation de ces marqueurs dans le cadre de paléo reconstitution.

En outre, contrairement aux milieux marins où les paramètres influençant les organismes sont restreints et en règle générale bien contraints (précipitations, pH de l'eau, température de l'eau, paramètres vitaux dans le cas de marqueurs « biologiques » ...), les paramètres d'influences sur les organismes sont beaucoup plus nombreux en milieu continental : effet de la latitude, de la topographie, de la végétation, du couvert pédologique, de la température du sol versus la température de l'air... En conséquence, les incertitudes sur les reconstitutions effectuées en milieu continental restent élevées.

De ces constats, découle l'enjeu majeur de ce travail de thèse : proposer de nouveaux marqueurs complémentaires et indépendants d'origine bactérienne en milieu terrestre. Nous nous sommes donc particulièrement intéressés aux acides gras 3-hydroxylés (3-AGH), composés synthétisés par les bactéries Gram-négatif et avons comparé leur réponse à des paramètres environnementaux à celle des brGDGTs.

En préambule de cette thèse, plusieurs grands objectifs avaient été proposés :

(1) Comprendre l'influence de la température de l'air et du pH du sol, ainsi que d'autres variables environnementales, sur la distribution des 3-AGH et des brGDGTs

(2) Obtenir des informations sur la vitesse d'adaptation des bactéries produisant des 3-AGH et des brGDGTs dans les sols.

(3) Tester l'applicabilité des 3-AGH comme marqueurs de température de l'air et de pH du sol en milieu terrestre à l'échelle mondiale. Développer une calibration globale

CONCLUSIONS & PERSPECTIVES

reliant l'abondance relative des 3-AGH avec la température de l'air d'une part et le pH des sols d'autre part.

(4) Proposer une nouvelle calibration globale de température de l'air basée sur les brGDGTs en milieu terrestre.

(5) Valider l'applicabilité des calibrations globales proposées sur les 3-AGH et les brGDGTs en les testant sur des archives sédimentaires connues.

(1) Comprendre l'influence de la température de l'air et du pH du sol, ainsi que d'autres variables environnementales, sur la distribution des 3-AGH et des brGDGTs

Afin de répondre au premier objectif de ma thèse, la distribution des 3-AGH et des brGDGTs a été déterminée au sein de 49 sols prélevés le long de transects altitudinaux (climatoséquences) dans les Alpes françaises (Massifs des Lautaret et des Bauges). L'intérêt majeur de ce site d'étude était de disposer d'un échantillonnage bien documenté dans le cas d'une Zone Atelier (Zone Atelier Alpes) (voir Chapitre 3.1). Des analyses de redondances ont révélé que l'influence des paramètres locaux (pH et, dans une moindre mesure, humidité du sol et granulométrie, liés à la végétation et aux types de sol) était prédominante sur la distribution des 3-AGH et des brGDGT. Plus important encore, dans cette zone des Alpes, nous avons montré que l'influence de ces différents paramètres était supérieure à celle de la TMAA. Or, ces lipides bactériens sont utilisés classiquement comme proxies de température ce qui explique la faiblesse des relations entre les TMAA et les indices basés sur les 3-AGH (RAN₁₅ & RAN₁₇) et dans une moindre mesure sur les brGDGT (MBT'_{5Me}). Les Alpes françaises apparaissaient comme un site représentatif de microenvironnements très contrastés, permettant d'explorer plus en détail les sources des incertitudes associés à l'utilisation de ces marqueurs. Ainsi, à l'aide des analyses de redondances, nous avons pu identifier les composés spécifiquement influencés par la température de l'air ou le pH du sol. **Grâce à cette approche statistique, nous avons développé des calibrations locales, basés sur ces molécules spécifiques et pu souligner l'importance de contraindre les facteurs environnementaux affectant la distribution des biomarqueurs lipidiques avant toute reconstruction paléoenvironnementale.**

(2) *Obtenir des informations sur la vitesse d'adaptation des bactéries produisant des 3-AGH et des brGDGTs dans les sols.*

Les 3-AGH et les brGDGTs sont biosynthétisés par des microorganismes, afin de maintenir une fluidité appropriée de leurs membranes. Peu d'informations sont disponibles à ce jour sur ces microorganismes, notamment sur la vitesse d'adaptation des bactéries produisant des 3-AGH et des brGDGTs, ainsi que sur le temps de renouvellement de ces composés dans les sols, en fonction de variations de températures de l'air. C'est pourquoi nous avons cherché à (1) comprendre la distribution des lipides bactériens en réponse aux variations du seul paramètre de température, et (2) fournir des informations sur la dynamique d'adaptation des communautés bactériennes et de leurs lipides membranaires face à un changement de température de l'air et donc du milieu de vie de ces microorganismes. Pour cela, nous avons étudié la distribution des 3-AGH et des brGDGTs au sein de sols incubés en conditions contrôlées à différentes températures durant 12 mois (Chapitre 3.2). Bien que les bactéries sources de 3-AGH aient pu modifier leur profil lipidique en réponse aux changements de température, aucuns liens entre les changements de températures et les abondances relatives des 3-AGH n'ont été observé. En effet, il est fait l'hypothèse que les lipides nouvellement produits au cours de l'expérience n'ont pas été détectés au sein du stock de lipides fossiles initialement présents dans les sols. **Ce résultat suggère que le temps de renouvellement du stock « fossile » de 3-AGH présent initialement dans les sols est supérieur à 12 mois, avant de refléter les nouvelles conditions de température. Concernant les tetraéthers bactériens, les proportions de tetraéthers tetraméthylés acycliques et de l'indice MBT'_{5Me} tendent à augmenter au bout de 12 mois, en accord avec la littérature.**

(3) *Tester l'applicabilité des 3-AGH comme marqueurs de température de l'air et de pH des sols en milieu terrestre à l'échelle mondiale. Développer une calibration globale reliant l'abondance relative des 3-AGH avec la température d'une part et le pH d'autre part.*

Les 3-AGH ont été analysés dans 54 échantillons de sols provenant de différents transects altitudinaux ont été analysés et les données ont été ajoutées à celles déjà disponibles dans la littérature, fournissant un jeu de données étendu de 168 échantillons. La réponse des 3-AGHs aux changements de températures et de pH dans ces sols a été

comparée à celle des brGDGTs, utilisés comme marqueurs «de référence» (Chapitre 4). Nous avons observé localement des relations linéaires fortes entre les indices dérivés des 3-AGH (RAN₁₅, RAN₁₇ et RIAN) et la TMAA/pH pour certains transects individuels, mais pas globalement (contrairement aux brGDGTs). Ce résultat suggère que ces marqueurs ne peuvent pas être utilisés à l'échelle globale, en utilisant des régressions linéaires simples.

Or, dans la première partie de cette thèse (Chapitre 3.1), nous avons montré l'intérêt majeur d'utiliser, non plus des indices, mais de travailler à l'échelle des abondances relatives pour reconstituer les températures ou le pH. Suivant cette idée, nous avons utilisé l'utilisation de différents modèles mathématiques multivariés (régression multiple, random forest et méthode des plus proches voisins (k-NN)). Ces modèles mettent en relation les abondances relatives des lipides avec la TMAA/pH et permettent de s'affranchir de l'utilisation des indices, et de prendre en compte les influences environnementales sur chaque composé spécifique. De plus, les modèles random forest et k-NN permettent aussi de prendre en compte les potentielles influences non-linéaires de la TMAA/pH sur les composés. En effet, dans l'environnement, différents paramètres peuvent influencer de façon concomitante les microorganismes, et s'influencer entre eux. Par exemple, il est classiquement considéré que l'humidité relative des sols est liée au régime de précipitation. Or, l'humidité relative des sols est également influencée par la granulométrie des sols (et donc le couvert pédologiques), la présence d'un système racinaire (et donc la végétation), la faune, Les modèles linéaires et paramétriques (ex : régression linéaire simple) ne permettent pas de refléter l'importante complexité des variables environnementales.

De plus, les mécanismes d'adaptation membranaire peuvent être dépendant des espèces bactériennes, de leur métabolisme, mais aussi à plus grande échelle des communautés bactériennes. Il est donc difficile de proposer un seul modèle d'adaptation homéovisqueuse, applicable à l'ensemble des microorganismes. Tout ceci justifiait l'utilisation préférentielle de modèles multivariés et non paramétriques pour estimer des valeurs de TMAA et de pH à partir de lipides membranaires comme les 3-AGH et les brGDGTs.

L'utilisation de ces différents modèles a mis en évidence de fortes relations globales entre les abondances relatives de 3-AGH et de TMAA/pH. Par la suite, nous

avons testé sur une archive de spéléothème ces nouvelles calibrations terrestres de TMAA. Cette archive avait fait l'objet d'une étude paléoclimatique basée sur une calibration locale des indices dérivés des 3-AGH. Seul le modèle random forest est apparu comme robuste et a montré des tendances similaires aux paléo-reconstructions précédentes et aux variations climatiques connues durant l'Holocène. Ceci démontre la nécessité de tester les nouvelles calibrations sur des archives bien documentées afin d'étudier leur fiabilité. Ce nouveau modèle global est prometteur pour les reconstructions de paléotempératures en milieu terrestre et pourrait être encore amélioré en analysant les 3-AGH dans un plus grand nombre de sols répartis dans le monde. **Ce travail de thèse a démontré le potentiel majeur des 3-AGH en tant que marqueurs TMAA/pH dans les environnements terrestres, ainsi que l'utilisation de modèles autres que les régressions linéaires simples.**

(4) Proposer une nouvelle calibration globale de température basée sur les brGDGTs en milieu terrestre.

L'ensemble des connaissances accumulées dans les chapitres 3 et 4 ont permis d'affiner notre compréhension de l'influence de différentes variables environnementales sur la distribution des lipides bactériens utilisés comme marqueurs d'environnements. J'ai utilisé l'ensemble des échantillons de sols et de tourbières disponibles dans la littérature dans lesquels ont été analysés ces composés ($n=767$) afin d'améliorer la calibration de température globale basée sur les brGDGTs. Les questionnements étaient les suivants :

(1) Quelles sont les variables environnementales qui influencent la structure et la distribution de ces composés ?

(2) Comment expliquer la dispersion des points sur les calibrations globales de TMAA, responsable d'une erreur quadratique moyenne importante sur les calibrations ($>4^{\circ}\text{C}$) ?

(3) Est-il possible de proposer une calibration améliorée de température, en se basant sur un modèle autre que la régression linéaire simple et l'utilisation de l'indice $\text{MBT}'_{5\text{Me}}$?

Nous avons pu confirmer que si dans certains cas, la température était le paramètre prédominant d'influence sur la distribution des brGDGTs, d'autres facteurs – le pH, les précipitations et la saisonnalité – pouvaient également influencer de façon majeure la

CONCLUSIONS & PERSPECTIVES

distribution de ces composés. Ceci permet alors d'expliquer l'erreur quadratique moyenne importante associée aux calibrations de brGDGTs. Il est classiquement considéré que l'abondance relative des isomères 5-méthyles des brGDGTs est majoritairement influencée par la température. Cependant, des analyses de redondances ont montré que ces composés étaient aussi influencés par d'autres variables environnementales, et que d'autres brGDGTs (comme par exemple les isomères 6-méthyles) pouvaient être influencés de façon majoritaire par la température. L'utilisation d'un indice n'utilisant que les composés 5-méthyles (MBT'_{5Me}) apparaissait alors restrictive. Ces éléments mettaient en lumière l'intérêt d'utiliser d'autres modèles que les régressions linéaires simples, et de s'affranchir de l'utilisation d'un indice. Nous proposons donc une nouvelle calibration globale des brGDGTs pour estimer les TMAA, basée sur le modèle random forest: random **F**orest **R**egression for Pale**O**TMAA using brGDGTs (**FROG**). Cette approche avait déjà montré son intérêt et sa robustesse pour déterminer une calibration globale des 3-AGH au sein de ce travail de thèse.

Le modèle FROG, multifactoriel et non-paramétrique, permet de s'affranchir de l'utilisation d'un indice unique et est davantage représentatif de la complexité environnementale. **Le modèle FROG est plus précis que les précédentes calibrations globales présentées dans la littérature** (De Jonge et al., 2014; Naafs et al., 2017b; Dearing Crampton-Flood et al., 2020) **tout en étant proposé sur un jeu de données plus étendu. D'autre part, ce modèle est tout aussi performant lorsqu'il considère l'ensemble des abondances relatives de brGDGTs au lieu des seuls brGDGT 5-méthyle, utilisés dans l'indice MBT'_{5Me}.**

(5) Valider l'applicabilité des calibrations globales proposées sur les 3-AGH et les brGDGTs en les testant sur des archives sédimentaires connues.

Comme présenté précédemment, les nouvelles calibrations globales de températures basées sur les 3-AGH et les brGDGTs ont pu être appliqués et validés avec succès sur plusieurs archives paléoclimatiques. Notamment, le modèle FROG a été appliqué avec succès à deux paléo-archives, démontrant ainsi l'utilité majeure de ce modèle, et de nouvelles approches statistiques pour établir des calibrations.

Ce travail de thèse a permis de répondre aux différents objectifs proposés. Il reste toutefois des questions en suspens auxquelles il sera nécessaire de répondre et ainsi poursuivre l'acquisition de connaissances.

Perspectives

Une première question en suspens concerne les températures « réelles » estimées grâce aux marqueurs de températures en milieu terrestre. Ainsi, il existe un biais potentiel entre les températures rencontrées par les microorganismes dans les sols et les températures de l'air. En première approximation, il est souvent considéré que ces températures covarient pour les premiers centimètres de sol (1-10 cm). Cependant, l'effet des précipitations, et notamment des chutes de neige, peuvent découpler ces systèmes sol/atmosphère, en isolant le sol à une température de 0°C alors même que les températures de l'air peuvent être beaucoup plus froides (par exemple, Choler et al., 2018). Dans le cadre de cette thèse, nous avons cherché à réduire ce biais, en utilisant des capteurs de températures du sol. Ces capteurs permettent d'accéder à des données de températures hautes résolutions (heure par heure), et de s'approcher au maximum des températures du sol dans lequel les microorganismes produisant les 3-AGH et les brGDGTs évoluent. Néanmoins, il reste difficile d'avoir accès à ce type de données pour de nombreux échantillons, dans le cadre d'une calibration par exemple (il faudrait pouvoir disposer de capteurs partout). En outre, un intérêt des marqueurs d'environnement pour les climatologues est de pouvoir reconstituer les températures de l'air atmosphérique, et non la température du sol. Il serait alors nécessaire de pouvoir estimer précisément la différence entre les températures estimées à l'aide de ces marqueurs (basées sur des températures atmosphériques), et les températures « réelles » du sol qui ont une influence directe sur les microorganismes. Cette différence pourrait expliquer en partie les incertitudes associées aux estimations de températures effectuées à l'aide de ces marqueurs. Pour cela, il serait envisageable de travailler sur les échantillons des massifs des Bauges et du Lautaret pour lesquelles des mesures de la température de l'air et de la température du sol sont disponibles. Nous pourrions alors étudier le décalage entre les températures estimés à l'aide des lipides bactériens et les températures réelles, pour ensuite appliquer cette correction à d'autres calibrations de températures basées sur ces mêmes lipides.

De plus, de précédentes études suggéraient par le passé que les brGDGTs pouvaient être influencé par les variations saisonnières de températures. Au cours de ce travail, nous avons aussi pu montrer un effet potentiel de la saisonnalité sur la distribution des brGDGTs. On peut supposer qu'au cours de la période hivernale, les microorganismes

sont moins actifs, et qu'ils sont donc plus sensibles aux variations de températures en saisons estivales. Les températures estimées à l'aide des lipides bactériens seraient donc des températures saisonnières plutôt que des températures annuelles moyennes. Il s'agit d'une question majeure, source d'une partie de l'incertitude sur les paléo-reconstitutions utilisant ces marqueurs. Ceci nécessite de mieux contraindre l'effet de la saisonnalité sur les microorganismes. Une approche serait de travailler sur des estimations de températures saisonnières (température estivale versus température hivernale ; température du mois le plus chaud versus température du mois le froid). Ainsi, plutôt que d'établir des calibrations de températures moyennes annuelles, il serait nécessaire de proposer des calibrations de températures saisonnières, en prenant en compte l'optimum de vie et de croissances des microorganismes sources des composés lipidiques. Aussi et comme nous l'avons proposé de façon préliminaire dans cette thèse, il serait possible d'utiliser des indicateurs de la saisonnalité, comme le nombre de jours de gelées dans l'année (FRS) ou encore le degré-jour de croissance (Growing Degree Days ; GDD). Ces indicateurs présentent l'avantage majeur de montrer directement les périodes durant lesquelles les organismes (macro ou micro) peuvent être actifs au sein de l'environnement. Par exemple, les bactéries présentes dans un sol tropical et tempéré sont susceptibles de se développer tout au long de l'année avec des températures toujours au-dessus de 0°C. A l'inverse, il apparaît difficile d'envisager que dans les sols situés aux latitudes tempérée à polaires, où les hivers sont caractérisés par des températures inférieures à 0°C, les plantes ou les microorganismes puissent se développer pendant toute l'année. Ainsi, pour ces échantillons, les lipides bactériens ne sont sans doute pas produits en quantités équivalentes tout au long de l'année.

Pour infirmer ou confirmer ceci, lors des expériences de microcosmes de sol, la distinction entre le stock de lipides intacts (fraichement produits) et le stock de lipides de cœur (lipides fossiles), pour les tétraéthers bactériens et pour les 3-AGH est possible. Cette approche permettrait d'observer l'évolution de la distribution et de la concentration des lipides intacts au sein d'échantillons de sols, placés en incubation, à différents pas de temps et à différentes températures. Néanmoins, cette méthode de séparation existe déjà dans la littérature pour les brGDGTs, mais pas pour les 3-AGH. En outre, il serait nécessaire d'utiliser des au moins des triplicats plutôt que des duplicats d'échantillons pour proposer des analyses statistiques robustes. De plus, il serait nécessaire d'analyser

les autres pas de temps des microcosmes réalisés pendant cette thèse (1, 6 et 9 mois) que réaliser des analyses isotopiques et élémentaires sur ces échantillons afin d'étudier l'évolution de ces paramètres au cours de l'incubation.

Il serait aussi possible d'évaluer l'activité, la croissance et le temps de renouvellement des communautés bactériennes et des lipides membranaires associés (3-AGH et tétraéthers bactériens) dans ces échantillons de sols par marquage isotopique (Stable Isotope Probing, SIP). Cette méthode permettrait de suivre l'incorporation de substrats marqués au sein des lipides bactériens, et de cibler les bactéries actives dans le sol et de s'affranchir des cultures de souches pures (par exemple Boschker et al., 1998, Kopf et al., 2016). Par exemple par Huguet et al. (2017), ont proposé une approche en double SIP, avec des ajouts d'eau deutérée (D₂O) et de carbone inorganique dissous marqué au carbone ¹³C. Cette approche « double SIP » permettrait dans un premier temps de quantifier la proportion et la vitesse de production des lipides nouvellement produit au cours de l'expérience. L'incorporation du substrat inorganique marqué au carbone ¹³C permettrait également l'obtention d'information sur les voies de biosynthèse des lipides (autotrophie vs hétérotrophie).

En parallèle, des cultures de souches pures de bactéries à Gram-négatif, sources des 3-AGHs, pourraient être réalisées. L'étude de souches pures de bactéries à Gram-négatifs permettrait d'acquérir des connaissances plus fines sur les voies métaboliques et de biosynthèse de ces microorganismes. Concernant les tétraéthers bactériens, leurs organismes sources sont encore largement inconnus (Weijers et al., 2009a; Sinninghe Damsté et al., 2018). Ce manque de connaissance empêche de travailler à l'heure actuelle à l'échelle des souches pures, excluant l'étude approfondie de la réponse des brGDGTs en fonction de paramètres environnementaux spécifiques en milieu contrôlé.

Dans le cadre de cette thèse, un autre questionnement a été soulevé : les variations de distribution des lipides bactériens en réponse aux changements environnementaux proviennent-elles du seul mécanisme homéovisqueux et/ou d'un changement des communautés bactériennes ? Le manque de connaissances sur les organismes sources des 3-AGH et des brGDGT ne permet pas de statuer de façon définitive à cette question. Ceci est d'autant plus complexe que toutes les souches de bactéries ne biosynthétisent pas les mêmes homologues (différences de longueurs de chaînes, insaturations, ramifications...),

que la synthèse de lipides ne dépend pas forcément des réponses aux variations environnementales (adaptation pathogène des bactéries Gram-négatifs par exemple). Néanmoins, un travail de recensement de la diversité bactérienne au sein des 49 sols des Alpes françaises est en cours au sein du laboratoire METIS. Ce travail permettra de lier la diversité bactérienne présente dans les sols (à l'échelle de l'espèce ou du genre) avec les variations de distributions en 3-AGH et en brGDGT.

Enfin, nous nous sommes concentrés sur le développement de marqueurs bactériens en milieu terrestre. Dans le futur, il pourrait être intéressant de développer ces marqueurs en milieu lacustres, voir en milieu marin dans le cas des 3-AGH. En effet, l'avantage de ces composés est qu'ils sont ubiquistes, et cela permettrait d'utiliser un seul et même marqueur pour différents types d'environnements. Néanmoins, il est possible que cette approche se heurte à une limite. En effet, Yang et al., (2020) ont récemment démontré que les 3-AGH pouvaient être utilisés comme marqueurs de température en milieu marin, mais moyennant l'utilisation d'un autre indice (indice RAN₁₃). Se pose alors la question de la multiplication des indices, afin de pouvoir utiliser ces lipides dans tous les milieux. Un début de réponse pourrait être l'utilisation d'autres modèles que les régressions linéaires simples basés sur les indices, comme proposée dans cette thèse. Une perspective future pourrait alors être de développer de nouvelles calibrations de températures ou de pH. Ces calibrations pourraient être basés par exemple sur l'algorithme random forest (issu du machine learning), nécessitant simplement d'intégrer les abondances relatives de ces composés. Cette approche pourrait alors être étendue à d'autres marqueurs moléculaires, comme les GDGTs archéens. Il serait aussi possible de reconstituer d'autres variables que la température ou le pH, comme par exemple les précipitations, ou les paléovégétations (Peaple et al., 2021). L'utilisation d'autres modèles basés sur d'autres algorithmes n'est pas exclue, comme par exemple les modèles bayésiens, déjà proposés par quelques auteurs (Tierney and Tingley, 2014; Dearing Crampton-Flood et al., 2020). Ces modèles ont permis de proposer des calibrations plus précises et robustes. Enfin, afin d'améliorer la fiabilité des reconstitutions paléoclimatiques, il apparaît nécessaire de systématiser l'utilisation et la comparaison de marqueurs autres que moléculaires, indépendants et complémentaires (assemblages polliniques, marqueurs inorganiques...).

CONCLUSIONS & PERSPECTIVES

Pour finir, on peut émettre l'hypothèse que la multiplicité des variables environnementales pouvant influencer les organismes sources des brGDGTs et des 3-AGH, ainsi que la méconnaissance du mécanisme d'adaptation homéoviscieux au sein des membranes de ces microorganismes, sont à l'origine d'une "sensibilité limite " de ces composés en tant que marqueur de TMAA. On peut supposer qu'en raison de toute cette complexité, il sera difficile de continuer à augmenter la précision associée aux calibrations de ces marqueurs de la TMAA, ce fait étant mis en évidence par la redondance des calibrations avec une EQM de plus ou moins 4°C. Peut-être que cette précision est un seuil pour l'instant indépassable, étant donné la complexité des environnements continentaux et la sensibilité de l'adaptation membranaire des microorganismes.

Bien sûr, il s'agit d'une hypothèse, et dans le futur, des développements analytiques conduisant à une nouvelle séparation des isomères de position des brGDGTs comme proposé par exemple par De Jonge et al., (2013, 2014) ou Ding et al., (2016), ou une compréhension plus fine des changements de distributions des 3-AGH face aux variations des paramètres environnementaux permettraient de limiter les incertitudes sur les reconstitutions utilisant ces marqueurs. Cela pourrait alors permettre d'abaisser ce seuil de précision des calibrations de marqueurs terrestres en dessous de 4°C. Par ailleurs, des études plus approfondies sur les organismes sources des brGDGTs et des 3-AGH et leurs mécanismes d'adaptation membranaire face aux variables environnementales, en conditions contrôlées, pourront également permettre d'améliorer cette précision. Le développement d'autres marqueurs de température et de pH, complémentaires et indépendants, des brGDGTs et des 3-AGH est également une piste à envisager pour améliorer et rendre plus robuste l'utilisation des marqueurs organiques de température en milieu terrestre.

RÉFÉRENCES BIBLIOGRAPHIQUES

Adl S. M., Simpson A. G. B., Farmer M. A., Andersen R. A., Anderson O. R., Barta J. R., Bowser S. S., Brugerolle G., Fensome R. A., Fredericq S., James T. Y., Karpov S., Kugrens P., Krug J., Lane C. E., Lewis L. A., Lodge J., Lynn D. H., Mann D. G., Mccourt R. M., Mendoza L., Moestrup Ø., Mozley-Standridge S. E., Nerad T. A., Shearer C. A., Smirnov A. V., Spiegel F. W. and Taylor M. F. J. R. (2005) The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *Journal of Eukaryotic Microbiology* **52**, 399–451.

Adl S. M., Simpson A. G. B., Lane C. E., Lukeš J., Bass D., Bowser S. S., Brown M. W., Burki F., Dunthorn M., Hampl V., Heiss A., Hoppenrath M., Lara E., Gall L. le, Lynn D. H., McManus H., Mitchell E. A. D., Mozley-Stanridge S. E., Parfrey L. W., Pawlowski J., Rueckert S., Shadwick L., Schoch C. L., Smirnov A. and Spiegel F. W. (2012) The Revised Classification of Eukaryotes. *Journal of Eukaryotic Microbiology* **59**, 429–514.

Adl S. M., Bass D., Lane C. E., Lukeš J., Schoch C. L., Smirnov A., Agatha S., Berney C., Brown M. W., Burki F., Cárdenas P., Čepička I., Chistyakova L., Campo J. del, Dunthorn M., Edvardsen B., Eglit Y., Guillou L., Hampl V., Heiss A. A., Hoppenrath M., James T. Y., Karnkowska A., Karpov S., Kim E., Kolisko M., Kudryavtsev A., Lahr D. J. G., Lara E., Gall L. L., Lynn D. H., Mann D. G., Massana R., Mitchell E. A. D., Morrow C., Park J. S., Pawlowski J. W., Powell M. J., Richter D. J., Rueckert S., Shadwick L., Shimano S., Spiegel F. W., Torruella G., Youssef N., Zlatogursky V. and Zhang Q. (2019) Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology* **66**, 4–119.

Albers S.-V. and Meyer B. H. (2011) The archaeal cell envelope. *Nature Reviews Microbiology* **9**, 414–426.

Amann R. I., Ludwig W. and Schleifer K. H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Mol. Biol. Rev.* **59**, 143–169.

Anand P., Elderfield H. and Conte M. H. (2003) Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18**.

Anderson V. J., Shanahan T. M., Saylor J. E., Horton B. K. and Mora A. R. (2014a) Sources of local and regional variability in the MBT'/CBT paleotemperature proxy:

REFERENCES BIBLIOGRAPHIQUES

Insights from a modern elevation transect across the Eastern Cordillera of Colombia. *Organic Geochemistry* **69**, 42–51.

Badger M. P. S., Chalk T. B., Foster G. L., Bown P. R., Gibbs S. J., Sexton P. F., Schmidt D. N., Pälke H., Mackensen A. and Pancost R. D. (2019) Insensitivity of alkenone carbon isotopes to atmospheric CO₂ at low to moderate CO₂ levels. *Climate of the Past* **15**, 539–554.

Ballantyne A. P., Greenwood D. R., Damsté J. S. S., Csank A. Z., Eberle J. J. and Rybczynski N. (2010) Significantly warmer Arctic surface temperatures during the Pliocene indicated by multiple independent proxies. *Geology* **38**, 603–606.

Balleza D., Garcia-Arribas A. B., Sot J., Ruiz-Mirazo K. and Goñi F. M. (2014) Ether- versus Ester-Linked Phospholipid Bilayers Containing either Linear or Branched Apolar Chains. *Biophysical Journal* **107**, 1364–1374.

Bangham A. D. and Horne R. W. (1964) Negative staining of phospholipids and their structural modification by surface-active agents as observed in the electron microscope. *Journal of Molecular Biology* **8**, 660-IN10.

de Bar M. W., Rampen S. W., Hopmans E. C., Sinninghe Damsté J. S. and Schouten S. (2019) Constraining the applicability of organic paleotemperature proxies for the last 90 Myrs. *Organic Geochemistry* **128**, 122–136.

Barns S. M., Delwiche C. F., Palmer J. D. and Pace N. R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences* **93**, 9188–9193.

Bartlett M. G., Chapman D. S. and Harris R. N. (2006) A Decade of Ground–Air Temperature Tracking at Emigrant Pass Observatory, Utah. *J. Climate* **19**, 3722–3731.

Beales N. (2004) Adaptation of Microorganisms to Cold Temperatures, Weak Acid Preservatives, Low pH, and Osmotic Stress: A Review. *Comprehensive Reviews in Food Science and Food Safety* **3**, 1–20.

Bhandari J. and Zhang Y. (2019) Effect of altitude and soil properties on biomass and plant richness in the grasslands of Tibet, China, and Manang District, Nepal. *Ecosphere* **10**, e02915.

Bhat U. R. and Carlson R. W. (1992) A new method for the analysis of amide-linked hydroxy fatty acids in lipid-As from gram-negative bacteria. *Glycobiology* **2**, 535–539.

Bianchi T. S. and Canuel E. A. (2011) *Chemical biomarkers in aquatic ecosystems.*, Princeton University Press, Princeton.

Bidigare R. R., Fluegge A., Freeman K. H., Hanson K. L., Hayes J. M., Hollander D., Jasper J. P., King L. L., Laws E. A., Milder J., Millero F. J., Pancost R., Popp B. N., Steinberg P. A. and Wakeham S. G. (1997) Consistent fractionation of ^{13}C in nature and in the laboratory: Growth-rate effects in some haptophyte algae. *Global Biogeochemical Cycles* **11**, 279–292.

Bini M., Zanchetta G., Persoiu A., Cartier R., Catala A., Cacho I., Dean J. R., Di Rita F., Drysdale R. N., Finné M., Isola I., Jalali B., Lirer F., Magri D., Masi A., Marks L., Mercuri A. M., Peyron O., Sadori L., Sicre M.-A., Welc F., Zielhofer C. and Brisset E. (2019) The 4.2 ka BP Event in the Mediterranean region : an overview. *Climate of the Past* **15**, 555–577.

Bлга C. I., Reichart G.-J., Heiri O. and Sinninghe Damsté J. S. (2009) Tetraether membrane lipid distributions in water-column particulate matter and sediments: a study of 47 European lakes along a north–south transect. *J Paleolimnol* **41**, 523–540.

Bлга C. I., Reichart G.-J., Lotter A. F., Anselmetti F. S. and Damsté J. S. S. (2013) A TEX86 lake record suggests simultaneous shifts in temperature in Central Europe and Greenland during the last deglaciation. *Geophysical Research Letters* **40**, 948–953.

Bonanomi G., Zotti M., Mogavero V., Cesarano G., Saulino L., Rita A., Tesei G., Allegrezza M., Saracino A. and Allevato E. (2020) Climatic and anthropogenic factors explain the variability of *Fagus sylvatica* treeline elevation in fifteen mountain groups across the Apennines. *Forest Ecosystems* **7**, 5.

Bond G., Kromer B., Beer J., Muscheler R., Evans M. N., Showers W., Hoffmann S., Lotti-Bond R., Hajdas I. and Bonani G. (2001) Persistent Solar Influence on North Atlantic Climate During the Holocene. *Science* **294**, 2130–2136.

Boschker H. T. S., Nold S. C., Wellsbury P., Bos D., de Graaf W., Pel R., Parkes R. J. and Cappenberg T. E. (1998) Direct linking of microbial populations to specific biogeochemical processes by ^{13}C -labelling of biomarkers. *Nature* **392**, 801–805.

Boudouresque C.-F. (2015) Taxonomy and Phylogeny of Unicellular Eukaryotes. In *Environmental Microbiology: Fundamentals and Applications: Microbial Ecology* (eds. J.-C. Bertrand, P. Caumette, P. Lebaron, R. Matheron, P. Normand, and T. Sime- Ngando). Springer Netherlands, Dordrecht. pp. 191–257.

REFERENCES BIBLIOGRAPHIQUES

Bounemoura, Z., Lambert, K., Cadel, G., Choler, P., Manneville, O., Michalet, R., 1998. Influence 814 des facteurs édapho-climatiques sur la distribution des pelouses alpines dans le massif du 815 Galibier (Alpes Françaises). *ÉCOLOGIE-BRUNOY* **29**, 53–58.

Braak C. J. F. ter and Smilauer P. (2002) *CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (version 4.5)*., www.canoco.com, Ithaca NY, USA.

Brandenburg K., Funari S. S., Koch M. H. J. and Seydel U. (1999) Investigation into the Acyl Chain Packing of Endotoxins and Phospholipids under Near Physiological Conditions by WAXS and FTIR Spectroscopy. *Journal of Structural Biology* **128**, 175–186.

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

Brassell S. C. and Dumitrescu M. (2004) Recognition of alkenones in a lower Aptian porcellanite from the west-central Pacific. *Organic Geochemistry* **35**, 181–188.

Breiman L. (2001) Random Forests. *Machine Learning* **45**, 5–32.

Briggs D. E. G. and Summons R. E. (2014) Ancient biomolecules: Their origins, fossilization, and role in revealing the history of life. *BioEssays* **36**, 482–490.

Brochier-Armanet C., Boussau B., Gribaldo S. and Forterre P. (2008) Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology* **6**, 245–252.

Brown S. J. and Elderfield H. (1996) Variations in Mg/Ca and Sr/Ca ratios of planktonic foraminifera caused by postdepositional dissolution: Evidence of shallow Mg-dependent dissolution. *Paleoceanography* **11**, 543–551.

Bryant C. E., Spring D. R., Gangloff M. and Gay N. J. (2010) The molecular basis of the host response to lipopolysaccharide. *Nature Reviews Microbiology* **8**, 8–14.

Buckles L. K., Weijers J. W. H., Tran X.-M., Waldron S. and Sinninghe Damsté J. S. (2014) Provenance of tetraether membrane lipids in a large temperate lake (Loch Lomond, UK): implications for glycerol dialkyl glycerol tetraether (GDGT)-based palaeothermometry. *Biogeosciences* **11**, 5539–5563.

Cao J., Rao Z., Jia G., Xu Q. and Chen F. (2017) A 15ka pH record from an alpine lake in north China derived from the cyclization ratio index of aquatic brGDGTs and its paleoclimatic significance. *Organic Geochemistry* **109**, 31–46.

Carlson B. Z., Corona M. C., Dentant C., Bonet R., Thuiller W. and Choler P. (2017) Observed long-term greening of alpine vegetation—a case study in the French Alps. *Environ. Res. Lett.* **12**, 114006.

Carter M. R., Gregorich E. G. and Gregorich E. G. (2007) *Soil Sampling and Methods of Analysis.*, CRC Press.

Castañeda I. S. and Schouten S. (2011) A review of molecular organic proxies for examining modern and ancient lacustrine environments. *Quaternary Science Reviews* **30**, 2851–2891.

Cavalier-Smith T. (1998) A revised six-kingdom system of life. *Biological Reviews* **73**, 203–266.

Chen Y., Zheng F., Chen S., Liu H., Phelps T. J. and Zhang C. (2018) Branched GDGT production at elevated temperatures in anaerobic soil microcosm incubations. *Organic Geochemistry* **117**, 12–21.

Cheng J. Y. W., Hui E. L. C. and Lau A. P. S. (2012) Bioactive and total endotoxins in atmospheric aerosols in the Pearl River Delta region, China. *Atmospheric Environment* **47**, 3–11.

Cheng-jim, J.I., Yang Y., Han W., HE Y.-F., Smith J. and SMITH P. (2014) Climatic and Edaphic Controls on Soil pH in Alpine Grasslands on the Tibetan Plateau, China: A Quantitative Analysis. *Pedosphere* **24**, 39–44.

Choler P. (2018) Winter soil temperature dependence of alpine plant distribution: Implications for anticipating vegetation changes under a warming climate. *Perspectives in Plant Ecology, Evolution and Systematics* **30**, 6–15.

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

Coffinet S. (2015) Validation et application de biomarqueurs lipidiques pour la reconstruction des changements environnementaux en Afrique de l'Est. phdthesis, Université Pierre et Marie Curie - Paris VI.

Coffinet S., Huguet A., Pedentchouk N., Bergonzini L., Omuombo C., Williamson D., Anquetil C., Jones M., Majule A., Wagner T. and Derenne S. (2017) Evaluation of branched GDGTs and leaf wax n-alkane $\delta^2\text{H}$ as (paleo) environmental proxies in East Africa. *Geochimica et Cosmochimica Acta* **198**, 182–193.

Coffinet S., Huguet A., Bergonzini L., Pedentchouk N., Williamson D., Anquetil C., Gałka M., Kołaczek P., Karpińska-Kołaczek M., Majule A., Laggoun-Défarge F., Wagner T. and Derenne S. (2018) Impact of climate change on the ecology of the Kyambangunguru crater marsh in southwestern Tanzania during the Late Holocene. *Quaternary Science Reviews* **196**, 100–117.

Comont L., Laggoun-Défarge F. and Disnar J.-R. (2006) Evolution of organic matter indicators in response to major environmental changes: The case of a formerly cut-over peat bog (Le Russey, Jura Mountains, France). *Organic Geochemistry* **37**, 1736–1751.

Conant R. T., Ryan M. G., Ågren G. I., Birge H. E., Davidson E. A., Eliasson P. E., Evans S. E., Frey S. D., Giardina C. P., Hopkins F. M., Hyvönen R., Kirschbaum M. U. F., Lavalley J. M., Leifeld J., Parton W. J., Steinweg J. M., Wallenstein M. D., Wetterstedt J. Å. M. and Bradford M. A. (2011) Temperature and soil organic matter decomposition rates – synthesis of current knowledge and a way forward. *Global Change Biology* **17**, 3392–3404.

Cortese G. and Abelman A. (2002) Radiolarian-based paleotemperatures during the last 160 kyr at ODP Site 1089 (Southern Ocean, Atlantic Sector). *Palaeogeography, Palaeoclimatology, Palaeoecology* **182**, 259–286.

Cranwell P. A. (1985) Long-chain unsaturated ketones in recent lacustrine sediments. *Geochimica et Cosmochimica Acta* **49**, 1545–1551.

Crave A. and Gascuel-Oudou C. (1997) The Influence of Topography on Time and Space Distribution of Soil Surface Water Content. *Hydrological Processes* **11**, 203–210.

Crosta X. and Koç N. (2007) Chapter Eight Diatoms: From Micropaleontology to Isotope Geochemistry. In *Developments in Marine Geology* (eds. C. Hillaire-Marcel and A. De Vernal). Proxies in Late Cenozoic Paleoceanography. Elsevier. pp. 327–369.

Dang X., Xue J., Yang H. and Xie S. (2016a) Environmental impacts on the distribution of microbial tetraether lipids in Chinese lakes with contrasting pH:

Implications for lacustrine paleoenvironmental reconstructions. *Sci. China Earth Sci.* **59**, 939–950.

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

Davidson E. A., Belk E. and Boone R. D. (1998) Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. *Global Change Biology* **4**, 217–227.

Davtian N., Ménot G., Bard E., Poulénard 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 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.

De Jonge C., Hopmans E. C., Zell C. I., Kim J.-H., Schouten S. and Sinninghe Damsté J. S. (2016) Corrigendum to “Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction” [Geochim. Cosmochim. Acta 141 (2014) 97-112]. *Geochimica et Cosmochimica Acta* **183**, 276–278.

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

Dearing Crampton-Flood E., Peterse F., Munsterman D. and Sinninghe Damsté J. S. (2018) Using tetraether lipids archived in North Sea Basin sediments to extract North

REFERENCES BIBLIOGRAPHIQUES

Western European Pliocene continental air temperatures. *Earth and Planetary Science Letters* **490**, 193–205.

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

Deconinck J.-F. (2014) *Paléoclimats: l'enregistrement des variations climatiques.*, Dedysch S. N., Pankratov T. A., Belova S. E., Kulichevskaya I. S. and Liesack W. (2006) Phylogenetic Analysis and In Situ Identification of Bacteria Community Composition in an Acidic Sphagnum Peat Bog. *Appl. Environ. Microbiol.* **72**, 2110–2117.

deMenocal P. B. (2001) Cultural Responses to Climate Change During the Late Holocene. *Science* **292**, 667–673.

Deng L., Jia G., Jin C. and Li S. (2016) Warm season bias of branched GDGT temperature estimates causes underestimation of altitudinal lapse rate. *Organic Geochemistry* **96**, 11–17.

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.

Denisko D. and Hoffman M. M. (2018) Classification and interaction in random forests. *Proceedings of the National Academy of Sciences USA* **115**, 1690–1692.

Ding S., Xu Y., Wang Y., He Y., Hou J., Chen L. and He J.-S. (2015) Distribution of branched glycerol dialkyl glycerol tetraethers in surface soils of the Qinghai–Tibetan Plateau: implications of brGDGTs-based proxies in cold and dry regions. *Biogeosciences* **12**, 3141–3151.

Ding S., Schwab V. F., Ueberschaar N., Roth V.-N., Lange M., Xu Y., Gleixner G. and Pohnert G. (2016) Identification of novel 7-methyl and cyclopentanyl branched glycerol dialkyl glycerol tetraethers in lake sediments. *Organic Geochemistry* **102**, 52–58.

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.

Djukic I., Zehetner F., Tatzber M. and Gerzabek M. H. (2010) Soil organic-matter stocks and characteristics along an Alpine elevation gradient. *Journal of Plant Nutrition and Soil Science* **173**, 30–38.

Dowhan W. (1997) Molecular basis for membrane phospholipid diversity: Why Are There So Many Lipids? *Annual Review of Biochemistry* **66**, 199–232.

Drotz S. H., Sparrman T., Nilsson M. B., Schleucher J. and Öquist M. G. (2010) Both catabolic and anabolic heterotrophic microbial activity proceed in frozen soils. *Proceedings of the National Academy of Sciences* **107**, 21046–21051.

Du B., Liu C., Kang H., Zhu P., Yin S., Shen G., Hou J. and Ilvesniemi H. (2014) Climatic Control on Plant and Soil $\delta^{13}\text{C}$ along an Altitudinal Transect of Lushan Mountain in Subtropical China: Characteristics and Interpretation of Soil Carbon Dynamics. *PLOS ONE* **9**, e86440.

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. *Climate of the Past* **16**, 2599–2617.

Durand Y., Laternser M., Giraud G., Etchevers P., Lesaffre B. and Mérindol L. (2009) Reanalysis of 44 Yr of Climate in the French Alps (1958–2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation. *Journal of Applied Meteorology and Climatology* **48**, 429–449.

Eglinton T. I., Conte M. H., Eglinton G. and Hayes J. M. (2001) Proceedings of a workshop on alkenone-based paleoceanographic indicators. *Geochemistry, Geophysics, Geosystems* **2**.

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

Elkins J. G., Podar M., Graham D. E., Makarova K. S., Wolf Y., Randau L., Hedlund B. P., Brochier-Armanet C., Kunin V., Anderson I., Lapidus A., Goltsman E., Barry K., Koonin E. V., Hugenholtz P., Kyrpides N., Wanner G., Richardson P., Keller M. and Stetter K. O. (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences* **105**, 8102–8107.

Elling F. J., Könneke M., Lipp J. S., Becker K. W., Gagen E. J. and Hinrichs K.-U. (2014) Effects of growth phase on the membrane lipid composition of the thaumarchaeon

REFERENCES BIBLIOGRAPHIQUES

Nitrosopumilus maritimus and their implications for archaeal lipid distributions in the marine environment. *Geochimica et Cosmochimica Acta* **141**, 579–597.

Emiliani C. (1955) Pleistocene Temperatures. *The Journal of Geology* **63**, 538–578.

Erez J. and Luz B. (1983) Experimental paleotemperature equation for planktonic foraminifera. *Geochimica et Cosmochimica Acta* **47**, 1025–1031.

Ernst N., Peterse F., Breitenbach S. F. M., Syiemlieh H. J. and Eglinton T. I. (2013) Biomarkers record environmental changes along an altitudinal transect in the wettest place on Earth. *Organic Geochemistry* **60**, 93–99.

Ernst R., Ejsing C. S. and Antonny B. (2016) Homeoviscous Adaptation and the Regulation of Membrane Lipids. *Journal of Molecular Biology* **428**, 4776–4791.

Feng X., D'Andrea W. J., Zhao C., Xin S., Zhang C. and Liu W. (2019) Evaluation of leaf wax δD and soil brGDGTs as tools for paleoaltimetry on the southeastern Tibetan Plateau. *Chemical Geology* **523**, 95–106.

Feng Z., Sun A., Abdusalih N., Ran M., Kurban A., Lan B., Zhang D. and Yang Y. (2017) Vegetation changes and associated climatic changes in the southern Altai Mountains within China during the Holocene. *The Holocene* **27**, 683–693.

Fierer N. and Jackson R. B. (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* **103**, 626–631.

Fitch W. M. and Upper K. (1987) The Phylogeny of tRNA Sequences Provides Evidence for Ambiguity Reduction in the Origin of the Genetic Code. *Cold Spring Harb Symp Quant Biol* **52**, 759–767.

Fournier G. P. and Poole A. M. (2018) A Briefly Argued Case That Asgard Archaea Are Part of the Eukaryote Tree. *Front Microbiol* **9**.

Gandouin E., Ponel P., Franquet E., Van Vliet-Lanoë B., Andrieu-Ponel V., Keen D. H., Brulhet J. and Brocandel M. (2007) Chironomid responses (Insect: Diptera) to Younger Dryas and Holocene environmental changes in a river floodplain from northern France (St-Momelin, St-Omer basin). *The Holocene* **17**, 331–347.

Gangopadhyay S., Harding B. L., Rajagopalan B., Lukas J. J. and Fulp T. J. (2009) A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles. *Water Resources Research* **45**.

Gao H., Zhu C. and Xu W. (2007) Environmental change and cultural response around 4200 cal. yr BP in the Yishu River Basin, Shandong. *J GEOGR SCI* **17**, 285–292.

- Gao L., Nie J., Clemens S., Liu W., Sun J., Zech R. and Huang Y. (2012) The importance of solar insolation on the temperature variations for the past 110kyr on the Chinese Loess Plateau. *Palaeogeography, Palaeoclimatology, Palaeoecology* **317–318**, 128–133.
- Gibbons A. (1993) How the Akkadian Empire Was Hung Out to Dry. *Science* **261**, 985.
- Goldman E. and Green L. H. (2015) *Practical Handbook of Microbiology.*, CRC Press.
- Gómez-Plaza A., Martínez-Mena M., Albaladejo J. and Castillo V. M. (2001) Factors regulating spatial distribution of soil water content in small semiarid catchments. *Journal of Hydrology* **253**, 211–226.
- Gorham E. (1991) Northern Peatlands: Role in the Carbon Cycle and Probable Responses to Climatic Warming. *Ecological Applications* **1**, 182–195.
- Gornitz V. (2008) *Encyclopedia of Paleoclimatology and Ancient Environments.*, Springer Science & Business Media.
- Gray M. W. (2017) Lynn Margulis and the endosymbiont hypothesis: 50 years later. *MBoC* **28**, 1285–1287.
- Greenwood D. and West C. (2016) A fossil coryphoid palm from the Paleocene of western Canada. *Review of Palaeobotany and Palynology* **239**, 55–65.
- Gutiérrez-Girón A., Díaz-Pinés E., Rubio A. and Gavilán R. G. (2015) Both altitude and vegetation affect temperature sensitivity of soil organic matter decomposition in Mediterranean high mountain soils. *Geoderma* **237–238**, 1–8.
- Guy L. and Ettema T. J. G. (2011) The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends in Microbiology* **19**, 580–587.
- Haque M. A. and Russell N. J. (2004) Strains of *Bacillus cereus* vary in the phenotypic adaptation of their membrane lipid composition in response to low water activity, reduced temperature and growth in rice starch. *Microbiology*, **150**, 1397–1404.
- Harris I., Jones P. D., Osborn T. J. and Lister D. H. (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset: Updated high-resolution grids of monthly climatic observations. *Int. J. Climatol.* **34**, 623–642.
- Hasle G. R., Syvertsen E. E., Steidinger K. A., Tangen K. and Tomas C. R. (1996) *Identifying Marine Diatoms and Dinoflagellates.*, Elsevier.

REFERENCES BIBLIOGRAPHIQUES

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

Hazel J. R. and Eugene Williams E. (1990) The role of alterations in membrane lipid composition in enabling physiological adaptation of organisms to their physical environment. *Progress in Lipid Research* **29**, 167–227.

Hemkemeyer M., Dohrmann A. B., Christensen B. T. and Tebbe C. C. (2018) Bacterial Preferences for Specific Soil Particle Size Fractions Revealed by Community Analyses. *Front. Microbiol.* **9**.

Ho T. K., (1995) Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* Proceedings of 3rd International Conference on Document Analysis and Recognition. pp. 278–282 vol.1.

Hofmann K., Lamprecht A., Pauli H. and Illmer P. (2016) Distribution of Prokaryotic Abundance and Microbial Nutrient Cycling Across a High-Alpine Altitudinal Gradient in the Austrian Central Alps is Affected by Vegetation, Temperature, and Soil Nutrients. *Microb Ecol* **72**, 704–716.

Holmes J. A. (2000) Non-Marine Ostracods and Quaternary Palaeoenvironments. *Quaternary Research Association Technical Guide*.

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

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.

Horne D. J., Cohen A. and Martens K. (2002) Taxonomy, morphology and biology of Quaternary and living ostracoda. *Washington DC American Geophysical Union Geophysical Monograph Series* **131**, 5–36.

Howes E. L., Joos F., Eakin M. and Gattuso J.-P. (2015) An updated synthesis of the observed and projected impacts of climate change on the chemical, physical and biological processes in the oceans. *Front. Mar. Sci.* **2**.

Hu C., Henderson G. M., Huang J., Xie S., Sun Y. and Johnson K. R. (2008) Quantification of Holocene Asian monsoon rainfall from spatially separated cave records. *Earth and Planetary Science Letters* **266**, 221–232.

Huber R., Wilharm T., Huber D., Trincone A., Burggraf S., König H., Reinhard R., Rockinger I., Fricke H. and Stetter K. O. (1992) *Aquifex pyrophilus* gen. nov. sp. nov., Represents a Novel Group of Marine Hyperthermophilic Hydrogen-Oxidizing Bacteria. *Systematic and Applied Microbiology* **15**, 340–351.

Huber H., Hohn M. J., Stetter K. O. and Rachel R. (2003) The phylum Nanoarchaeota: Present knowledge and future perspectives of a unique form of life. *Research in Microbiology* **154**, 165–171.

Huguet A., Fosse C., Laggoun-Défarge F., Toussaint M.-L. and Derenne S. (2010) Occurrence and distribution of glycerol dialkyl glycerol tetraethers in a French peat bog. *Organic Geochemistry* **41**, 559–572.

Huguet A., Fosse C., Laggoun-Défarge F., Delarue F. and Derenne S. (2013) Effects of a short-term experimental microclimate warming on the abundance and distribution of branched GDGTs in a French peatland. *Geochimica et Cosmochimica Acta* **105**, 294–315.

Huguet A., Francez A.-J., Jusselme M. D., Fosse C. and Derenne S. (2014) A climatic chamber experiment to test the short term effect of increasing temperature on branched GDGT distribution in Sphagnum peat. *Organic Geochemistry* **73**, 109–112.

Huguet A., Meador T. B., Laggoun-Défarge F., Könneke M., Wu W., Derenne S. and Hinrichs K.-U. (2017) Production rates of bacterial tetraether lipids and fatty acids in peatland under varying oxygen concentrations. *Geochimica et Cosmochimica Acta* **203**, 103–116.

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.

Huguet C., Hopmans E. C., Febo-Ayala W., Thompson D. H., Sinninghe Damsté J. S. and Schouten S. (2006) An improved method to determine the absolute abundance of glycerol dibiphytanyl glycerol tetraether lipids. *Organic Geochemistry* **37**, 1036–1041.

REFERENCES BIBLIOGRAPHIQUES

Idso S. B., Schmugge T. J., Jackson R. D. and Reginato R. J. (1975) The utility of surface temperature measurements for the remote sensing of surface soil water status. *Journal of Geophysical Research (1896-1977)* **80**, 3044–3049.

IPCC, Intergovernmental Panel on Climate Change, 5^{ème} rapport, (2014) -en français : GIEC, Groupe d'Experts Intergouvernemental sur l'Evolution du Climat

IPCC: *Climate Change (2014): Synthesis Report*.

IPCC: *Climate Change 2014, the physical science basis*

Jasper J. P. and Hayes J. M. (1990) A carbon isotope record of CO₂ levels during the late Quaternary. *Nature* **347**, 462–464.

Jia C., Zhang C. L., Xie W., Wang J.-X., Li F., Wang S., Dong H., Li W. and Boyd E. S. (2014) Differential temperature and pH controls on the abundance and composition of H-GDGTs in terrestrial hot springs. *Organic Geochemistry* **75**, 109–121.

Jiang W., Guo Z., Sun X., Wu H., Chu G., Yuan B., Hatté C. and Guiot J. (2006) Reconstruction of climate and vegetation changes of Lake Bayanchagan (Inner Mongolia): Holocene variability of the East Asian monsoon. *Quaternary Research* **65**, 411–420.

Kaniewski D., Campo E. V., Guiot J., Burel S. L., Otto T. and Baeteman C. (2013) Environmental Roots of the Late Bronze Age Crisis. *PLOS ONE* **8**, e71004.

Keeling C. D., Bacastow R. B., Bainbridge A. E., Jr C. A. E., Guenther P. R., Waterman L. S. and Chin J. F. S. (1976) Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. *Tellus* **28**, 538–551.

Keinänen M. M., Korhonen L. K., Martikainen P. J., Vartiainen T., Miettinen I. T., Lehtola M. J., Nenonen K., Pajunen H. and Kontro M. H. (2003) Gas chromatographic–mass spectrometric detection of 2- and 3-hydroxy fatty acids as methyl esters from soil, sediment and biofilm. *Journal of Chromatography B* **783**, 443–451.

Killops S. (2005) Introduction to Organic Geochemistry, 2nd edn (paperback) ed. V. Killops. *Geofluids* **5**, 236–237.

Kim J.-H., Schouten S., Hopmans E. C., Donner B. and Sinninghe Damsté J. S. (2008) Global sediment core-top calibration of the TEX₈₆ paleothermometer in the ocean. *Geochimica et Cosmochimica Acta* **72**, 1154–1173.

Kim J.-H., van der Meer J., Schouten S., Helmke P., Willmott V., Sangiorgi F., Koç N., Hopmans E. C. and Damsté J. S. S. (2010) New indices and calibrations derived from

the distribution of crenarchaeal isoprenoid tetraether lipids: Implications for past sea surface temperature reconstructions. *Geochimica et Cosmochimica Acta* **74**, 4639–4654.

Kling S. A. (1998) 9 - Radiolaria. In *Introduction to Marine Micropaleontology (Second Edition)* (eds. B. U. Haq and A. Boersma). Elsevier Science B.V., Amsterdam. pp. 203–244.

Knapp A. B. and Manning S. W. (2016) Crisis in Context: The End of the Late Bronze Age in the Eastern Mediterranean. *American Journal of Archaeology* **120**, 99–149.

Knowles J. R. (1980) Enzyme-Catalyzed Phosphoryl Transfer Reactions. *Annual Review of Biochemistry* **49**, 877–919.

Koga Y., Kyuragi T., Nishihara M. and Sone N. (1998) Did Archaeal and Bacterial Cells Arise Independently from Noncellular Precursors? A Hypothesis Stating That the Advent of Membrane Phospholipid with Enantiomeric Glycerophosphate Backbones Caused the Separation of the Two Lines of Descent. *J Mol Evol* **46**, 54–63.

Koga Y. and Morii H. (2007) Biosynthesis of Ether-Type Polar Lipids in Archaea and Evolutionary Considerations. *Microbiol. Mol. Biol. Rev.* **71**, 97–120.

Koga Y. et al (2008) Did Archaeal and Bacterial Cells Arise Independently from Noncellular Precursors? A Hypothesis Stating That the Advent of Membrane Phospholipid with Enantiomeric Glycerophosphate Backbones Caused the Separation of the Two Lines of Descent.

Kopf S. H., Sessions A. L., Cowley E. S., Reyes C., Sambeek L. V., Hu Y., Orphan V. J., Kato R. and Newman D. K. (2016) Trace incorporation of heavy water reveals slow and heterogeneous pathogen growth rates in cystic fibrosis sputum. *Proceedings of the National Academy of Sciences* **113**, E110–E116.

Kraśnik L., Szponar B., Walczak M., Larsson L. and Gamian A. (2006) Routine clinical laboratory tests correspond to increased serum levels of 3-hydroxy fatty acids, markers of endotoxins, in cardiosurgery patients. *Arch. Immunol. Ther. Exp.* **54**, 55–60.

Kristen I., Wilkes H., Vieth A., Zink K.-G., Plessen B., Thorpe J., Partridge T. C. and Oberhänsli H. (2010) Biomarker and stable carbon isotope analyses of sedimentary organic matter from Lake Tswaing: evidence for deglacial wetness and early Holocene drought from South Africa. *J Paleolimnol* **44**, 143–160.

REFERENCES BIBLIOGRAPHIQUES

Ladbroke B. D. and Chapman D. (1969) Thermal analysis of lipids, proteins and biological membranes a review and summary of some recent studies. *Chemistry and Physics of Lipids* **3**, 304–356.

Langworthy T. A., Smith P. F. and Mayberry W. R. (1972) Lipids of *Thermoplasma acidophilum*. *Journal of Bacteriology* **112**, 1193–1200.

Langworthy T. A. (1977) Long-chain diglycerol tetraethers from *Thermoplasma acidophilum*. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **487**, 37–50.

Langworthy T. A., Holzer G., Zeikus J. G. and Tornabene T. G. (1983) Iso- and Anteiso-Branched Glycerol Diethers of the Thermophilic Anaerobe *Thermodesulfotobacterium commune*. *Systematic and Applied Microbiology* **4**, 1–17.

Lattaud J., Lo L., Huang J., Chou Y., Gorbarenko S. A., Sinninghe Damsté J. S. and Schouten S. (2018) A Comparison of Late Quaternary Organic Proxy-Based Paleotemperature Records of the Central Sea of Okhotsk. *Paleoceanography and Paleoclimatology* **33**, 732–744.

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. *Appl. Environ. Microbiol.* **75**, 5111–5120.

Lecointre G. and Guyader H. L. (2006) *The Tree of Life: A Phylogenetic Classification.*, Harvard University Press.

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.

de Leeuw J. W., Irene W., Rijpstra C. and Schenck P. A. (1981) The occurrence and identification of C30, C31 and C32 alkan-1, 15-diols and alkan-15-one-1-ols in Unit I and Unit II Black Sea sediments. *Geochimica et Cosmochimica Acta* **45**, 2281–2285.

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.

Leng M. J. and Barker P. A. (2006) A review of the oxygen isotope composition of lacustrine diatom silica for palaeoclimate reconstruction. *Earth-Science Reviews* **75**, 5–27.

Lepage C., Fayolle F., Hermann M. and Vandecasteele J. P. (1987) Changes in Membrane Lipid Composition of *Clostridium acetobutylicum* during Acetone-Butanol Fermentation: Effects of Solvents, Growth Temperature and pH. *Microbiology*, **133**, 103–110.

Li C.-H., Li Y.-X., Zheng Y.-F., Yu S.-Y., Tang L.-Y., Li B.-B. and Cui Q.-Y. (2018) A high-resolution pollen record from East China reveals large climate variability near the Northgrippian-Meghalayan boundary (around 4200 years ago) exerted societal influence. *Palaeogeography, Palaeoclimatology, Palaeoecology* **512**, 156–165.

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

Lincoln S. A., Wai B., Eppley J. M., Church M. J., Summons R. E. and DeLong E. F. (2014) Planktonic Euryarchaeota are a significant source of archaeal tetraether lipids in the ocean. *Proceedings of the National Academy of Sciences* **111**, 9858–9863.

Lisiecki L. E. and Raymo M. E. (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* **20**.

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

Liu Z., Zhu J., Rosenthal Y., Zhang X., Otto-Bliesner B. L., Timmermann A., Smith R. S., Lohmann G., Zheng W. and Elison Timm O. (2014) The Holocene temperature conundrum. *Proc Natl Acad Sci U S A* **111**, E3501–E3505.

Ljungqvist F. C. (2010) A new reconstruction of temperature variability in the extra-tropical northern hemisphere during the last two millennia. *Geografiska Annaler: Series A, Physical Geography* **92**, 339–351.

Lombard J., López-García P. and Moreira D. (2012) The early evolution of lipid membranes and the three domains of life. *Nature Reviews Microbiology* **10**, 507–515.

Longo W. M., Huang Y., Yao Y., Zhao J., Giblin A. E., Wang X., Zech R., Haberzettl T., Jardillier L., Toney J., Liu Z., Krivonogov S., Kolpakova M., Chu G.,

REFERENCES BIBLIOGRAPHIQUES

D'Andrea W. J., Harada N., Nagashima K., Sato M., Yonenobu H., Yamada K., Gotanda K. and Shinozuka Y. (2018) Widespread occurrence of distinct alkenones from Group I haptophytes in freshwater lakes: Implications for paleotemperature and paleoenvironmental reconstructions. *Earth and Planetary Science Letters* **492**, 239–250.

Loomis S. E., Russell J. M. and Sinninghe Damsté J. S. (2011) Distributions of branched GDGTs in soils and lake sediments from western Uganda: Implications for a lacustrine paleothermometer. *Organic Geochemistry* **42**, 739–751.

Loomis S. E., Russell J. M., Ladd B., Street-Perrott F. A. and Sinninghe Damsté J. S. (2012) Calibration and application of the branched GDGT temperature proxy on East African lake sediments. *Earth and Planetary Science Letters* **357–358**, 277–288.

Loomis S. E., Russell J. M., Eggermont H., Verschuren D. and Sinninghe Damsté J. S. (2014) Effects of temperature, pH and nutrient concentration on branched GDGT distributions in East African lakes: Implications for paleoenvironmental reconstruction. *Organic Geochemistry* **66**, 25–37.

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

Madigan, M., & Martinko, J. (2018). Brock. Biologie des micro-organismes. 11 e édition. *Edition Person Education France*, 599-601.

Makarova K. S., Aravind L., Galperin M. Y., Grishin N. V., Tatusov R. L., Wolf Y. I. and Koonin E. V. (1999) Comparative Genomics of the Archaea (Euryarchaeota): Evolution of Conserved Protein Families, the Stable Core, and the Variable Shell. *Genome Res.* **9**, 608–628.

Malhi Y., Silman M., Salinas N., Bush M., Meir P. and Saatchi S. (2010) Introduction: Elevation gradients in the tropics: laboratories for ecosystem ecology and global change research. *Global Change Biology* **16**, 3171–3175.

Mann M. E., Zhang Z., Hughes M. K., Bradley R. S., Miller S. K., Rutherford S. and Ni F. (2008) Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences* **105**, 13252–13257.

Margesin R., Jud M., Tscherko D. and Schinner F. (2009) Microbial communities and activities in alpine and subalpine soils: Communities and activities in alpine and subalpine soils. *FEMS Microbiology Ecology* **67**, 208–218.

Marlowe I. T., Brassell S. C., Eglinton G. and Green J. C. (1984) Long chain unsaturated ketones and esters in living algae and marine sediments. *Org. geochem* **6**, 135–141.

Marr A. G. and Ingraham J. L. (1962) Effect of Temperature on the Composition of Fatty Acids in *Escherichia Coli*. *Journal of Bacteriology* **84**, 1260–1267.

Martens Koen, Schön I., Meisch C. and Horne D. J. (2008) Global diversity of ostracods (Ostracoda, Crustacea) in freshwater. In *Freshwater Animal Diversity Assessment* (eds. E. V. Balian, C. Lévêque, H. Segers, and K. Martens). Developments in Hydrobiology. Springer Netherlands, Dordrecht. pp. 185–193.

Martin C., Ménot G., Thouveny N., Davtian N., Andrieu-Ponel V., Reille M. and Bard E. (2019) Impact of human activities and vegetation changes on the tetraether sources in Lake St Front (Massif Central, France). *Organic Geochemistry* **135**, 38–52.

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

Mayewski P. A., Rohling E. E., Curt Stager J., Karlén W., Maasch K. A., David Meeker L., Meyerson E. A., Gasse F., van Kreveland S., Holmgren K., Lee-Thorp J., Rosqvist G., Rack F., Staubwasser M., Schneider R. R. and Steig E. J. (2004) Holocene climate variability. *Quaternary Research* **62**, 243–255.

McMaster G. and Wilhelm W. (1997) Growing degree-days: one equation, two interpretations. *Publications from USDA-ARS / UNL Faculty*.

de Mendoza D. and Cronan J. E. (1983) Thermal regulation of membrane lipid fluidity in bacteria. *Trends in Biochemical Sciences* **8**, 49–52.

Menges J., Huguet C., Alcañiz J. M., Fietz S., Sachse D. and Rosell-Melé A. (2014) Influence of water availability in the distributions of branched glycerol dialkyl glycerol tetraether in soils of the Iberian Peninsula. *Biogeosciences* **11**, 2571–2581.

Metson A. J. (1957) Methods of Chemical Analysis for Soil Survey Samples. *Soil Science* **83**, 245.

REFERENCES BIBLIOGRAPHIQUES

- Meyers P. A. (2003) Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. *Organic Geochemistry* **34**, 261–289.
- 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.
- Miller D. R., Habicht M. H., Keisling B. A., Castañeda I. S. and Bradley R. S. (2018) A 900-year New England temperature reconstruction from in situ seasonally produced branched glycerol dialkyl glycerol tetraethers (brGDGTs). *Clim. Past* **14**, 1653–1667.
- Milliman J. D. (1993) Production and accumulation of calcium carbonate in the ocean: Budget of a nonsteady state. *Global Biogeochemical Cycles* **7**, 927–957.
- Mitsuguchi T., Matsumoto E., Abe O., Uchida T. and Isdale P. J. (1996) Mg/Ca Thermometry in Coral Skeletons. *Science* **274**, 961–963.
- Mueller-Niggemann C., Utami S. R., Marxen A., Mangelsdorf K., Bauersachs T. and Schwark L. (2016) Distribution of tetraether lipids in agricultural soils – differentiation between paddy and upland management. *Biogeosciences* **13**, 1647–1666.
- Mykytczuk N. C. S., Trevors J. T., Leduc L. G. and Ferroni G. D. (2007) Fluorescence polarization in studies of bacterial cytoplasmic membrane fluidity under environmental stress. *Progress in Biophysics and Molecular Biology* **95**, 60–82.
- Naafs B. D. A., Gallego-Sala A. V., Inglis G. N. and Pancost R. D. (2017a) Refining the global branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration. *Organic Geochemistry* **106**, 48–56.
- Naafs B. D. A., Inglis G. N., Zheng Y., Amesbury M. J., Biester H., Bindler R., Blewett J., Burrows M. A., del Castillo Torres D., Chambers F. M., Cohen A. D., Evershed R. P., Feakins S. J., Gałka M., Gallego-Sala A., Gandois L., Gray D. M., Hatcher P. G., Honorio Coronado E. N., Hughes P. D. M., Huguet A., Könönen M., Laggoun-Défarge F., Lähteenoja O., Lamentowicz M., Marchant R., McClymont E., Pontevedra-Pombal X., Ponton C., Pourmand A., Rizzuti A. M., Rochefort L., Schellekens J., De Vleeschouwer F. and Pancost R. D. (2017b) Introducing global peat-specific temperature and pH calibrations based on brGDGT bacterial lipids. *Geochimica et Cosmochimica Acta* **208**, 285–301.

Naeher S., Peterse F., Smittenberg R. H., Niemann H., Ziegler P. K. and Schubert C. J. (2014) Sources of glycerol dialkyl glycerol tetraethers (GDGTs) in catchment soils, water column and sediments of Lake Rotsee (Switzerland) – Implications for the application of GDGT-based proxies for lakes. *Organic Geochemistry* **66**, 164–173.

Navarre W. W. and Schneewind O. (1999) Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. *Microbiol. Mol. Biol. Rev.* **63**, 174–229.

Nedwell D. B. (1999) Effect of low temperature on microbial growth: lowered affinity for substrates limits growth at low temperature. *FEMS Microbiol Ecol* **30**, 101–111.

Ning D., Zhang E., Shulmeister J., Chang J., Sun W. and Ni Z. (2019) Holocene mean annual air temperature (MAAT) reconstruction based on branched glycerol dialkyl glycerol tetraethers from Lake Ximenglongtan, southwestern China. *Organic Geochemistry* **133**, 65–76.

Nishihara M. and Koga Y. (1987) Extraction and Composition of Polar Lipids from the Archaeobacterium, *Methanobacterium thermoautotrophicum* : Effective Extraction of Tetraether Lipids by an Acidified Solvent. *J Biochem* **101**, 997–1005.

Nottingham A. T., Whitaker J., Turner B. L., Salinas N., Zimmermann M., Malhi Y. and Meir P. (2015) Climate Warming and Soil Carbon in Tropical Forests: Insights from an Elevation Gradient in the Peruvian Andes. *BioScience* **65**, 906–921.

Nottingham A. T., Fierer N., Turner B. L., Whitaker J., Ostle N. J., McNamara N. P., Bardgett R. D., Leff J. W., Salinas N., Silman M. R., Kruuk L. E. B. and Meir P. (2018) Microbes follow Humboldt: temperature drives plant and soil microbial diversity patterns from the Amazon to the Andes. *Ecology* **99**, 2455–2466.

Pagani M. (2002) The alkenone-CO₂ proxy and ancient atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **360**, 609–632.

Pagani M., Zachos J. C., Freeman K. H., Tindle C. T. and Bohaty S. (2005) Marked Decline in Atmospheric Carbon Dioxide Concentrations During the Paleogene. *Science* **309**, 600–603.

REFERENCES BIBLIOGRAPHIQUES

Peaple M. D., Tierney J. E., McGee D., Lowenstein T. K., Bhattacharya T. and Feakins S. J. (2021) Identifying plant wax inputs in lake sediments using machine learning. *Organic Geochemistry*, 104222.

Pearson A. and Ingalls A. E. (2013) Assessing the Use of Archaeal Lipids as Marine Environmental Proxies. *Annu. Rev. Earth Planet. Sci.* **41**, 359–384.

Pearson E. J., Juggins S. and Farrimond P. (2008) Distribution and significance of long-chain alkenones as salinity and temperature indicators in Spanish saline lake sediments. *Geochimica et Cosmochimica Acta* **72**, 4035–4046.

Pearson E. J., Juggins S., Talbot H. M., Weckström J., Rosén P., Ryves D. B., Roberts S. J. and Schmidt R. (2011) A lacustrine GDGT-temperature calibration from the Scandinavian Arctic to Antarctic: Renewed potential for the application of GDGT-paleothermometry in lakes. *Geochimica et Cosmochimica Acta* **75**, 6225–6238.

Peterse F., Meer M. T. J. van der, Schouten S., Jia G., Ossebaar J., Blokker J. and Sinninghe Damsté J. S. (2009a) Assessment of soil *n*-alkane δD and branched tetraether membrane lipid distributions as tools for paleoelevation reconstruction. *Biogeosciences* **6**, 2799–2807.

Peterse Francien, Kim J.-H., Schouten S., Kristensen D. K., Koç N. and Sinninghe Damsté J. S. (2009b) Constraints on the application of the MBT/CBT palaeothermometer at high latitude environments (Svalbard, Norway). *Organic Geochemistry* **40**, 692–699.

Peterse F., Nicol G. W., Schouten S. and Sinninghe Damsté J. S. (2010) Influence of soil pH on the abundance and distribution of core and intact polar lipid-derived branched GDGTs in soil. *Organic Geochemistry* **41**, 1171–1175.

Peterse F., Prins M. A., Beets C. J., Troelstra S. R., Zheng H., Gu Z., Schouten S. and Damsté J. S. S. (2011) Decoupled warming and monsoon precipitation in East Asia over the last deglaciation. *Earth and Planetary Science Letters* **301**, 256–264.

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.

Peterse F., Martínez-García A., Zhou B., Beets C. J., Prins M. A., Zheng H. and Eglinton T. I. (2014) Molecular records of continental air temperature and monsoon

precipitation variability in East Asia spanning the past 130,000 years. *Quaternary Science Reviews* **83**, 76–82.

Peterse F., Moy C. M. and Eglinton T. I. (2015) A laboratory experiment on the behaviour of soil-derived core and intact polar GDGTs in aquatic environments. *Biogeosciences* **12**, 933–943.

Peterse F. and Eglinton T. I. (2017) Grain Size Associations of Branched Tetraether Lipids in Soils and Riverbank Sediments: Influence of Hydrodynamic Sorting Processes. *Front. Earth Sci.* **5**.

Pitcher A., Hopmans E. C., Mosier A. C., Park S.-J., Rhee S.-K., Francis C. A., Schouten S. and Damsté J. S. S. (2011) Core and Intact Polar Glycerol Dibiphytanyl Glycerol Tetraether Lipids of Ammonia-Oxidizing Archaea Enriched from Marine and Estuarine Sediments. *Appl. Environ. Microbiol.* **77**, 3468–3477.

Poger D., Caron B. and Mark A. E. (2014) Effect of Methyl-Branched Fatty Acids on the Structure of Lipid Bilayers. *J. Phys. Chem. B* **118**, 13838–13848.

Popp B. N., Bidigare R. R., Deschenes B., Laws E. A., Prahl F. G., Tanimoto J. K. and Wallsgrove R. J. (2006) A new method for estimating growth rates of alkenone-producing haptophytes. *Limnology and Oceanography: Methods* **4**, 114–129.

Power M. J., Marlon J., Ortiz N., Bartlein P. J., Harrison S. P., Mayle F. E., Ballouche A., Bradshaw R. H. W., Carcaillet C., Cordova C., Mooney S., Moreno P. I., Prentice I. C., Thonicke K., Tinner W., Whitlock C., Zhang Y., Zhao Y., Ali A. A., Anderson R. S., Beer R., Behling H., Briles C., Brown K. J., Brunelle A., Bush M., Camill P., Chu G. Q., Clark J., Colombaroli D., Connor S., Daniau A.-L., Daniels M., Dodson J., Doughty E., Edwards M. E., Finsinger W., Foster D., Frechette J., Gaillard M.-J., Gavin D. G., Gobet E., Haberle S., Hallett D. J., Higuera P., Hope G., Horn S., Inoue J., Kaltenrieder P., Kennedy L., Kong Z. C., Larsen C., Long C. J., Lynch J., Lynch E. A., McGlone M., Meeks S., Mensing S., Meyer G., Minckley T., Mohr J., Nelson D. M., New J., Newnham R., Noti R., Oswald W., Pierce J., Richard P. J. H., Rowe C., Sanchez Goñi M. F., Shuman B. N., Takahara H., Toney J., Turney C., Urrego-Sanchez D. H., Umbanhowar C., Vandergoes M., Vanniére B., Vescovi E., Walsh M., Wang X., Williams N., Wilmshurst J. and Zhang J. H. (2008) Changes in fire regimes since the Last Glacial Maximum: an assessment based on a global synthesis and analysis of charcoal data. *Clim Dyn* **30**, 887–907.

REFERENCES BIBLIOGRAPHIQUES

Powers L., Werne J. P., Vanderwoude A. J., Sinninghe Damsté J. S., Hopmans E. C. and Schouten S. (2010) Applicability and calibration of the TEX86 paleothermometer in lakes. *Organic Geochemistry* **41**, 404–413.

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

Prahl F. G., Sparrow M. A. and Wolfe G. V. (2003) Physiological impacts on alkenone paleothermometry. *Paleoceanography* **18**.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for 974 Statistical Computing, Vienna, Austria.

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.

Rampen S. W., Datema M., Rodrigo-Gámiz M., Schouten S., Reichart G.-J. and Sinninghe Damsté J. S. (2014a) Sources and proxy potential of long chain alkyl diols in lacustrine environments. *Geochimica et Cosmochimica Acta* **144**, 59–71.

Rampen S. W., Willmott V., Kim J.-H., Rodrigo-Gámiz M., Uliana E., Mollenhauer G., Schefuß E., Sinninghe Damsté J. S. and Schouten S. (2014b) Evaluation of long chain 1,14-alkyl diols in marine sediments as indicators for upwelling and temperature. *Organic Geochemistry* **76**, 39–47.

Randlett M.-È., Coolen M. J. L., Stockhecke M., Pickarski N., Litt T., Balkema C., Kwiecien O., Tomonaga Y., Wehrli B. and Schubert C. J. (2014) Alkenone distribution in Lake Van sediment over the last 270 ka: influence of temperature and haptophyte species composition. *Quaternary Science Reviews* **104**, 53–62.

Rao Z., Huang C., Xie L., Shi F., Zhao Y., Cao J., Gou X., Chen J. and Chen F. (2019) Long-term summer warming trend during the Holocene in central Asia indicated by alpine peat α -cellulose $\delta^{13}\text{C}$ record. *Quaternary Science Reviews* **203**, 56–67.

Rapp J. M. and Silman M. R. (2012) Diurnal, seasonal, and altitudinal trends in microclimate across a tropical montane cloud forest. *Climate Research* **55**, 17–32.

Ravelo A. C. and Hillaire-Marcel C. (2007) Chapter Eighteen The Use of Oxygen and Carbon Isotopes of Foraminifera in Paleoceanography. In *Developments in Marine*

Geology (eds. C. Hillaire-Marcel and A. De Vernal). Proxies in Late Cenozoic Paleooceanography. Elsevier. pp. 735–764.

Reille M. and de Beaulieu J. L. (1990) Pollen analysis of a long upper Pleistocene continental sequence in a Velay maar (Massif Central, France). *Palaeogeography, Palaeoclimatology, Palaeoecology* **80**, 35–48.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2014). *Campbell biology* (No. s 1309). Boston: Pearson.

Renard M., Lagabriele Y., Martin E. and Saint-Sauveur (de) M. R. (2015) *Eléments de Géologie. 15 ème édition.*, Dunod.

Rodrigo-Gámiz M., Rampen S. W., Schouten S. and Sinninghe Damsté J. S. (2016) The impact of oxic degradation on long chain alkyl diol distributions in Arabian Sea surface sediments. *Organic Geochemistry* **100**, 1–9.

Round F. E., Crawford R. M. and Mann D. G. (2007) *Diatoms: Biology and Morphology of the Genera.*, Cambridge University Press.

Ruddiman W. F. (2001) Earth's Climate: Past and Future. , 412.

Ruddiman W. F. (2003) The Anthropogenic Greenhouse Era Began Thousands of Years Ago. *Climatic Change* **61**, 261–293.

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

Russell N. J. (1984) Mechanisms of thermal adaptation in bacteria: blueprints for survival. *Trends in Biochemical Sciences* **9**, 108–112.

Russell N. J. (1989) Adaptive modifications in membranes of halotolerant and halophilic microorganisms. *J Bioenerg Biomembr* **21**, 93–113.

Russell N. J., Evans R. I., ter Steeg P. F., Hellemons J., Verheul A. and Abee T. (1995) Membranes as a target for stress adaptation. *International Journal of Food Microbiology* **28**, 255–261.

Sachs J. P., Anderson R. F. and Lehman S. J. (2001) Glacial Surface Temperatures of the Southeast Atlantic Ocean. *Science* **293**, 2077–2079.

Sali W. (2014) Physico-chimie des lipopolysaccharides et réponse inflammatoire : rôle des lipoprotéines. thesis, Dijon.

REFERENCES BIBLIOGRAPHIQUES

Sánchez Goñi M. F. (2006) Interactions végétation-climat au cours des derniers 425.000 ans en Europe occidentale. Le message du pollen des archives marines. *Quaternaire. Revue de l'Association française pour l'étude du Quaternaire*, 3–25.

Saraf A., Larsson L., Burge H. and Milton D. (1997) Quantification of ergosterol and 3-hydroxy fatty acids in settled house dust by gas chromatography-mass spectrometry: comparison with fungal culture and determination of endotoxin by a *Limulus* amoebocyte lysate assay. *Appl. Environ. Microbiol.* **63**, 2554–2559.

Scalercio S., Bonacci T., Mazzei A., Pizzolotto R. and Brandmayr P. (2014) Better up, worse down: bidirectional consequences of three decades of climate change on a relict population of *Erebia cassioides*. *J Insect Conserv* **18**, 643–650.

Schimel J. P. and Clein J. S. (1996) Microbial response to freeze-thaw cycles in tundra and taiga soils. *Soil Biology and Biochemistry* **28**, 1061–1066.

Schimel J., Balser T. C. and Wallenstein M. (2007) Microbial Stress-Response Physiology and Its Implications for Ecosystem Function. *Ecology* **88**, 1386–1394.

Schouten S., Hopmans E. C., Schefuß E. and Sinninghe Damsté J. S. (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., Middelburg J. J., Hopmans E. C. and Sinninghe Damsté J. S. (2010) Fossilization and degradation of intact polar lipids in deep subsurface sediments: A theoretical approach. *Geochimica et Cosmochimica Acta* **74**, 3806–3814.

Schouten S., Rijpstra W. I. C., Durisch-Kaiser E., Schubert C. J. and Sinninghe Damsté J. S. (2012) Distribution of glycerol dialkyl glycerol tetraether lipids in the water column of Lake Tanganyika. *Organic Geochemistry* **53**, 34–37.

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.

Schreuder L. T., Beets C. J., Prins M. A., Hatté C. and Peterse F. (2016) Late Pleistocene climate evolution in Southeastern Europe recorded by soil bacterial membrane lipids in Serbian loess. *Palaeogeography, Palaeoclimatology, Palaeoecology* **449**, 141–148.

Schulz H. N. and Jørgensen B. B. (2001) Big Bacteria. *Annual Review of Microbiology* **55**, 105–137.

Shaw E. C., Phinn S. R., Tilbrook B. and Steven A. (2015) Natural in situ relationships suggest coral reef calcium carbonate production will decline with ocean acidification. *Limnology and Oceanography* **60**, 777–788.

Shen C., Shi Y., Fan K., He J.-S., Adams J. M., Ge Y. and Chu H. (2019) Soil pH dominates elevational diversity pattern for bacteria in high elevation alkaline soils on the Tibetan Plateau. *FEMS Microbiol Ecol* **95**.

Siles J. A. and Margesin R. (2016) Abundance and Diversity of Bacterial, Archaeal, and Fungal Communities Along an Altitudinal Gradient in Alpine Forest Soils: What Are the Driving Factors? *Microb Ecol* **72**, 207–220.

Sinensky M. (1974) Homeoviscous Adaptation—A Homeostatic Process that Regulates the Viscosity of Membrane Lipids in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **71**, 522–525.

Singer S. J. and Nicolson G. L. (1972) The Fluid Mosaic Model of the Structure of Cell Membranes. *Science* **175**, 720–731.

Sinninghe Damsté J. S., Hopmans E. C., Pancost R. D., Schouten S. and Geenevasen J. A. J. (2000) Newly discovered non-isoprenoid glycerol dialkylglycerol tetraether lipids in sediments. *Chem. Commun.*, 1683–1684.

Sinninghe Damsté J. S., Ossebaar J., Schouten S. and Verschuren D. (2008) Altitudinal shifts in the branched tetraether lipid distribution in soil from Mt. Kilimanjaro (Tanzania): Implications for the MBT/CBT continental palaeothermometer. *Organic Geochemistry* **39**, 1072–1076.

Sinninghe Damsté J. S., Ossebaar J., Abbas B., Schouten S. and Verschuren D. (2009) Fluxes and distribution of tetraether lipids in an equatorial African lake: Constraints on the application of the TEX86 palaeothermometer and BIT index in lacustrine settings. *Geochimica et Cosmochimica Acta* **73**, 4232–4249.

Sinninghe Damsté J. S., Rijpstra W. I. C., Hopmans E. C., Weijers J. W. H., Foesel B. U., Overmann J. and Dedysh S. N. (2011) 13,16-Dimethyl Octacosanedioic Acid (iso-Diabolic Acid), a Common Membrane-Spanning Lipid of Acidobacteria Subdivisions 1 and 3. *Appl. Environ. Microbiol.* **77**, 4147–4154.

REFERENCES BIBLIOGRAPHIQUES

Sinninghe Damsté J. S., Rijpstra W. I. C., Hopmans E. C., Foesel B. U., Wüst P. K., Overmann J., Tank M., Bryant D. A., Dunfield P. F., Houghton K. and Stott M. B. (2014) Ether- and Ester-Bound iso-Diabolic Acid and Other Lipids in Members of Acidobacteria Subdivision 4. *Appl. Environ. Microbiol.* **80**, 5207–5218.

Sinninghe Damsté J. S. (2016) Spatial heterogeneity of sources of branched tetraethers in shelf systems: The geochemistry of tetraethers in the Berau River delta (Kalimantan, Indonesia). *Geochimica et Cosmochimica Acta* **186**, 13–31.

Sinninghe Damsté J. S., Rijpstra W. I. C., Foesel B. U., Huber K. J., Overmann J., Nakagawa S., Kim J. J., Dunfield P. F., Dedysh S. N. and Villanueva L. (2018) An overview of the occurrence of ether- and ester-linked iso-diabolic acid membrane lipids in microbial cultures of the Acidobacteria: Implications for brGDGT paleoproxies for temperature and pH. *Organic Geochemistry* **124**, 63–76.

Sjögren J., Magnusson J., Broberg A., Schnürer J. and Kenne L. (2003) Antifungal 3-Hydroxy Fatty Acids from *Lactobacillus plantarum* MiLAB 14. *Appl. Environ. Microbiol.* **69**, 7554–7557.

Smith J. L., Halvorson J. J. and Bolton H. (2002) Soil properties and microbial activity across a 500m elevation gradient in a semi-arid environment. *Soil Biology and Biochemistry* **34**, 1749–1757.

Sonesson A., Larsson L., Schütz A., Hagmar L. and Hallberg T. (1990) Comparison of the limulus amebocyte lysate test and gas chromatography-mass spectrometry for measuring lipopolysaccharides (endotoxins) in airborne dust from poultry-processing industries. *Appl. Environ. Microbiol.* **56**, 1271–1278.

Spang A., Saw J. H., Jørgensen S. L., Zaremba-Niedzwiedzka K., Martijn J., Lind A. E., van Eijk R., Schleper C., Guy L. and Ettema T. J. G. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179.

Spang A., Eme L., Saw J. H., Caceres E. F., Zaremba-Niedzwiedzka K., Lombard J., Guy L. and Ettema T. J. G. (2018) Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLOS Genetics* **14**, e1007080.

Sprott G. D., Meloche M. and Richards J. C. (1991) Proportions of diether, macrocyclic diether, and tetraether lipids in *Methanococcus jannaschii* grown at different temperatures. *Journal of Bacteriology* **173**, 3907–3910.

Staubwasser M., Sirocko F., Grootes P. M. and Segl M. (2003) Climate change at the 4.2 ka BP termination of the Indus valley civilization and Holocene south Asian monsoon variability. *Geophysical Research Letters* **30**, 1425.

Steinhilber F., Beer J. and Fröhlich C. (2009) Total solar irradiance during the Holocene. *Geophysical Research Letters* **36**.

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.

Szponar B., Kraśnik L., Hryniewiecki T., Gamian A. and Larsson L. (2003) Distribution of 3-Hydroxy Fatty Acids in Tissues after Intraperitoneal Injection of Endotoxin. *Clin Chem* **49**, 1149–1153.

Theroux S., D'Andrea W. J., Toney J., Amaral-Zettler L. and Huang Y. (2010) Phylogenetic diversity and evolutionary relatedness of alkenone-producing haptophyte algae in lakes: Implications for continental paleotemperature reconstructions. *Earth and Planetary Science Letters* **300**, 311–320.

Tierney J. E. and Russell J. M. (2009) Distributions of branched GDGTs in a tropical lake system: Implications for lacustrine application of the MBT/CBT paleoproxy. *Organic Geochemistry* **40**, 1032–1036.

Tierney J. E., Russell J. M., Eggermont H., Hopmans E. C., Verschuren D. and Sinninghe Damsté J. S. (2010) Environmental controls on branched tetraether lipid distributions in tropical East African lake sediments. *Geochimica et Cosmochimica Acta* **74**, 4902–4918.

Tierney J. E. (2012) GDGT Thermometry: Lipid Tools for Reconstructing Paleotemperatures. *Paleontol. Soc. pap.* **18**, 115–132.

Tierney J. E. and Tingley M. P. (2014) A Bayesian, spatially-varying calibration model for the TEX86 proxy. *Geochimica et Cosmochimica Acta* **127**, 83–106.

Todaro L., Andreu-Hayles L., D'Alessandro C., Gutiérrez E., Cherubini P. and Saracino A. (2007) Response of *Pinus leucodermis* to climate and anthropogenic activity in the National Park of Pollino (Basilicata, Southern Italy). *Biological Conservation* **137**, 507–519.

U.S. Global Change Research Program (2017). "Climate Science Special Report. Chapter 12: Sea Level Rise"

REFERENCES BIBLIOGRAPHIQUES

- Valentine D. L. (2007) Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nature Reviews Microbiology* **5**, 316–323.
- Van Dyk M. S., Kock J. L. F. and Botha A. (1994) Hydroxy long-chain fatty acids in fungi. *World J Microbiol Biotechnol* **10**, 495–504.
- Vellai T. and Vida G. (1999) The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**, 1571–1577.
- 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 Discussions*, 1–40.
- Véquaud P., Derenne S., Anquetil C., Collin S., Poulenard J., Sabatier P. and Huguet A. (2021) Influence of environmental parameters on the distribution of bacterial lipids in soils from the French Alps: Implications for paleo-reconstructions. *Organic Geochemistry* **153**, 104194.
- Versteegh G. J. M., Bosch H.-J. and De Leeuw J. W. (1997) Potential palaeoenvironmental information of C₂₄ to C₃₆ mid-chain diols, keto-ols and mid-chain hydroxy fatty acids; a critical review. *Organic Geochemistry* **27**, 1–13.
- Versteegh G. J. M., Jansen J. H. F., Schneider R. R. and De Leeuw J. W. (2000) Mid-chain diols and keto-ols in se atlantic sediments: a new tool for tracing past sea surface water masses? *Geochimica et Cosmochimica Acta* **64**, 1879–1892.
- Villanueva L., Besseling M., Rodrigo-Gámiz M., Rampen S. W., Verschuren D. and Sinninghe Damsté J. S. (2014) Potential biological sources of long chain alkyl diols in a lacustrine system. *Organic Geochemistry* **68**, 27–30.
- Villanueva L., von Meijenfildt F. A. B., Westbye A. B., Yadav S., Hopmans E. C., Dutilh B. E. and Damsté J. S. S. (2020) Bridging the membrane lipid divide: bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. *The ISME Journal*, 1–15.
- Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., . . . & Willemet, J.M. (2012). 1031 The detailed snowpack scheme Crocus and its implementation in SURFEX v7. 2.

Volkman J. K., Barrett S. M., Dunstan G. A. and Jeffrey S. W. (1992) C30–C32 alkyl diols and unsaturated alcohols in microalgae of the class Eustigmatophyceae. *Organic Geochemistry* **18**, 131–138.

Volkman John K., Barrett S. M. and Blackburn S. I. (1999a) Eustigmatophyte microalgae are potential sources of C29 sterols, C22–C28 n-alcohols and C28–C32 n-alkyl diols in freshwater environments. *Organic Geochemistry* **30**, 307–318.

Volkman John K., Barrett S. M. and Blackburn S. I. (1999b) Fatty Acids and Hydroxy Fatty Acids in Three Species of Freshwater Eustigmatophytes. *Journal of Phycology* **35**, 1005–1012.

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., Bendle J. A., Yang H., Yang Y., Hardman A., Yamoah A., Thorpe A., Mandel I., Greene S. E., Huang J. and Xie S. (2021) Global calibration of novel 3-hydroxy fatty acid based temperature and pH proxies. *Geochimica et Cosmochimica Acta*.

Wang H., Liu W., Zhang C. L., Wang Z., Wang J., Liu Z. and Dong H. (2012) Distribution of glycerol dialkyl glycerol tetraethers in surface sediments of Lake Qinghai and surrounding soil. *Organic Geochemistry* **47**, 78–87.

Wang H., Liu W. and Lu H. (2016) Appraisal of branched glycerol dialkyl glycerol tetraether-based indices for North China. *Organic Geochemistry* **98**, 118–130.

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**, 106172.

REFERENCES BIBLIOGRAPHIQUES

Wang J.-T., Cao P., Hu H.-W., Li J., Han L.-L., Zhang L.-M., Zheng Y.-M. and He J.-Z. (2015) Altitudinal Distribution Patterns of Soil Bacterial and Archaeal Communities Along Mt. Shegyla on the Tibetan Plateau. *Microb Ecol* **69**, 135–145.

Wang M., Zheng Z., Man M., Hu J. and Gao Q. (2017) Branched GDGT-based paleotemperature reconstruction of the last 30,000years in humid monsoon region of Southeast China. *Chemical Geology* **463**, 94–102.

Watanabe Y. and Takakuwa M. (1984) Effect of Sodium Chloride on Lipid Composition of *Saccharomyces rouxii*. *Agricultural and Biological Chemistry* **48**, 2415–2422.

WCRP Global Sea Level Budget Group (2018) Global sea-level budget 1993–present. *Earth System Science Data* **10**, 1551–1590.

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.

Wei K. and Jia G. (2009) Soil n-alkane $\delta^{13}\text{C}$ along a mountain slope as an integrator of altitude effect on plant species $\delta^{13}\text{C}$. *Geophysical Research Letters* **36**.

Weijers J. W. H., Schouten S., van der Linden M., van Geel B. and Sinninghe Damsté J. S. (2004) Water table related variations in the abundance of intact archaeal membrane lipids in a Swedish peat bog. *FEMS Microbiol Lett* **239**, 51–56.

Weijers J. W. H., Schouten S., Hopmans E. C., Geenevasen J. A. J., David O. R. P., Coleman J. M., Pancost R. D. and Damsté J. S. S. (2006a) Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environmental Microbiology* **8**, 648–657.

Weijers J. W. H., Schouten S., Spaargaren O. C. and Sinninghe Damsté J. S. (2006b) Occurrence and distribution of tetraether membrane lipids in soils: Implications for the use of the TEX86 proxy and the BIT index. *Organic Geochemistry* **37**, 1680–1693.

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.

Weijers J. W. H., Panoto E., Bleijswijk J. van, Schouten S., Rijpstra W. I. C., Balk M., Stams A. J. M. and Damsté J. S. S. (2009a) Constraints on the Biological Source(s) of the Orphan Branched Tetraether Membrane Lipids. *Geomicrobiology Journal* **26**, 402–414.

Weijers J. W. H., Schouten S., Schefuß E., Schneider R. R. and Sinninghe Damsté J. S. (2009b) Disentangling marine, soil and plant organic carbon contributions to continental margin sediments: A multi-proxy approach in a 20,000 year sediment record from the Congo deep-sea fan. *Geochimica et Cosmochimica Acta* **73**, 119–132.

Weijers J. W. H., Wiesenberg G. L. B., Bol R., Hopmans E. C. and Pancost R. D. (2010) Carbon isotopic composition of branched tetraether membrane lipids in soils suggest a rapid turnover and a heterotrophic life style of their source organism(s). *Biogeosciences* **7**, 2959–2973.

Weijers J. W. H., Bernhardt B., Peterse F., Werne J. P., Dungait J. A. J., Schouten S. and Sinninghe Damsté J. S. (2011a) Absence of seasonal patterns in MBT–CBT indices in mid-latitude soils. *Geochimica et Cosmochimica Acta* **75**, 3179–3190.

Weijers J. W. H., Steinmann P., Hopmans E. C., Schouten S. and Sinninghe Damsté J. S. (2011b) Bacterial tetraether membrane lipids in peat and coal: Testing the MBT–CBT temperature proxy for climate reconstruction. *Organic Geochemistry* **42**, 477–486.

Weiss B. (1982) The decline of Late Bronze Age civilization as a possible response to climatic change. *Climatic Change* **4**, 173–198.

de Wet G. (2017) Arctic and North Atlantic paleo-environmental reconstructions from lake sediments. *Doctoral Dissertations*.

Whitaker J., Ostle N., Nottingham A. T., Ccahuana A., Salinas N., Bardgett R. D., Meir P. and McNamara N. P. (2014) Microbial community composition explains soil respiration responses to changing carbon inputs along an Andes-to-Amazon elevation gradient. *Journal of Ecology* **102**, 1058–1071.

Whittaker R. H. and Margulis L. (1978) Protist classification and the kingdoms of organisms. *Biosystems* **10**, 3–18.

Wilkinson, S.G., 1988. Gram-negative bacteria. In: Ratledge, C., Wilkinson, S.G. (Eds.), *Microbial Lipids*, 1063, vol. 1. Academic Press, New York, pp. 199–488

REFERENCES BIBLIOGRAPHIQUES

Woese C. R. and Fox G. E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences* **74**, 5088–5090.

Woese C. R., Kandler O. and Wheelis M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**, 4576–4579.

Wollenweber H.-W., Broady K. W., Luderitz O. and Rietschel E. T. (1982) The Chemical Structure of Lipid A. *European Journal of Biochemistry* **124**, 191–198.

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

Wu D., Cao J., Jia G., Guo H., Shi F., Zhang X. and Rao Z. (2020) Peat brGDGTs-based Holocene temperature history of the Altai Mountains in arid Central Asia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **538**, 109464.

Xiao W., Xu Y., Ding S., Wang Y., Zhang X., Yang H., Wang G. and Hou J. (2015) Global calibration of a novel, branched GDGT-based soil pH proxy. *Organic Geochemistry* **89–90**, 56–60.

Yang H., Ding W., He G. and Xie S. (2010) Archaeal and bacterial tetraether membrane lipids in soils of varied altitudes in Mt. Jianfengling in South China. *J. Earth Sci.* **21**, 277–280.

Yang H., Ding W., Zhang C. L., Wu X., Ma X., He G., Huang J. and Xie S. (2011) Occurrence of tetraether lipids in stalagmites: Implications for sources and GDGT-based proxies. *Organic Geochemistry* **42**, 108–115.

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 H., Lü X., Ding W., Lei Y., Dang X. and Xie S. (2015) The 6-methyl branched tetraethers significantly affect the performance of the methylation index (MBT') in soils from an altitudinal transect at Mount Shennongjia. *Organic Geochemistry* **82**, 42–53.

Yang Y. and Hansson L. (2016) Data compilation on the biological response to ocean acidification: an update. , 9.

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.

Zaremba-Niedzwiedzka K., Caceres E. F., Saw J. H., Bäckström D., Juzokaite L., Vancaester E., Seitz K. W., Anantharaman K., Starnawski P., Kjeldsen K. U., Stott M. B., Nunoura T., Banfield J. F., Schramm A., Baker B. J., Spang A. and Ettema T. J. G. (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358.

Zell C., Kim J.-H., Abril G., Sobrinho R., Dorhout D., Moreira-Turcq P. and Sinninghe Damsté J. (2013) Impact of seasonal hydrological variation on the distributions of tetraether lipids along the Amazon River in the central Amazon basin: implications for the MBT/CBT paleothermometer and the BIT index. *Front. Microbiol.* **4**.

Zell C., Kim J.-H., Balsinha M., Dorhout D., Fernandes C., Baas M. and Sinninghe Damsté J. S. (2014) Transport of branched tetraether lipids from the Tagus River basin to the coastal ocean of the Portuguese margin: consequences for the interpretation of the MBT'/CBT paleothermometer. *Biogeosciences* **11**, 5637–5655.

Zelles L., Bai Q. Y., Rackwitz R., Chadwick D. and Beese F. (1995) Determination of phospholipid- and lipopolysaccharide-derived fatty acids as an estimate of microbial biomass and community structures in soils. *Biol Fertil Soils* **19**, 115–123.

Zelles L. (1999) Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation of microbial communities in soil: a review. *Biol Fertil Soils* **29**, 111–129.

Zeng X., Chen X. and Zhuang J. (2015) The positive relationship between ocean acidification and pollution. *Marine Pollution Bulletin* **91**, 14–21.

Zhang Y.-M. and Rock C. O. (2008) Membrane lipid homeostasis in bacteria. *Nature Reviews Microbiology* **6**, 222–233.

Zhao M., Mercer J. L., Eglinton G., Higginson M. J. and Huang C.-Y. (2006) Comparative molecular biomarker assessment of phytoplankton paleoproductivity for the last 160kyr off Cap Blanc, NW Africa. *Organic Geochemistry* **37**, 72–97.

Zheng F., Chen Y., Xie W., Chen S., Liu H., Phelps T. J. and Zhang C. (2019) Diverse biological sources of core and intact polar isoprenoid GDGTs in terrace soils from southwest of China: Implications for their use as environmental proxies. *Chemical Geology* **522**, 108–120.

REFERENCES BIBLIOGRAPHIQUES

Zheng Y., Li Q., Wang Z., Naafs B. D. A., Yu X. and Pancost R. D. (2015) Peatland GDGT records of Holocene climatic and biogeochemical responses to the Asian Monsoon. *Organic Geochemistry* **87**, 86–95.

Zilber-Rosenberg I. and Rosenberg E. (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev* **32**, 723–735.

Zink K.-G., Leythaeuser D., Melkonian M. and Schwark L. (2001) Temperature dependency of long-chain alkenone distributions in recent to fossil limnic sediments and in lake waters. Associate editor: J. B. Fein. *Geochimica et Cosmochimica Acta* **65**, 253–265.

Zuckerandl E. and Pauling L. (1965) Evolutionary Divergence and Convergence in Proteins. In *Evolving Genes and Proteins* (eds. V. Bryson and H. J. Vogel). Academic Press. pp. 97–166.

LISTE DES FIGURES

Chapitre 1

Figure 1.1 Les différents compartiments du système climatique terrestre (d'après Ruddiman, 2001).....	26
Figure 1.2 Bilan radiatif global du système Terre (d'après Ruddiman, 2001).	27
Figure 1.3 Températures annuelles au cours des 2000 dernières années pour (a) l'hémisphère nord et (b) l'hémisphère sud, et (c) les températures annuelles moyennes globales. Les différentes reconstitutions utilisées sont présentées comme indiqué dans les légendes, et regroupées par couleur en fonction de leur représentation spatiale (rouge : terre seulement – toutes latitudes ; orange : latitudes extratropicales uniquement terrestres ; bleu clair : latitudes extratropicales terrestres et maritimes ; bleu foncé: latitudes terrestres et maritimes toutes les latitudes) et les températures instrumentales sont indiquées en noir (d'après le 5 ^{ème} rapport de l'IPCC, <i>Climate Change 2014, the physical science basis</i>).	28
Figure 1.4 Evaluation des contributions au changement global (hausse des températures) pour la période 1951-2010 et erreurs associées. Adapté du rapport IPCC, 2014 : <i>Climate Change 2014 : Synthesis Report</i>	29
Figure 1.5 Concentrations mensuelles en CO ₂ atmosphérique mesurées à l'observatoire de Mauna Loa (Hawaï) depuis 1958 jusqu'en juin 2020 (NOAA ; https://gml.noaa.gov/).....	30
Figure 1.6 Observations du changement global et climatique. a) Anomalies de la température de surface des terres et des océans combinées, moyennes annuelles et mondiales par rapport à la moyenne sur la période 1986-2005. b) Variation annuelle et mondiale moyenne du niveau de la mer par rapport à la moyenne sur la période 1986-2005. c) Émissions mondiales de CO ₂ anthropique provenant de la sylviculture et d'autres utilisations des terres ainsi que de la combustion de combustibles fossiles, de la production de ciment et des torchères. d) Concentrations atmosphériques des gaz à effet de serre dioxyde de carbone (CO ₂ , vert), le méthane (CH ₄ , orange) et l'oxyde nitreux (N ₂ O, rouge) déterminées à partir de données de carottes de glace (points) et de mesures atmosphériques directes (lignes). Adapté du rapport IPCC, 2014: <i>Climate Change 2014: Synthesis Report</i>	32

LISTE DES FIGURES

Figure 1.7 Différents modèles de l'évolution possible des a) la température moyenne à la surface du globe ; b) l'étendue de la banquise de l'hémisphère Nord en septembre ; c) l'élévation du niveau moyen des mers à l'échelle du globe. Adapté d'après IPCC 2014, the physical science basis, 5 ^{ème} rapport.	33
Figure 1.8 De l'intérêt des marqueurs : reconstituer de façon indirecte des conditions environnementales et/ ou des paléoenvironnements (source de l'image : http://www.bvsm.ca)	36
Figure 1.9 Représentation de l'arbre du vivant, basé sur l'étude des séquences de nucléotides des sous unités d'ARN ribosomique. Cet arbre illustre les trois grands domaines du vivant et leurs composés lipidiques caractéristiques. Adapté de Bianchi et Canuel, 2011 ; Briggs et Summons, 2014 et références incluses.....	40
Figure 1.10 Schéma d'une cellule eucaryote (source : Chantal Proulx).	42
Figure 1.11 Schéma d'une cellule procaryote (source : Chantal Proulx).....	43
Figure 1.12 Quelques exemples de la diversité morphologiques des organismes procaryotes.....	45
Figure 1.13 Schéma de la structure de la membrane des bactéries a) Gram-positif et b) Gram-négatif (d'après Madigan et Martinko 2018).....	48
Figure 1.14 Structure des lipides membranaires des bactéries et eucaryotes. a) Acide phosphatidique. b) Configuration en bicouche de la membrane des bactéries et des eucaryotes.	50
Figure 1.15 Structure des lipides membranaires archéens a) acide archaetidique b) Configuration en monocouche de la membrane des archées.	52
Figure 1.16 Illustration de la transformation de lipides intacts en lipides de cœur par hydrolyse enzymatique. L'exemple donné pour les lipides bactériens ne concerne pas toutes les bactéries, certaines peuvent produire des tétraéthers de glycérol à chaînes alkyles ramifiées (partie 1.3.2.1).....	53
Figure 1.17 Schéma illustrant l'influence de la structure, de la composition et de l'agencement des lipides sur la fluidité de la membrane cytoplasmique (adapté de Ernst et al., 2016)	55
Figure 1.18 Structures des alcénones en C _{37:2} et C _{37:3}	58
Figure 1.19 Photographies au microscope électronique de a) <i>Emiliana huxleyi</i> et b) <i>Gephyrocapsa oceanica</i> . (source : https://www.mikrotax.org/)	59

Figure 1.20 Chromatogrammes des alcénones détectées au sein des cultures d' <i>Emiliana huxleyi</i> à 10°C et 25°C (Prahl et Wakeham 1987).....	59
Figure 1.21 Calibration globale entre l'indice U^{k}_{37} et la température de surfaces marine (SST) (592 échantillons de sédiments de surface répartis au sein de différents océans) proposée par Conte et al. (2006).	60
Figure 1.22 Structures des diols à longues chaînes : a) C_{28} 1,13-diol ; b) C_{30} 1,13-diol ; c) C_{30} 1,15-diol.....	62
Figure 1.23 Structure des tétraéthers isopréniques (isoGDGTs), avec la nomenclature associée ainsi que les rapports masse/charge (m/z) correspondants.	65
Figure 1.24 Régression linéaire entre l'indice TEX_{86} et la température de surface marine (n= 44), correspondant aux équations 1.5 et 1.6.	67
Figure 1.25 . Structure et nomenclature des tétraéthers ramifiés bactériens, avec les isomères de position 5- et 6-méthyles, les nomenclatures associés (en chiffres romain) et les rapports masse/charge (m/z) des différents composés (d'après De Jonge et al., 2014)	72
Figure 1.26 Calibrations proposées par Weijers et al. (2007) entre les indices MBT/CBT et la TMAA et le pH respectivement.....	73
Figure 1.27 Régression linéaire multiple entre les indices MBT, CBT et la TMAA établie pour 134 sols (Eq. 1.16 ; Weijers et al., 2007).	74
Figure 1.28 Structure schématique du LPS (lipopolysaccharide) ancré dans la membrane par l'intermédiaire du lipide A (d'après Brandenburg et al., 1999). Sur le lipide A sont montrées les chaînes d'acides gras (de différentes longueurs).	81
Figure 1.29 Quelques exemples de structures du lipide A, constitué d'unités glucosamines, ancrées dans la membrane par l'intermédiaire d'acides gras (notamment les acides gras 3-hydroxylés). Adapté de Bryant et al. (2010).	82
Figure 1.30 Les différents isomères des acides gras 3-hydroxylés.....	83
Figure 1.31 Régression linéaire entre l'indice RIAN et les valeurs de pH mesurées au sein des échantillons de sols prélevés le long des transects du Mont Shennongjia (Chine; C.Wang et al., 2016), et des Mont Majella (Italie) et Rungwe (Tanzanie ; Huguet et al., 2019). La droite noire est la régression globale pour l'ensemble des échantillons de sols proposée par Huguet et al. (2019).	84

Figure 1.32 Régressions linéaires entre a) l'indice RAN_{15} et b) l'indice RAN_{17} et les valeurs de températures moyennes annuelles de l'air au sein des échantillons de sols prélevés le long des transects du Mont Shennongjia (Chine ; C.Wang et al., 2016), et des Mont Majella (Italie) et Rungwe (Tanzanie ; Huguet et al., 2019). La droite noire sur le graphique b correspond à la régression globale pour l'ensemble des échantillons de sols proposée par Huguet et al. (2019). 85

Chapitre 2

Figure 2.1 Emplacement des différents transects altitudinaux étudiés dans cette thèse. 91

Figure 2.2 Quelques photographies illustrant la diversité et la beauté des paysages rencontrés au sein des Alpes françaises. 93

Figure 2.3 Cartes des (a) Alpes françaises, (b) du massif des Bauges et (c) du Lautaret-Galibier. Sources Google Satellites. 94

Figure 2.4 Transect altitudinal composite des massifs des Bauges et du Lautaret-Galibier. 96

Figure 2.5 Régression linéaire du modèle de TMAA (°C) permettant la reconstruction des températures sur le transect altitudinal composite des Alpes françaises. Les lignes pointillées représentent l'intervalle de prédiction à 95% et la zone colorée représente l'intervalle de confiance à 95%. 98

Figure 2.6 Schéma d'une plaque de respiration et répartitions des différents substrats carbonés. 105

Figure 2.7 Schéma illustrant le protocole d'incubation des 4 échantillons de sols en microcosme 107

Figure 2.8 Structure interne du tétraéther synthétique en C_{46} utilisé pour la semi-quantification des tétraéthers. 110

Figure 2.9 Structure chimique du standard C_{14} deutéré, utilisé pour la quantification des acides gras 3-hydroxylés. 113

Figure 2.10 Illustration de l'effet du séchage sous flux d'azote des échantillons sur un standard en C_{14} deutéré, méthylé et silylé, avant analyse en GC-MS 114

Figure 2.11 Test de répétabilité et de reproductibilité sur un triplicat du sol des Alpes françaises BOU11 (232 m d'altitude) analysé en GC-MS. Des valeurs aberrantes

ont été observées sur un des triplicats d'injection, dû à une erreur analytique. Ces erreurs sont représentées sous la forme de points sur le graphique.	115
Figure 2.12 Rapport des fragments m/z 106 (standard C ₁₄ deutéré non silylé) et 178 (standard C ₁₄ deutéré silylé) dans l'échantillon BOU11 analysé en triplicat (3 vials), chaque vial étant injecté 3 fois.	116
Figure 2.13 Illustration de la présence de coélutions d'acides gras non hydroxylés sur les acides gras 3-hydroxylés.	117
Figure 2.14 Principe de l'analyse de redondance ou RDA.	120
Figure 2.15 Principe de la méthode k-NN. L'échantillon inconnu x appartient vraisemblablement au groupe rouge selon cette méthode.	121
Figure 2.16 Principe de la méthode Random Forest pour les régressions. D'après Breiman (2001)	122

Chapitre 3

Figure 3.1 Average distribution of (a) brGDGTs and (b) 3-OH FAs in the 49 soil samples of the French Alps.	133
Figure 3.2 (a) Linear regression between CBT and pH along the two massifs in the French Alps. The black dotted line corresponds to the global linear regression between CBT and pH from De Jonge et al. (2014). (b) Linear regression between RIAN and pH along the two massifs in the French Alps. The black dotted line corresponds to the linear regression between RIAN and pH from Huguet et al. (2019). Dotted and colored lines represent the 95% prediction interval for each regression and colored areas represent the 95% confidence interval for each regression.	135
Figure 3.3 (a) Linear regression between MAAT and MBT' _{5Me} along the two massifs in the French Alps. The black dotted line corresponds to global the linear regression between MBT' _{5Me} and MAAT from De Jonge et al. (2014). Linear regressions between (b) RAN ₁₅ and (c) RAN ₁₇ along the two massifs in the French Alps. Dotted and colored lines represent the 95% prediction interval for each linear regression and colored areas represent the 95% confidence interval for each regression.	138
Figure 3.4 Correlation matrix between the different environmental variables and soil properties. Correlation coefficients in red (negative correlation) or blue (positive	

LISTE DES FIGURES

- correlation) are statistically significant. Red variables have been chosen as explanatory variables for statistical tests.139
- Figure 3.5 Triplot presenting the results of the RDA carried out on the relative abundances of (a) brGDGTs and (b) 3-OH FAs and the non-redundant physicochemical parameters: pH, MAAT, soil moisture, sand and clay contents.147
- Figure 3.6 MAAT estimated using the new calibration based on a linear combination of brGDGT relative abundances (Eq. 3.1) vs. MAAT observed along the the Bauges and Lautaret-Galibier massifs. The roman numerals correspond to the different GDGT structures presented in De Jonge et al. (2014).149
- Figure 3.7 (a) pH estimated using the new calibration based on a linear combination of 3-OH FA relative abundances (Eq. 7) vs. pH observed along the Bauges and Lautaret-Galibier massifs. (b) MAAT estimated using the new calibration based on a linear combination of 3-OH relative abundances (Eq. 8) vs. MAAT observed along the Bauges and Lautaret-Galibier massifs. The roman numerals correspond to the different GDGT structures presented in De Jonge et al. (2014).150
- Figure 3.8 Données de viabilité des bactéries Gram-négatif dans les 4 échantillons incubés aux différentes températures et pour les différents pas de temps (0, 1, 6, 9 et 12 mois). Les concentrations d'unités formant des colonies sont exprimées en utilisant une notation logarithmique, où la valeur indiquée est le logarithme en base 10 de la concentration.157
- Figure 3.9 Concentrations des 3-AGH ($\mu\text{g/g}$ sol sec) au sein de l'échantillon FP88 aux différents pas de temps et températures d'incubation considérées. Après 3 et 12 mois d'incubation les moyennes de deux réplicats avec les écarts types associés sont montrées.158
- Figure 3.10 Abondances relatives des 3-AGH pour les échantillons initiaux ($t = 0$ mois) KS48, FP88, MOU32 et BOU11.159
- Figure 3.11 Degré de ramification des 3-AGH mesurés au sein des 49 échantillons de sols des Alpes françaises en fonction de l'altitude. L'écart type est compris dans la taille du point sur cette figure.160
- Figure 3.12 Degré de ramification des acides gras 3-hydroxylés (Eq 1) aux différentes températures d'incubation et aux différents pas de temps pour les échantillons

a) KS48, b) FP88, c) MOU32 et d) BOU11. Les résultats présentés correspondent à la moyenne sur les deux répliquats.	162
Figure 3.13 Abondances relatives des acides gras 3-hydroxylés aux différentes températures d'incubation et aux différents pas de temps pour les échantillons a) KS48, b) FP88, c) MOU32 et d) BOU11. Les résultats présentés correspondent à la moyenne des abondances relatives sur les deux répliquats.	164
Figure 3.14 Concentrations des tetraéthers bactériens (ng/g sol sec) dans l'échantillon FP88 aux différents pas de temps et températures d'incubation considérés. A 3 mois sont montrées les moyennes des deux répliquats avec les écarts types associés.	165
Figure 3.15 Abondances relatives des brGDGTs au pas de temps initial pour les échantillons KS48, FP88, MOU32 et BOU11.	166
Figure 3.16 Abondances relatives des tetraéthers bactériens aux différentes températures d'incubations et aux différents pas de temps pour les échantillons a) KS48, b) FP88, c) MOU32 et d) BOU11. Les résultats présentés correspondent à la moyenne des abondances relatives sur les deux répliquats.	169
Figure 3.17 Indice MBT' _{5Me} pour les 4 échantillons des sols, aux différentes températures d'incubation et pour les différents pas de temps considérés. Les traits noirs représentent les moyennes de l'indice MBT' _{5Me} des 4 échantillons de sols pour les différentes modalités d'incubations.	170

Chapitre 4

Figure 4.1 Average distribution of 3-OH FAs along the eight altitudinal transects investigated in this study. Data from Mt. Majella and Mt. Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Wang et al. (2016). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021).	181
Figure 4.2 Average distribution of 5- and 6-methyl brGDGTs, along Mt. Shegyla, Mt. Pollino Majella, Lautaret-Bauges, the Peruvian Andes and the Chilean Andes. Data from Mt. Majella were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Yang et al. (2015). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021).	182

LISTE DES FIGURES

Figure 4.3 PCA biplot of (a) 3-OH FA fractional abundances in soil samples from the eight altitudinal transects and (b) brGDGT fractional abundances in soil samples from seven of the eight altitudinal transects. The brGDGT data from Mt. Rungwe, for which 5- and 6-methyl isomers were not separated, were not included in the PCA.186

Figure 4.4 Linear regressions between (a) pH and RIAN and (b) pH and CBT' along the eight altitudinal transects investigated. Dotted lines represent the 95% confidence interval for each regression and coloured areas represent the 95% confidence interval for each regression. Data for Mt. Majella and Mt. Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Yang et al. (2015) and C. Wang et al. (2016). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021). Only significant regressions ($p < 0.05$) are shown.189

Figure 4.5 Linear regressions between (a) MAAT and RAN15 and (b) MAAT and RAN17 along the eight altitudinal transects investigated. Dotted lines represent the 95% confidence interval for each regression and coloured areas represent the 95% confidence interval for each regression. Data from Mt. Majella and Mt. Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from C. Wang et al. (2016). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021). Only significant regressions ($p < 0.05$) are shown.191

Figure 4.6 Linear regressions between (a) MAAT and MBT'_{5Me} along seven of the eight altitudinal transects investigated. Data from Mt. Rungwe (Coffinet et al., 2014), for which 5- and 6-methyl brGDGTs were not separated, were not included in this graph. Dotted lines represent the 95% confidence interval for each regression and colored areas represent the 95% confidence interval for each regression. Data from Mt. Majella were taken from Huguet et al. (2019). Data from Mt. Banges and Mt. Lautaret-Galibier were taken from Véquaud et al. (2021). Data from Mt. Shennongjia were taken from Yang et al. (2015). The global soil calibration by De Jonge et al. (2014) was applied to all these transects. Only significant regressions ($p < 0.05$) are shown.193

Figure 4.7 Results of the three different models tested to reconstruct the MAAT from 3-OH FA distribution: observed MAAT (°C) vs Predicted MAAT (°C) for (a) the multiple linear regression model, (b) the random forest model and (c) the k-NN method. MAAT residuals plotted against the predicted MAAT for (d) the multiple linear regression model, (e) the random forest model and (f) the k-NN method.197

Figure 4.8 Results of the three different models tested to reconstruct the pH from 3-OH FA distribution: observed pH vs predicted pH for (a) the multiple linear regression model, (b) the random forest model and (c) the k-NN method. The pH residuals plotted against the predicted pH for (d) the multiple linear regression model, (e) the random forest model and (f) the k-NN method. 199

Figure 4.9 Importance (arbitrary unit) of the 3-OH FAs used to estimate (a) MAAT and (b) pH in the random forest models proposed in this study according to the permutation importance method (Breiman, 2001). 200

Figure 4.10 Comparison of the 3-OH FA model-MAAT record with other time series and proxy records for the HS4 speleothem (Wang et al., 2018). (a) RAN₁₅-MAAT record reconstructed using a local Chinese calibration (C. Wang et al., 2016; Wang et al., 2018). (b) 3-OH FA random forest model-MAAT. (d) The CaCO₃ oxygen isotope record (Hu et al., 2008b). (e) Total solar irradiance (TSI; W/m²) during the Holocene (past 9300 years) based on a composite described in Steinhilber et al. (2009). 206

Chapitre 5

Figure 5.1 Principal Component Analysis performed on the global dataset (n=767) with the k-means clustering. (a) PC2 vs PC1, (b) PC3 vs PC1, (c) Variance explained (%) for each component, (d) Optimal number of clusters according to the k-means method. 215

Figure 5.2 Boxplot showing the distribution of the 4 environmental variables considered: (a) pH, (b) MAAT(°C), (c) MAP (mm/year), (d) Average frost days per year (FRS). Interquartile range (IQR) = $Q_3 - Q_1$ where Q_3 is the 75th percentile and Q_1 is the 25th percentile. Outliers are defined with a 1.5 coefficient on the IQR. Letters on the panel show the differences between each cluster according to Kruskal-Wallis and Dunn post-hoc tests. 217

Figure 5.3 Spatial distribution of samples in the global dataset. The colors correspond to the different clusters. 218

Figure 5.4 Linear regressions between MAAT (°C) and MBT'_{5Me} for (a) cluster A, (b) cluster B, (c) cluster C, (d) cluster D and (e) cluster E. Grey dots correspond to soils and green dots to peats samples. 219

LISTE DES FIGURES

Figure 5.5 Redundancy analysis between brGDGT distribution and environmental variables for (a) cluster A, (b) cluster B, (c) cluster C, (d) cluster D and (e) cluster E. 221

Figure 5.6 Comparison of the linear regressions between MAAT (°C) and MBT'_{5Me} in the two subgroups derived from the extended dataset: “cold” cluster (CI < 0.64) and “warm” cluster (CI > 0.64). The CI index (Eq. 5.2) was defined by De Jonge et al. (2019).
.....225

Figure 5.7 Average values of the Community Index (CI; Eq. 3) for each cluster of the global soil dataset defined after PCA analysis (Fig. 2). The threshold of 0.64 for the CI separating “warm” and “cold” groups (De Jonge et al., 2019) is represented.227

Figure 5.8 MAAT predicted by the random forest models using the relative abundances of (a) 5-methyl brGDGTs (FROG_{5Me}) and (b) all brGDGTs. Residuals of the (c) FROG_{5Me} model and (d) FROG model with all brGDGTs. Importance of the individual brGDGTs in (e) the FROG_{5Me} model and (f) the FROG model with all brGDGTs, according to the permutation importance method (Breiman, 2001). Grey samples correspond to soils and green samples to peats.229

Figure 5.9 Comparison of the global MBT'_{5Me} calibration proposed by (a) De Jonge et al. (2014) and (b) Naafs et al. (2017a) using the same dataset than the FROG model.
.....232

Figure 5.10 Comparison of the MAAT estimates derived from the MBT'_{5Me} peat calibration by Naafs et al. (2017; red) and the FROG random forest model (blue) with the peat α -cellulose $\delta^{13}C$ record (green) from core ATM10-C7 of the SSP (Rao et al., 2019; Wu et al., 2020).237

Figure 5.11 Reconstructed MAAT for the Pliocene marine sediment sequence from the Hank core located in the Netherlands (Dearing Crampton-Flood et al., 2018, 2020) (a) derived from the calibrations by De Jonge et al. (2014), Naafs et al. (2017a), Dearing Crampton-Flood et al. (2020; BayMBT) and from the FROG model and (b) derived from the calibrations by Dearing Crampton-Flood et al., (2020; BayMBT₀) and from the FROG₀ model. Grey zones are the 95% intervals for the FROG and FROG₀ models .240

LISTE DES TABLEAUX

Chapitre 2

Tableau 2.1 Données disponibles sur les différents transects altitudinaux étudiés. Les marques rouges correspondent à des données obtenues au cours de cette thèse.....	92
Tableau 2.2 Liste des échantillons de sol prélevés le long des massifs des Bauges et du Lautaret-Galibier, associés au type de végétation et aux différents couverts pédologiques.	97
Tableau 2.3 Liste des échantillons de sol le long du transect du Mont Pollino et les caractéristiques environnementales et physico-chimiques associées.....	99
Tableau 2.4 Liste des échantillons de sol le long du transect du Mont Shegyla et les caractéristiques environnementales et physico-chimiques associées (Wang et al., 2015)	100
Tableau 2.5 Liste des échantillons de sols prélevés au niveau du transect du Kosñipata et caractéristiques environnementales et physico-chimiques associées. (Nottingham et al., 2017). (n.d. : non disponible).....	101
Tableau 2.6 Liste des échantillons de sols prélevés au niveau du transect Araucanos et les caractéristiques environnementales et physico-chimiques associée. (n.d. : non disponible)	102
Tableau 2.7 Emplacement, nombre et références des échantillons de sols et de tourbières dans le jeu de données global de brGDGT. Les paramètres disponibles pour les différents sites d'échantillonnage sont indiqués : TMAA (Température moyenne annuelle de l'air (°C), pH, PMA (Précipitation moyenne annuelle (mm/an), FRS (Nombre de jours de gel par an).....	103
Tableau 2.8 Comparaison des protocole d'extraction, séparation et d'analyse des 3-AGH. En bleu sont indiquées les modifications apportées au protocole original (C. Wang et al., 2016).	118

Chapitre 3

Table 3.1 Physicochemical characteristics of the 49 samples collected along the Bauges and Lautaret-Galibier massifs. The numbering of the samples corresponds to that	
--	--

LISTE DES TABLEAUX

used in Table 2.2. Samples with stars represent sites instrumented with temperature sensors.....	132
Table 3.2 RDA correlation coefficients of the environmental variables selected along axes 1 and 2.	142
Table 3.3 Quantification of the influence of the different environmental variables on brGDGTs and 3-OH FAs. Statistical significancies are shown as follows: *** (0<p<0.001); ** (0.001<p<0.01); * (0.01<p<0.05); ° (0.05 <p<0.1).	143
Tableau 3.4 Proportion des isomères d'acides gras 3-hydroxylés (en %) et degré de ramification pour les échantillons BOU11, FP88, MOU32 et KS48 aux différents pas de temps et températures d'incubation considérés. Les écart-types entre réplicats sont indiqués en italique.....	163
Tableau 3.5 Abondances relatives des tétraéthers bactériens individuels (en %) pour les échantillons BOU11, FP88, MOU32 et KS48 aux différents pas de temps et températures d'incubation considérés. Les écart-types entre réplicats sont indiqués en italiques.....	168

Chapitre 4

Table 4.1 List of the soil samples collected along Mts. Shegyla, Pollino, Peruvian Andes and Chilean Andes, with corresponding altitude (m), MAAT (°C), pH and 3-OH FA/brGDGT-derived indices.	184
Table 4.2 Characteristics of the different models proposed in this study to estimate MAAT and pH: R ² , RMSE, variance of the residuals, mean absolute error (MAE) and the upper and lower limits of estimation. The "training" samples were used to develop the different machine learning models, which were then tested on a "test" sample set.	195

Chapitre 5

Table 5.1 Quantitative description of the 5 clusters obtained after the k-means on PCA on the brGDGT distribution of the total dataset.	216
Table 5.2 RDA correlation coefficients of the selected environmental variables along axes 1 and 2 for each cluster, and quantification of the influence of the different environmental variables on brGDGT relative abundances. Statistical significance (p<0.05) is shown with an asterisk.....	222

Table 5.3 Characteristics of the different models proposed in this study to estimate MAAT: R^2 , RMSE, variance of the residuals and the upper and lower limits of estimation. The "training" samples were used to develop the different machine learning models, which were then tested on the remaining sample set.228

ANNEXES

Annexe 1 – Données supplémentaires du Chapitre 3 –

Supplementary Table 3.1. Relative abundances (%) of the *normal*, *iso* and *anteiso* 3-OH FAs in soils from the French Alps and corresponding indices.

Samples	Relative abundances (%)																				Indices		
	<i>nC10</i>	<i>iC11</i>	<i>aC11</i>	<i>nC11</i>	<i>iC12</i>	<i>nC12</i>	<i>iC13</i>	<i>aC13</i>	<i>nC13</i>	<i>iC14</i>	<i>nC14</i>	<i>iC15</i>	<i>aC15</i>	<i>nC15</i>	<i>iC16</i>	<i>nC16</i>	<i>iC17</i>	<i>aC17</i>	<i>nC17</i>	<i>nC18</i>	RAN ₁₅	RAN ₁₇	RIAN
1	4.05	0.36	0.05	0.23	0.11	12.22	3.08	0.69	0.44	1.69	21.57	5.75	4.33	0.50	2.28	14.43	10.28	4.00	0.69	13.23	8.61	5.77	0.31
2	3.17	0.26	0.13	0.23	0.18	10.15	3.02	1.00	0.90	2.26	19.84	4.36	5.59	0.91	1.91	17.41	8.37	2.70	1.09	16.52	6.12	2.48	0.37
3	5.09	0.24	0.10	0.25	0.13	14.51	2.69	0.75	0.70	1.84	21.79	4.68	4.89	0.59	1.39	14.49	6.58	2.38	0.65	16.25	8.27	3.66	0.46
4	3.72	0.30	0.12	0.25	0.16	11.18	2.93	0.72	1.09	2.09	20.60	4.13	4.70	0.88	1.95	17.24	8.85	2.76	1.14	15.18	5.34	2.42	0.39
5	5.29	1.71	0.27	0.28	0.10	13.99	2.46	0.33	0.60	1.32	23.25	5.19	3.39	0.46	0.90	13.15	6.10	1.96	1.14	18.13	7.45	1.72	0.51
6	5.24	0.32	0.05	0.26	0.10	12.80	2.73	0.53	0.57	1.68	22.67	5.58	4.29	0.55	1.55	16.12	7.10	2.29	0.84	14.74	7.78	2.72	0.45
7	4.69	0.48	0.06	0.29	0.06	14.58	2.50	0.35	0.57	1.34	24.09	5.22	2.92	0.65	1.01	13.53	7.26	2.37	1.26	16.77	4.49	1.87	0.51
8	2.77	0.67	0.15	0.15	0.27	8.55	4.73	0.83	0.41	1.94	19.58	6.74	4.58	0.45	2.37	15.22	11.34	3.24	1.09	14.94	10.25	2.98	0.23
9	5.10	0.28	0.05	0.39	0.06	13.25	2.46	0.30	1.25	1.43	22.64	5.23	3.24	0.83	0.77	15.06	5.93	1.77	0.99	18.99	3.92	1.79	0.56
10	5.41	0.53	0.03	0.33	0.12	0.87	2.63	0.51	0.65	1.46	27.39	6.07	3.86	0.64	1.14	15.67	7.95	2.57	1.21	20.96	6.06	2.12	0.43
11	5.64	0.48	0.06	0.31	0.06	12.52	3.27	0.63	0.70	1.55	21.01	5.59	4.08	0.77	1.87	15.98	6.70	3.53	1.23	14.02	5.32	2.86	0.41
12	5.65	0.82	0.06	0.28	0.12	9.06	2.94	0.50	0.76	1.53	16.97	6.79	3.53	1.58	2.34	16.32	10.79	4.89	3.43	11.63	2.23	1.43	0.28
13	5.75	0.33	0.04	0.50	0.16	8.84	2.72	0.59	1.18	1.59	15.54	5.03	2.66	1.85	2.59	17.95	11.33	4.24	3.53	13.55	1.44	1.20	0.34
14	2.82	0.31	0.04	0.23	0.07	13.47	1.68	0.39	0.45	1.62	23.43	6.42	4.19	0.71	0.99	12.42	9.40	3.17	1.10	17.10	5.86	2.89	0.40
15	3.14	0.40	0.05	0.23	0.11	12.98	2.19	0.54	0.48	1.73	23.95	5.60	4.41	0.46	1.30	13.35	9.52	2.92	0.75	15.89	9.61	3.87	0.39

ANNEXES

16	3.70	0.25	0.05	0.21	0.15	12.07	2.75	0.53	0.48	1.91	22.65	4.92	4.82	0.48	1.43	16.33	8.19	2.51	0.74	15.82	10.03	3.40	0.42
17	4.34	0.75	0.07	0.24	0.13	11.98	2.52	0.43	0.43	1.50	25.42	6.46	3.71	0.49	1.33	12.96	9.11	3.10	0.76	14.26	7.64	4.06	0.39
18	3.99	0.55	0.04	0.24	0.11	11.23	2.16	0.57	0.39	1.33	24.06	5.65	4.77	0.50	1.54	12.96	9.09	4.04	0.77	16.02	9.45	5.28	0.37
19	2.40	0.35	0.03	0.14	0.07	10.28	1.66	0.45	0.26	1.47	23.89	5.28	4.66	0.40	1.67	13.00	8.35	4.20	0.65	20.79	11.62	6.42	0.41
20	4.28	0.52	0.17	0.21	0.31	10.96	3.62	0.80	0.42	1.86	19.59	7.44	4.88	0.71	2.24	13.65	12.34	3.86	0.97	11.17	6.86	3.97	0.21
21	4.21	0.43	0.11	0.12	0.28	9.38	3.67	1.01	0.35	2.12	16.96	6.48	5.26	0.60	2.90	14.44	12.65	4.17	0.88	14.00	8.76	4.76	0.19
22	5.15	0.67	0.07	0.22	0.27	11.34	3.74	0.81	0.46	2.06	18.25	6.80	4.01	0.64	2.40	13.92	13.16	3.38	0.83	11.80	6.30	4.05	0.22
23	3.77	0.76	0.11	0.18	0.31	10.73	3.69	0.90	0.43	1.87	16.72	7.01	4.11	0.61	2.35	14.01	15.98	3.75	0.95	11.79	6.77	3.94	0.16
24	3.89	1.29	0.21	0.49	1.73	9.92	2.96	1.36	0.75	1.99	12.87	6.29	3.00	0.96	2.78	14.33	16.65	4.08	2.99	11.47	3.12	1.37	0.13
25	7.19	0.88	0.13	0.32	0.43	11.36	3.32	0.84	0.54	1.63	17.34	6.37	3.93	0.82	1.96	13.97	11.64	3.27	1.84	12.22	4.77	1.78	0.28
26	5.63	0.45	0.06	0.16	0.08	10.92	2.18	0.40	0.29	1.12	23.08	6.35	3.72	0.30	1.41	12.32	7.18	2.64	0.53	21.18	12.26	4.97	0.46
27	5.24	1.19	0.10	0.27	0.20	8.85	2.87	0.64	0.58	1.76	15.63	8.25	3.95	1.04	1.65	14.66	11.42	3.46	2.37	15.88	3.80	1.46	0.26
28	3.63	0.32	0.03	0.16	0.04	9.35	1.44	0.34	0.47	1.18	21.45	6.20	4.08	0.50	0.84	12.50	5.84	2.72	1.06	27.84	8.22	2.57	0.52
29	4.37	0.87	0.07	0.17	0.18	11.18	2.89	0.80	0.44	1.67	19.59	7.15	3.24	0.58	0.81	13.10	15.14	4.02	0.79	12.92	5.56	5.11	0.23
30	4.44	0.96	0.07	0.14	0.18	11.90	2.81	0.52	0.45	1.80	19.94	7.01	3.30	0.58	1.21	14.45	14.03	3.56	0.84	11.82	5.70	4.26	0.26
31	3.43	1.38	0.06	0.15	0.14	12.01	2.91	0.46	0.50	2.67	20.63	6.67	3.50	0.49	0.65	16.03	13.61	3.20	0.63	10.91	7.10	5.11	0.26
32	5.14	0.42	0.04	0.20	0.06	10.38	2.07	0.40	0.41	1.33	23.06	6.37	3.51	0.52	0.97	13.65	6.90	2.30	0.63	21.65	6.76	3.65	0.49
33	5.95	0.84	0.07	0.27	0.10	10.96	2.07	0.38	0.43	1.24	23.68	6.19	3.45	0.43	1.13	12.35	8.19	2.70	0.74	18.83	7.99	3.65	0.45
34	4.55	0.65	0.06	0.24	0.09	12.30	2.83	0.41	0.44	1.53	22.92	6.27	3.93	0.52	1.37	13.12	9.66	2.61	0.61	15.89	7.56	4.29	0.38
35	2.18	0.70	0.04	0.13	0.06	9.11	1.36	0.19	0.51	1.98	23.73	7.74	3.13	0.41	0.80	14.38	9.22	2.04	0.56	21.75	7.65	3.65	0.43
36	1.65	1.01	0.03	0.16	0.07	11.77	1.84	0.26	0.46	1.63	23.19	6.89	2.44	0.47	0.63	14.91	9.56	2.57	0.95	19.52	5.21	2.71	0.43
37	1.85	0.56	0.04	0.11	0.15	11.52	2.57	0.44	0.42	1.97	22.48	7.00	3.54	0.52	1.21	14.62	11.32	2.31	0.57	16.79	6.79	4.02	0.35
38	2.55	0.82	0.05	0.11	0.18	10.44	3.35	0.78	0.42	1.64	21.41	7.01	3.67	0.54	1.41	15.97	12.88	2.88	1.02	12.87	6.86	2.83	0.28

39	1.91	0.46	0.11	0.15	0.21	7.50	3.70	0.78	0.45	2.13	17.96	7.57	5.57	0.67	1.59	15.34	11.82	3.95	1.13	17.03	8.36	3.49	0.21
40	5.45	0.42	0.07	0.15	0.12	7.66	3.49	0.47	0.36	1.43	19.74	7.32	4.50	0.51	2.05	13.76	9.24	3.67	0.85	18.75	8.88	4.32	0.31
41	2.13	0.62	0.12	0.10	0.19	7.30	4.01	0.75	0.38	1.66	18.12	6.07	4.27	0.45	1.94	17.13	12.14	3.52	0.98	18.13	9.57	3.57	0.26
42	3.19	0.41	0.06	0.13	0.18	7.24	2.88	0.51	0.46	1.42	17.32	6.20	4.17	0.54	2.00	17.19	10.46	4.14	1.21	20.31	7.65	3.41	0.32
43	1.12	0.46	0.10	0.12	0.17	6.44	3.52	0.75	0.49	1.95	17.30	7.18	4.96	0.83	2.25	16.75	10.63	3.89	1.80	19.28	5.96	2.16	0.25
44	1.22	0.19	0.07	0.20	0.29	3.87	4.10	0.87	0.64	2.05	19.29	7.57	4.05	0.74	1.05	11.91	18.73	1.92	2.15	19.09	5.50	0.89	0.16
45	0.62	0.15	0.03	0.16	0.19	3.57	3.77	0.45	0.51	1.75	20.82	9.02	3.96	0.46	1.25	15.57	11.29	2.46	1.56	22.42	8.62	1.57	0.28
46	1.73	0.45	0.17	0.20	0.15	7.55	3.94	0.95	0.54	2.24	18.34	7.38	5.79	0.73	1.78	15.39	10.46	3.20	1.26	17.76	7.89	2.54	0.24
47	2.34	0.57	0.13	0.21	0.23	6.57	4.17	0.87	0.48	2.20	18.01	7.30	4.58	0.84	1.72	17.52	10.03	3.16	1.54	17.54	5.45	2.04	0.27
48	3.55	0.65	0.13	0.17	0.19	5.58	3.86	0.58	0.51	1.85	18.63	7.54	3.75	0.70	1.67	16.01	10.23	3.21	1.39	19.79	5.35	2.32	0.29
49	1.23	0.53	0.09	0.22	0.24	3.79	2.97	0.74	0.64	2.30	18.36	9.54	4.99	1.30	2.36	18.51	12.22	3.49	2.01	14.44	3.85	1.73	0.19

Supplementary Table 3.2. Relative abundances (%) of the brGDGTs in soils from the French Alps and corresponding indices.

Samples	Relative abundances (%)															Indices	
	<i>Ia</i>	<i>Ib</i>	<i>Ic</i>	<i>Ila</i>	<i>Ila'</i>	<i>Ilb</i>	<i>Ilb'</i>	<i>Ilc</i>	<i>Ilc'</i>	<i>Illa</i>	<i>Illa'</i>	<i>Illb</i>	<i>Illb'</i>	<i>Illc</i>	<i>Illc'</i>	MBT' _{5ME}	CBT
1	23.97	0.81	0.00	35.04	13.43	1.00	0.69	0.39	0.38	20.97	3.30	0.00	0.00	0.00	0.00	0.30	1.46
2	13.68	6.37	1.78	19.09	22.20	8.25	5.97	1.46	0.13	9.54	10.05	0.55	0.85	0.09	0.00	0.36	0.43
3	30.23	3.45	0.69	30.88	16.43	3.98	1.60	1.08	0.00	8.18	3.21	0.15	0.12	0.00	0.00	0.44	0.93
4	17.47	7.21	1.54	17.90	19.08	7.28	7.78	1.24	0.00	10.00	9.57	0.51	0.43	0.00	0.00	0.42	0.39
5	29.72	2.19	0.00	39.27	6.74	0.00	0.00	0.00	0.00	20.88	0.61	0.33	0.25	0.00	0.00	0.35	1.54
6	28.33	1.48	0.19	34.50	5.97	1.00	0.13	0.38	0.00	13.88	13.89	0.15	0.11	0.00	0.00	0.38	1.42
7	26.49	0.47	0.00	43.39	4.59	0.00	0.00	0.00	0.00	23.51	1.20	0.26	0.08	0.00	0.00	0.29	2.20
8	24.19	0.00	0.00	32.32	16.32	1.26	0.29	0.00	0.00	22.84	2.28	0.00	0.00	0.50	0.00	0.30	1.67
9	32.78	1.43	0.26	36.29	6.47	1.03	0.09	0.28	0.28	19.89	0.93	0.16	0.12	0.00	0.00	0.37	1.47
10	25.88	1.41	1.21	40.99	7.34	1.10	0.15	0.31	0.05	20.31	1.06	0.11	0.08	0.00	0.00	0.31	1.45
11	29.30	1.12	0.78	24.86	19.42	1.62	0.60	0.94	0.95	14.76	5.62	0.00	0.00	0.00	0.00	0.43	1.34
12	18.69	1.15	0.54	26.84	9.60	0.90	0.57	0.00	0.00	35.72	5.20	0.49	0.29	0.00	0.00	0.24	1.32
13	19.58	4.39	1.29	20.06	3.61	4.09	8.77	0.00	0.00	14.15	20.85	1.50	1.71	0.00	0.00	0.40	0.40
14	26.05	2.69	3.64	41.32	7.04	2.15	0.29	0.47	0.12	14.56	1.66	0.00	0.00	0.00	0.00	0.36	1.16
15	23.27	2.17	1.18	39.49	13.72	1.97	0.82	0.00	0.00	14.92	2.16	0.18	0.11	0.00	0.00	0.32	1.19
16	21.52	3.96	1.03	30.13	17.87	5.59	1.31	1.01	0.17	12.93	3.96	0.33	0.18	0.00	0.00	0.35	0.81
17	32.65	2.26	1.87	37.10	11.16	1.21	0.33	0.32	0.00	11.11	1.99	0.00	0.00	0.00	0.00	0.43	1.33
18	28.74	4.51	4.12	31.37	16.46	2.62	1.11	0.30	0.00	8.88	1.89	0.00	0.00	0.00	0.00	0.46	0.97
19	29.50	3.32	2.86	36.29	13.73	1.96	0.31	0.34	0.34	9.82	1.49	0.00	0.00	0.00	0.03	0.42	1.15
20	14.59	7.16	0.77	23.43	18.04	3.85	9.34	0.00	0.00	12.17	9.81	0.84	0.00	0.00	0.00	0.36	0.44
21	9.53	6.40	0.76	28.65	7.42	21.03	0.00	0.00	0.00	3.42	18.94	3.24	0.60	0.00	0.00	0.24	0.22
22	13.78	4.83	0.47	15.84	28.08	2.12	7.64	0.00	0.00	8.61	17.24	0.54	0.84	0.00	0.00	0.42	0.60

23	13.46	8.34	0.00	14.31	27.56	3.68	10.10	0.00	0.35	8.35	12.35	0.49	1.01	0.00	0.00	0.45	0.40
24	18.40	5.91	0.00	15.66	28.89	3.38	4.64	0.00	0.00	7.97	13.36	0.44	0.71	0.00	0.65	0.47	0.66
25	14.95	8.45	1.42	22.43	19.63	5.44	7.79	0.70	0.30	10.02	7.76	0.42	0.71	0.00	0.00	0.39	0.42
26	37.77	2.63	0.61	40.47	7.48	0.73	0.36	0.00	0.00	9.50	0.44	0.00	0.00	0.00	0.00	0.45	1.36
27	20.62	7.70	2.60	31.41	16.85	4.75	3.71	0.27	0.15	9.33	2.29	0.16	0.18	0.00	0.00	0.40	0.63
28	34.63	2.07	1.02	38.39	8.46	0.78	0.48	0.19	0.19	12.64	1.14	0.00	0.00	0.00	0.00	0.42	1.39
29	24.69	5.25	0.93	22.36	20.77	2.09	4.10	0.06	0.00	9.15	9.30	0.84	0.46	0.00	0.00	0.48	0.77
30	36.41	3.98	0.65	25.32	13.24	1.46	3.76	0.00	0.00	8.42	5.41	0.86	0.50	0.00	0.00	0.54	0.91
31	38.67	1.39	0.00	37.66	10.76	0.58	0.91	0.00	0.00	6.69	2.96	0.24	0.14	0.01	0.00	0.47	1.48
32	38.92	1.72	0.65	36.52	9.24	0.74	0.44	0.17	0.00	11.01	0.60	0.00	0.00	0.00	0.00	0.46	1.47
33	32.52	1.92	1.07	40.47	8.75	0.87	0.40	0.00	0.13	12.23	1.41	0.07	0.15	0.00	0.01	0.40	1.41
34	36.67	1.86	0.78	36.26	11.07	1.08	0.85	0.24	0.00	8.98	1.79	0.27	0.16	0.00	0.00	0.46	1.35
35	65.69	2.25	0.34	27.60	0.00	0.00	0.00	0.00	0.00	3.97	0.15	0.00	0.00	0.00	0.00	0.68	1.62
36	56.17	0.55	0.00	36.19	0.00	0.00	0.00	0.08	0.00	7.02	0.00	0.00	0.00	0.00	0.00	0.57	2.23
37	49.64	2.19	0.00	26.14	14.45	0.48	0.66	0.00	0.00	4.95	1.49	0.00	0.00	0.00	0.00	0.62	1.43
38	37.18	8.70	1.52	24.35	9.38	1.96	4.25	0.00	0.00	4.89	6.84	0.20	0.75	0.00	0.00	0.60	0.68
39	6.71	7.72	0.97	5.84	19.27	9.20	20.20	0.66	0.00	6.27	19.07	0.78	3.30	0.00	0.00	0.41	-0.07
40	15.77	10.35	5.24	20.22	15.14	7.80	8.63	1.10	0.23	6.62	7.68	0.39	0.85	0.00	0.00	0.47	0.28
41	8.62	12.99	2.07	5.38	16.25	8.66	20.14	1.31	4.40	4.18	9.64	3.00	2.68	0.69	0.00	0.55	-0.14
42	19.26	14.84	9.58	21.21	12.99	4.89	8.29	0.00	0.00	4.54	3.09	0.48	0.74	0.04	0.03	0.59	0.28
43	15.14	17.07	6.63	13.08	14.15	6.33	13.20	0.51	0.57	4.47	5.06	0.52	1.50	0.90	0.85	0.61	0.06
44	7.15	13.21	2.35	15.33	9.69	13.11	19.11	4.86	0.00	3.53	6.84	2.87	1.95	0.00	0.00	0.38	-0.15
45	6.67	10.03	1.31	17.85	9.22	8.98	21.31	1.21	3.82	3.84	10.70	3.07	2.00	0.00	0.00	0.36	-0.08
46	9.31	11.04	1.81	7.59	15.14	9.02	18.84	1.25	1.22	6.41	13.43	3.32	0.86	0.24	0.55	0.48	-0.08
47	7.29	17.42	2.92	11.57	8.16	11.17	19.22	1.06	6.28	3.60	4.63	2.28	2.57	0.04	1.81	0.50	-0.25
48	8.57	17.24	3.24	15.71	8.10	7.22	17.61	2.34	4.83	3.65	6.82	2.41	1.89	0.38	0.00	0.50	-0.11
49	9.33	16.18	2.03	17.52	8.48	7.20	19.66	1.03	4.22	4.34	5.15	0.71	2.21	0.32	1.61	0.48	-0.09

Supplementary Figure 3.1. Location of the sampling sites along the a) Bauges and b) Lautaret-Galibier massifs.

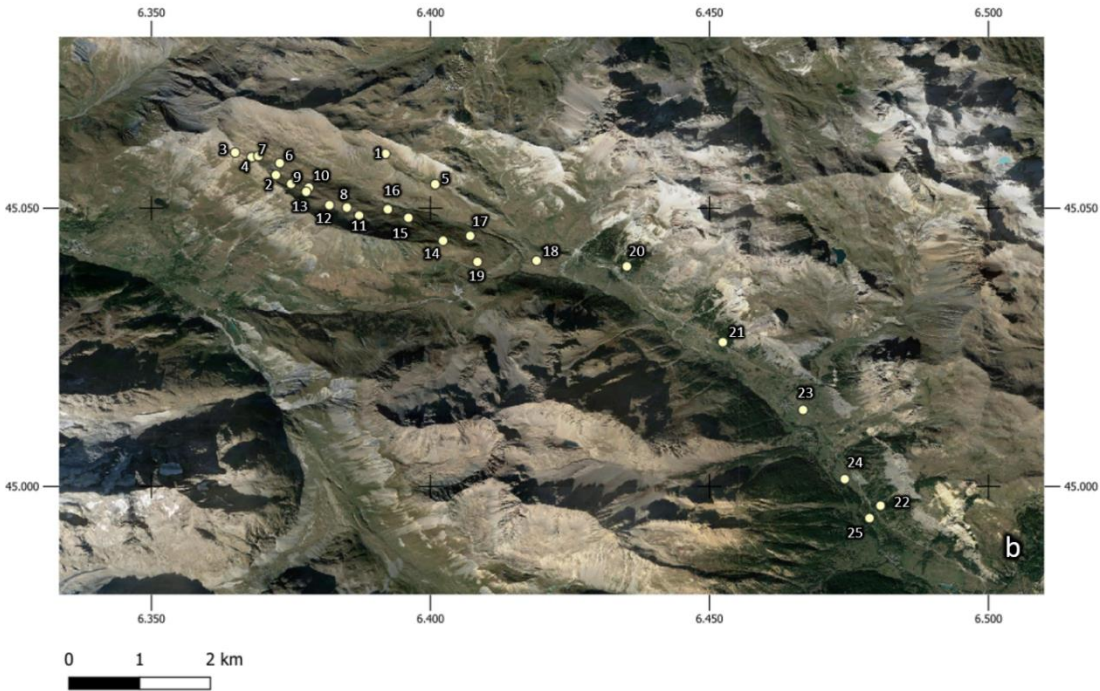
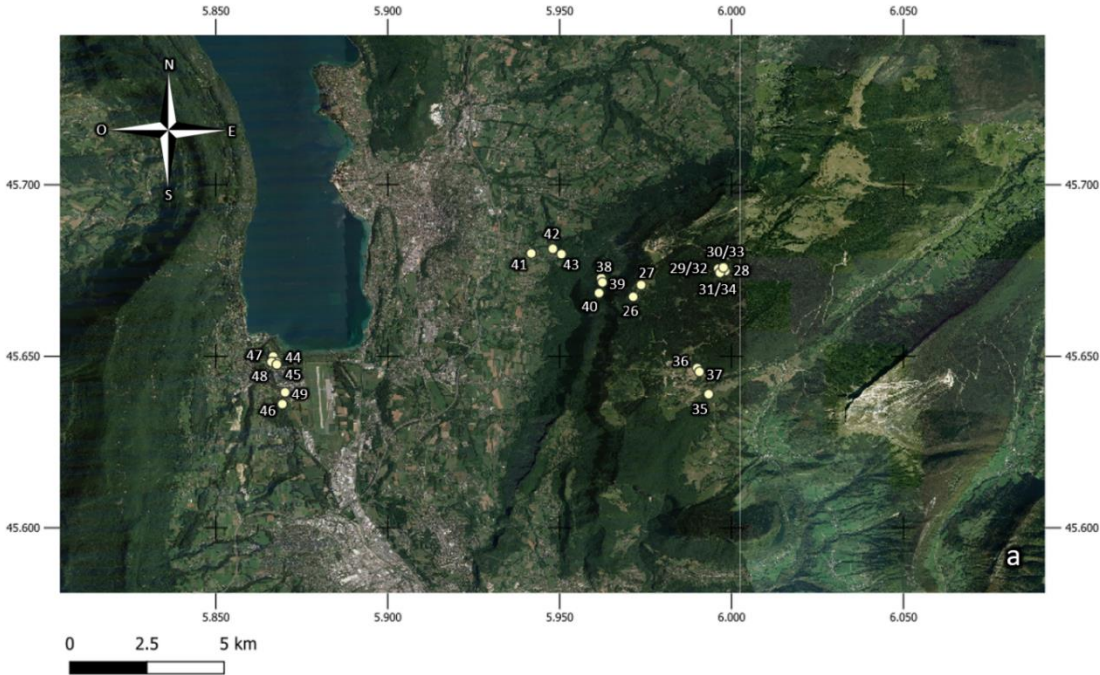


Tableau supplémentaire 3.3. Concentrations et abondances relatives (%) des 3-AGH pour les 4 échantillons de sols placés en microcosme, aux différents pas de temps et aux différentes températures

				Abondances relatives (%)																				
Temps	Température	Réplikat	Concentration (µg/g sol)	nC10	iC11	aC11	nC11	iC12	nC12	iC13	aC13	nC13	iC14	nC14	iC15	aC15	nC15	iC16	nC16	iC17	aC17	nC17	nC18	
KS48	0 mois		2.87	3.72	0.30	0.12	0.25	0.16	11.18	2.93	0.72	1.09	2.09	20.60	4.13	4.70	0.88	1.95	17.24	8.85	2.76	1.14	15.18	
		5°C	A	2.05	5.14	0.49	0.13	0.28	0.21	11.86	3.20	0.70	0.88	1.98	20.46	4.91	4.25	0.79	2.10	16.22	10.47	3.12	0.98	11.83
			B	2.69	3.18	0.27	0.09	0.18	0.16	10.46	2.83	0.83	0.60	2.05	19.06	4.57	5.24	0.57	1.98	16.92	9.51	3.12	1.04	17.32
		15°C	A	1.63	5.20	0.39	0.10	0.31	0.16	12.55	3.06	0.59	0.93	1.73	21.39	5.04	3.94	0.72	1.82	15.30	11.09	3.14	0.90	11.63
			B	2.18	3.42	0.31	0.09	0.20	0.11	12.17	3.04	0.70	0.66	1.91	21.41	4.85	4.67	0.62	1.89	16.80	9.76	3.01	1.00	13.39
		25°C	A	2.16	4.67	0.34	0.13	0.06	0.14	12.24	2.49	0.80	0.59	2.07	21.07	4.65	5.16	0.55	1.68	16.09	7.45	2.53	0.79	16.50
			B	2.02	2.65	0.28	0.09	0.15	0.12	10.93	2.73	0.78	0.52	1.96	21.52	4.79	5.15	0.57	1.74	17.13	8.50	2.68	0.85	16.85
		35°C	A	2.15	4.73	0.40	0.12	0.32	0.20	12.03	3.01	0.81	0.91	1.99	20.62	5.11	4.36	0.84	1.94	15.88	10.29	3.13	0.95	12.37
			B	4.35	2.31	0.22	0.08	0.15	0.12	11.31	2.74	0.86	0.52	2.15	20.89	4.76	5.59	0.52	1.85	16.48	8.39	2.68	0.80	17.58
		5°C	A	9.34	4.07	0.35	0.12	0.21	0.17	11.48	3.02	0.77	0.69	1.97	20.21	4.66	4.71	0.58	1.94	16.78	9.54	2.87	0.86	14.99
			B	1.38	3.26	0.28	0.11	0.18	0.13	12.51	3.08	1.03	0.59	2.19	21.26	4.85	5.48	0.54	1.64	16.11	8.10	2.71	0.71	15.25
		15°C	A	1.70	3.59	0.49	0.12	0.20	0.14	11.68	3.04	0.73	0.63	1.82	21.14	4.50	4.26	0.60	1.93	16.73	9.91	3.04	0.97	14.49

ANNEXES

	B	2.23	2.40	0.34	0.10	0.15	0.11	11.89	3.07	0.82	0.54	2.02	21.72	4.99	5.18	0.53	1.70	16.93	9.10	2.85	0.75	14.84
25°C	A	4.45	2.53	0.34	0.08	0.14	0.14	11.14	2.73	0.75	0.48	1.90	21.74	4.69	5.16	0.53	1.71	16.93	8.02	2.54	0.80	17.66
	B	1.20	2.71	0.31	0.08	0.15	0.11	12.23	2.70	0.73	0.49	1.98	22.89	4.83	5.34	0.47	1.56	16.55	7.34	2.44	0.70	16.39
35°C	A	5.49	2.13	0.34	0.12	0.15	0.14	10.19	3.28	0.87	0.45	2.15	19.41	4.60	5.10	0.54	2.22	16.98	11.40	3.29	0.91	15.73
	B	2.66	2.18	0.31	0.11	0.15	0.11	10.68	3.55	0.93	0.49	2.15	19.44	4.75	5.23	0.53	2.06	17.06	11.14	3.20	0.84	15.07

FP88	0 mois		0.24	2.40	0.35	0.03	0.14	0.07	10.28	1.66	0.45	0.26	1.47	23.89	5.28	4.66	0.40	1.67	13.00	8.35	4.20	0.65	20.79	
		5°C	A	0.31	3.49	0.47	0.07	0.24	0.12	11.18	2.01	0.53	0.48	1.29	24.65	5.26	4.31	0.52	1.58	12.92	8.43	3.80	1.31	17.35
		B	0.63	2.47	0.24	0.03	0.18	0.06	10.10	1.68	0.57	0.40	1.37	23.64	5.08	5.07	0.45	1.54	12.28	7.14	3.84	0.71	23.12	
	3 mois	15°C	A	0.77	2.37	0.37	0.05	0.16	0.10	10.50	1.78	0.58	0.38	1.39	23.38	5.27	5.00	0.47	1.52	12.22	7.64	4.13	0.65	22.06
			B	0.43	3.25	0.44	0.04	0.18	0.05	10.53	1.95	0.42	0.36	1.26	24.40	5.36	4.60	0.38	1.63	13.11	8.78	4.26	0.72	18.26
		25°C	A	0.80	4.92	0.55	0.05	0.23	0.13	10.30	1.87	0.51	0.41	1.46	22.70	5.35	4.62	0.46	1.75	12.50	8.51	4.36	0.96	18.34
			B	0.67	3.72	0.50	0.05	0.21	0.12	10.74	1.83	0.52	0.39	1.29	23.60	5.43	4.49	0.42	1.74	13.14	8.87	4.05	0.91	17.99
		35°C	A	0.55	3.81	0.46	0.07	0.20	0.08	11.14	2.07	0.54	0.41	1.45	24.62	5.34	4.61	0.45	1.70	12.57	7.73	3.87	0.77	18.11
			B	0.14	5.74	0.63	0.08	0.30	0.11	12.26	2.11	0.57	0.53	1.33	25.84	5.16	4.04	0.49	1.59	12.56	7.86	3.26	1.02	14.51
	12 mois	5°C	A	0.73	3.44	0.35	0.05	0.16	0.08	10.60	1.74	0.48	0.36	1.35	23.55	5.04	4.74	0.49	1.47	12.14	7.38	3.93	0.68	21.99
			B	0.15	3.43	0.51	0.05	0.21	0.10	11.07	2.17	0.55	0.44	1.23	25.67	5.37	4.15	0.53	1.55	13.25	8.86	3.77	1.13	15.97

	15°C	A	0.68	3.96	0.63	0.06	0.23	0.12	11.44	1.97	0.59	0.45	1.33	23.85	5.11	4.55	0.93	1.57	12.07	8.21	4.08	0.82	18.05
		B	0.12	4.72	0.63	0.05	0.24	0.10	11.23	2.13	0.54	0.45	1.22	25.47	5.45	4.22	0.49	1.51	13.09	8.81	3.63	0.82	15.20
	25°C	A	0.68	3.72	0.61	0.06	0.23	0.13	11.08	1.88	0.44	0.44	1.18	24.15	5.26	4.38	0.44	1.67	13.05	7.84	3.83	0.96	18.65
		B	0.20	6.21	0.66	0.05	0.27	0.10	11.47	2.02	0.45	0.44	1.24	25.36	5.57	4.08	0.44	1.59	13.13	8.17	3.59	0.70	14.46
	35°C	A	0.36	4.47	0.35	0.06	0.33	0.14	11.91	1.26	0.54	0.48	1.38	25.61	3.67	4.41	0.49	1.42	12.37	4.80	2.56	0.56	23.19
		B	0.15	4.04	0.47	0.05	0.19	0.08	12.37	1.97	0.51	0.39	1.30	27.49	4.94	4.49	0.37	1.55	11.84	6.84	3.50	0.51	17.11

MOU32	3 mois	0 mois		2.49	1.91	0.46	0.11	0.15	0.21	7.50	3.70	0.78	0.45	2.13	17.96	7.57	5.57	0.67	1.59	15.34	11.82	3.95	1.13	17.03
			5°C	A	1.09	3.63	0.68	0.16	0.27	0.25	8.32	4.40	0.84	0.49	1.90	19.40	6.83	4.80	0.51	2.09	14.20	11.44	3.93	1.00
		B		0.69	3.57	0.65	0.15	0.21	0.27	7.47	4.53	0.83	0.45	1.75	19.50	6.92	4.19	0.63	2.10	14.98	12.27	3.28	1.26	14.99
		15°C	A	0.83	3.59	0.69	0.15	0.21	0.29	7.56	4.23	0.87	0.44	1.75	19.89	7.03	4.36	0.61	2.01	14.61	11.98	3.55	1.05	15.12
			B	0.52	3.69	0.70	0.16	0.16	0.21	8.81	4.69	0.85	0.43	1.84	19.74	6.71	4.89	0.43	1.98	14.07	11.08	3.73	0.84	15.02
		25°C	A	1.11	3.65	0.74	0.12	0.26	0.28	8.82	4.76	0.88	0.41	1.88	18.91	6.76	4.67	1.06	2.34	14.52	10.69	3.09	0.69	15.47
			B	0.80	4.00	0.74	0.15	0.20	0.27	8.06	4.56	0.84	0.48	1.79	18.83	7.00	4.31	0.56	2.24	14.91	11.48	3.10	0.89	15.60
		35°C	A	0.75	4.08	0.72	0.16	0.17	0.24	8.59	4.57	0.81	0.40	1.96	18.91	6.55	4.92	0.57	2.46	14.62	11.00	3.24	0.85	15.18
			B	0.86	3.99	0.96	0.14	0.19	0.24	8.91	5.00	0.71	0.40	1.85	18.85	6.52	4.56	0.50	2.25	14.83	11.03	2.94	0.84	15.28

ANNEXES

12 mois	5°C	A	1.19	2.65	0.52	0.12	0.20	0.21	8.10	4.77	0.74	0.44	1.43	20.35	6.82	4.42	0.43	2.32	15.12	12.40	3.20	1.05	14.70
		B	0.15	4.24	0.71	0.11	0.24	0.27	7.37	5.15	0.84	0.56	1.54	22.89	7.52	4.23	0.57	1.88	14.09	11.06	3.03	1.00	12.70
	15°C	A	0.60	2.35	0.51	0.10	0.14	0.23	9.04	4.88	0.91	0.39	1.75	19.33	6.53	5.39	0.50	2.33	13.99	10.89	3.29	0.94	16.51
		B	3.17	3.67	0.74	0.15	0.20	0.27	9.13	4.79	1.16	0.52	2.04	19.42	7.52	5.11	0.61	1.82	14.36	11.44	3.40	0.86	12.78
	25°C	A	0.83	2.45	0.64	0.12	0.13	0.25	8.19	4.65	0.90	0.46	1.84	19.48	6.96	4.67	0.53	2.45	15.46	11.41	3.13	1.02	15.28
		B	3.60	2.43	0.58	0.12	0.12	0.21	9.36	5.00	1.05	0.40	2.02	19.04	6.96	5.92	0.55	2.48	14.19	9.47	3.28	0.76	16.05
	35°C	A	2.62	2.35	0.49	0.10	0.15	0.17	9.95	4.51	0.77	0.45	1.62	20.25	5.55	4.93	0.43	2.02	17.18	7.36	2.46	0.79	18.45
		B	3.22	3.10	0.70	0.12	0.15	0.16	10.64	5.43	0.96	0.54	1.88	20.95	5.99	5.55	0.51	1.97	14.92	7.34	2.50	0.68	15.90

BOU11	3 mois	5°C	A	0.11	1.22	0.19	0.07	0.20	0.29	3.87	4.10	0.87	0.64	2.05	19.29	7.57	4.05	0.74	1.05	11.91	18.73	1.92	2.15	19.09
			B	1.79	2.80	0.59	0.19	0.21	0.22	7.96	4.85	0.96	0.57	2.67	18.08	7.12	5.74	0.71	1.84	15.08	9.59	3.27	1.08	16.48
		15°C	A	0.62	2.38	0.48	0.11	0.28	0.32	4.82	4.92	0.68	0.59	1.78	18.27	7.89	4.03	0.80	1.59	16.26	10.93	2.70	1.89	19.28
			B	0.85	2.79	0.53	0.15	0.21	0.25	6.99	4.17	0.78	0.55	2.36	19.02	7.46	4.59	0.75	1.74	16.19	10.31	2.91	1.43	16.81
		25°C	A	0.21	2.44	0.16	0.06	0.13	0.32	4.32	3.75	0.79	0.45	1.77	18.90	6.82	3.79	0.53	1.15	12.39	17.17	1.73	2.19	21.17
			B	0.95	3.97	0.78	0.23	0.26	0.27	8.88	4.87	0.98	0.55	2.54	19.85	7.43	5.03	0.66	1.70	15.50	9.01	2.65	1.04	13.79
		35°C	A	2.80	2.71	0.51	0.19	0.17	0.20	7.49	4.53	0.91	0.46	2.59	17.38	6.68	5.83	0.82	1.84	15.46	8.91	3.34	1.16	18.84
			B	0.10	1.81	0.00	0.00	0.00	0.00	5.30	3.74	1.05	0.43	2.29	26.16	6.24	3.42	1.12	1.33	14.07	11.88	2.34	6.25	12.57

	B	0.39	2.48	0.50	0.06	0.23	0.25	4.90	4.37	0.66	0.53	1.65	19.84	7.39	3.92	0.81	1.65	18.03	10.11	2.47	2.07	18.10	
12 mois	5°C	A	0.15	1.87	0.00	0.00	0.00	3.56	4.34	0.98	0.66	1.30	15.97	6.66	2.96	0.54	1.59	15.97	18.83	2.33	1.26	21.17	
		B	0.13	2.96	0.48	0.07	0.24	0.25	5.29	5.01	0.60	0.49	1.71	19.48	8.24	4.14	0.80	1.31	16.54	10.56	2.56	1.57	17.71
	15°C	A	1.19	3.30	0.55	0.14	0.18	0.19	7.18	4.57	0.77	0.44	2.14	19.10	6.72	5.02	0.51	1.53	16.67	9.31	2.87	1.07	17.75
		B	0.20	2.82	0.60	0.11	0.22	0.26	6.30	4.33	0.75	0.52	1.87	21.23	7.61	4.53	0.73	1.42	16.10	11.28	2.63	1.30	15.41
	25°C	A	0.16	0.96	0.30	0.09	0.17	0.33	2.92	2.34	0.39	0.65	1.25	14.73	5.52	2.97	0.71	1.39	14.30	14.72	2.37	3.98	29.89
		B	1.43	2.10	1.12	0.15	0.16	0.18	8.10	4.56	0.92	0.52	2.37	19.42	7.05	5.47	0.75	1.60	16.80	8.81	2.90	1.13	15.92
	35°C	A	0.11	0.73	0.00	0.00	0.00	0.00	3.28	3.67	0.66	0.92	1.20	22.17	5.17	3.32	0.78	1.97	23.46	6.72	2.11	4.69	19.15
		B	0.01	20.72	0.10	0.00	0.11	0.11	2.62	3.27	0.50	0.42	0.54	37.57	3.06	1.91	0.49	0.89	10.63	5.02	0.54	3.69	7.82

Tableau supplémentaire 3.4. Concentrations et abondances relatives (%) des brGDGT pour les 4 échantillons de sols placés en microcosme, aux différents pas de temps et aux différentes températures

Temps d'incubation	Température d'incubation (°C)	Réplicat	Concentration (ng/g sol)	Abondances relatives (%)															
				1022a	1020a	1018	1036a	1036b	1034a	1034b	1032a	1032b	1050a	1050b	1048a	1048b	1046a	1046b	
				Ia	Ib	Ic	IIa	IIa'	IIb	IIb'	IIc	IIc'	IIIa	IIIa'	IIIb	IIIb'	IIIc	IIIc'	
KS48	0 mois		1922.02	17.47	7.21	1.54	17.90	19.08	7.28	7.78	1.24	0.00	10.00	9.57	0.51	0.43	0.00	0.00	
		5°C	A	700.75	22.26	5.14	0.00	23.82	19.99	5.45	3.44	1.13	0.00	9.16	8.83	0.25	0.52	0.00	0.00
	B		596.80	24.76	4.38	0.96	26.36	19.58	4.94	2.58	1.00	0.00	8.98	6.45	0.00	0.00	0.00	0.00	
	12 mois	15°C	A	849.34	22.34	4.49	0.96	24.82	19.81	5.97	2.60	1.25	0.00	10.23	6.93	0.26	0.34	0.00	0.00
			B	848.65	23.67	4.27	0.96	26.66	19.79	5.59	2.46	1.19	0.00	9.00	5.92	0.21	0.29	0.00	0.00
	35°C	A	831.22	27.19	4.71	0.80	27.38	16.95	4.84	1.90	0.95	0.00	9.90	4.90	0.20	0.29	0.00	0.00	
		B	898.69	30.00	4.26	0.88	27.69	16.96	4.14	1.50	1.14	0.00	8.99	4.00	0.24	0.20	0.00	0.00	
	35°C	A	771.84	36.90	4.13	0.94	19.24	16.71	4.72	2.33	1.07	0.00	7.22	6.05	0.26	0.28	0.06	0.09	
		B	320.73	44.40	3.63	0.75	19.29	12.39	4.57	2.17	0.90	0.00	6.25	4.89	0.26	0.49	0.00	0.00	
	FP88	0 mois		794.73	29.50	3.32	2.86	36.29	13.73	1.96	0.31	0.34	0.34	9.82	1.49	0.00	0.00	0.00	0.03
12 mois			5°C	A	721.27	30.87	3.44	3.47	31.94	12.56	2.43	2.44	0.43	0.00	10.03	1.70	0.00	0.00	0.34
		B		747.35	31.38	3.68	3.58	34.05	11.81	2.14	0.49	0.38	0.00	10.69	1.81	0.00	0.00	0.00	0.00

	15°C	A	471.20	32.29	3.95	3.71	33.77	11.86	2.35	0.45	0.46	0.00	9.79	1.36	0.00	0.00	0.00	0.00
		B	717.48	32.16	4.07	3.84	34.87	10.26	2.31	0.47	0.38	0.00	10.19	1.44	0.00	0.00	0.00	0.00
	25°C	A	301.84	31.31	4.12	4.13	34.67	11.67	2.26	0.44	0.53	0.00	9.49	1.21	0.07	0.03	0.04	0.03
		B	686.79	33.64	4.18	3.96	34.52	9.84	2.26	0.34	0.42	0.00	9.41	1.42	0.00	0.00	0.00	0.00
	35°C	A	69.05	39.37	3.94	4.32	31.18	9.57	1.72	0.51	0.79	0.00	7.43	1.17	0.00	0.00	0.00	0.00
		B	707.40	36.59	3.71	3.94	31.97	10.44	1.95	0.35	0.37	0.51	8.86	1.32	0.00	0.00	0.00	0.00

MOU3 2	0 mois		6160.47	6.71	7.72	0.97	5.84	19.27	9.20	20.20	0.66	0.00	6.27	19.07	0.78	3.30	0.00	0.00	
	5°C	A	734.47	7.74	7.98	0.89	4.67	22.02	6.04	19.30	0.21	0.88	6.40	20.11	0.57	2.88	0.00	0.29	
		B	2026.63	7.85	8.14	0.95	5.86	20.84	7.91	19.03	0.22	0.00	5.73	19.85	0.57	3.06	0.00	0.00	
	15°C	A	384.99	9.45	18.17	3.07	6.98	12.90	11.77	21.53	0.95	0.00	3.94	8.22	0.62	2.41	0.00	0.00	
		B	1427.72	8.05	8.35	0.92	5.15	22.37	6.62	18.41	0.21	0.00	6.24	20.08	0.54	3.08	0.00	0.00	
	12 mois	25°C	A	620.66	9.07	19.45	3.35	6.42	11.94	10.06	24.31	0.93	0.00	3.74	7.18	0.76	2.80	0.00	0.00
			B	1404.87	7.80	8.96	1.10	5.16	20.63	7.00	20.69	0.26	0.00	6.37	18.27	0.70	3.05	0.00	0.00
	35°C	A	191.00	11.71	11.57	1.68	5.34	19.58	5.51	18.98	0.24	0.90	5.70	15.76	0.53	2.50	0.00	0.00	
		B	440.25	14.25	9.12	1.32	4.92	19.44	6.73	18.11	0.33	0.00	6.26	16.11	0.68	2.72	0.00	0.00	

BOU11	0 mois		40.39	14.58	11.01	2.22	9.60	17.18	9.73	11.65	0.57	0.49	7.26	14.03	0.74	0.74	0.20	0.00
	12 mois	5°C	A	220.45	9.38	17.74	2.98	5.11	14.41	10.52	22.24	0.98	1.56	3.66	8.19	0.73	2.50	0.00
B			1166.18	9.00	17.55	2.86	6.35	13.45	9.82	23.10	0.82	1.42	3.84	8.43	0.71	2.65	0.00	0.00

ANNEXES

15°C	A	950.19	8.74	12.78	1.93	5.78	18.04	8.62	20.16	0.57	0.00	5.09	14.85	0.65	2.71	0.00	0.08
	B	586.99	9.13	17.82	2.97	6.62	12.89	9.94	22.83	0.86	1.54	3.84	8.28	0.72	2.57	0.00	0.00
25°C	A	436.87	7.61	9.29	1.22	5.41	19.81	8.38	20.67	0.31	0.00	6.53	17.07	0.62	3.09	0.00	0.00
	B	699.19	8.59	18.16	3.03	5.16	12.55	11.13	24.89	0.84	1.47	3.55	7.53	0.68	2.41	0.00	0.00
35°C	A	215.55	20.25	26.44	1.36	4.88	9.60	5.34	19.20	0.57	1.09	2.93	5.64	0.54	1.81	0.09	0.26
	B	365.29	14.29	20.79	3.42	6.10	10.97	8.41	21.12	0.72	1.41	3.86	6.04	0.58	2.29	0.00	0.00

Annexe 2 – Données supplémentaires du Chapitre 4 –

Supplementary table 4.1. Relative abundances (%) of the normal, iso and anteiso 3-OH FAs in soils from the Peruvian Andes, Chilean Andes, Mts Pollino and Shegyla.

Samples	Localisation	<i>nC10</i>	<i>iC11</i>	<i>aC11</i>	<i>nC11</i>	<i>iC12</i>	<i>nC12</i>	<i>iC13</i>	<i>aC13</i>	<i>nC13</i>	<i>iC14</i>	<i>nC14</i>	<i>iC15</i>	<i>aC15</i>	<i>nC15</i>	<i>iC16</i>	<i>nC16</i>	<i>iC17</i>	<i>aC17</i>	<i>nC17</i>	<i>nC18</i>
1	Peruvian Andes	2.11	0.75	0.06	0.22	0.15	7.64	3.16	0.31	0.56	1.22	23.14	7.02	2.09	0.85	0.84	20.60	8.76	1.07	1.11	18.35
2	Peruvian Andes	1.06	0.28	0.04	0.11	0.02	5.18	1.23	0.10	0.56	0.92	24.88	9.17	1.59	0.62	0.47	18.27	5.84	0.59	0.97	28.10
3	Peruvian Andes	1.39	0.40	0.05	0.09	0.03	4.80	1.21	0.18	0.47	1.17	25.14	9.36	2.05	0.59	0.54	20.13	6.80	0.64	0.91	24.05
4	Peruvian Andes	1.98	0.48	0.05	0.15	0.06	6.36	1.67	0.24	0.53	1.23	24.98	7.92	2.56	0.62	0.58	20.30	7.73	0.95	1.02	20.60
5	Peruvian Andes	1.60	0.18	0.02	0.11	0.03	5.99	1.38	0.12	0.51	1.03	26.14	7.78	2.80	0.53	0.63	17.88	8.44	1.22	0.93	22.66
6	Peruvian Andes	1.27	0.16	0.03	0.11	0.02	5.42	0.90	0.03	0.44	1.05	24.80	8.06	2.96	0.43	0.54	17.81	7.45	1.07	0.87	26.59
7	Peruvian Andes	1.26	0.16	0.01	0.08	0.03	5.15	0.88	0.84	0.38	1.08	24.88	7.68	3.29	0.47	0.60	16.49	6.86	0.96	0.80	28.09
8	Peruvian Andes	1.60	0.11	0.04	0.11	0.03	6.11	1.19	0.18	0.40	1.11	24.68	7.83	3.90	0.46	0.56	16.89	6.78	1.14	0.72	26.15
9	Peruvian Andes	1.32	0.08	0.03	0.11	0.02	6.35	1.12	0.18	0.44	1.18	23.88	8.11	3.87	0.46	0.55	15.63	8.07	1.47	0.85	26.29
10	Peruvian Andes	0.97	0.12	0.03	0.09	0.01	6.94	1.42	0.19	0.42	1.34	23.51	8.81	4.67	0.34	0.67	15.53	7.53	1.65	0.75	24.99
11	Peruvian Andes	1.67	0.18	0.03	0.10	0.04	7.61	1.74	0.26	0.45	1.56	23.12	9.17	4.03	0.58	0.94	15.89	9.67	2.13	0.91	19.93
12	Peruvian Andes	1.45	0.21	0.02	0.10	0.03	7.65	1.06	0.21	0.45	1.12	23.31	7.98	3.43	0.39	0.60	16.73	7.06	1.59	0.91	25.69
13	Peruvian Andes	1.26	0.15	0.02	0.12	0.04	7.64	1.56	0.27	0.47	1.61	22.37	8.71	4.46	0.49	0.82	15.33	8.81	2.09	0.88	22.91
14	Peruvian Andes	4.71	0.06	0.03	0.23	0.03	7.77	0.99	0.12	0.44	1.06	22.26	4.98	4.37	0.49	0.41	12.08	3.94	1.70	0.84	33.49
15	Mt. Shegyla, Tibet	6.20	0.36	0.05	0.33	0.09	10.16	2.27	0.47	0.58	1.35	21.29	5.22	3.81	0.61	0.85	16.90	7.39	1.60	0.79	19.69
16	Mt. Shegyla, Tibet	2.83	0.43	0.06	0.22	0.16	9.97	3.48	0.71	0.73	1.62	20.70	5.53	3.92	0.88	1.36	17.79	11.27	2.00	1.07	15.27
17	Mt. Shegyla, Tibet	2.89	0.31	0.04	0.30	0.12	9.95	3.08	0.56	0.85	1.43	23.67	5.41	3.72	0.91	1.24	17.87	9.14	1.91	1.11	15.47
18	Mt. Shegyla, Tibet	3.62	0.34	0.08	0.30	0.13	9.64	3.21	0.58	0.66	1.59	21.79	6.76	4.03	0.75	1.48	17.33	10.28	2.83	1.01	13.60
19	Mt. Shegyla, Tibet	2.15	0.19	0.08	0.33	0.21	7.99	2.83	0.58	0.80	1.66	19.61	9.12	3.75	0.92	1.81	18.24	13.58	3.23	1.19	11.72
20	Mt. Shegyla, Tibet	3.39	0.25	0.06	0.38	0.17	8.47	3.02	0.51	0.69	1.79	20.09	9.00	3.69	0.95	1.79	17.20	12.53	2.96	1.38	11.65
21	Mt. Shegyla, Tibet	3.69	0.35	0.05	0.24	0.07	12.80	2.27	0.41	0.54	1.73	23.45	6.95	3.80	0.46	0.96	15.96	8.20	2.11	0.58	15.38

ANNEXES

22	Mt. Shegyla, Tibet	3.51	0.30	0.04	0.29	0.08	12.29	1.92	0.49	0.57	1.44	22.96	6.65	3.78	0.45	0.94	16.77	6.60	2.08	0.69	18.14
23	Mt. Shegyla, Tibet	3.89	0.21	0.04	0.21	0.06	13.47	2.26	0.34	0.56	1.75	22.38	6.98	3.90	0.44	1.01	18.19	6.03	1.52	0.62	16.14
24	Mt. Shegyla, Tibet	3.89	0.20	0.05	0.21	0.07	13.03	2.42	0.35	0.61	1.83	21.84	6.93	3.91	0.40	0.99	18.55	6.38	1.61	0.70	16.01
25	Mt. Shegyla, Tibet	4.12	0.22	0.05	0.20	0.04	13.00	1.94	0.25	0.51	1.55	21.11	6.42	3.44	0.40	0.95	19.44	5.50	1.36	0.76	18.75
26	Mt. Shegyla, Tibet	3.11	0.11	0.04	0.21	0.06	12.84	2.38	0.38	0.67	1.93	23.44	7.12	4.72	0.47	0.72	15.03	4.89	1.58	0.57	19.75
27	Mt. Shegyla, Tibet	3.10	0.12	0.06	0.26	0.05	12.75	2.47	0.45	0.66	1.97	22.66	6.20	4.91	0.40	0.66	15.74	5.39	1.63	0.42	20.10
28	Mt. Shegyla, Tibet	3.65	0.22	0.05	0.26	0.06	12.96	2.55	0.40	0.71	2.09	23.63	6.75	4.56	0.45	0.78	15.41	6.55	1.95	0.61	16.39
29	Mt. Shegyla, Tibet	4.01	0.19	0.06	0.27	0.06	12.16	2.09	0.43	0.69	1.78	22.18	4.91	5.01	0.50	0.86	16.81	5.95	1.99	0.58	19.47
30	Mt. Shegyla, Tibet	4.22	0.17	0.06	0.36	0.06	12.22	2.19	0.29	1.25	1.92	21.30	5.75	4.44	0.78	0.87	16.08	5.95	2.29	0.76	19.05
31	Mt. Shegyla, Tibet	3.48	0.22	0.06	0.31	0.06	12.08	2.40	0.33	0.71	2.25	22.79	6.35	4.70	0.60	0.79	16.55	8.27	2.26	0.62	15.18
32	Mt. Pollino, Italy	1.97	0.60	0.05	0.18	0.24	8.74	3.99	0.61	0.55	1.62	17.23	8.63	2.55	0.94	2.17	15.45	18.79	2.32	1.95	11.43
33	Mt. Pollino, Italy	3.22	0.57	0.09	0.17	0.24	9.29	4.16	0.87	1.34	1.79	17.21	5.19	3.46	1.43	2.23	18.13	12.81	2.10	1.64	14.08
34	Mt. Pollino, Italy	3.97	0.86	0.06	0.16	0.30	10.04	3.65	0.77	0.56	1.29	19.04	6.95	2.71	0.64	2.03	16.64	16.19	2.66	1.16	10.34
35	Mt. Pollino, Italy	4.43	0.71	0.07	0.12	0.25	9.12	3.45	0.64	0.45	1.30	17.80	6.58	2.95	0.71	2.19	16.21	16.20	3.01	1.27	12.53
36	Mt. Pollino, Italy	4.33	0.69	0.06	0.29	0.10	11.55	3.11	0.53	0.80	1.66	20.93	5.89	3.34	1.00	1.78	14.65	11.53	2.73	0.98	14.05
37	Mt. Pollino, Italy	4.18	0.49	0.08	0.41	0.22	8.71	3.88	0.67	0.68	1.83	18.38	6.64	3.76	1.23	2.42	16.78	11.50	2.95	1.61	13.58
38	Mt. Pollino, Italy	3.11	0.65	0.09	0.12	0.27	10.37	3.58	0.84	0.51	1.69	17.40	5.80	3.48	0.83	2.26	17.40	15.22	2.81	1.47	12.08
39	Mt. Pollino, Italy	4.53	1.60	0.06	0.15	0.35	11.08	4.03	0.83	0.37	1.53	18.99	7.66	2.90	0.50	1.75	15.26	16.07	3.00	0.72	8.61
40	Mt. Pollino, Italy	6.15	0.41	0.10	0.22	0.22	11.35	3.82	0.73	0.58	1.66	21.18	7.62	4.03	0.53	1.60	13.52	12.05	2.79	0.62	10.84
41	Mt. Pollino, Italy	6.45	0.53	0.06	0.37	0.20	10.87	3.51	0.57	0.85	1.72	19.53	7.22	3.17	0.92	2.07	13.96	13.73	3.16	1.00	10.12
42	Mt. Pollino, Italy	4.67	0.67	0.09	0.22	0.34	11.88	4.50	0.99	0.70	1.82	18.01	6.77	3.72	0.59	1.98	13.34	14.85	3.31	0.73	10.83
43	Mt. Pollino, Italy	3.14	0.46	0.11	0.14	0.22	9.66	4.52	0.99	0.38	2.26	16.09	5.55	5.85	0.57	2.84	14.75	12.52	3.91	1.08	14.95
44	Mt. Pollino, Italy	6.31	0.52	0.07	0.12	0.17	9.89	3.43	0.65	0.53	1.87	17.35	4.74	3.75	0.60	2.75	17.29	12.16	2.99	1.06	13.76
45	Chilean Andes	6.04	0.57	0.05	0.25	0.15	12.13	2.94	0.60	0.52	1.52	22.27	6.19	3.41	0.65	1.14	14.15	9.96	2.20	0.91	14.35
46	Chilean Andes	5.54	1.03	0.05	0.25	0.14	14.77	3.10	0.56	0.55	1.36	24.79	5.85	2.73	0.54	1.04	12.62	9.71	2.15	0.61	12.60
47	Chilean Andes	6.47	1.32	0.01	0.23	0.08	16.80	1.99	0.31	0.50	1.69	22.87	5.05	2.10	0.55	0.91	12.85	9.99	2.17	0.52	13.58
48	Chilean Andes	5.59	0.58	0.07	0.21	0.24	12.84	3.93	1.20	0.44	1.76	21.09	5.86	3.45	0.73	1.32	14.57	11.34	2.28	0.79	11.71
49	Chilean Andes	6.15	1.21	0.05	0.23	0.18	12.61	2.76	0.59	0.79	2.08	16.11	7.05	3.47	0.90	1.40	16.01	12.59	2.92	0.76	12.15
50	Chilean Andes	4.23	1.22	0.04	0.25	0.07	15.22	1.55	6.88	0.67	1.66	23.34	4.26	2.23	0.46	0.46	11.13	8.72	1.90	1.00	14.72

51	Chilean Andes	4.43	0.90	0.04	0.25	0.10	14.97	1.93	0.31	0.61	1.62	24.86	5.18	2.82	0.44	0.90	13.17	8.49	2.52	0.81	15.64
52	Chilean Andes	5.02	0.55	0.04	0.19	0.16	13.02	3.19	0.51	0.64	1.70	23.00	6.27	3.07	0.54	1.34	15.22	12.12	2.59	0.64	10.19
53	Chilean Andes	6.91	0.80	0.05	0.28	0.19	13.98	2.06	0.32	0.51	1.22	25.76	5.07	2.55	0.49	0.72	12.75	8.07	1.91	0.71	15.64
54	Chilean Andes	6.78	0.47	0.07	0.18	0.27	10.64	4.07	0.98	0.67	1.86	17.94	6.58	3.43	0.74	2.33	14.33	14.79	3.55	0.83	9.48

Supplementary table 4.2. Relative abundances (%) of the brGDGTs (with separation of the 5- and 6-methyl isomers) from the Peruvian Andes, Chilean Andes, Mts. Pollino and Shegyla (n.d., not detected).

Samples	Localisation	<i>IIIa</i>	<i>IIIa'</i>	<i>IIIb</i>	<i>IIIb'</i>	<i>IIIc</i>	<i>IIIc'</i>	<i>IIa</i>	<i>IIa'</i>	<i>IIb</i>	<i>IIb'</i>	<i>IIc</i>	<i>IIc'</i>	<i>Ia</i>	<i>Ib</i>	<i>Ic</i>
1	Peruvian Andes	0.21	0.23	n.d	n.d	n.d	n.d	2.84	3.10	0.39	0.93	0.16	n.d	81.50	8.08	2.56
2	Peruvian Andes	0.11	0.09	n.d	n.d	n.d	n.d	2.21	0.16	0.13	0.00	0.14	n.d	94.87	1.37	0.93
3	Peruvian Andes	0.08	0.02	n.d	n.d	n.d	n.d	2.14	0.28	0.08	0.04	0.07	n.d	93.56	2.41	1.33
4	Peruvian Andes	0.37	0.12	0.02	0.04	0.13	n.d	8.27	1.14	0.34	0.07	0.16	n.d	84.71	3.39	1.23
5	Peruvian Andes	0.35	0.09	n.d	n.d	n.d	0.02	7.19	0.75	0.26	0.04	0.15	n.d	87.11	2.68	1.36
6	Peruvian Andes	0.15	0.02	n.d	n.d	n.d	n.d	4.11	0.39	0.10	0.04	0.05	0.01	92.20	1.71	1.22
7	Peruvian Andes	0.16	0.04	n.d	n.d	n.d	n.d	4.66	0.41	0.19	0.05	0.09	n.d	90.62	2.30	1.50
8	Peruvian Andes	1.77	0.32	n.d	n.d	n.d	0.02	22.46	1.47	1.09	0.15	0.47	n.d	68.49	2.21	1.55
9	Peruvian Andes	0.82	0.21	n.d	n.d	n.d	0.01	14.67	1.21	0.63	0.08	0.26	n.d	78.37	1.94	1.81
10	Peruvian Andes	0.39	0.39	0.01	0.01	n.d	n.d	14.61	1.56	0.94	0.13	0.46	0.13	74.96	3.08	3.33
11	Peruvian Andes	1.50	0.25	n.d	n.d	n.d	n.d	25.49	1.25	1.30	0.16	0.58	0.07	66.03	2.05	1.31
12	Peruvian Andes	0.62	0.15	n.d	n.d	n.d	n.d	16.58	1.01	0.51	0.07	0.22	n.d	77.88	2.10	0.85
13	Peruvian Andes	0.37	0.38	n.d	n.d	n.d	n.d	26.07	1.44	1.50	0.14	0.54	0.11	65.26	2.54	1.64
14	Peruvian Andes	5.58	0.21	0.03	0.03	n.d	n.d	33.77	3.02	0.77	0.08	0.22	n.d	51.89	2.15	2.24
15	Mt. Shegyla, Tibet	3.57	1.10	n.d	n.d	n.d	n.d	31.39	9.82	1.09	0.69	n.d	n.d	47.91	3.75	0.67
16	Mt. Shegyla, Tibet	4.98	4.83	0.22	0.47	n.d	n.d	23.43	17.85	2.95	3.44	n.d	n.d	33.88	6.58	1.37
17	Mt. Shegyla, Tibet	3.95	3.28	n.d	n.d	n.d	n.d	23.69	16.22	2.85	2.74	n.d	n.d	40.78	5.46	1.04
18	Mt. Shegyla, Tibet	8.61	1.48	n.d	n.d	n.d	n.d	31.94	10.53	2.00	0.43	0.49	n.d	36.74	3.87	3.90
19	Mt. Shegyla, Tibet	11.16	4.40	0.13	0.23	n.d	n.d	27.51	19.35	1.79	2.06	0.47	0.17	28.40	3.53	0.79
20	Mt. Shegyla, Tibet	12.70	2.54	n.d	n.d	n.d	n.d	34.72	11.49	1.41	1.08	0.52	0.11	31.93	2.87	0.63
21	Mt. Shegyla, Tibet	7.50	0.41	n.d	n.d	n.d	n.d	38.98	6.13	0.71	0.19	n.d	n.d	44.22	1.43	0.44
22	Mt. Shegyla, Tibet	6.02	0.43	n.d	n.d	n.d	n.d	37.82	5.05	0.88	0.21	0.51	0.05	46.84	1.74	0.46
23	Mt. Shegyla, Tibet	14.10	0.54	n.d	n.d	n.d	n.d	38.23	3.59	0.56	0.10	0.46	n.d	41.49	0.73	0.18

24	Mt. Shegyla, Tibet	17.32	n.d	n.d	n.d	n.d	n.d	41.78	n.d	0.14	0.16	0.47	n.d	38.97	0.92	0.24
25	Mt. Shegyla, Tibet	13.66	n.d	n.d	n.d	n.d	n.d	40.31	n.d	0.18	0.26	0.51	0.06	43.79	0.98	0.26
26	Mt. Shegyla, Tibet	12.59	n.d	n.d	n.d	n.d	n.d	41.65	n.d	0.13	0.20	0.39	0.33	42.77	1.57	0.36
27	Mt. Shegyla, Tibet	13.85	0.64	0.16	0.16	n.d	n.d	39.49	5.57	0.12	0.34	0.36	0.40	36.45	1.97	0.49
28	Mt. Shegyla, Tibet	12.01	0.47	0.07	n.d	n.d	n.d	38.23	5.39	0.15	0.22	0.46	0.22	40.83	1.66	0.30
29	Mt. Shegyla, Tibet	18.74	0.56	n.d	n.d	n.d	n.d	42.39	3.74	0.94	n.d	0.42	0.24	30.95	1.64	0.38
30	Mt. Shegyla, Tibet	19.19	1.55	n.d	n.d	n.d	n.d	36.73	10.14	0.72	0.36	0.40	n.d	29.41	1.18	0.32
31	Mt. Shegyla, Tibet	22.30	0.63	n.d	n.d	n.d	n.d	40.94	5.47	0.38	0.19	0.17	n.d	29.25	0.45	0.21
32	Mt. Pollino, Italy	7.68	20.93	0.99	n.d	n.d	n.d	9.68	35.29	0.98	6.31	n.d	n.d	13.58	3.85	0.70
33	Mt. Pollino, Italy	4.12	17.06	1.53	n.d	n.d	n.d	7.65	35.81	1.61	9.41	n.d	n.d	16.64	6.17	n.d
34	Mt. Pollino, Italy	4.50	16.73	0.48	1.77	n.d	n.d	8.11	30.26	3.46	11.91	0.13	0.28	14.56	7.31	0.50
35	Mt. Pollino, Italy	3.89	19.17	0.69	1.69	n.d	n.d	9.04	29.02	2.88	14.01	n.d	n.d	11.19	7.30	1.12
36	Mt. Pollino, Italy	7.75	6.38	0.21	0.23	0.21	0.16	22.76	25.66	1.03	2.32	0.16	n.d	29.91	2.96	0.26
37	Mt. Pollino, Italy	5.67	9.83	0.42	1.07	n.d	n.d	15.82	24.46	8.09	7.30	n.d	n.d	15.15	8.88	3.31
38	Mt. Pollino, Italy	2.48	24.34	0.50	1.60	0.36	n.d	9.53	33.66	2.22	7.50	0.07	0.15	13.03	4.56	n.d
39	Mt. Pollino, Italy	3.82	23.37	0.28	1.41	n.d	n.d	10.41	30.85	2.04	9.86	0.11	0.25	12.26	5.08	0.28
40	Mt. Pollino, Italy	10.58	7.83	0.22	0.15	0.10	0.18	20.41	27.53	3.36	1.96	0.33	0.07	22.06	3.69	1.53
41	Mt. Pollino, Italy	10.49	10.05	0.48	0.30	n.d	n.d	19.27	29.82	1.50	2.98	n.d	n.d	21.67	3.12	0.30
42	Mt. Pollino, Italy	5.81	21.77	0.37	1.04	n.d	n.d	6.93	31.71	2.40	10.05	0.08	0.18	13.01	6.19	0.46
43	Mt. Pollino, Italy	7.97	19.61	0.37	1.96	n.d	n.d	8.81	27.14	4.94	12.98	0.16	0.48	9.23	5.61	0.72
44	Mt. Pollino, Italy	7.60	16.47	1.44	1.19	0.11	0.14	9.31	25.75	5.63	10.20	0.83	0.37	9.69	7.75	3.52
45	Chilean Andes	18.40	1.89	n.d	n.d	n.d	n.d	34.19	4.26	1.78	1.78	0.08	0.01	29.75	2.77	5.09
46	Chilean Andes	7.44	4.35	n.d	n.d	n.d	n.d	31.69	15.68	0.83	2.09	0.08	0.03	34.36	3.34	0.12
47	Chilean Andes	9.89	1.52	n.d	n.d	n.d	n.d	39.82	8.06	0.15	0.60	0.09	0.02	38.47	1.39	n.d
48	Chilean Andes	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d
49	Chilean Andes	10.48	2.58	n.d	n.d	n.d	n.d	33.25	13.90	0.33	1.01	0.09	0.01	36.91	1.33	0.10
50	Chilean Andes	13.80	0.88	n.d	n.d	n.d	n.d	43.65	0.39	0.37	0.15	0.04	0.01	40.18	0.53	n.d
51	Chilean Andes	14.09	1.40	n.d	n.d	n.d	n.d	39.47	6.48	0.31	0.02	n.d	n.d	36.93	0.62	0.68
52	Chilean Andes	10.96	7.28	n.d	n.d	n.d	n.d	25.23	20.55	3.57	1.88	n.d	n.d	26.25	3.92	0.36

ANNEXES

53	Chilean Andes	12.94	5.32	n.d	n.d	n.d	n.d	26.58	20.60	1.66	1.40	n.d	n.d	29.81	1.70	n.d
54	Chilean Andes	14.14	0.28	n.d	n.d	n.d	n.d	44.54	0.24	n.d	n.d	n.d	n.d	40.81	n.d	n.d

Supplementary table 4.3. Characteristics of the linear regressions (R^2 , RMSE, p-value, equation) between pH and RIAN/CBT'; MAAT and MBT_{5Me}/RAN₁₅/RAN₁₇ for the different altitudinal transect investigated

RIAN	Site	Study	R^2	p-value	RMSE	Equation
	Peruvian Andes	Present study	0.04	4.70×10^{-1}	None	None
	Chilean Andes	Present study	0.66*	4.24×10^{-3}	0.42	$\text{pH} = -5.50 \times \text{RIAN} + 7.59$
	Mt. Majella	Huguet et al., 2019	0.15	2.64×10^{-1}	None	None
	Mt. Pollino	Present study	0.29	5.60×10^{-2}	None	None
	Mt. Rungwe	Huguet et al., 2019	0.11	8.34×10^{-2}	None	None
	Mt. Shegyla	Present study	0.46*	2.84×10^{-3}	0.4	$\text{pH} = -3.77 \times \text{RIAN} + 6.98$
	Mt. Shennongjia	Wang et al., 2016	0.71*	7.40×10^{-8}	0.54	$\text{pH} = -6.79 \times \text{RIAN} + 9.37$
	Mts. Lautaret-Bauges	Vequaud et al., 2021	0.43*	3.06×10^{-7}	0.83	$\text{pH} = -6.59 \times \text{RIAN} + 8.04$
	Global	Present study	0.34*	7.39×10^{-17}	0.99	$\text{pH} = -5.40 \times \text{RIAN} + 7.86$

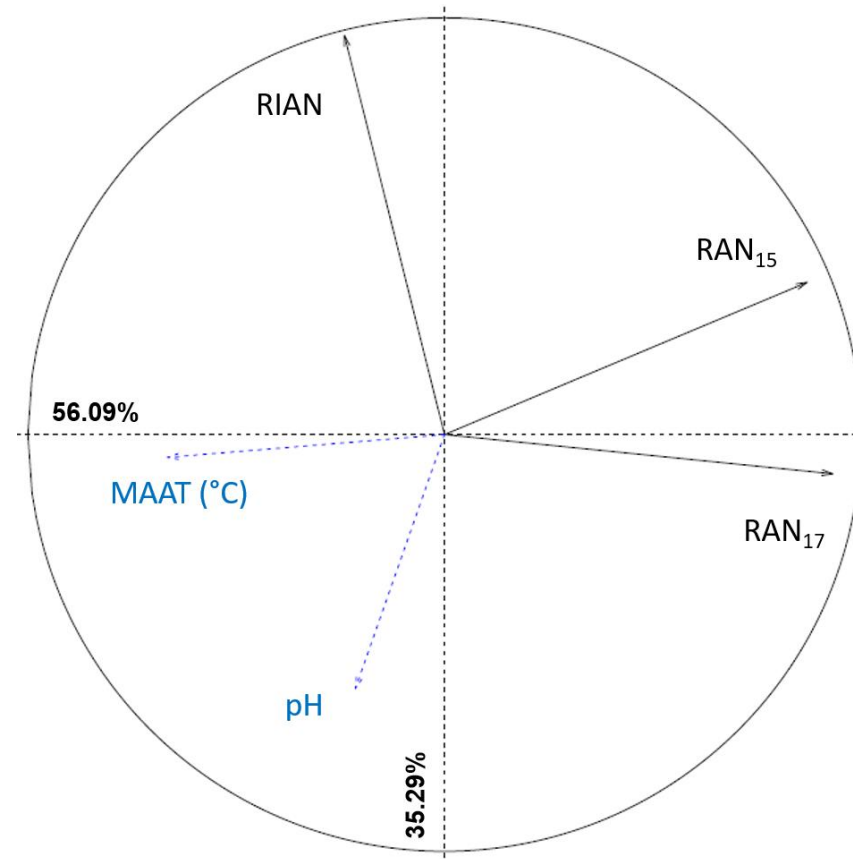
RAN15	Site	Study	R^2	p-value	RMSE	Equation
	Peruvian Andes	Present study	0.71*	1.52×10^{-4}	3.57	$\text{MAAT} = -1.73 \times \text{RAN}_{15} + 26.5$
	Chilean Andes	Present study	0.03	7.07×10^{-1}	None	None
	Mt. Majella	Huguet et al., 2019	0.49*	3.33×10^{-2}	3.34	$\text{MAAT} = -1.92 \times \text{RAN}_{15} + 13.6$
	Mt. Pollino	Present study	0.54*	4.39×10^{-3}	2.64	$\text{MAAT} = -1.21 \times \text{RAN}_{15} + 18$
	Mt. Rungwe	Huguet et al., 2019	0.79*	2.14×10^{-10}	2.13	$\text{MAAT} = -3.81 \times \text{RAN}_{15} + 29.7$
	Mt. Shegyla	Present study	0.57*	4.33×10^{-4}	1.98	$\text{MAAT} = -0.86 \times \text{RAN}_{15} + 11.1$
	Mt. Shennongjia	Wang et al., 2016	0.53*	2.67×10^{-5}	2.63	$\text{MAAT} = -1.60 \times \text{RAN}_{15} + 16.3$
	Mts. Lautaret-Bauges	Vequaud et al., 2021	0	7.31×10^{-1}	None	None
	Global	Present study	0.37*	4.75×10^{-18}	5.46	$\text{MAAT} = -1.55 \times \text{RAN}_{15} + 18.3$

RAN17	Site	Study	R^2	p-value	RMSE	Equation
	Peruvian Andes	Present study	0.81*	1.03×10^{-5}	2.87	$\text{MAAT} = -9.61 \times \text{RAN}_{17} + 29$
	Chilean Andes	Present study	0.04	6.4×10^{-1}	None	None
	Mt. Majella	Huguet et al., 2019	0.27	1.03×10^{-1}	None	None
	Mt. Pollino	Present study	0.56*	3.22×10^{-3}	2.57	$\text{MAAT} = -2.43 \times \text{RAN}_{17} + 18.9$
	Mt. Rungwe	Huguet et al., 2019	0.53*	3.18×10^{-6}	3.19	$\text{MAAT} = -9.86 \times \text{RAN}_{17} + 29.3$
	Mt. Shegyla	Present study	0.54*	8.42×10^{-4}	2.07	$\text{MAAT} = -3.11 \times \text{RAN}_{17} + 13.1$
	Mt. Shennongjia	Wang et al., 2016	0.49*	5.52×10^{-5}	2.71	$\text{MAAT} = -4.68 \times \text{RAN}_{17} + 18$
	Mts. Lautaret-Bauges	Vequaud et al., 2021	0	4.05×10^{-1}	None	None
	Global	Present study	0.41*	4.75×10^{-18}	5.28	$\text{MAAT} = -3.49 \times \text{RAN}_{17} + 18.1$

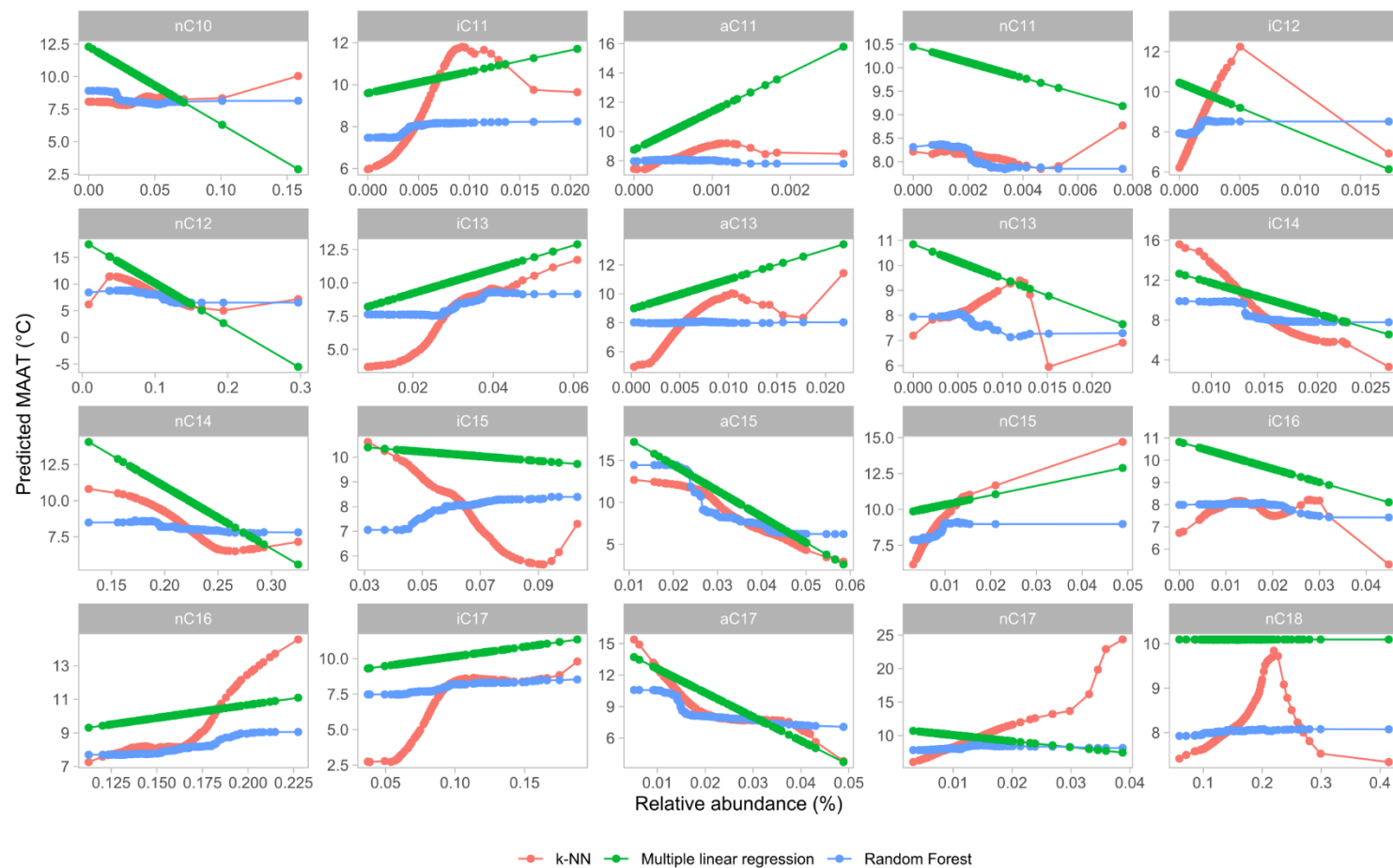
MBT _{5Me}	Site	Study	R^2	p-value	RMSE	Equation
	Peruvian Andes	Present study	0.68*	2.77×10^{-4}	3.74	$\text{MAAT} = 42.5 \times \text{MBT}_{5\text{Me}} - 21.4$
	Chilean Andes	Present study	0.24	2.21×10^{-1}	None	None
	Mt. Majella	Huguet et al., 2019	0.75*	5.47×10^{-4}	2.34	$\text{MAAT} = 59.2 \times \text{MBT}_{5\text{Me}} - 24.6$
	Mt. Pollino	Present study	0.34*	3.50×10^{-2}	3.14	$\text{MAAT} = 35.2 \times \text{MBT}_{5\text{Me}} - 5.89$
	Mt. Shegyla	Present study	0.58*	3.78×10^{-4}	1.97	$\text{MAAT} = 26.9 \times \text{MBT}_{5\text{Me}} - 7.70$
	Mt. Shennongjia	Wang et al., 2016	0.89*	4.51×10^{-13}	1.55	$\text{MAAT} = 34.7 \times \text{MBT}_{5\text{Me}} - 13$
	Mts. Lautaret-Bauges	Vequaud et al., 2021	0.35*	8.19×10^{-6}	3.29	$\text{MAAT} = 24.4 \times \text{MBT}_{5\text{Me}} - 4.78$
	Global	Present study	0.57*	5.86×10^{-27}	3.39	$\text{MAAT} = 24.3 \times \text{MBT}_{5\text{Me}} - 5.03$

CBT'	Site	Study	R^2	p-value	RMSE	Equation
	Peruvian Andes	Present study	0.18	1.33×10^{-1}	None	None
	Chilean Andes	Present study	0.89*	1.01×10^{-2}	0.27	$\text{pH} = 0.74 \times \text{CBT}' + 5.98$
	Mt. Majella	Huguet et al., 2019	0.18	1.92×10^{-1}	None	None
	Mt. Pollino	Present study	0.69*	3.91×10^{-4}	0.37	$\text{pH} = 2.29 \times \text{CBT}' + 5.18$
	Mt. Shegyla	Present study	0.50*	1.42×10^{-3}	0.38	$\text{pH} = 0.61 \times \text{CBT}' + 6.01$
	Mt. Shennongjia	Wang et al., 2016	0.84*	6.98×10^{-11}	0.41	$\text{pH} = 1.91 \times \text{CBT}' + 6.99$
	Mts. Lautaret-Bauges	Vequaud et al., 2021	0.76*	5.71×10^{-16}	0.51	$\text{pH} = 1.72 \times \text{CBT}' + 6.55$
	Global	Present study	0.68*	1.94×10^{-35}	0.71	$\text{pH} = 1.43 \times \text{CBT}' + 6.47$

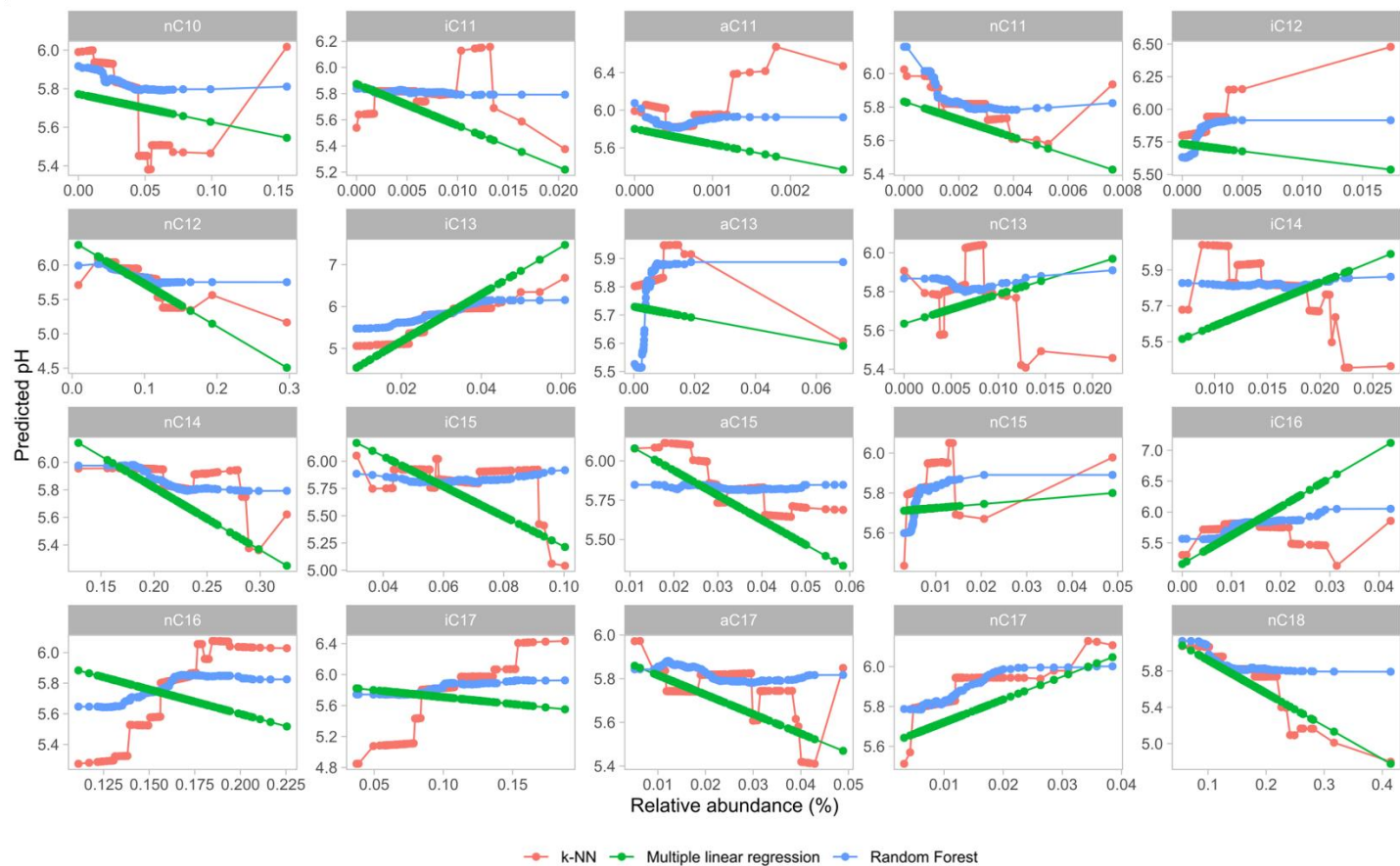
Supplementary Figure 4.1. PCA between pH, MAAT and 3-OH FA-derived proxies (RAN₁₅, RAN₁₇, RIAN) along Mts Rungwe, Pollino, Majella, Shegyla, Shennongjia, Lautaret-Galibier Peruvian Andes and Chilean Andes.



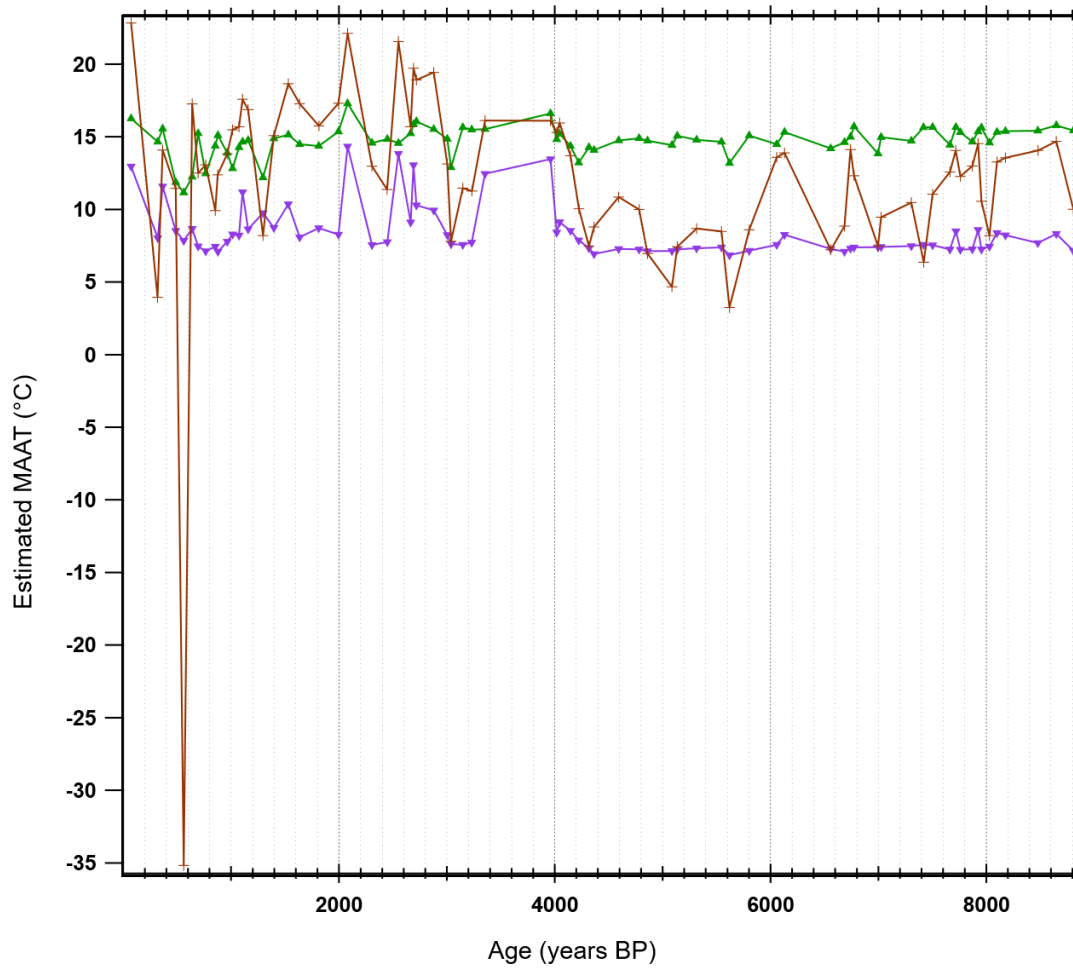
Supplementary Figure 4.2. Partial plots of the individual 3-OH FA variations in multiple regression, random forest and k-NN models proposed to estimate MAAT.



Supplementary Figure 4.3. Partial plots of the individual 3-OH FA variations in multiple regression, random forest and k-NN models proposed to estimate pH.

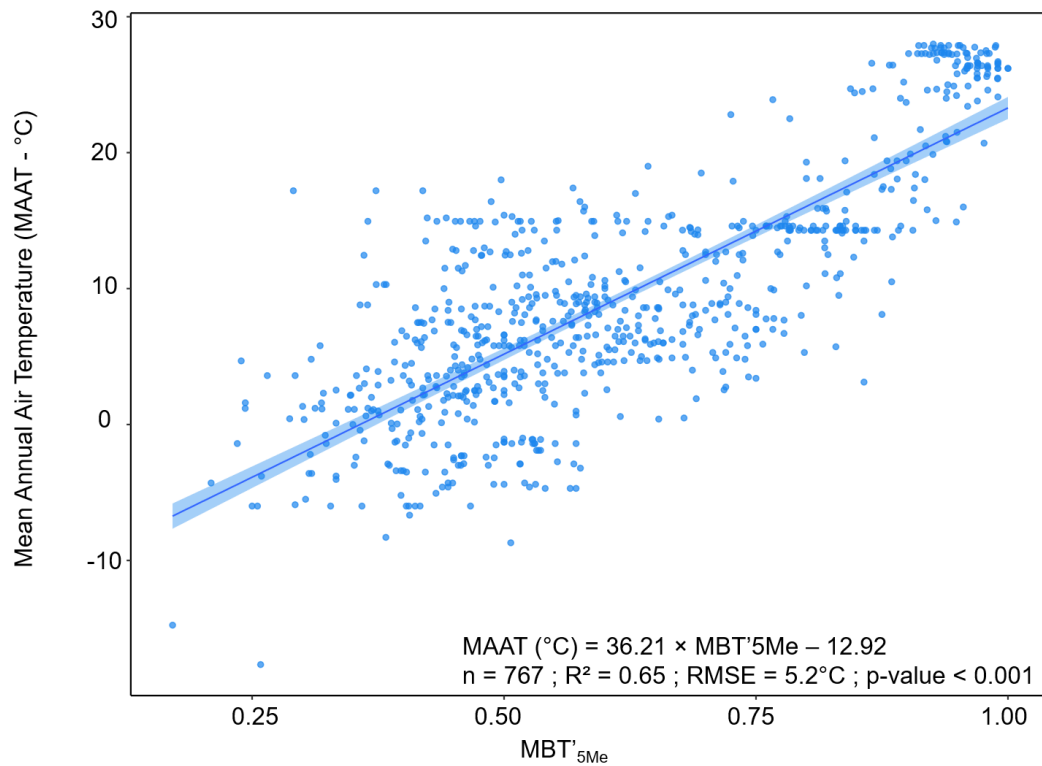


Supplementary Figure 4.4. Application of the three MAAT models proposed in this study and applied to the HS4 speleothem: random forest calibration (green); k-NN calibration (purple); multiple linear regression (brown).

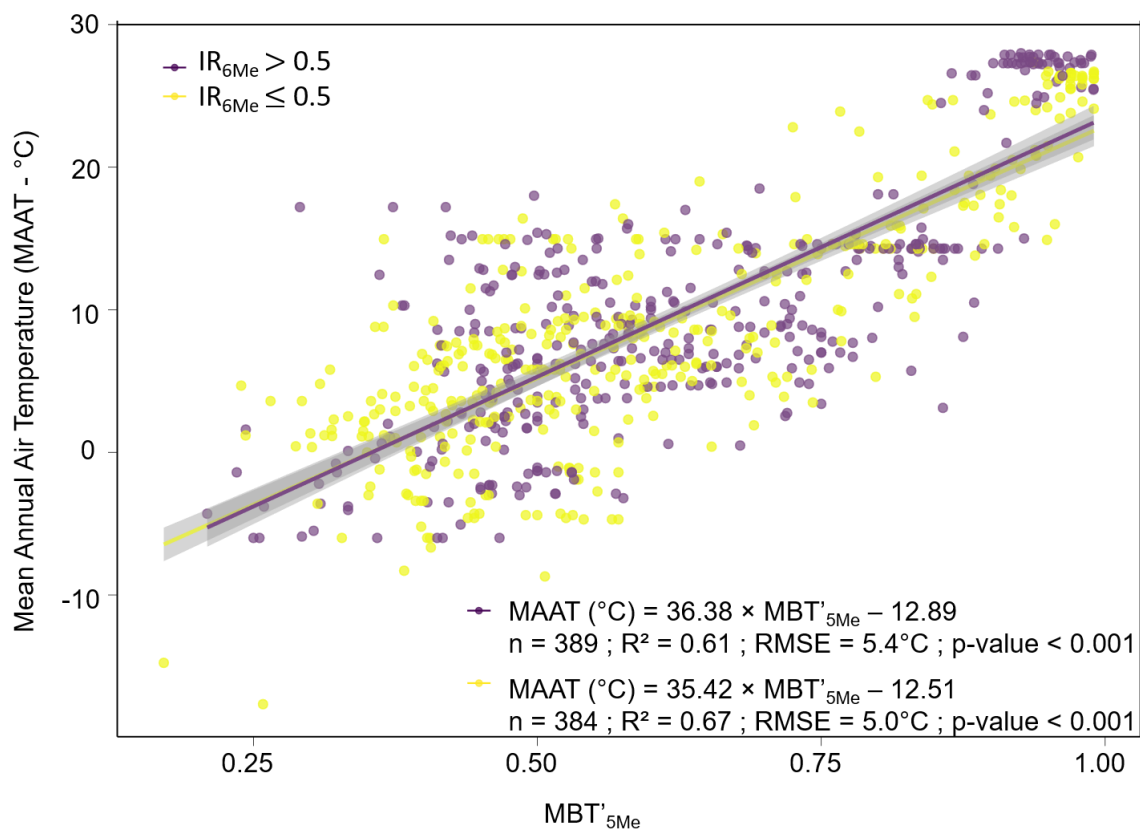


Annexe 3 – Données supplémentaires du Chapitre 5 –

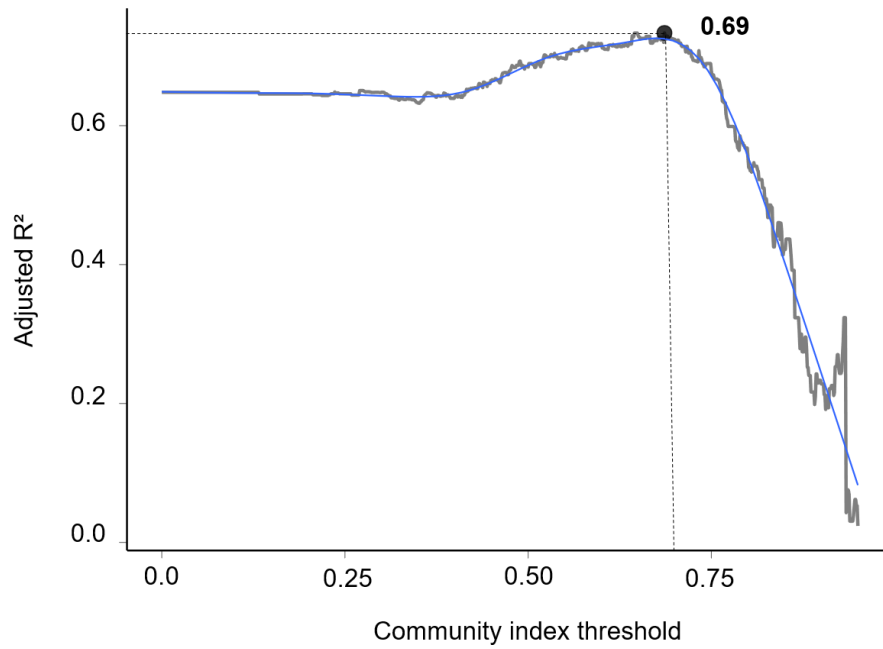
Supplementary Figure 5.1 Linear relationship between MAAT (°C) and MBT'_{5Me} in the global dataset used in this study ($n=775$).



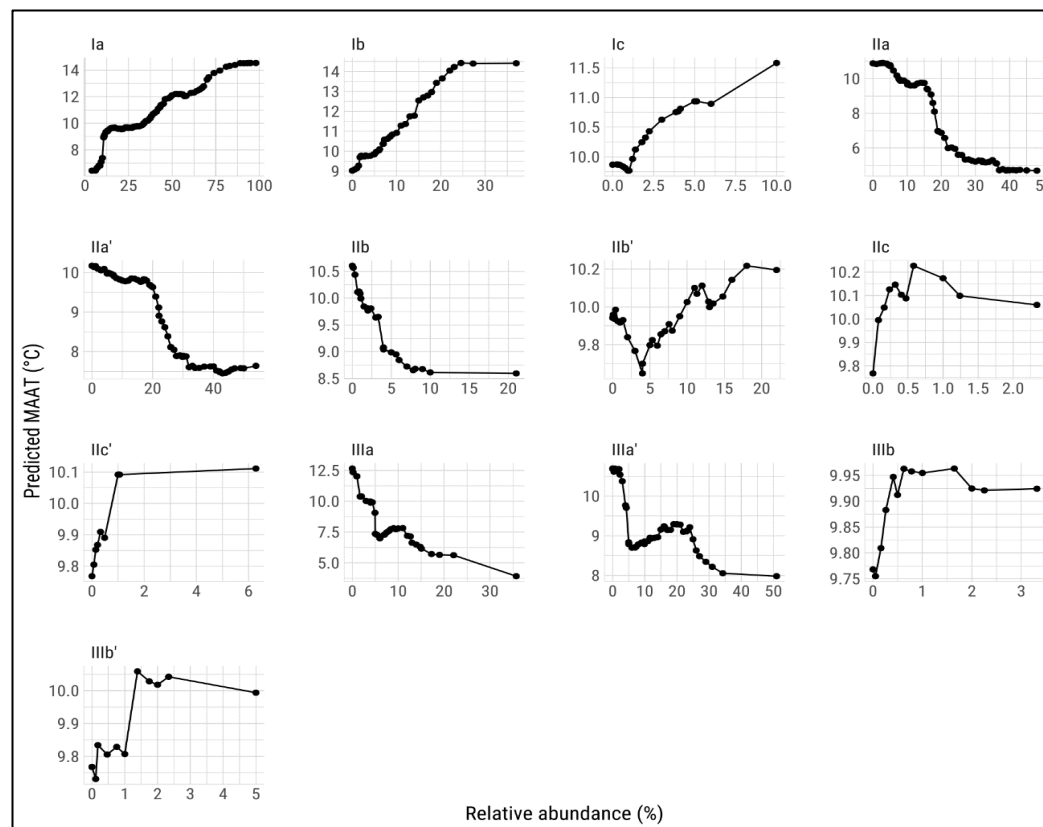
Supplementary Figure 5.2. Comparison of the linear regressions between MAAT (°C) and MBT'_{5Me} in the global dataset for samples with $IR_{6Me} > 0.5$ and $IR_{6Me} \leq 0.5$.



Supplementary Figure 5.3 Estimation of the community index thresholds, for each linear regression between MAAT and MBT'_{5Me}



Supplementary figure 5.4. Partial plots of the individual brGDGT variations in the FROG model proposed to estimate MAAT.



Annexe 4 – Listes des publications et communications en lien avec cette thèse

Publications:

Véquaud P., Derenne S., Anquetil C., Collin S., Poulénard J., Sabatier P. and Huguet A. (2021) **Influence of environmental parameters on the distribution of bacterial lipids in soils from the French Alps: Implications for paleo-reconstructions.** *Organic Geochemistry* **153**, 104194. (Annexe 6)

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 Discussions*, 1–40.

Communications :

Oraux

1. European Geosciences Union (EGU), Conférence virtuelle, mai 2020

Influence of environmental parameters on bacterial lipids in soils from the french alps: implications for paleo-reconstructions. Véquaud P., Derenne S., Collin S., Anquetil C., Poulénard J., Sabatier P., Huguet A.

2. Goldschmidt, Conférence virtuelle, janvier 2020

Influence of environmental parameters on bacterial lipids in soils from the french alps: implications for paleo-reconstructions. Derenne S., Véquaud P., Collin S., Anquetil C., Poulénard J., Sabatier P & Huguet A

3. Journées scientifiques UMR METIS, Paris, France, janvier 2020

Influence des paramètres environnementaux sur les lipides bactériens dans les sols des alpes françaises : implications pour les reconstructions paléoenvironnementales. Véquaud P., Derenne S., Anquetil C., Collin S., Poulénard J., Sabatier P., Huguet A.

4. 7^{ème} International Symposium on Soil Organic Matters (SOM), Adelaïde, Australie, octobre 2019

Bacterial 3-hydroxy fatty acids: applicability as temperature and pH proxies in soils from the french alps. Véquaud P., Derenne S., Collin S., Anquetil C., Poulénard J., Sabatier P., Huguet A.

5. Journée Des Doctorants (JDD2019 GRNE), Paris, France, mars 2019

Applicabilité d'un nouveau marqueur bactérien d'environnement (pH et T°) dans les sols des Alpes Françaises. Véquaud P.

6. 4^{ème} réunion des géochimistes organiciens français (FROG), Lyon, France, juillet 2018

Les acides gras hydroxylés d'origine bactérienne : développement et application de nouveaux marqueurs d'environnement en milieu continental. Véquaud P., Huguet A., Collin S. Anquetil C., Seder-Colomina M., Poulénard J., Sabatier P. FROG

Posters

1. 29^{ème} International Meeting on Organic Geochemistry (IMOG 2019), Goteborg, Suède, septembre 2019

Bacterial 3-hydroxy fatty acids: applicability as temperature and pH proxies in soils from the french alps. Véquaud P., Collin S., Anquetil C., Poulénard J., Sabatier P., Derenne S., Choler P., Huguet A.

2. 26^{ème} Réunion des Sciences de la Terre (RST), Lille, France, octobre 2018

Les acides gras hydroxylés d'origine bactérienne : applicabilité comme marqueurs de température et de pH dans des sols des Alpes françaises. Véquaud P., Collin S., Anquetil C., Seder-Colomina M., Poulénard J., Sabatier P., Derenne S., Huguet A.

3. Journées scientifiques UMR METIS, Paris, France, novembre 2018

Les acides gras hydroxylés d'origine bactérienne : développement et application de nouveaux marqueurs d'environnement en milieu continental. Véquaud P., Collin S., Anquetil C., Seder-Colomina M., Poulénard J., Sabatier P., Derenne S., Huguet A.

Isolement de bactéries Gram-négatif dans la zone atelier Alpes : Seder-Colomina M., Millot-Cornette L., Véquaud P., Kish A., Poulénard J., Sabatier P., Huguet A., Collin S.

4. Journées des Doctorants (JDD2018 GRNE), Paris, France, mars 2018

Développement d'un nouveau marqueur d'environnement en milieu continental. Approche et premiers résultats. Véquaud P.

Annexe 5 – Enseignements

	Unités d'Enseignements (UE)	2017-2018	2018-2019	2019-2020
Licence 1 (Semestre 1)	Géosciences 1 - 1T001 (Sciences de l'atmosphère / Océanographie / Hydrologie / Paléoclimatologie / Paléoenvironnements / Minéralogie / Géomorphologie)	22.5 h TP	42 h TP	45 h TP
Licence 1 (Semestre 2)	Géosciences 2 - 1T002 (Cartographie/ Minéralogie/ Paléontologie)	9 h TD 18 h TP		
Licence 3 Pro	LTP01 Projet Tutoré (Chimie analytique)	15 h TD	18 h TP	26 h TP
	Total (Heure/année)	64.5 h	60 h	71 h