Novel hydrocarbon-utilizing soil mycobacteria synthesize unique mycocerosic acids at a Sicilian everlasting fire

Nadine T. Smit[1], Laura Villlanueva[1], Darci Rush[1], Fausto Grassa[2], Caitlyn R. Witkowski[1], Mira Holzheimer[3], Adriaan J. Minnaard[3], Jaap S. Sinninghe Damsté[1,4], Stefan Schouten[1,4]

[1] NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, P.O. Box 59, 1790 Ab Den Burg Texel, the Netherlands

[2] Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Palermo, Via Ugo La Malfa, 153, 90144, Palermo, Italy

[3] Stratingh Institute for Chemistry, University of Groningen, 9747 AG Groningen, the Netherlands

[4] Department of Earth Sciences, Faculty of Geosciences, Utrecht University, P.O. Box 80.121, 3508 TA Utrecht, the Netherlands

Correspondence to Nadine T. Smit (nadine.smit@nioz.nl)

Abstract. Soil bacteria rank among the most diverse groups of organisms on Earth and actively impact global processes of carbon cycling, especially in the emission of greenhouse gases like methane, CO_2 and higher gaseous hydrocarbons. An abundant group of soil bacteria are the mycobacteria, which colonize various habitats due to their impermeable cell envelope that contains remarkable lipids. These bacteria have been found to be highly abundant at petroleum and gas seep areas, where they might utilize the released hydrocarbons. However, the function and the lipid biomarker inventory of these soil mycobacteria are poorly studied. Here, soils from the Fuoco di Censo seep, an everlasting fire (gas seep) in Sicily, Italy, were investigated for the presence of mycobacteria via 16S rRNA gene sequencing and fatty acid profiling. The soils contained high relative abundances (up to 34 % of reads assigned) of mycobacteria, phylogenetically close to the *Mycobacterium simiae* complex and more distant to the well-studied *M. tuberculosis* and hydrocarbon-utilizing *M. paraffinicum*. The soils showed decreasing abundances of mycocerosic acids (MAs), fatty acids unique for mycobacteria, with increasing distance from the seep. The major MAs at this seep were tentatively identified as 2,4,6,8-tetramethyl tetracosanoic acid and 2,4,6,8,10-pentamethyl hexacosanoic acid. Unusual MAs with mid-chain methyl branches at positions C-12 and C-16 (i.e. 2,12-dimethyl eicosanoic acid and 2,4,6,8,16-pentamethyl tetracosanoic acid) were also present. The molecular structures of the Censo MAs are different from those of the well-studied mycobacteria like *M. tuberculosis* or *M. bovis* and have relatively ^{13}C-depleted values (-38 to -48 ‰), suggesting a direct or indirect utilization of the released seep gases like methane or ethane. The structurally unique MAs in combination with their depleted δ^{13}C values identified at the Censo seep offer a new tool to study the role of soil mycobacteria as hydrocarbon gas consumers in the carbon cycle.

30

1. Introduction

Soils harbor the largest diversity of microorganisms on our planet and have a large influence on the Earth's ecosystem as they actively impact nutrient and carbon cycling, plant production and the emissions of greenhouse gases (Tiedje et al.,
35   1999; Bardgett and van der Putten, 2014; Delgado-Baquerizo et al., 2018). Soil bacteria rank among the most diverse and abundant groups of organisms on Earth. However, numerous studies suggest that most of their function and diversity in our ecosystems is still undescribed (Tiedje et al., 1999; Bardgett and Van Der Putten, 2014). The assessment of soil bacterial diversity has mainly relied on 16S ribosomal RNA (rRNA) gene sequencing and has indicated that the most abundant bacterial phylotypes in global soils include Alphaproteobacteria, Gammaproteobacteria, Betaproteobacteria, Actinobacteria,
40   Acidobacteria and Planctomycetes (Fierer et al., 2012; Delgado-Baquerizo et al., 2018). Besides the use of DNA-based techniques, lipid biomarkers offer an additional tool to investigate soil bacterial communities, such as fatty acids specific for methanotrophs (Bull et al., 2000; Bodelier et al., 2009) or branched glycerol dialky glycerol tetraether (brGDGTs) believed to derive from soil acidobacteria (Weijers et al., 2009; Peterse et al., 2010; Sinninghe Damsté et al., 2018).

Mycobacteria of the genus *Mycobacterium* belonging to the phylum Actinobacteria form an abundant microbial
45   group in global soils (Falkinham, 2015; Walsh et al., 2019). Some members of the genus *Mycobacterium* are obligate pathogens (e.g. *Mycobacterium tuberculosis* and *Mycobacterium leprae*) and are the cause of more than 1.5 million annual human deaths worldwide through the diseases tuberculosis and leprosy (World Health Organization, 2019) and have consequently been more frequently studied than opportunistic pathogenic and non-pathogenic environmental mycobacteria. Interestingly, early studies from the 1950s reported high abundances of non-pathogenic hydrocarbon-consuming
50   mycobacteria (*M. paraffinicum*) in areas of oil and gas production, gas seeps, and in common garden soils (Davis et al., 1956; Dworkin and Foster, 1958; Davis et al., 1959). Cultivation and genomic studies show that mycobacteria can oxidize a range of greenhouse gases (ethane, propane, alkenes, carbon monoxide or hydrogen) and can degrade toxic polycyclic aromatic hydrocarbons (Miller et al., 2007; Hennessee et al., 2009; Coleman et al., 2012; Martin et al., 2014). Mycobacteria are able to colonize a wide variety of habitats from soils to aquatic and human-engineered environments (Brennan and
55   Nikaido, 1995; Falkinham, 2009). Their impermeable cell envelope may play an important role in their ecological dominance. It consists of a peptidoglycan polymer that is surrounded by a thick hydrophobic lipid-rich outer membrane. This impermeability favors the formation of biofilms, thus enabling mycobacteria to be often the first colonizers at environmental interfaces like air-water or surface-water (Brennan and Nikaido, 1995). Additionally, the impermeable cell membrane allows a resistance to acidic conditions, anoxic survival and the possibility to metabolize recalcitrant carbon compounds.
60   Mycobacteria feature an unusual lipid inventory such as extremely long fatty acids with chains up to 90 carbon atoms long and with numerous methyl groups, hydroxylations and/or methoxylations (Minnikin et al., 1985; Minnikin et al., 1993a; Donoghue et al., 2017; Daffé et al., 2019). Furthermore, mycobacteria are known to synthesize fatty acids like

tuberculostearic acid (10-Me $C_{18:0}$) and the multi methyl-branched mycocerosic acids (MAs) that can contain 3 to 5 methyl branches at regularly spaced intervals such as at positions C-2, C-4, C-6 and C-8 (Minnikin et al., 1985; Minnikin et al., 1993a; Minnikin et al., 2002; Redman et al., 2009). These unusual fatty acids are bound to complex glycolipids like phthiocerol dimycocerosates (PDIMs), diacyl trehalose (DATs) or phenolic glycolipids (Minnikin et al., 2002; Jackson et al., 2007). However, in contrast to the pathogenic and opportunistic pathogenic mycobacteria, the lipid biomarker inventory of non-pathogenic mycobacteria in soils and other environments remains poorly described.

In this study, we investigated soils near a continuous gas seep named "Fuoco di Censo" ("Everlasting Fire") in Sicily, Italy to explore the presence of non-pathogenic, potentially hydrocarbon-utilizing, mycobacterial species using 16S rRNA gene amplicon sequencing and fatty acid profiling. It resulted in the identification of potential biomarkers for the presence of mycobacteria in terrestrial environments and hydrocarbon seeps. Furthermore, their stable carbon isotopic composition provided hints for their role in the carbon cycle in this gas seepage environment.

2. Material and Methods

2.2 Study area

The Fuoco di Censo seep (Censo seep; 39º62.503'N, 13º38.751'E) is located at 800 m above sea level in the mountains of Southwestern Sicily, Italy (Etiope et al., 2002; Grassa et al., 2004). The area is part of the Alpine orogenic belt in the Mediterranean and located along the boundary of the African and European plates (Basilone, 2012). The Censo seep belongs to the Bivona area, which is characterized by a complex geological setting. The seep is located in an area with sandy clays, marls and evaporites from the Tortonian-Messinian that are covered by a thrusting limestone of Carnian-Rhetian age (Trincianti et al., 2015). The Censo seep is a typical example of a natural 'Everlasting Fire', which is characterized by the absence of water and the temporal production of flames, that can be several meters high, by a continuous gas flux (Etiope et al., 2002). The Censo seep gas consists mainly of CH_4 (76-86 %) and N_2 (10-17%) as well as some other minor gases like CO_2, O_2, ethane, propane, He and H_2 (Etiope et al., 2002; Grassa et al., 2004). A diffuse soil degassing is detectable within an area of 80 m^2 with an average CH_4 flux of $7*10^6$ mg m^{-2} d^{-1} and a total CH_4 emission of $6.2*10^3$ kg yr $^{-1}$ (Etiope et al., 2002; Etiope et al., 2007). The CH_4 is suggested to be generated by the thermal alteration of organic matter and is characterized by a stable carbon isotopic composition of $\delta^{13}C$ = -35 ‰ and δ^2H = -146 ‰ (Grassa et al., 2004). This thermogenic CH_4 possibly derives from mature marine source rocks (kerogen type II) with a thermal maturity beyond the oil window, resulting in a dry gas with $C_1/(C_2+C_3)$ ratios greater than 100 (Grassa et al., 2004).

2.2 Sample collection

Soil samples of the Censo seep were recovered during a field campaign in October 2017. The soil was collected from a horizon 5 to 10 cm below the surface and at three distances from the seep, i.e. 0 m (seep site), 0.8 m, and a control at 13.2 m

95 distance from the main vent. The *in-situ* temperature of the soils at the time of collection was ca. 18 ºC. The soils were directly transferred into a clean geochemical sampling bag and stored frozen at -20 ºC until freeze drying and extraction.

2.3 Extraction and saponification

Freeze dried Censo soils were extracted with a modified Bligh and Dyer extraction (Schouten et al., 2008; Bale et al., 2013).
100 Soil samples (ca. 12 g) were ultrasonically extracted (10 min) with a solvent mixture containing methanol (MeOH), dichloromethane (DCM) and phosphate buffer (2: 1: 0.8, v: v: v). After centrifugation, the solvent was collected, combined and the residues re-extracted twice. A biphasic separation was achieved by adding additional DCM and phosphate buffer to a ratio of MeOH, DCM and phosphate buffer (1: 1: 0.9, v: v: v). The aqueous layer was washed two more times with DCM and the combined organic layers dried over a Na_2SO_4 column followed by drying under N_2.
105 Saponification (base hydrolysis) was conducted on aliquots (1-7 mg) of the Bligh Dyer extracts (BDEs) by the addition of 2 ml 1N KOH in MeOH solution and refluxing for 1 h at 130 °C. After cooling, the pH was adjusted to 5 by using a 2N HCL in MeOH solution, separated with 2 ml bidistilled water and 2 ml DCM, and the organic bottom layer was collected. The aqueous layer was washed two more times with DCM and the combined organic layers dried over a Na_2SO_4 column followed by drying under N_2.

110

2.4 Derivatization of fatty acids

2.4.1 Preparation of fatty acid methyl esters using BF₃

Aliquots of the saponified Censo seep BDEs and aliquots of a mycocerosic acid standard (2,4,6-trimethyl-tetracosanoic acid; C_{27} MA standard) synthesized by hydrogenation with palladium and charcoal from mycolipenic acid (Holzheimer et al.,
115 2020), were esterified with 0.5 ml of a boron trifluoride-methanol solution (BF₃ solution) for 10 min at 60 °C. After cooling, 0.5 ml bidistilled water and 0.5 ml DCM were added and shaken, and the DCM bottom layer pipetted off. The water layer was extracted twice with DCM and the combined DCM layers were dried over an $MgSO_4$ column. The soil extracts were eluted over a small silica gel column with ethyl acetate as an eluent to remove polar material. Extracts were subsequently separated using a small column packed with activated aluminum oxide into two fractions. The first fraction (fatty acid
120 methyl ester fraction) was eluted with 4 column volumes of DCM followed by a second fraction (polar fraction) eluted with 3 column volumes of DCM/MeOH (1:1). The fatty acid methyl ester fractions were dried under a continuous flow of N_2 and analyzed using gas chromatography-mass spectrometry (GC-MS) and GC-isotope ratio mass spectrometry (IRMS).

2.4.2 Preparation of fatty acid "picolinyl esters" derivatives using 3-pyridylcarbinol

125 Aliquots of saponified Censo seep BDEs, as well as aliquots of the C_{27} MA standard, were derivatized into picolinyl esters. This technique enhances the abundance of diagnostic fragment ions in the mass spectrum, such as those of methyl branching points in fatty acids, enabling an improved structural identification (Christie, 1998; Harvey, 1998). Different 'picolinyl' derivatization protocols were tested on the C_{27} MA standard and the highest yields were achieved by the procedure in Harvey, 1998. In this procedure, 0.5 ml of thionyl chloride was added using a 1 ml disposable syringe to 1 mg aliquot of the

130 dried saponified Censo seep BDEs in a pressure vial and left for ca. 2 min at room temperature. The vials were then dried by a continuous flow of N_2. 0.5 ml of a 1% 3-pyridylcarbinol in acetonitrile solution was added in the reaction vials and left at room temperature for 2 min. The volumes of reagents in this protocol were reduced (0.1 ml) for 0.1 mg of the MA standard. The 'picolinyl' esters were transferred with acetonitrile to 2 ml analysis vials and the concentration was adjusted to 1 mg/ml with acetonitrile. The 'picolinyl' esters were analyzed using GC-MS with acetonitrile as injection solvent.

135

2.4.3 Preparation of fatty acid methyl sulfide esters using dimethyl disulfide (DMDS)

To determine the position of the double bonds in unsaturated fatty acids, dimethyl disulfide (DMDS) derivatization was used (Francis, 1981; Nichols et al., 1986). For this, 100 μl of hexane, 100 μl of DMDS solution (Merck ≥ 99%) and 20 μl of I_2/ether were added to the dry aliquot and heated overnight at 40 ºC. The mixture was left to room temperature and 400 μl of

140 hexane and 200 μl of a 5% aqueous solution of $Na_2S_2O_3$ (for iodine deactivation) were added and mixed. The upper hexane layer was removed, and the aqueous layer washed twice with hexane. The three hexane layers were combined and dried over a Na_2SO_4 column before GC-MS analysis with hexane as injection solvent.

2.5 Instrumental analysis

145 ### 2.5.1 Gas chromatography-mass spectrometry (GC-MS)

GC-MS was performed using an Agilent Technologies GC-MS Triple Quad 7000C in full scan mode. A CP-Sil5 CB column (25 m x 0.32 mm with a film of 0.12 μm, Agilent Technologies) was used for the chromatography with He as carrier gas (constant flow 2 ml min^{-1}). The samples (1 μl) were injected on column at 70 °C, the temperature was increased at 20 °C min^{-1} to 130 °C, raised further by 4 °C min^{-1} to 320 °C, at which it was held for 20 min. The mass spectrometer was operated

150 over a mass range of *m/z* 50 to 850, the gain was set on 3, with a scan time of 700 ms.

2.5.2 Gas chromatography-isotope ratio mass spectrometry (GC-IRMS)

GC-IRMS was carried out with a Thermo Scientific Trace 1310 with a GC-Isolink II, a ConFlo IV and a Delta Advantage IRMS. The gas chromatography was performed on a CP-Sil5 CB column (25 m x 0.32 mm with a film thickness of 0.12 μm,

155 Agilent) with He as carrier gas (constant flow 2 ml min^{-1}). The BF_3 methylated samples (dissolved in ethyl acetate) were on-

column injected at 70 ºC and subsequently, the oven was programmed to 130 ºC at 20 ºC min⁻¹, and then at 4 ºC min⁻¹ to 320 ºC, which was held for 10 min. Stable carbon isotope ratios are reported in delta-notation against VPDB ^{13}C standard. Values were determined by two analysis and results averaged to a mean value.

160 2.6 DNA extraction, 16S rRNA gene amplification, analysis and phylogeny

DNA was extracted from sediments using the PowerMax soil DNA isolation kit (Qiagen). DNA extracts were stored at -80 °C until further analysis. The 16S rRNA gene amplicon sequencing and analysis was performed with the general 16S rRNA archaeal and bacterial primer pair 515F and 806RB targeting the V4 region (Caporaso et al., 2012; Besseling et al., 2018). Polymerase chain reaction (PCR) products were gel purified using the QIAquick Gel-Purification kit (Qiagen), pooled and

165 diluted. Sequencing was performed at the Utrecht Sequencing Facility (Utrecht, the Netherlands) using an Illumina MiSeq sequencing platform. The 16S rRNA gene amplicon sequences were analyzed by an in-house pipeline (Asbun et al., 2019) that includes quality assessment by FastQC (Andrews, 2010), assembly of the paired-end reads with Pear (Zhang et al., 2013), and assignment of taxonomy (including picking representative set of sequences with 'longest' method) with blast by using the Silva 128 release as a reference database (https://www.arb-silva.de/). Representative operational taxonomic unit

170 (OTU) sequences (assigned with OTU picking method based on 97 % nucleotide similarity with Uclust) (Edgar, 2010), attributed to the family Mycobacteriaceae were aligned by using Muscle (Edgar, 2004) implemented in MEGA6, and then used to construct a phylogeny together with 16S rRNA gene sequences of characterized *Mycobacterium* species and closely related uncultured Mycobacteriaceae 16S rRNA gene sequences. The phylogenetic tree was inferred using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The analysis involved 32

175 nucleotide sequences with 294 base pairs positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).

3. Results and discussion

3.1 Microbial diversity in the Censo seep soils

180 Soils were sampled at the Censo seep and with increasing distance from the seep (Table 1). To investigate the microbial diversity, 16S rRNA gene libraries were generated from extracted DNA using 16S rRNA gene amplicon sequencing. This analysis showed a high relative abundance of 16S rRNA gene reads attributed to Mycobacteriaceae ranging from 0.7 to 34.1 % of assigned bacterial plus archaeal reads in the soils with relative abundances increasing with decreasing distance to the seep (Table 1). Sequences assigned to known methanotrophs are Gammaproteobacteria (Methylococcales),

185 Alphaproteobacteria (Methylocystaceae and Methylobacteriaceae) and Verrucomicrobia ('*Candidatus* Methylacidiphilum') but only accounted for 0.2 to 5.1 % of the total number of reads assigned (Table 1). Phylogenetic analysis indicated that there are two sequences representative for operational taxonomic units (OTUs) attributed to mycobacteria (i.e. sequences

Censo seep 1 and Censo seep 2) present in the soils (Fig. 1). Both OTUs are phylogenetically most closely related to sequences of the *Mycobacterium simiae* complex (Tortoli, 2014) (Fig. 2; >98 % identical considering the 294 bp sequence

190   fragment analyzed), which include *M. simiae*, *M. europaeum*, *M. kubicae* and *M. heidelbergense* (Hamieh et al., 2018). Previously described cultivated mycobacteria of the *M. simiae* complex are slow-growing mycobacterium species isolated from environmental niches but also associated to infections in humans as opportunistic pathogens (Lévy-Frébault et al., 1987; Heap, 1989; Bouam et al., 2018). The Censo seep sequences are more distantly related (94-95 % identical) to frequently studied pathogenic mycobacteria (such as *M. tuberculosis* and *M. leprae*) and other environmental mycobacteria

195   like hydrocarbon-utilizers (e.g. *M. paraffinicum* and *M. vanbaalenii*) (Fig. 1). To the best of our knowledge the hydrocarbon-utilizing bacteria have not been isolated from humans or animals (e.g. *M. vanbaalenii*) and are mostly able to degrade aromatic hydrocarbons (Kweon et al., 2015). These data reveal the presence and relevance of uncultured mycobacteria in the soils around the Censo seep. This is in line with previous reports of the occurrence of mycobacteria near petroleum seeps and gas fields (Davis et al., 1956; Davis et al., 1959).

200

3.2 Fatty acid composition of Censo seep soils

Analysis of the fatty acid fractions of the Censo seep soils reveal a distinct pattern that changes with increasing distance from the main seep (Fig. 2). Common fatty acids such as $C_{16:0}$, $C_{16:1\omega6}$, $C_{16:1\omega7}$, C_{18}, $C_{18:1\omega9}$ and $C_{18:1\omega7}$ as well as the longer chain C_{22} and C_{24} fatty acids occur in all three soils. C_{16} and C_{18} fatty acids are abundant lipids in soils and are synthesized by

205   diverse bacteria and fungi, whereas the longer chain (C_{22}-C_{24}) fatty acids originate commonly from higher plants (Řezanka and Sigler, 2009; Frostegård et al., 2011). These fatty acids could also derive from mycobacteria which can produce fatty acids (C_{14} to C_{26}) with high amounts of C_{16} and C_{18} fatty acids and their unsaturated homologues (Chou et al., 1996; Torkko et al., 2003). Besides mycobacteria which are abundant in the soils close to the main seepage (Table 1), the C_{16} fatty acids may also originate from Type I methanotrophs (Gammaproteobacteria), whereas C_{18} fatty acids could derive from type II

210   methanotrophs (Alphaproteobacteria), present in these Censo seep soils (Fig. 2 and Table 1) (Bull et al., 2000; Bowman et al., 1993; Bodelier et al., 2009). Although the relative abundances of 16S rRNA gene reads of this Type I and II methanotrophs are only minor in the Censo soils (Table 1).

The Censo seep soils also feature C_{31}-C_{33} 17β,21β(H)-homohopanoic acids, the most abundant of which is the C_{32} 17β,21β(H)-hopanoic acid (bishomohopanoic acid) (Fig. 2). Hopanoic acids are common components in terrestrial

215   environments (Ourisson et al., 1979; Rohmer et al., 1984; Ries-Kautt and Albrecht, 1989 ;Crossman et al., 2005) and can be derived from a range of bacteria, including Alpha-and Gammaproteobacteria, Planctomycetes and Acidobacteria (Thiel et al., 2003; Sinninghe Damsté et al., 2004; Birgel and Peckmann, 2008; Sinninghe Damsté et al., 2017). Explorative searches of genomic databases for the biosynthetic gene encoding squalene-hopane-cyclase (shc) in mycobacteria from the *M. simiae* complex revealed a potential for biohopanoid production. Therefore, mycobacteria may be an additional source for hopanoic

220   acids in the Censo seep soils.

Interestingly, at the seep (0 m) the FA pattern is dominated by unusual FAs ranging from C_{19} to C_{31}, which are absent further away from the main seepage (Fig. 2). The mass spectra of the three most abundant representatives of these fatty acids are shown in Figure 3. Mass spectra of the methyl ester derivatives of these fatty acids show major fragment ions of m/z 88 and 101. These fragments result from "McLafferty" rearrangements associated with the presence of the carboxylic

225   acid methyl ester group (Lough, 1975; Ran-Ressler et al., 2012). The presence of the even-numbered m/z 88 fragment ion, rather than the typical fragment ion at m/z 74 in the mass spectra of methyl esters of n-FAs, strongly suggests a methyl at position C-2 (Fig. 3). One FA shows also high fragment ions at m/z 213 and m/z 241 (Fig. 3A). This difference of 28 Da hints at a second methyl group at position C-12 (Fig. 3A). Two of these fatty acids show a fragment ion at m/z 129 (Figs. 3C and E), suggesting the presence of an additional methyl at position C-4 of the fatty acids. The apparent methyl branches in

230   these fatty acids are in agreement with the relatively early retention times of these FAs compared to the regular straight-chain counterparts (Fig. 2). Other fragment ions, including those potentially revealing the positions of additional methyl groups, were only present in low abundance, complicating further structural identification. Nevertheless, the presence of methyl branches at C-2 and C-4 in a number of these fatty acids does suggest that they may be related to mycobacteria-derived MAs, which share the same structural characteristics (Alugupalli et al., 1998; Nicoara et al., 2013). Indeed, the mass

235   spectrum of the methyl ester of a synthetic C_{27} MA standard (2,4,6-trimethyl-tetracosanoic acid) (Holzheimer et al., 2020) shows identical mass spectral features (i.e. m/z 88 and 129; Fig. 4A). However, full structural interpretation of the mass spectrum of this authentic standard is also complicated by the low abundances of diagnostic fragment ions indicative for the position of the methyl branches in the alkyl chain.

To enhance the diagnostic fragmentation patterns of these potential MAs, the fatty acids were also transformed into

240   a "picolinyl ester" (Harvey, 1998). The potential of this technique is revealed by the mass spectrum of the synthetic MA standard (2,4,6-trimethyl-tetracosanoic acid) "picolinyl ester" derivative (Fig. 4B), which shows fragment ions revealing all positions of methylation of the fatty acid n-alkyl chain. The high intensity of the fragment ion of m/z 165 indicates the presence of a methyl group at position C-2, while the presence of the fragments ions at m/z 178 and 206 combined with the absence of an m/z 192 fragment ion indicates the presence of a methyl at C-4. Similarly, the presence of the third methyl at

245   position C-6 is revealed by the fragment ions at m/z 220 and 248 and the low abundance of the fragment ion at m/z 234. Thus, the "picolinyl derivatization" technique substantially increases the confidence in the structural identification of multi-methyl-branched fatty acids using mass spectrometry. Therefore, this "picolinyl ester" derivatization technique was also applied to determine the methylation pattern of the potentially novel MAs in the Censo seep soils (Fig. 3).

To illustrate this approach, we discuss the identification of the three major MAs. When analyzed as "picolinyl ester

250   derivatives" (Fig. 3B, D and F), these MAs respectively showed molecular ions (M+) of m/z 431, 515, and 557, indicating C_{22}, C_{28}, and C_{31} MAs. The mass spectrum of the "picolinyl ester derivative" C_{28} MA (Fig. 3D), the most abundant MA in the Censo seep soils, confirms the methylation at C-2 with the fragment ion of m/z 165. Furthermore, this spectrum also shows abundant fragment ions at m/z 178, 206, 220, 248, 262 and 290. Combined with the absence of the fragment ions at m/z 192, 234 and 276, this strongly suggests the presence of three additional methyl groups at position C-4, C-6 and C-8.

255 The mass spectrum of the C_{22} "picolinyl ester derivative" (Fig. 3B) also confirms the methyl branch at position C-2 through the mass ion m/z 165. Elevated fragment ions at m/z 290 and 318 in combination with the low intensity of the fragment ion at m/z 304 suggests a methyl group at position C-12. Further mass spectral interpretations can be made for the C_{31} MA, with mass spectrum similar to that of the C_{28} MA but including an additional methyl group at position C-10, as indicated by the presence of fragment ions at m/z 304 and 332 and the absence of a fragment ion at m/z 318 (Fig. 3F). Thus, we tentatively

260 identified these MAs as 2,12-dimethyl-eicosanoic acid (C_{22} 2,12-dimethyl MA), 2,4,6,8-tetramethyl-tetracosanoic acid (C_{28} 2,4,6,8-tetramethyl MA) and 2,4,6,8,10-pentamethyl-hexacosanoic acid (C_{31} 2,4,6,8,10-pentamethyl MA), respectively (Figs. 2, 3 and Table 2). Other abundant MAs tentatively identified include 2-methyl-octadecanoic acid (C_{19} 2-methyl MA), 2-methyl-nonadecanoic acid (C_{20} 2-methyl MA), 2-methyl-eicosanoic(C_{21} 2-methyl MA), 2,4,6-trimethyl-docosanoic acid (C_{25} 2,4,6-trimethyl MA), 2,4,6,8-tetramethyl-pentacosanoic acid (C_{29} 2,4,6,8-tetramethyl) and 2,4,6,8,16-pentamethyl-

265 tetracosanoic acid (C_{29} 2,4,6,8,16-pentamethyl MA) MAs (Fig. 2 and Table 2).

At the seep (0 m), the MAs have a high relative abundance, representing ca. 44% of the total FAs. Their abundance decreases to ca. 20% in the soil at 1.8 m from the seep, whereas MAs were not detected in the soil at 13.2 m distance from the seep (Fig. 2). These lipids show a similar distribution trend as the 16S rRNA gene sequencing results, which show high relative abundances of sequences from mycobacteria at the seep (ca. 34.1 % at 0 m), decrease to 8.5 % at 1.8 m, and are <1

270 % at 13.2 m (Table 1). Therefore, both the specific structure and the 16S rRNA gene data strongly suggest that the unusual FAs are derived from mycobacteria.

3.3 Mycocerosic acids as biomarkers for mycobacteria in the environment

Mycocerosic acids have been mainly studied as biomarker for diseases from pathogenic (e.g. *M. tuberculosis* or *M. leprae*)

275 or opportunistic pathogenic (e.g. the *M. simiae* complex) mycobacteria in the last decades (e.g. Minnikin et al., 1993a; Minnikin et al., 1993b; Torkko et al., 2003). These studies revealed a high structural variability of MAs with distribution patterns characteristic for different mycobacterial species. For example, the frequently studied *M. tuberculosis* shows a major C_{32} 2,4,6,8-tetramethyl MA and *M. leprae* a C_{34} 2,4,6,8-tetramethyl MA (Minnikin et al., 1993a; Minnikin et al., 1993b), whereas other mycobacteria feature shorter chain major MAs like C_{21} 2-methyl MA in *M. palustre* and C_{22} 2,4-dimethyl MA

280 in *M. intermedium* (Chou et al., 1996; Torkko et al., 2002) (Fig. 2 and Table 2).

The Censo seep soils reveal a high number of tentatively identified MAs which have not been reported previously (Fig. 1 and Table 2), e.g. those biosynthesized by pathogenic mycobacteria like *M. tuberculosis* and *M. leprae* and by mycobacteria belonging to the more closely related *M. simiae* complex like *M. heidelbergense* (Minnikin et al., 1993a; Minnikin et al., 1993b; Torkko et al., 2003). The MA distribution of the Censo seep soils is characterized by a dominant C_{28}

285 2,4,6,8-tetramethyl MA, while the MA distribution of *M. heidelbergense* or *M. palustre* from the *M. simiae* complex is dominated by the C_{21} 2-methyl MA. Other more distantly related environmental opportunistic pathogens besides those of the *M. simiae* complex, like *M. marinum* or *M. intermedium,* produce a dominant 2,4,6- C_{27} trimethyl or C_{22} 2,4-dimethyl MA.

As mentioned earlier, pathogenic mycobacteria like *M. tuberculosis* feature a major C_{32} 2,4,6,8-tetramethyl MA as well as *M. leprae* produces a dominant C_{34} 2,4,6,8-tetramethyl MA, clearly different from the major MA in the Censo soils (Fig. 1 and

290   Table 2). Possibly, these unusual MAs could help to differentiate environmental Censo mycobacteria from opportunistic pathogenic and pathogenic mycobacteria in various modern and past environments.

       Interestingly, the Censo mycobacteria show relatively high abundances of pentamethylated MAs (C_{29} 2,4,6,8,16-pentamethyl MA and C_{31} 2,4,6,8,10-pentamethyl MA) compared to other studied mycobacteria. *M. kansasii* has a dominant pentamethyl MA (C_{33} 2,4,6,8,10-pentamethyl MA; Table 2), which was also been reported in *M. tuberculosis* and *M. leprae*

295   albeit in very low abundances, while *M. botniense* features a partially identified pentamethylated C_{27} (2,4,6,x,x) MA (Minnikin et al., 1985; Daffé and Laneelle, 1988; Torkko et al., 2003). Shorter chain MAs are also abundant in the Censo soils, some of which have been identified in other mycobacterial species (Fig. 1 and Table 2): C_{20} 2-methyl MA (*M. palustre*), C_{21} 2-methyl MA (e.g. *M. palustre*, *M. heidelbergense* or *M. interjectum*) and C_{25} 2,4,6-trimethyl MA (*M. bohemicum*, *M. szulgai* and *M. intermedium*) (Torkko et al., 2001; Torkko et al., 2002; Torkko et al., 2003). The presence of

300   C_{20} 2-methyl and C_{21} 2-methyl MAs in both Censo mycobacteria and mycobacteria from the closely related *M. simiae* complex indicate that these MAs might be a common feature in the *M. simiae* complex. However, these MAs have been also found in more distantly related mycobacterial species like *M. interjectum* and *M. malmoense*, while common pathogenic mycobacteria like *M. tuberculosis* or *M. bovis* do not produce these shorter chain MAs. These pathogenic mycobacteria contain a C_{27} 2,4,6-methyl MA (*M. tuberculosis*) and a C_{26} 2,4-methyl (*M. bovis*) as the shortest chain MAs (Minnikin et al.,

305   1993a; Redman et al., 2009) which are not present in the Censo soils. Some more distantly related mycobacteria can even contain much shorter chain fatty acids like C_{11} 2-methyl MA (*M. interjectum* or *M. intermedium*), C_{15} 2-methyl MA (e.g. *M. kansaii* or *M. intermedium*) or C_{16} 2,4-dimethyl MA (*M. gastrii and M. kansasii*) (Torkko et al., 2003), but these are not found in the Censo MA inventory.

       The most unique feature that distinguishes the MAs of the Censo mycobacteria from cultivated mycobacterial

310   species is the occurrence of methyl groups in the middle of the fatty acid chain at positions C-12 and C-16 in C_{22} 2,12-dimethyl and C_{29} 2,4,6,8,16-pentamethyl MAs, respectively. To the best of our knowledge, this mid-chain methyl branching has only been reported once before, in the mycobacterial species *M. palustre,* also from the *M. simiae* complex (Torkko et al., 2002), which is closely related to the species living in the Censo soil. However, the methyl branching in *M. palustre* is at position C-9 (C_{22} 2,9-dimethyl MA) (Torkko et al., 2002).

315   The fatty acid profile of the Censo soils shows longer chain MAs (e.g. C_{28} 2,4,6,8-tetramethyl and C_{31} 2,4,6,8,10-pentamethyl MAs) which are even more abundant than C_{24} and C_{26} long-chain n-alkyl fatty acids. This feature has not been previously reported in mycobacteria including mycobacteria from the closely-related *M. simiae* complex like *M. heidelbergense* and *M. palustre*, which synthesize much higher amounts of regular fatty acids over MAs (Torkko et al., 2002; Torkko et al., 2003). Some mycobacterial species from the *M. simiae* complex (i.e. *M. lentiflavum, M. florentinum* and

320   *M. genavense*) and other more distantly related mycobacteria (e.g. *M. paraffinicum* and *M. smegmatis*) (Torkko et al., 2002; Torkko et al., 2003; Fernandes and Kolattukudy, 1997; Chou et al., 1998) do not even contain MAs.

In conclusion, the MA patterns in the Censo soil mycobacteria are clearly different from those of previously cultivated mycobacterial species. This could be caused by environmental conditions near the Censo seep, which may have induced adaptions and regulation processes within the biosynthesis systems of MAs in the Censo mycobacteria or may just
325   be a chemotaxonomic feature. Further studies of other soils that contain mycobacteria should reveal how unique the MAs detected in the Censo soils are.

3.4 Role of the mycobacteria at the Censo seep

The high relative abundances of mycobacteria and MAs based on both the relative 16S rRNA gene abundance and FA
330   composition in the soil close to the main Censo seep (Table 1 and Fig. 2), combined with the decrease of these abundances in soils further away from the seep, hint to the potential involvement of mycobacteria in gas oxidation processes at the gas seep system. To further investigate this, the $\delta^{13}C$ values of the MAs, as well as regular fatty acids and hopanoic acids, were analyzed in the Censo seep soils (Fig. 5). The $\delta^{13}C$ values of methane (-30 to -35 ‰) and ethane (-25 ‰) in the released gases at the Censo seep have previously been reported by Grassa et al. (2004).

335   At the seep site, regular and unsaturated C_{16}, C_{18}, C_{22} and C_{24} fatty acids showed no significant depletion in their carbon isotopic composition ($\delta^{13}C$ = -25 to -30 ‰), while at 1.8 m distance these FAs feature a bit more depleted $\delta^{13}C$ values ranging from -33 to -37 ‰ (Fig. 5). As mentioned before, the C_{16} and C_{18} FAs could originate from Type I and Type II methanotrophs (e.g. Bowman, 2006; Dedysh et al., 2007; Bodelier et al., 2009) although larger depletion of ca. 10 to 20 ‰ relative to the methane source is generally expected for fatty acids of aerobic methanotrophs (Jahnke et al., 1999;
340   Blumenberg et al., 2007; Berndmeyer et al., 2013). Thus, other soil microbes using soil organic matter as a carbon source are likely to contribute to the pool of these fatty acids at Censo 0 m and 1.8 m. This agrees with the typical bulk $\delta^{13}C$ values of -25 to -30 ‰ in temperate soils (Balesdent et al., 1987; Huang et al., 1996) and the presence of saturated and unsaturated C_{16} and C_{18} fatty acids even further away from the seep, at 13.2 m (Fig. 2).

The C_{32} 17β,21β(H)-hopanoic acid shows more depleted $\delta^{13}C$ values ranging from -42 to -48 ‰, at Censo 0 m and
345   1.8 m, respectively (Fig. 5), suggesting an origin from bacteria involved in the cycling of a ^{13}C depleted carbon source like methane at this gas seep. The C_{32} hopanoic acid is a diagenetic product of bacteriohopanepolyols (Rohmer et al., 1984; Ries-Kautt and Albrecht, 1989; Farrimond et al., 2002), which could be produced by some of the aerobic methanotrophs (e.g. Methylocystaceae or Methylococcales) (Zundel and Rohmer, 1985; van Winden et al., 2012) identified in the Censo seep soils (Table 1). However, as discussed above, the mycobacteria in the soil, which are closely related to the *M. simiae*
350   complex (Fig. 1), might also be able to synthesize hopanoids and therefore could be contributing to the hopanoid pool.

Depleted $\delta^{13}C$ values are observed in the MAs (-38 to -48 ‰) close to the Censo seep at 0 m and 1.8 m (Fig. 5), indicating that these are likely synthesized by organisms that use a ^{13}C-depleted carbon source rather than soil organic matter. The Censo seep releases high amounts of methane (76-86 % of total released gas) and minor amounts of higher gaseous hydrocarbons (ethane, propane etc.) as well as CO_2 and N_2 (Etiope et al., 2002; Grassa et al., 2004). Thus, it would

355   appear that the Censo mycobacteria are using ^{13}C-depleted methane as their carbon source as it is the major released gas at the Censo seep. This is in agreement with the decreasing relative abundance of mycobacteria and MAs away from the main seepage according to the decrease in the released gas. Furthermore, the $\delta^{13}C$ values of the MAs are more negative than the $\delta^{13}C$ value of the released methane, as expected for methanotrophs, and dissimilar to bulk soil organic matter and the simple FAs likely derived from heterotrophic bacteria.

360   However, these results are not completely in agreement with previous incubation and genetic studies, which showed that mycobacteria are not able to utilize methane but rather use other gaseous hydrocarbons like ethane and propane as well as alkenes, methanol, and carbon monoxide as their carbon source (Park et al., 2003; Coleman et al., 2011; Coleman et al., 2012; Martin et al., 2014). Studies from the 1950s reported high abundances of mycobacteria in soils from areas of oil and gas production and in areas of petroliferous gas seeps, hinting to their potential involvement in gas oxidation processes

365   (Davis et al., 1956; Davis et al., 1959). Cultivation experiments of those soils confirmed that mycobacteria did not utilize methane but higher gaseous hydrocarbons (ethane and propane) (Davis et al., 1956; Dworkin and Foster, 1958). These results suggest that the mycobacteria in the Censo soils are perhaps not using methane, but possibly other gaseous hydrocarbons in the seep, like ethane or propane. However, it should be noted that two previous studies have described mycobacterial species *Mycobacterium flavum var. methanicum*, *Mycobacterium methanicum n. sp.* and *Mycobacterium* ID-Y

370   that were able to oxidize methane (Nechaeva, 1954; Reed and Dugan, 1987). Alternatively, mycobacteria at the Censo seep could act as indirect methane utilizers by using secondary products of methane oxidation performed by other methanotrophs, like methanol. Indeed, some studies have shown that cultured pathogenic mycobacteria were able to utilize methanol (Reed and Dugan, 1987; Park et al., 2003; Park et al., 2010). However, this is difficult to reconcile with their very high abundances (up to 34.1 %) compared to the low abundance of typical methanotrophs like Methylococcales or Methylocystaceae (up to

375   5.1 %) near the seep (Table 1).

          Overall, based on the clear abundance of mycobacterial 16S rRNA sequences in the Censo seep soils, the novel ^{13}C depleted MAs identified here may be useful biomarkers for the presence of hydrocarbon-oxidizing mycobacteria in soils. These unique MAs in combination with ^{13}C depletion could be used to trace mycobacteria in present and past environments, specifically those influenced by hydrocarbon seepage. Longer chain fatty acids and branched fatty acids like MAs have been

380   shown to be more resistant than other biomolecules (e.g. short-chain fatty acids) to diagenetic changes in diverse studies of fossil forests, sediment cores from the Gulf of California, and petroleum systems (Staccioli et al., 2002; Wenger et al., 2002; Camacho-Ibar et al., 2003). Indeed, studies have indicated the presence of MAs of *M. tuberculosis* on ancient bones from a 17,000 year old bison and from a ca. 200 year old human skeletons (Redman et al., 2009; Lee et al., 2012), suggesting a high preservation potential of these lipids.

385   Nevertheless, future research should investigate the presence of MAs in ancient soils, soil cores, and ancient terrestrial hydrocarbon seeps. Additionally, further detailed incubation studies and genomic analysis of the Censo mycobacteria are required to elucidate the exact role of the mycobacteria in gas oxidation processes at the Censo seep.

Biogeosciences
Discussions
Open Access
EGU

4. Conclusion

390   Soils from the Fuoco di Censo Everlasting Fire show high relative abundances (up to 34 %) of uncultivated mycobacterial 16S rRNA gene sequences. These Censo mycobacteria are phylogenetically distant from the typical pathogenic mycobacteria *Mycobacterium tuberculosis* or *M. leprae,* and more closely related to the *M. simiae* complex like *M. heidelbergense* and *M. palustre*. At the main seep, Censo soils feature a unique MA pattern especially in the longer chain MAs. The most abundant MAs were tentatively identified as 2,4,6,8-tetramethyl-tetracosanoic acid (C_{28} 2,4,6,8-tetramethyl

395   MA) and 2,4,6,8,10-pentamethyl-hexacosanoic acid (C_{31} 2,4,6,8,10-pentamethyl MA). The Censo soils also contained MAs with novel mid-chain methyl branching at positions C-12 and C-16 (C_{22} 2,12-dimethyl and C_{29} 2,4,6,8,16-pentamethyl MAs). The MA pattern in the Censo seep soils are clearly different from those reported for the well-studied mycobacteria like *M. tuberculosis* or *M. leprae* and from the closely related *M. simiae* complex. Only C_{20} 2-methyl, C_{21} 2-methyl and C_{25} 2,4,6-trimethyl MAs have been found previously in other mycobacteria from the *M. simiae* complex (e.g. *M. heidelbergense*) and

400   three more distantly related mycobacteria (e.g. *M. interjectum*). These MAs have relatively low δ^{13}C values, suggesting that Censo mycobacteria use a carbon source depleted in ^{13}C, such as methane, higher gaseous hydrocarbons or secondary products of gas oxidation processes, like methanol. The novel identified MAs in the Censo samples offer a new tool, besides DNA-based techniques, to investigate soils from present and past terrestrial environments for the presence of mycobacteria potentially involved in the cycling of gases.

405

Data availability

Data will be made available on request to the corresponding author.

Author contribution

NTS, DR and SS planned research. NTS, FG and CRW collected samples. MH and AJM provided the synthetic C_{27}
410   mycocerosic acid standard. NTS performed lipid analysis. LV analyzed 16S rRNA gene sequencing data. NTS, SS and LV interpreted the data. NTS wrote the paper with input from all authors.

Competing interests

The authors declare that they have no conflict of interest.

415

Acknowledgements

References

Alugupalli, S., Sikka, M. K., Larsson, L., and White, D. C.: Gas chromatography–mass spectrometry methods for the
425 analysis of mycocerosic acids present in Mycobacterium tuberculosis, Journal of Microbiological Methods, 31, 143-150, 1998.

Andrews, S.: FastQC: a quality control tool for high throughput sequence data. 2010.

Asbun, A. A., Besseling, M. A., Balzano, S., van Bleijswijk, J., Witte, H., Villanueva, L., and Engelmann, J. C.: Cascabel: a flexible, scalable and easy-to-use amplicon sequence data analysis pipeline, BioRxiv, 809384, 2019.

430 Bale, N., Villanueva, L., Hopmans, E., Schouten, S., and Sinninghe Damsté, J. S.: Different seasonality of pelagic and benthic Thaumarchaeota in the North Sea, Biogeosciences, 10, 7195-7206, 2013.

Balesdent, J., Mariotti, A., and Guillet, B.: Natural ^{13}C abundance as a tracer for studies of soil organic matter dynamics, Soil Biology and Biochemistry, 19, 25-30, 1987.

Bardgett, R. D., and van der Putten, W. H.: Belowground biodiversity and ecosystem functioning, Nature, 515, 505, 2014.

435 Basilone, L.: Litostratigrafia della Sicilia, Dipartimento di scienze della terra e del mare, Università degli studi, 2012.

Berndmeyer, C., Thiel, V., Schmale, O., and Blumenberg, M.: Biomarkers for aerobic methanotrophy in the water column of the stratified Gotland Deep (Baltic Sea), Organic Geochemistry, 55, 103-111, 2013.

Besseling, M. A., Hopmans, E. C., Boschman, C. R., Sinninghe Damsté, J. S., and Villanueva, L.: Benthic archaea as potential sources of tetraether membrane lipids in sediments across an oxygen minimum zone, Biogeosciences, 15, 4047-
440 4064, 2018.

Birgel, D., and Peckmann, J.: Aerobic methanotrophy at ancient marine methane seeps: a synthesis, Organic Geochemistry, 39, 1659-1667, 2008.

Blumenberg, M., Seifert, R., and Michaelis, W.: Aerobic methanotrophy in the oxic–anoxic transition zone of the Black Sea water column, Organic Geochemistry, 38, 84-91, 2007.

445   Bodelier, P. L., Gillisen, M.-J. B., Hordijk, K., Sinninghe Damsté, J. S., Rijpstra, W. I. C., Geenevasen, J. A., and Dunfield, P. F.: A reanalysis of phospholipid fatty acids as ecological biomarkers for methanotrophic bacteria, The ISME journal, 3, 606, 2009.

Bouam, A., Armstrong, N., Levasseur, A., and Drancourt, M.: Mycobacterium terramassiliense, Mycobacterium rhizamassiliense and Mycobacterium numidiamassiliense sp. nov., three new Mycobacterium simiae complex species
450   cultured from plant roots, Scientific reports, 8, 1-13, 2018.

Bowman, J.: The methanotrophs - the families Methylococcaceae and Methylocystaceae, The Prokaryotes, 5, 266-289, 2006.

Bowman, J. P., Sly, L. I., Nichols, P. D., and Hayward, A.: Revised taxonomy of the methanotrophs: description of Methylobacter gen. nov., emendation of Methylococcus, validation of Methylosinus and Methylocystis species, and a proposal that the family Methylococcaceae includes only the group I methanotrophs, International Journal of Systematic and
455   Evolutionary Microbiology, 43, 735-753, 1993.

Brennan, P. J., and Nikaido, H.: The envelope of mycobacteria, Annual Review of Biochemistry, 64, 29-63, 1995.

Bull, I. D., Parekh, N. R., Hall, G. H., Ineson, P., and Evershed, R. P.: Detection and classification of atmospheric methane oxidizing bacteria in soil, Nature, 405, 175, 2000.

Camacho-Ibar, V. c. F., Aveytua-Alcázar, L., and Carriquiry, J. D.: Fatty acid reactivities in sediment cores from the
460   northern Gulf of California, Organic Geochemistry, 34, 425-439, 2003.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., and Bauer, M.: Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, The ISME journal, 6, 1621, 2012.

Chou, S., Chedore, P., Haddad, A., Paul, N., and Kasatiya, S.: Direct identification of Mycobacterium species in Bactec
465   7H12B medium by gas-liquid chromatography, Journal of Clinical Microbiology, 34, 1317-1320, 1996.

Chou, S., Chedore, P., and Kasatiya, S.: Use of gas chromatographic fatty acid and mycolic acid cleavage product determination to differentiate among Mycobacterium genavense, Mycobacterium fortuitum, Mycobacterium simiae, and Mycobacterium tuberculosis, Journal of Clinical Microbiology, 36, 577-579, 1998.

Christie, W. W.: Gas chromatography-mass spectrometry methods for structural analysis of fatty acids, Lipids, 33, 343-353,
470   1998.

Coleman, N. V., Yau, S., Wilson, N. L., Nolan, L. M., Migocki, M. D., Ly, M. a., Crossett, B., and Holmes, A. J.: Untangling the multiple monooxygenases of Mycobacterium chubuense strain NBB4, a versatile hydrocarbon degrader, Environmental Microbiology Reports, 3, 297-307, 2011.

Coleman, N. V., Le, N. B., Ly, M. A., Ogawa, H. E., Mc Carl, V., Wilson, N. L., and Holmes, A. J.: Hydrocarbon
475 monooxygenase in Mycobacterium: recombinant expression of a member of the ammonia monooxygenase superfamily, The ISME journal, 6, 171, 2012.

Crossman, Z., Ineson, P., and Evershed, R.: The use of 13 C labelling of bacterial lipids in the characterisation of ambient methane-oxidising bacteria in soils, Organic Geochemistry, 36, 769-778, 2005.

Daffé, M., and Laneelle, M.: Distribution of phthiocerol diester, phenolic mycosides and related compounds in
480 mycobacteria, Microbiology, 134, 2049-2055, 1988.

Daffé, M., Quémard, A., and Marrakchi, H.: Mycolic Acids: From Chemistry to Biology, Biogenesis of Fatty Acids, Lipids and Membranes, 181-216, 2019.

Davis, J., Chase, H., and Raymond, R.: Mycobacterium paraffinicum n. sp., a bacterium isolated from soil, Applied Microbiology, 4, 310, 1956.

485 Davis, J., Raymond, R., and Stanley, J.: Areal contrasts in the abundance of hydrocarbon oxidizing microbes in soils, Applied microbiology, 7, 156, 1959.

Dedysh, S. N., Belova, S. E., Bodelier, P. L., Smirnova, K. V., Khmelenina, V. N., Chidthaisong, A., Trotsenko, Y. A., Liesack, W., and Dunfield, P. F.: Methylocystis heyeri sp. nov., a novel type II methanotrophic bacterium possessing 'signature' fatty acids of type I methanotrophs, International Journal of Systematic and Evolutionary Microbiology, 57, 472-
490 479, 2007.

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

Donoghue, H., Taylor, G., Stewart, G., Lee, O., Wu, H., Besra, G., and Minnikin, D.: Positive diagnosis of ancient leprosy and tuberculosis using ancient DNA and lipid biomarkers, Diversity, 9, 46, 2017.

495 Dworkin, M., and Foster, J.: Experiments with some microorganisms which utilize ethane and hydrogen, Journal of Bacteriology, 75, 592, 1958.

Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research, 32, 1792-1797, 2004.

Edgar, R. C.: Search and clustering orders of magnitude faster than BLAST, Bioinformatics, 26, 2460-2461, 2010.

500 Etiope, G., Caracausi, A., Favara, R., Italiano, F., and Baciu, C.: Methane emission from the mud volcanoes of Sicily (Italy), Geophysical Research Letters, 29, 56-51-56-54, 2002.

Etiope, G., Martinelli, G., Caracausi, A., and Italiano, F.: Methane seeps and mud volcanoes in Italy: gas origin, fractionation and emission to the atmosphere, Geophysical Research Letters, 34, 2007.

Falkinham, J. O.: The biology of environmental mycobacteria, Environmental Microbiology Reports, 1, 477-487, 2009.

505 Falkinham, J. O.: Environmental sources of nontuberculous mycobacteria, Clinics in Chest Medicine, 36, 35-41, 2015.

Farrimond, P., Griffiths, T., and Evdokiadis, E.: Hopanoic acids in Mesozoic sedimentary rocks: their origin and relationship with hopanes, Organic Geochemistry, 33, 965-977, 2002.

Fernandes, N. D., and Kolattukudy, P. E.: Methylmalonyl coenzyme A selectivity of cloned and expressed acyltransferase and beta-ketoacyl synthase domains of mycocerosic acid synthase from Mycobacterium bovis BCG, Journal of Bacteriology, 510 179, 7538-7543, 1997.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., Owens, S., Gilbert, J. A., Wall, D. H., and Caporaso, J. G.: Cross-biome metagenomic analyses of soil microbial communities and their functional attributes, Proceedings of the National Academy of Sciences, 109, 21390-21395, 2012.

Francis, G. W.: Alkylthiolation for the determination of double-bond position in unsaturated fatty acid esters, Chemistry and 515 Physics of Lipids, 29, 369-374, 1981.

Frostegård, Å., Tunlid, A., and Bååth, E.: Use and misuse of PLFA measurements in soils, Soil Biology and Biochemistry, 43, 1621-1625, 2011.

Grassa, F., Capasso, G., Favara, R., Inguaggiato, S., Faber, E., and Valenza, M.: Molecular and isotopic composition of free hydrocarbon gases from Sicily, Italy, Geophysical Research Letters, 31, 2004.

520   Hamieh, A., Tayyar, R., Tabaja, H., EL Zein, S., Bou Khalil, P., Kara, N., Kanafani, Z. A., Kanj, N., Bou Akl, I., and Araj, G.: Emergence of Mycobacterium simiae: A retrospective study from a tertiary care center in Lebanon, PloS One, 13, e0195390, 2018.

Harvey, D. J.: Picolinyl esters for the structural determination of fatty acids by GC/MS, Molecular Biotechnology, 10, 251-260, 1998.

525   Heap, B.: Mycobacterium simiae as a cause of intra-abdominal disease: a case report, Tubercle, 70, 217-221, 1989.

Hennessee, C. T., Seo, J.-S., Alvarez, A. M., and Li, Q. X.: Polycyclic aromatic hydrocarbon-degrading species isolated from Hawaiian soils: Mycobacterium crocinum sp. nov., Mycobacterium pallens sp. nov., Mycobacterium rutilum sp. nov., Mycobacterium rufum sp. nov. and Mycobacterium aromaticivorans sp. nov, International Journal of Systematic and Evolutionary Microbiology, 59, 378-387, 2009.

530   Holzheimer, M., Reijneveld, J. F., Ramnarine, A. K., Misiakos, G., Young, D. C., Ishikawa, E., Cheng, T.-Y., Yamasaki, S., Moody, D. B., Van Rhijn, I. and Minnaard A. J.: Asymmetric Total Synthesis of Mycobacterial Diacyl Trehaloses Demonstrates a Role for Lipid Structure in Immunogenicity, ACS Chemical Biology, 2020.

Huang, Y., Bol, R., Harkness, D. D., Ineson, P., and Eglinton, G.: Post-glacial variations in distributions, 13C and 14C contents of aliphatic hydrocarbons and bulk organic matter in three types of British acid upland soils, Organic Geochemistry,
535   24, 273-287, 1996.

Jackson, M., Stadthagen, G., and Gicquel, B.: Long-chain multiple methyl-branched fatty acid-containing lipids of Mycobacterium tuberculosis: biosynthesis, transport, regulation and biological activities, Tuberculosis, 87, 78-86, 2007.

Jahnke, L. L., Summons, R. E., Hope, J. M., and Des Marais, D. J.: Carbon isotopic fractionation in lipids from methanotrophic bacteria II: The effects of physiology and environmental parameters on the biosynthesis and isotopic
540   signatures of biomarkers, Geochimica et Cosmochimica Acta, 63, 79-93, 1999.

Kweon, O., Kim, S.-J., Blom, J., Kim, S.-K., Kim, B.-S., Baek, D.-H., Park, S. I., Sutherland, J. B., and Cerniglia, C. E.: Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus Mycobacterium, BMC Evolutionary Biology, 15, 21, 2015.

Lee, O. Y., Wu, H. H., Donoghue, H. D., Spigelman, M., Greenblatt, C. L., Bull, I. D., Rothschild, B. M., Martin, L. D.,
545   Minnikin, D. E., and Besra, G. S.: Mycobacterium tuberculosis complex lipid virulence factors preserved in the 17,000-year-old skeleton of an extinct bison, Bison antiquus, PloS One, 7, 2012.

Lévy-Frébault, V., Pangon, B., Buré, A., Katlama, C., Marche, C., and David, H.: Mycobacterium simiae and Mycobacterium avium- M. intracellulare mixed infection in acquired immune deficiency syndrome, Journal of Clinical Microbiology, 25, 154-157, 1987.

550 Lough, A.: The chemistry and biochemistry of phytanic, pristanic and related acids, Progress in the Chemistry of Fats and other Lipids, 14, 1-48, 1975.

Martin, K. E., Ozsvar, J., and Coleman, N. V.: SmoXYB1C1Z of Mycobacterium sp. strain NBB4: a soluble methane monooxygenase (sMMO)-like enzyme, active on C2 to C4 alkanes and alkenes, Applied Environmental Microbiology, 80, 5801-5806, 2014.

555 Miller, C., Child, R., Hughes, J., Benscai, M., Der, J., Sims, R., and Anderson, A.: Diversity of soil mycobacterium isolates from three sites that degrade polycyclic aromatic hydrocarbons, Journal of Applied Microbiology, 102, 1612-1624, 2007.

Minnikin, D., Dobson, G., Goodfellow, M., Magnusson, M., and Ridell, M.: Distribution of some mycobacterial waxes based on the phthiocerol family, Microbiology, 131, 1375-1381, 1985.

Minnikin, D., Besra, G., Bolton, R., Datta, A., Mallet, A., Sharif, A., Stanford, J., Ridell, M., and Magnusson, M.:
560 Identification of the leprosy bacillus and related mycobacteria by analysis of mycocerosate profiles, Annales-Societe Belge De Medecine Tropicale, 1993a, 25-25,

Minnikin, D., Bolton, R., Hartmann, S., Besra, G., Jenkins, P., Mallet, A., Wilkins, E., Lawson, A., and Ridell, M.: An integrated procedure for the direct detection of characteristic lipids in tuberculous patients, Annales-Societe Belge De Medecine Tropicale, 1993b, 13-13,

565 Minnikin, D. E., Kremer, L., Dover, L. G., and Besra, G. S.: The methyl-branched fortifications of Mycobacterium tuberculosis, Chemistry & Biology, 9, 545-553, 2002.

Nechaeva, N.: Two species of methane-oxidizing mycobacteria, Associated Technical Services, 1954.

Nei, M., and Kumar, S.: Molecular evolution and phylogenetics, Oxford University Press, 2000.

Nichols, P. D., Guckert, J. B., and White, D. C.: Determination of monosaturated fatty acid double-bond position and
570 geometry for microbial monocultures and complex consortia by capillary GC-MS of their dimethyl disulphide adducts, Journal of Microbiological Methods, 5, 49-55, 1986.
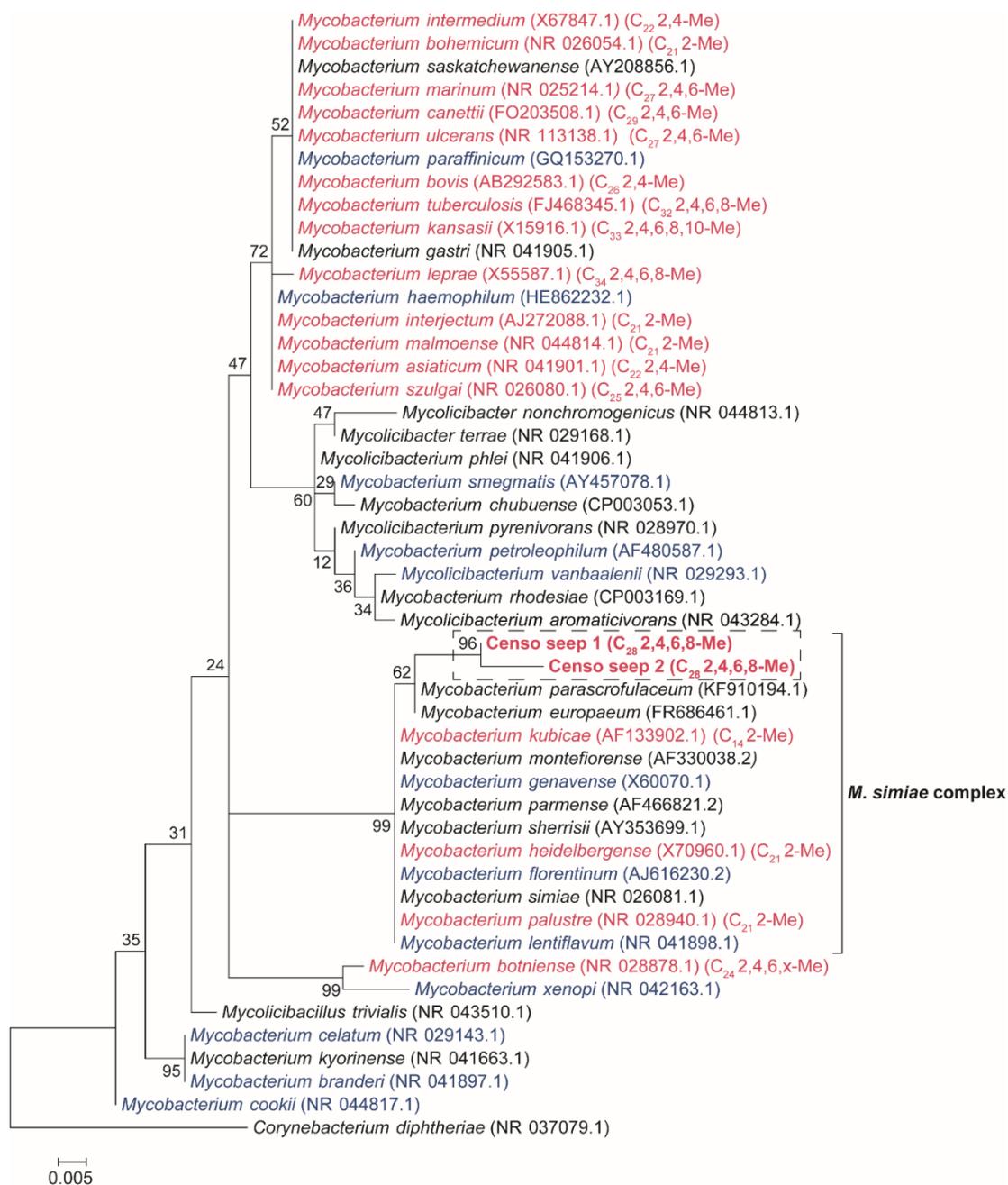
Nicoara, S. C., Minnikin, D. E., Lee, O. C., O'Sullivan, D. M., McNerney, R., Pillinger, C. T., Wright, I. P., and Morgan, G. H.: Development and optimization of a gas chromatography/mass spectrometry method for the analysis of thermochemolytic degradation products of phthiocerol dimycocerosate waxes found in Mycobacterium tuberculosis, Rapid Communications in

575   Mass Spectrometry, 27, 2374-2382, 2013.

Ourisson, G., Albrecht, P., and Rohmer, M.: The hopanoids: palaeochemistry and biochemistry of a group of natural products, Pure and Applied Chemistry, 51, 709-729, 1979.

Park, H., Lee, H., Ro, Y. T., and Kim, Y. M.: Identification and functional characterization of a gene for the methanol: N, N′-dimethyl-4-nitrosoaniline oxidoreductase from Mycobacterium sp. strain JC1 (DSM 3803), Microbiology, 156, 463-471,

580   2010.

Park, S. W., Hwang, E. H., Park, H., Kim, J. A., Heo, J., Lee, K. H., Song, T., Kim, E., Ro, Y. T., and Kim, S. W.: Growth of mycobacteria on carbon monoxide and methanol, Journal of Bacteriology, 185, 142-147, 2003.

Peterse, F., Nicol, G. W., Schouten, S., and Damsté, J. S. S.: 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, 2010.

585   Ran-Ressler, R. R., Lawrence, P., and Brenna, J. T.: Structural characterization of saturated branched chain fatty acid methyl esters by collisional dissociation of molecular ions generated by electron ionization, Journal of Lipid Research, 53, 195-203, 2012.

Redman, J. E., Shaw, M. J., Mallet, A. I., Santos, A. L., Roberts, C. A., Gernaey, A. M., and Minnikin, D. E.: Mycocerosic acid biomarkers for the diagnosis of tuberculosis in the Coimbra Skeletal Collection, Tuberculosis, 89, 267-277, 2009.

590   Reed, W. M., and Dugan, P. R.: Isolation and characterization of the facultative methylotroph Mycobacterium ID-Y, Microbiology, 133, 1389-1395, 1987.

Řezanka, T., and Sigler, K.: Odd-numbered very-long-chain fatty acids from the microbial, animal and plant kingdoms, Progress in Lipid Research, 48, 206-238, 2009.

Ries-Kautt, M., and Albrecht, P.: Hopane-derived triterpenoids in soils, Chemical Geology, 76, 143-151, 1989.

595   Rohmer, M., Bouvier-Nave, P., and Ourisson, G.: Distribution of hopanoid triterpenes in prokaryotes, Microbiology, 130, 1137-1150, 1984.

Schouten, S., Hopmans, E. C., Baas, M., Boumann, H., Standfest, S., Könneke, M., Stahl, D. A., and Sinninghe Damsté, J. S.: Intact membrane lipids of "Candidatus Nitrosopumilus maritimus," a cultivated representative of the cosmopolitan mesophilic group I crenarchaeota, Applied and Environmental Microbiology, 74, 2433-2440, 2008.

600   Sinninghe Damsté, J. S., Rijpstra, W. I. C., Schouten, S., Fuerst, J. A., Jetten, M. S., and Strous, M.: The occurrence of hopanoids in planctomycetes: implications for the sedimentary biomarker record, Organic Geochemistry, 35, 561-566, 2004.

Sinninghe Damsté, J. S., Rijpstra, W. I. C., Dedysh, S. N., Foesel, B. U., and Villanueva, L.: Pheno-and genotyping of hopanoid production in Acidobacteria, Frontiers in Microbiology, 8, 968, 2017.

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.: 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, 2018.

Staccioli, G., McMillan, N., Meli, A., and Bartolini, G.: Chemical characterisation of a 45-million-year bark from Geodetic Hills fossil forest, Axel Heiberg Island, Canada, Wood Science and Technology, 36, 419-427, 2002.

610   Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S.: MEGA6: molecular evolutionary genetics analysis version 6.0, Molecular Biology and Evolution, 30, 2725-2729, 2013.

Thiel, V., Blumenberg, M., Pape, T., Seifert, R., and Michaelis, W.: Unexpected occurrence of hopanoids at gas seeps in the Black Sea, Organic Geochemistry, 34, 81-87, 2003.

Tiedje, J. M., Asuming-Brempong, S., Nüsslein, K., Marsh, T. L., and Flynn, S. J.: Opening the black box of soil microbial diversity, Applied Soil Ecology, 13, 109-122, 1999.

Torkko, P., Suomalainen, S., Iivanainen, E., Suutari, M., Paulin, L., Rudbäck, E., Tortoli, E., Vincent, V., Mattila, R., and Katila, M.-L.: Characterization of Mycobacterium bohemicum isolated from human, veterinary, and environmental sources, Journal of Clinical Microbiology, 39, 207-211, 2001.

Torkko, P., Suomalainen, S., Iivanainen, E., Tortoli, E., Suutari, M., Seppänen, J., Paulin, L., and Katila, M.-L.: Mycobacterium palustre sp. nov., a potentially pathogenic, slowly growing mycobacterium isolated from clinical and veterinary specimens and from Finnish stream waters, International Journal of Systematic and Evolutionary Microbiology, 52, 1519-1525, 2002.

Torkko, P., Katila, M.-L., and Kontro, M.: Gas-chromatographic lipid profiles in identification of currently known slowly growing environmental mycobacteria, Journal of Medical Microbiology, 52, 315-323, 2003.

625 Tortoli, E.: Microbiological features and clinical relevance of new species of the genus Mycobacterium, Clinical Microbiology Reviews, 27, 727-752, 2014.

Trincianti, E., Frixa, A., and Sartorio, D.: Palynology and stratigraphic characterization of subsurface sedimentary successions in the Sicanian and Imerese Domains—Central Western Sicily, Review of Palaeobotany and Palynology, 218, 48-66, 2015.

630 Valero-Guillén, P., Martín-Luengo, F., Larsson, L., Jimenez, J., Juhlin, I., and Portaels, F.: Fatty and mycolic acids of Mycobacterium malmoense, Journal of Clinical Microbiology, 26, 153-154, 1988.

van Winden, J. F., Talbot, H. M., Kip, N., Reichart, G.-J., Pol, A., McNamara, N. P., Jetten, M. S., Op den Camp, H. J., and Sinninghe Damsté, J. S.: Bacteriohopanepolyol signatures as markers for methanotrophic bacteria in peat moss, Geochimica et Cosmochimica Acta, 77, 52-61, 2012.

635 Walsh, C. M., Gebert, M. J., Delgado-Baquerizo, M., Maestre, F., and Fierer, N.: A global survey of mycobacterial diversity in soil, bioRxiv, 562439, 2019.

Weijers, J. W., Panoto, E., van Bleijswijk, J., Schouten, S., Rijpstra, W. I. C., Balk, M., Stams, A. J., and Damste, J. S. S.: Constraints on the biological source (s) of the orphan branched tetraether membrane lipids, Geomicrobiology Journal, 26, 402-414, 2009.

640 Wenger, L. M., Davis, C. L., and Isaksen, G. H.: Multiple controls on petroleum biodegradation and impact on oil quality, SPE Reservoir Evaluation & Engineering, 5, 375-383, 2002.

World Health Organization: World health statistics 2019: monitoring health for the SDGs, sustainable development goals9241565705, 2019.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A.: PEAR: a fast and accurate Illumina Paired-End reAd mergeR, 645 Bioinformatics, 30, 614-620, 2013.

Zundel, M., and Rohmer, M.: Prokaryotic triterpenoids: 1. 3β-Methylhopanoids from Acetobacter species and Methylococcus capsulatus, European Journal of Biochemistry, 150, 23-27, 1985.

Mycobacterium intermedium (X67847.1) (C$_{22}$ 2,4-Me)
Mycobacterium bohemicum (NR 026054.1) (C$_{21}$ 2-Me)
Mycobacterium saskatchewanense (AY208856.1)
Mycobacterium marinum (NR 025214.1) (C$_{27}$ 2,4,6-Me)
Mycobacterium canettii (FO203508.1) (C$_{29}$ 2,4,6-Me)
Mycobacterium ulcerans (NR 113138.1) (C$_{27}$ 2,4,6-Me)
Mycobacterium paraffinicum (GQ153270.1)
Mycobacterium bovis (AB292583.1) (C$_{26}$ 2,4-Me)
Mycobacterium tuberculosis (FJ468345.1) (C$_{32}$ 2,4,6,8-Me)
Mycobacterium kansasii (X15916.1) (C$_{33}$ 2,4,6,8,10-Me)
Mycobacterium gastri (NR 041905.1)
Mycobacterium leprae (X55587.1) (C$_{34}$ 2,4,6,8-Me)
Mycobacterium haemophilum (HE862232.1)
Mycobacterium interjectum (AJ272088.1) (C$_{21}$ 2-Me)
Mycobacterium malmoense (NR 044814.1) (C$_{21}$ 2-Me)
Mycobacterium asiaticum (NR 041901.1) (C$_{22}$ 2,4-Me)
Mycobacterium szulgai (NR 026080.1) (C$_{25}$ 2,4,6-Me)
Mycolicibacter nonchromogenicus (NR 044813.1)
Mycolicibacter terrae (NR 029168.1)
Mycolicibacterium phlei (NR 041906.1)
Mycobacterium smegmatis (AY457078.1)
Mycobacterium chubuense (CP003053.1)
Mycolicibacterium pyrenivorans (NR 028970.1)
Mycobacterium petroleophilum (AF480587.1)
Mycolicibacterium vanbaalenii (NR 029293.1)
Mycobacterium rhodesiae (CP003169.1)
Mycolicibacterium aromaticivorans (NR 043284.1)
Censo seep 1 (C$_{28}$ 2,4,6,8-Me)
Censo seep 2 (C$_{28}$ 2,4,6,8-Me)
Mycobacterium parascrofulaceum (KF910194.1)
Mycobacterium europaeum (FR686461.1)
Mycobacterium kubicae (AF133902.1) (C$_{14}$ 2-Me)
Mycobacterium montefiorense (AF330038.2)
Mycobacterium genavense (X60070.1)
Mycobacterium parmense (AF466821.2)
Mycobacterium sherrisii (AY353699.1)
Mycobacterium heidelbergense (X70960.1) (C$_{21}$ 2-Me)
Mycobacterium florentinum (AJ616230.2)
Mycobacterium simiae (NR 026081.1)
Mycobacterium palustre (NR 028940.1) (C$_{21}$ 2-Me)
Mycobacterium lentiflavum (NR 041898.1)
Mycobacterium botniense (NR 028878.1) (C$_{24}$ 2,4,6,x-Me)
Mycobacterium xenopi (NR 042163.1)
Mycolicibacillus trivialis (NR 043510.1)
Mycobacterium celatum (NR 029143.1)
Mycobacterium kyorinense (NR 041663.1)
Mycobacterium branderi (NR 041897.1)
Mycobacterium cookii (NR 044817.1)
Corynebacterium diphtheriae (NR 037079.1)

M. simiae complex

0.005

650 **Figure 1 Maximum likelihood (ML) phylogenetic tree of the Mycobacterial 16S rRNA gene fragments (i.e. 294 bp; in bold) generated by amplicon sequencing and representative for the two OTUs present in the soils from the Censo seep "everlasting fire". The 16S rRNA gene sequence of *Corynebacterium diphtheriae* was used as an outgroup and other Mycobacterial 16S rRNA gene sequence are plotted for reference. The ML tree is based on the General Time Reversible model with gamma distribution plus invariable sites. Mycobacterial species biosynthesizing MAs are indicated in red font, species not containing MAs are shown in**
655 **blue and species for which MAs have not been analyzed are shown in black. The mycobacterial species producing MAs (in red) are labelled with their dominant MA in brackets (total carbon number, Me= methyl, x = unidentified position of methyl group).**
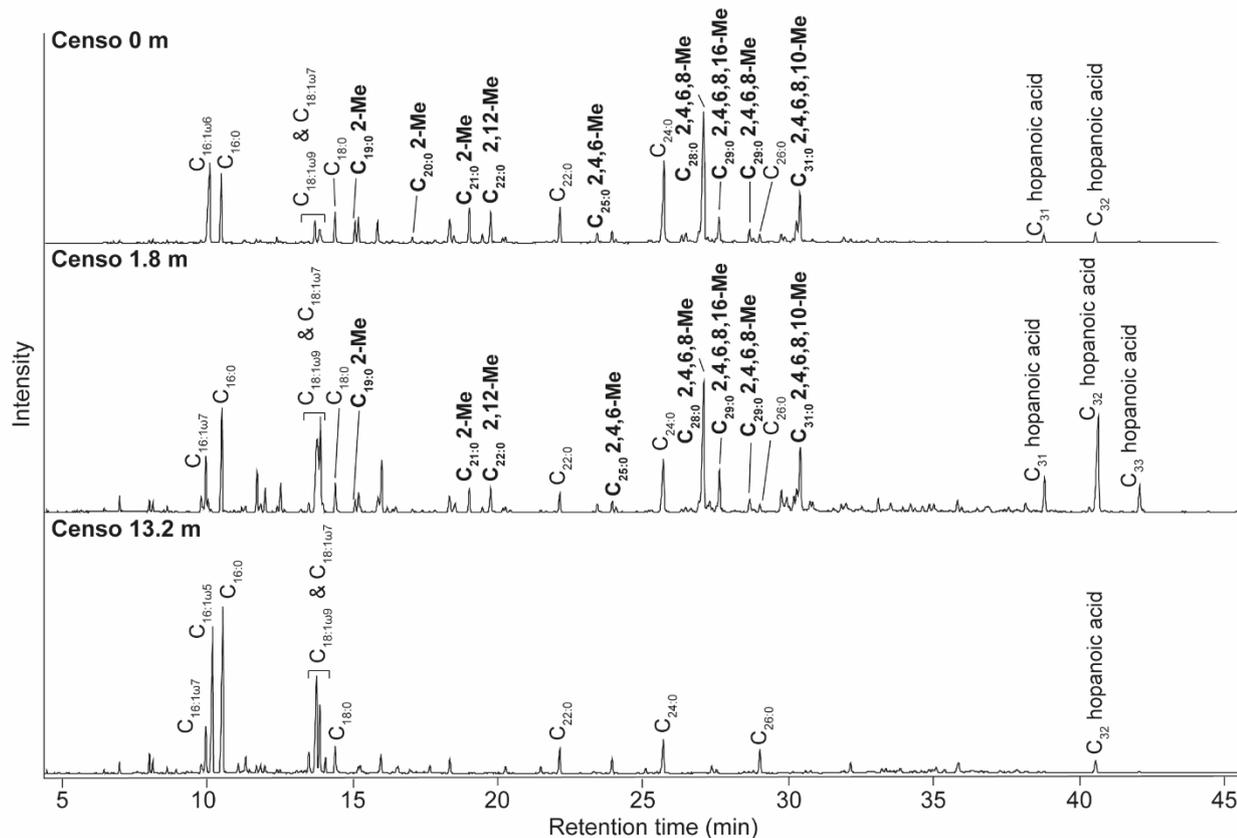
660 **Figure 2 Total ion chromatograms of the saponified and derivatized (BF₃) fatty acid fractions from the Censo seep soils in increasing distance from the main seepage showing the distributions of PLFAs, hopanoic acids, and MAs. The black bold annotations show the tentatively identified MAs and the carbon position of their methyl groups (Me).**
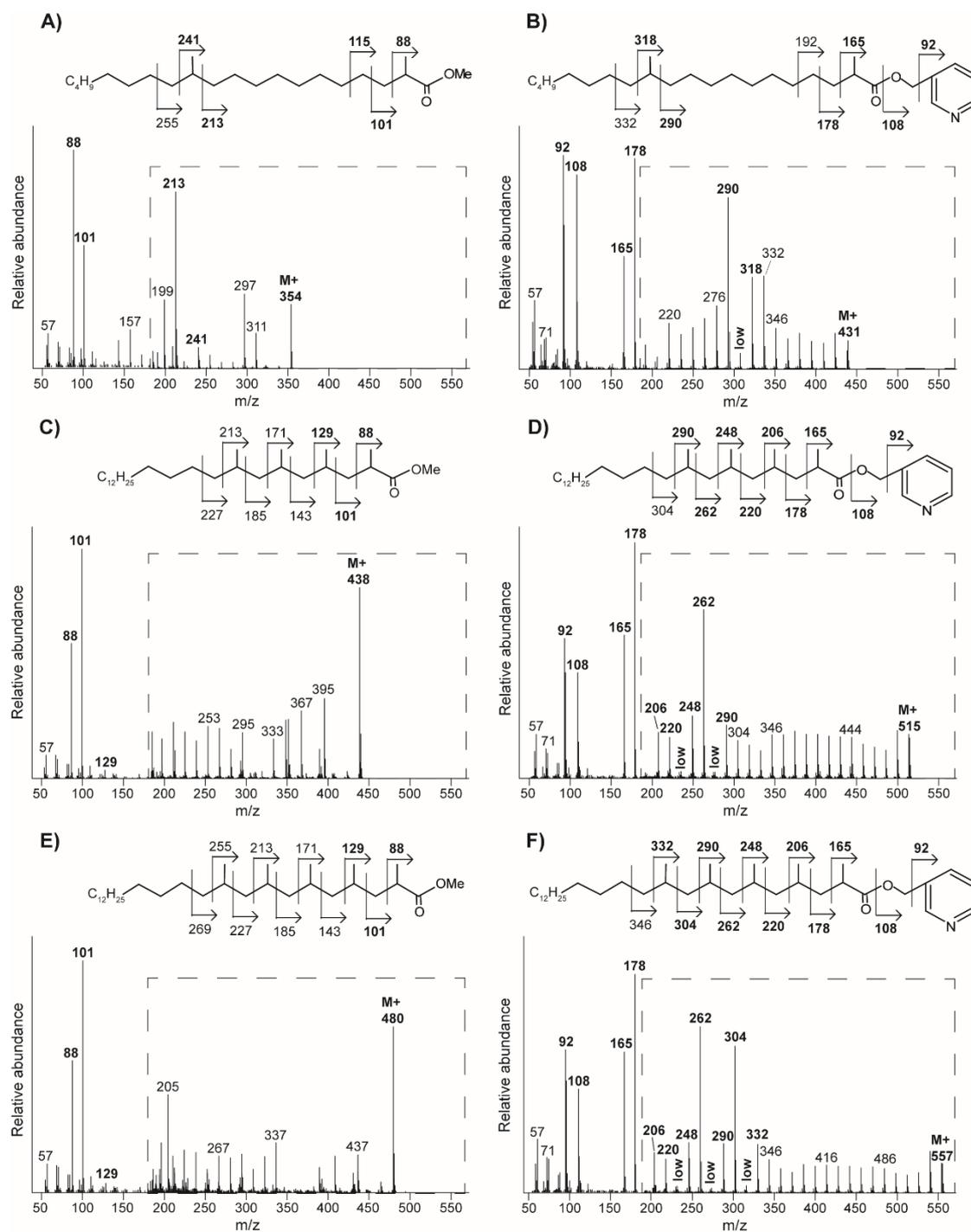
Figure 3 Mass spectra of the methyl ester (left panels) and 'picolinyl'-ester (right panels) derivatized MAs of the Censo 0 m soil sample, with proposed molecular structures and fragmentation patterns. 2,12-dimethyl-eicosanoic acid (C_{22} 2,12-Me MA) (A and B), 2,4,6,8-tetramethyl-tetracosanoic acid (C_{28} 2,4,6,8-Me MA) (C and D), and 2,4,6,8,10-pentamethyl-hexacosanoic acid (C_{31} 2,4,6,8,10-Me MA) (E and F). The dashed boxes show a 10 times exaggerated view into the indicated area of the mass spectrum.
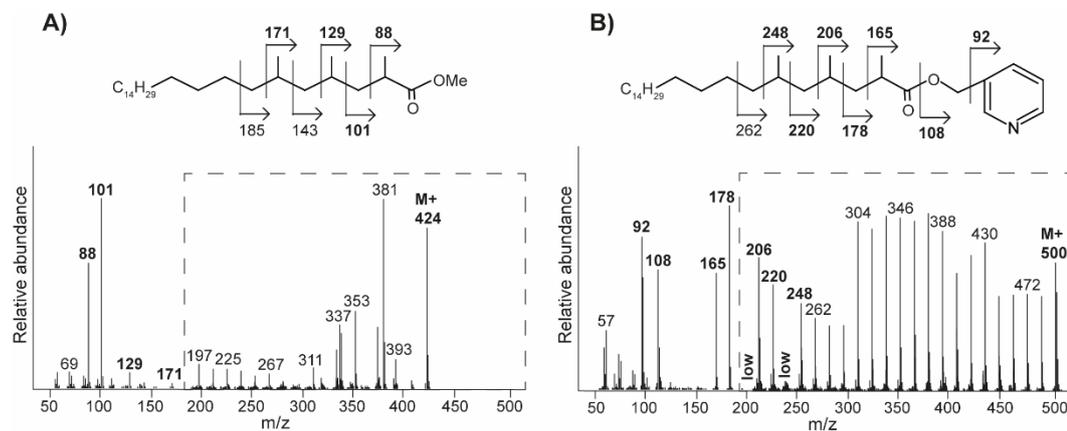
Figure 4 Mass spectra with fragmentation and annotated molecular structures of the A) methyl ester and B) 'picolinyl' ester synthetic 2,4,6-trimethyl-tetracosanoic acid (C_{27} 2,4,6-Me MA standard). The dashed boxes show a 10 times exaggerated view into the indicated area of the mass spectrum.
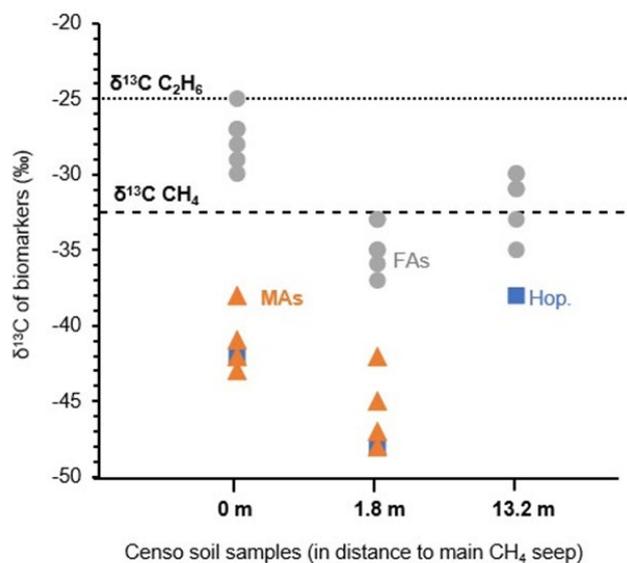
Figure 5 The stable carbon isotopic composition ($\delta^{13}C$) of biomarkers in the Censo soils at increasing distance from the main gas seepage. Biomarkers shown are fatty acids (FAs = grey circle), mycocerosic acids (MAs = orange triangle), and the C_{32} hopanoic acid (Hop. = blue square). Data points represent the mean average of two analysis. The $\delta^{13}C$ values of the released methane ($CH_4 \sim$ -32.5 ‰) and ethane (C_2H_6 = -25 ‰) are indicated by dashed lines in the plot.

Table 1 Distribution of the main microbial groups (relative abundance % of assigned reads) based on 16S rRNA gene amplicon sequencing at three distances from the main gas seep in the Censo soils.

685

	0 m	1.8 m	13.2 m
Archaea; Euryarchaeota	53.5	0.0	0.0
Archaea; Thaumarchaeota	0.0	0.0	2.5
Bacteria; Acidobacteria	0.2	3.0	10.5
Bacteria; Actinobacteria	36.8	23.6	46.0
Acidimicrobiales, other	0.0	0.3	2.3
Corynebacteriales, Mycobacteriaceae, Mycobacterium	**34.1**	**8.5**	**0.7**
Frankiales, Geodermatophilaceae, Geodermatophilus	0.0	5.1	0.1
Micrococcales, Microbacteriaceae, Humibacter	0.1	3.0	0.0
Micromonosporales, Micromonosporaceae, Micromonospora	1.0	0.0	0.1
Pseudonocardiales, Pseudonocardiaceae, Pseudonocardia	0.0	0.0	1.6
Rubrobacterales, Rubrobacteriaceae, Rubrobacter	0.0	0.0	6.2
Gaiellales, Gaiellaceae, Gaiella	0.0	0.0	5.1
Solirubrobacterales, 288-2, uncultured bacterium	0.0	0.0	2.2
Solirubrobacterales, Elev-16S-1332, uncultured bacterium	0.0	0.6	4.3
Solirubrobacterales, Solirubrobacteraceae, Solirubrobacter	0.0	0.0	1.8
others	1.6	6.0	21.5
Bacteria; Armatimonadetes	0.0	0.0	0.2
Bacteria; Bacteroidetes	0.3	1.2	2.1
Bacteria; Chloroflexi	0.2	30.5	10.5
Anaerolineales, Anaerolineae, Anaerolineales	0.0	7.9	0.2
Ktedonobacteria	0.1	17.8	0.1
Bacteria; Firmicutes	5.2	15.7	0.6
Bacteria; Gemmatimonadetes	0.0	0.0	3.1
Bacteria; Latescibacteria	0.0	0.0	0.2
Bacteria; Nitrospirae	0.0	0.0	0.2
Bacteria; Planctomycetes	0.2	1.5	3.7
Bacteria; Proteobacteria	2.9	21.1	17.2
Alphaproteobacteria, Rhizobiales, Methylobacteriaceae	0.0	0.3	0.8
Alphaproteobacteria, Rhizobiales, Methylocystaceae	0.0	4.2	0.0

27

	0.2	5.1	0.2
Gammaproteobacteria, Methylococcales	0.2	5.1	0.2
Bacteria; Saccharibacteria	0.1	2.1	0.2
Bacteria; Tectomicrobia	0.0	0.0	0.7
Bacteria; Verrucomicrobia	0.3	0.3	1.7
Verrucomicrobia Incertae Sedis, Candidatus Methylacidiphilum	0.2	0.0	0.0
Number of reads assigned	**140,206**	**63,916**	**259,714**

690

Table 2 Chemical variability and occurrence of MAs in the Censo seep soils and in the most relevant mycobacterial species. The underlined names of the mycobacterial species indicate the major MA configuration in the mycobacterial species. The x in the position of methylations in the *n*-alkyl chain features an unidentified position of the methyl group. MAs indicated in bold typeface are MAs identified in the Censo seep soils

Chemical structure of MAs		Occurrence
Length of *n*-alkyl chain	Position of methyl group(s)	
C_{16}	2,4	*M. marinum* [1]
C_{18}	**2**	**Censo**
C_{19}	2	**Censo,** *M. palustre* [2]
C_{20}	**2**	**Censo**, *M. bohemicum*, *M. heidelbergense*, *M. malmoense*, *M. interjectum*, *M. palustre* [1,2,3,4,5]
C_{20}	2,4	*M. asiaticum*, *M. szulgai*, *M. intermedium*, *M. heidelbergense*, *M. malmoense* [1,2,4,5]
C_{20}	2,9	*M. palustre* [2]
C_{20}	**2,12**	**Censo**
C_{20}	2,4,6,x	*M. botniense* [4]
C_{22}	**2,4,6**	**Censo**, *M. bohemicum*, *M. szulgai*, *M. intermedium* [1,2,4]
C_{22}	2,4,6,x,x	*M. botniense* [4]
C_{24}	2,4	*M. bovis* [6,7]
C_{24}	2,4,6	*M. tuberculosis*, *M. bovis*, *M. kansasii*, *M. marinum*, *M. ulcerans*, *M. bohemicum*, *M. heidelbergense*, *M. malmoense*, *M. interjectum* [2,3,4,5,6,7,8]

C_{24}	**2,4,6,8**	<u>**Censo**</u>
C_{26}	2,4,6	*M. tuberculosis, M. leprae*, M. bovis, *M. kansasii, M. marinum, M. ulcerans* [6,7]
C_{24}	**2,4,6,8,16**	**Censo**
C_{25}	**2,4,6,8**	**Censo**
C_{26}	2,4,6,8	*M. tuberculosis, M.* leprae, *M. bovis, M. kansasii, M. marinum, M. ulcerans* [6,7,8,9,10]
C_{26}	**2,4,6,8,10**	**Censo**
C_{28}	2,4,6,8	<u>*M. tuberculosis*</u>*, M. leprae, M. bovis, M. kansasii, M. marinum* [6,7,8,11]
C_{28}	2,4,6,8,10	*M. tuberculosis, M. leprae,* <u>*M. kansasii*</u> [6,7]
C_{30}	2,4,6,8	<u>*M. leprae*</u> [6]

[1]Chou et al. (1996); [2]Torkko et al. (2002), [3]Torkko et al. (2001), [4]Torkko et al. (2003), [5]Valero-Guillén et al. (1988), [6]Minnikin et al. (1993a), [7]Minnikin et al. (1985), [8]Daffé and Laneelle (1988), [9]Donoghue et al. (2017), [10]Redman et al. (2009), [11]Minnikin et al. (2002).