



Organic carbon compounds associated with deep soil carbon stores

Podjane Sangmanee · Bernard Dell ·
Richard J. Harper · Suman George ·
David J. Henry

Received: 30 December 2021 / Accepted: 24 July 2022 / Published online: 5 August 2022
© The Author(s) 2022

Abstract

Aims Organic carbon has been reported in deep regolithic profiles to depths of tens of metres, but the composition of the carbon compounds is unknown.

Methods Residual carbon in the form of non-volatile low molecular weight compounds (LMWC) was characterised in three deep soil profiles to a depth of 19 m under farmland in south-western Australia following extraction with ethyl acetate and analysis by GC/MS. Pyrolysis and off-line thermochemolysis were used to characterise macromolecular organic carbon (MOC) to a depth of 29 m at a fourth site.

Results Three compound classes occurred across the three different field locations: (1) terpenes, (2)

fatty acids, amides and alcohols, and (3) plant steroids; indicating the influence of input of the past and present vegetation. Compounds related to fatty acids were the predominant residual carbon species in deep soils, and may be derived from plants and microorganisms. Biomarkers such as lignin, polysaccharides, proteins and terpenes at 0–0.1 m implied influences of vegetation, fire events and microorganisms. Pyrolysis found that polysaccharides were distributed mainly from 0 to 0.1 m, while aromatic compounds were consistently detected down to 29 m.

Conclusions Carbon was stabilised in the form of aromatic compounds in deep soil, whereas other carbon sources such as cellulose, chitin, and N-containing compounds were confined to the surface soil. LMWC (*Z*)-docos-13-enamide and bis(6-methylheptyl) phthalate, were the main components throughout the soil profiles representing 53–81% of the LMWC, and were a greater proportion of the organic matter at depths of 18–19 m.

Responsible Editor: Wen-Feng Cong.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11104-022-05627-7>.

P. Sangmanee
Faculty of Agriculture, Uttaradit Rajabhat University,
53000 Uttaradit, Thailand

P. Sangmanee · B. Dell · R. J. Harper (✉) · S. George ·
D. J. Henry
Food Futures Institute, Murdoch University, South Street,
6150 Murdoch, WA, Australia
e-mail: r.harper@murdoch.edu.au

S. George
School of Earth and Environment, The University
of Western Australia, 6907 Crawley, WA, Australia

Keywords Deep soil carbon · Low molecular weight compounds · Macromolecular organic carbon · Thermochemolysis · Pyrolysis

Introduction

Soil is a component of the terrestrial system that is recognised for its vast reservoir of carbon, however, only the carbon in the upper soil horizons is used to

estimate the global soil organic carbon (SOC) stock (Batjes 1996). There are many deep soils in the world, and organic carbon has been reported in deep regolithic profiles to depths of tens of metres (Harper and Tibbett 2013; Jackson et al. 2002; Moreland et al. 2021). These stocks of carbon can be substantial; in Harper and Tibbett's (2013) study carbon storage to bedrock (mean 21 m, range 5 to 38 m depth) was 2–5 times of that reported from the surface 0.5 m.

The composition of the carbon compounds that occur in the deep profiles is unknown. The study of low molecular weight organic carbon compounds (LMWC) will provide an understanding of the sources of soil organic matter, microbial activity and the pathways of degradation and stabilisation of soil organic matter (Bull et al. 2000; Feng and Simpson 2007). This will be of particular importance where this carbon occurs within deep soils, because at present little is known about the distribution or dynamics of this material, and particularly in response to global change.

For instance, components such as aliphatic lipids, steroids, terpenoids, glycerols and carbohydrates (mono- and disaccharides), extracted sequentially by a series of dichloromethane and methanol solvent mixtures, were used to study soil organic matter biomarkers such as plant waxes, bacteria and fungi (Otto and Simpson 2007). In strongly podsolized laterites, an abundance of free lipids has been interpreted to indicate the conditions of the area such as combinations of acidity and waterlogging, limitation of complex elements (Fe, Al and Si) and decreases in microbial activity (Bardy et al. 2008).

Meanwhile, macromolecular organic carbon (MOC) is usually composed of large-non-volatile organic substrates that are difficult to analyse directly using conventional GC/MS. However, pyrolysis and thermochemolysis have been developed for characterisation of large organic substrates because they lead to extensive fragmentation, which enables easy separation and identification of fragments by GC/MS (Kaal et al. 2009; Klingberg et al. 2005; Shadkani and Helleur 2010). Pyrolysis involves degradation of MOC by elevated temperatures in the absence of oxygen. Signature fragments of MOC are produced during pyrolysis that identifies the form(s) of MOC present. Several carbon materials have been qualified by this technique such as polysaccharides, humic acid, degraded wood or charred materials (Buurman et al. 2007, 2009; del Río et al. 2001; Kaal et al. 2009; Kaal and Rumpel 2009).

Furthermore, thermochemolysis involves assisted fragmentation with an alkylating reagent at elevated temperatures. TMAH is a common reagent used to assist the thermochemolysis reactions. This procedure requires lower temperatures than pyrolysis to degrade MOC because the TMAH cleaves the MOC selectively at ester and ether bonds (Clifford et al. 1995; Shadkani and Helleur 2010) and the fragment compounds are easily characterised by GC/MS. Again, the carbon fragments identified by this method can be used to imply the presence of compounds such as lignin, cellulose and chitin, including proteinaceous materials.

Previous work has focused on samples with relatively large amounts of organic carbon, mostly in near surface soils, and this carbon has been well characterised (e.g. Kleber and Johnson 2010; Kögel-Knabner and Rumpel 2018). Whereas there is increasing recognition of the occurrence of deep soil carbon stores to depth of tens of metres (Harper and Tibbett 2013; Jackson et al. 1996) the LMWC and MOC in this organic matter have not been characterised and quantified. The main objectives of this paper are thus to (i) identify and quantify the LMWC in deep soil profiles from three field locations in south-western Australia where soil organic carbon stores have been previously reported by Harper and Tibbett (2013), and (ii) examine the utility of the thermochemolysis and pyrolysis techniques for characterisation of MOC of deep soil carbon.

Materials and methods

Identification and quantification of non-volatile low molecular weight compounds (experiment 1)

Samples

Samples from three field sites located in south-western Australia and previously reported by Harper and Tibbett (2013) were analysed namely GL03 (-34° 21' 56", 117° 26' 31"), PT06 (-31° 57' 19", 117° 5' 25") and ST04 (-33° 36' 23", 116° 39' 30"). Carbon concentrations of each profile at depths of 0.5–1, 10–11 and 19–20 m were 0.55, 0.05, 0.04% total organic carbon (TOC) (GL03); 0.17, 0.04, 0.03% TOC (PT06) and 0.16, 0.04, 0.04% TOC (ST04). Soil pH values (0.01 M CaCl₂) for these samples were 5.6, 6.7, 6.5 (GL03), 4.8, 4.5, 5.8 (PT06) and 5.9, 7.4, 5.9 (ST04). These profiles have sandy surface horizons,

with lower horizons to bedrock having sandy clay to medium clay textures.

Briefly, the samples were taken from farmland in south-western Australia where deep weathering profiles have formed on the Archean granites and gneisses of the Yilgarn Craton (Gilkes et al. 1973). One profile from each location was selected at random. Vegetation prior to agriculture was deep rooted native plants which had been removed 50–80 years previously and replaced by shallow-rooted annual crops such as cereals or annual pastures. In the year prior to sampling, *Pinus pinaster* plantations had been established (Harper and Tibbett 2013) with these still containing the pasture plants within their inter-row areas.

Extraction of low molecular weight compounds

Samples were weighed into Erlenmeyer flasks, ethyl acetate was added and the flasks closed with glass stoppers. The extraction was carried out for 2 h using a shaker. The samples were then filtered through glass fibre filter (Whatman GF/A) and the solutions transferred to 100 ml round bottom flasks (or 50 ml size in case of surface soil). The solution was concentrated to approximately 1 ml using a rotary evaporator, with water bath temperatures of 60–65 °C. Then, the solution was transferred to a 5 ml round bottom flask and concentrated until dried. In the final step, the sample was redissolved with 1 ml of internal standard, n-decane in ethyl acetate, before analysis by GC/MS and identification of compounds. The samples were prepared in duplicate.

Quality control and reproducibility of procedures

A soil reference sample was prepared by mixing 100 g of sample CU03 at depth increment of 12–13 (0.02% TOC) and 27–28 m (0.01% TOC). The extraction procedure efficiency was tested by measuring the recovery of a spiked commercial standard, (Z)-docos-13-enamide from this sample. An average recovery of 95% (SD=7, n=3) was obtained. Furthermore, this reference sample was re-analysed at the same time as the field samples in order to monitor the quality and reproducibility of procedure.

Analytical gas chromatography/mass spectrometry

Samples were analyzed by capillary GC on a Shimadzu GC 2010. The column used for GC separation

was 30 m in length, 0.25 µm in diameter and 0.25 µm thick (SGE BPX5) and helium was used as the carrier gas. The column oven temperature was set to 60 °C and injection temperature was 310 °C using splitless injection mode. The column was heated to 60 °C for 1 min then ramped at 30 °C min⁻¹ to 100 °C, held at this temperature for 1 min before heating at a rate of 4 °C min⁻¹ to a final temperature of 300 °C then held at this temperature for 5 min. Compound identification was achieved by comparison with the National Institute of Standards and Technology (NIST) database with a threshold match of > 80%. Quantification of compounds was achieved by comparing peak area of a selected compound to the peak area of an internal standard of known concentration. The response factor was assumed to be 1 for all compounds. Concentration of compounds was then normalized to sample weight before the carbon concentration of each compound was determined.

Characterisation of deep soil macromolecular organic carbon (experiment 2)

Sample preparation

A full profile (CU03, Harper and Tibbett 2013) with depths of 0–0.1 (4.68% TOC), 1–2 (0.23% TOC), 10–11 (0.02% TOC), 19–20 (0.02% TOC) and 28–29 m (0.01% TOC) was selected for this experiment. The location of this sample site was –32° 35'3", 117° 6'15". Soil pH values (0.01 M CaCl₂) for these samples were 5.1, 5.5, 6.2, 6.3 and 6.4, respectively.

Samples were pre-concentrated by using 5 g of sample placed into a 25 ml plastic tube with 15 ml of 0.1 M NaOH and flushed with nitrogen gas before extraction. The suspension was shaken for 24 h and then centrifuged for 15 min. Following this the solution was decanted and two further extractions were done by adding 15 ml of 0.1 M NaOH and shaking the tube by hand for 10 min. After each of the extractions the solution was separated by centrifugation as mentioned above. The combined extracts were then freeze-dried for 4 days (Hetosicc CD4 freeze dryer).

Thermochemolysis with TMAH (TC(TMAH)-GC/MS)

Approximately 5 mg of solid extract was placed in glass tubes (1 cm diameter and 18 cm long) treated with 200 µl of TMAH (25% in methanol) and flushed

with nitrogen gas. Samples were allowed to stand for 30 min then methanol was evaporated completely. The tubes were sealed under vacuum then placed in an oven at 250 °C for 30 min. After cooling to room temperature, condensation of the samples was observed. The tubes were then opened and extracted with ethyl acetate (3×1 ml). The residues were filtered through 0.45 µm polytetrafluoroethylene (PTFE) membrane before concentrating the samples under a nitrogen stream. Finally, the samples were analysed by GC/MS.

Samples were analyzed by capillary GC on a Shimadzu GC 2010. The column used for GC separation was 30 m in length, 0.25 µm in diameter and 0.25 µm thick (SGE BPX5) and helium was used as the carrier gas. Conditions were used as follows; the column was set to 40 °C for 1 min and then a temperature ramp was applied at 8 °C min⁻¹ to a final temperature of 300 °C which was held for 30 min. Aliquots of 1 µl were injected at 310 °C using splitless injection mode.

Mass spectra were obtained by a Shimadzu QP2012S mass spectrometer with ion source temperature and interface temperature of 200 °C. Compounds were scanned at the rate of 2000 amu sec⁻¹ and range of 45–1000 m/z. Peaks were identified by comparison with the NIST database.

Pyrolysis-GC/MS for laboratory standard samples and deep soil carbon

In this study, analysis was conducted on 0.5 mg samples and three standards, lignin, cellulose and chitin mixed with kaolinite at 3.85% TOC, without pretreatment. Pyrolysis was performed at two temperature ranges (i) 250–340 °C and (ii) 250–600 °C, for 10 s with a heating rate of 10 °C min⁻¹ using a pyrolysis-GC/MS from an Agilent 6890 GC interfaced to an Agilent 5973b mass selective detector. The GC was fitted with a fused silica DB5 phase capillary column (60 m length x 0.25 mm i.d. x 0.25 µm film thickness) and helium was used as the carrier gas. Split injection modes of 15:1 and 10:1 were applied for standard and field samples, respectively. The GC oven temperature was programmed from an initial temperature of 40 °C and held for 2 min, then increased at a rate of 4 °C min⁻¹ to a final temperature of 280 °C isothermal for 20 min, 70 eV full scan (50–550 amu sec⁻¹) and selected ion (m/z 57, 91, 123, 156, 170, 184, 191, 192, 217) analyses were simultaneously recorded at

a scan speed of ~2 scans sec⁻¹. The identification of individual compounds was based on comparison of mass spectral and retention time data to laboratory standards, the NIST mass spectral library and other published data.

The relative abundance of each compound was calculated as percentage of peak area and normalised to 100%. However, absolute quantification was not carried out because this would require separate verification of each of the compound identities using pure standards and then determination of the response factors for each of these compounds. Nevertheless, the relative abundance is suitable for comparing samples in a semi-quantified way (Buurman et al. 2007; González-Pérez et al. 2012; Vancampenhout et al. 2012).

Results

Identification of low molecular weight compounds in deep soil profiles (experiment I)

Compound identification

Generally, the diversity of compounds declined with depth. However, three main compound classes were identified in the chromatograms: (1) terpenes (RT=6.70–21.74 min), (2) functionalised long chain alkanes including alkanolic acids, alkyl amides and alkanols (RT=22.42–44.22 min), and (3) bioactive compounds and plant sterols (RT=54.49–58.12 min) (Table 1).

Chromatograms obtained from profiles GL03, PT06 and ST04 revealed small amounts of terpenes only in the surface soil (0.5–1 m) (Table 1). Several terpenes were common between the different locations such as camphor (9) in location GL03 and ST04; aromadendrene (46) and epiglobulol (68) in PT06 and ST04; and α -selinene (82) and cubenol (83) in GL03 and PT06. However, other terpenes were only detected at a single location. For example, borneol (14) from GL03; patchoulane (38), globulol (76) and aromadendrene oxide (86) from PT06; and isothujol (11), isoborneol (15), germacrene (52) and β -eudesmol (85) from ST04.

Fatty acids, commonly observed in all chromatograms, were divided into 2 groups according to their occurrence. The first group were fatty acids such as hexadecanoic acid (115), (Z)-octadec-9-enoic acid (124) and octadecanoic acid (126) which were

Table 1 Concentration of LMWC ($\mu\text{g C g soil}^{-1}$) for locations GL03, PT06 and ST04 at 3 depths (n = 2)

No.	RT (min)	Compound	GL03			PT06			ST04		
			0.5–1 m	11–12 m	18–19 m	0.5–1 m	11–12 m	18–19 m	0.5–1 m	11–12 m	18–19 m
9	6.40	Camphor	0.02	-	-	-	-	-	0.02	-	-
11	6.70	Isothujol	-	-	-	-	-	-	0.02	-	-
13	6.83	Octanoic acid	-	-	-	0.07	-	-	-	-	-
14	6.86	Borneol	0.10	-	-	-	-	-	-	-	-
15	6.89	Isoborneol	-	-	-	-	-	-	0.09	-	-
30	9.10	Nonanoic acid	0.27	-	-	0.22	-	-	-	-	-
32	9.32	Tridecene	-	0.03	-	-	-	-	-	-	0.06
37	10.35	γ -Elemene	-	-	-	-	-	-	-	-	-
38	10.36	Patchoulane	-	-	-	0.04	-	-	-	-	-
41	11.45	Decanoic acid	0.08	-	-	-	-	-	-	-	-
46	12.41	Aromadendrene	-	-	-	0.01	-	-	0.01	-	-
48	12.64	Espatulenol	-	-	-	-	-	-	0.02	-	-
50	12.84	Caryophyllene	-	-	-	-	-	-	0.03	-	-
52	13.13	Germacrene	-	-	-	-	-	-	0.03	-	-
54	13.98	2,6-Ditert-butylcyclohexa-2,5-diene-1,4-dione	-	0.11	-	-	-	-	-	-	-
56	14.42	1-Heptanol	-	-	-	-	-	-	-	0.05	-
65	15.99	Ethyl 4-ethoxybenzoate	0.19	0.32	0.09	0.13	-	-	-	-	-
68	16.94	Epiglobulol	-	-	-	0.12	-	-	0.07	-	-
70	17.18	Ethyl iso-allocholate	-	-	-	0.10	-	-	-	-	-
71	17.31	3-Hydroxy-4,4-dimethyldihydro-2(3 H)-furanone	-	-	-	-	-	-	-	-	0.01
72	17.32	1- <i>o</i> -(2-methylpropyl) 4- <i>o</i> -propan-2-yl 2,2-dimethyl-3-propan-2-ylbutanedicarboxylate	-	-	-	-	-	-	-	0.06	-
75	17.64	Ledol	-	-	-	-	-	-	0.29	-	-
76	17.67	Globulol	-	-	-	0.45	-	-	-	-	-
80	18.90	Eremophylene	-	-	-	-	-	-	0.08	-	-
82	19.43	α -Selinene	0.11	-	-	0.07	-	-	-	-	-
83	19.51	Cubanol	0.11*	-	-	0.06	-	-	-	-	-
85	19.63	β -Eudesmol	-	-	-	-	-	-	0.20	-	-
86	19.68	Aromadendrene oxide	-	-	-	0.12	-	-	-	-	-
92	21.74	α -Limonene diepoxide	-	-	-	-	-	-	0.09	-	-
95	22.42	Eicosanoic acid	0.27	-	-	-	-	-	-	-	-

Table 1 (continued)

No.	RT (min)	Compound	GL03			PT06			ST04			
			0.5–1 m	11–12 m	18–19 m	0.5–1 m	11–12 m	18–19 m	0.5–1 m	11–12 m	18–19 m	
114	27.34	Tritetracontane	1.58	-	-	-	-	-	-	-	-	-
115	27.57	Hexadecanoic acid	0.97	-	-	1.22	-	-	0.26	-	-	-
118	30.15	Hexadecan-1-ol	0.55	0.58	0.44	-	-	-	-	-	-	-
119	30.29	(9E)-9-Hexadecen-1-ol	-	-	-	0.36	1.81	0.74	0.17	0.62	0.65	-
124	31.51	(Z)-Octadec-9-enoic acid	0.33	-	-	0.47	-	-	-	-	-	-
126	32.07	Octadecanoic acid	0.49	-	-	0.49	0.09	-	0.09	-	-	-
127	32.40	Butyl palmitate	-	-	-	-	-	-	-	0.18	0.17	-
128	32.41	2-Methyl-2-propanyl palmitate	-	-	-	-	0.27	-	0.11	-	-	-
129	34.19	4,5-Epoxycholestan-3-one	-	-	-	0.12	0.05	-	-	-	-	-
133	36.70	Diethyl adipate	-	-	-	-	0.23	-	0.09	0.15	0.11	-
134	36.74	Hepta decyl heptadecanoate	-	-	-	0.17	-	-	-	-	-	-
135	38.87	(9Z)-9-octadecenitrile (Oleanitrile)	-	0.34	0.29	-	0.50	0.41	-	0.19	0.24	-
137	39.58	Bis(6-methylheptyl) phthalate	1.42	1.44	3.49	0.27	0.42	0.38	0.19	0.30	0.39	-
138	40.49	(9Z)-9-Octadecenamide	0.13	0.12	0.09	-	0.23	0.15	-	0.12	0.21	-
143	42.54	2,6,10-Trimethyldodecane	-	-	-	-	-	-	0.18	-	-	-
144	42.57	Tetratetracontane	-	-	-	0.37	-	-	-	-	-	-
146	44.22	(Z)-Docos-13-enamide	1.31	0.99	1.09	3.72	10.58	7.36	1.00	1.00	3.47	-
147	54.49	3-Acetoxy-7,8-epoxylanostan-11-ol	0.06	-	-	-	-	-	-	-	-	-
148	54.61	Pollinastanol	-	-	-	0.28	-	-	-	-	-	-
149	54.94	26-(Acetyloxy)-cholest-4-en-3-one	0.06	-	-	-	-	-	-	-	-	-
150	57.05	4,4'-[[4-Hydroxy-5-(2-methyl-2-propenyl)-1,3-phenylene]bis(methylene)]bis[2,6-bis(2-methyl-2-propenyl)phenol]*	0.61	0.70	0.62	-	0.10	0.08	0.05	1.02	0.96	-
152	58.12	14-Methyl-(3 β)-cholest-7-en-3-ol	-	-	-	0.06	-	-	0.03	-	-	-
		Total LMWC ($\mu\text{C g soil}^{-1}$)	8.68	4.63	6.10	8.93	14.27	9.12	3.15	3.69	6.27	-

* Tentative identification with 70% match to database

commonly observed in 0–1 m. The second group consisted of alkyl amides and alkanols, which were observed throughout the profiles. Examples were hexadecan-1-ol (118), (9E)-9-hexadecen-1-ol (119), (9Z)-9-octadecenamide (138) and (Z)-docos-13-enamide (146). Furthermore, bis(6-methylheptyl) phthalate (137) was observed in all depths. In contrast, nitrogen containing compounds, such as (9Z)-9-octadecenitrile (oleanitrile) (135), were only observed in deeper layers (11–12 and 18–19 m) (Table 1).

Plant steroids, such as pollinastanol (148) and cholesterol (129, 149, 152), were observed at 0.5–1 m. Another steroid, 4,4'-[4-Hydroxy-5-(2-methyl-2-propanyl)-1,3-phenylene]bis(methylene) bis[2,6-bis(2-methyl-2-propanyl) phenol] (150) was found throughout most profiles and locations but with only a 70% match to the database (Table 1), so this is only a tentative identification.

Compound quantification

Clearly, (Z)-docos-13-enamide (146) made the largest contribution over the whole depth of the profile at locations PT06 and ST04 (Table 1). Additionally, the relative contribution of this compound tended to increase with depth representing 42% of LMWC at 0.5–1 m, 74% at 11–12 m and 81% (18–19 m) at the PT06 site; and 32% of LMWC at 0–1 m, 27% at 11–12 m and 53% at 18–19 m at the ST04 site. However, bis(6-methylheptyl) phthalate (137) was the dominant compound in profile GL03 (Table 1) which represented 16% of LMWC at 0.5–1 m, 31% at 11–12 m and 57% at 18–19 m.

Quantitative analysis revealed that LMWC concentrations ranged from 3.15 to 14.27 $\mu\text{g C g soil}^{-1}$. Although fatty acids, amides esters and alcohols were typically observed in all locations and depths, the dominant compounds differed among locations. For example, in the 0–1 m sample from location GL03, fatty acids (115+126), long chain alkyl amides (138+146), and bis(6-methylheptyl) phthalate (137) had similar concentrations (1.42, 1.44 and 1.46 $\mu\text{g C g soil}^{-1}$, respectively). Bis(6-methylheptyl) phthalate (137) had a similar concentration (1.44 $\mu\text{g C g soil}^{-1}$) at a depth of 11–12 m but a slightly higher concentration (3.49 $\mu\text{g C g soil}^{-1}$) at 18–19 m. At the PT06 and ST04 locations, (Z)-docos-13-enamide (146) was the most abundant compound throughout the profiles (Table 1). On the other hand, several terpenes,

bioactive compounds and plant sterols were minor species, sharing approximately the same magnitude of concentrations in the first bands of all locations (Table 1).

Characterisation of deep soil macromolecular organic carbon (experiment 2)

TC(TMAH)-GC/MS: Procedure application for standards and surface layer

For TMAH thermochemolysis of pure lignin, the dominant peaks identified in the GC-MS arose from dimethoxy benzenes (Table S1) such as 3,4-dimethoxybenzaldehyde, 1-(2,4-dimethoxyphenyl) ethanone, methyl 3,4-dimethoxybenzoate, 1,2-dimethoxy-4-[(1E)-3-methoxy-1-propen-1-yl] benzene, 1,2-dimethoxy-4-(2-methoxyvinyl) benzene, 1,2-dimethoxybenzene, 1,2,4-trimethoxybenzene, vanillinor4-hydroxy-3-methoxybenzaldehyde and 1,2-dimethoxy-4-(2-propenyl)-benzene, which are typically obtained from methylation of lignin. Furthermore, various kinds of alkyl benzenes were also identified from the chromatogram (Fig S1).

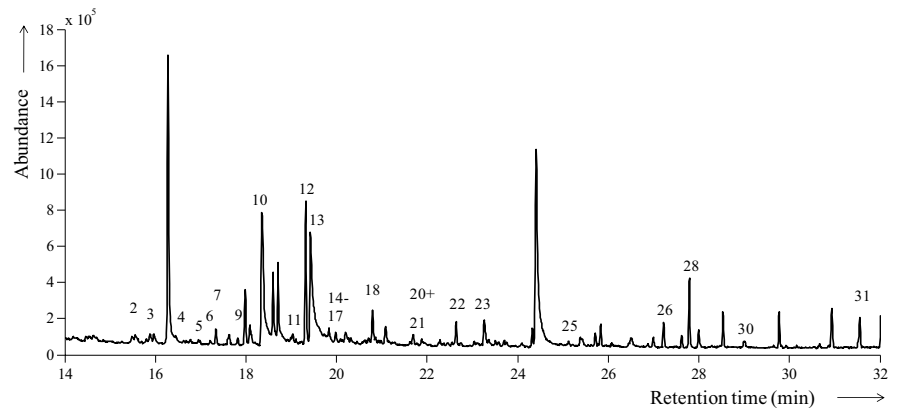
The chromatogram obtained following TMAH thermochemolysis for pure cellulose and chitin (Fig S2a) indicated the presence of many benzene derivatives (Table S2). Analysis of extracted cellulose/kaolinite samples revealed furans, while pyridine compounds were obtained from extracts of chitin/kaolinite samples. However, the number of peaks derived from this method was low compared to the number of compounds derived from methylation of pure cellulose and chitin.

A wide variety of MOC was obtained in the surface 0–0.1 m layer (Fig. 1) and the analysis indicated the presence of lignin, terpenes, polysaccharides and proteins. Aromatic compounds of non-unique origin were also identified including naphthalene, toluene and phenol (Table 2). Unfortunately, compounds detected from the deeper samples using this technique could not be categorically identified due to very low similarity index with the compounds in the NIST database.

Pyrolysis-GC/MS: for standards and deep soil carbon characterisation

The main pyrolysis products of lignin combined with kaolinite (Tables S3 and S4) were (a) simple

Fig. 1 Chromatogram of soil MOC in the 0–0.1 m layer assigned by GC/MS. Table 2 contains the full list of peak identifications



aromatics such as benzene, toluene, styrene and phenol; (b) methoxy-substituted phenol and benzene; (c) naphthalenes and phenanthrenes with methyl-substituents; and (d) unsaturated hydrocarbons, mainly alkenes (Table S4). Pyrolysis of cellulose generated signature products such as 1,4:3,6-dianhydro- α -D-glucopyranose, furans, ketones, alcohols and acids. On the other hand, pyrolysis of chitin produced compounds containing nitrogen such as pyridine, pyrole, nitriles and amides (Table S4).

Hexadecanoic acid was the only compound observed from pyrolysis of the 0–0.1 m band at 250–340 °C. This is likely to be a natural component of the soil carbon rather than a pyrolysis product from MOC degradation. No detectible compounds were obtained from pyrolysis of the deeper samples at 250–340 °C. However, 72 compounds were identified from pyrolysis at 250–600 °C from throughout the profile (Table 3). Representative chromatograms of MOC at 0–0.1 m and 28–29 m are given in Fig. 2.

Generally, the identified compounds were most abundant in the 0–0.1 m layer and declined in abundance with depth. Simple aromatic forms such as benzene, toluene, ethylbenzene and styrene were observed throughout the profile. By contrast, pyridine and benzonitrile species mostly occurred in the 0–0.1 m horizon (Table 3).

Relative abundance of deep soil pyrolysis products

Relative abundances of compounds at each profile depth showed that pyrolysis of the carbonaceous component of the soil resulted predominantly in aromatic compounds (50–80%) including toluene, benzene and styrene at all depths (Table 3). These compounds

can arise from the pyrolysis of lignin but the absence of corresponding methoxy and phenol derivatives suggests that they are more likely evidence of carbon black from past fire events. There was also a large proportion of compounds of unknown origin (13–42%). Polysaccharides were present in small proportions near the surface (17% in surface soil and 6% at 1–2 m) but declined in deeper layers.

Discussion

Identification of non-volatile LMWCs

Three distinct compound classes, namely (1) terpenes, (2) functionalised long chain alkanes including alkanolic acids, alkyl amides and alkanols, and (3) plant steroids were obtained from all profiles. Terpenes are a large and diverse class of compounds produced by all plants. The previous vegetation dominated by trees and shrubs and its replacement over the past 50–80 years by agricultural pursuits with annual crops are likely to be the main source of the observed terpene compounds. For example, many of the terpenols are derived from eucalypts (de Blas et al. 2013), which have dominated the landscape in south-western Australia for thousands of years (Brooker and Kleinig 1990). As pines were only established in the year preceding sampling, it is unlikely that they contributed to the organic compounds in the deeper profiles.

In addition to the indigenous eucalypts, there are many other genera of native plants in south-western Australia that contain terpenes in resins or essential oils (Dell and McComb 1979). This is in addition to the terpenes that plants produce as photosynthetic pigments,

Table 2 TMAH products of soil at 0–0.1 m depth

No.	RT (min)	Compound
Hydrocarbons (possibly from lignin)		
2	15.56	1,2,4-Trimethoxybenzene
3	15.80	4-Hydroxycoumarin
4	16.44	1,2,3-Trimethoxy-5-methylbenzene
5	16.68	1,2,3,4-Tetramethoxybenzene
9	17.82	Methyl 3,4-dimethoxybenzoate
10	18.35	2-Methylcoumarin
13	19.43	3-Methylcoumarin
14	19.83	1,2-Dimethoxy-4-(2-methoxy-1-propenyl)benzene
19	21.62	3-Methoxybenzoic acid
21	21.88	1,2-Dimethoxy-4-(1,2,3-trimethoxypropenyl)benzene
25	25.12	Decahydro-1,6-dimethyl-4-(1-methylethyl)naphthalene
29	28.42	3-(3,4-Dimethoxyphenoxy)-4-methoxybenzoic acid
31	31.55	1-[3,4,5-Trimethoxyphenyl]-2-methylnaphth[1,2-d]imidazole-4,5-dione
33	36.93	Octyl methoxyacetate
Polysaccharides		
26	26.99	Methyl 2,3,4,6-tetra- <i>o</i> -methyl- α -D-glucopyranoside
28	27.79	Methyl 2,3,5,6-tetra- <i>o</i> -methyl- α -D-glucofuranoside
30	29.02	Methyl 2,3,4,5-tetra- <i>o</i> -methyl- α -D-galactoseptanoside
Protein		
15	19.98	9 H-xanthene-9-carboxylic acid-(pyridin-4-yl)amide
20	21.70	N'-hydroxy-N-(3,4,5-trimethoxyphenyl)propanediamide
32	32.36	2,3-Diphenyl-1 H-pyrrolo[2,3-b]pyridine-6-amine
Terpenes		
6	17.22	Patchouli alcohol
11	19.04	Globulol
22	22.24	Yalangene
Compounds of unknown origin		
1	6.25	Phenyl butyrate
7	17.34	2,4,6-Trimethoxytoluene
8	17.63	3,5-Bis(2-methyl-2-propanyl)phenol
12	19.32	3,3,6,8-Tetramethyl-3,4-dihydro-1(2 H)-naphthalenone
16	20.09	2-(4a-Methyl-8-methylidene-1,2,3,4,5,6,7,8a-octahydronaphthalen-2-yl)propan-2-ol
17	20.20	9 H-xanthene
18	21.09	1-(3-Biphenyl)ethanone
23	23.26	Xanthone
24	23.36	Methyl 8-nonynoate
27	27.40	4-(3-Hydroxyphenyl)-oxobutyric acid

and hormones. The half-life of soil terpenes originating from the original endemic vegetation is unknown.

However, the current study suggests that LMWC derived from the current above ground vegetation

occurs mainly in the surface horizon and does not percolate to the deeper layers. This could be due to the hydrophobicity of these compounds (de Blas et al. 2013; Franco et al. 2000) and also the shallow root

Table 3 Pyrolytic products (250–600 °C) and percentage of peak area of soil samples at 5 depths

No.	RT (min)	Compound	Depth (m)				
			0–0.1	1–2	10–11	19–20	28–29
Aromatic hydrocarbons (possibly from lignin or charred wood)							
1	7.04	Benzene	6.91	22.32	16.90	12.86	26.81
7	9.20	Toluene	10.73	34.12	28.20	25.39	20.68
14	11.66	Ethylbenzene	1.79	3.85	4.76	4.82	4.43
18	12.50	Styrene	10.33	12.50	31.24	35.94	-
26	14.23	Propylbenzene	1.11	-	-	-	-
27	14.45	1-Ethyl-3-methylbenzene	0.51	1.04	-	-	-
39	16.18	1,3,5-Trimethylbenzene	0.56	-	-	-	-
40	16.19	1,2,4-Trimethylbenzene	-	1.54	-	-	-
41	16.23	1-Methyl-3-(1-methylethyl)benzene	0.32	-	-	-	-
42	16.32	1-Propenylbenzene	0.60	0.72	-	-	-
50	17.45	4-Methylphenol	4.66	-	-	-	-
68	19.70	(1-Methyl-2-cyclopropen-1-yl)benzene	1.99	-	-	-	-
72	20.17	1,4-Dihydronaphthalene	0.84	-	-	-	-
75	20.49	Dodec-1-ene	1.44	-	-	-	2.83
79	20.77	Naphthalene	1.99	-	3.35	2.39	2.60
94	22.48	6-Methyl – 1,2-dihydronaphthalene	0.86	-	-	-	-
97	22.58	3-Methyl-1,2-dihydronaphthalene	0.23	-	-	-	-
104	23.46	1-Methylnaphthalene	1.04	-	-	-	-
110	23.88	2-Methylnaphthalene	1.20	-	0.59	-	-
121	25.34	2-Ethenylnaphthalene	1.04	-	-	-	-
127	26.28	2,6-Dimethylnaphthalene	0.81	-	-	-	-
128	26.38	1,5-Dimethylnaphthalene	0.44	-	-	-	-
141	29.25	1,6,7-Trimethylnaphthalene	0.67	-	-	-	-
154	32.60	1,4,5,8-Tetramethylnaphthalene	0.25	-	-	-	-
			50.31	76.10	85.04	81.40	57.35
Hydrocarbons (possibly from cellulose)							
48	16.91	Indene	2.60	0.68	1.80	1.51	-
70	19.89	1-Methyl-1 H-indene	1.30	-	-	-	-
			3.91	0.68	1.80	1.51	0.00
N-containing compounds							
12	10.44	2-Methylpyridine	1.18	-	-	-	-
15	11.69	3-Methylpyridine	1.13	-	-	-	-
22	13.64	2,5-Dimethylpyridine	0.40	-	-	-	-
23	13.70	3,4-Dimethylpyridine	0.40	-	-	-	-
31	15.18	Benzonitrile	5.12	5.41	-	-	-
54	18.10	2-Methylbenzonitrile	1.07	-	-	-	-
63	19.32	1-Isocyano-2-methylbenzene	1.55	-	-	-	-
88	21.96	Benzeneacetonitrile, α -methylene	0.56	-	-	-	-
102	23.32	(2E)-3-Phenylacrylonitrile	0.88	-	-	-	-
105	23.58	Indolizine	0.74	-	-	-	-
167	35.12	Hexadecanenitrile	0.67	-	-	-	-
			13.70	5.41	0.00	0.00	0.00

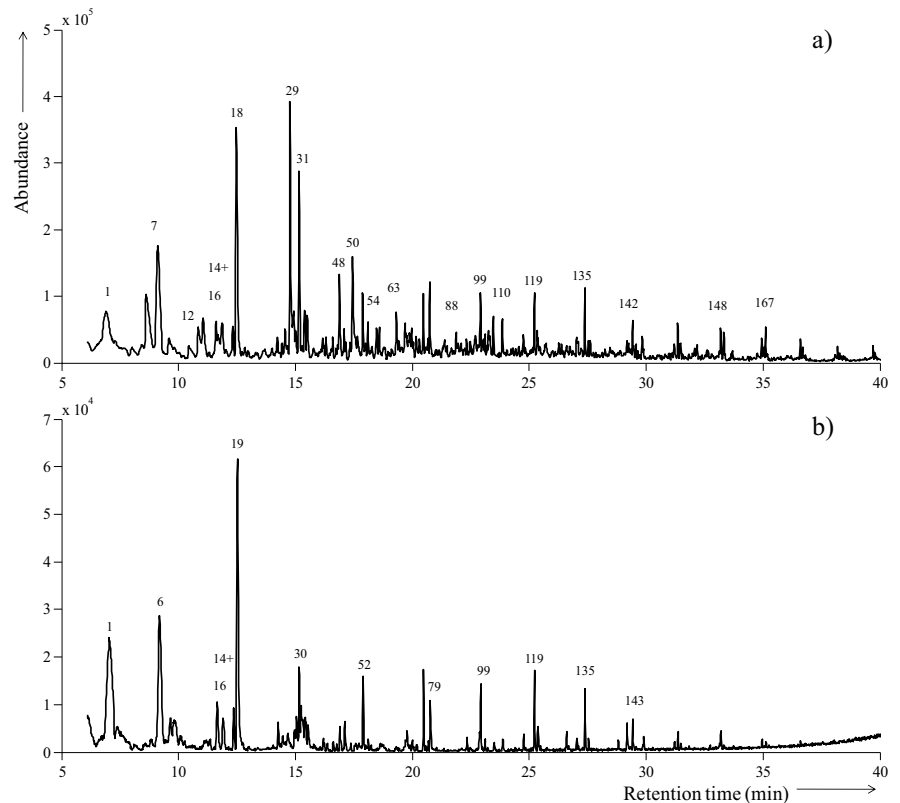
Table 3 (continued)

No.	RT (min)	Compound	Depth (m)				
			0–0.1	1–2	10–11	19–20	28–29
Compounds of unknown origin							
8	9.49	2-Methylthiophene	-	4.52	-	-	-
9	9.60	2-Octene	1.69	-	-	-	-
16	11.92	<i>p</i> -Xylene	2.20	4.95	5.32	4.34	3.05
17	12.33	Non-1-ene	1.39	-	-	-	-
19	12.52	Benzocyclobutene	-	-	-	-	23.31
28	14.55	Benzaldehyde	1.20	-	-	-	-
29	14.88	Phenol	7.67	1.15	-	-	-
30	15.17	Dec-1-ene	-	-	-	-	3.58
33	15.40	1-Propen-1-ylbenzene	1.92	-	-	-	-
34	15.52	Benzofuran	1.46	4.19	-	-	-
49	17.10	Butylbenzene	0.86	-	-	-	-
52	17.90	Undec-1-ene	-	-	-	3.27	3.16
53	17.90	Dodec-4-ene	-	-	2.37	-	-
57	18.51	2-Methyl-1-benzofuran	0.72	1.51	-	-	-
60	18.96	1-Phenyl-1-butene	0.21	-	-	-	-
78	20.68	Dodecane	0.42	-	-	-	-
92	22.32	Hexylbenzene	0.58	-	-	-	-
99	22.93	Tridec-1-ene	1.74	-	1.75	2.42	3.44
101	23.21	Tetradec-3-ene	0.37	-	-	-	-
114	24.55	8-Prop-1-en-2-ylbicyclo[4.2.1]nona-2,4,7-triene	0.49	-	-	-	-
119	25.23	Tetradec-1-ene	1.37	0.24	1.86	3.01	2.92
124	26.01	Cyclohexa-2,5-dien-1-ylbenzene	0.55	-	-	-	-
131	27.03	Octylbenzene	1.13	-	-	-	-
135	27.39	Hexadecan-1-ol	1.62	-	1.15	2.01	2.10
136	27.47	Pentadecane	-	0.88	-	-	-
137	27.53	Nonadecane	0.44	0.37	-	-	-
142	29.42	Tetradec-2-ene	1.02	-	-	-	-
143	29.44	Heptadecan-1-ol	-	-	-	1.21	1.08
144	29.44	Dodecan-2-ol	-	-	0.71	-	-
145	29.56	Hexadecane	0.35	-	-	-	-
148	31.37	Eicos-3-ene	0.76	-	-	-	-
150	31.47	Tridecane	0.58	-	-	-	-
157	33.20	Nonadec-1-ene	0.70	-	-	0.82	-
162	33.70	Anthracene	0.26	-	-	-	-
165	34.76	Cyclotetradecane	0.39	-	-	-	-
			32.08	17.80	13.16	17.09	42.65

systems of both the annual pasture plants and the young *Pinus pinaster* trees. In contrast, the root systems of the original natural vegetation are likely to have extended to bedrock (Dell et al. 1983), while leaching may be limited due to the relative low rainfall (450 mm year⁻¹) in the area (Harper and Tibbett 2013).

Similarly, plant steroids were mostly observed at 0.5–1 m depth. The behaviour of these compounds was also similar to the terpene group as they do not extend into deeper layers. Plant sterols are highly resistant to biodegradation (Bull et al. 2000) and are constituents of agricultural crops as well as the native vegetation.

Fig. 2 Chromatogram after pyrolysis of MOC in the 0–0.1 m (a) and 28–29 m (b) layers assigned by GC/MS. Table 3 contains the full list of peak identifications



Fatty acids were dominant in top soil layers and may originate from plants and microorganisms (Andreetta et al. 2013; Franco et al. 2000; Spielvogel et al. 2014). Fatty acids can occur as cutin monomers in plant waxes and have been detected in leaves, roots and wood of forest trees (Andreetta et al. 2013; Martínez-Iñigo et al. 2001; Spielvogel et al. 2014). Fatty acids can also be an indicator of biological influence in various environments, such as temperate grassland, lateritic formation processes or even in deep seas, as reported previously (Bardy et al. 2008; Feng and Simpson 2007; Huang et al. 2013). The novel finding of this study was that fatty acid amides (138+146) were commonly observed throughout the deep profiles ($1.11\text{--}10.81\ \mu\text{g C g soil}^{-1}$) while fatty acids were not detectable in the deeper layers.

In particular, (Z)-docos-13-enamide (146) occurred throughout these profiles. This compound has been identified in wood and bark and is the main component of stem oil of *Eucalyptus urophylla* x *grandis* hybrids (Xu et al. 2013). Its presence in these profiles is likely to be a fingerprint for the previous vegetation. Furthermore, the persistence of (Z)-docos-13-enamide (146) in fire-affected soils was reported by Atanassova et al. (2012). Their study

found that (Z)-docos-13-enamide was still detected even when the soil was heated up to 300 °C in air.

Although, (Z)-docos-13-enamide is mostly considered as a compound extracted from woody parts of eucalyptus trees, it can also be derived from a broad range of microorganisms including *Trichoderma harzianum*, a ubiquitous soil fungus (Siddiquee et al. 2012), *Paecilomyces* sp., an endophytic fungus of *Panax ginseng* (Xu et al. 2009) and *Bacillus* sp., a halophilic bacteria collected from solar salt works (Donio et al. 2013).

Fatty alcohols, containing carbon numbers ranging from 7 to 35, have been discovered in various environments such as soil, peatland and marine sediment (Huang et al. 2013; Treignier et al. 2006; Yang et al. 2014). However, fatty alcohols including hexadecan-1-ol (118) and (9E)-9-hexadecan-1-ol (119) have not previously been reported in deep soil layers. It is likely that these fatty alcohols were derived from hexadecanoic acid (115), one of the fatty acids commonly found in many organisms (Kaneda 1991). Significant amounts of hexadecanoic acid (115) and (9Z)-hexadecan-9-ol (119) have also been isolated from halophilic bacteria from extreme environments such a deep-sea carbonate rock at a methane cold seep in Japan (Hua et al. 2007).

Other long chain compounds including, (9Z)-9-oc-tadecenenitrile (135) and (9Z)-9-Octadecenamide (138) have also been detected from *Paecilomyces* sp. and *Bacillus* sp., respectively (Donio et al. 2013; Xu et al. 2009). Interestingly, the latter compound is produced by a wide range of microorganisms and has antimicrobial properties as discussed by Donio et al. (2013). This could be a reason for persistency of this compound in most of the samples studied ($2.25 \times 10^{-4} - 5.75 \times 10^{-4}$ g C g TOC⁻¹). It can be concluded that fatty acid amides and alcohols, including nitrogen containing compounds present in deep soils of all locations, are derived from microorganisms and plants.

It has been reported that phthalate compounds are components of wood that resist decomposition by organisms such as fungi and termites (Xu et al. 2015). This could be a reason that bis(6-methylheptyl) phthalate (137) was observed abundantly at all locations and depths in this study. Although bis(6-methylheptyl) phthalate (137) has not been reported as a compound in eucalypts and pines, several related compounds have been identified including, phthalic acid (Liu and Xu 2012; Xu et al. 2015).

Contribution of LMWC in deep soil

The presence of LMWC will be largely influenced by vegetation and in situ decomposition of plant roots. For example, terpenes and plant steroids were minor contributions in surface soils. In general, the amount of soil carbon arising from LMWC from the three profiles was in the microgram per gram range, which is significantly less than the TOC content determined from dry combustion or wet chemistry (Harper and Tibbett 2013). Even when the LMWC concentrations for species that were below the detection limit or residual compounds that may not have met the 80% match criterion of this study are factored in, it seems likely that the LMWC represents only a small proportion (0.92–3.56%) of the TOC in deep soil profiles.

Thermochemolysis (TMAH-GC/MS) of surface soil profile

An abundance of MOC was identified in the surface soil layer. Various polysaccharide compounds indicated the decomposition of organic material produced by recent vegetation. For example, furan and pyran are released from carbohydrates by wildfire or biodegradation (Atanassova et al. 2012; Schellekens et al. 2009). The

contribution from vegetation could also be proposed from the presence of 1,2,3-trimethoxy-5-methylbenzene which could be the syringyl unit of lignin. This lignin subunit could represent the residue of grass (Buurman et al. 2009). It is likely some of the syringyl measured was the result of the recent degradation from current agricultural practices: annual pasture and crop rotation.

Terpene species such as globulol and yalangene are constituents of the essential oils of native eucalypts and exotic pines, respectively (Dob et al. 2005; Tan et al. 2008; Xu et al. 2013) and provide evidence of the link to vegetation. Furthermore, condensed aromatic rings such as naphthalene observed from surface soil are the signature compounds of fire events (del Rio et al. 2001, 2002). Fire is widely used for prescribed burning and land clearing, and wildfires are common where fuel loads are high enough (Burrrows 2008).

Compound characterisation and distribution in deep soil profiles

This is the first study to characterise soil organic carbon compounds in profiles to a depth of 29 m. This study exemplified the persistence and distribution of different carbon compounds over the entire regolith depth. It is highly possible that many of the compounds identified are root-derived as evidenced by the early studies of root distribution in this region (Carbon et al. 1980; Dell et al. 1983). The main pyrolysis products of lignin identified in this study are methyl or methoxy substituted phenols, which agrees well with the study of Jiang et al. (2010). Naphthalenes and phenanthrenes were also obtained from this study which has previously been reported by Asmadi et al. (2011).

The aromatic compounds that were identified could result from the different pyrolysis pathways for lignin. Asmadi et al. (2011) demonstrated that the degradation pathway occurs after the ether bonds of lignin are broken. Consequently, the aromatic compounds are released and react via two main pathways. The first pathway is radical induced rearrangement followed by decomposition to cresols and xylenols, and subsequently phenol and benzofuran are generated. Secondly, aromatics are further reacted via homolysis of the O-CH₃ bonds to produce catechols and pyrogallols before these intermediates are decomposed to produce naphthalenes and phenanthrenes. For cellulose,

a key pyrolysis product was 1,4:3,6-dianhydro- α -D-glucopyranose, which corresponds with the studies of Lu et al. (2011) and Stefanidis et al. (2014). Although the signature compounds are mainly furans and cyclopentanones, there is no evidence of levoglucosan, which may be influenced by the pyrolysis temperature.

The detection of aromatic compounds over the entire depth implies that lignin is a recalcitrant biopolymer compared to cellulose and chitin. This is consistent with the observation that lignin comprises a major portion of root tissue and has a low decay rate (Kögel-Knabner 2002; Rasse et al. 2005). This is also consistent with the study of Moreland et al. (2021) which investigated the age of carbon in deep saprolite in the Sierra Nevada, with this increasing with age. The interaction between SOM and the soil matrix has been found to be important in recent studies (e.g. Lehmann and Kleber 2015; Schmidt et al. 2011). This bears investigation in future studies, particularly as it will relate to the rates of decomposition of deep carbon stores following deforestation.

Cellulose and N-containing compounds appear to be mostly distributed in the surface soil at 0–0.1 m and obviously linked to above ground inputs. The results also suggest that these materials are decomposed rather than leached to greater depths. The abundance of compounds of unidentified origin suggests that soil carbon also includes many residual compounds from different stages of degradation.

Performance of thermochemolysis and pyrolysis methods

Thermochemolysis with TMAH and pyrolysis methods were found to be robust for characterisation of carbon determined from pure carbon substrate products. However, the performance of these methods when dealing with samples with small concentrations was different. When comparing both methods applied to the same surface soil sample, TC(TMAH) resulted in selective fragmentation of MOC resulting in products with complex structures. In comparison, pyrolysis gave a large number of simple compounds.

Secondly, terpene compounds, indicative of plant species were observed when samples were analysed using TC(TMAH) but were lost during heating in the pyrolysis technique. The difference demonstrates the advantage of using both techniques to characterise soil organic components.

For Py-GC/MS, the pyrolysis temperature had a big effect on the number of compounds observed, with one (hexadecanoic acid) at 250–340 °C and 72 at 250–600 °C. To achieve a greater understanding of MOC in deep soil profiles at least three complementary improvements are recommended. Firstly, pre-concentration of samples with 0.1 M NaOH before analysis with Py-GC/MS. Secondly, employing multi-shot Py-GC/MS in the temperature range 250–600 °C. Thirdly, exploiting selective fragmentation with TMAH combined with Py-GC/MS at temperatures lower than thermal degradation may resolve loss of compound information.

Conclusions

The vertical distribution of LMWC and MOC was investigated for the first time in three deep soil profiles in south-western Australia. Three common compound classes of LMWC were obtained from surface soils (1) terpenes, (2) fatty acids and amides of fatty acids and (3) plant steroids, but only alkyl amides and alkanols were dominant in deep soils. The concentration of LMWC ranged from 3.15 (0–1 m) to 14.27 (11–12 m) $\mu\text{g C g soil}^{-1}$. The LMWC terpenes and plant steroids in surface soils may result from vegetation dominated by eucalypts and pine. The compounds typically isolated from deep soils such as (9Z)-9-octadecenitrile, (9Z)-9-octadecenamide, hexadecan-1-ol and (9E)-9-hexadecen-1-ol. Overall, (Z)-docos-13-enamide and bis(6-methylheptyl) phthalate may come from microorganisms and previous native vegetation.

Analysis of both LMWC and MOC indicated that the present vegetation influences only the surface soil, which was revealed by a range of biomarkers including terpenes, methoxy benzenes, pyrans (cellulose), pyridines and the occurrence of yalangene and patchouli alcohol. The presence of soil biota, such as bacteria and fungi, was revealed by the distribution of N-containing compounds not derived from chitin. Furthermore, naphthalene compounds indicated the imprint of past fire in this area. This study suggests that carbon in the form of aromatic compounds was stabilised in deep soil, whereas other carbon sources such as cellulose, chitin, and N-containing compounds were confined to the surface soil.

Author contributions Podjane Sangmanee undertook this study as part of a PhD program at Murdoch University, supervised by Bernard Dell, David Henry and Richard Harper. Suman George provided assistance with specialized analysis. The first draft of the manuscript was written by Podjane Sangmanee and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions Podjane Sangmanee was supported by a Royal Thai PhD Scholarship. The authors declare that no other funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Competing interests Financial interests: All authors declare that they have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andreetta A, Dignac M-F, Carnicelli S (2013) Biological and physico-chemical processes influence cutin and suberin biomarker distribution in two Mediterranean forest soil profiles. *Biogeochemistry* 112:41–58. <https://doi.org/10.1007/s10533-011-9693-9>
- Asmadi M, Kawamoto H, Saka S (2011) Thermal reactions of guaiacol and syringol as lignin model aromatic nuclei. *J Anal Appl Pyrol* 92:88–98. <https://doi.org/10.1016/j.jaap.2011.04.011>
- Atanassova I, Doerr SH, Bryant SH (2012) Changes in organic compound composition in soil following heating to maximum soil water repellency under anoxic conditions. *Environ Chem* 9:369–378. <https://doi.org/10.1071/EN11122>
- Bardy M, Derenne S, Fritsch E (2008) Evolution of lipid abundance and molecular composition during the podzolisation of laterites in the upper Amazon basin. *Biogeochemistry* 92:95. <https://doi.org/10.1007/s10533-008-9240-5>
- Batjes NH (1996) Total carbon and nitrogen in the soils of the world. *Eur J Soil Sci* 47:151–163
- Brooker MIH, Kleinig DA (1990) Field guide to eucalypts, vol 2. Inkata Press, Melbourne
- Bull ID, van Bergen PF, Nott CJ, Poulton PR, Evershed RP (2000) Organic geochemical studies of soils from the Rothamsted classical experiments—V. The fate of lipids in different long-term experiments. *Org Geochem* 31:389–408. [https://doi.org/10.1016/S0146-6380\(00\)00008-5](https://doi.org/10.1016/S0146-6380(00)00008-5)
- Burrows ND (2008) Linking fire ecology and fire management in south-west Australian forest landscapes. *For Ecol Manag* 255:2394–2406. <https://doi.org/10.1016/j.foreco.2008.01.009>
- Buurman P, Nierop KGJ, Kaal J, Senesi N (2009) Analytical pyrolysis and thermally assisted hydrolysis and methylation of EUROSOL humic acid samples — A key to their source. *Geoderma* 150:10–22. <https://doi.org/10.1016/j.geoderma.2008.12.012>
- Buurman P, Peterse F, Almendros Martin G (2007) Soil organic matter chemistry in allophanic soils: a pyrolysis-GC/MS study of a Costa Rican Andosol catena. *Eur J Soil Sci* 58:1330–1347. <https://doi.org/10.1111/j.1365-2389.2007.00925.x>
- Carbon BA, Bartle GA, Murray AM, Macpherson DK (1980) The distribution of root length, and the limits to flow of soil water to roots in a dry sclerophyll forest. *For Sci* 26:656–664. <https://doi.org/10.1093/forestscience/26.4.656>
- Clifford DJ, Carson DM, McKinney DE, Bortiatynski JM, Hatcher PG (1995) A new rapid technique for the characterization of lignin in vascular plants: thermochemolysis with tetramethylammonium hydroxide (TMAH). *Org Geochem* 23:169–175. [https://doi.org/10.1016/0146-6380\(94\)00109-E](https://doi.org/10.1016/0146-6380(94)00109-E)
- de Blas E, Almendros G, Sanz J (2013) Molecular characterization of lipid fractions from extremely water-repellent pine and eucalyptus forest soils. *Geoderma* 206:75–84. <https://doi.org/10.1016/j.geoderma.2013.04.027>
- del Río JC, Gutiérrez A, Martínez MJ, Martínez AT (2001) Py-GC/MS study of *Eucalyptus globulus* wood treated with different fungi. *J Anal Appl Pyrol* 58–59:441–452. [https://doi.org/10.1016/S0165-2370\(00\)00184-4](https://doi.org/10.1016/S0165-2370(00)00184-4)
- del Río JC, Speranza M, Gutiérrez A, Martínez MJ, Martínez AT (2002) Lignin attack during eucalypt wood decay by selected basidiomycetes: a Py-GC/MS study. *J Anal Appl Pyrol* 64:421–431. [https://doi.org/10.1016/S0165-2370\(02\)00043-8](https://doi.org/10.1016/S0165-2370(02)00043-8)
- Dell B, Bartle JR, Tacey WH (1983) Root occupation and root channels of jarrah forest subsoils. *Aust J Bot* 31:615–627. <https://doi.org/10.1071/BT9830615>
- Dell B, McComb AJ (1979) Plant resins—their formation, secretion and possible functions. *Adv Bot Res* 6:277–316. [https://doi.org/10.1016/S0065-2296\(08\)60332-8](https://doi.org/10.1016/S0065-2296(08)60332-8)
- Dob T, Berramdane T, Chelghoum C (2005) Analysis of essential oil from the needles of *Pinus pinaster* growing in Algeria. *Chem Nat Compd* 41:545–548. <https://doi.org/10.1007/s10600-005-0202-z>
- Donio MBS, Ronica SFA, Viji VT, Velmurugan S, Jenifer JA, Michaelbabu M, Citarasu T (2013) Isolation and characterization of halophilic *Bacillus* sp. BS3 able to produce pharmacologically important biosurfactants. *Asian Pac J Trop Med* 6:876–883. [https://doi.org/10.1016/S1995-7645\(13\)60156-X](https://doi.org/10.1016/S1995-7645(13)60156-X)
- Feng X, Simpson MJ (2007) The distribution and degradation of biomarkers in Alberta grassland soil profiles. *Org Geochem* 38:1558–1570. <https://doi.org/10.1016/j.orggeochem.2007.05.001>
- Franco CMM, Clarke PJ, Tate ME, Oades JM (2000) Hydrophobic properties and chemical characterisation of natural

- water repellent materials in Australian sands. *J Hydrol* 231–232:47–58. [https://doi.org/10.1016/S0022-1694\(00\)00182-7](https://doi.org/10.1016/S0022-1694(00)00182-7)
- Gilkes RJ, Scholz G, Dimmock GM (1973) Lateritic deep weathering of granite. *J Soil Sci* 24:523–536
- González-Pérez M, Buurman P, Vidal-Torrado P, Martín-Neto L (2012) Pyrolysis-Gas Chromatography/Mass Spectrometry Characterization of Humic Acids in Coastal Spodosols from Southeastern Brazil. *Soil Sci Soc Am J* 76:961–971. <https://doi.org/10.2136/sssaj2011.0178>
- Harper RJ, Tibbett M (2013) The hidden carbon in deep mineral soils. *Plant Soil* 368:641–648. <https://doi.org/10.1007/s11104-013-1600-9>
- Hua N-P, Kanekiyo A, Fujikura K, Yasuda H, Naganuma T (2007) *Halobacillus profundus* sp. nov. and *Halobacillus kuroshimensis* sp. nov., moderately halophilic bacteria isolated from a deep-sea methane cold seep. *Int J Syst Evol Microbiol* 57:1243–1249. <https://doi.org/10.1099/ijs.0.64817-0>
- Huang X, Zeng Z, Chen S, Yin X, Wang X, Zhao H, Yang B, Rong K, Ma Y (2013) Component characteristics of organic matter in hydrothermal barnacle shells from Southwest Indian Ridge. *Acta Oceanol Sin* 32:60–67. <https://doi.org/10.1007/s13131-013-0388-z>
- Jackson RB, Banner JL, Jobbágy EG, Pockman WT, Wall DH (2002) Ecosystem carbon loss with woody plant invasion of grasslands. *Nature* 418:623–626. <https://doi.org/10.1038/nature00910>
- Jackson RB, Canadell J, Ehleringer JR, Mooney HA, Sala OE, Schulze ED (1996) A global analysis of root distributions for terrestrial biomes. *Oecologia* 108:389–411
- Jiang G, Nowakowski DJ, Bridgwater AV (2010) Effect of the temperature on the composition of lignin pyrolysis products. *Energy Fuels* 24:4470–4475. <https://doi.org/10.1021/ef100363c>
- Kaal J, Martínez Cortizas A, Nierop KGJ (2009) Characterisation of aged charcoal using a coil probe pyrolysis-GC/MS method optimised for black carbon. *J Anal Appl Pyrol* 85:408–416. <https://doi.org/10.1016/j.jaap.2008.11.007>
- Kaal J, Rumpel C (2009) Can pyrolysis-GC/MS be used to estimate the degree of thermal alteration of black carbon? *Org Geochem* 40:1179–1187. <https://doi.org/10.1016/j.orggeochem.2009.09.002>
- Kaneda T (1991) Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev* 55:288–302. <https://doi.org/10.1128/mr.55.2.288-302.1991>
- Kleber M, Johnson MG (2010) Advances in understanding the molecular structure of soil organic matter: implications for interactions in the environment. *Adv Agron* 106:77–142. [https://doi.org/10.1016/S0065-2113\(10\)06003-7](https://doi.org/10.1016/S0065-2113(10)06003-7)
- Klingberg A, Odermatt J, Meier D (2005) Influence of parameters on pyrolysis-GC/MS of lignin in the presence of tetramethylammonium hydroxide. *J Anal Appl Pyrol* 74:104–109. <https://doi.org/10.1016/j.jaap.2004.11.023>
- Kögel-Knabner I (2002) The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biol Biochem* 34:139–162. [https://doi.org/10.1016/S0038-0717\(01\)00158-4](https://doi.org/10.1016/S0038-0717(01)00158-4)
- Kögel-Knabner I, Rumpel C (2018) Advances in molecular approaches for understanding soil organic matter composition, origin, and turnover: a historical overview. *Adv Agron* 149:1–48. <https://doi.org/10.1016/bs.agron.2018.01.003>
- Lehmann J, Kleber M (2015) The contentious nature of soil organic matter. *Nature* 528:60–68. <https://doi.org/10.1038/nature16069>
- Liu R, Xu B (2012) Characterization of essential oil in pine nut shells from commodity waste in China by steam distillation and GC-MS. *Food Anal Methods* 5:435–440. <https://doi.org/10.1007/s12161-011-9264-7>
- Lu Q, Yang X-c, Dong C-q, Zhang Z-f, Zhang X-m, Zhu X-f (2011) Influence of pyrolysis temperature and time on the cellulose fast pyrolysis products: Analytical Py-GC/MS study. *J Anal Appl Pyrol* 92:430–438. <https://doi.org/10.1016/j.jaap.2011.08.006>
- Martínez-Iñigo M, Gutiérrez A, del Río J, Martínez M, Martínez A (2001) Time course of fungal removal of lipophilic extractives from *Eucalyptus globulus* wood. *J Biotechnol* 84:119–126. [https://doi.org/10.1016/S0168-1656\(00\)00345-x](https://doi.org/10.1016/S0168-1656(00)00345-x)
- Moreland K, Tian Z, Berhe AA, McFarlane KJ, Hartsough P, Hart SC, Bales R, O'Geen AT (2021) Deep in the Sierra Nevada critical zone: saprock represents a large terrestrial organic carbon stock. *Environ Res Lett* 16:124059. <https://doi.org/10.1088/1748-9326/ac3bfe>
- Otto A, Simpson MJ (2007) Analysis of soil organic matter biomarkers by sequential chemical degradation and gas chromatography – mass spectrometry. *J Sep Sci* 30:272–282. <https://doi.org/10.1002/jssc.200600243>
- Rasse DP, Rumpel C, Dignac M-F (2005) Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant Soil* 269:341–356. <https://doi.org/10.1007/s11104-004-0907-y>
- Schellekens J, Buurman P, Pontevedra-Pombal X (2009) Selecting parameters for the environmental interpretation of peat molecular chemistry – A pyrolysis-GC/MS study. *Org Geochem* 40:678–691. <https://doi.org/10.1016/j.orggeochem.2009.03.006>
- Schmidt MWI, Torn MS, Abiven S, Dittmar T, Guggenberger G, Janssens IA, Kleber M, Kögel-Knabner I, Lehmann J, Manning DAC, Nannipieri P, Rasse DP, Weiner S, Trumbore SE (2011) Persistence of soil organic matter as an ecosystem property. *Nature* 478:49–56. <https://doi.org/10.1038/nature10386>
- Shadkani F, Helleur R (2010) Recent applications in analytical thermochemolysis. *J Anal Appl Pyrol* 89:2–16. <https://doi.org/10.1016/j.jaap.2010.05.007>
- Siddiquee S, Cheong BE, Taslima K, Kausar H, Hasan MM (2012) Separation and identification of volatile compounds from liquid cultures of *Trichoderma harzianum* by GC-MS using three different capillary columns. *J Chromatogr Sci* 50:358–367. <https://doi.org/10.1093/chromsci/bms012>
- Spielvogel S, Prietzel J, Leide J, Riedel M, Zemke J, Kögel-Knabner I (2014) Distribution of cutin and suberin biomarkers under forest trees with different root systems. *Plant Soil* 381:95–110. <https://doi.org/10.1007/s11104-014-2103-z>
- Stefanidis SD, Kalogiannis KG, Iliopoulou EF, Michailof CM, Pilavachi PA, Lappas AA (2014) A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin. *J Anal Appl Pyrol* 105:143–150. <https://doi.org/10.1016/j.jaap.2013.10.013>
- Tan M, Zhou L, Huang Y, Wang Y, Hao X, Wang J (2008) Antimicrobial activity of globulol isolated from the fruits

- of *Eucalyptus globulus* Labill. Nat Prod Res 22:569–575. <https://doi.org/10.1080/14786410701592745>
- Treignier C, Derenne S, Saliot A (2006) Terrestrial and marine n-alcohol inputs and degradation processes relating to a sudden turbidity current in the Zaire canyon. Org Geochem 37:1170–1184. <https://doi.org/10.1016/j.orggeochem.2006.03.010>
- Vancampenhout K, De Vos B, Wouters K, Swennen R, Buurman P, Deckers J (2012) Organic matter of subsoil horizons under broadleaved forest: Highly processed or labile and plant-derived? Soil Biol Biochem 50:40–46. <https://doi.org/10.1016/j.soilbio.2012.03.005>
- Xu K, Feng J, Zhong T, Zheng Z, Chen T (2015) Effects of volatile chemical components of wood species on mould growth susceptibility and termite attack resistance of wood plastic composites. Int Biodeterior Biodegrad 100:106–115. <https://doi.org/10.1016/j.ibiod.2015.02.002>
- Xu K, Li K, Yun H, Zhong T, Cao X (2013) A comparative study on the inhibitory ability of various wood-based composites against harmful biological species. 8:12. <https://doi.org/10.15376/biores.8.4.5749-5760>
- Xu L-L, Han T, Wu J-Z, Zhang Q-Y, Zhang H, Huang B-K, Rahman K, Qin L-P (2009) Comparative research of chemical constituents, antifungal and antitumor properties of ether extracts of Panax ginseng and its endophytic fungus. Phytomedicine 16:609–616. <https://doi.org/10.1016/j.phymed.2009.03.014>
- Yang H, Ding W, Xie S (2014) Distribution of microbial fatty acids and fatty alcohols in soils from an altitude transect of Mt. Jianfengling in Hainan, China: Implication for paleoaltimetry and paleotemperature reconstruction. Sci China Earth Sci 57:999–1012. <https://doi.org/10.1007/s11430-013-4729-8>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.