Thermochemical Oil Formation and the Search for a Subsurface Biosphere Hydrocarbon Fingerprint in Hydrothermal Vent Sediments at Guaymas Basin, Gulf of California

By

Connor John Dalzell

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Applied Science

August 2020, Halifax, Nova Scotia

Approved: Dr. Todd Ventura
Supervisor

Approved: Dr. Jacob Hanley
Supervisory Committee

Approved: Dr. Christa Brosseau
Supervisory Committee

Approved: Dr. Thomas Oldenburg
External Examiner

Date: August 17, 2020

Abstract

Thermochemical Oil Formation and the Search for a Subsurface Biosphere Hydrocarbon Fingerprint in Hydrothermal Vent Sediments at Guaymas Basin, Gulf of California

By: Connor John Dalzell

For this study a biomarker survey and multi-molecular chemometric analysis was performed on petroleum-rich sediments from the Cathedral Hill hydrothermal vent site in Guaymas Basin, Gulf of California. The objective of this study was to monitor the progression of petrogenesis, biological activity, and migration of hydrocarbons along a natural geothermal gradient (spanning 0–155 °C) intersected by a push core transect. Biodegradation increases down-core to variable degrees across the transect, while decoupled ratios of bacterial-sourced lipids such as hopene and hopane mark the transition zone of subsurface bacterial habitability. Various proxies show maturity increases with depth and proximity to the vent-center. These parameters, along with chemometric models, subtracted chromatograms, and hydrocarbon diversity measurements reveals that hydrocarbon generation likely begins where vent temperatures surpass 110 °C. This results in even shallower, 8-10 centimeters below seafloor, generation from what is predicted by kinetic models. The results suggest the vent temperatures may vary over time or that other applied geochemical factors (e.g. vent chemistry) may influence catagenesis. Overall, the results indicate marked spatial diversity and complexity in this sedimented hydrothermal site.

August 17, 2020

Acknowledgements

I'd like to offer my sincerest thanks to my thesis supervisor, Dr. Todd Ventura, for his guidance and steadfast support. I have learned much and benefited greatly from his tutelage, and am grateful for his persistence and enthusiasm in all aspects of the project. I thank my committee members Dr. Christa Brosseau and Dr. Jacob Hanley for their valuable feedback, suggestions, and support, which was integral for my thesis. I would also like to extend thanks to Dr. Jason Clyburne for his time on my committee, and wish him a speedy recovery. Equally, I would like to thank my co-authors/collaborators Robert Nelson, Dr. Christopher Reddy, Dr. Jeffrey Seewald, and Dr. Stefan Sievert of Woods Hole Oceanographic Institute, along with Dr. Clifford Walters from ExxonMobil Research and Engineering. Thanks for taking the time to review my first manuscript and for providing such thoughtful feedback. Thank you all for helping make this project the best it could be!

I thank and acknowledge the financial support from the following organizations: Province of Nova Scotia and Faculty of Graduate Studies and Research at Saint Mary's University (NSGS scholarship), Natural Sciences and Engineering Research Council (NSERC scholarship), Nova Scotia Department of Energy and Pengrowth Energy Corporation (Pengrowth Energy-Innovation grant).

Finally, I would like to extend thanks to my colleagues at OGL, my fellow graduate students in the geology department, and my friends and family at and beyond St. Mary's. Thank you Christopher, Jeremy, and Carl for your help, insight, and for tolerating all my complaining. Thank you Phillip, James, Kelsey, Corwin, Matt, Lee, Richard, and Paige for all the fun along the way and for listening to my ramblings. Finally, I would like to thank my family and my brother Cameron, whom I tortured daily with my derangement. None of this would be possible or worthwhile without you all. Thank you, everyone!

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
1D	one-dimensional
2D	two-dimensional
ANME	anaerobic methanotrophic (or methane oxidizing) Archaea and/or bacteria
AP	apolar extract
Biphy	biphytane
°C	degrees Celsius
cmbsf	centimeters below sea floor
CO_2	carbon dioxide
DCM	dichloromethane
DH	Diels' hydrocarbon
FID	flame ionization detector
GB	Guaymas Basin
GC	gas chromatography
GC×GC	comprehensive two-dimensional gas chromatography
H	hopane
HCA	hierarchical cluster analysis
HCl	hydrochloric acid
HMW	high molecular weight
LMW	low molecular weight
Hz	hertz
M	moretane
MeOH	methanol
mm/yr	millimeters per year
MPCA	multiway principal components analysis
mW m^2	milliWatt per square metre
m/z	mass to charge ratio
N	north
OM	organic matter
P	polar extract
PAH	polycyclic aromatic hydrocarbon
PC	principal component
Ph	phytane
Pr	pristane
Re	vitrinite reflectance equivalent
R_o	vitrinite reflectance
S	south
S^0	elemental sulphur
SRCC	Spearman's rank correlation coefficient
TLE	total lipid extract
ToFMS	time of flight mass spectrometer
UCM	unresolved complex mixture
W	west

Thesis Structure

This thesis includes four chapters. Chapter 1 is an introduction to the research topic and objectives, providing information about the study area and other background knowledge. Chapter 2 provides a comprehensive biomarker analysis to characterize the thermal and biological regime of the study site, elucidating the effects and range of the hydrocarbon fingerprints left behind by subsurface microbial communities and how this co-exists within a setting of progressive petrogenesis. Chapter 3 encompasses a multi-molecular, chemometric analysis of hydrothermal oils from the study area using comprehensive two-dimensional gas chromatography. Because Chapters 2 and 3 are prepared as independent manuscripts for journal submission, there is notable overlap between the introduction, geological setting, methods, and some results. Finally, Chapter 4 consists of a brief summary of key findings and recommendations for future work in this area.

Chapter 1: Background and Introduction

1.1 Introduction

In 1972, gasoline-range hydrocarbons were recovered in drill core from the Guaymas Basin spreading center, which alluded to petrogenesis within young marine sediments due to the high regional geothermal gradient (Goldhaber, 1974; Kalil, 1976; Simoneit et al., 1979). Further studies of the Basin revealed more diverse hydrocarbons and an extensive network of hydrothermal features (vents, chimneys, etc.), which are linked to mid-ocean spreading ridge activity and fluid-flow in the overlying sediments (Einsele et al., 1980; Curray et al., 1982a; Lonsdale and Becker, 1985; Kawka and Simoniet, 1987; Fisher and Becker 1991). These conditions promote a unique form of hydrothermal alteration, where hot, circulating fluids pyrolyze hosted sedimentary organic matter to form complex petroleum-based fluids over relatively short time spans (i.e. days-years) and shallow intervals of the seafloor (Simoneit and Lonsdale, 1982; Kawka and Simoneit, 1987; Bazylinski et al., 1988; Didyk and Simoneit, 1990). This greatly differs from conventional petroleum-producing systems, where oil generation is a slow process linked to sediment compaction and geologic heating rates during burial (~3 °C/Ma) that takes anywhere from 5-100 million years at over 5 to 6 kilometers depth of basin fill (Hunt 1979; Tissot and Welte, 1984; Selley, 1997) to accumulate. As such, the compression of space and time within hydrothermal environments creates a "natural hydrous pyrolysis experiment" to study oil-forming processes.

Hydrothermal oil is found at many ocean spreading centers, including the Escanaba Trough, along the Gorda Ridge and Middle Valley, located off-axis of the Juan de Fuca Ridge, both in the Northeast Pacific, (Simoneit, 1993b; Simoneit et al., 1997; Rushdi and Simoneit, 2002a b; Gieskes et al., 2002); Bransfield Strait in Antarctica (Brault and

Simoneit, 1989), and the Kebrit, Shaban, and Atlantis II deeps in the Red Sea (Simoneit et al., 1987; Michaelis et al., 1990). In these settings, organic matter (OM) alteration occurs from 60 to >400 °C aided by fluid chemistry, which can be rich in sulphur and metal radicals that promote carbon-carbon bond cleavage (Lewan, 1998; Fouquet et al., 1998; Zierenberg and Miller, 2000). At higher temperatures, aromatiziation and conjugation of smaller aromatic molecules results in the formation of higher molecular weight polycyclic aromatic hydrocarbons (PAHs), which greatly contributes to the overall molecular composition of hydrothermal petroleum (Kawka and Simoneit, 1990; Simoneit, 1993b). The presence of CH_4 and CO_2 (Simoneit and Galimov 1984; Von Damm et al., 1985; Welhan and Lupton 1987; Whelan et al., 1988), in addition to the thermal gradient and rapid fluid flow, further promotes migration of these petroleum-products (Simoneit, 1984; Simoneit et al., 1992; Simoneit and Fetzer, 1996; Ventura et al., 2012). The entrained oils often condense at the seafloor or near the vent site once in contact with the cold sea-bottom (Simoneit, 1984; Simoneit et al., 1992; Simoneit and Fetzer, 1996; Ishibashi et al., 2002). Overall, the quantity of petroleum can range from trace levels (found in poorly sedimented regions like the East Pacific Rise) to seabed pools of oil (in highly organic-rich sedimented regions like Guaymas Basin, Escanaba Trough, and Middle Valley) (Simoneit, 1982; Bazylinski et al., 1988; Kvenvolden and Simoneit, 1990; Kawka and Simoneit, 1994).

Much like these hydrothermal environments, laboratory pyrolysis experiments can demonstrate the effects of heating on kerogen over relatively short timespans (e.g. Hoering, 1968; Ishiwatari et al., 1977; Van Graas et al., 1981; Horsfield, 1984; Lewan, 1985; Lewan et al., 1986; Eglinton and Douglas, 1988; Leif and Simoneit, 1995; McCollom et al., 1999; Zárate-del Valle et al., 2006). However, they have thus far not been able to replicate the diagenesis of OM (Tissot and Welte, 1984), and therefore do not realistically reflect the

true complexity of natural petroleum-forming systems. In this regard, the hydrothermal environments at Guaymas Basin may provide an opportunity to study these natural processes, as temperatures reaching ~120 °C within a meter below the seafloor (e.g. Bazylinski et al. 1988; Magenheim and Gieskes, 1992; Elsgaard et al., 1994; Weber and Jørgensen, 2002; McKay et al., 2012; Teske et al., 2016). Even higher temperatures (>400 °C) have been recorded in some active vent sites (Simoniet, 1993a). This may allow the entire process of OM diagenesis, catagenesis, and metagenesis to be framed within the confines of a single scientific study, as the rapidly accumulating, near-uniform hemiplegic sediments of the basin (Van Andel. 1964; Calvert, 1966; Curray et al., 1982a; Kelts et al., 1982; Simoneit and Lonsdale, 1982) are progressively altered with burial and proximity to the vent-center (i.e. with increasing temperature). As such, this study examines the hydrocarbons along a transect of hydrothermal sediments from an active vent site, in order to monitor petrogenesis and microbial alteration within these systems.

1.1.1 Considerations for oil characterization

One of the most common and simple approaches to classify and compare oil samples is to use diagnostic hydrocarbon ratios, which can be used to assess petroleum compositional similarities, thermal histories, various degradation processes, and the biological and geochemical conditions within paleo-depositional environments (e.g. Boehm et al., 1997; Peters et al., 2005, 2007; Wang and Stout, 2007; Christensen et al., 2016). This ultimately relies on the comparison of two or more integrated compound peak-areas, that are traditionally collected from gas chromatograms obtained by the analysis of a hydrocarbon-based matrix using gas chromatography–flame-ionization detection (GC–FID) or gas chromatography–mass spectrometry (GC-MS) (Peters et al., 2007; Wang and

Stout, 2007; Christensen et al., 2016). However, more advanced techniques such as comprehensive two-dimensional gas chromatography (GC×GC) have distinct advantages over traditional "1D" methods as they provide enhanced peak capacity and resolving power by separating hydrocarbons onto two chromatographic dimensions, having unique column separation capabilities (e.g. Barman and Membrado, 2000; Reddy et al., 2002; Frysinger et al., 2002; Arey et al., 2005, 2007; Ventura et al., 2007, 2008, 2012; Tran et al., 2010). Multi-dimensional gas chromatography can be used to characterize oils which are normally too complex to resolve with 1D methods (i.e. Booth et al., 2007; Ventura, 2007, 2012; Tran et al., 2010; Wang et al., 2010; Zeng et al., 2012), by reducing peak co-elution in oils that have significantly reduced paraffin contents; especially by contributions of normal alkanes (Tissot et al., 1971; Rowland and Maxwell, 1984; Killops and Al-Juboori, 1990; Warton et al., 1999). This may be especially important for the analysis of Guaymas Basin sediments, which have reportedly high rates of heterotrophy (e.g., Edgcomb et al., 2002; Biddle et al., 2012; Liu et al., 2016).

Studies that document oil transformations are invariably limited by the difficulty of tracking the multitude of hydrocarbons that comprise petroleum, which can encompass thousands of individual compounds (Brooks et al., 1957; Tissot and Welte, 1984; Marshall et al., 1988; Mullins et al., 2006). Furthermore, relying on a limited number of molecular ratios and hydrocarbons may not accurately reflect the holistic, multi-molecular changes occurring in these complex fluids. However, in recent years the application of chemometric techniques to GC×GC analysis of petroleum has allowed for the comparison of thousands of individual hydrocarbons across datasets, mitigating this longstanding problem with traditional oil fingerprinting techniques (e.g. Wang et al., 2014; Christensen et al., 2016; Peters et al., 2016, Zhan et al., 2018).

First described by Svante Wold in 1971, chemometrics is the application of mathematical and statistical methods to chemistry, allowing for more extensive and informative analyses of chemical datasets (Kowalski, 1975, 1977; Meglen, 1988; Massart and Buydens, 1988; Christensen and Tomasi, 2007; Héberger, 2008). These encompass several techniques, including cluster analysis, principal component analysis, partial least squares regression, pattern recognition, parallel factor analysis, and neural networks (Wold, 1995; Christensen et al., 2005, 2016; Christensen and Tomasi, 2007; Héberger, 2008; Sarker and Nahar, 2015). For oils, which are among the most complex chemical mixtures known (Marshall and Rodgers, 2003), the use of multivariate statistics in chemometrics can elucidate correlations between these large, intricate datasets, and identify/describe patterns associated with physicochemical processes such as thermal maturity, weathering, migration, and genesis of petroleums (e.g. Christensen et al., 2005; Christensen and Tomasi, 2007; Wang, et al., 2018).

One chemometric technique that has had relative success in characterizing oils analyzed by GC×GC is multi-way principal components analysis (MPCA), which is a multivariate dimensionality-reducing statistic (Ventura et al., 2011; Mogollón et al., 2017; Prata et al., 2016). Another technique, hierarchical cluster analysis (HCA), is an agglomerative clustering method that has been used to link oil samples based on geochemical data (Peters and Fowler, 2002, Pavón et al., 2006, Barbeira et al., 2007, Peters et al., 2007, Wang et al., 2018; Zhan et al., 2018) and whole GC analysis (Hashemi-Nasab et al., 2019). Together, the application of these methods may identify holistic changes between oil samples that, along with traditional biomarker analysis, may help assess the oil/kerogen characteristics of near-surface hydrothermal vent sediments.

As such, the characterization of hydrothermal oils in this study will be conducted via GC×GC to enable a survey of hydrocarbon ratios, along with a comprehensive chemometric assessment, as the primary focus of Chapters 2 and 3 (respectively). The synthesis of these chapters will provide a detailed overview of this hydrothermal system, with the more established hydrocarbon ratios and techniques creating a foundation of understanding that can be further improved by more exploratory, holistic methodology.

1.1.2 Objectives

Here, we apply traditional biomarker analysis and chemometric techniques to a transect of petroleum-impregnated sediments from the Cathedral Hill hydrothermal vent site at Guaymas Basin, Gulf of California. At this site, high temperature fluids projected to reach up to 155 °C by 21 cm sediment depths may be capable of pyrolyzing the organic-rich sediments into petroleum hydrocarbons. This dual analytical approach will elucidate the degree to which the transect sediments are affected by petrogenesis, microbial activity, and other processes (**Chapter 2**), along with how these changes are reflected in the multi-molecular suite of hydrocarbons (i.e. **Chapter 3**). Through this, we aim to capture the stepwise progression of thermal maturity, and the interactions of a microbial community and their modifications to the sedimentary organic matter, along a natural geothermal gradient in this near-surface hydrothermal vent system, which may identify to what degree the multi-molecular matrix-changes are attributed to in situ petrogenesis/thermochemical oxidation, biodegradation, and migration.

1.2 Geological Context

Guaymas Basin is a submarine depression in the Gulf of California, formed by extension of the East Pacific Rise oceanic spreading center (Rusnak et al., 1964; Larson et

al., 1968; Moore et al., 1973). The basin is comprised of two axial rift valleys to the north and south (Moore, 1973; Lonsdale and Lawver, 1980, Lonsdale and Becker, 1985), which show substantial hydrothermal activity resulting from the emplacement of basaltic dykes and sills in the overlying sediments (Einsele et al., 1980; Curray et al., 1982a; Fisher and Becker 1991; Lonsdale and Becker, 1985; Kawka and Simoneit, 1987; Bazylinski et al. 1988). These hydrothermal fluids can reach over 300 °C in some localities, circulating through the thick, organic-rich cover of the basin (Lonsdale et al., 1980; Lonsdale and Becker, 1985; Von Damm et al., 1985; Bazylinski et al., 1988). This rapidly pyrolyzes the detrital OM, generating petroleum based fluids (Simoneit and Lonsdale, 1982; Simoneit, 1984, 1990) and other by-products (i.e. methane, CO_2, ammonia, organic acids; Von Damm et al., 1985; Whelan et al., 1988; Martens, 1990).

Cathedral Hill is a hydrothermal vent site in the Southern Trough of the Guaymas Basin, located 27°00.696 N, 111°24.265 W at 1996m depth (Fig. 1). The area comprises a central complex of active white smoker chimneys covered with giant tube worms (*Riftia pachyptila*) and a microbial mat (likely sulfide oxidizing, filamentous *Beggiatoa*; McKay et al., 2012; Teske et al., 2016). Previously reported porewater temperature measurements reach 120 °C within 45 centimeters below seafloor (cmbsf) of the organic-rich sediments (Teske et al., 2016), which are comprised of diatomaceous oozes and turbidite muds intermittently mixed with calcareous shells and buried benthic macrofauna (see section 2.1). These sediments are sourced from continental run-off and highly-productive overlying waters (Van Andel. 1964; Calvert, 1966; Curray et al., 1982a; Kelts et al., 1982), with an inferred sedimentation rate of 1–2 mm/yr (though this has not been directly measured at Cathedral Hill; Williams et al., 1979; Teske et al., 2002). Vent-fluid chemistry, subsurface temperatures, and dissolved organic carbon content are comparable to other vent sites in

the Southern Trough (Bazylinski et al. 1988; Weber and Jørgensen, 2002; Reeves, 2010; McDermott et al., 2015; Teske et al., 2016; Zhuang et al., 2019). The regional heat-flow regime at Cathedral Hill (500-1000 mW m^2) is also similar to other hydrothermal features in the Southern Trough (i.e. Rebeca's Roost, Big Pagoda, Robin's Roost, etc.), which itself has a background reading of ~660 mW m^2 (Lonsdale and Becker, 1985; Teske et al., 2016)

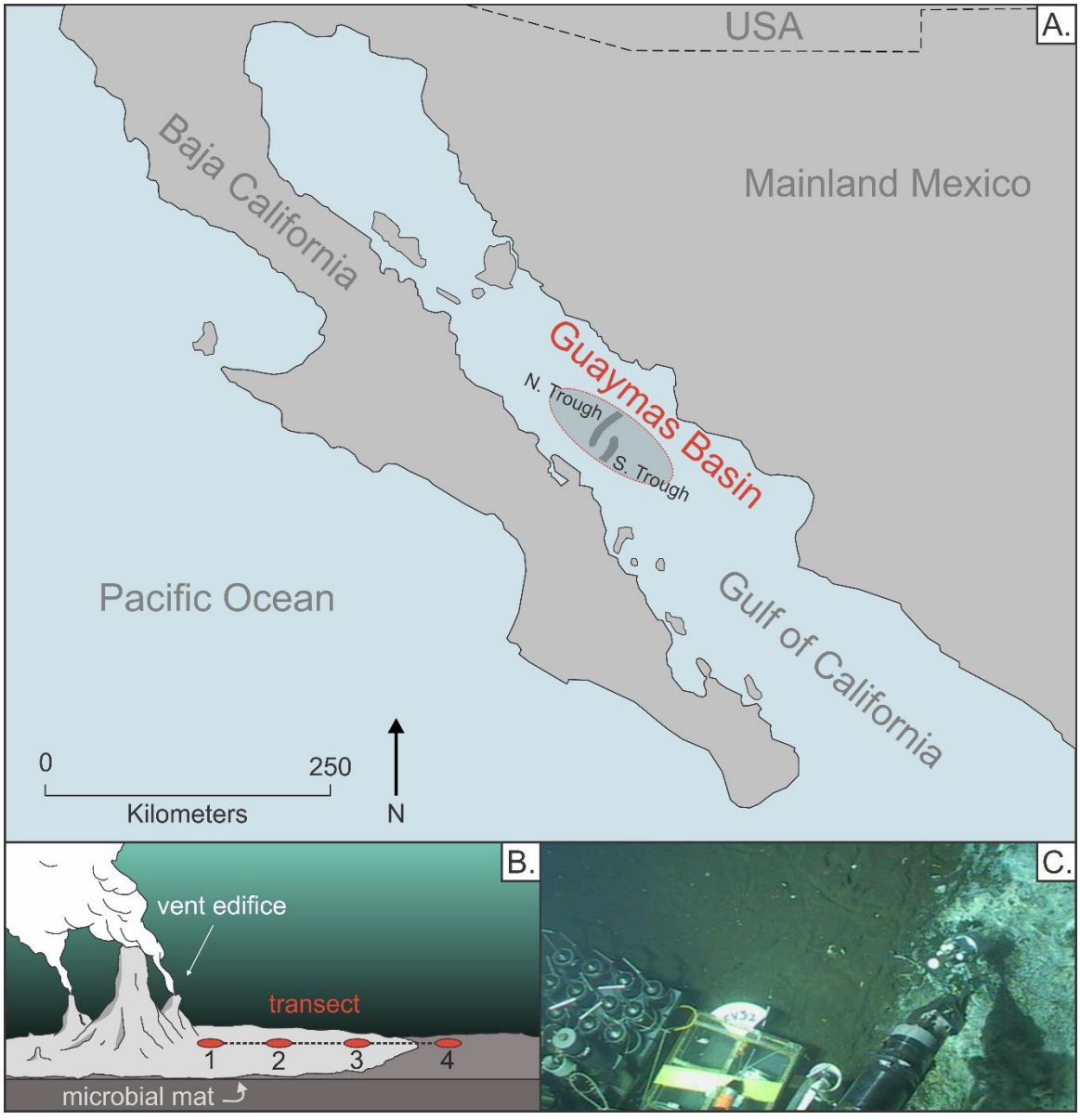


Fig. 1. (A.) Guaymas Basin study site (edited from Castillo et al., 2002). (B.) Model of the push core transect at Cathedral Hill. (C.) Photograph of Cathedral Hill sediments and microbial mat, showing thermal probe and push core basket.

1.2.1 Sedimentation and sources of organic material

Guaymas Basin is characterized by rapid deposition of organic-rich sedimentary material from both highly productive overlying waters and terrigenous (continental) inputs (Van Andel. 1964; Calvert, 1966; Curray et al., 1982a, 1982b; Kelts et al., 1982; Simoneit and Lonsdale, 1982). These sources culminate into thick sedimentary layers comprised of roughly 2–4% organic carbon sourced primarily from diatomaceous and microbial detritus (Curray et al., 1982a, 1982b; Jannasch and Taylor, 1984; De la Lanza-Espino and Soto, 1999). The Basin also acts as a depocenter for periodic turbidity flows (Moore, 1973; Curray et al., 1982a, 1982b; Kluesner et. al, 2014). Sedimentation rate varies from 1–2 mm/yr., and the total thickness ranges from 300–500 m (Van Andel, 1964; Calvert, 1966; Williams et al., 1979; Curray et al., 1979; Teske et al., 2002). The oldest dated sediments in the basin are early Miocene to Pliocene (Moore, 1973; Kelts et al., 1982).

1.2.1.1 Terrigenous input

Continental sources can comprise up to ~50–90% of the sedimentary material within the Gulf of California, which varies depending on the proximity to the coast and major river systems (Van Andel, 1964; Williams et al., 1979; Dean, 2006). Terrigenous input is highest in the northern Gulf and is more clay-rich, affected by extreme deposition from the Colorado River delta (Curray, 1982a, Williams 1979). Five major rivers additionally supply the Guaymas Basin (Curray et al., 1982b; Dean, 2006), though in modern times, much of this has ceased due to dam construction (Baba et al., 1991, Dean, 2006).

1.2.1.2 Marine input

The remaining material in the Gulf is autochthonous ocean sediments, comprised of mainly siliceous particulate with some coccoliths (Van Andel, 1964; Calvert, 1966; Curray, 1982a). Wind-driven surface currents drive upwelling of nutrients and dissolved silica from the Pacific Ocean bottom waters, promoting diatom activity in the euphotic zone of the Gulf and supplying the seafloor with diatomaceous sediments (Van Andel, 1964; Calvert, 1966; Sancetta, 1995; Thunell, 1998). These marine materials increase towards the central area of the Gulf, caused by the high organic productivity (Van Andel, 1964; Calvert, 1966)

1.2.1.3 "Localized" sources of organic material

1.2.1.3.1 Microbial Mats

Microbial mats from the genus *Beggiatoa* colonize the sulfide-rich sediments of active hydrothermal systems at Guaymas Basin, including Cathedral Hill (Jannasch et al., 1989; Nelson et al., 1989; McKay et al., 2012; Teske et al., 2016). These are individually large, filamentous sulfur-oxidizing bacteria of a white, yellow, and/or orange colouration that appears to be temperature-dependant (Jannasch and Taylor, 1984; Nelson et al., 1989; Van Dover, 2000; Teske et al., 2016). These mats are generally 1–10 cm thick, can extend "root-like" structures several centimeters into the substrate for stability, and will typically migrate upward through the sediment column to avoid burial, which could limit their detrital contribution to the sediments (Van Dover, 2000; Judd and Hovland 2007; McKay et al., 2012). Importantly, microbial mats can alter the subsurface environment by preventing fluid and gas exchange from the seafloor to ocean bottom-waters, which can create extreme

chemical microenvironments that are hotter and contain more ions than the surrounding porewaters (Dymond et al., 1989; Judd and Hovland, 2007).

1.2.1.3.2 Subsurface Microorganisms

A variety of other microbial communities thrive in the hydrothermal sediments at Guaymas Basin (e.g. Karl et al., 1988; Teske et al., 2002; Edgcomb et al., 2002; Guerrero-Barajas et al., 2011; Biddle et al., 2012; Vigneron et al., 2014; Liu et al., 2016). Examples of this microfaunal diversity include the ANME-1 and -2 archaeal lineages, which inhabit a range of temperatures (2–96 °C; Teske et al., 2002, Biddle et al., 2012), SEEP-SRB2 delta proteobacteria, and HotSeep-1 bacterial groups, with some lineages only described at Guaymas Basin (Dowell et al., 2016). In general, the microfauna of hot vent systems are likely to be more diverse than those at cold seeps, as the temperature range of these environments (>300 °C to ~0 °C) supports a variety of extremophiles (Alain, 2002; Judd and Hovland, 2007; Holden, 2009; Teske, 2009).

1.2.1.3.3 Benthic Macrofauna

Heavy bioturbation in surface sediments at Guaymas Basin (Curray et al., 1982a, 1982b) suggest that benthic macrofaunal remains could contribute to the sedimentary OM. Some seafloor-dwelling organisms include tubeworms and bivalves, which encrust hydrothermal vent edifices (Cruaud et al., 2015; Judd and Hovland, 2007) or scale worms that feed off microbial mats (Teske et al., 2016).

1.2.2 Hydrothermal petroleum character

Petroleum compositions at Guaymas Basin are well documented (e.g. Simoneit, 1984, 1985; Kawka and Simoneit, 1987, 1990, 1994; Bazylinski, et al., 1988; Didyk and Simoneit, 1990). These oils are distinct from typical reservoir petroleum, as reflected in their high diversity and greater overall polar content (Tissot and Welte, 1984; Kawka and Simoneit 1987; Kvenvolden and Simoneit 1990; Kawka, 1990). Guaymas Basin oils are typically characterised by mature biomarkers (e.g., αβ-hopanes; triaromatic steranes), olefins, PAHs, C_{21} Diels' hydrocarbon, and broad suites of *n*-alkanes with no carbon number preference (e.g. Simoneit, 1984, 1993b; Kawka and Simoneit, 1987; Kawka 1990; Kawka and Simoneit, 1994). These oils are often reworked by biodegradation or water washing (Kawka, 1990; Lonsdale and Becker, 1985), with surveys of seabed petroleums showing marked spatial heterogeneity on the scale of centimeters attributed to these processes along with mixing from multiple sources (Bazylinski, et al. 1988).

1.3 Methods

1.3.1 Comprehensive two-dimensional gas chromatography (GC×GC)

The method of GC×GC was first introduced by Phillips et al. (1996, 1999), and has since become a widespread analytical tool for the separation and identification of hydrocarbons from complex mixtures (e.g. Frysinger et al., 2002; Reddy et al., 2002; Dallüge et al., 2003; Dimandja, 2004; Arey et al., 2005, 2007; Ventura et al., 2007, 2008, 2012; Adahchour et al., 2008; Tran et al., 2010; Marriott et al., 2012). This technique evolved from heartcutting multidimensional gas chromatography, which improved upon conventional methods via selective, independent separations of the eluent using a primary and secondary column joined by a diversion valve (Schomburg, 1995; Dimandja, 2004;

Marriott et al., 2012). In this approach, enhanced resolution of select chromatographic regions are made by diverting the more complicated/relevant portions of the eluent onto the secondary column, allowing these regions to be further separated and analyzed independently on another detector (Dimandja, 2004; Adahchour et al., 2008; Marriott et al., 2012). This differs from GC×GC, which is akin to "continuous" heart-cutting and is truly comprehensive in that the entire eluent is transferred onto the secondary column, allowing for maximized peak capacity (Beens et al., 2001; Dimandja, 2004; Marriott et al., 2012). Importantly, GC×GC is able to mitigate the high time-requirement that effectively prevents a comprehensive analysis via heartcutting (Beens et al., 2001).

The principle of GC×GC is based on two independent and mechanically distinct separations of the sample, provided via linkage of two GC columns that have unique stationary phases which are linked by a modulator (Dallüge et al., 2003; Dimandja, 2004; Marriott et al., 2012). The first stage of separation is achieved by passage of the eluent through a 1st dimension capillary column, which is typically non-polar and driven via thermal gradient elution and a carrier-gas (Fig.2). In this regard, 1st dimension separation is carried out like traditional gas chromatography, via partitioning and separation of the sample on the stationary phase (Heftmann, 2004; Shellie, 2013). The 1st dimension eluting analytes are then refocused in the modulator and sequentially injected onto the 2nd dimension column (Dallüge et al., 2003; Dimandja, 2004; Marriott et al., 2012). Most modern systems employ two-jet cryogenic modulators, which use pulsed flow of heated vs. cryogenically cooled gas to trap (freeze) and release (thaw) the eluent (Beens et al., 2001; Dallüge et al., 2003). Traditionally, the 2nd dimension column is shorter and of a narrower bore diameter than the 1stdimension, resulting in a final stage of separation that elutes quite fast, generating narrow peaks that require a high sampling-rate detector of at least ~20 Hz

to resolve (Dallüge et al., 2003; Dimandja, 2004; Marriott et al., 2012). Flame-ionization detection (FID) is widely used for this purpose, which has a high scan frequency (50–400 Hz) well-suited for such rapid analysis (Beens et al., 1998a, 2005; Dimandja, 2004). For compound identification, mass spectral analysis can be carried out using time-of-flight mass spectrometry (ToF-MS), which has a high data acquisition rate (Van Deursen et al., 2000; Shellie et al., 2001).

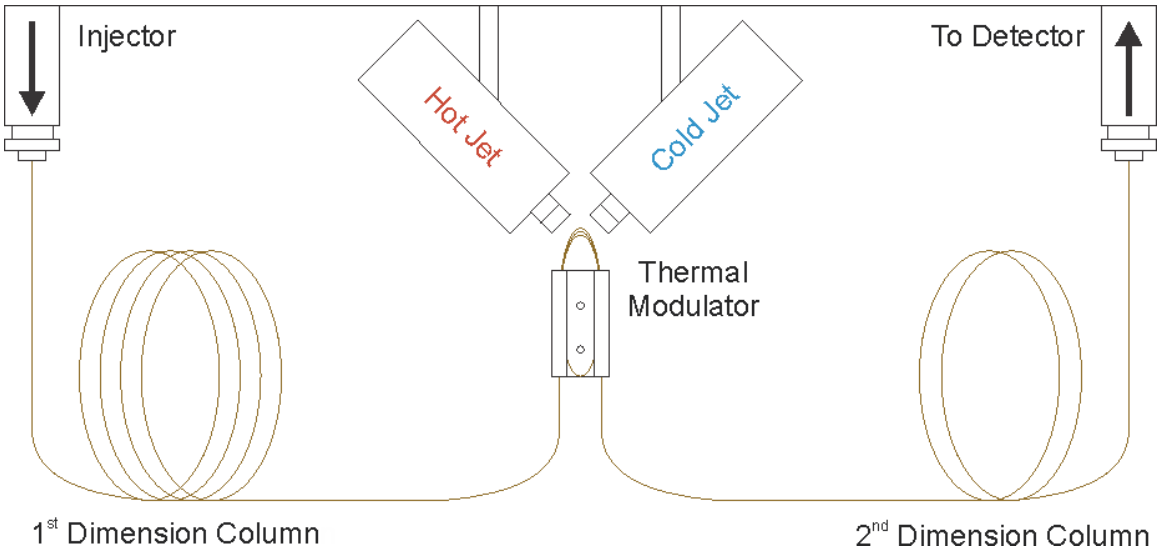


Fig. 2. Overview of a GC×GC instrumental system.

The final output of GC×GC analysis is a continuous data stream from the 2nd dimension column separation, which is segmented and stacked to form a 2D matrix array based on the modulation frequency (Dallüge et al., 2003; Dimandja, 2004; Marriott et al., 2012). The resulting chromatogram can be visualized as three-dimensional plots or on a 2D plane using colours, contour lines, etc. to indicate signal intensity, where the x and y axis represents the separations achieved on the 1st and 2nd dimension columns respectively (Fig. 3). Importantly, GC×GC allows for hydrocarbons to be binned into compound-classes based on their elution-position in chromatographic space, which is dependant on their

chemical interactions with the 1st and 2nd dimension columns such that similar molecules will elute or cluster in predictable patterns (Beens et al., 1998b; Dimandja, 2004). Compared to traditional gas chromatography, GC×GC has enhanced resolving power/peak capacity, and greatly reduces the co-elution problems inherent to 1D-GC methods (Frysinger et al., 2002; Dallüge et al., 2003; Adahchour et al., 2008; Barman and Membrado, 2000; Marriott et al., 2012). This is demonstrated in Fig. 3 by the unresolved complex mixture (UCM) depicted in the 1st dimension reconstructed 1D gas chromatogram. The 1D chromatogram is an example of how peaks become hidden by co-elution in GC space.

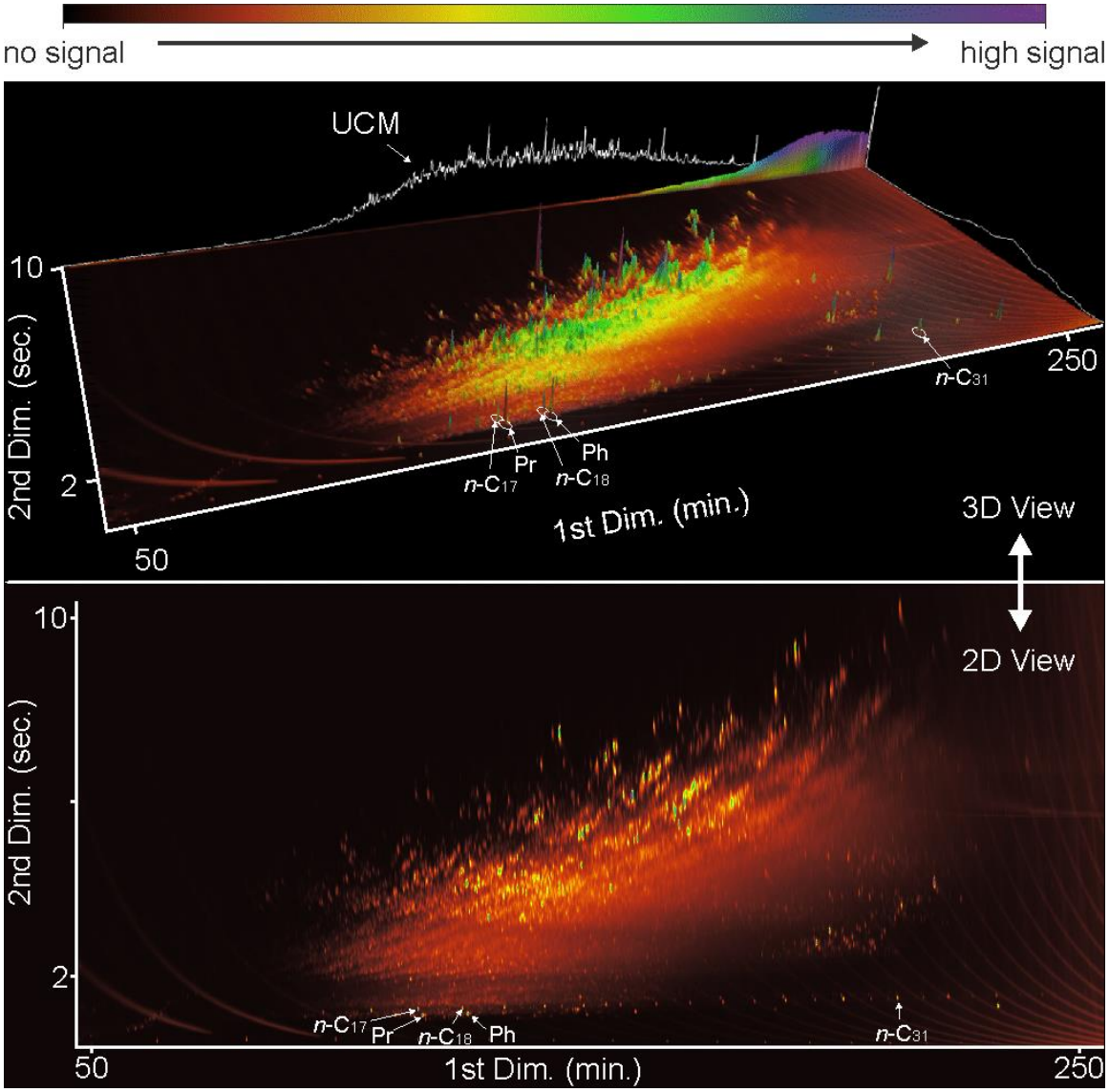


Fig. 3. GC×GC-FID chromatograms of Cathedral Hill oils in both isometric (3D) and plan view (2D). The UCM is identified in isometric view, while the biomarkers Pristane (Pr), Phytane (Ph), *n*-C_{17}, *n*-C_{18}, and *n*-C_{31} are indicated on both planes.

1.4 Works Cited

Adahchour M., Beens J. and Brinkman, U. A. T. (2008). Recent developments in the application of comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **1186**(1-2), 67–108.

Alain K. (2002) Phylogenetic characterization of the bacterial assemblage associated with mucous secretions of the hydrothermal vent polychaete Paralvinella palmiformis. *FEMS Microbiol. Ecol.* **42**(3), 463–476.

Arey J. S., Nelson R. K., Xu L. and Reddy C. M. (2005) Using comprehensive two dimensional gas chromatography retention indices to estimate environmental partitioning properties for a complete set of diesel fuel hydrocarbons. *Anal. Chem*. **77**, 7172–7182.

Arey J. S., Nelson R. K. and Reddy C. M. (2007) Disentangling oil weathering using GC GC. 1. Chromatogram analysis. *Environ. Sci. Technol*. **41**, 5738–5746.

Baba J., Peterson C. D. and Schrader H. J. (1991) Modern fine-grained sediments in the Gulf of California. In *The Gulf and Peninsular Province of the Californias* (eds. J. P. Dauphin and B. R. T. Simoneit). American Association of Petroleum Geologists, Tulsa. pp. 569–587.

Barbeira P. J. S., Pereira R. C. C. and Corgozinho C. N. C. (2007) Identification of gasoline origin by physical and chemical properties and multivariate analysis. *Energ. Fuel* **21**(4), 2212–2215.

Barman B. and Membrado V. (2000) Chromatographic techniques for petroleum and related products. *Crit. Rev. Anal. Chem.* **30**(2-3), 75–120.

Bazylinski A., Farrington W. and Jannasch W. (1988) Hydrocarbons in surface sediments from a Guaymas Basin hydrothermal vent site. *Org. Geochem*. **12**, 547-558.

Beens J., Adahchour M., Vreuls R. J. J., van Altena K. and Brinkman U. A. T. (2001) Simple, non-moving modulation interface for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **919**(1), 127–132.

Beens J., Boelens H., Tijssen R. and Blomberg J. (1998a). Quantitative Aspects of Comprehensive Two-Dimensional Gas Chromatography (GC×GC). *J. High Resolut. Chromatogr*. **21**(1), 47–54.

Beens J., Brinkman, U. A. T. (2005) Comprehensive two-dimensional gas chromatography - a powerful and versatile technique. Analyst **130**, 123–127.

Beens J., Tijssen R. and Blomberg J. (1998b). Prediction of comprehensive two-dimensional gas chromatographic separations. *J. Chromatogr. A* **822**(2), 233–251.

Biddle J. F., Cardman Z., Mendlovitz H., Albert D.B., Lloyd K.G., Boetius A. and Teske A. (2012) Anaerobic oxidation of methane at different temperature regimes in Guaymas Basin hydrothermal sediments. *ISME J.* **6**(5), 1018–1031.

Boehm P. D., Douglas G. S., Burns W. A., Mankiewicz P. J., Page, D. S. and Bence A. E. (1997) Application of petroleum hydrocarbon chemical fingerprinting and allocation techniques after the Exxon Valdez oil spill. *Mar. Pollut. Bull.* **34**(8), 599–613.

Booth A. M., Sutton P. A., Lewis C. A., Lewis A. C., Scarlett A., Chau W., Widdows J. and Rowland S. J. (2007) Unresolved Complex Mixtures of Aromatic Hydrocarbons: Thousands of Overlooked Persistent, Bioaccumulative, and Toxic Contaminants in Mussels. *Environ. Sci. Technol.* **41**(2), 457–464.

Brault M. and Simoneit B. R. T. (1989) Trace petroliferous organic matter associated with hydrothermal minerals from the Mid-Atlantic Ridge at the trans-Atlantic geotraverse 26°N site. *J. Geophys. Res.* **94**(C7), 9791.

Brooks B. T., Bood C. E., Kurtz Jr. S. S. and Schmerling L. (1957) *The Chemistry of Petroleum Hydrocarbons*, Vol. 1. Reinhold Publishing, New York.

Castillo P. R., Hawkins J. W., Lonsdale P. F., Hilton D. R., Shaw A. M. and Glascock M. D. (2002) Petrology of Alarcon Rise lavas, Gulf of California: Nascent intracontinental ocean crust. *J. Geophys. Res.* **107**, 1-15

Cruaud P., Vigneron A., Pignet P., Caprais J., Lesongeur F., Toffin L., Godfroy A. and Cambon-Bonavita M. (2015) Microbial communities associated with benthic faunal assemblages at cold seep sediments of the Sonora Margin, Guaymas Basin. *Front. Mar. Sci.***2**, 53–69.

Curray J. R., Moore D. G. et al. (1979) Leg 64 seeks evidence on development of basins. *Geotimes* **24**(7), 18–20.

Curray J., Moore D., and the DSDP Scientific Party. (1982a) Guaymas Basin: Sites 477, 478, and 481. In *Initial reports of the Deep Sea Drilling Project-LXIV.* U.S. Govt. Printing Office, Washington DC.

Curray J., Moore D., and the DSDP Scientific Party. (1982b) Guaymas Basin Slope: Sites 479 and 480. In *Initial reports of the Deep Sea Drilling Project-LXIV.* U.S. Govt. Printing Office, Washington DC.

Calvert S. E. (1966) Accumulation of diatomaceous silica in the sediments of the Gulf of California. *Geol. Soc. Am. Bull.* **77**, 569–596.

Christensen J. H. and Tomasi G. (2016) A multivariate approach to oil hydrocarbon fingerprinting and spill source identification. In *Standard Handbook Oil Spill Environmental Forensics: Fingerprinting and Source Identification, Second Edition* (eds. S. A. Stout and Z. Wang). Academic Press, Massachusetts, pp. 747–788.

Christensen J. H. and Tomasi G. (2007) Practical aspects of chemometrics for oil spill fingerprinting. *J. Chromatogr. A* **1169**(1-2), 1–22.

Christensen J. H., Tomasi G. and Hansen A. B. (2005) Chemical Fingerprinting of Petroleum Biomarkers Using Time Warping and PCA. *Environ. Sci.* **39**(1), 255–260.

Dallüge J., Beens J. and Brinkman, U. A. T. (2003). Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. *J. Chromatogr. A* **1000**(1-2), 69–108.

Dean W. (2006) The geochemical record of the last 17,000 years in the Guaymas Basin, Gulf of California. *Chem. Geol.* **232**, 87–98.

De la Lanza-Espino G. and Soto L.A. (1999) Sedimentary geochemistry of hydrothermal vents in Guaymas Basin, Gulf of California, Mexico. *Appl. Geochem.* **14**(4), 499–510.

Didyk M. and Simoneit B. R. T. (1990) Petroleum characteristics of the oil in a Guaymas Basin hydrothermal chimney. *Appl. Geochem.* **5**, 29–40.

Dimandja J. M. D. (2004) GC x GC. *Anal. Chem.* **76**, 167A–174A.

Dowell, F., Cardman, Z., Dasarathy, S., Kellermann, M., Lipp, J., Ruff, S., Biddle, J., McKay, J., MacGregor, B., Lloyd, K., Albert, D., Mendlovitz, H., Hinrichs,U.-K., Teske, A. (2016) Microbial communities in methane and short chain alkane rich hydrothermal sediments of Guaymas Basin. *Front. Microbiol.* **7**, 1–16.

Dymond J., Collier R.W. and Watwood M.E. (1989) Bacterial mats from Crater Lake, Oregon and their relationship to possible deep-lake hydrothermal venting. *Nature* **342**, 673–675.

Edgcomb V.D., Kysela A., Teske A., de Vera Gomez, Sogin M. L., (2002) Benthic eukaryotic diversity in the Guaymas Basin, a hydrothermal vent environment. *P.N.A.S.* **99**, 7658–7662.

Eglinton T. I. and Douglas A. G. (1988) Quantitative study of biomarker hydrocarbons released from kerogens during hydrous pyrolysis. *Energ. Fuel.* **2**(1), 81–88.

Einsele G., Gieskes J., Curray J., Moore D., Aguayo E., Aubry M., Fornari D., Guerrero J., Kastner M., Kelts K., Lyle M., Matoba Y., Molina-Cruz A., Niemitz J., Rueda J., Saunders A., Schrader H., Simoneit B. and Vacquier V. (1980) Intrusion of basaltic sills into highly porous sediments and resulting hydrothermal activity. *Nature* **283**, 441–445.

Elsgaard L., Isaksen M. F., Jørgensen B. B., Alayse A.-M. and Jannasch H. W. (1994) Microbial sulfate reduction in deep-sea sediments at the Guaymas Basin hydrothermal vent area: Influence of temperature and substrates. *Geochim. Cosmochim. Acta* **58**(16), 3335–3343.

Fisher A. T. and Becker K. (1991) Heat flow, hydrothermal circulation and basalt intrusions in the Guaymas Basin, Gulf of California. *Earth Planet. Sci. Lett.* **103**(1– 4), 84–99.

Fouquet Y., Zierenberg R. A., Miller D. J., Bahr J. M., Baker P. A., Bjerkgården T., Brunner C. A., Duckworth R.C., Gable R., Gieskes J., Goodfellow W. D., Gröschel-Becker H. M., Guèrin G., Ishibashi J., Iturrino G., James R. H., Lakschewitz K. S., Marquez L. L., Nehlig P., Peter J. M., Rigsby C. A., Simoneit B. R. T., Schultheiss P., Shanks W. C., Summit M., Teagle D. A. H., Urbat M. and Zuffa G. G. (1998) *Proceedings of the Ocean Drilling Program, Initial Reports,* Vol. 169. US Government Printing Office, Washington DC.

Frysinger G., Gaines R. and Reddy C. (2002) GC×GC—A new analytical tool for environmental forensics. *Environ. Forensics* **3**(1), 27–34.

Garrigues P., Budzinski H., Manitz M. P. and Wise S. A. (1995) Pyrolytic and petrogenic inputs in recent sediments: A definitive signature through phenanthrene and chrysene compounds distribution. *Polycycl. Aromat. Comp.* **7**(4), 275–284.

Gieskes J., Simoneit B., Shanks W., Goodfellow W., James R., Baker P. and Ishibashi J. (2002) Geochemistry of fluid phases and sediments: relevance to hydrothermal circulation in Middle Valley, ODP Legs 139and 169. *Appl. Geochem*. **17**, 1381–1399.

Goldhaber M. B. (1974) Equilibrium of dynamic aspects of the marine geochemistry of sulphur. Ph.D. Thesis, Univ. of California, Los Angeles.

Guerrero-Barajas C., Garibay-Orijel C. and Rosas-Rocha L. E. (2011) Sulfate reduction and trichloroethylene biodegradation by a marine microbial community from hydrothermal vents sediments. *Int. Biodeterior. Biodegr* **65**(1), 116–123.

Hashemi-Nasab F. and Parastar H. (2019) Pattern recognition analysis of gas chromatographic and infrared spectroscopic fingerprints of crude oil for source identification. *Microchem. J.* **111**, 67–81.

Heftmann E. (2004) *Chromatography: Fundamentals and applications of chromatography and related differential migration methods - Part A: Fundamentals and techniques.* Elsevier, Amsterdam.

Héberger K. (2008) Chemoinformatics—multivariate mathematical–statistical methods for data evaluation. In *Medical Applications of Mass Spectrometry*. Elsevier, Netherlands pp. 141–169.

Hoering T. C. (1968) Reactions of the organic matter in a recent marine sediment. *Carnegie Inst. Wash. Yearbook* **67**, 199–201.

Holden J. F. (2009) Extremophiles: Hot Environments. In *Encyclopedia of Microbiology*, Academic Press, Massachusetts, pp. 127–146.

Horsfield B. (1984) Pyrolysis studies and petroleum exploration. In *Advances in Petroleum Geochemistry* (eds. J. Brooks and D. H. Welte). Academic Press, London, pp. 247–292.

Hunt J. M. (1979) *Petroleum Geochemistry and Geology*. W.H. Freeman and Co., California.

Ishibashi J. I., Sato M., Sano Y., Wakita H., Gamo T. and Shanks W. C. (2002) Helium and carbon gas geochemistry of pore fluids from the sediment-rich hydrothermal system in Escanaba Trough. *Appl. Geochem.* **17**, 1457–1466.

Ishiwatari R., Ishiwatari M., Rohrback B. and Kaplan I. (1977) Thermal alteration experiments on organic matter from recent marine sediments in relation to petroleum genesis. *Geochim. Cosmochim. Acta,* **41**(6), 815–828.

Jannasch H., Nelson D. and Wirsen C. (1989) Massive natural occurrence of unusually large bacteria (*Beggiatoa* sp.) at a hydrothermal deep-sea vent site. *Nature*. **342**, 834–836.

Jannasch, H., Taylor, C. (1984) Deep-Sea Microbiology. *Annu. Revs. Microbiol.* **38**, 487–514.

Judd A. and Hovland M. (2007) *Seabed Fluid Flow: The impact on geology, biology and the marine environment*. Cambridge University Press, UK.

Kalil E. K. (1976) The distribution and geochemistry of uranium in Recent and Pleistocene sediments. Ph.D. Thesis, Univ. of California, Los Angeles.

Karl D. M., Taylor G. T., Novitsky J. A., Jannasch H. W., Wirsen C. O., Pace N. R., Lane D. J., Olsen G. J. and Giovannoni S. J. (1988) A microbiological study of Guaymas Basin high temperature hydrothermal vents. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **35**(5), 777–791.

Kawka E. (1990) Hydrothermal alteration of sedimentary organic matter in Guaymas Basin, Gulf of California. Ph.D. Thesis, Oregon State Univ. Corvallis.

Kawka E. and Simoneit B. (1987) Survey of hydrothermally-generated petroleums from the Guaymas Basin spreading center. *Org. Geochem.* **11**, 3 1 l–328.

Kawka E., Simoneit, B. (1990) Polycyclic aromatic hydrocarbons in hydrothermal petroleums from the Guaymas Basin spreading center. *Appl. Geochem*. **5**, 17–27.

Kawka E. and Simoneit B. (1994) Hydrothermal pyrolysis of organic matter in Guaymas Basin: I. Comparison of hydrocarbon distributions in subsurface sediments and seabed petroleums. *Org. Geochem.* **22**(6), 947–978.

Kelts K., Curray J. and Moore, D. (1982) Introduction and explanatory notes, DSDP leg 64 (California Basins) In *Initial reports of the Deep Sea Drilling Project-LXIV*. U.S. Govt. Printing Office, Washington DC.

Killops S. D. and Al-Juboori, M. A. H. A. (1990) Characterization of the unresolved complex mixture (UCM) in the gas chromatograms of biodegraded petroleums. *Org. Geochem.* **15**, 147–160.

Kluesner J., Lonsdale P. and González-Fernández A. (2014) Late Pleistocene cyclicity of sedimentation and spreading-center structure in the Central Gulf of California. *Mar. Geol.* **347**, 58–68.

Kowalski B. R. (1977) *Chemometrics: Theory and Application*. ACS Publications, Washington.

Kowalski, B. R. (1975). Chemometrics: Views and Propositions. Journal of Chemical Information and Modeling, 15(4), 201–203.

Kvenvolden K. and Simoneit B. (1990) Hydrothermally derived petroleum: Examples from Guaymas Basin, Gulf of California and Escanaba Trough, Northeast Pacific. *A.A.P.G. Bull.* **74**, 223–237.

Larson R. L., Menard H. W. and Smith S. M. (1968) Gulf of California: A Result of Ocean-Floor Spreading and Transform Faulting. *Science* **161**(3843), 781–784.

Leif R. N. and Simoneit, B. R. T. (1995) Confined-pyrolysis as an experimental method for hydrothermal organic synthesis. *Origins Life Evol. B.* **25**(5), 417–429.

Liu X-. K., Birgel D., Elling F. J., Sutton P. A., Lipp J. S., Zhu R., Zhang C., Könneke M., Peckmann J., Rowland S. J., Summons R. E. and Hinrichs K-. U. (2016) From ether to acid: a plausible degradation pathway of glycerol dialkyl glycerol tetraethers. *Geochim. Cosmochim. Acta* **183**,138–152.

Lewan M. D. (1985) Evaluation of petroleum generation by hydrous pyrolysis experimentation. *Phil Trans. Roy. Soc. London, Series A* **315**, 123-134.

Lewan M. (1998) Sulfur-radical control on petroleum formation rates. *Nature* **391**, 164–166.

Lewan M. D., Bjorøy M. and Dolcater D. L. (1986) Effects of thermal maturation on steroid hydrocarbons as determined by hydrous pyrolysis of Phosphoria Retort Shale. *Geochim. Cosmochim. Acta* **50**, 1977–1987.

Lonsdale P., Bischoff J. L., Burns V. M., Kastner M. and Swccney R. E. (1980) A high-temperature thermal deposit on the seabed at a Gulf of California spreading center. *Earth Planet. Sci. Lett.* **49**, 8–20.

Lonsdale P. and Becker K. (1985) Hydrothermal plumes, hot springs, and conductive heat flow in the southern trough of Guaymas Basin. *Earth Planet. Sci. Lett*. **73**, 211–225.

Lonsdale P. and Lawver L. (1980) Immature plate boundary zones studied with a submersible in the Gulf of California. *Geol. Soc*. **91**(9), 555.

Magenheim A. J. and Gieskes J. M. (1992) Hydrothermal discharge and alteration in near-surface sediments from the Guaymas Basin, Gulf of California. *Geochim. Cosmochim. Acta,* **56**(6), 2329–2338.

Marriott P. J., Chin S.-T., Maikhunthod B., Schmarr H.-G. and Bieri S. (2012). Multidimensional gas chromatography. *Trends Analyt. Chem*. **34**, 1–21.

Marshall A.G. and Rodgers R.P. (2003) Petroleomics: The next grand challenge for chemical analysis. *Acc. Chem. Res.* **37**, 53–59.

Marshall A. G., Hendrickson C. L. and Jackson G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17**(1), 1–35.

Martens C. S. (1990) Generation of short chain organic acid anions in hydrothermally altered sediments of the Guaymas Basin, Gulf of California. *Appl. Geochem*. **5**, 71–76.

Massart D.L. and Buydens L. (1988) Chemometrics in pharmaceutical analysis. *J. Pharm. Biomed. Anal. **6**, 535-545.*

McCollom T., Simoneit B. and Shock E. (1999) Hydrous pyrolysis of polycyclic aromatic hydrocarbons and implications for the origin of PAH in hydrothermal petroleum. *Energ. Fuel.* **13** (2), 401–410.

McDermott J. M., Ono S., Tivey M. K., Seewald J. S., Shanks W. C. and Solow A. R. (2015) Identification of sulfur sources and isotopic equilibria in submarine hot-springs using multiple sulfur isotopes. *Geochim. Cosmochim. Acta* **160**, 169–187.

McKay L., MacGregor B., Biddle J., Albert D., Mendlovitz H., Hoer D., Lipp J., Lloyd K. and Teske A. (2012) Spatial heterogeneity and underlying geochemistry of phylogenetically diverse orange and white Beggiatoa mats in Guaymas Basin hydrothermal sediments. *Deep-Sea Res.* **67**, 21–31.

Meglen R. R. (1988). Chemometrics: Its role in chemistry and measurement sciences. *Chemometr. Intell. Lab.* **3**(1-2), 17–29.

Michaelis W., Jenisch A. and Richnow H. H. (1990) Hydrothermal petroleum generation in Red Sea sediments from the Kebrit and Shaban deeps. *Appl. Geochem.* **5**(1-2), 103–114.

Mogollón N., Ribeiro F., Poppi R., Quintana A., Chávez J., Agualongo D., Aleme H. and Augusto F. (2017) Exploratory analysis of biodiesel by combining comprehensive two-dimensional gas chromatography and multiway principal component analysis. *J. Braz. Chem. Soc*. **28**(5), 740–746.

Moore D. (1973) Plate-edge deformation and crustal growth, Gulf of California Structural Province. *Geol. Soc. Am. Bull.* **84**(6), 1883–1905.

Mullins O. C., Rodgers R. P., Weinheber P., Klein G. C., Venkataramanan L., Andrews A. B. and Marshall A. G. (2006) Oil reservoir characterization via crude oil analysis by downhole fluid analysis in oil wells with visible−Near-Infrared Spectroscopy and by Laboratory Analysis with electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Energ. Fuel.* **20**(6), 2448–2456.

Nelson D., Wirsen C. and Jannasch H. (1989) Characterization of large, autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas Basin. *Appl. Environ. Microbiol.* **55**, 2909–2917.

Pavón J. L. P., Peña A. G., Pinto C. G. and Cordero B. M. (2006) Differentiation of types of crude oils in polluted soil samples by headspace-fast gas chromatography–mass spectrometry. *J. Chromatogr. A* **1137**(1), 101–109.

Peters K. E. and Fowler M.G. (2002) Applications of petroleum geochemistry to exploration and reservoir management. *Org. Geochem*. **33**(1), 5–36.

Peters K., Walters C. and Moldowan J. M. (2007) *The Biomarker Guide: Biomarkers and Isotopes in Petroleum Systems and Earth History*, Vol. 2. Cambridge University Press, UK.

Peters K., Walters C. and Moldowan, J. (2005) *The Biomarker Guide. Biomarkers and Isotopes in the Environment and Human History,* Vol. 1. Cambridge University Press, UK.

Peters K., Wright T., Ramos L., Zumberge J. and Magoon L. (2016) Chemometric recognition of genetically distinct oil families in the Los Angeles basin, California *A.A.P.G. Bull.* **100**, 115–135.

Phillips J. B. and Beens J. (1999). Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions. *J. Chromatogr. A* **856**(1-2), 331–347.

Phillips J. B. and Ledford E. B. (1996). Thermal modulation: A chemical instrumentation component of potential value in improving portability. *Field Anal. Chem. Technol.* **1**(1), 23–29.

Prata P. S., Alexandrino G. L., Mogollón N. G. S. and Augusto F. (2016) Discriminating Brazilian crude oils using comprehensive two-dimensional gas chromatography–mass spectrometry and multiway principal component analysis. *J. Chromatogr. A* **1472**, 99–106.

Reddy C. M., Eglinton T. I., Hounshell A., White H. K., Xu L., Gaines R. B. and Frysinger G. S. (2002) The West Falmouth oil spill after thirty years: the persistence of petroleum hydrocarbons in marsh sediments. *Environ. Sci. Technol.* **36**, 4754-4760.

Reeves E. (2010) Laboratory and Field-Based Investigations of Subsurface Geochemical Processes in Seafloor Hydrothermal Systems. Ph. D. thesis, Massachusetts Institute of Technology.

Rowland S. J. and Maxwell J. R. (1984) Reworked triterpenoid and steroid hydrocarbons in a recent sediment. *Geochim. Cosmochim. Acta* **48**, 617–624.

Rushdi A. I. and Simoneit B. R. T. (2002a) Hydrothermal alteration of organic matter in sediments of the northeastern Pacific Ocean: Part 1. Middle Valley, Juan de Fuca Ridge. *Appl. Geochem*. **17**, 1401–1428.

Rushdi A. I. and Simoneit B. R. T. (2002b) Hydrothermal alteration of organic matter in sediments of the northeastern Pacific Ocean: Part 2. Escanaba Trough, Gorda Ridge. *Appl. Geochem*. **17**, 1467–1494.

Rusnak G., Fisher R. and Shepard F. (1964) Bathymetry and Faults of Gulf of California. in, Andel, Tj. H. Van, And Shor, G. G., eds., Marine geology of the Gulf of California: *Am. Assoc. Petroleum Geologists Mem*. **3**, 59–75.

Sancetta C. (1995) Diatoms in the Gulf of California: seasonal flux patterns and the sediment record for the last 15,000 years. *Paleoceanography*, **10**(1), 67–84.

Sarker S. D. and Nahar L. (2015). Applications of High Performance Liquid Chromatography in the Analysis of Herbal Products. In *Evidence-Based Validation of Herbal Medicine.* Elsevier, Netherlands pp. 405–425.

Schomburg G. (1995) Two-dimensional gas chromatography: Principles, instrumentation, methods. *J. Chromatogr. A*. **703**(1-2), 309–325.

Selley R. (1997) *Elements of Petroleum Geology*. Academic Press, Massachusetts.

Shellie R. A. (2013). Gas Chromatography. In *Encyclopedia of Forensic Sciences* (eds. J. A. Siegel, P. J. Saukko and M. M. Houck). Academic Press, Massachusetts, pp. 579–585.

Shellie R., Marriott P. J. and Morrison P. (2001) Concepts and preliminary observations on the triple dimensional analysis of complex volatile samples by using GC GC-TOFMS. *Anal. Chem.* **73**, 1336–1344.

Simoneit B. R. T. (1990) Petroleum generation, an easy and widespread process in hydrothermal systems: an overview. *Appl. Geochem*. **5**(1-2), 3–15.

Simoneit B. R. T. (1993a) Aqueous high-temperature and high-pressure organic geochemistry of hydrothermal vent systems. *Geochim. Cosmochim. Acta*. **57**, 3231–3243.

Simoneit B. R. T. (1993b) Hydrothermal activity and its effects on sedimentary organic matter. In *Bitumens in Ore Deposits* (eds. J. Parnell, H. Kucha and P. Landais) Springer-Verlag, Berlin, pp. 81–96.

Simoneit B. R. T. (1982) Organic geochemistry, Leg 64: Introduction and summary. In *Initial Reports of the Deep Sea Drilling Project, Vol. 64, Part II*. U.S. Government Printing Office, Washington, pp. 717–721.

Simoneit B. R. T. (1984) Hydrothermal effects on organic matter-high versus low temperature components. *Org. Geochem.* **6** 857–864.

Simoneit B. R. T. (1985) Hydrothermal petroleum: genesis, migration, and deposition in Guaymas Basin, Gulf of California. *Can. J. Earth Sci*. **22**, 1919–1929.

Simoneit B. R. T. and Fetzer J. C. (1996) High molecular weight polycyclic aromatic hydrocarbons in hydrothermal petroleums from the Gulf of California and northeast Pacific Ocean. *Org. Geochem.* **24**, 1065–1077.

Simoneit B. R. T. and Galimov E. M. (1984) Geochemistry of interstitial gases in Quaternary sediments of the Gulf of California. *Chem. Geol.* **43**, 151–166.

Simoneit B. R. T., Grimalt J. O., Hayes J. and Hartman H. (1987) Low temperature hydrothermal maturation of organic matter in sediments from the Atlantis II Deep, Red Sea. *Geochim. Cosmochim. Acta* **51**(4), 879–894.

Simoneit B. R. T., Leif R. N., Sturz A. A., Sturdivant A. E. and Gieskes J. M. (1992) Geochemistry of shallow sediments in Guaymas Basin, Gulf of California: hydrothermal gas and oil migration and effects of mineralogy. *Org. Geochem.* **18**, 765–784.

Simoneit B. R. T. and Lonsdale P. (1982) Hydrothermal petroleum in mineralized mounds at the seabed of Guaymas Basin. *Nature* **295**, 198–202.

Simoneit B. R. T. Mazurek M., Brenner S., Crisp P. and Kaplan I. (1979) Organic geochemistry of recent sediments from Guaymas Basin, Gulf of California. *Deep Sea Res. Part I Oceanogr. Res. Pap*. **26**, 879–891.

Simoneit B. R. T, Schoell M. and Kvenvolden K. (1997) Carbon isotope systematics of individual hydrocarbons in hydrothermal petroleum from Escanaba Trough, Northeastern Pacific Ocean. *Org. Geochem.* **26**, 511–515.

Teske A. (2009) Deep-Sea Hydrothermal Vents. In *Encyclopedia of Microbiology*. Academic Press, Massachusetts, pp. 80–90.

Teske A., de Beer D., McKay L., Tivey M., Biddle J., Hoer D., Lloyd K., Lever M., Roy H., Albert D., Mendlovitz H. and MacGregor B. (2016) The Guaymas Basin hiking guide to hydrothermal mounds, chimneys, and microbial mats: complex seafloor expressions of subsurface hydrothermal circulation. *Front. Microbiol.* **7**, 75–98.

Teske A., Hinrichs K.-U., Edgcomb V. P., de Vera Gomez A., Kysela D., Sylva S. P., Sogin M. and Jannasch H. (2002) Microbial diversity in hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl. Environ. Microbiol.* **68**, 1994–2007.

Thunell, R. C. (1998) Seasonal and annual variability in particle fluxes in the Gulf of California: a response to climate forcing. *Deep-Sea Res. I* **45**, 2059–2083.

Tissot B., Califet-Debyser Y., Deroo G. and Oudin J. L. (1971) Origin and evolution of hydrocarbons in early Toarcian shales, Paris Basin, France. *A.A.P.G. Bull.* **55**, 2177–2193.

Tissot B. and Welte D. (1984) *Petroleum Formation and Occurrence.* Springer-Verlag, Berlin.

Tran T., Logan G., Grosjean E., Ryan D. and Marriott P. (2010) Use of comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for the characterization of biodegradation and unresolved complex mixtures in petroleum. *Geochim. Cosmochim. Acta* **74**, 6468–6484.

Van Andel Tj. H. (1964) Recent marine sediments in the Gulf of California. In *Marine Geology of the Gulf of California: a symposium* (eds. Tj. H. van Andel and G. G. Shor Jr.) American Association of Petroleum Geologists, Oklahoma. pp. 216–311.

Van Deursen M., Beens J., Reijenga J., Lipman P., Cramers C. and Blomberg J. (2000) Group-Type Identification of Oil Samples Using Comprehensive Two-

Dimensional Gas Chromatography Coupled to a Time-of-Flight Mass Spectrometer (GC×GC-TOF). *J. High Resolut. Chromatogr.* **23**(7-8), 507–510.

Van Dover C. L. (2000) *The ecology of deep-sea hydrothermal vents*. Princeton University Press, Princeton.

Van Graas G., de Leeuw J., Schenck P. and Haverkamp J. (1981) Kerogen of Toarcian shales of the Paris Basin. A study of its maturation by flash pyrolysis techniques. *Geochim. Cosmochim. Acta* **45**(12), 2465–2474.

Ventura G. T., Hall G. J., Nelson R. K., Frysinger G. S., Raghuraman B., Pomerantz A. E., Mullins O. C. and Reddy C. M. (2011) Analysis of petroleum compositional similarity using multiway principal components analysis (MPCA) with comprehensive two-dimensional gas chromatographic data. *J. Chromatogr. A* **1218**(18), 2584–2592.

Ventura G. T., Kenig F., Reddy, C. M., Schieber J., Frysinger G. S., Nelson R. K., Dinel E., Gaines R. B. and Schaeffer P. (2007) Molecular evidence of Late Archean archaea and the presence of a subsurface hydrothermal biosphere. *P.N.A.S.* **104**(36), 14260–14265.

Ventura G. T., Kenig F., Reddy C. M., Frysinger G. S., Nelson R. K., Mooy B. V., and Gaines R. B. (2008) Analysis of unresolved complex mixtures of hydrocarbons extracted from Late Archean sediments by comprehensive two-dimensional gas chromatography (GC×GC). *Org. Geochem.* **39**(7), 846–867.

Ventura G. T., Simoneit B. R. T., Nelson R. K. and Reddy C. M. (2012) The composition, origin and fate of complex mixtures in the maltene fractions of hydrothermal petroleum assessed by comprehensive two-dimensional gas chromatography. *Org. Geochem.* **45**, 48–65.

Vigneron A., Cruaud P., Pignet P., Caprais J. C., Gayet N., Cambon−Bonavita M. A., Godfroy A. and Toffin L. (2014) Bacterial communities and syntrophic associations involved in anaerobic oxidation of methane process of the Sonora Margin cold seeps, Guaymas Basin. *Environ. Microbiol*. **16**, 2777–2790.

Von Damm K. L., Edmond J. M., Measures C. I. and Grant B. (1985) Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochim. Cosmochim. Acta* **49**(11), 2221–2237.

Wang Z. and Stout S. (2007) *Oil Spill Environmental Forensics. Fingerprinting and Source Identification.* Elsevier, Amsterdam.

Wang Y., Chen Q., Norwood D. L. and McCaffrey J. (2010) Recent Development in the Applications of Comprehensive Two-Dimensional Gas Chromatograph. *J. Liq. Chromatogr. Relat. Technol.* **33**(9-12), 1082–1115.

Wang Y., Peters K., Moldowan J., Bird K. and Magoon L. (2014) Cracking, mixing, and geochemical correlation of crude oils, North Slope, Alaska. *A.A.P.G. Bull.* **98**, 1235–1267.

Wang Y.-P., Zou Y.-R., Shi J.-T. and Shi J. (2018) Review of the chemometrics application in oil-oil and oil-source rock correlations. *J. Nat. Gas. Geo*. **3**, 217–232.

Warton B., Alexander R. and Kagi R. I. (1999) Characterization of the ruthenium tetroxide oxidation products from the aromatic unresolved complex mixture of a biodegraded crude oil. Org. Geochem. **30**, 1255–1272.

Weber A. and Jørgensen B. (2002) Bacterial sulfate reduction in hydrothermal sediments of the Guaymas Basin, Gulf of California, Mexico. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **49**(5), 827–841.

Welhan J. A. and Lupton J. E. (1987) Light hydrocarbon gases in Guaymas Basin hydrothermal fluids: thermogenic versus abiogenic origin. *A.A.P.G. Bull.* **71**, 215–223.

Whelan J. K., Simoneit B. R. T. and Tarafa M. E. (1988) C1-C8 hydrocarbons in sediments from Guaymas Basin, Gulf of California – Comparison to Peru Margin, Japan Trench, and California Borderlands. *Org. Geochem*. **12**, 171–194.

Williams D., Becker K., Lawver L. and Von Herzen R. (1979) Heat flow at the spreading centers of the Guaymas Basin, Gulf of California. *J. Geophys. Res.* **84**, 6757–6796.

Wold S. (1995). Chemometrics; what do we mean with it, and what do we want from it? *Chemometr. Intell. Lab*. **30**(1), 109–115.

Zárate-del Valle P. F., Rushdi A. I. and Simoneit B. R. T. (2006) Hydrothermal petroleum of Lake Chapala, Citala Rift, western Mexico: Bitumen compositions from source sediments and application of hydrous pyrolysis. *Appl. Geochem.* **21**(4), 701–712.

Zeng H., Zou F., Lehne E., Y., J. and Zhang D. (2012) Gas Chromatograph Applications in Petroleum Hydrocarbon Fluids. In *Advanced Gas Chromatography - Progress in Agricultural, Biomedical and Industrial Applications* (eds. M. A. Mohd). InTech, Croatia, pp. 363–388.

Zierenberg R. A., Miller D. J. (2000) Overview of ocean drilling program leg 169:

Sedimented ridges II. In *Proceedings of the Ocean Drilling Program, Scientific Results*, Vol. 169. ODP, Texas.

Zhan Z.-W., Lin X.-H., Zou Y.-R., Li Z., Wang D., Liu C. and Peng P. (2018) Chemometric differentiation of crude oil families in the southern Dongying Depression, Bohai Bay Basin, China. *Org. Geochem.* **127**, 37–49.

Zhuang G.-C., Montgomery A., Samarkin V. A., Song M., Liu J., Schubotz F., Teske A., Hinrichs K.-U. and Joye S. B. (2019) Generation and utilization of volatile fatty acids and alcohols in hydrothermally altered sediments in the Guaymas Basin, Gulf of California. *Geophys. Res. Lett.* **46**(5), 2637–2646.

Chapter 2: Thermochemical and Microbial Biomarker Transformations in Hydrothermally-Altered Sediments at Cathedral Hill, Guaymas Basin, Gulf of California

Connor J. Dalzell[1], G. Todd Ventura[1*], Robert K., Nelson[2], Clifford C. Walters[3], Christopher M. Reddy[2], Jeffrey S. Seewald[2], Stefan M. Sievert[2]

Affiliations:

[1] Saint Mary's University, Canada.

[2] Woods Hole Oceanographic Institution, USA.

[3] ExxonMobil Research and Engineering, USA.

[*] Corresponding author: todd.ventura@smu.ca

Target Journal: Organic Geochemistry

Abstract

Accelerated oil production naturally occurs in hydrothermal environments, where circulating high-temperature fluids rapidly pyrolyze sedimentary organic matter into petroleum products. The Guaymas Basin in the Gulf of California is one such area of significant hydrothermal activity, where evidence of active oil formation in the uniform, rapidly accumulating hemipelagic sediments has been documented within 12–30 meters below the seafloor. However, studies near active vent sites at Guaymas Basin have reported ~120 °C temperatures within 50 cm of the seabed, which is well within the threshold predicted for hydrothermal alterations of organic matter. As such, active petroleum formation may be occurring in shallow sediments proximal to hydrothermal vents, which could record the progression of immature sedimentary organic matter into petroleum in a stepwise fashion (i.e. as a function of heat/distance to the vent center). Equally, the range and activity of subsurface microorganisms operating within these systems may be shown in a similar fashion. To test these hypotheses, a push core transect of petroleum-impregnated sediments was collected from the Cathedral Hill hydrothermal vent site in the Southern Trough of the Guaymas Basin. At this site, projected porewater temperatures range from 0-153 °C within 20 cm of the seafloor. The apolar fraction of these seabed hydrocarbons were analyzed using comprehensive two-dimensional gas chromatography (GC×GC), which has enhanced resolving power and is better suited for analysis of biodegraded oils compared to traditional gas chromatography. A systematic survey of biomarkers and homologous series reveals increasing levels of sediment thermal maturity with depth and distance to the hydrothermal vent edifice. However, conflicting thermal maturity proxies suggests these results reflect variable temperature fluctuations or staining from oil migration. Microbial activity is pervasive in the subsurface, with evidence of an

active biosphere in almost all sediments within the temperature limit of life. Bacterial life may also persist in sediments beyond the 80 °C temperature threshold, alluding to a zone of microbial "metastability," where cells linger intact before their eventual demise. All together, these results describe a complex and rapidly evolving system, where sediment transformations are apparent on the scale of centimeters.

2.1 Introduction

There are few places outside sedimented hydrothermal systems where natural thermal maturation trends can be observed entirely from beginning to end. In these locales, the conversion of young organic matter (OM) into petroleum rapidly occurs via hydrous pyrolysis of sedimentary biomaterial by circulating high-temperature fluids (Didyk and Simoneit, 1989, 1990; Simoneit and Lonsdale, 1982; Simoneit, 1984; Kvenvolden et al., 1986; Kvenvolden and Simoneit, 1990; Simoneit, 1993), involving short generation times (i.e. days-years) in relatively shallow sediments (i.e. within 30 m below the seafloor; Simoneit and Lonsdale, 1982; Simoneit, 1985, 1993; Peter et al., 1991; Simoneit and Kvenvolden, 1993). Embedded in these systems are active and diverse microbial communities, which can exhibit high rates of heterotrophy (Edgcomb et al., 2002; Biddle et al., 2012; Liu et al., 2016) and have been shown to significantly affect the compositions of hosted seabed oils (e.g. Kawka and Simoneit, 1987, 1990, 1994; Bazylinski et al., 1988; Kawka, 1990; Marchand et al., 1994). As such, these hydrothermal systems provide an opportunity to capture a complete thermal maturation trend in the same conditions where subsurface microbial hydrocarbon degradation is happening, allowing both processes to be captured in tandem.

Hydrothermal fluid flow is the primary mechanism controlling generation and transport of petroleum in these systems (Didyk and Simoneit 1989, 1990 Simoneit, 1993). This process is further promoted by fluid chemistry, which contains high amounts of CH_4, CO_2 (Simoneit and Galimov 1984; Von Damm et al., 1985; Welhan and Lupton 1987; Whelan et al., 1988) and metals (Fouquet et al., 1998; Zierenberg and Miller, 2000) that aid in the creation and solubilisation of the pyrolysate. Hydrothermal alteration of OM occurs at ~60 to >400 °C via breaking of C-C bonds (Tissot and Welte, 1984; Simoneit, 1993; McCollom et al., 1999; Peters et al., 2007). Higher temperatures can also promote thermal resynthesis (i.e. bond reformation) of polycyclic aromatic compounds (PAHs) (Kawka and Simoneit, 1990; Simoneit, 1993). If entrained in circulating hydrothermal fluids, these oils are often discharged into the water column or accumulate near the seabed, due to quenching via interaction with cooler ambient seawaters (Simoneit, 1984, 1993; Kawka, 1990; Peter et al. 1990).

Active hydrocarbon formation has been reported within 12-30 m below the seabed at Guaymas Basin (Peter et al. 1991). However, surveys of hydrothermal sediments near active vent sites have revealed temperatures reaching ~120 °C within 50 cm of the seafloor (e.g. Bazylinski et al. 1988; Magenheim and Gieskes, 1992; Elsgaard et al., 1994; Weber and Jørgensen, 2002; McKay et al., 2012; Teske et al., 2016), which is well within the temperature range predicted for OM alterations in hydrothermal systems (~60 to >400 °C, Simoniet 1993). This suggests that active petroleum formation may be occurring over narrow depth intervals in proximity to hydrothermal vents, which could be younger than ~1000 years based on the 1-2 mm/yr sedimentation rate at Guaymas Basin (Williams et al., 1979; Teske et al., 2002). As such, a transect of sediments radiating from an active hydrothermal site may capture a natural, stepwise progression of OM alteration into

petroleum with depth and distance to the vent center (i.e. as a function of temperature). Equally, the effects of biodegradation and microbial activity may be preserved in a similar fashion, which could provide insight to the habitable range and effects of microorganisms within these systems. Therefore, this project will examine the thermal maturity and microbial weathering profiles along such a transect, in order to assess the progression of petrogenesis and microbial activity within these near-surface hydrothermal vent sediments.

2.1.1 Considerations for studying seabed oils

Studies evaluating petroleum characteristics often rely on molecular ratios to infer reaction rates, OM compositions, and thermal histories of oils (e.g. Boehm et al., 1997; Peters et al., 2005, 2007; Wang and Stout, 2007; Christensen et al., 2016). However, this endeavour can be limited by the use of traditional analytical methods, such as capillary gas chromatography, which are incapable of resolving the full complexity of oils (Tissot and Welte, 1984; Marshall et al., 1998; Mullins et al., 2006) and suffer from co-elution that can make identification and quantitation of these diagnostic compounds difficult. This is further complicated by the appearance of unresolved complex mixtures (UCMs) in petroleum, which can manifest in these traditional "1-D" analytical methods if the molecular composition of oils becomes too complex to resolve, which can obscure compounds and is common in samples that are biodegraded (Tissot et al., 1971; Rowland and Maxwell, 1984; Killops and Al-Juboori, 1990; Warton et al., 1999; Reddy et al., 2002). However, these effects can be mitigated using comprehensive two-dimensional gas chromatography (GC×GC), which resolves the hydrocarbon eluent onto two dimensions thus eliminating the problem of co-elution/UCMs and enhancing compound resolution compared to traditional gas chromatography (e.g. Reddy et al., 2002; Frysinger et al., 2002; Ventura et

al., 2007, 2008, 2012; Tran et al., 2010; Barman and Membrado, 2000; Arey et al., 2005, 2007).

2.1.2 Objectives

In this study, a comprehensive molecular survey is conducted on the apolar extracts of hydrothermal petroleum sourced from a pushcore transect at the Cathedral Hill vent site in Guaymas Basin. At this site, thermal probe measurements estimate temperatures spanning 0-155 °C based on sediment depth and proximity to the vent, reaching a maximum of 21 cm below seafloor (cmbsf). Well-known biomarkers and aromatic hydrocarbons that are common in oils are compared to elucidate thermal maturity levels and biodegradation along the natural thermal gradient. Source parameters and other indicators are also documented to further characterize and understand the system. Through this, the process of oil generation, along with a host of secondary weathering processes (i.e. water washing and hydrocarbon biodegradation) are evaluated to provide improved insight into how these exotic petroleum systems operate.

2.2 Geological Setting

The Guaymas Basin spreading center in the Gulf of California is a tectonically active site formed via diverging of the Pacific and North American plates (Rusnak et al., 1964; Larson et al., 1968; Moore et al., 1973; Curray et al., 1979). Heat flow is very high, locally exceeding 1.2Wm^{-2} (Lawver et al., 1975; Curray et al., 1982) with significant hydrothermal activity occurring in the Northern and Southern Troughs of the basin (Lonsdale and Becker 1985; Kawka and Simoneit, 1987). Plate activity promotes intrusion of shallow dikes and sills into the recent hemipelagic sediments (Einsele et al., 1980;

Curray et al., 1982; Fisher and Becker 1991), elevating regional heat flow and creating hydrothermal circulation and alteration of these organic-rich muds (Curray et al., 1982; Lonsdale and Becker, 1985; Kawka and Simoniet, 1987). These hydrothermal processes allow for rapid petroleum genesis via hydrous pyrolysis of hosted organic material (Simoneit and Lonsdale, 1982; Simoneit, 1984, 1990, 1992), with fluids reaching 270–350 °C at chimneys and 50 °C through porous sediments (Lonsdale et al., 1980; Simoneit, 1984; Lonsdale and Becker, 1985).

Cathedral Hill is an active white smoker complex in the Southern Trough of the Guaymas Basin, located 1996 meters deep at 27°00.696 N, 111°24.265 W (Fig. 1). Sedimentation at this site is likely similar to the rest of the basin (i.e. 1–2 mm/yr; Calvert, 1966; Williams et al., 1979; Teske et al., 2002), which is predominantly sourced from terrestrial run-off and highly productive surface waters (Van Andel. 1964; Calvert, 1966; Curray et al., 1982; Simoneit and Lonsdale, 1982; Kelts et al., 1982). These sources, along with periodic turbidity flows (Moore, 1973; Kluesner et. al, 2014) result in uniform compositions comprising diatomaceous oozes and turbidite muds (Curray et al., 1982) that are sparsely mixed with buried calcareous shells and benthic macrofauna. Additionally, giant tube worms (*Riftia pachyptila*) and a sulfide oxidizing Beggiatoa microbial mat inhabit this site (McKay et al., 2012; Teske et al., 2016).Cathedral Hill is similar to other hydrothermal features in the Southern Trough, showing comparable subsurface temperatures (120 °C within 45 cmbsf; Bazylinski et al. 1988; Weber and Jørgensen, 2002; Teske et al., 2016), vent fluid chemistries, and dissolved organic carbon content (Reeves, 2010; McDermott et al., 2015 Zhuang et al., 2019), occupying a heat-flow regime (500-1000 mW m^2) host to many other vent sites (Lonsdale and Becker, 1985; Teske et al., 2016).

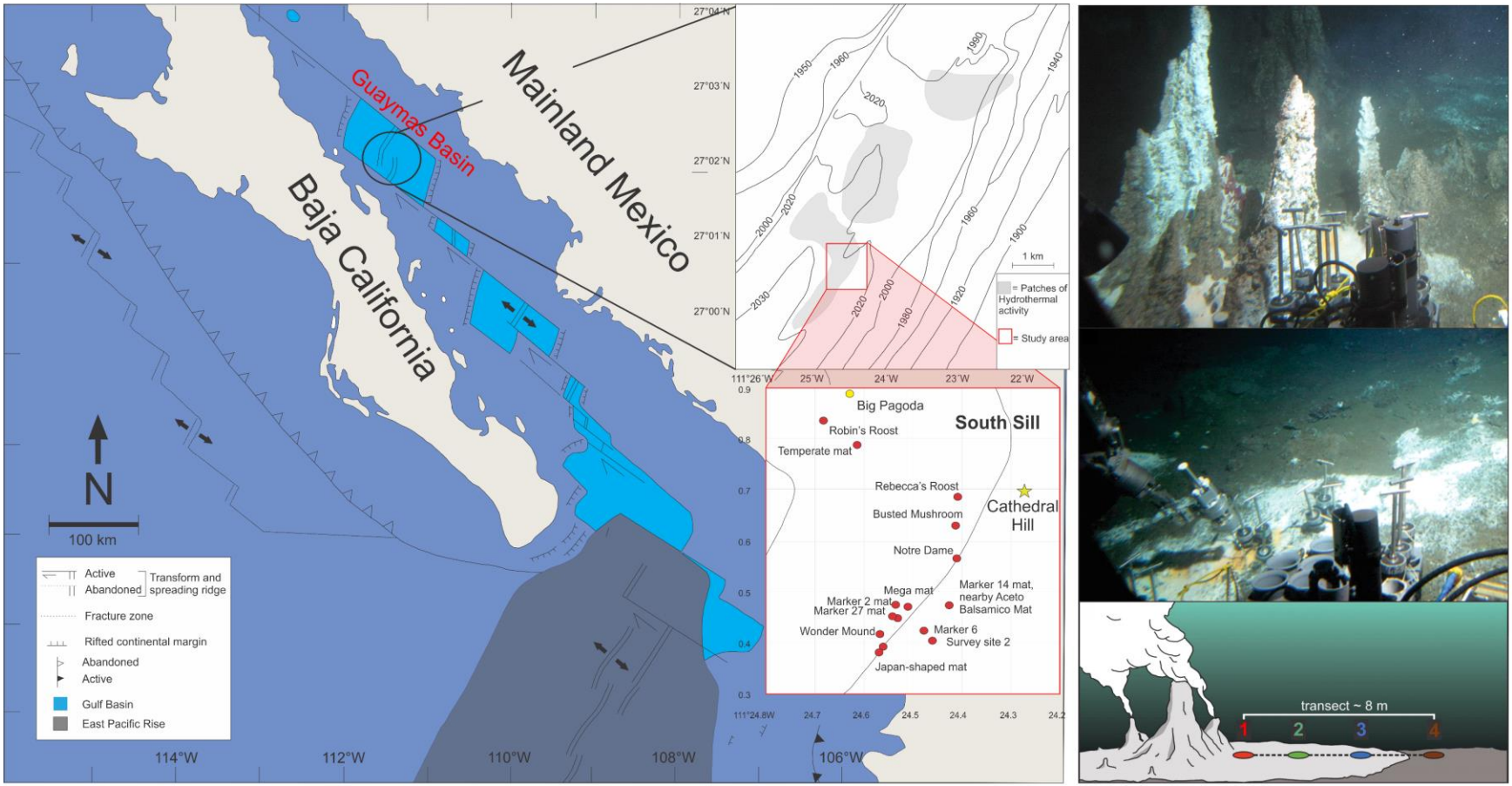


Fig. 1. Overview of Guaymas Basin showing the tectonic regime, Southern Trough, and Cathedral Hill. The red box marks the cluster of vents (red circles) in which Cathedral Hill (yellow star) is located. The right are photos of Cathedral Hill, Alvin push core sampling, and use of the thermal probe along with an illustration depicting the push core transect (modified from Castillo et al., 2002 and Teske et al., 2016).

.

2.3 Methods

2.3.1 Sample collection

Using the human operated vehicle *Alvin* of Woods Hole Oceanographic Institution four near-equidistantly spaced (~2 m) push cores of petroleum-impregnated, unconsolidated sediments were collected along a transect running from the vent center to beyond the edge of the surrounding microbial mat (Fig. 1; Table 1). Measurements of porewater temperatures were collected alongside each push core using a thermal probe (analytical precision ±2 °C) in ~5 cm increments up to ~20 cmbsf (Table 1).

2.3.2 Sample preparation

Once the push cores were brought to the surface, they were immediately sectioned into 2-3 cm increments, transferred to combusted gas vials, and frozen at -40 °C. The samples were later freeze dried and stored at -80 °C. Each sample was extracted three times via a CEM Mars Microwave Digestion System, which had a 15-20 minute ramp time to 115°C, with a hold of 15 min at 1030-1800 V, using a 3:1 mixture of dichloromethane (DCM) and methanol (MeOH). Apolar fractions of the total lipid extracts were collected via flash chromatography, using 60 ml of hexane to rinse the apolar (AP) fraction through a Pasteur pipette column filled with 30 g of activated silica gel (60–200 mesh). Elemental sulphur was then removed by passing the sample through a Pasteur pipette column filled with activated Cu (HCl 6N) powder using hexane as the eluent. Samples were then dried under a continuous flow of dry N_2 at 30 °C.

Table 1. Temperature probe, sediment characteristics, and geochemical extract data of Cathedral Hill samples.

Sample location and name[a]	Alvin Dive # and core ID	Porewater temperature (°C)	Interpolated porewater temperature (°C)	Sulfate (mmol/kg)[1]	Description/lithology[b]	Extracted sediment wt (g)	TLE (mg)	Polar (mg)	Apolar + S^0 (mg)	Apolar (mg)
Microbial mat closest to vent										
Core 1 0-2 cmbsf	GB4462-5	19	19	20.1	Black mud with microbial mat filaments	1.78	50.16	40.31	3.54	2.27
Core 1 2-4 cmbsf	GB4462-5	-	67	10.8	Brownish-green diatomaceous mud	2	47.04	27.38	2.73	1.98
Core 1 4-6 cmbsf	GB4462-5	85	85	3.41	Brownish-green diatomaceous mud	2.68	29.14	19.81	1.25	1.24
Core 1 6-8 cmbsf	GB4462-5	-	105	3.41	Brownish-green diatomaceous mud	3.23	31.76	22.5	1.24	1.04
Core 1 8-10 cmbsf	GB4462-5	-	117	0.736	Brownish-green diatomaceous mud	4.56	28.24	12.61	0.98	1.1
Core 1 10-12 cmbsf	GB4462-5	121, 124	125	-	Grayish-green mud	5.18	17.77	8.49	0.94	0.99
Core 1 12-15 cmbsf	GB4462-5	-	135	0.718	Brownish-green consolidated mud with clay shards	4.34	14.19	9.93	0.94	0.86
Core 1 15-18 cmbsf	GB4462-5	142	145	1.01	Brownish-green consolidated clay	4.13	16.73	16.72	0.81	0.79
Core 1 18-21 cmbsf	GB4462-5	153	153	-	Brownish-green consolidated clay	3.74	9.75	2.31	0.65	0.32
Microbial mat further from vent										
Core 2 0-2 cmbsf	GB4462-6	9, 13	11	13.66	Black mud with microbial mat filaments	2.61	36.78	1.18	1.06	0.45
Core 2 2-4 cmbsf	GB4462-6	-	22	4.099	Black mud with microbial mat filaments	1.77	18.48	3.16	0.53	0.47
Core 2 4-6 cmbsf	GB4462-6	20	20	-	Brownish-green diatomaceous mud	3.01	22.16	2.41	0.53	0.29
Core 2 6-8 cmbsf	GB4462-6	-	47	1.202	Brownish-green diatomaceous mud	3.77	13.87	2.07	0.35	0.26
Core 2 8-10 cmbsf	GB4462-6	-	60	0.61	Brownish-green diatomaceous mud	4.87	11.21	4.25	0.58	0.54
Core 2 10-12 cmbsf	GB4462-6	69, 77	73	-	Brownish-green diatomaceous mud	6.44	9.74	2.46	0.38	0.18
Core 2 12-15 cmbsf	GB4462-6	-	87	-	Brownish-green diatomaceous mud	6.58	11.16	1.99	0.7	0.3

Core 2 15-18 cmbsf	GB4462-6	118	105	2.53	Brownish-green diatomaceous mud	4.21	10.72	5.56	0.69	0.38
Core 2 18-21 cmbsf	GB4462-6	109	125	-	Brownish-green diatomaceous mud	5.16	10.01	2.68	1.25	1
Just inside microbial mat perimeter										
Core 3 0-2 cmbsf	GB4462-3	3.2	3.2	-	Black mud with microbial mat filaments	2.71	38.73	1.8	1.18	0.47
Core 3 2-4 cmbsf	GB4462-3	-	8	16.1	Brownish-green diatomaceous mud	2.89	24.25	2.45	1.24	0.16
Core 3 4-6 cmbsf	GB4462-3	15	15	-	Brownish-green diatomaceous mud	2.04	19.55	2.59	0.72	0.41
Core 3 6-8 cmbsf	GB4462-3	-	26	-	Brownish-green diatomaceous mud	3.57	20.24	1.23	0.75	0.28
Core 3 8-10 cmbsf	GB4462-3	34	34	-	Brownish-green diatomaceous mud	5.08	12.71	2.38	0.82	0.26
Core 3 10-12 cmbsf	GB4462-3	-	43	-	Brownish-green diatomaceous mud	6.55	9.77	4.99	0.23	0.3
Core 3 12-15 cmbsf	GB4462-3	-	54	-	Brownish-green diatomaceous mud	6.6	6.73	1.81	0.58	0.22
Core 3 15-18 cmbsf	GB4462-3	61	66	-	Brownish-green diatomaceous mud	6.46	5.42	3.22	0.26	0.1
Core 3 18-21 cmbsf	GB4462-3	83	80	0.688	Brownish-green diatomaceous mud	3.73	6.94	1.21	0.28	0.23
Just outside microbial mat perimeter										
Core 4 0-2 cmbsf	GB4462-8	0	0	24.2	Black mud	1.89	37.04	1.61	0.47	0.37
Core 4 2-4 cmbsf	GB4462-8	1.5	8	-	Brownish-green diatomaceous mud	2.8	20.21	2.27	1.03	0.47
Core 4 4-6 cmbsf	GB4462-8	16	16	-	Brownish-green diatomaceous mud	3.1	23.05	1.87	0.88	0.28
Core 4 6-8 cmbsf	GB4462-8	-	18	4.01	Brownish-green diatomaceous mud	4.18	20.16	2.99	1.64	0.9
Core 4 8-10 cmbsf	GB4462-8	-	21	-	Brownish-green diatomaceous mud	4.99	13.9	4.44	3.27	1.67
Core 4 10-12 cmbsf	GB4462-8	-	23	-	Brownish-green diatomaceous mud	6.08	10.6	3.63	1.51	0.75
Core 4 12-15 cmbsf	GB4462-8	-	25	0.469	Brownish-green diatomaceous mud	4.42	1.78	0.6	0.24	0.11
Core 4 15-18 cmbsf	GB4462-8	-	27	-	-	-	-	-	-	-
Core 4 18-21 cmbsf	GB4462-8	29	29	-	-	-	-	-	-	-

[1] Pore water sulfate concentrations (with permission from Jeff Seewald, WHOI).

[a] Collected core numbers are relabelled in the sample name to reflect a relative transect position (1-4).
[b] Sediment lithology based on freeze-dried sediments.

2.3.3 Sample analysis

2.3.3.1 Comprehensive two-dimensional gas chromatography (GC×GC)

The S^0-free apolar fractions were analyzed for their hydrocarbon contents using a Leco GC×GC flame ionization detector (GC×GC-FID) equipped with a dual stage cryogenic modulator (Leco, Saint Joseph, Michigan) at Woods Hole Oceanographic Institution. For GC×GC-FID analysis, 1st dimension separation was performed on a nonpolar Restek Rtx-1 Crossbond (20-m, 0.25-mm ID, 0.25-µm film thickness) that was held at 60 °C for 12 min then ramped to 315 °C at 1.5 °C/min, with 2nd dimension separation on a 50% phenyl polysilphenylene-siloxane column (SGE BPX50, 1-m, 0.10-mm i.d., 0.1-µm film thickness) that was held at 70 °C for 12 min and then ramped to 325 °C at 1.5 °C/min. The inlet temperature was 300 °C. The modulation column was pulsed by the hot jet modulator at 60 °C above the temperature of the primary GC oven for 0.4 s every 10.0 s. with a 5.6 s cooling period between stages. The GC was configured for splitless auto-injection, with an inlet temperature of 300 °C, and purge vent opening at 0.5min. Carrier gas (H_2) flow rate was 1mL/min (constant flow mode). The FID detector signal was sampled at a rate of 200 data points s^{-1}.

Mass spectral analysis was performed using an Agilent 7890B Gas Chromatograph, equipped with a ZX1 – LN_2 Cooled Loop Modulator (Zoex Corporation, Houston, Texas), interfaced with an Agilent 7200B quadrupole mass spectrometer (GC×GC-qToFMS) run in electron impact mode at Saint Mary's University. For GC×GC-qToFMS analysis, 1st dimension separation was performed on a nonpolar (5%-phenyl)-methylpolysiloxane phase (Agilent, HP-5ms Ultra Inert, 60-m; 0.25-mm ID; 0.25-µm film thickness), with 2nd dimension separation on a polar 50% phenyl equivalent polysiloxane (BPX-50, SGE, 1.31-

m; 0.10-mm ID; 0.1l-m film thickness) column. The modulation column was deactivated fused silica (1.37-m, 0.10-mm ID), which was pulsed by the hot jet modulator at 200 ºC for 0.4 s every 14.0 s. The GC was configured for 20:1 split auto-injection. The injection inlet temperature was 300 °C, and the purge vent was opened at 0.6 min. Oven temperature ramped from 60 °C to 340 °C at 1.15 °C/min with a carrier gas flow rate of 2.25 mL/min (constant flow mode). The transfer line temperature was 280 °C. The ToF-MS data collection was 50 scans/s at 4 GHz sampling rate.

2.3.3.2 GC×GC post data processing

The GC×GC-FID and GC×GC-qToFMS chromatograms were digitalized using GCImage Version 2.7 software (Zoex, Houston, USA). All samples were baseline subtracted with a rolling ball method and then phase shifted in the same manner to align and maximize a coherent spread of hydrocarbons on the produced chromatogram. Individual hydrocarbons and the range of various compound classes were then identified by GC×GC-ToFMS mass spectral analysis. The identified compounds were translated to GC×GC-FID chromatograms and template matched to other samples.

2.4 Results and Discussion

2.4.1 Geothermal gradients in the vent

A heat-map of the transect (Fig. 2) was created using interpolated thermal probe measurements (Table 1). Cores 1 and 2 record the hottest porewater temperatures, while

cores 3 and 4 at the transect periphery host more ambient conditions. In general, temperatures increase with depth and proximity to the hydrothermal vent edifice.

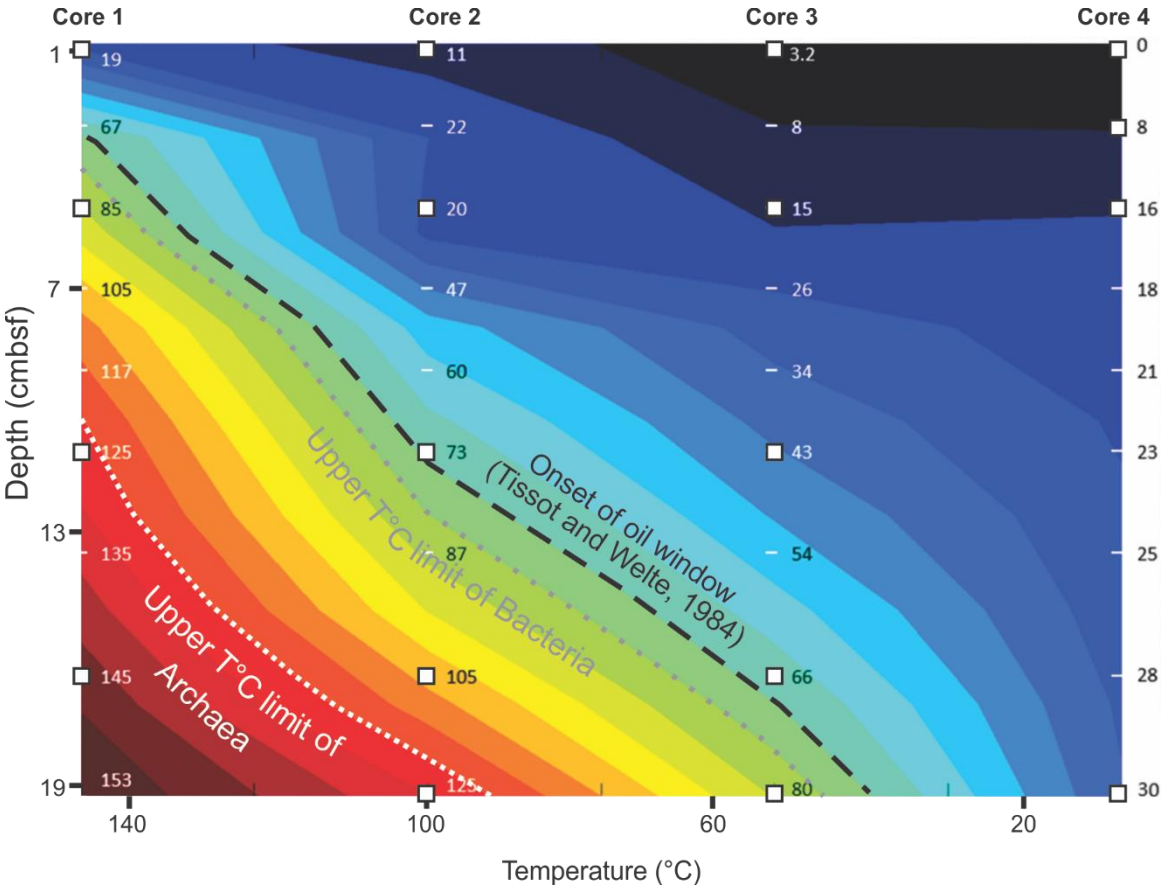


Fig. 2. Thermal heat map of the transect profile. White squares indicate recorded temperatures (Table 1). The onset of oil generation is given by Tissot and Welte (1984) for conventional petroleum basins, while the temperature limit of Archaea is given by Kashefi and Lovley (2003) and bacteria by Holden (2009).

2.4.2 Sediment extract yields

Total lipid extract (TLE) yields are highest proximal to the vent site and at the top of the ocean floor. The extracts progressively decrease with sediment depth (Table 1; Fig. 3). The polar and apolar fractions are elevated/maximized at 6-8 to 10-12 cmbsf in all cores,

though core 1 produces a maximum polar/apolar yield at ~12-18 cmbsf. These elevated yields may coincide with in situ hydrocarbon generation and the migration of more mature oil to the subsurface.

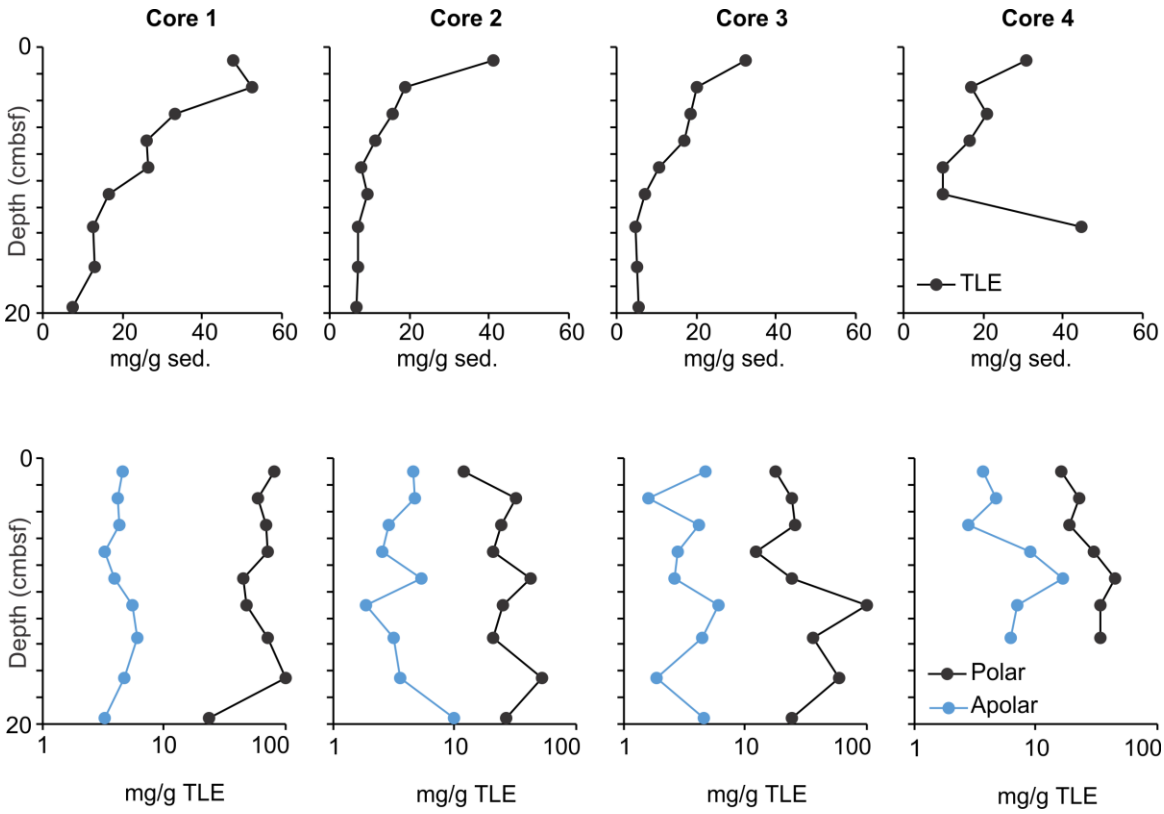


Fig. 3. Recoveries of the total lipid extract (top row) and polar and apolar yields (bottom row).

2.4.3 Petroleum hydrocarbons

A variety of compounds were identified via retention time and elution order as well as by mass spectral analysis (Fig. 4; Table 2). These compounds with their remaining apolar matrix contain up to 5,700 unique hydrocarbons across a range of compound classes that include alkanes, substituted/unsubstituted polycyclic aromatic hydrocarbons (PAHs),

perhydroPAHs, benzothiophenes, and saturate/unsaturated biomarkers (see also Chapter 3).

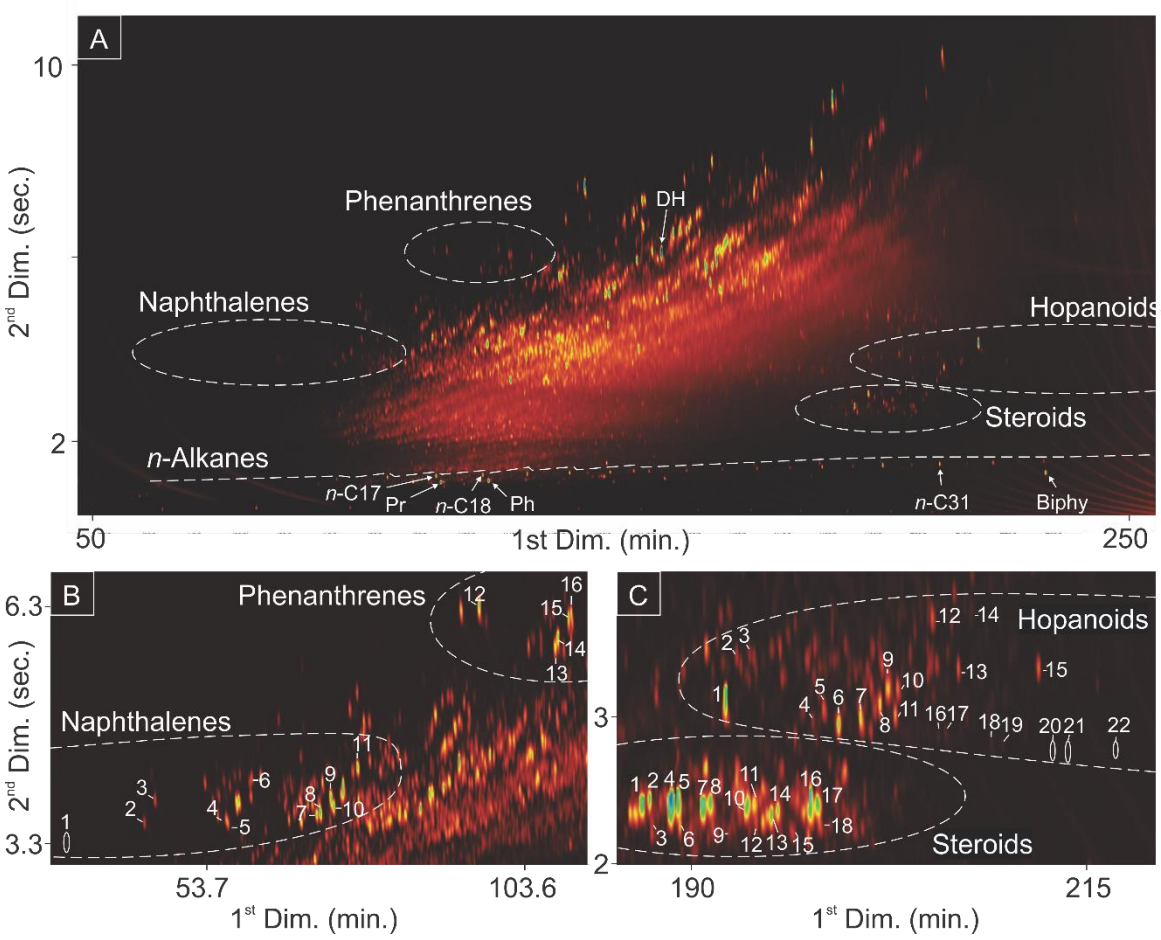


Fig. 4: GC×GC-FID chromatogram with (A) superimposed hydrocarbon groups, including alkanes (normal and branched), naphthalenes, phenanthrenes, hopanoids, and steroids. The biomarkers pristane (Pr), phytane (Ph), *n*-C_{17}, *n*-C_{18}, *n*-C_{31}, Diels hydrocarbon (DH), and biphytane (Biphy) are also indicated. (B) Elution area of naphthalene and phenanthrenes. (C) Elution area of steroids and hopanoids.

Table 2. Compounds identified in the chromatogram shown in Fig. 4.

Peak no. or abbr.	Compound name (steroids)	Formula	M.W.	Diagnostic fragment ions *m/z*	Peak no. or abbr.	Compound name (hopanoids)	Formula	M.W.	Diagnostic fragment ions *m/z*
1	cholest-4-ene	$C_{27}H_{46}$	370	215, 257	1	unknown compound	-	-	-
2	cholest-4,22-diene	$C_{27}H_{44}$	368	215, 257	2	17α(H)-22,29,30-trisnorhopane (Ts)	$C_{27}H_{46}$	370	191, 149
3	5α,14β,17β-cholestane (20R)	$C_{27}H_{48}$	372	217, 259	3	18β(H)-22,29,30-trisnorneohopane (Tm)	$C_{27}H_{46}$	370	191, 149
4	cholest-4-ene	$C_{27}H_{46}$	370	215, 257	4	30-norhopane	$C_{29}H_{50}$	398	191, 205
5	cholest-5-ene	$C_{27}H_{46}$	370	215, 257	5	unknown hopa-diene	$C_{30}H_{48}$	408	187, 202
6	5α,14α,17α-cholestane (20R)	$C_{27}H_{48}$	372	217, 259	6	hop-17(21)-ene	$C_{30}H_{50}$	410	191, 231
7	24-methylcholesta-3,5-diene	$C_{28}H_{46}$	382	213, 255	7	17α,21β-hopane	$C_{30}H_{52}$	412	191
8	24-methylcholest-4,22-diene	$C_{28}H_{46}$	382	215, 257	8	unknown compound	-	-	-
9	5α,14β,17β-ergostane (20R)	$C_{28}H_{50}$	386	217, 259	9	unknown compound	-	-	-
10	24-methylcholest-4-ene	$C_{28}H_{48}$	384	215, 257	10	unknown compound	-	-	-
11	24-methylcholest-5-ene	$C_{28}H_{48}$	384	215, 257	11	17β,21α-moretane	$C_{30}H_{52}$	412	191
12	5α,14α,17α-ergostane (20R)	$C_{28}H_{50}$	386	217, 259	12	unknown compound	-	-	-
13	24-Ethylcholesta-3,5-diene	$C_{29}H_{48}$	396	213, 255	13	unknown compound	-	-	-
14	24-Ethylcholest-4,22-diene	$C_{29}H_{48}$	396	215, 257	14	diploptene	$C_{30}H_{50}$	410	191, 367
15	5α,14β,17β-stigmastane (20R)	$C_{29}H_{52}$	400	217, 259	15	unknown compound	-	-	-
16	24-ethylcholest-4-ene	$C_{29}H_{50}$	398	215, 257	16	17α,21β-homohopane (22S)	$C_{31}H_{54}$	426	191, 205
17	24-ethylcholest-5-ene	$C_{29}H_{50}$	398	215, 257	17	17α,21β-homohopane (22R)	$C_{31}H_{54}$	426	191, 205
18	5α,14α,17α-stigmastane (20R)	$C_{29}H_{52}$	400	217, 259	18	17α,21β-bishomohopane (22S)	$C_{32}H_{56}$	440	191
					19	17α,21β-bishomohopane (22R)	$C_{32}H_{56}$	440	191
					20	17α,21β-trishomohopane (22S)	$C_{33}H_{58}$	454	191
					21	17α,21β-trishomohopane (22R)	$C_{33}H_{58}$	454	191
					22	17α,21β-tetrakishomohopane (22S)	$C_{34}H_{60}$	468	191

Peak no. or abbr.	Compound name (naphthalene and phenanthrenes)	Formula	M.W.	Diagnostic fragment ions *m/z*	Peak no. or abbr.	Compound name (miscellaneous)	Formula	M.W.	Diagnostic fragment ions *m/z*
1	naphthalene	$C_{10}H_8$	128	76	Pr	pristane	$C_{19}H_{40}$	268	57, 183, 253
2	2-methyl naphthalene	$C_{11}H_{10}$	142	115	Ph	phytane	$C_{20}H_{42}$	282	57, 183, 267
3	3-methyl naphthalene	$C_{11}H_{10}$	142	115	Biphy	biphyatne	$C_{40}H_{80}$	561	57, 183, 253, 546
4	2,6-dimethyl naphthalene	$C_{12}H_{12}$	156	141, 127	DH	Diels' hydrocarbon	$C_{18}H_{16}$	232	202, 217
5	2,7-dimethyl naphthalene	$C_{12}H_{12}$	156	141, 127					
6	1,5-dimethyl naphthalene	$C_{12}H_{12}$	156	141, 127					
7	1,3,6-trimethyl naphthalene	$C_{13}H_{14}$	170	155, 141					
8	1,3,7-trimethyl naphthalene	$C_{13}H_{14}$	170	155, 141					
9	1,4,6-trimethyl naphthalene	$C_{13}H_{14}$	170	155, 141					
10	2,3,6-trimethyl naphthalene	$C_{13}H_{14}$	170	155, 141					
11	1,2,5-trimethyl naphthalene	$C_{13}H_{14}$	170	155, 141					
12	phenanthrene	$C_{14}H_{10}$	178	126					
13	3-methyl phenanthrene	$C_{15}H_{12}$	192	177					
14	2-methyl phenanthrene	$C_{15}H_{12}$	192	177					
15	9-methyl phenanthrene	$C_{15}H_{12}$	192	177					
16	1-methyl phenanthrene	$C_{15}H_{12}$	192	177					

2.4.3.1 Normal alkanes and acyclic isoprenoids

n-Alkanes and acyclic isoprenoids were monitored using *m/z* 57 and 183 mass chromatograms, respectively, and identified based on well established elution patterns. All samples contain *n*-alkanes ranging between *n*-C_{11} to *n*-C_{36}, with higher molecular weight homologous displaying distinct odd-over-even carbon number preferences, which indicates contributions of terrigenous OM (i.e. > C_{-24}; Tissot and Welte, 1984) and higher plant waxes. Lower ranged *n*-alkanes (*n*-C_8 to *n*-C_{10}) were likely lost by water washing in the vent site and/or by evaporative lost during sample preparation. The acyclic isoprenoids, pristane (Pr) and phytane (Ph) were present in all samples. These compounds predominantly source from diagenetically altered photosynthetic carotenoid inputs of chlorophyll a (phytol) that has undergone differential oxidative and reductive steps due to the redox conditions within the depositional environment (e.g. Didyk et al., 1978). Pristine/phytane ratios ranged from 0.2 to 1.6 (σ = 0.36), which is low compared to other hydrothermal petroleums (e.g. Kvenvolden and Simoneit, 1990; Simoneit, 1994; Yamanaka et al., 2000). These relatively low values (<1) suggest anoxic (reducing) conditions in the subsurface sediments (Table 4; e.g. Didyk et al., 1978; Peters et al., 2005). Ratios of Pr/*n*-C_{17} and Ph/*n*-C_{18} are elevated (>1; Table 4), which further indicates contributions of terrigenous plant material (Didyk et al. 1978).

2.4.3.2 Tricyclic terpenoids

The tricyclic terpenoids were monitored using *m/z 123* and 191 along with a comparison of their reference elution patterns as published in the literature. These compounds are not particularly abundant. At Cathedral Hill, tricyclic terpanes form a

psuedohomologous series of chelanthanes ranging from C_{19}–C_{29}. These compounds are commonly found in conventional oils and bitumens sourced from more thermally mature source rocks. Additionally, very low abundances of modern terrestrial sourced plant resins such as simonellite and retene were also tentatively identified in low concentrations.

2.4.3.3 Steroids

Steroids are tetracyclic terpenoids formed with Eukaryotes that serve an array of biological function commonly associated with, but not limited to cellular membrane regidification, signaling, and energy stores. Steroidal biomarkers were monitored using *m/z* 213, 215, 217, and 218 mass chromatograms, and tentatively identified by mass spectral analysis in association with their expected elution patterns as described in the literature (Fig. 4, SI Fig. 1-4). The associated detected compounds were then quantitatively measured using GC×GC-FID with identification by template matching.

The tentatively identified steroids were detected and quantified in all of the transect samples (n=34). The C_{27} to C_{29} steroid concentrations were tightly coupled with distributions consistently favoring C_{27}>C_{29}>C_{28} (Fig. 5). The steroids also display a general pattern of increasing concentration with burial (Fig. 5). The C_{27} to C_{29} steroids concentrations progressively increase with depth in cores 1 to 2 (~15-18 cmbsf) before precipitously falling to lower concentrations at the bottom of core 1. For cores 3 and 4 the steroid profiles show more variability, with core 3 having a near uniform low abundance of steroids and core 4 containing a sterane spike at ~8-12 cmbsf (Fig. 5). Collectively, this

indicates that these compounds were likely either tightly linked by a common source organism(s) or by equivalent inputs of sterols into the depositional environment.

During sediment burial, sterols are diagenetically altered to become sterenes. Sterenes then become progressively saturated and stereochemically transformed into thermodynamically more favorable forms of steranes, which makes these compounds both source specific and thermal maturity indicators (Wang et al., 1964; Lloyd et al., 1982; Hazra et al. 2019). As such, a ratio of steranes and sterenes was created to track variation in the relative abundances of steroid as a function of sediment depth (Fig. 6). Down core changes in the ratio are similar for cholest-, ergo-, and stigmasteroids. Collectively these ratios show systematic down core trends. For cores 1 and 2, the sterene/sterane ratio increases until reaching ~10 cmbsf. At this point the ratio produces and inflection point and becomes at or near the same values with depth for core 1 and with slightly lower values in core 2 suggesting higher conversion of sterenes to steranes within these hotter regions of the studied transect site. In the much more ambient core 3, increasing sterene abundance likely signals a more gradual diagenetically degraded sterol concentrations with depth. Together, these results indicate increasing levels of diagenesis and catagenesis with sediment depth and proximity to the vent-center. In core 4, a sharp decrease occurs from 6-8 to 10-12 cmbsf, owed to a high input of steranes at this structural level that appears independent of sterene abundance (Fig. 6), which may indicate migration of a "mature" oil.

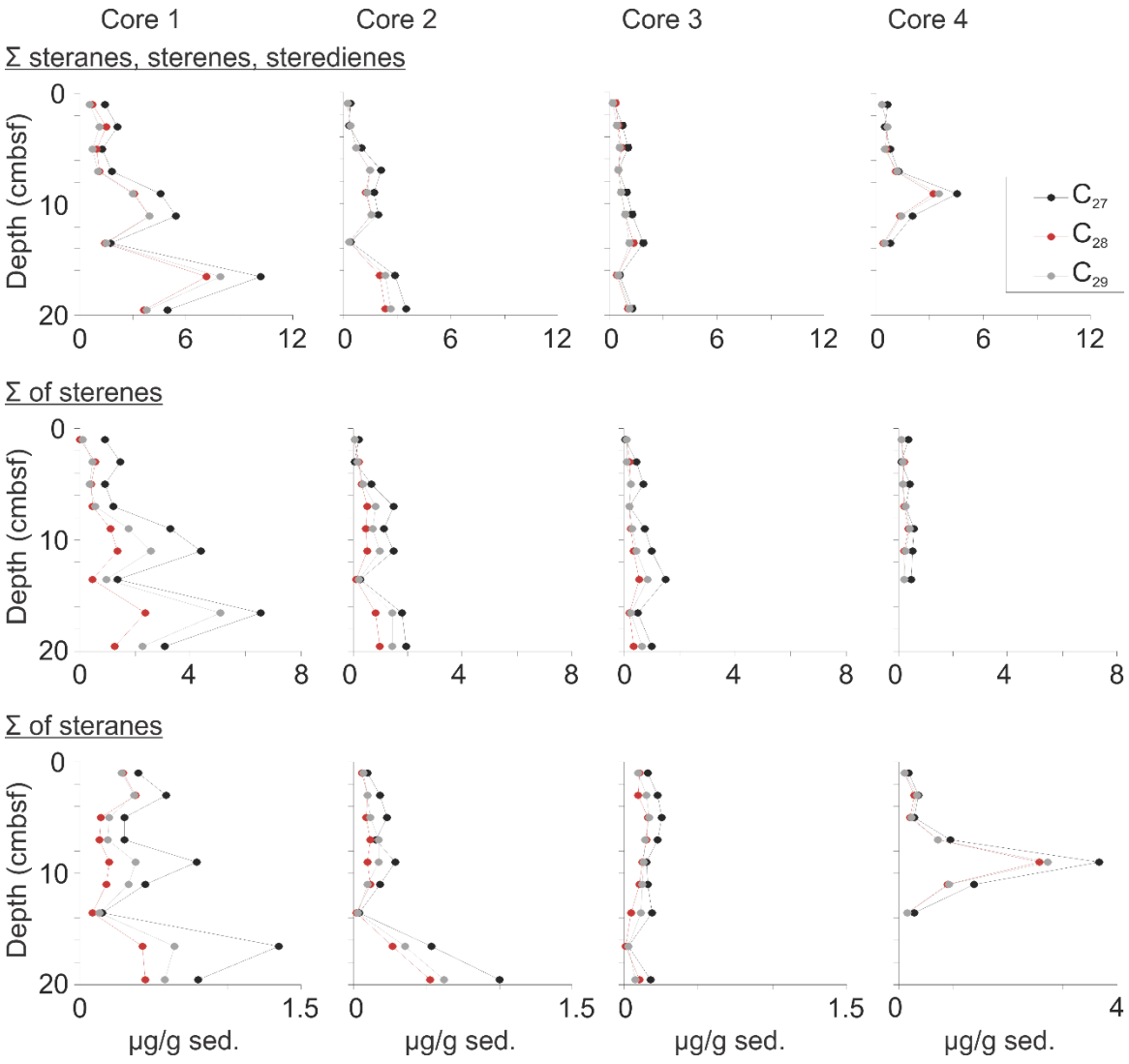


Fig. 5. C_{27}-, C_{28}-, and C_{29}- summed steradiene, sterene, and sterane, abundance with sediment depth.

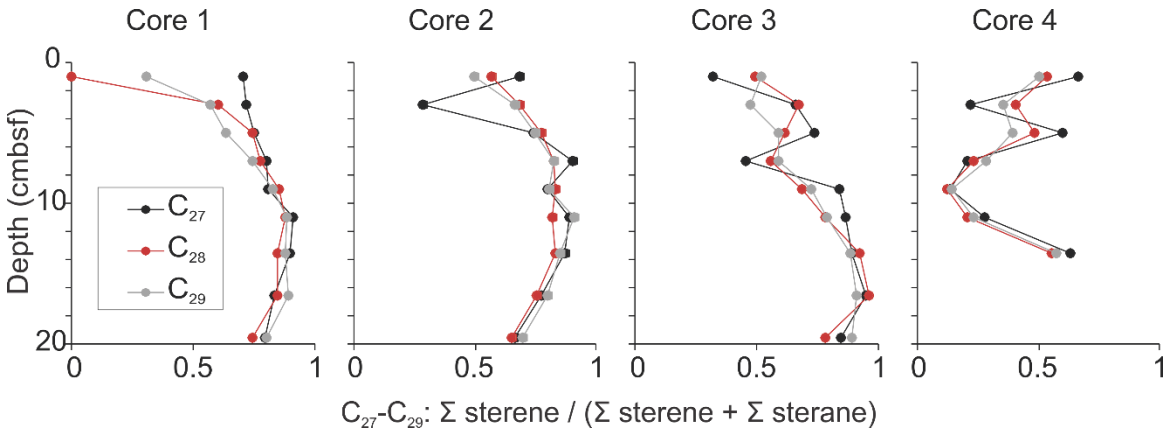


Fig. 6: Down core ratios of C_{27}, C_{28}, and C_{29} sterenes/steranes.

2.4.3.4 Hopanoids

Hopanoids are pentacyclic triterpenoid lipids that function to support bacterial membrane stability and functioning. Their diagenetic derivatives are widely observed as molecular fossils in the geosphere. Hopanoids including hopadienes, hopenes, and hopanes, were monitored using *m/z* 123, and 163, 187, 189, 187, 189, 191, and 205 mass chromatograms. The existence of hop-17(21)-ene, hop-22(29)-ene (diploptene), and hop-21(22)-ene was further confirmed by coelution experiments and mass spectra comparison following serial and doped injections of authentic standards (SI Fig. 5). The hopanoids were then quantitatively measured using GC×GC-FID with identification by template matching. These hopenes are known to form in living bacteria (De Rosa et al., 1971). For example, anaerobically grown Geobacter species *G. metallireducens* and G. *sulfurreducens* contain a range of hopanoids, including diploptene, hop-21(22)-ene, and complex, elongated hopanoids (Härtner et al., 2005). Acetobacter are also known to host diploptene as well as hydroxylated derivatives of hop-22(29)-ene (Rohmer et al., 1980). The strictly fermentative

Gram-negative bacterium Zymomonas mobilis likewise produces diploptene and hop-21-ene (Douka et al., 2001). Hopenes are indicative of recent biogenic inputs and have been identified in cyanobacterial mats and deltaproteobacteria (Härtner et al., 2005; Venkatesan, 1987; Philp, 1985). As such, hopene abundance may reflect the diagenetic conditions where bacteria are living within the transect sediments.

A plot of hopene abundance is provided in Fig. 7. The identified hopenes generally decrease in abundance with depth for cores 1 and 2. For core 3 these compounds steadily increase from the near surface sediment loading and then sharply decline to very low concentration towards the bottom of the core. Core 4 has variably high concentrations throughout the core depth, with no clear systematic trends.

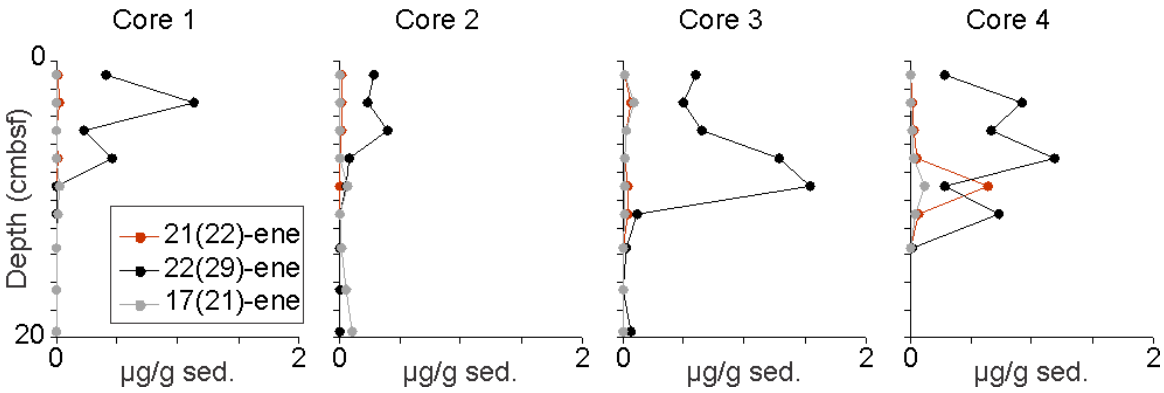


Fig. 7. Changes in hopene abundance with depth across the Cathedral Hill transect.

Similar to sterenes, hopenes are also thermally-unstable molecules that loose their π-bonds and transform into hopanes with rising temperatures. The ubiquity and ease of quantitatively measuring these compounds makes them common thermal maturity indicators (Ourisson et al., 1979; Peters et al., 2005, 2007). These compounds, should

therefore similarly become saturated with depth as observed with the steroid biomarkers (see section 4.3.2) as they become more deeply buried and exposed to hotter vent fluids. A plot of the hopene:hopane (termed hopanoid ratio) to assess maturity level changes in the sediments is provided in Fig. 8. This ratio appears consistent with the thermal gradients recorded by the thermal probe (see section 4.1) with sharply declining values in cores 1 and 2 and values remaining near 1.0 within sediments exposed to more ambient-temperatures. However, closer inspection of the hopane concentrations reveals that the thermally more mature compound remains consistently low irrespective of increasing porewater temperatures and is completely decoupled from the observed hopene losses. This indicates the hopenes and hopanes are not only derived from different sources, but that the hopenes also cannot be detrital molecules that become diagenetically altered with increasing burial (Fig. 8). As such, the only alternative source for these compounds is the living bacteria communities confined to less hydrothermally impacted sites in the vent complex. In this way, the occurrence of hopenes, and in particular diploptene, may delineate the microbial zone of habitation.

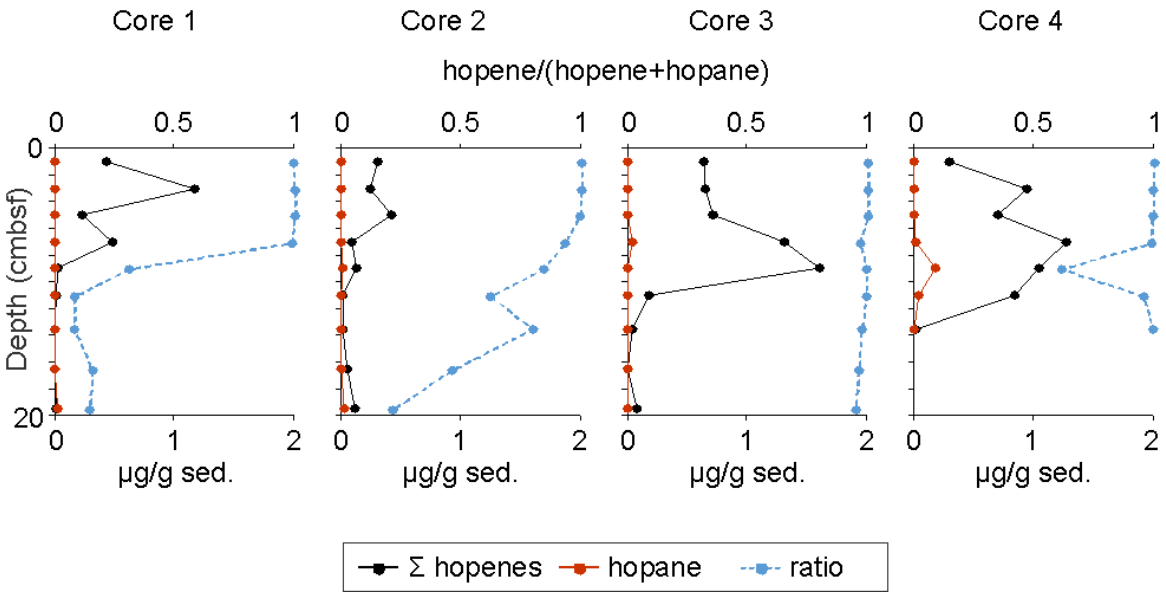


Fig. 8: Hopane and the sum of hopene concentrations along side the hopenes:hopane ratios.

2.4.3.5 Targeted aromatic compounds

Naphthalene and methylnaphthalenes were monitored using *m/z* 142, 156, and 170 mass chromatograms. Phenanthrene, anthracene, and methylphenanthrenes were monitored on the *m/z* 178, 192, and 206 mass chromatogram. The associated detected compound were then quantitatively measured using GC×GC-FID with identification by template matching. These compounds are compared, alongside a variety of hopanoids, as ratios to elucidate maturity (Table 3 and Fig. 11). The Diels' hydrocarbon and various triaromatic steranes, monitored on the *m/z* 217 and 231 mass chromatograms, respectively; were similarly recorded due to their structures representing thermal endpoint of sterane thermal degradation. These compounds are also highly abundant in Guaymas Basin oils - having been first proposed as a marker for high degrees of hydrothermal alteration of OM (Simoneit et al., 1992a; 1992b).

2.4.4 Hydrocarbon Biodegradation

Microbial degradation levels were assessed using the Peters and Moldowan (1993) biodegradation scale after Wegner et al. (2002), which works off the principal that straight chain *n*-alkanes are very susceptible to biodegradation and are normally broken-down before other hydrocarbons (e.g., acyclic isoprenoids, cycloalkanes, etc.) in a quasi-stepwise order (Fig. 9). In general, biodegradation levels increase with sediment depth. Cores 1 and 3 reach their maximum levels of biodegradation at ~8-10 cmbsf. Core 2 reaches its maximum level at 4-6 cmbsf with stretches of unbiodegraded sediments spanning 12-15 to 18-21 cmbsf. The overall level of biodegradation is light, approaching levels 1-2 in the Peters and Moldowan (1993) reflecting a general loss of the low molecular weight (LMW) *n*-alkanes that could also potentially be linked to water washing. Core 4, which hosts ambient sediments, shows a varied trend with depth and is the least biodegraded of all cores (approaching level 1; Peters and Moldowan 1993). Together, these data suggest that most activity takes place within the upper 10 cmbsf, potentially attenuated by the thermal gradient, and that conditions near the vent site that are promoting microbial activity.

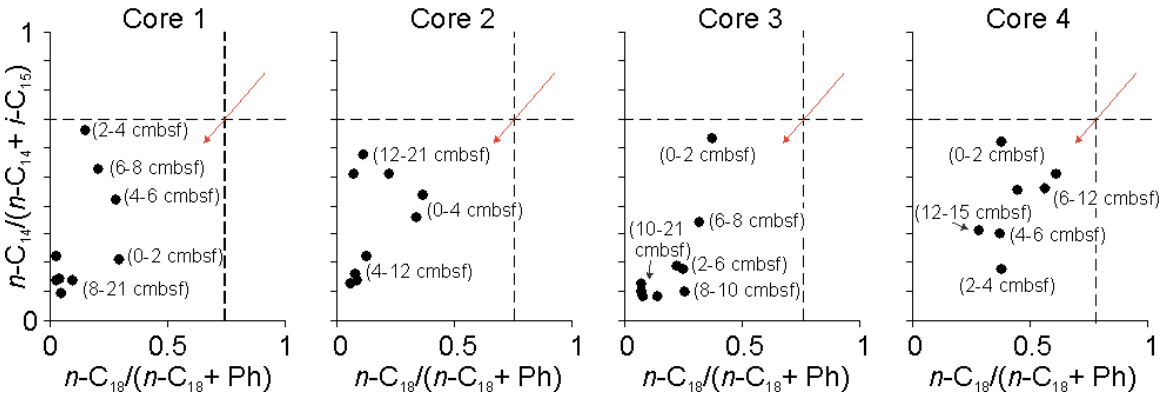


Fig. 9: Ratios of *n*-alkanes to acyclic isoprenoids, monitoring levels of biodegradation with depth. Biodegradation level 2-3 on the Peters and Moldowan (1993) scale are located

within the dotted line cross-section. Red arrow indicates direction of increasing biodegradation.

2.4.5 Biomarker proxies of organic matter Thermal Maturity

2.4.5.1 Carbon Preference Index

During kerogen maturation, odd-carbon-numbered *n*-alkanes predominantly sourced from recent sedimentary inputs of higher plant waxes, become progressively cracked to loose their carbon number preference (Eglinton and Hamilton, 1966; Douglas and Eglinton, 1966). This process can be monitored using the carbon preference index (CPI; Bray and Evans 1961). For this indicator, recent or unaltered high molecular weight (HMW) *n*-alkanes have high CPIs that decrease to approach parity (1) with increasing thermal maturation (Table 3 and Fig. 10).

For cores 1 to 3, CPI values rises with depth suggests decreasing levels of sediment maturity. This runs counter to the high thermal gradient of the vent site and does not reflect the trends observed for nearly all other thermal maturations parameters (see section 4.5.2 below). The increase in CPI, however, can be explained if new additions of odd-numbered *n*-alkanes are being added to the apolar fractions of the Cathedral Hill oils by the defunctionalisation of secondary alcohols, ketones, and aldehydes that are likewise sourced from higher plant-cuticular waxes (Jeffree, 1996; Millar, 1999; Kunst, 2003; Bernard and Joubès, 2013; Barbero, 2016), with alkyl tails that similarly share a carbon number preference. As such, elevated CPI may conform to the expected diagenetic depth trends, increasing with higher temperatures.

Furthermore, in core 4, several intervals have CPI values approaching 1.0 (Fig. 10). These low values also correspond to sediments that contained elevated polar and apolar material and likely marks more mature oil formed deeper in the sediments advected into the cooler flanks of the vent site (see also Chapter 3 for further multi-molecular evidence of a migrated oil lens that extends across the transect at ~10 cmbsf).

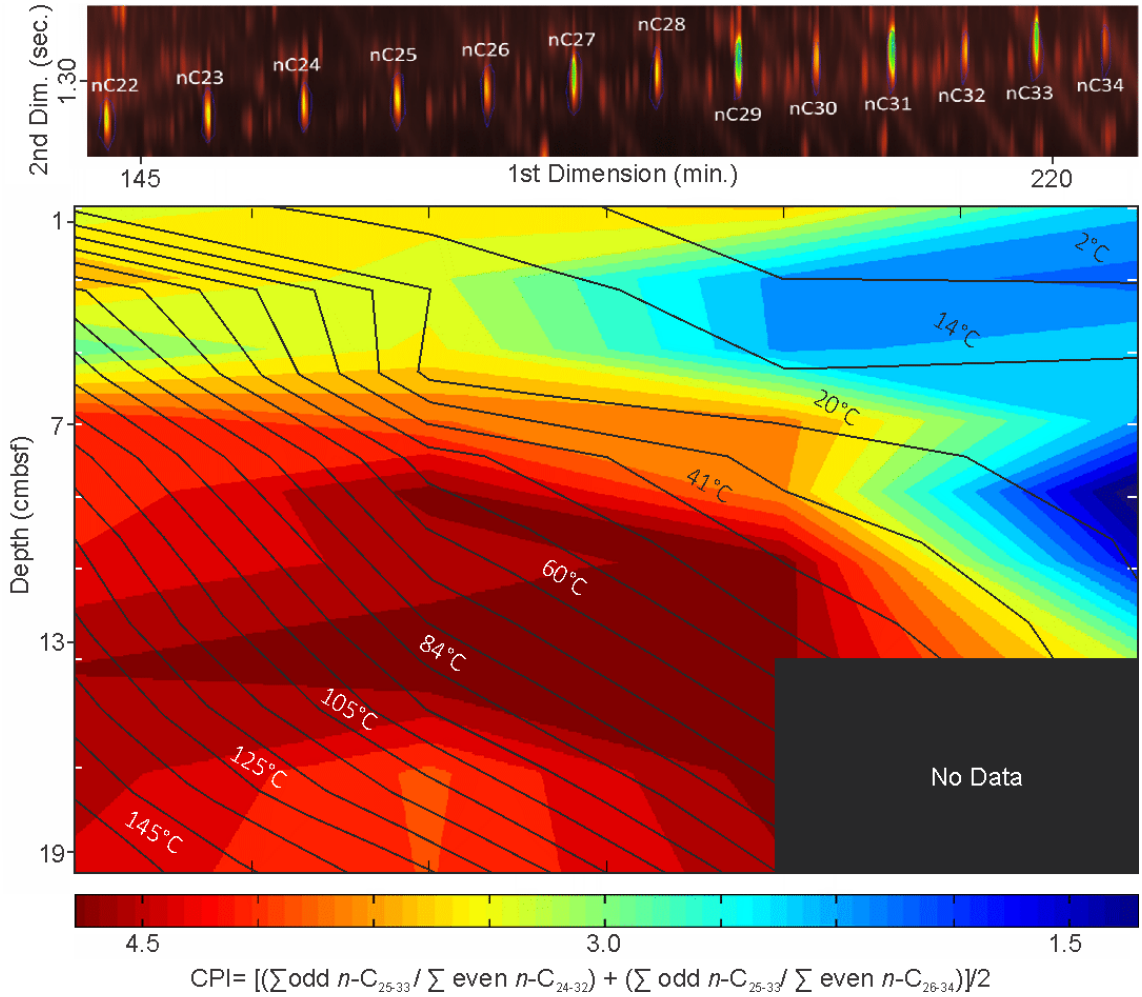


Fig. 10. Heat map of CPI at depth with superimposed thermal gradient and associated *n*-alkane series.

2.4.5.2 Aromatic and terpenoid-based estimates of thermal maturity

Nearly all measured ratios show an increase in thermal maturity with sediment depth for cores 1 and 2, and comparatively lower maturities with depth for cores 3 and 4 (Fig. 11; Table 3). However, most proxies conflict on the exact stage of oil generation/level of maturity, and some show trends that are incompatible with previous observations (ex: late oil in cores 1 and 2, even though sterenes are still present). Out of all proxies, MH, Ts/Tm, and the Diels' hydrocarbon:C_{26} triaromatic cholestane ratio conform best to the expected diagenetic depth trends, showing clear increase of thermal maturity with depth in the hotter pushcores 1-2 and relative immaturity in the other, more ambient samples. Discrepancies include spikes in Ts/Tm at ~8-10 cmbsf in core 4, which aligns with the potentially migrated oils. Collectively, the results indicate that cores 1 and 2 enter higher levels of thermal maturity with sediment depth, and provide evidence of trapped oil within core 4 at approximately 8-10 cmbsf.

Table 3. Cathedral Hill transect thermal maturation parameters [†]

Sample location and name	CPI [a]	TMN [b]	MPR-1 [c]	MPI-1 [d]	TNR-2 [e]	Diels'/C_{26} TriSterane [f]	Ts/Tm [g]	M/H [h]	C_{27} sterene/ C_{27} sterane [i]	C_{28} sterene/ C_{28} sterane [j]	C_{29} sterene/ C_{29} sterane [k]	hopene/ hopane [l]
Microbial mat closest to vent												
Core 1 0-2 cmbsf	3.21	0.38	0.50	0.63	1.01	0.84	0.30	4.24	0.72	0.50	0.53	0.99
Core 1 2-4 cmbsf	3.40	0.4	0.88	0.31	0.88	0.80	0.19	5.79	0.79	0.59	0.63	1.00
Core 1 4-6 cmbsf	2.82	0.49	1.04	0.53	0.89	0.55	0.14	6.2	0.85	0.68	0.74	1.00
Core 1 6-8 cmbsf	4.05	0.54	1.2	0.75	1.01	0.70	0.15	2.54	0.87	0.72	0.80	0.99
Core 1 8-10 cmbsf	4.05	0.57	1.51	1.01	1.10	0.90	0.34	2.79	0.92	0.83	0.84	0.31
Core 1 10-12 cmbsf	3.89	0.59	1.69	1.05	0.99	0.96	0.44	0.6	0.94	0.89	0.91	0.08
Core 1 12-15 cmbsf	4.35	0.57	1.46	1.08	0.94	0.98	0.44	0.99	0.92	0.88	0.91	0.08
Core 1 15-18 cmbsf	4.19	0.59	1.63	1.08	1.00	0.92	0.56	0.69	0.88	0.83	0.87	0.15
Core 1 18-21 cmbsf	4.01	0.57	1.56	1.05	1.10	0.98	0.65	0.67	0.85	0.80	0.82	0.14
Microbial mat further from vent												
Core 2 0-2 cmbsf	3.12	0.57	0.75	0.76	0.90	0.83	0.31	4.67	0.76	0.53	0.59	1.00
Core 2 2-4 cmbsf	3.02	0.48	1.26	0.24	0.68	0.18	0.27	4.46	0.82	0.64	0.71	1.00
Core 2 4-6 cmbsf	3.29	0.49	1.28	0.57	0.71	0.46	0.29	2.01	0.81	0.70	0.79	0.99
Core 2 6-8 cmbsf	3.71	0.53	1.45	0.69	0.72	0.79	0.36	1.43	0.91	0.81	0.83	0.93
Core 2 8-10 cmbsf	4.83	0.54	1.33	0.8	0.96	0.86	0.37	2.46	0.93	0.86	0.90	0.84
Core 2 10-12 cmbsf	4.28	0.53	1.5	0.83	0.66	0.60	0.34	2.59	0.93	0.84	0.90	0.62
Core 2 12-15 cmbsf	4.65	0.56	1.36	1.03	0.83	0.87	0.38	1.67	0.91	0.85	0.88	0.79
Core 2 15-18 cmbsf	3.90	0.6	1.64	1.00	0.85	0.99	0.52	1.73	0.84	0.78	0.82	0.46
Core 2 18-21 cmbsf	4.04	0.62	1.65	1.05	1.17	0.98	0.53	1.11	0.72	0.67	0.68	0.21
Just inside microbial mat perimeter												
Core 3 0-2 cmbsf	3.17	0.66	1.88	1.21	0.94	0.21	0.19	6.58	0.77	0.54	0.60	1.00
Core 3 2-4 cmbsf	2.39	0.61	1.28	0.9	1.22	0.02	0.09	5.23	0.70	0.53	0.57	1.00
Core 3 4-6 cmbsf	2.64	0.49	1.1	0.87	1.17	0.13	0.08	3.5	0.81	0.62	0.64	1.00
Core 3 6-8 cmbsf	3.43	0.39	0.79	0.18	0.73	0.18	0.17	0.4	0.76	0.52	0.61	0.97
Core 3 8-10 cmbsf	3.65	0.34	0.72	0.46	0.70	0.19	0.11	3.05	0.84	0.72	0.76	0.99

Core 3 10-12 cmbsf	4.65	0.35	0.8	0.39	0.73	0.17	0.10	2.04	0.84	0.76	0.78	0.99
Core 3 12-15 cmbsf	4.65	0.45	0.89	0.35	0.74	0.31	0.06	2.14	0.92	0.85	0.85	0.97
Core 3 15-18 cmbsf	4.66	0.47	0.89	0.48	0.68	0.57	0.11	2.45	0.93	0.85	0.89	0.96
Core 3 18-21 cmbsf	4.60	0.40	0.46	0.40	0.65	0.50	0.11	2.82	0.93	0.88	0.86	0.95
Just outside microbial mat perimeter												
Core 4 0-2 cmbsf	2.54	0.52	2.31	1.00	0.90	0.03	0.20	7.04	0.71	0.46	0.46	1.00
Core 4 2-4 cmbsf	1.92	0.46	0.88	0.73	0.76	0.22	0.19	3.58	0.67	0.44	0.43	1.00
Core 4 4-6 cmbsf	2.27	0.4	1.23	0.93	0.82	0.09	0.23	4.34	0.60	0.42	0.36	0.99
Core 4 6-8 cmbsf	2.14	0.47	0.91	0.78	0.63	0.05	0.36	3.89	0.40	0.19	0.29	0.98
Core 4 8-10 cmbsf	0.94	0.53	0.94	0.78	0.77	0.02	0.44	1.58	0.44	0.28	0.32	0.61
Core 4 10-12 cmbsf	1.53	0.46	1.05	0.8	0.63	0.03	0.33	2.71	0.48	0.45	0.62	0.95
Core 4 12-15 cmbsf	2.83	0.47	0.94	0.66	0.65	0.11	0.13	1.57	0.74	0.49	0.55	0.99

† See Fig. 11 for select depth profile plots.

All compounds quantified in their respective GC×GC–FID sample chromatograms.

[a] CPI= [(∑odd *n*-C_{25-33} / ∑ even *n*-C_{24-32}) + (∑ odd *n*-C_{25-33}/ ∑ even *n*-C_{26-34})]/2 (Bray and Evans 1961)

[b] Trimethylnaphthalene ratio (TMN) = 1,3,7-trimethylnaphthalene /(1,3,7-trimethylnaphthalene + 1,2,5-trimethylnaphthalene), identified via monitoring of the *m/z* 170 mass chromatograms.

[c] Methylphenanthrene ratio 1 (MPR-1) = 2-methylphenanthrene / 1-methylphenanthrene identified via monitoring of the *m/z* 191 mass chromatograms and quantified in their respective GC×GC–FID sample chromatograms.

[d] Methylphenanthrene Index 1 (MPI-1) = (3-methylphenanthrene + 2-methylphenanthrene)/(9-methylphenanthrene + 1-methylphenanthrene) (Radke et al., 1982).

[e] Trimethylnaphthalene ratio (TNR-2) = (2,3,6-trimethylnaphthalene + 1,3,7-trimethylnaphthalene) /(1,3,5-trimethylnaphthalene +1,4,6-trimethylnaphthalene +1,3,6-trimethylnaphthalene) (Radke et al., 1986).) Identified via monitoring of the *m/z* 170 mass chromatograms.

[f] Diels'/C_{26} TriSterane = Deils' hydrocarbon/(Deils' hydrocarbon + C_{26} triaromatic cholestane 20S) (Ventura et al., 2012), identified via monitoring of the m/z 217 mass chromatograms.

[g] Ts/Tm = C_{27} 17α(H)-22,29,30-trisnorhopane /(C_{27} 17α(H)-22,29,30-trisnorhopane + 18α(H)-22,29,30-trisnorneohopane) (Table 2 compounds 1 and 2) identified via monitoring of the *m/z* 191 mass chromatograms.

[h] M/H = βα-hopane/ αβ-hopane (Table 2 compounds 11 and 7), identified via monitoring of the *m/z* 191 mass chromatograms.

[i, j, k] Steroid ratios = C_{27}-C_{29} sterenes / (C_{27}-C_{29} steranes + C_{27}-C_{29} sterenes) marking compounds 3-6, 9-12, and 15-18 from Table 2. Compounds were identified via monitoring of the *m/z* 217 and 215 mass chromatograms.

[l] hopene/hopane = C_{30} hopene / (C_{30} hopene + C_{30} hopane), identified via monitoring of the *m/z* 191 mass chromatograms.

Table 4. Cathedral Hill transect source and biodegradation parameters

Sample location and name	Biodegradation		Source		
	n-C_{18}/(n-C_{18}+Ph) [a]	n-C_{14}/(n-C_{14}+i-C_{15}) [b]	Pr/Ph [c]	Pr/ n-C_{17} [d]	Ph/ n-C_{18} [e]
Microbial mat closest to vent					
Core 1 0-2 cmbsf	0.30	0.21	0.93	2.05	2.38
Core 1 2-4 cmbsf	0.15	0.66	0.39	1.74	5.64
Core 1 4-6 cmbsf	0.28	0.42	0.86	1.38	2.59
Core 1 6-8 cmbsf	0.20	0.52	0.46	1.99	3.89
Core 1 8-10 cmbsf	0.10	0.14	0.31	2.95	9.18
Core 1 10-12 cmbsf	0.05	0.10	0.33	5.77	18.87
Core 1 12-15 cmbsf	0.04	0.15	0.32	7.04	23.40
Core 1 15-18 cmbsf	0.03	0.14	0.33	11.19	34.41
Core 1 18-21 cmbsf	0.03	0.23	0.33	8.14	34.76
Microbial mat further from vent					
Core 2 0-2 cmbsf	0.34	0.36	1.46	1.62	1.98
Core 2 2-4 cmbsf	0.36	0.44	1.61	1.80	1.76
Core 2 4-6 cmbsf	0.12	0.22	0.28	2.26	7.28
Core 2 6-8 cmbsf	0.08	0.14	0.33	3.53	11.30
Core 2 8-10 cmbsf	0.07	0.17	0.34	3.63	12.55
Core 2 10-12 cmbsf	0.05	0.13	0.32	5.39	17.81
Core 2 12-15 cmbsf	0.22	0.51	0.60	0.49	3.57
Core 2 15-18 cmbsf	0.11	0.57	0.50	1.13	8.20
Core 2 18-21 cmbsf	0.07	0.51	0.46	1.61	13.06
Just inside microbial mat perimeter					
Core 3 0-2 cmbsf	0.37	0.63	1.20	1.41	1.71
Core 3 2-4 cmbsf	0.25	0.18	0.78	2.78	3.04
Core 3 4-6 cmbsf	0.22	0.19	0.46	4.71	3.59
Core 3 6-8 cmbsf	0.31	0.34	0.71	1.67	2.19

Core 3 8-10 cmbsf	0.25	0.10	0.50	1.86	2.95
Core 3 10-12 cmbsf	0.14	0.09	0.24	2.22	6.31
Core 3 12-15 cmbsf	0.07	0.13	0.28	3.25	13.47
Core 3 15-18 cmbsf	0.08	0.09	0.34	3.01	12.06
Core 3 18-21 cmbsf	0.07	0.10	0.39	3.84	13.17
Just outside microbial mat perimeter					
Core 4 0-2 cmbsf	0.38	0.62	1.08	1.72	1.64
Core 4 2-4 cmbsf	0.38	0.18	0.85	1.86	1.66
Core 4 4-6 cmbsf	0.37	0.30	0.67	1.78	1.71
Core 4 6-8 cmbsf	0.56	0.46	1.02	0.90	0.79
Core 4 8-10 cmbsf	0.61	0.51	1.17	0.92	0.64
Core 4 10-12 cmbsf	0.45	0.46	0.88	1.23	1.23
Core 4 12-15 cmbsf	0.28	0.31	0.73	2.14	2.58

All compounds quantified in their respective GC×GC–FID sample chromatograms.

[a,b] components of the Wegner et al. (2002) biodegradation scale: *n*-C_{18}/(*n*-C_{18}+Phytane) and *n*-C_{14}/(*n*-C_{14}+*i*-C_{15}). Approaches 0 with increasing biodegradation. Identified via monitoring of the *m/z* 57 and 183 mass chromatograms.

[c] Pr/Ph = pristine/phytane, identified via monitoring of the *m/z* 183 mass chromatogram.

[d] Pr/ *n*-C_{17} = pristine/ *n*-C_{17}, identified via monitoring of the *m/z* 57 and 183 mass chromatograms.

[e] Ph/ *n*-C_{18} = phytane/ *n*-C_{18}, identified via monitoring of the *m/z* 57 and 183 mass chromatograms.

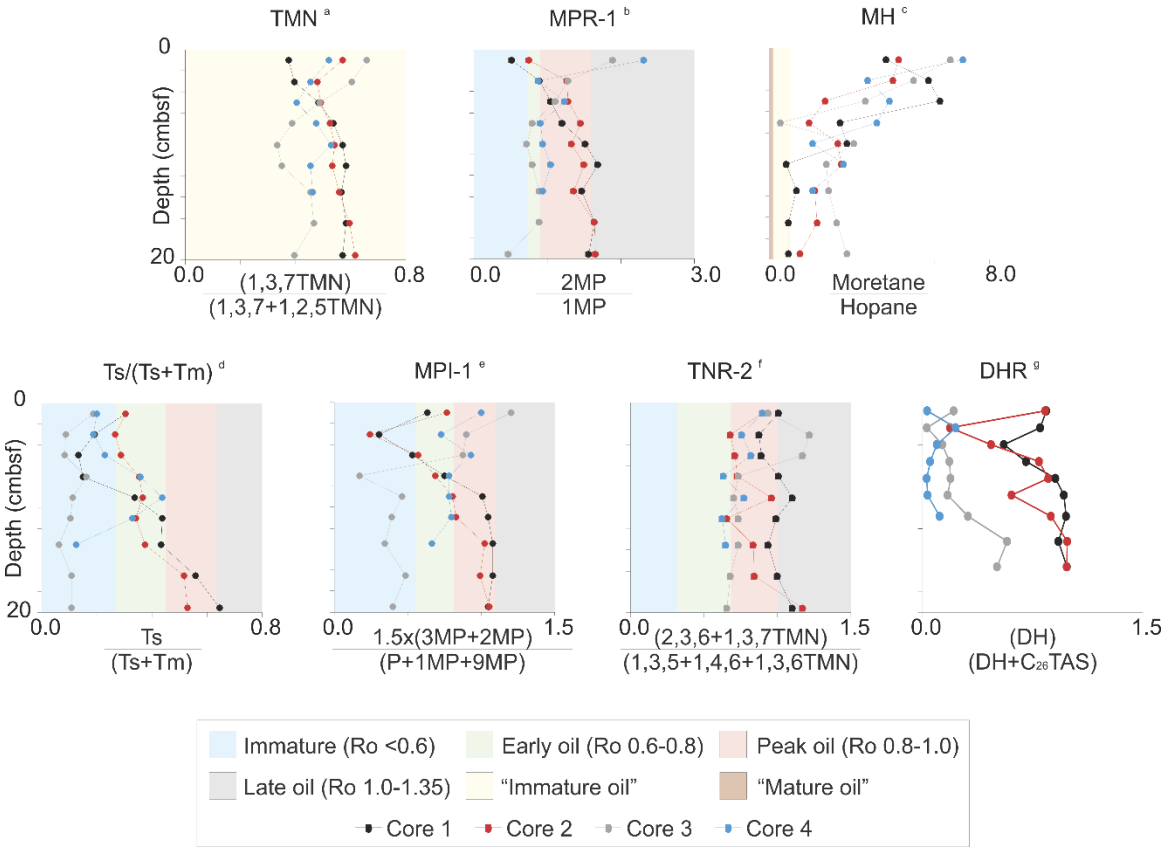


Fig. 11: Thermal maturity parameters. Relationship of each parameter to vitrinite relectance (R_o) or oil maturity as provided by Van Aarssen et al. (1999)[a], Radke (1988)[b], Peters et al. (2007)[c], Raju et al. (2014)[d], Boreham et al. (1988)[e], Radke et al. (1994)[f], Ventura et al., 2012[g]. The abbreviations of ratios and associated compounds is provided in Table 3).

2.4.6 Synthesis of Cathedral Hill subsurface biodegradation and thermal maturation profile

Marine hydrothermal sediments at Guaymas Basin host a wide array of heterotopic microorganisms (e.g. Teske et al., 2002; Edgcomb et al., 2002; Guerrero-Barajas et al., 2011; Biddle et al., 2012; Liu et al., 2016). At Cathedral Hill, active microbial degradation appears near-ubiquitous in sediments within zone of subsurface bacterial activity (Fig. 12). Independent patterns between detrital steroids (which are attenuated by the thermal

gradient) and hopanoids (which are uncoupled from the expected diagenetic depth trends) help mark the range of microbial habitability, as hopene abundances, such as diplotene, are likely controlled by the presence of living microbes in the subsurface. However, Fig. 12 shows an offset between the temperature-limit of bacterial life, recorded by in situ pore water temperatures, and this hopene lipid biomarker proxy used to delineate the habitable zone. In core 1, the range of hopenes extends beyond this threshold, while in cores 2-3 bacterial life appears attenuated before the temperature limit is reached. This may indicate that the thermal probe readings are misrepresentative of the subsurface temperature profile, reflecting a mere snapshot a dynamic hydrothermal system, which are known to have variable temperature fluctuations over short periods (Johnson et al., 1988). It may also show that the hopene biomarker can persist in the geosphere, and thus may not delineate the range of active microbes. Equally, the presence of hopenes past the bacterial temperature threshold could also depict a zone of microbial "metastability," where the community is able to persist in lethal conditions for long periods (up to ~25 years based on the 1-2 mm/yr sedimentation rate; Williams et al., 1979; Teske et al., 2002) before eventual cell death. Such hardiness is not inconceivable considering the persistence of terrestrial (forest soil) organotrophic microbes within deep-sea lignite beds for millions of years (Inagaki et al., 2015), or the revival of sporulating bacteria preserved in amber for 25-40 million years (Cano and Burucki, 1995), along with increased heat tolerance of microbial spores in general (up to ~46 °C higher than the active cells; Warth 1978).

The reducing conditions at Cathedral Hill may promote anaerobic metabolic strategies such as bacterial sulfate reduction (BSR), methanogenesis, nitrate and nitrite reduction, sulfide oxidation, and anaerobic oxidation of ammonia (e.g. Rueter et al., 1994;

Teske et al., 2002; Amend and Teske, 2005; MacGregor et al., 2013; Dombrowski et al., 2018) by microorganisms such as *Thermodesulfobacterium* (Jeanthon et al., 2002), *Desulfobulbaceae* (Teske et al., 2002; Kleindienst et al., 2012) and *Beggiatoa* (Jannasch et al., 1989; Nelson et al., 1989; MacGregor et al., 2013). Overall, low levels of biodegradation are achieved within the first 10 cmbsf (indicated via Wegner and Peters and Moldowan (1993) biodegradation indices) and appears attenuated by the thermal gradient below. This is consistent with other microbial surveys in hydrothermally-active sediments at Guaymas, showing temperature-controlled confinement of microbes near the surface and a sharp decrease with depth within the upper 3–30 cm of the seafloor (Elsgaard et al., 1994; Guezennec et al., 1996; Weber and Jørgensen, 2002; Edgcomb et al., 2002; Teske et al., 2002; Amend and Teske, 2005). More fundamental is that these processes are commonly mediated by the heterotrophic process of BSR. The recorded decreases in porewater sulphate concentrations (Table 1) proximal to rapid reductions in the bacterial sourced biomarker, diploptene; that equally occupies sediments that have had some level of biodegradation of their aliphatic portions of OM is consistent with the presence of bacterial sulphate reducers (Fig. 12).

Several maturity proxies predict early generation of immature oil within 10 cmbsf in cores 1 and 2, which is cohesive with elevated CPI (Fig. 10) and steroid ratios at depth (Fig. 6). If accurate, this represents in situ hydrocarbon production inside 40-80 years of burial, which is significantly lower than the youngest oil radiocarbon ages established at Guaymas Basin (~6,000 years; Peter et al., 1991; Simoneit and Kvenvolden, 1993). However, some areas of inferred generation are below the temperature threshold for both conventional and hydrothermal oil-production (Tissot and Welte, 1984; Simoneit, 1993).

Again, this may suggest higher temperature flux in the past (Johnson et al., 1988), or that the maturity proxies are potentially being skewed via biodegradation or water washing (Lonsdale and Becker, 1985; Kawka, and Simoneit, 1987; Kawka, 1990; Simoneit, 1993). Migration could equally be overprinting the maturity proxies, with evidence supporting oil impregnations in core 4 (i.e. elevated polar/apolar material, Ts/Tm, steranes, and low CPI), which may have travelled through and stained the transect sediments with mature hydrocarbons. However, independent of these trends are steadily increasing levels of thermal maturity with depth and proximity to the vent center, which overprint these relict/migrated signals in the hottest sections of the transect. As such, it is inconclusive if/where in situ oil generation is occurring within the sediment profile, though thermal transformations with the temperature gradient are evident.

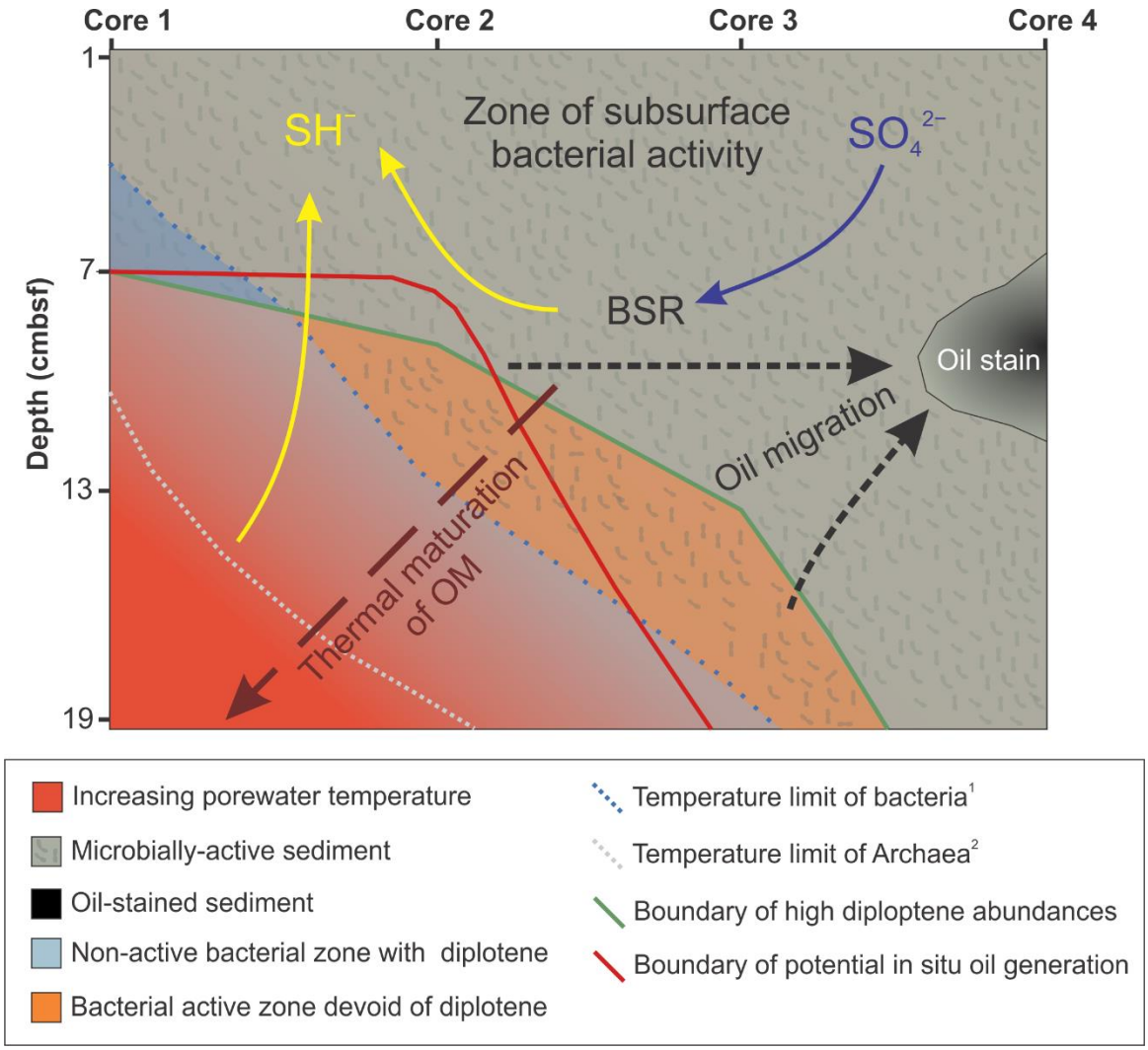


Fig. 12. Model of subsurface microbiological and thermochemical activity at Cathedral Hill. The boundary of oil generation is based on thermal maturation proxies (Table 3). The boundary of high diploptene abundance is given by Fig. 7. Blue-shaded regions indicate recovery of diploptene past the temperature limit of life for bacteria (Holden, 2009)[1]. Likewise, orange-shaded regions indicate no diploptene recovery despite being within this threshold. Migrated oils likely originate from advecting hydrothermal fluids, with evidence for an oil stain in core 4 (Fig. 3, 6, Table 3). Decreases in porewater sulphate (SO_4^{2-}) concentrations with depth (Table 1) are in agreement with BSR being confined to the more ambient regions of the subsurface, which likewise display low levels of hydrocarbon biodegradation (Fig. 9). Elevated concentration of reduced sulphur species (SH-) likely emanate from both the vent fluids and as a waste product of BSF. Thermal limits of archaea provided by Kashefi and Lovley (2003)[2].

2.5 Conclusion

This study documents the degree of thermal maturity and biodegradation among the petroliferous, near-surface hydrothermal vent sediments at Cathedral Hill. The results show evidence for in situ transformation in the subsurface, which may be overprinted by the migration of more deeply generated and thermally altered hydrocarbons. Thermal maturity levels increase with depth and proximity to the hydrothermal vent edifice, with various proxies roughly indicating oil generation within 20 cmbsf at 60-153 ˚C and burial times of 40-80 years. However, conflict between these proxies, evidence of migrated oil, and the fact that some regions are beyond the conventional/hydrothermal oil window suggest a more comprehensive analysis is needed to conclusively identify in-situ production in these sediments. An active subsurface community is predicted to be operating within almost all samples below the 80 ˚C bacterial temperature threshold, where sediments quickly become microbially altered. Uncoupling of hopane from hopene:hopane ratio indicates that the distribution of biological form of this lipid is controlled by the distribution of living cells in the subsurface, helping to delineate the range of the habitable zone. Evidence of hopene beyond the temperature threshold of bacterial life may also allude to a zone of microbial "metastability,"or that the reconstructed porewater temperature profile may not accurately reflect the true thermal dynamics of this system. All together, the results show marked spatial heterogeneity resolvable on the scale of centimeters, demonstrating that the shallow, hydrothermal environment at Cathedral Hill is a complex and rapidly evolving system.

2.6 References

Amend J. P. and Teske A. (2005). Expanding frontiers in deep subsurface microbiology. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **219**(1-2), 131–155.

Arey J. S., Nelson R. K., Xu L. and Reddy C. M. (2005) Using comprehensive two dimensional gas chromatography retention indices to estimate environmental partitioning properties for a complete set of diesel fuel hydrocarbons. *Anal. Chem*. **77**, 7172–7182.

Arey J. S., Nelson R. K. and Reddy C. M. (2007) Disentangling oil weathering using GC GC. 1. Chromatogram analysis. *Environ. Sci. Technol*. **41**, 5738–5746.

Barbero F. (2016) Cuticular Lipids as a Cross-Talk among Ants, Plants and Butterflies. *Int. J. Mol. Sci.* **17**(12), 1966.

Barman B. and Membrado V. (2000) Chromatographic techniques for petroleum and related products. *Crit. Rev. Anal. Chem.* **30**(2-3), 75–120.

Bazylinski A., Farrington W. and Jannasch W. (1988) Hydrocarbons in surface sediments from a Guaymas Basin hydrothermal vent site. *Org. Geochem.***12**, 547–558.

Bernard A. and Joubès J. (2013) Arabidopsis cuticular waxes: Advances in synthesis, export and regulation. *Prog. Lipid Res*. **52**(1), 110–129.

Biddle J. F., Cardman Z., Mendlovitz H., Albert D.B., Lloyd K.G., Boetius A. and Teske A. (2012) Anaerobic oxidation of methane at different temperature regimes in Guaymas Basin hydrothermal sediments. *ISME J.* **6**(5), 1018–1031.

Boehm P. D., Douglas G. S., Burns W. A., Mankiewicz P. J., Page, D. S. and Bence A. E. (1997) Application of petroleum hydrocarbon chemical fingerprinting and allocation techniques after the Exxon Valdez oil spill. *Mar. Pollut. Bull.* **34**(8), 599–613.

Boreham C. J., Crick I. H., and Powell T. G. (1988) Alternative calibration of the Methylphenanthrene Index against vitrinite reflectance: Application to maturity measurements on oils and sediments. *Org. Geochem.* **12**(3), 289–294.

Bray E. E. and Evans E. D. (1961) Distribution of n-paraffins as a clue to recognition of source beds. *Geochim. Cosmochim. Acta* **22**, 2–15.

Calvert S. E. (1966) Accumulation of diatomaceous silica in the sediments of the Gulf of California. *Geol. Soc. Am. Bull.* **77**, 569–596.

Cano R. J. and Burucki M. K. (1995) Revival and identification of bacterial spores in 25- to 40- million year old dominican amber. *Sciences* **268**, 1060–1064.

Castillo P. R., Hawkins J. W., Lonsdale P. F., Hilton D. R., Shaw A. M. and Glascock M. D. (2002) Petrology of Alarcon Rise lavas, Gulf of California: Nascent intracontinental ocean crust. *J. Geophys. Res.* **107**, 1-15

Christensen J. H. and Tomasi G. (2016) A multivariate approach to oil hydrocarbon fingerprinting and spill source identification. In *Standard Handbook Oil Spill Environmental Forensics: Fingerprinting and Source Identification, Second Edition* (eds. S. A. Stout and Z. Wang). Academic Press, Massachusetts, pp. 747–788.

Curray J., Moore D., and the DSDP Scientific Party. (1982) Guaymas Basin: Sites 477, 478, and 481. In *Initial reports of the Deep Sea Drilling Project-LXIV.* U.S. Govt. Printing Office, Washington DC.

Curray J. R., Moore D. G. et al. (1979) Leg 64 seeks evidence on development of basins. *Geotimes* **24**(7), 18–20.

De Rosa M., Gambacorta A. and Minale L.(1971) Bacterial Triterpanes. *ChemComm.* 12, 619-620.

Didyk M., and Simoneit B. R. T. (1989) Hydrothermal oil of Guaymas Basin and implications for petroleum formation mechanisms. *Nature* **342**(6245), 65–69.

Didyk M. and Simoneit B. R. T. (1990) Petroleum characteristics of the oil in a Guaymas Basin hydrothermal chimney. *Appl. Geochem.* **5**, 29–40.

Didyk .M., Simoneit B. R. T., Brassell S. C. and Eglinton G. (1978) Organic geochemical indicators of paleoenvironmental conditions of sedimentation. *Nature* **272**, 216–222.

Dombrowski N., Teske A. P. and Baker B. J. (2018) Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* **9**(1), 1–13.

Douglas A. G. and Eglinton, G. (1966) The distribution of alkanes. In: Swain, T. (Ed.). Comparative Phytochemistry. Academic Press, London, pp. 57–78.

Douka E., Koukkou A.-I., Drainas C., Grosdemange-Billiard C. and Rohmer M. (2001) Structural diversity of the triterpenic hydrocarbons from the bacterium Zymomonas mobilis: the signature of defective squalene cyclization by the squalene/hopene cyclase. *FEMS Microbiol. Lett.* **199**(2), 247–251.

Edgcomb V.D., Kysela A., Teske A., de Vera Gomez, Sogin M. L., (2002) Benthic eukaryotic diversity in the Guaymas Basin, a hydrothermal vent environment. *P.N.A.S.* **99**, 7658–7662.

Einsele G., Gieskes J., Curray J., Moore D., Aguayo E., Aubry M., Fornari D., Guerrero J., Kastner M., Kelts K., Lyle M., Matoba Y., Molina-Cruz A., Niemitz J., Rueda J., Saunders A., Schrader H., Simoneit B. and Vacquier V. (1980) Intrusion of basaltic sills into highly porous sediments and resulting hydrothermal activity. *Nature* **283**, 441–445.

Eglinton G. and Hamilton R. J. (1967) Leaf epicuticular waxes. *Science* 156, 1322–1335.

Elsgaard L., Isaksen M. F., Jørgensen B. B., Alayse A.-M. and Jannasch H. W. (1994) Microbial sulfate reduction in deep-sea sediments at the Guaymas Basin hydrothermal vent area: Influence of temperature and substrates. *Geochim. Cosmochim. Acta* **58**(16), 3335–3343.

Fisher A. T. and Becker K. (1991) Heat flow, hydrothermal circulation and basalt intrusions in the Guaymas Basin, Gulf of California. *Earth Planet. Sci. Lett.* **103**(1– 4), 84–99.

Frysinger G., Gaines R. and Reddy C. (2002) GC×GC—A new analytical tool for environmental forensics. *Environ. Forensics* **3**(1), 27–34.

Fouquet Y., Zierenberg R. A., Miller D. J., Bahr J. M., Baker P. A., Bjerkgården T., Brunner C.A., Duckworth R.C., Gable R., Gieskes J., Goodfellow W. D., Gröschel-Becker H. M., Guèrin G., Ishibashi J., Iturrino G., James R. H., Lakschewitz K.S., Marquez L. L., Nehlig P., Peter J. M., Rigsby C. A., Simoneit B. R. T., Schultheiss P., Shanks W. C., Summit M., Teagle D. A. H., Urbat M. and Zuffa G. G. (1998) *Proceedings of the Ocean Drilling Program, Initial Reports,* Vol. 169. US Government Printing Office, Washington DC.

Guerrero-Barajas, C., Garibay-Orijel, C. and Rosas-Rocha, L. E. (2011) Sulfate reduction and trichloroethylene biodegradation by a marine microbial community from hydrothermal vents sediments. *Int. Biodeterior. Biodegr* **65**(1), 116–123.

Guezennec J. G., Dussauze, J. and Bian M. (1996) Bacterial community structure in sediments from Guaymas Basin, Gulf of California, as determined by analysis of phospholipid ester-linked fatty acids. *J. Mar. Biotechnol.* **4**, 165–175.

Härtner T., Straub K. L. and Kannenberg E. (2005) Occurrence of hopanoid lipids in anaerobic Geobacter species. *FEMS Microbiol. Lett.* **243**, 59–64.

Hazra B., Wood D. A., Mani D., Singh P. K. and Singh A. K. (2019) Sedimentary Biomarkers and Their Stable Isotope Proxies in Evaluation of Shale Source and Reservoir Rocks. In *Evaluation of Shale Source Rocks and Reservoirs*. Springer, USA, pp. 85–106.

Holden J. F. (2009) Extremophiles: Hot Environments. In *Encyclopedia of Microbiology*. Academic Press, Massachusetts, pp. 127–146.

Inagaki F., Hinrichs K.-U., Kubo Y., Bowles M. W., Heuer V. B., Hong W.-L., … Yamada, Y. (2015). Exploring deep microbial life in coal-bearing sediment down to 2.5 km below the ocean floor. *Science* **349**(6246), 420–424.

Jannasch H., Nelson D. and Wirsen C. (1989) Massive natural occurrence of unusually large bacteria (Beggiatoa sp.) at a hydrothermal deep-sea vent site. *Nature*. **342**, 834–836.

Jeanthon C., L'Haridon S., Cueff V., Banta A., Reysenbach A.-L. and Prieur D. (2002) Thermodesulfobacterium hydrogenophilum sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a deep-sea hydrothermal vent at Guaymas Basin, and emendation of the genus Thermodesulfobacterium. *Int. J. Syst. Evol. Microbiol.* **52**, 765–772.

Jeffree C. E. (1996) Structure and ontogeny of plant cuticles. In *Plant Cuticles: An Integrated Functional Approach*. BIOS Scientific, Oxford, pp. 33–85.

Johnson K., Childress J. and Beehler C. (1988) Short-term temperature variability in the Rose Garden hydrothermal vent field: an unstable deep-sea environment. Deep Sea Research Part A. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **35**(10-11), 1711–1721.

Kashefi, K. and Lovley D. R. (2003) Extending the upper temperature limit for life. *Science*. **301**, 934–935.

Kawka E. (1990) Hydrothermal alteration of sedimentary organic matter in Guaymas Basin, Gulf of California. Ph.D. Thesis, Oregon State Univ. Corvallis.

Kawka E. and Simoneit B. (1987) Survey of hydrothermally-generated petroleums from the Guaymas Basin spreading center. *Org. Geochem.* **11**, 3 1 l–328.

Kawka E., Simoneit, B. (1990) Polycyclic aromatic hydrocarbons in hydrothermal petroleums from the Guaymas Basin spreading center. *Appl. Geochem*. **5**, 17–27.

Kawka E. and Simoneit B. (1994) Hydrothermal pyrolysis of organic matter in Guaymas Basin: I. Comparison of hydrocarbon distributions in subsurface sediments and seabed petroleums. *Org. Geochem.* **22**(6), 947–978.

Kelts K., Curray J. and Moore, D. (1982) Introduction and explanatory notes, DSDP leg 64 (California Basins). In *Initial reports of the Deep Sea Drilling Project-LXIV*. U.S. Govt. Printing Office, Washington DC.

Killops S. D. and Al-Juboori, M. A. H. A. (1990) Characterization of the unresolved complex mixture (UCM) in the gas chromatograms of biodegraded petroleums. *Org. Geochem.* **15**, 147–160.

Kleindienst S., Ramette A., Amann R. and Knittel K. (2012) Distribution and in situ abundance of sulfate-reducing bacteria in diverse marine hydrocarbon seep sediments. *Environ. Microbiol.* **14**, 2689–2710

Kluesner J., Lonsdale P. and González-Fernández A. (2014) Late Pleistocene cyclicity of sedimentation and spreading-center structure in the Central Gulf of California. *Mar. Geol.* **347**, 58–68.

Kunst L. (2003) Biosynthesis and secretion of plant cuticular wax. *Prog. Lipid Res.* **42**(1), 51–80.

Kvenvolden K. A., Rapp J. B., Hostettler F. D., Morton J. L., King J. D. and Claypool G. E. (1986) Petroleum Associated with Polymetallic Sulfide in Sediment from Gorda Ridge. *Science* **234**(4781), 1231–1234.

Kvenvolden K. and Simoneit B. (1990) Hydrothermally derived petroleum: Examples from Guaymas Basin, Gulf of California and Escanaba Trough, Northeast Pacific. *AAPG Bull.* **74**, 223–237.

Larson R. L., Menard H. W. and Smith S. M. (1968) Gulf of California: A Result of Ocean-Floor Spreading and Transform Faulting. *Science* **161**(3843), 781–784.

Lawver L. A., Williams D. L. and Herzen R. P. von (1975) A major geothermal anomaly in the Gulf of California. *Nature* **257**, 23–28.

Liu X-. K., Birgel D., Elling F. J., Sutton P. A., Lipp J. S., Zhu R., Zhang C., Könneke M., Peckmann J., Rowland S. J., Summons R. E. and Hinrichs K-. U. (2016) From ether to acid: a plausible degradation pathway of glycerol dialkyl glycerol tetraethers. *Geochim. Cosmochim. Acta* **183**,138–152.

Lloyd K. G., Schreiber L., Petersen D. G., Kjeldsen K. U., Lever M. A., Steen A. D., Stepanauskas R., Richter M., Kleindienst, S., Lenk S., Schramm A., Jørgensen B.

B., Mackenzie, A. S., Brassell, S. C., Eglinton, G. and Maxwell, J. R. (1982) Chemical Fossils: The Geological Fate of Steroids. *Science* **217**(4559), 491–504.

Lonsdale P. and Becker K. (1985) Hydrothermal plumes, hot springs, and conductive heat flow in the southern trough of Guaymas Basin. *Earth Planet. Sci. Lett*. **73**, 211–225.

Lonsdale P., Bischoff J. L., Burns V. M., Kastner M. and Swccney R. E. (1980) A high-temperature thermal deposit on the seabed at a Gulf of California spreading center. *Earth Planet. Sci. Lett.* **49**, 8–20.

MacGregor B. J., Biddle J. F., Harbort C., Matthysse A. G. and Teske A. (2013). Sulfide oxidation, nitrate respiration, carbon acquisition, and electron transport pathways suggested by the draft genome of a single orange Guaymas Basin Beggiatoa (Cand. Maribeggiatoa) sp. filament. *Mar. Genomics* **11**, 53–65

Magenheim A. J. and Gieskes J. M. (1992) Hydrothermal discharge and alteration in near-surface sediments from the Guaymas Basin, Gulf of California. *Geochim. Cosmochim. Acta,* **56**(6), 2329–2338.

Marchand, M., Caprais, J.C., Corre, S., Jacq, E. and Hussein, D. (1994) Utilisation des hydrocarbures par la microflore bacterienne du site hydrothermal du bassin de Guaymas (golfe de Californie). *Oceanologica Acta* **17**, 177–189.

Marshall A. G., Hendrickson C. L. and Jackson G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17**(1), 1–35.

McCollom T., Simoneit B. and Shock E. (1999) Hydrous pyrolysis of polycyclic aromatic hydrocarbons and implications for the origin of PAH in hydrothermal petroleum. *Energ. Fuel.* **13** (2), 401–410.

McDermott J. M., Ono S., Tivey M. K., Seewald J. S., Shanks W. C. and Solow A. R. (2015) Identification of sulfur sources and isotopic equilibria in submarine hot-springs using multiple sulfur isotopes. *Geochim. Cosmochim. Acta* **160**, 169–187.

McKay L., MacGregor B., Biddle J., Albert D., Mendlovitz H., Hoer D., Lipp J., Lloyd K. and Teske A. (2012) Spatial heterogeneity and underlying geochemistry of phylogenetically diverse orange and white Beggiatoa mats in Guaymas Basin hydrothermal sediments. *Deep-Sea Res.* **67**, 21–31.

Millar A. A. (1999) CUT1, an Arabidopsis Gene Required for Cuticular Wax Biosynthesis and Pollen Fertility, Encodes a Very-Long-Chain Fatty Acid Condensing Enzyme. *Plant Cell* **11**(5), 825–838.

Moore, D.(1973) Plate-Edge Deformation and Crustal Growth, Gulf of California Structural Province. *Geol. Soc. Am. Bull*. **84**(6), 1883–1905.

Mullins O. C., Rodgers R. P., Weinheber P., Klein G. C., Venkataramanan L., Andrews A. B. and Marshall A. G. (2006) Oil reservoir characterization via crude oil analysis by downhole fluid analysis in oil wells with visible−Near-Infrared Spectroscopy and by Laboratory Analysis with electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Energ. Fuel.* **20**(6), 2448–2456.

Nelson D., Wirsen C. and Jannasch H. (1989) Characterization of large, autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas Basin. *Appl. Environ. Microbiol.* **55**, 2909–2917.

Ourisson G., Albrecht P. and Rohmer M. (1979) The Hopanoids: palaeochemistry and biochemistry of a group of natural products. *Pure Appl. Chem.* **51**, 709–729.

Peter J. M., Simoneit B. R. T., Kawka E. and Scott S.D. (1990) Liquid hydrocarbon-bearing inclusions in modern hydrothermal chimneys and mounds from the southern trough of Guaymas Basin. *Appl Geochem.* **5**, 51–63.

Peter J., Peltonen P., Scott S., Simoneit B. and Kawka E. (1991) Carbon-14 ages of hydrothermal petroleum and carbonate in Guaymas Basin, Gulf of California Implications for oil generation, expulsion, and migration. *Geology* **19**, 253–256.

Peters K. E. and Moldowan J. M. (1993) *The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments*. Prentice Hall, New Jersey.

Peters K., Walters C. and Moldowan J. M. (2007) *The Biomarker Guide: Biomarkers and Isotopes in Petroleum Systems and Earth History,* Vol. 2. Cambridge University Press, UK.

Peters K., Walters C. and Moldowan, J. (2005) *The Biomarker Guide. Biomarkers and Isotopes in the Environment and Human History,* Vol. 1. Cambridge University Press, UK.

Philp R. (1985) *Fossil Fuel Biomarkers: Applications and Spectra*. Elsevier, Amsterdam.

Radke, M. (1988) Application of aromatic compounds as maturity indicators in source rocks and crude oils. *Mar. Pet. Geol.* **5**(3), 224–236.

Radke M., Rullkötter J. and Vriend S. P. (1994) Distribution of naphthalenes in crude oils from the Java Sea: Source and maturation effects. *Geochimica et Cosmochimica Acta* **58**(17), 3675–3689.

Radke M., Welte D. H. and Willsch H. (1986) Maturity parameters based on aromatic hydrocarbons: Influence of the organic matter type. Org. Geochem. **10**(1-3), 51–63.

Radke M., Welte D.H., Willsch H., (1982). Geochemical study of a well in the Western Canada Basin: relation of the aromatic distribution pattern to maturity of organic matter. *Geochim. Cosmochim. Acta* **46**, 1–10.

Raju, S. V., Mathur N. and Sarmah M. K. (2014) Geochemical characterization of Neoproterozoic heavy oil from Rajasthan, India: implications for future exploration of hydrocarbons. *Curr. Sci.* **107**(8), 1298–1305.

Reddy C. M., Eglinton T. I., Hounshell A., White H. K., Xu L., Gaines R. B. and Frysinger G. S. (2002) The West Falmouth oil spill after thirty years: the persistence of petroleum hydrocarbons in marsh sediments. *Environ. Sci. Technol.* **36**, 4754–4760.

Reeves E. (2010) Laboratory and Field-Based Investigations of Subsurface Geochemical Processes in Seafloor Hydrothermal Systems. Ph. D. thesis, Massachusetts Institute of Technology.

Rohmer M., Anding C., and Ourisson G. (2005) Non-specific Biosynthesis of Hopane Triterpenes by a Cell-Free System from Acetobacter pasteurianum. *Eur. J. Biochem*. **112**(3), 541–547.

Rowland S. J. and Maxwell J. R. (1984) Reworked triterpenoid and steroid hydrocarbons in a recent sediment. *Geochim. Cosmochim. Acta* **48**, 617–624.

Rueter P., Rabus R., Wilkest H., Aeckersberg F., Rainey F. A., Jannasch H. W. and Widdel, F. (1994) Anaerobic oxidation of hydrocarbons in crude oil by new types of sulphate-reducing bacteria. *Nature*, **372**(6505), 455–458

Rusnak G., Fisher R. and Shepard F. (1964) Bathymetry and Faults of Gulf of California. in, Andel, Tj. H. Van, And Shor, G. G., eds., Marine geology of the Gulf of California: *Am. Assoc. Petroleum Geologists Mem*. **3**, 59–75.

Simoneit B. R. T. (1984) Hydrothermal effects on organic matter-high versus low temperature components. *Org. Geochem.* **6** 857–864.

Simoneit B. R. T. (1985) Hydrothermal petroleum: genesis, migration, and deposition in Guaymas Basin, Gulf of California. *Can. J. Earth Sci*. **22**, 1919–1929.

Simoneit B. R. T. (1990) Petroleum generation, an easy and widespread process in hydrothermal systems: an overview. *Appl. Geochem*. **5**(1-2), 3–15.

Simoneit B. R. T. (1992) Natural hydrous pyrolysis: Petroleum generation in submarine hydrothermal systems. In *Organic Matter: Productivity, Accumulation, and*

Preservation in Recent and Ancient Sediments (eds. J. Whelan and J. W. Farrington). Columbia University Press, New York, pp. 368–403.

Simoneit B. R. T. (1993) Hydrothermal activity and its effects on sedimentary organic matter. In *Bitumens in Ore Deposits* (eds. J. Parnell, H. Kucha and P. Landais). Springer-Verlag, Berlin, pp. 81–96.

Simoneit, B.R.T. (1994) Lipid/bitumen maturation by hydrothermal activity in sediments of Middle Valley, Leg 139. In *Proceedings of the Ocean Drilling Program, Scientific Results,* Vol. 139 (eds. M. J. Mottl, E. E. Davis, A. T. Fisher, and J. F. Slack). Texas A & M University Ocean Drilling Program, USA. pp. 447–465.

Simoneit B. R. T. and Galimov E. M. (1984) Geochemistry of interstitial gases in Quaternary sediments of the Gulf of California. *Chem. Geol.* **43**, 151–166

Simoneit B. R. T. and Kvenvolden, K. (1993) Comparison of ^{14}C ages of hydrothermal petroleums. *Org. Geochem.* **21**, 525–529.

Simoneit B. R. T. and Lonsdale P. (1982) Hydrothermal petroleum in mineralized mounds at the seabed of Guaymas Basin. *Nature* **295**, 198–202.

Simoneit B. R. T., Leif R. N., Sturz A. A., Sturdivant A. E. and Gieskes J. M. (1992a) Geochemistry of shallow sediments in Guaymas Basin, gulf of California: hydrothermal gas and oil migration and effects of mineralogy. *Org. Geochem.* **18**(6), 765–784.

Simoneit B. R. T., Kawka O. E. and Wang G.-M. (1992b) Biomarker maturation in contemporary hydrothermal systems, alteration of immature organic matter in zero geological time. In *Biological Markers in Sediments and Petroleum, A Tribute to W. K. Seifert* (eds J. M. Moldowan, P. Albrecht and R. P. Philp). Prentice Hall, New Jersey, pp. 124-141.

Teske A., de Beer D., McKay L., Tivey M., Biddle J., Hoer D., Lloyd K., Lever M., Roy H., Albert D., Mendlovitz H. and MacGregor B. (2016) The Guaymas Basin hiking guide to hydrothermal mounds, chimneys, and microbial mats: complex seafloor expressions of subsurface hydrothermal circulation. *Front. Microbiol.* **7**, 75–98.

Teske A., Hinrichs K.-U., Edgcomb V. P., de Vera Gomez A., Kysela D., Sylva S. P., Sogin M. and Jannasch H. (2002) Microbial diversity in hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl. Environ. Microbiol.* **68**, 1994–2007.

Tissot B., Califet-Debyser Y., Deroo G. and Oudin J. L. (1971) Origin and evolution of hydrocarbons in early Toarcian shales, Paris Basin, France. *AAPG Bull.* **55**, 2177–2193.

Tissot B. and Welte D. (1984) *Petroleum Formation and Occurrence.* Springer-Verlag, Berlin.

Tran T., Logan G., Grosjean E., Ryan D. and Marriott P. (2010) Use of comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for the characterization of biodegradation and unresolved complex mixtures in petroleum. *Geochim. Cosmochim. Acta* **74**, 6468–6484.

Van Aarssen B. G. ., Bastow T. P., Alexander R. and Kagi R. I. (1999) Distributions of methylated naphthalenes in crude oils: indicators of maturity, biodegradation and mixing. *Org. Geochem.* **30**(10), 1213–1227.

Van Andel Tj. H. (1964) Recent marine sediments in the Gulf of California. In *Marine Geology of the Gulf of California: a symposium* (eds. Tj. H. van Andel and G. G. Shor Jr.) American Association of Petroleum Geologists, Oklahoma. pp. 216–311.

Ventura G. T., Kenig F., Reddy, C. M., Schieber J., Frysinger G. S., Nelson R. K., Dinel E., Gaines R. B. and Schaeffer P. (2007) Molecular evidence of Late Archean archaea and the presence of a subsurface hydrothermal biosphere. *PNAS* **104**(36), 14260–14265.

Ventura G. T., Kenig F., Reddy C. M., Frysinger G. S., Nelson R. K., Mooy B. V., and Gaines R. B. (2008) Analysis of unresolved complex mixtures of hydrocarbons extracted from Late Archean sediments by comprehensive two-dimensional gas chromatography (GC×GC). *Org. Geochem.* **39**(7), 846–867.

Ventura G. T., Simoneit B. R. T., Nelson R. K. and Reddy C. M. (2012) The composition, origin and fate of complex mixtures in the maltene fractions of hydrothermal petroleum assessed by comprehensive two-dimensional gas chromatography. *Org. Geochem.* **45**, 48–65.

Venkatesan M. (1987) Diploptene in Antarctic sediments. *Geochim. Cosmochim. Acta* **52**, 217–222.

Von Damm K. L., Edmond J. M., Measures C. I. and Grant B. (1985) Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochim. Cosmochim. Acta* **49**(11), 2221–2237.

Wang Z., Fingas M., Yang C. and Christensen J. H. (1964) Crude Oil and Refined Product Fingerprinting: Principles. In *Environmental Forensics: Contaminant Specific Guide*. Elsevier Academic Press, Boston. 339–407.

Wang, Z. and Stout S. (2007) *Oil Spill Environmental Forensics. Fingerprinting and Source Identification*. Elsevier, Amsterdam.

Warth A. D. (1978) Relationship between the Heat Resistance of Spores and the Optimum and Maximum Growth Temperatures of Bacillus Species. *J. Bacteriol*. **134**(3), 699–705.

Warton B., Alexander R. and Kagi R. I. (1999) Characterization of the ruthenium tetroxide oxidation products from the aromatic unresolved complex mixture of a biodegraded crude oil. Org. Geochem. **30**, 1255–1272.

Weber A. and Jørgensen B. (2002) Bacterial sulfate reduction in hydrothermal sediments of the Guaymas Basin, Gulf of California, Mexico. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **49**(5), 827–841.

Wenger L. M., Davis C. L. and Isaksen G. H. (2002) Multiple Controls on Petroleum Biodegradation and Impact on Oil Quality. *SPE Reserv. Eval. Eng.* **5**, 375–383.

Welhan J. A. and Lupton J. E. (1987) Light hydrocarbon gases in Guaymas Basin hydrothermal fluids: thermogenic versus abiogenic origin. *AAPG Bull*. **71**, 215–223.

Whelan J. K., Simoneit B. R. T. and Tarafa M. E. (1988) C1-C8 hydrocarbons in sediments from Guaymas Basin, Gulf of California – Comparison to Peru Margin, Japan Trench, and California Borderlands. *Org. Geochem*. **12**, 171–194.

Williams D., Becker K., Lawver L. and Von Herzen R. (1979) Heat flow at the spreading centers of the Guaymas Basin, Gulf of California. *J. Geophys. Res.* **84**, 6757–6796.

Yamanaka, T., Ishibashi J. and Hashimoto J. (2000) Organic geochemistry of hydrothermal petroleum generated in the submarine Wakamiko caldera, southern Kyushu, Japan. *Org. Geochem.* **31**, 1117–1132.

Zierenberg R. A., Miller D. J. (2000) Overview of ocean drilling program leg 169: Sedimented ridges II. In *Proceedings of the Ocean Drilling Program, Scientific Results,* Vol. 169. ODP, Texas.

Zhuang G.-C., Montgomery A., Samarkin V. A., Song M., Liu J., Schubotz F., Teske A., Hinrichs K.-U. and Joye S. B. (2019) Generation and utilization of volatile fatty acids and alcohols in hydrothermally altered sediments in the Guaymas Basin, Gulf of California. *Geophys. Res. Lett.* **46**(5), 2637–2646.

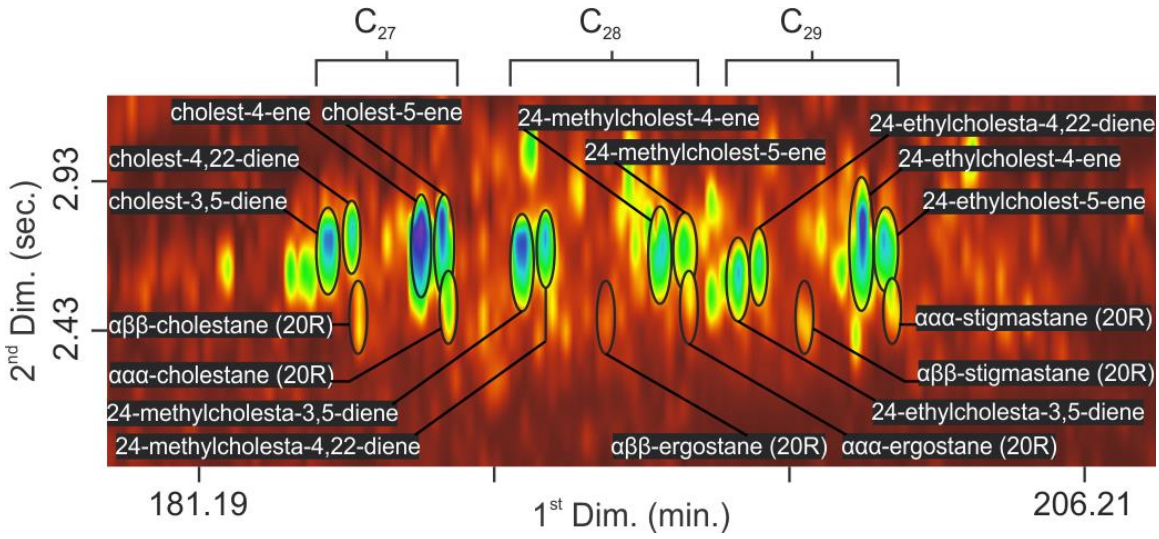
Chapter 2 Supplementary: Thermochemical and Microbial Biomarker Transformations in Hydrothermally-Altered Sediments at Cathedral Hill, Guaymas Basin, Gulf of California

Connor J. Dalzell[1], G. Todd Ventura[1*], Robert K., Nelson[2], Clifford C. Walters[3], Christopher M. Reddy[2], Jeffrey S. Seewald[2], Stefan M. Sievert[2]

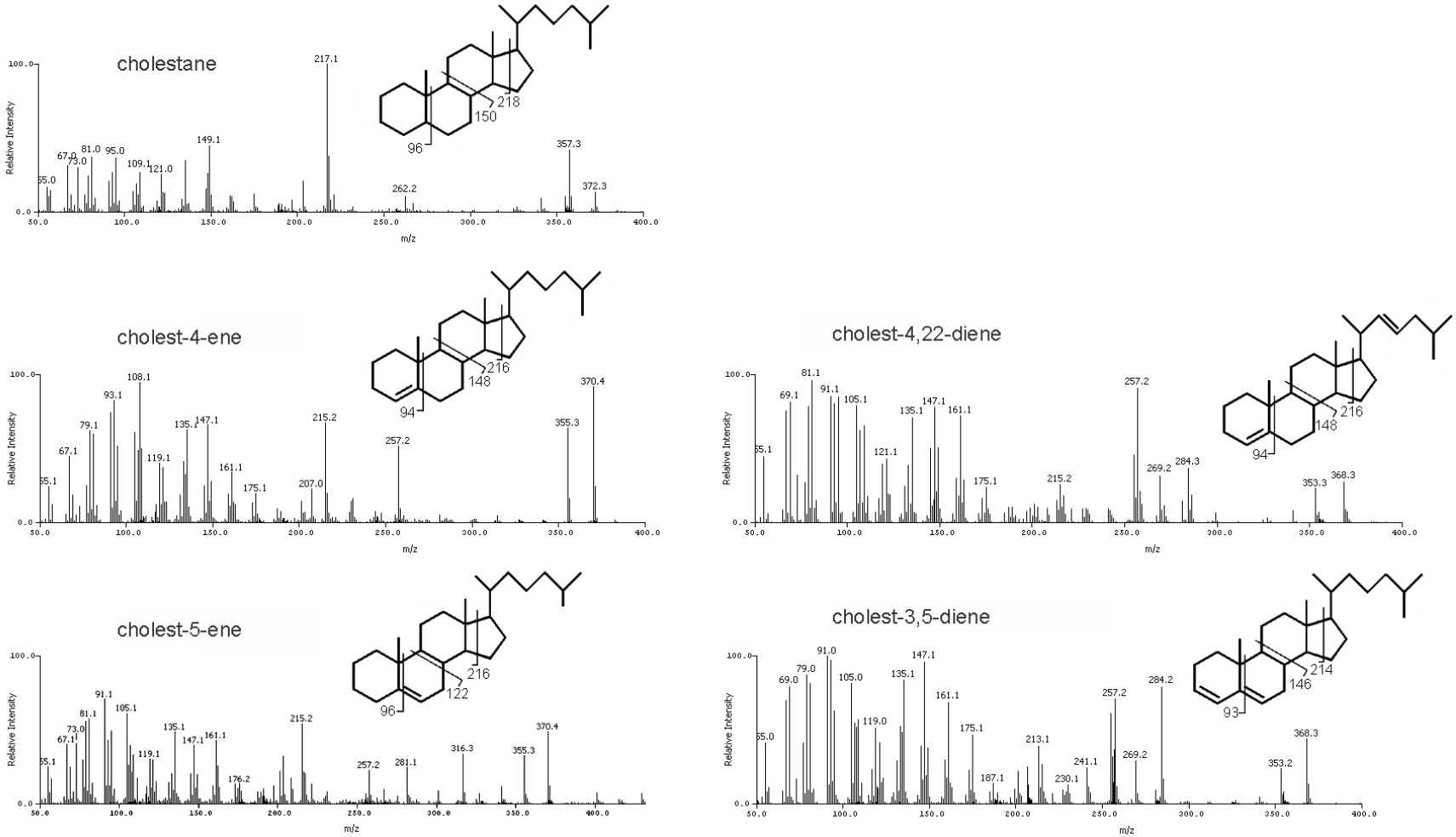
Affiliations:
[1] Saint Mary's University, Canada.
[2] Woods Hole Oceanographic Institution, USA.
[3] ExxonMobil Research and Engineering, USA.
[*] Corresponding author: todd.ventura@smu.ca

Supplemental Information

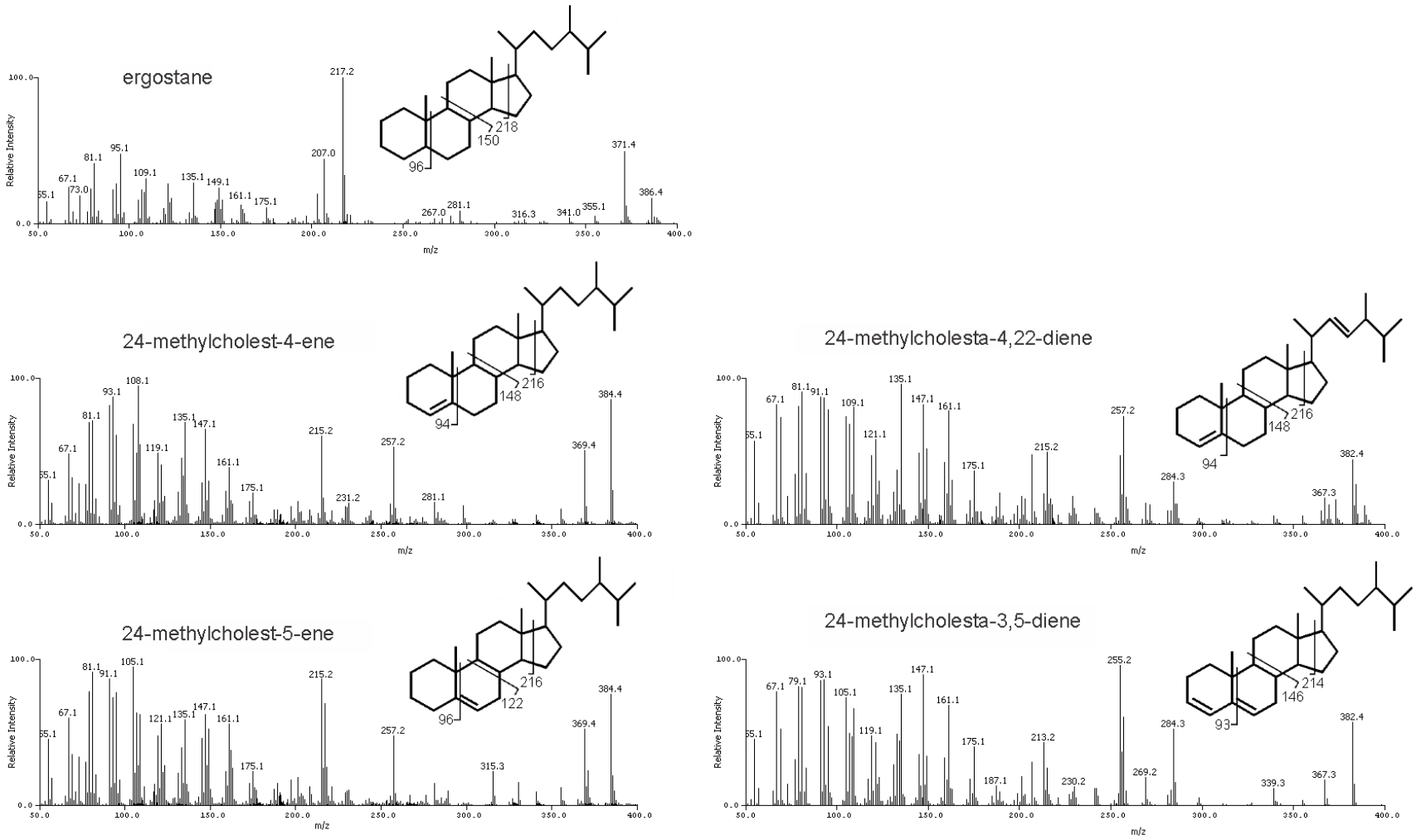
SI Figures:



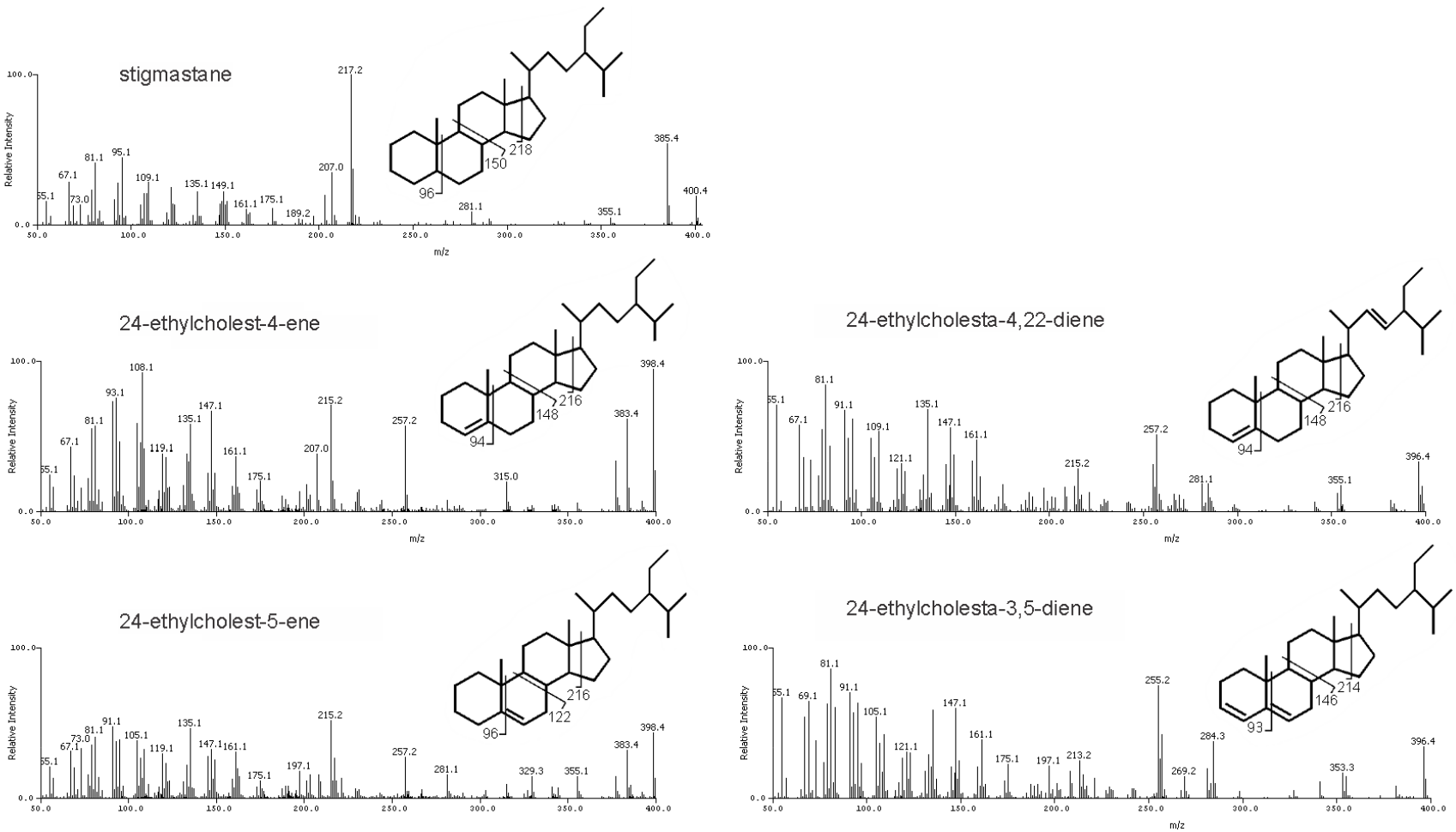
SI Fig. 1. Elution profile of the tentatively identified steroid biomarkers.



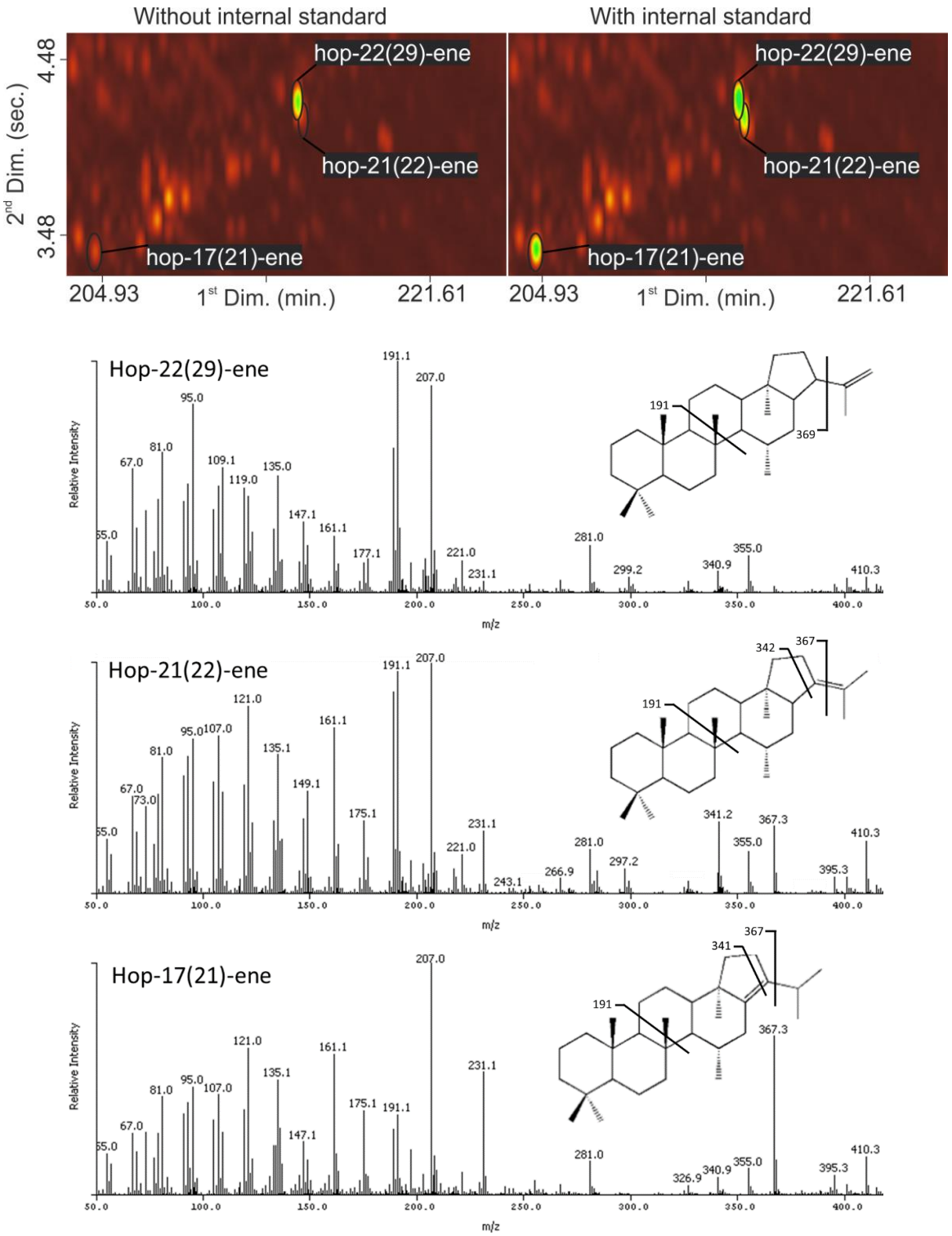
SI Fig. 2. Mass spectra of C_{27} steroids (note: *m/z* 207 is a contaminant from column bleed).



SI Fig. 3. Mass spectra of C_{28} steroids (note: *m/z* 207 is a contaminant from column bleed).



SI Fig. 4. Mass spectra of C_{29} steroids (note: m/z 207 is a contaminant from column bleed).



SI Fig. 5. GC×GC-FID chromatogram marking hopanoids both with and without authentic internal standards elution profiles of hopenes (top) and their associated matrix identified mass spectra (bottom; note: m/z 207 is a contaminant from column bleed).

Chapter 3: Resolution of multi-molecular hydrocarbon transformation in petroleum-bearing sediments from the Cathedral Hill hydrothermal vent complex at Guaymas Basin, Gulf of California by comprehensive two-dimensional gas chromatography and chemometric analyses

Connor J. Dalzell[1], G. Todd Ventura[1*], Robert K., Nelson[2], Clifford C. Walters[3], Christopher M. Reddy[2], Jeffrey S. Seewald[2], Stefan M. Sievert[2]

Affiliations:

[1] Saint Mary's University, Canada.

[2] Woods Hole Oceanographic Institution, USA.

[3] ExxonMobil Research and Engineering, USA.

[*] Corresponding author: todd.ventura@smu.ca

Target Journal: Geochimica et Cosmochimica Acta or Organic Geochemistry or Chemical Geology

Abstract

Guaymas Basin is a submarine depression at the northern end of the East Pacific Rise, mid-oceanic spreading ridge in the Gulf of California. The basin receives high inputs of organic matter from elevated productivity in the overlying surface waters and runoff from the surrounding continent. This high sedimentation rate produces near-uniform compositions of sedimentary organic matter (SOM). Various hydrothermal vent complexes occur along this margin. One of these is Cathedral Hill, which comprises a cluster of white smokers that are themselves partially overlain and surrounded by a *Beggiatoa*-dominated sulfur oxidizing microbial mat. Porewater temperatures near the vent edifice are projected to reach 155 °C by 21 cm sediment depth below the sea floor (cmbsf). Based on a vitrinite reflectance equivalent kinetic model, this environment is capable of pyrolyzing the SOM into petroleum forming hydrocarbons as shallow as 15-18 cmbsf, with metagenesis commencing at less than 60 cmbsf. Although many studies have examined how these petroleums are formed and modified, few have examined the holistic hydrocarbon transformations that occur. In recent years, application of chemometric techniques to comprehensive two-dimensional gas chromatography (GC×GC) analyses of oils has allowed for the comparison of thousands of individual hydrocarbons across datasets. At Cathedral Hill, apolar fractions with up to 5,700 unique compounds spanning a range of saturate and unsaturated biomarkers, alkanes, branched alkanes, substituted and unsubstituted polycyclic aromatic hydrocarbons (PAHs), perhydroPAHs, and benzothiophenes with up to four ring-cycles (i.e. chrysenes). We applied multiway principal components analysis (MPCA) and hierarchical cluster analysis (HCA) directly to GC×GC chromatograms of samples sourced from four push cores (n=34) along a transect stretching from near the vent center to the exterior of the microbial mat to identify multi-molecular changes in the subsurface and elucidate how these hydrocarbon matrices are attenuated by in situ hydrocarbon production, migration, or thermochemical weathering.

From these measures as well as a systematic survey of subtracted GC×GC chromatograms, measurements of hydrocarbon diversities, and thermal maturation parameters a reconstruction of the shallow subsurface petroleum system is resolved. Collectively, these data indicate that whole or partial in situ pyrolysis likely begins at 8-10 cmbsf and 15-18 cmbsf in push cores nearest to the vent center (marking vent porewater temperatures of ~110 °C). These depths are shallow with respect to the kinetic model, indicating higher temperature flux in the past or additional mechanisms promoting in-situ pyrolysis within these layers. Independent of these sites of generation and thermochemical destruction is a ubiquitous staining of the sediments from advected oil that shows systematic temperature-dependent, stereochemical (resolved from biomarker abundances) and structural alteration across the transect area. In this regard, the MPCA models along with the validation techniques show evidence of decreasing diversities and abundances of alkylated aromatic hydrocarbons and the production of dealkylated "parent" molecules, steroids, and hopanoid biomarkers in more hydrothermally impacted sediments. Overprinting this profile is a band of migrated oil that extends across the transect at ~8-10 cmbsf. Overall, the results indicate that even at relatively small spatial scales the petroleum systems of sedimented hydrothermal vent sites can be highly complex.

3.1 Introduction

Since their initial discovery at the Galapagos rift (Lonsdale, 1977; Corliss et al. 1979), deep-sea hydrothermal vents have been intensely studied environments for biology, geology, and geochemistry, owed to their unique ecological, morphological, and chemical properties and relative commonality associated with ocean spreading centers. These sites are of particular interest to petroleum geologists due to the production of oils in sedimented deep-sea hydrothermal systems, which has been shown to be a geologically rapid processes

(e.g., days-years) that can be confined to a narrow sediment interval near the seafloor surface (e.g. Simoneit and Lonsdale, 1982; Simoneit, 1985, 1993; Bazylinski et al., 1988; Peter et al., 1991; Rushdi and Simoneit, 2002a, b; Simoneit and Kvenvolden, 1993). In contrast, typical reservoir crude oils normally require millions of years to form, involving basin subsidence, kilometers of sediment accumulation, and large-scale migration processes (Hunt 1979; Tissot and Welte, 1984; Selley, 1997). The contrasting miniaturization in both space and time of hydrothermal systems provides a unique opportunity to study the natural multi-molecular changes that occurs to organic matter (OM) over millions of years in normal marine sedimentary basins.

The circulating high-temperature fluids of hydrothermal systems induces generation and mass transfer of petroleum within unconsolidated sediments (Didyk and Simoneit, 1989, 1990; Simoneit and Lonsdale, 1982; Simoneit, 1984; Kvenvolden et al., 1986; Didyk and Simoneit, 1989; Kvenvolden and Simoneit, 1990; Simoneit, 1993). Rapid, hydrothermally-driven alteration of OM into petroleum occurs from ~60 to >400 °C via cleavage of molecular bonds, or bond reformation in the case of polycyclic aromatic hydrocarbons (PAHs) (Peters et al., 2007; Simoneit, 1993; McCollom et al., 1999). These thermogenic products are soluble in the surrounding fluids at high temperatures (Connolly 1966; Sanders 1986; Ventura et al., 2012), and are subsequently removed via convection (Simoneit, 1993). These oils are either discharged into the water column or collect near the seafloor, where low temperatures promote condensation and retention of the pyrolysate (Didyk and Simoneit, 1989; Simoneit, 1984, 1993; Kawka, 1990; Peter et al. 1990). In this way, the production of hydrothermal oils feeds into the carbon cycle dynamics of Guaymas

Basin, whereby buried SOM is reintroduced to the water column/benthic environment (Pearson et al., 2005; LaRowe et al., 2020).

Laboratory pyrolysis experiments (e.g. Hoering, 1968; Ishiwatari et al., 1977; Van Graas et al., 1981; Horsfield, 1984; Lewan, 1985; Lewan et al., 1986; Eglinton and Douglas, 1988; Leif and Simoneit, 1995; McCollom et al., 1999; Zárate-del Valle et al., 2006) and surveys of naturally generated hydrothermal petroleums (e.g. Simoneit and Lonsdale, 1982; Simoneit, 1984, 1985, 1993; Kawka and Simoneit, 1987, 1990, 1994) have extensively documented the effects of thermal maturation, migration, and other processes on oils and kerogen. However, studies of this nature are invariably limited by the high complexity of naturally forming petroleums, which can be comprised of thousands of individual compounds (Brooks et al., 1957; Tissot and Welte, 1984; Marshall et al., 1988; Mullins et al., 2006), thus impeding the holistic tracking of hydrocarbons. In this regard, diagnostic compound ratios are often relied on to infer reaction rates, OM compositions, and thermal histories of oils, which are applied to the whole matrix of hydrocarbons (e.g. Boehm et al., 1997; Peters et al., 2005, 2007; Wang and Stout, 2007; Christensen et al., 2016). While this may be a valid approach for many compound classes, such scaling-up may not accurately reflect how hydrocarbon matrices holistically change. To our understanding, this has not been fully investigated. Therefore, this project will explore if multi-molecular transformations can be holistically tracked within hydrothermal oils, and if these changes can be attributed to migration, oil formation, or other processes within these systems.

3.1.1 Considerations for studying multi-molecular change

Additional issues arise from the fact that petroleums are approximately three orders of magnitude more complex than traditional analytical techniques, such as gas chromatography–mass spectrometry (GC–MS), are capable of resolving (Tissot and Welte, 1984; Marshall et al., 1998; Mullins et al., 2006), which could obscure multi-molecular changes and oil-modifying processes. To mitigate these effects, and holistically identify multi-molecular changes, a combination of multi-dimensional gas chromatography (GC×GC) and multivariate statistics can be used in the analysis of hydrothermal oils. The former provides enhanced peak capacity and resolving power relative to traditional gas chromatography by separating hydrocarbons onto two dimensions, eliminating the problem of co-elution inherent to traditional "1D" GC methods (e.g. Barman and Membrado, 2000; Reddy et al., 2002; Frysinger et al., 2002; Arey et al., 2005, 2007; Ventura et al., 2007, 2008, 2012; Tran et al., 2010). The latter has been proven as a capable technique to objectively compare and classify oils/oil mixtures, typically using the entire suite of hydrocarbons rather than relying on specific marker ratios or concentrations (e.g. Ventura et al., 2011; Christensen et al., 2016).

3.1.1.1 Multi-way principal components analysis

Multi-way principal components analysis (MPCA) is an exploratory multivariate technique used to reduce the dimensionality of multi-dimensional datasets (Wold et al., 1987; Wise et al., 2006; Haiping et al., 2008). The method has been shown to be an effective tool for the molecular compositional analysis of oils analyzed by GC×GC (Ventura et al., 2011; Prata et al., 2016; Mogollón et al., 2017). The algorithm models variation within a

data arrays via its decomposition into scores and loadings matrices, which are based upon the eigenvectors describing the greatest source of variance within the dataset (Wise et al., 2006). When applied to GC×GC data, chromatograms are stacked into three-dimensional arrays ($\bar{X}$) and unfolded row-wise (X), mean-centered, and further decomposed into score (t), loading (P), and residual matrices (E) in a processes summarized by:

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{t}_r \otimes \mathbf{P}_r + \mathbf{E} \qquad (1)$$

where r is the ideal factor number and $\otimes$ indicates the outer product operator (Wise et al., 2006; Mogollón et al., 2017).

The data outputs are scores and factor loadings plots, which are calculated with respect to a principal component (PC) that accounts for a fraction of the total variance resolved by the model. Scores reflect a sample's multivariate relatedness expressed as

$$t_{ac} = (X_{1a}P_{1c} - e_1) + (X_{2a}P_{2c} - e_2) + \ldots + (X_{ia}P_{ic} - e_i) \qquad (2)$$

where t is the score value for the a^{th} sample on the c^{th} principal component; X_1 to X_i are variables (i.e. hydrocarbon peak intensities) within the a^{th} sample; P_1 to P_i are their weights (loadings) with respect to the c^{th} PC and e marks the degree of error associate with each measurement (after Manly, 1994). When applied to GC×GC data, comparisons of very similar organic matrices result in near-zero loadings, having little influence on any PC generated by the model. Organic matrices with different compositions produce loadings that depart from zero, with the sign indicating how the original variables contribute to a PC. For oils analyzed by GC×GC, a net-positive score on any PC requires a sample to contain a greater overall abundance of positively-weighted hydrocarbons and lower abundances of negatively-weighted hydrocarbons. Thus, score values summarize overall

hydrocarbon distributions within samples as described by the loadings on a given PC. Linked to this are factor loadings plots, which identify variables most responsible for influencing a sample's cumulative score. The resulting factor loadings plots are themselves reconstructed multi-dimensional chromatograms with the z-axis marking P_1 to P_i of a particular PC.

3.1.1.2 Hierarchical cluster analysis

Hierarchical cluster analysis (HCA) is a multivariate, agglomerative technique used to calculate the similarity between samples that are initially assumed to be part of a lone cluster. In this regard, samples are linked by a distance metric (i.e. Euclidean distance, Manhattan distance) and a linkage criterion (i.e. Ward's method, centroid method, etc.) that calculates the distances between linked sets as a function of dissimilarity (Johnson, 1967, Lance and Williams, 1967, Aldenderfer and Blashfield, 1984; Rencher, 2003). This technique has previously been used to analyze oil similarity using geochemical data (Peters and Fowler, 2002; Pavón et al., 2006; Barbeira et al., 2007; Peters et al., 2007; Wang et al., 2018; Zhan et al., 2019) and whole GC fingerprints of oil samples (Hashemi-Nasab et al., 2019), which uses a combination of Euclidean distance and Ward's method.

Euclidean distance (d) is defined as:

$$d = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2]^{1/2} \quad (3)$$

When applied to GC×GC data, x_n and y_n correspond to GC×GC signal intensities from separate chromatograms that are captured at the same 1st and 2nd dimension elution coordinates. This is followed by a linkage criterion calculation such as Ward's Method, which establishes linkages by minimizing squared Euclidean distances of samples within-

clusters with respect to each cluster's centroid (Ward, 1963). HCA therefore defines sample relatedness in the form of a dendrogram of linkages across a Euclidean space, that when applied to GC×GC data sets, groups sample chromatograms based on their overall compositional similarity.

3.1.2 Objectives

In this study, GC×GC is coupled with MPCA and HCA to identify multi-molecular changes within the apolar fractions of hydrothermal petroleum extracts sourced from a shallow sediments push core transect at Cathedral Hill, Guaymas Basin, Gulf of California. The transect sediments were exposed to a range of geothermal temperatures 0-155 °C based on proximity to the vent site and depth of sediment penetrated by the push core. The chemometric results are compared with alternative techniques for visualizing multi-molecular change (i.e. difference chromatograms, classification groups, and numerical molecular complexity proxies) in order to validate the statistical model and elucidate additional trends missed by the multivariate statistical analyses. The results are further compared to modeled SOM kinetic maturation calculations predicted by EasyR_o software (Burnham 2019) to help identify, if or to what degree, multi-molecular changes in the Cathedral Hill transect sediments are linked to in situ petroleum forming processes or advection of hydrocarbons from more deeply buried sediments.

3.2 Geological Setting

Guaymas Basin is a marginal rift basin marking the northernmost extension of the East Pacific Rise oceanic spreading center (Rusnak et al., 1964; Larson et al., 1968; Moore et al., 1973). The basin is comprised of a northern and southern trough (Fig. 2) formed by

axial-parallel faulting and extension along the graben (Moore, 1973; Lonsdale and Lawver, 1980, Lonsdale and Becker, 1985). Basaltic sills intrude into the overlying unconsolidated sediments (Einsele et al., 1980; Curray et al., 1982; Fisher and Becker 1991) resulting in extensive elevation of regional heat flow that in the Southern Trough is expresses as an extensive network of hydrothermal vents, seeps, mounds, and chimneys (Fig. 2; Curray et al., 1982; Lonsdale and Becker, 1985). In some localities circulating vent fluids reach ~350 °C near seafloor (Simoneit, 1984). The high temperature fluids rapidly pyrolyze hosted SOM thereby forming petroleum based fluids (Simoneit and Lonsdale, 1982; Simoneit, 1984, 1990, 1992), along with methane (Whelan et al., 1988), CO_2, ammonia (Von Damm et al., 1985), and low-molecular weight organic acids (Martens, 1990).

Cathedral Hill is located at 1996m depth in the Southern Trough (27°00.696 N, 111°24.265 W; Fig. 2). At this site, seafloor sediments are mainly composed of diatomaceous oozes and turbidite muds (Curray et al., 1982) sourced from the overlying highly productive euphotic zone with significant contributions of continental run-off from the Mexican mainland (Van Andel. 1964; Calvert, 1966; Curray et al., 1982; Kelts et al., 1982). Mixed with these sediments are sparse occurrences of calcareous shells and buried benthic macrofauna. Although the sedimentation rate is unknown at Cathedral Hill, it is likely similar to the 1-2 mm/yr estimated for other portions of Guaymas Basin (Williams et al., 1979; Teske et al., 2002). The vent complex hosts active white smoker chimneys, encrusted with giant tube worms (*Riftia pachyptila*) that are partially covered with and surrounded by a sulfide oxidizing, filamentous *Beggiatoa* microbial mat (McKay et al., 2012; Teske et al., 2016). Cathedral Hill occupies a regional heat-flow regime (500-1000 mW m^2), which is comparable to the background heat-flow of the Southern Trough floor

(660 mW m^2) and other nearby hydrothermal features (Fig. 2; Lonsdale and Becker, 1985; Teske et al., 2016). Previous temperature probe measurements taken at the site recorded pore water temperatures reaching 120 °C within 45 centimeters below seafloor (cmbsf; Teske et al., 2016), which is similar to the thermal gradients observed at other vent sites in the Southern trough (Bazylinski et al. 1988; Weber and Jørgensen, 2002; Teske et al., 2016). Dissolved organic carbon content (~0.5-1.5 mM within 40 cmbsf) and vent fluid chemistry are also comparable to other sites in the Guaymas Basin (Reeves, 2010; McDermott et al., 2015 Zhuang et al., 2019).

†

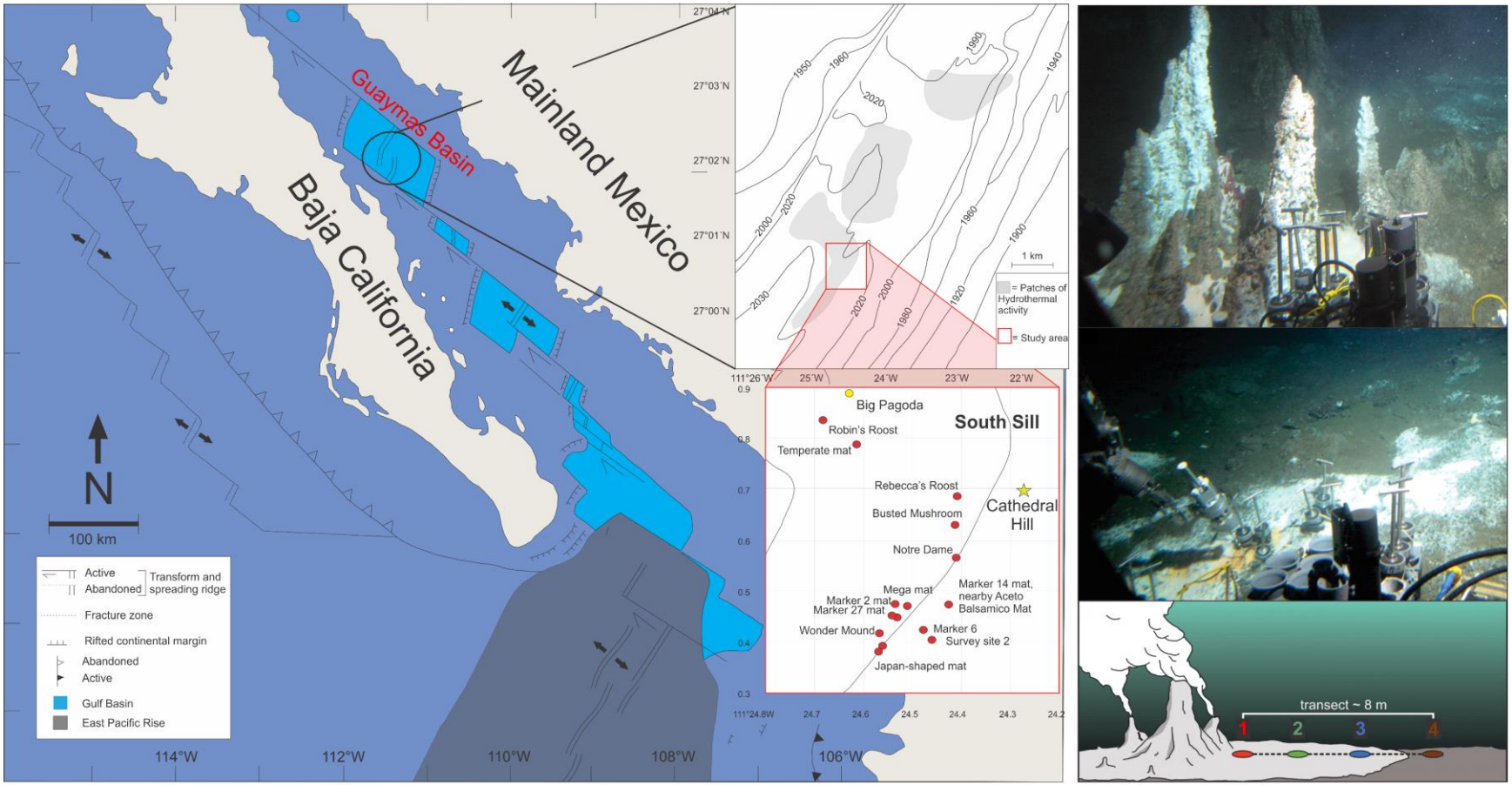


Fig. 1. Overview of Guaymas Basin showing the tectonic regime, Southern Trough, and Cathedral Hill. The red box marks the cluster of vents (red circles) in which Cathedral Hill (yellow star) is located. The right are photos of Cathedral Hill, Alvin push core sampling, and use of the thermal probe along with an illustration depicting the push core transect (modified from Castillo et al., 2002 and Teske et al., 2016).

3.3 Methods

3.3.1 Sample Collection

Using the human operated vehicle *Alvin* of Woods Hole Oceanographic Institution four near-equidistantly spaced (~2 m) push cores of petroleum-impregnated, unconsolidated sediments were collected along a transect running from the vent center to beyond the edge of the surrounding microbial mat (Fig. 1; Table 1). Measurements of porewater temperatures were collected alongside each push core using a thermal probe (analytical precision ±2 °C) in ~5 cm increments up to ~20 cmbsf (Table 1).

3.3.2 Sample Preparation

Once the push cores were brought to the surface, they were immediately sectioned into 2-3 cm increments, transferred to combusted gas vials, and frozen at -40 °C. The samples were later freeze dried and stored at -80 °C. Each sample was extracted three times via a CEM Mars Microwave Digestion System, which had a 15-20 minute ramp time to 115°C, with a hold of 15 min at 1030-1800 V, using a 3:1 mixture of dichloromethane (DCM) and methanol (MeOH). Apolar fractions of the total lipid extracts were collected via flash chromatography, using 60 ml of hexane to rinse the apolar (AP) fraction through a Pasteur pipette column filled with 30 g of activated silica gel (60-200 mesh). Elemental sulphur was then removed by passing the sample through a Pasteur pipette column filled with activated Cu (HCl 6N) powder using hexane as the eluent. Samples were then dried under a continuous flow of dry N_2 at 30 °C.

Table 1. *Temperature probe, sediment characteristics, and geochemical extract data of Cathedral Hill samples.*

Sample location and name[a]	Alvin Dive # and core ID	Porewater temperature (°C)	Interpolated porewater temperature (°C)	Description/lithology[b]	Extracted sediment wt (g)	TLE (mg)	Polar (mg)	Apolar + S^0 (mg)	Apolar (mg)
Microbial mat closest to vent									
Core 1 0-2 cmbsf	GB4462-5	19	19	Black mud with microbial mat filaments	1.78	50.16	40.31	3.54	2.27
Core 1 2-4 cmbsf	GB4462-5	-	67	Brownish-green diatomaceous mud	2	47.04	27.38	2.73	1.98
Core 1 4-6 cmbsf	GB4462-5	85	85	Brownish-green diatomaceous mud	2.68	29.14	19.81	1.25	1.24
Core 1 6-8 cmbsf	GB4462-5	-	105	Brownish-green diatomaceous mud	3.23	31.76	22.5	1.24	1.04
Core 1 8-10 cmbsf	GB4462-5	-	117	Brownish-green diatomaceous mud	4.56	28.24	12.61	0.98	1.1
Core 1 10-12 cmbsf	GB4462-5	121, 124	125	Grayish-green mud	5.18	17.77	8.49	0.94	0.99
Core 1 12-15 cmbsf	GB4462-5	-	135	Brownish-green consolidated mud with clay shards	4.34	14.19	9.93	0.94	0.86
Core 1 15-18 cmbsf	GB4462-5	142	145	Brownish-green consolidated clay	4.13	16.73	16.72	0.81	0.79
Core 1 18-21 cmbsf	GB4462-5	153	153	Brownish-green consolidated clay	3.74	9.75	2.31	0.65	0.32
Microbial mat further from vent									
Core 2 0-2 cmbsf	GB4462-6	9, 13	11	Black mud with microbial mat filaments	2.61	36.78	1.18	1.06	0.45
Core 2 2-4 cmbsf	GB4462-6	-	22	Black mud with microbial mat filaments	1.77	18.48	3.16	0.53	0.47
Core 2 4-6 cmbsf	GB4462-6	20	20	Brownish-green diatomaceous mud	3.01	22.16	2.41	0.53	0.29
Core 2 6-8 cmbsf	GB4462-6	-	47	Brownish-green diatomaceous mud	3.77	13.87	2.07	0.35	0.26

Core 2 8-10 cmbsf	GB4462-6	-	60	Brownish-green diatomaceous mud	4.87	11.21	4.25	0.58	0.54
Core 2 10-12 cmbsf	GB4462-6	69, 77	73	Brownish-green diatomaceous mud	6.44	9.74	2.46	0.38	0.18
Core 2 12-15 cmbsf	GB4462-6	-	87	Brownish-green diatomaceous mud	6.58	11.16	1.99	0.7	0.3
Core 2 15-18 cmbsf	GB4462-6	118	105	Brownish-green diatomaceous mud	4.21	10.72	5.56	0.69	0.38
Core 2 18-21 cmbsf	GB4462-6	109	125	Brownish-green diatomaceous mud	5.16	10.01	2.68	1.25	1
Just inside microbial mat perimeter									
Core 3 0-2 cmbsf	GB4462-3	3.2	3.2	Black mud with microbial mat filaments	2.71	38.73	1.8	1.18	0.47
Core 3 2-4 cmbsf	GB4462-3	-	8	Brownish-green diatomaceous mud	2.89	24.25	2.45	1.24	0.16
Core 3 4-6 cmbsf	GB4462-3	15	15	Brownish-green diatomaceous mud	2.04	19.55	2.59	0.72	0.41
Core 3 6-8 cmbsf	GB4462-3	-	26	Brownish-green diatomaceous mud	3.57	20.24	1.23	0.75	0.28
Core 3 8-10 cmbsf	GB4462-3	34	34	Brownish-green diatomaceous mud	5.08	12.71	2.38	0.82	0.26
Core 3 10-12 cmbsf	GB4462-3	-	43	Brownish-green diatomaceous mud	6.55	9.77	4.99	0.23	0.3
Core 3 12-15 cmbsf	GB4462-3	-	54	Brownish-green diatomaceous mud	6.6	6.73	1.81	0.58	0.22
Core 3 15-18 cmbsf	GB4462-3	61	66	Brownish-green diatomaceous mud	6.46	5.42	3.22	0.26	0.1
Core 3 18-21 cmbsf	GB4462-3	83	80	Brownish-green diatomaceous mud	3.73	6.94	1.21	0.28	0.23
Just outside microbial mat perimeter									
Core 4 0-2 cmbsf	GB4462-8	0	0	Black mud	1.89	37.04	1.61	0.47	0.37
Core 4 2-4 cmbsf	GB4462-8	1.5	8	Brownish-green diatomaceous mud	2.8	20.21	2.27	1.03	0.47
Core 4 4-6 cmbsf	GB4462-8	16	16	Brownish-green diatomaceous mud	3.1	23.05	1.87	0.88	0.28

Core 4 6-8 cmbsf	GB4462-8	-	18	Brownish-green diatomaceous mud	4.18	20.16	2.99	1.64	0.9
Core 4 8-10 cmbsf	GB4462-8	-	21	Brownish-green diatomaceous mud	4.99	13.9	4.44	3.27	1.67
Core 4 10-12 cmbsf	GB4462-8	-	23	Brownish-green diatomaceous mud	6.08	10.6	3.63	1.51	0.75
Core 4 12-15 cmbsf	GB4462-8	-	25	Brownish-green diatomaceous mud	4.42	1.78	0.6	0.24	0.11
Core 4 15-18 cmbsf	GB4462-8	-	27	-	-	-	-	-	-
Core 4 18-21 cmbsf	GB4462-8	29	29	-	-	-	-	-	-

[a] Collected core numbers are relabelled in the sample name to reflect a relative transect position (1-4).
[b] Sediment lithology based on freeze-dried sediments.

3.3.3 Sample Analysis

3.3.3.1 Comprehensive two-dimensional gas chromatography (GC×GC)

The ^{0}S-free apolar fractions were analyzed for their hydrocarbon contents using a Leco GC×GC flame ionization detector (GC×GC-FID) equipped with a dual stage cryogenic modulator (Leco, Saint Joseph, Michigan) at Woods Hole Oceanographic Institution. For GC×GC-FID analysis, 1st dimension separation was performed on a nonpolar Restek Rtx-1 Crossbond (20-m, 0.25-mm ID, 0.25-µm film thickness) that was held at 60 °C for 12 min then ramped to 315 °C at 1.5 °C/min, with 2nd dimension separation on a 50% phenyl polysilphenylene-siloxane column (SGE BPX50, 1-m, 0.10-mm i.d., 0.1-µm film thickness) that was held at 70 °C for 12 min and then ramped to 325 °C at 1.5 °C/min. The inlet temperature was 300 °C. The modulation column was pulsed by the hot jet modulator at 60 °C above the temperature of the primary GC oven for 0.4 s every 10.0 s. with a 5.6 s cooling period between stages. The GC was configured for splitless auto-injection, with an inlet temperature of 300 °C, and purge vent opening at 0.5min. Carrier gas (H_2) flow rate was 1mL/min (constant flow mode). The FID detector signal was sampled at a rate of 200 data points s^{-1}.

Mass spectral analysis was performed using an Agilent 7890B Gas Chromatograph, equipped with a ZX1 – LN_2 Cooled Loop Modulator (Zoex Corporation, Houston, Texas), interfaced with a Agilent 7200B quadrupole mass spectrometer (GC×GC-qToFMS) run in electron impact mode at Saint Mary's University. For GC×GC-qToFMS analysis, 1st dimension separation was performed on a nonpolar (5%-phenyl)-methylpolysiloxane phase (Agilent, HP-5ms Ultra Inert, 60-m; 0.25-mm ID; 0.25-µm film thickness), with 2nd dimension separation on a polar 50% phenyl equivalent polysiloxane (BPX-50, SGE, 1.31-

m; 0.10-mm ID; 0.1l-m film thickness) column. The modulation column was deactivated fused silica (1.37-m, 0.10-mm ID), which was pulsed by the hot jet modulator at 200 ºC for 0.4 s every 14.0 s. The GC was configured for 20:1 split auto-injection. The injection inlet temperature was 300 °C, and the purge vent was opened at 0.6 min. Oven temperature ramped from 60 °C to 340 °C at 1.15 °C/min with a carrier gas flow rate of 2.25 mL/min (constant flow mode). The transfer line temperature was 280 °C. The ToF-MS data collection was 50 scans/s at 4 GHz sampling rate.

3.3.3.2 GC×GC post data processing

The GC×GC-FID and GC×GC-qToFMS chromatograms were digitalized using GCImage Version 2.7 software (Zoex, Houston, USA). All samples were baseline subtracted with a rolling ball method and then phase shifted in the same manner to align and maximize a coherent spread of hydrocarbons on the produced chromatogram. Individual hydrocarbons and the range of various compound classes were then identified by GC×GC-ToFMS mass spectral analysis. The identified compounds were translated to GC×GC-FID chromatograms and template matched to other samples.

3.3.3.3 Chemometrics

Baseline-corrected GC×GC-FID data files were exported in csv format. The chromatographic space of each sample was then divided into two discrete data sets comprising the elution region of (1) aromatic compounds and (2) polycyclic triterpenoid

biomarkers (Fig. 2). Normal and branched, cycloalkanes, and acyclic isoprenoids were omitted from both data regions due to the high volatility and elevated signal intensities of these compounds, which were found to mask substrate-level pattern changes in the overall hydrocarbon matrix. The excision and division of corrected chromatographic regions improved the focus of the respective chemometric calculations for targeted compound classes and helped to further exclude column bleed and other chromatographic artifacts from interfering with the statistical analysis. Partitioned data sets were imported into MatLab 7.12 (R2011a) (MathWorks, Natick-MA, USA) to perform MPCA and HCA analysis using PLS Toolbox (Version 8.5.2 (2017); Eigenvector Research, Inc., Manson, WA USA 98831; software available at http://www.eigenvector.com). For these analyses, the data was mean-centered for each respective calculation. Various outliers were removed from the model (SI-Text 1.0, SI-Fig. 2). Data normalization was then attempted using total sum normalization (SI-Text 2.0). However, attempts to normalize the data were ultimately abandoned due to biases introduced to the ensuing chemometric model or because the normalization resulted in lower numbers of statistically significant correlations with proxy data used to validate the chemometric model (See SI Fig. 4, SI Table 1).

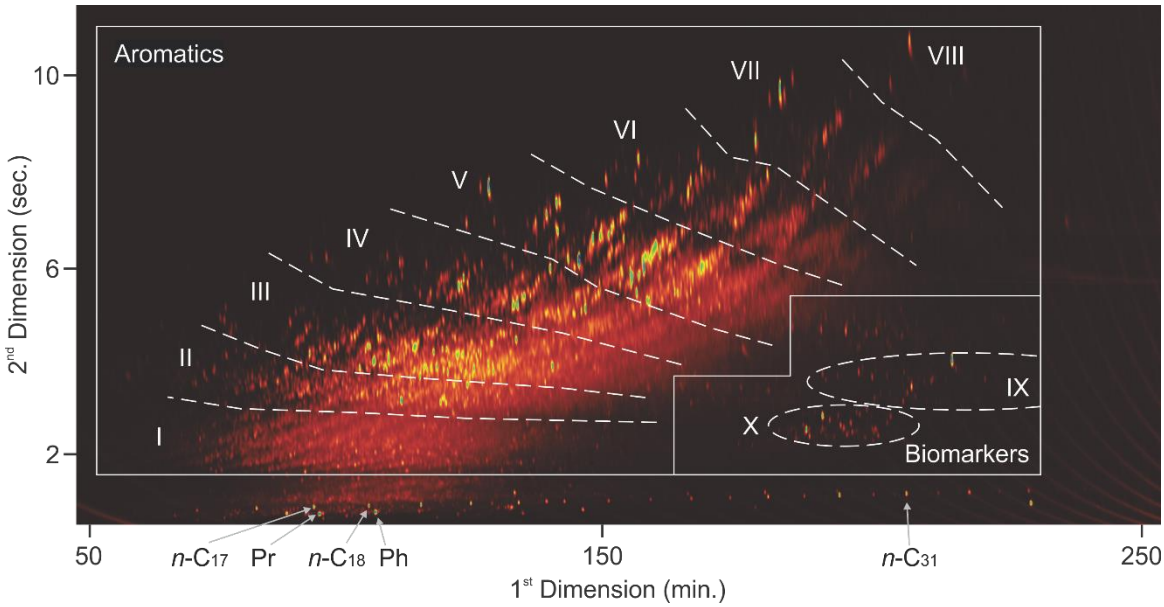


Fig 2. GC×GC-FID chromatogram partitioned into the two analytical regions (denoted aromatic and biomarkers). Hydrocarbon classes included in these regions: I: alkylbenzenes, triterpanes, and indenes II: naphthalenes, and benzothiophenes III: fluorenes, IV: phenanthrenes and dibenzothiophenes, V: fluoranthenes and pyrenes, VI: chrysenes, VII: benzofluoranthrenes, benzopyrenes, VIII: benzoperylenes, dibenzopyrenes, dibenzochrysenes, and indenopyrenes, IX: hopanoids, and X: steroids. Individual biomarkers pristane (Pr), phytane (Ph), n-C17, n-C18, and n-C31 are individually indicated for additional reference.

3.4 Results and Discussion

3.4.1 Geothermal gradients in the vent

Thermal probe porewater temperature measurements were interpolated (Table 1) to produce a heat map of the transect area (Fig. 3). Sediment porewater temperatures increase with depth and proximity to the vent center. Cores 1 and 2 record the hottest porewater temperatures (core 1; 153 °C within 21 cmbsf closest to the vent edifice), with more ambient conditions (core 4; 30 °C at 18-21 cmbsf) existing outside the microbial mat at the transect periphery (core 4).

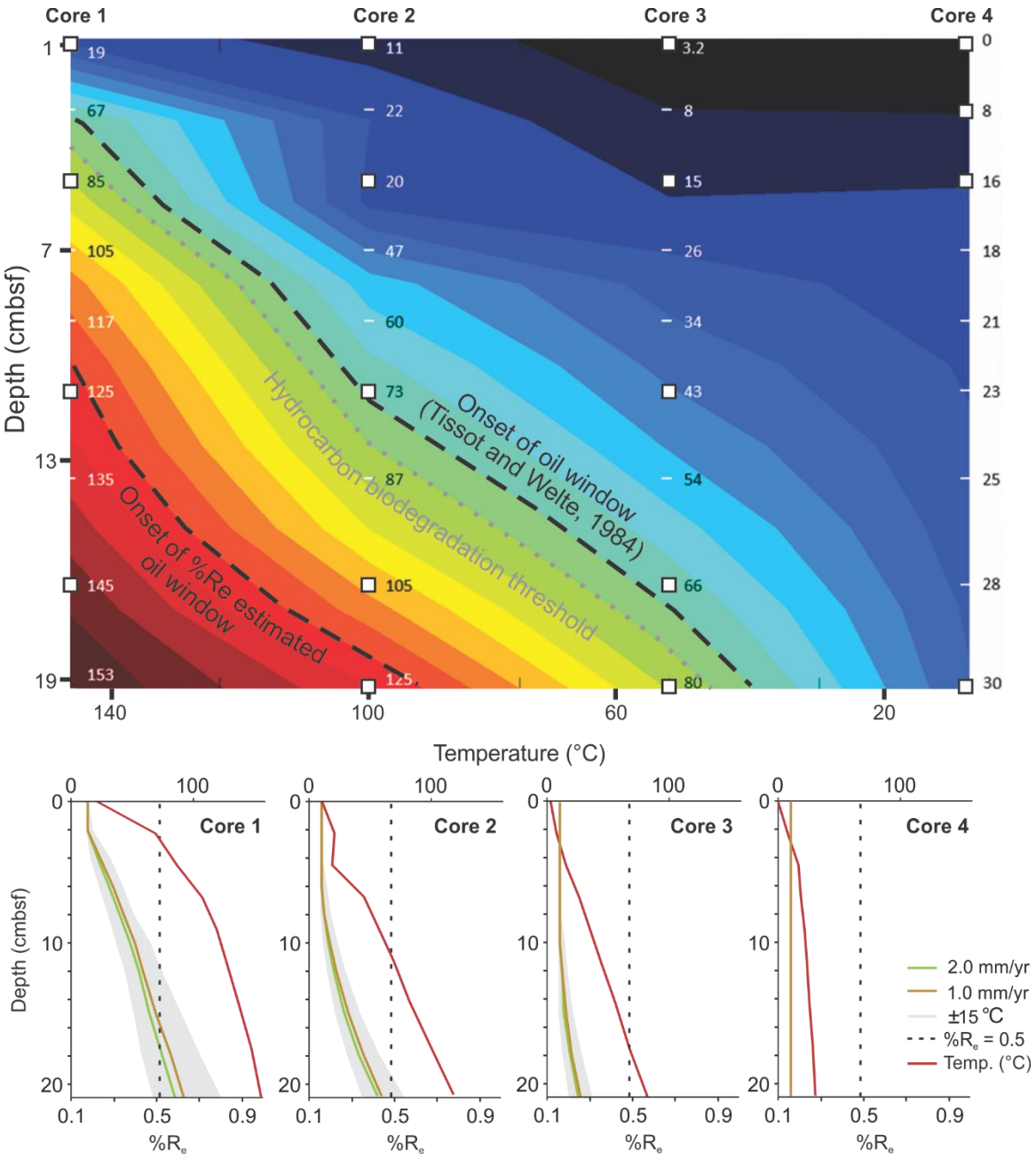


Fig. 3. Thermal heat map of the transect profile (top). White squares indicate recorded temperatures (Table 1). Vitrinite reflectance equivalent ($\%R_e$) estimates with depth projected for upper and lower sedimentation rates and maximum and minimum temperature-probe error (see Section 3.1).

3.4.2 Kinetic model of oil generation

For conventional basins, oil production begins at ~ 1 km burial depth commensurate with ~70-90 °C temperatures, taking anywhere from 5 million to >100 million years burial time (Tissot and Welte, 1984; Mesle et al., 2013). However, hydrothermal-petroleum generation involves relatively short exposure-times at extremely high temperatures in shallow sediment environments that are not reflective of the geological heating rates (~3 °C/Ma) and the high hydrostatic pressures of conventional petroleum systems (Simoneit and Lonsdale, 1982; Simoneit, 1985; Peter et al., 1991; Simoneit and Kvenvolden, 1993). As such, in the absence of a petrographic analyses and where heat flow and burial times are known, a kinetic model of maturation can be employed to more accurately express the location of the oil window at Cathedral Hill.

Using a modified version of the Easy%R_o kinetic model (Burnham 2019), vitrinite reflectance equivalent (%R_e) values were calculated for the range of porewater temperatures and exposure times (i.e. sedimentation rates). Kinetic maturation was calculated as either a linear and exponential model to match the respective ramp rates of a core's geothermal gradient. Given the steep geothermal gradients, probe measurements are considered to potentially have a ±15 °C/cm error in the hottest sections of the transect (cores 1 and 2) and ±5 °C/cm error within the outer more ambient sediments (cores 3 and 4) due to human error during the collection of the measurement. For conventional source rocks, a vitrinite reflectance of ~0.5 marks the onset of oil generation (Tissot and Espitalie, 1975). Based on these calculations, core 1 ≥15-18 cmbsf and core 2 >21 cmbsf reach oil generation (Fig. 3). The remaining cores have lower %R_e values reflecting thermally

immature sediments. By extrapolating the measured thermal gradients to greater depths, the onset of the oil window, peak oil, and wet gas formation is predicted to occur within 0.5 m below seafloor of cores 1-3 (Table 2).

Table 2. *Extrapolated depth of oil and gas windows at Cathedral Hill.*

Push core	Modeled temperature trend	Onset of oil window (~0.5%R_e)	Peak oil (~0.8-1.0 %R_e)	Onset of wet gas formation (~1.3 %R_e)
Core 1	exponential [a]	15-18 cmbsf	30-33 cmbsf	48-51 cmbsf
Core 2	linear [b]	22-23 cmbsf	27-28 cmbsf	31-32 cmbsf
Core 3	linear [b]	33-36 cmbsf	49-42 cmbsf	46-47 cmbsf

[a] trend is modeled after apparent exponential increase of porewater temperatures by depth for this core.
[b] trend is modeled after linear increase of porewater temperatures by depth for this core.

3.4.3 Sediment extract recoveries

Total lipid extract (TLE) yields are highest in the near-surface sediments closest to the vent center and decrease with depth (Table 2; Fig. 4). With the exception of core 1 which has maximized polar and apolar yields at ~12-18 cmbsf, all cores produce elevated/maximum apolar/polar yields across a range of 6-8 to 10-12 cmbsf.

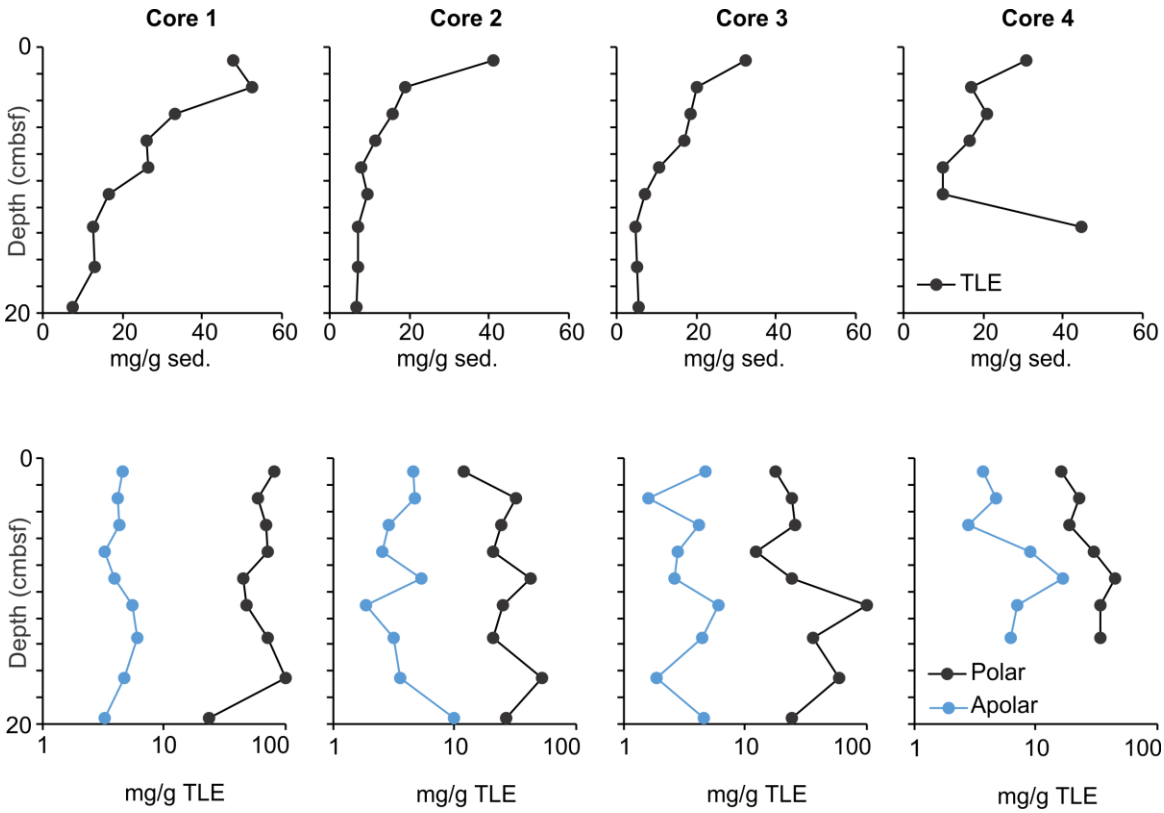


Fig. 4. Recovery of total lipid extract (top row) and the resulting polar and apolar subfraction yields (bottom row).

3.4.4 Multiway principal component analysis

3.4.4.1 Outlier identification

Various tests were performed to assess the coherence of a MPCA chemometric model and to ensure that systematic, whole compound class changes, rather than specific marker variations, were detected. To this end, several outliers were serially removed from the final model. Compositionally distinct outliers included core 4 sample chromatograms 8-10 and 10-12 cmbsf and core 3 sample chromatograms 6-8 cmbsf, which fell outside the 95% confidence limit in both the aromatic and biomarker scores plots (Fig. 5). For these

preliminary tests, the associated factor loadings plots indicated the source of the molecular disparities. Within the aromatic region, PC-1 factor loadings, accounting for ~40% of the matrix variability, were fundamentally derived from a single over-loaded peak (Fig. 5B) traced to core 3 sample 6-8 cmbsf. An additional ~30% of the matrix variability, captured in PC-2, was linked to a broad array of heavily loaded hydrocarbon signatures traced back to core 4 samples 8-10 and 10-12 cmbsf (Fig. 5C). For the biomarker region, the scores and factor loadings plots record similar anomalies across PCs -1 and -2 as observed in the aromatic region. Additional PCs for both regions reiterate similar disparity trends and did not conclusively identify new outliers (SI-Fig. 3). Additionally, in an iterative process, the chromatograms of core 2 samples 8-10 and 18-21 cmbsf; core 3 sample 0-2 cmbsf; and core 4 sample 6-8 cmbsf were likewise removed after showing to both cumulatively and independently obscure the base-level matrix relationships between the chromatographic stack (SI-Fig. 2).

Close inspection of the GC×GC-FID chromatograms for the sample outliers confirmed the presence of high intensity, irregularly shaped peaks that do not appear in the GC×GC-qToFMS chromatograms (i.e. core 3 sample 6-8 cmbsf). These peaks likely are analytical errors caused by sample overloading or anomalies acquired during an FID sample run. Additionally, the overprinting signatures may in some cases also represent environmental anomalies sourced from oil staining that substantially increased the diversity and intensity of the hydrocarbons measured in the respective chromatogram (i.e. core 2 samples 8-10 and 18-21 cmbsf; core 3 sample 0-2 cmbsf and core 4 samples 6-8, 8-10, and 10-12 cmbsf). In total, 7 of the 34 samples were removed from further MPCA analyses.

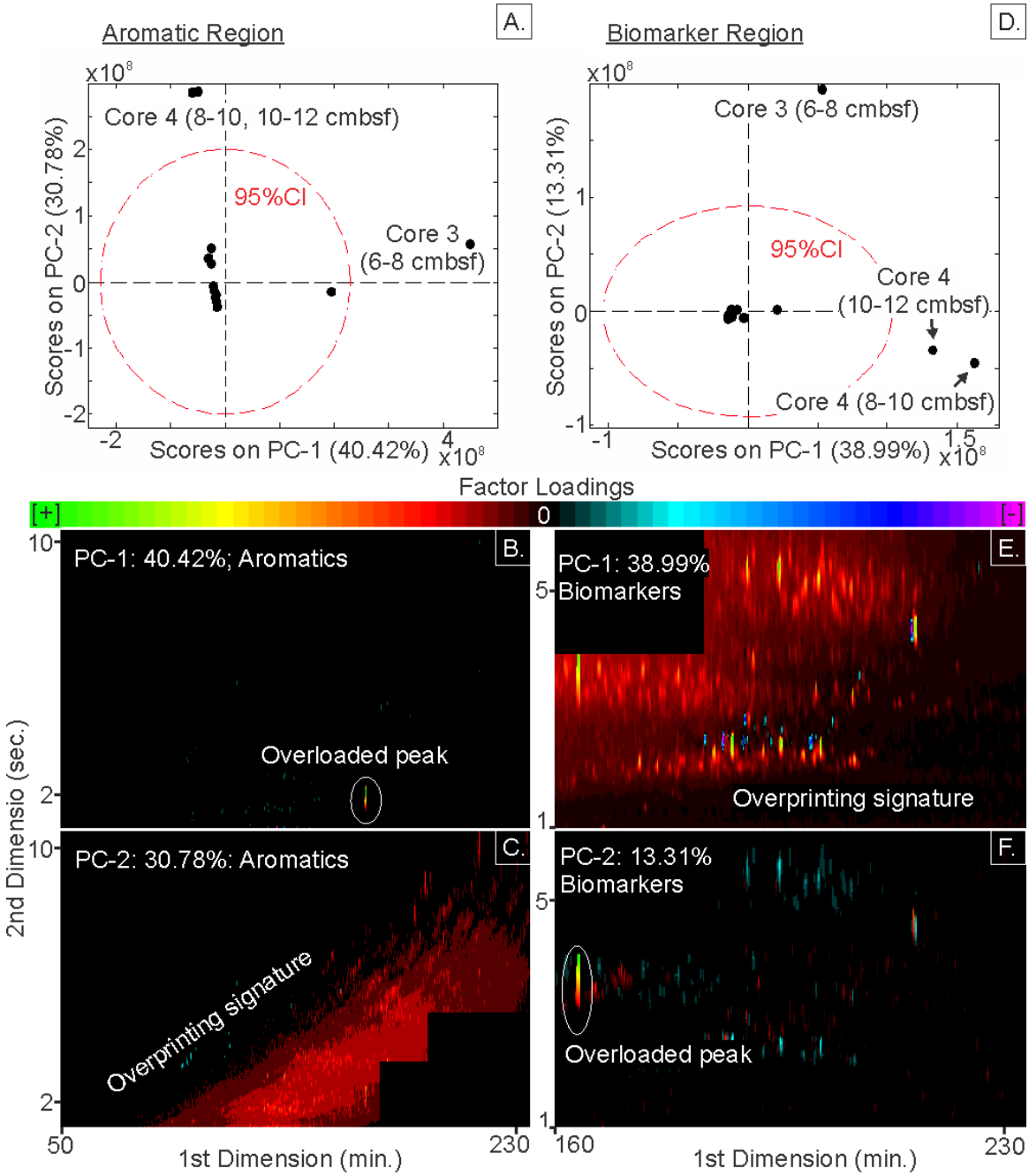


Fig. 5. MPCA scores and factor loadings plots of all Cathedral Hill sample chromatograms (n=34) showing (A, D) the outliers affecting sample clustering within the scores plots and

(B, C, E, F) the source of hydrocarbon variation between samples in their associated factor loadings plots.

3.4.4.2 Multi-molecular changes resolved by multiway principal component analysis

Following the removal of sample outliers, the multi-molecular variations within the aromatic and biomarker regions of the sample chromatograms were re-evaluated (Fig. 6). The first three PCs of the aromatic and biomarker regions account for respectively 83.03% and 76.47% of the variation calculated within the two data sets. Almost all 27 samples recorded negative PC-1 scores for the aromatic and biomarker regions (Fig. 6A and B). Samples with positive PC-1 scores were sourced from sediments having porewaters in excess of 70 °C and were limited to cores 1 and 2. Additional PCs did not provide meaningful information with respect to the relationships between sediment depths, exposure to in situ porewater temperatures, or allude to any systematic changes to hydrocarbon compound classes that may indicate production or migration (SI Fig. 3).

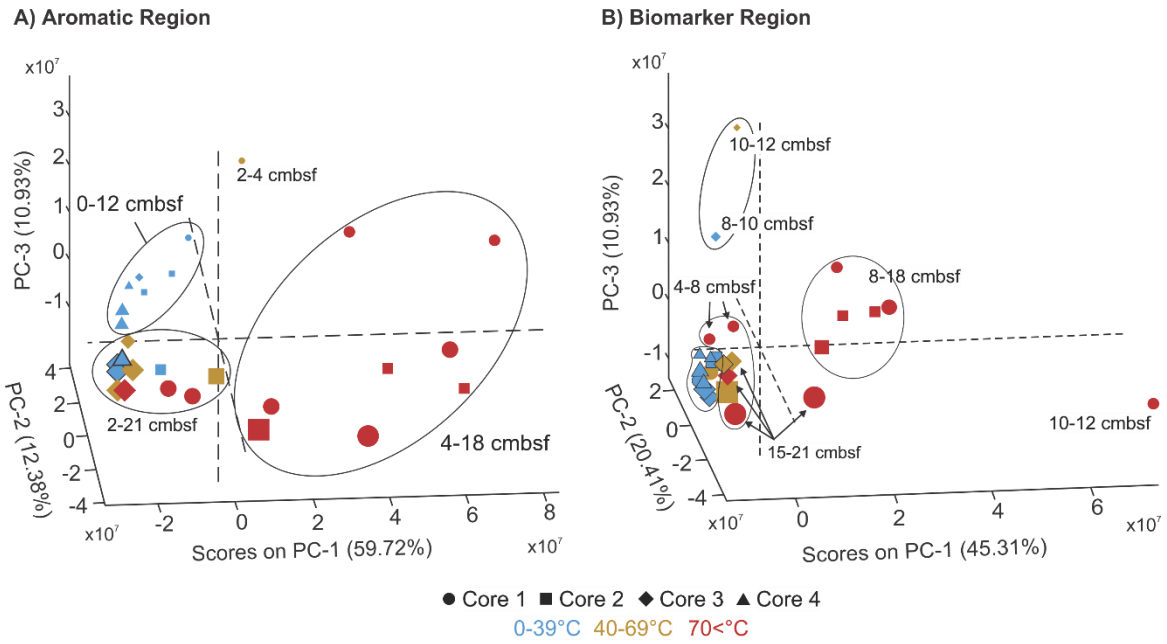


Fig 6. MPCA scores plots of the (A) aromatic and (B) biomarker regions (n=27). Ellipses mark sample clusters. Sample color indicates the porewater temperature range of sediment exposure.

3.4.4.3 Aromatic Region MPCA trends

3.4.4.3.1 Factor loadings

The factor loadings for PCs 1-3 display unique patterns for the aromatic region (Fig. 7A-C). Positive PC-1 loadings, accounting for 59.72% of the hydrocarbon matrix variability, include C_1-C_6 substituted alkylbenzenes, PAHs, benzothiophenes, and perhydroPAHs spanning compound classes I-VIII as well as a small region of negative loadings for C_7-C_8 substituted homologues of compound classes II-IV (Fig. 2 and 7A). Positive PC-2 loadings, marking 12.38% of the hydrocarbon matrix variability, comprise a sparse array of C_2-C_3 substituted homologues of compound classes III-V (along with >$C_{7\text{-}8}$

compounds of groups II-VIII that are extremely close to baseline), coupled with a larger domain of negative loadings comprising >C_4 substituted compounds of classes I-VIII (Fig. 7B). Positive PC-3 loadings account for 10.93% of the hydrocarbon matrix variability and include C_4-C_8 substituted homologues of compound classes I-VIII, with negative loadings on C_2-C_3 homologs spanning compound classes I-VI (Fig. 7C).

Additionally, the factor loadings plots of PCs 1-3 produce overlaps with positive and negative loadings coexisting on different PCs within the same compounds across the aromatic region. For example, $C_{4\text{-}6}$ substituted PAHs have both positive loadings on PC-1 and negative loadings on PC-2. For these conflicts, assignments of hydrocarbon change require the data to be ranked based on (1) the overall weight of the PC and (2) the magnitude difference between the opposing loadings on individual hydrocarbons. Therefore because $C_{4\text{-}6}$ substituted PAHs have an overall positive loading for 70.65% of the hydrocarbon matrix trends (based on PCs -1 and -3), the net expected expression of these compounds change is positive.

3.4.4.3.2 Transect trends of MPCA scores

To better track the relationships between macro-molecular variations in the Cathedral Hill maltenes, MPCA sample scores were plotted relative to their stratigraphic position within each push core (Fig. 7D-F). For core 1, surface-level sediments (<4 cmbsf) have negative to near zero scores on PCs 1-2 and positive scores on PC-3, indicating a slightly lower weighted relative abundance of PAHs and perhydroPAHs across classes' I-VIII and elevated C_{4+} substituted compounds. Between 6-12 cmbsf, scores progressively

reach their maximum positive (with respect to PCs 1 and 2) and negative (with respect to PC-3) score values, marking greater relative weighted abundance of C_1 to C_6 substituted compounds. At deeper sediment depths, PC scores gradually return to the surface sediment-level expression, in which more alkylated aromatic compounds abound.

For core 2, PCs 1-3 are more irregular with PC-1 reaching increasingly positive values with depth, PC-2 decreasing to negative values at >12 cmbsf, and PC-3 dipping to negative values in the mid-sections of the push core (Fig. 7D-F). With this, the overall weighted relative abundance of substituted aromatic hydrocarbons are shown to become more dominant with depth.

The more ambient temperature cores 3 and 4 have near constant negative/near-zero PCs -1 and -2 scores. These samples also have mostly negative PC-3 scores with the exception of positive swings at core 3 sample 8-10 cmbsf and core 4 sample 2-4 cmbsf (Fig. 7F). For cores 3 and 4, low abundance of C_1-C_4 compounds and highest relative abundance of > C_4 homologues predominate with depth.

Additionally, in both cores 1 and 2, the highest sample PC-1 score coincides with porewater temperatures that have reached ~100 °C. For cores 3 and 4, which have negative constant PC-1 scores, record porewater temperatures that do not rise above 80 °C.

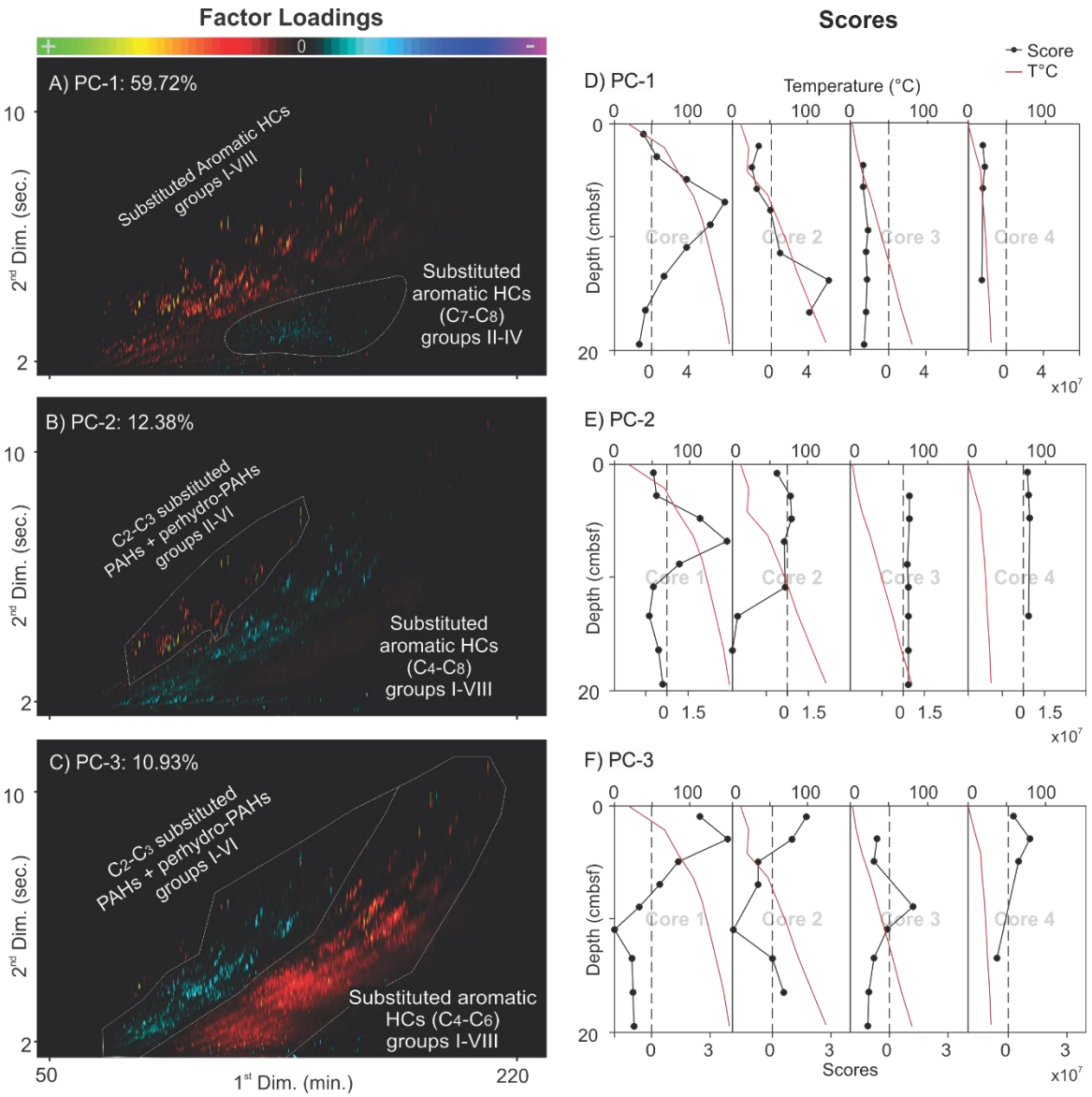


Fig. 7. Scores and factor loadings plots of the aromatics region (n=27). For the factor loadings plots (A-C), warm colours indicate positively weighted hydrocarbons (positive loadings) and cool colours mark the opposite (negative loadings). Sample scores replotted by push core depth alongside the thermal profile (red line: D-F).

3.4.4.4 Biomarker Region MPCA trends

3.4.4.4.1 Factor loadings

The MPCA PC-1 to -3 factor loadings plots display unique variations in biomarker abundances (Fig. 9). For PC-1, comprising of 45.31% of the matrix variability, positive loadings occur for C_{27}-C_{29} steroids, a group of unknown aromatics, and some faint hopanoids. This is coupled with anomalously high negative loading on diploptene (compound no. 14). Some of the positively loaded hopanoids include hop-17(21)-ene (6), 24-methylcholesta-3,5-diene (7), cholest-4-ene (4), and 24-methylcholest-4-ene (10), which are known to be early diagenetic products of biogenic compounds such as cholesterol and diploptene (Brault and Simoneit 1989). For PC-2 (Fig. 9D), representing 20.41% of the matrix variability, nearly all identified biomarkers have bifurcated peaks across the 1st dimension with the first half of an eluting peak marked by positive loadings followed by the later portion having negative loadings. Diploptene and the unknown aromatic hydrocarbons are also negatively loaded on this PC, but show no bifurcation. Inspection of the GC×GC FID chromatograms and ToF-MS data indicates slight 2nd dimension temporal offsets across the various modulations of peaks in the high temperature ramp of a few GC×GC sample runs. It appears these offsets are picked-up in PC-2, arguably making this factor loadings of little interpretative value. With respect to PC-3, corresponding to 10.93% of the matrix variation (Fig. 9F), positive loadings are produced for diploptene and various late 2nd dimension eluting unidentified steroid and hopanoid compounds. Furthermore, various C_{27} to C_{29} sterane isomers (likely corresponding to ααα- αββ- (20R) regular steranes) are also negatively loaded. To this end, all previously identified C_{27}-C_{29} steroids

(apart from cholest-4-ene) along with many of the hopanoids (no. 1, 4, 5, 6, 7, 12 from Table 3) have zero, or very near zero loadings and are absent from the PC-3 factor loadings plot.

3.4.4.4.2 Transect trends of MPCA scores

For core 1, the scores on PCs -1 and -3 record a unimodal stratigraphic profile with positive scores peaking at 10-12 cmbsf (PC-1) and 8-10 cmbsf (PC-3) coincident with ~125 °C and ~117 °C porewater temperatures, respectively (Fig. 8). The high scores are the result of particularly high concentrations of C_{27} to C_{29} sterenes and steradienes, coupled with lower relative abundance of diploptene. For PC-2, an extremely negative score on sample 10-12 cmbsf arises from highly negative loadings spanning a series of unknown aromatic compounds. The presence of these compounds appears to offset what may otherwise have been a unimodal stratigraphic scores trend similar to that of PCs -1 and -3. However, because a major portion of the variation captured by PC-2 is an analytical artifact and the associated stratigraphic profile of the scores is impacted by non-biomarker hydrocarbons, no systematic trends in biomarker distributions can reliably be made for the Cathedral Hill vent system.

For core 2, scores on PCs -1 and -2 are positively correlated to increasing porewater temperatures (Fig. 8). PC-1 scores are near-zero until 10-12cmbsf, indicating relatively lower C_{27} to C_{29} sterenes and steradienes alongside higher diploptene concentrations. Stratigraphically below this positive PC-1 scores dominate, indicating relatively low abundances of diploptene with high C_{27} to C_{29} sterenes and steradienes concentrations. PC-

3 scores are uniformly negative and of low values, indicating that the unidentified C_{27} to C_{29} sterane isomers are somewhat present and the remaining PC-3 hopanoid and steroid biomarkers are of lower abundance.

For cores 3 and 4, with the exception of core 3 PC-3 samples 8-10 cmbsf and 10-12 cmbsf, a persistent trend of negative or near-zero scores is recorded for the entire stratigraphic profile, which is in concert with recorded porewater temperatures that do not rise above 80 °C. Based on the factor loadings plots, these core samples have relatively high diploptene and low steroid/hopanoid biomarker concentrations. In core 3, anomalously high PC-3 scores at 8-10 cmbsf and 10-12 cmbsf record a jump in diploptene and unidentified steroid and hopanoid compound concentrations and lower abundances of unidentified C_{27} to C_{29} sterane isomers (Fig. 8).

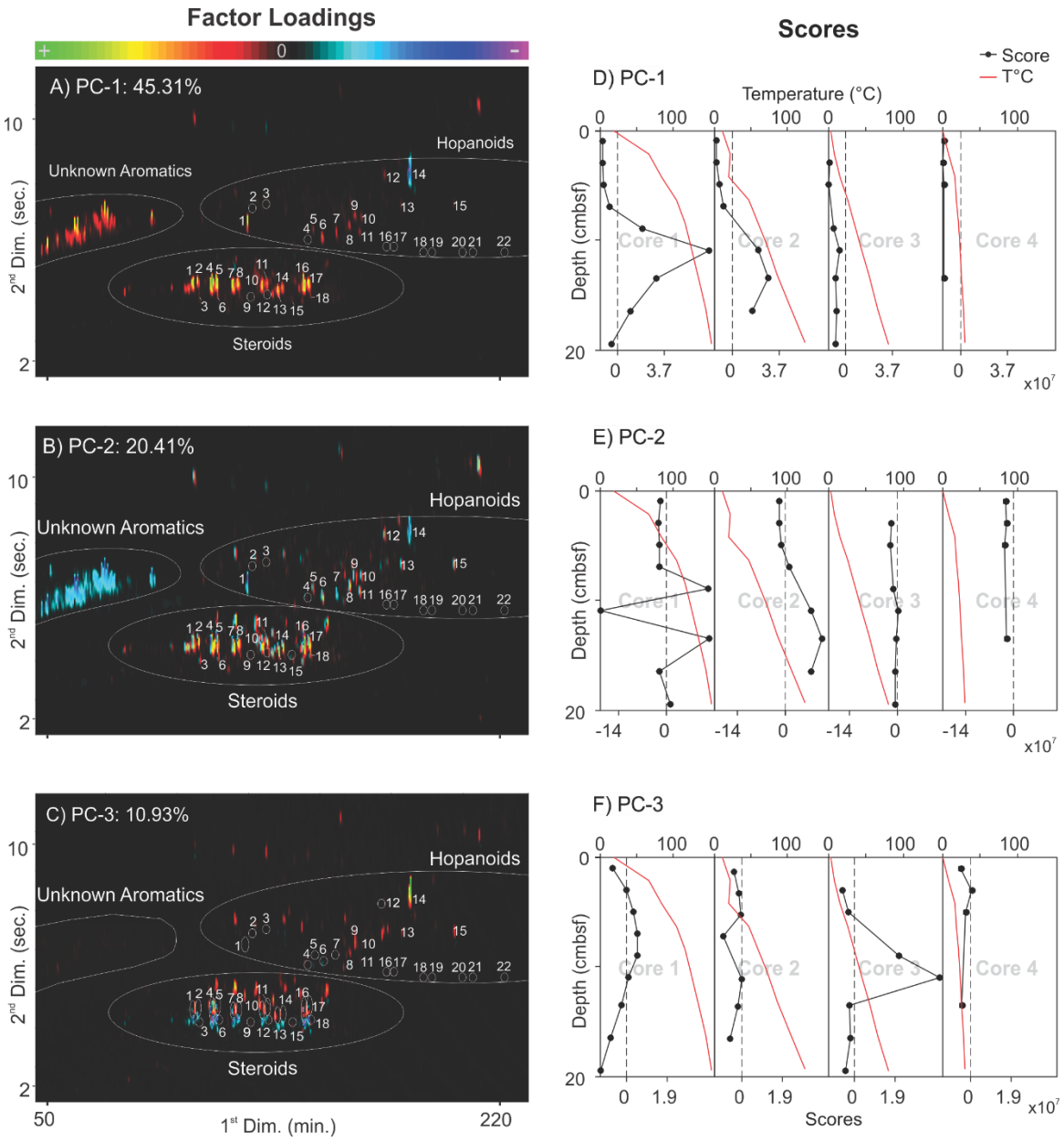


Fig. 8. Scores and factor loadings plots of the biomarker region (n=27). The original scores were replotted per core, by depth alongside the thermal profile (A, C, and E). For the factor loadings plots (B, D, and E), positive loadings (warm colours) indicate hydrocarbon enrichment. Negative loadings (cool colours) mark their relative depletion. Small white circles indicate no signal is generated in that specific elution-position.

Table 3. *Steroid and hopanoid compounds identified in the biomarker region (Figs. 3 and 10).*

Peak no.	Compound name (steroids)	Formula	M.W.	Diagnostic fragment ions *m/z*	Peak no.	Compound name (hopanoids)	Formula	M.W.	Diagnostic fragment ions *m/z*
1	cholest-4-ene	$C_{27}H_{46}$	370	215, 257	1	unknown compound	-	-	-
2	cholest-4,22-diene	$C_{27}H_{44}$	368	215, 257	2	17α(H)-22,29,30-trisnorhopane (TS)	$C_{27}H_{46}$	370	191, 149
3	5α,14β,17β-cholestane (20R)	$C_{27}H_{48}$	372	217, 259	3	18β(H)-22,29,30-trisnorneohopane (Tm)	$C_{27}H_{46}$	370	191, 149
4	cholest-4-ene	$C_{27}H_{46}$	370	215, 257	4	30-norhopane	$C_{29}H_{50}$	398	191, 205
5	cholest-5-ene	$C_{27}H_{46}$	370	215, 257	5	unknown hopa-diene	$C_{30}H_{48}$	408	187, 202
6	5α,14α,17α-cholestane (20R)	$C_{27}H_{48}$	372	217, 259	6	hop-17(21)-ene	$C_{30}H_{50}$	410	191, 231
7	24-methylcholesta-3,5-diene	$C_{28}H_{46}$	382	213, 255	7	17α,21β-hopane	$C_{30}H_{52}$	412	191
8	24-methylcholest-4,22-diene	$C_{28}H_{46}$	382	215, 257	8	unknown compound	-	-	-
9	5α,14β,17β-ergostane (20R)	$C_{28}H_{50}$	386	217, 259	9	unknown compound	-	-	-
10	24-methylcholest-4-ene	$C_{28}H_{48}$	384	215, 257	10	unknown compound	-	-	-
11	24-methylcholest-5-ene	$C_{28}H_{48}$	384	215, 257	11	17β,21α-moretane	$C_{30}H_{52}$	412	191
12	5α,14α,17α-ergostane (20R)	$C_{28}H_{50}$	386	217, 259	12	unknown compound	-	-	-
13	24-Ethylcholesta-3,5-diene	$C_{29}H_{48}$	396	213, 255	13	unknown compound	-	-	-
14	24-Ethylcholest-4,22-diene	$C_{29}H_{48}$	396	215, 257	14	diploptene	$C_{30}H_{50}$	410	191, 367
15	5α,14β,17β-stigmastane (20R)	$C_{29}H_{52}$	400	217, 259	15	unknown compound	-	-	-
16	24-ethylcholest-4-ene	$C_{29}H_{50}$	398	215, 257	16	17α,21β-homohopane (22S)	$C_{31}H_{54}$	426	191, 205
17	24-ethylcholest-5-ene	$C_{29}H_{50}$	398	215, 257	17	17α,21β-homohopane (22R)	$C_{31}H_{54}$	426	191, 205
18	5α,14α,17α-stigmastane (20R)	$C_{29}H_{52}$	400	217, 259	18	17α,21β-bishomohopane (22S)	$C_{32}H_{56}$	440	191
					19	17α,21β-bishomohopane (22R)	$C_{32}H_{56}$	440	191
					20	17α,21β-trishomohopane (22S)	$C_{33}H_{58}$	454	191
					21	17α,21β-trishomohopane (22R)	$C_{33}H_{58}$	454	191
					22	17α,21β-tetrakishomohopane (22S)	$C_{34}H_{60}$	468	191

3.4.5 Hierarchical Cluster Analysis

Hierarchical cluster analysis on the aromatic and biomarker chromatographic areas were used to elucidate the overall relatedness of sample matrices and to help identify areas of multi-molecular similarity and change. In this regard, HCA is used along side of MPCA to further validate detected patterns (i.e. Fig. 6). With respect to the aromatic region, the HCA dendrogram (Fig. 9A) produces two principle branch points with a mean distance between clusters at ~13×10^7. With a few exceptions, the majority of low temperature samples from cores 1-4 make-up the second arm of the dendrogram. The other arm forms a stair-stepped trend of high temperature samples from cores 1-3 being less related to the larger whole, which then progresses to greater matrix similarity at lower temperature with samples sourced from cores 2-4.

For the biomarker region, the HCA dendrogram forms third and fourth order clusters with a mean distance between clusters of ~2×10^7 (Fig. 9B). These clusters contain samples sourced from different cores of various porewater temperatures. Although less diagnostic when compared to the aromatic regions, a determinable pattern is evident with clusters loosely grouping by exposure to porewater temperatures. Collectively, the HCA dendrograms indicate complex multi-molecular changes are occurring in the transect profile across discrete spatial extents that may in part be the function of exposure to the hydrothermal vent fluids. As such, the HCA results compliment those made by MPCA.

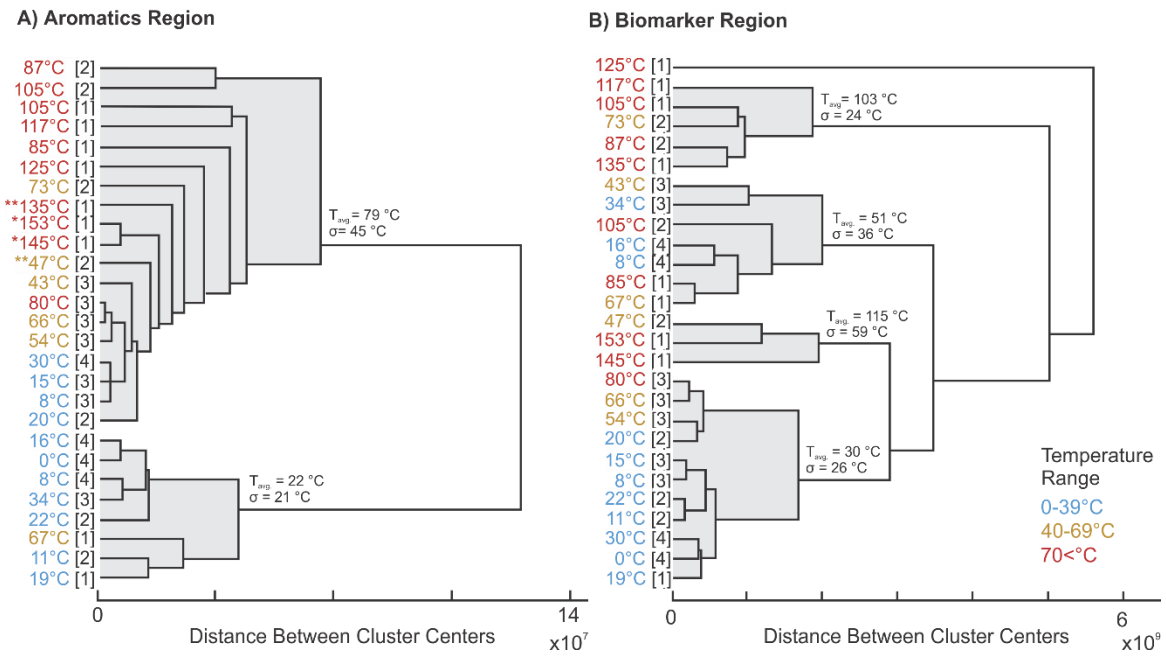


Fig. 9. HCA dendrograms of A) aromatics and B) biomarker regions (n=27). Sample names are replaced by their corresponding interpolated *in situ* porewater temperature value, followed by a bracketed number (ex: [1]) representing the specific core from which the sample was collected.

3.4.6 Validation of Chemometric Analyses

Validation of the MPCA and HCA model outcomes are provided by non-chemometric approaches that include comparisons of sample matrix variations determined by subtracted GC×GC chromatograms, the direct measurements of molecular complexity within the transect area, and statistical correlations between these results and physicochemical features of the transect area.

3.4.6.1 Subtracted GC×GC chromatograms

Multi-molecular changes within GC×GC data can be holistically tracked using difference chromatograms (e.g. Nelson et al., 2006, 2016; Reddy et al., 2007; Tran and Marriott, 2007; Wardlaw et al., 2008; Ventura et al, 2012). For the Cathedral Hill transect, a mean chromatographic hydrocarbon fingerprint ($\bar{x}$, n = 27) of the background oil profile used in the chemometric model was created. From this, individual core samples were systematically subtracted from the profile to assess relative hydrocarbon enrichments or depletions within the sample matrix.

For core 1 (Fig. 10A-I), difference chromatograms indicate the surface sediments are initially comprised of elevated contributions of $\geq C_4$ substituted PAHs, perhydroPAHs, and benzothiophenes spanning compound classes I-VIII, which transition into a much larger array of alkylated products by 6-8 cmbsf (Fig. 10A-D). Below this level, these compounds become progressively less diverse with short-chain homologs increasingly concentrated in the sample matrix. For this lower section, hopanoids and C_{27}-C_{29} steroids hydrocarbon biomarkers also become increasingly concentrated and diverse – the exception being diploptene, which marks an independent pattern of depletion with depth.

For core 2 (Fig. 11A-I), the concentration of $\sim C_1$-C_4 aromatic compounds is low in the near surface sediments and gradually increase with depth (excepting 8-10 and 18-21 cmbsf, which were excluded from the chemometric model). Concomitant with this trend is a general decline in the relative concentrations of $\geq C_4$ alkylated homologs. Hydrocarbon biomarkers begin to show elevated abundances at below 6-8 cmbsf. The exception being

diploptene, which, similar to core 1, is elevated at surface and systematically decreases towards the bottom of the push core.

For core 3 (Fig. 12A-I), with the exception of sample 8-10 cmbsf (and outliers 0-2 and 6-8 cmbsf), aromatic hydrocarbons and biomarkers are consistently below the mean hydrocarbon fingerprint. Diploptene concentrations remain elevated until ~12-15 cmbsf and below. Core 4 (Fig. 13A-D) has a relatively constant pattern of elevated $\geq C_4$ alkylated homologs across aromatic compound classes I-V (extending to VIII in sample 2-4 cmbsf) with overall low biomarker abundances, but elevated diploptene concentrations (this differs from excluded samples 6-8 to 10-12 cmbsf, which have above average concentration for all but the $< C_3$ substituted compounds).

In summary, the mass balance of hydrocarbons highlighted by subtracted chromatograms aligns near-perfectly with the MPCA chemometric model, and explains most of the scores trends with depth for both the aromatic and biomarker regions. Notably in core 1, an overall drain in hydrocarbon content appears responsible for the univariant trend seen in the PC-1 trends for the biomarker and aromatic regions, rather than additional input of diploptene or $\geq C_4$ substituted aromatics below ~10 cmbsf, which may allude loss by migration or water-washing. Additional structure the oil impregnated sediment profiles are also revealed and further discussed below.

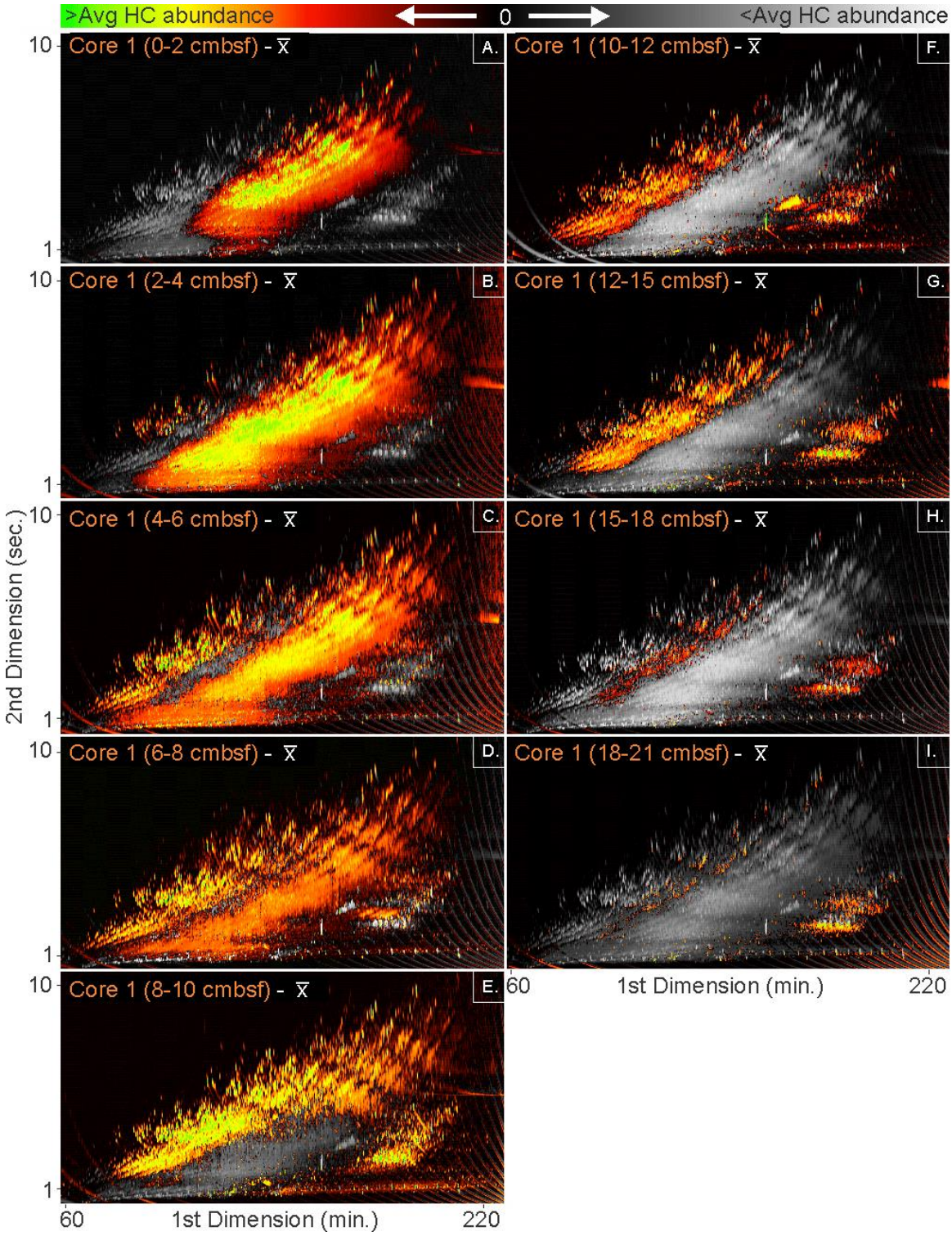


Fig. 10. Core 1 subtracted GC×GC FID chromatograms assessing the molecular differences between samples and the average hydrocarbon profile (x̄, n=27) from the chemometric

dataset. Warm colours indicate the hydrocarbon is at above-average concentration of a compound compared to its equivalent value in the transect mean (grey colors).

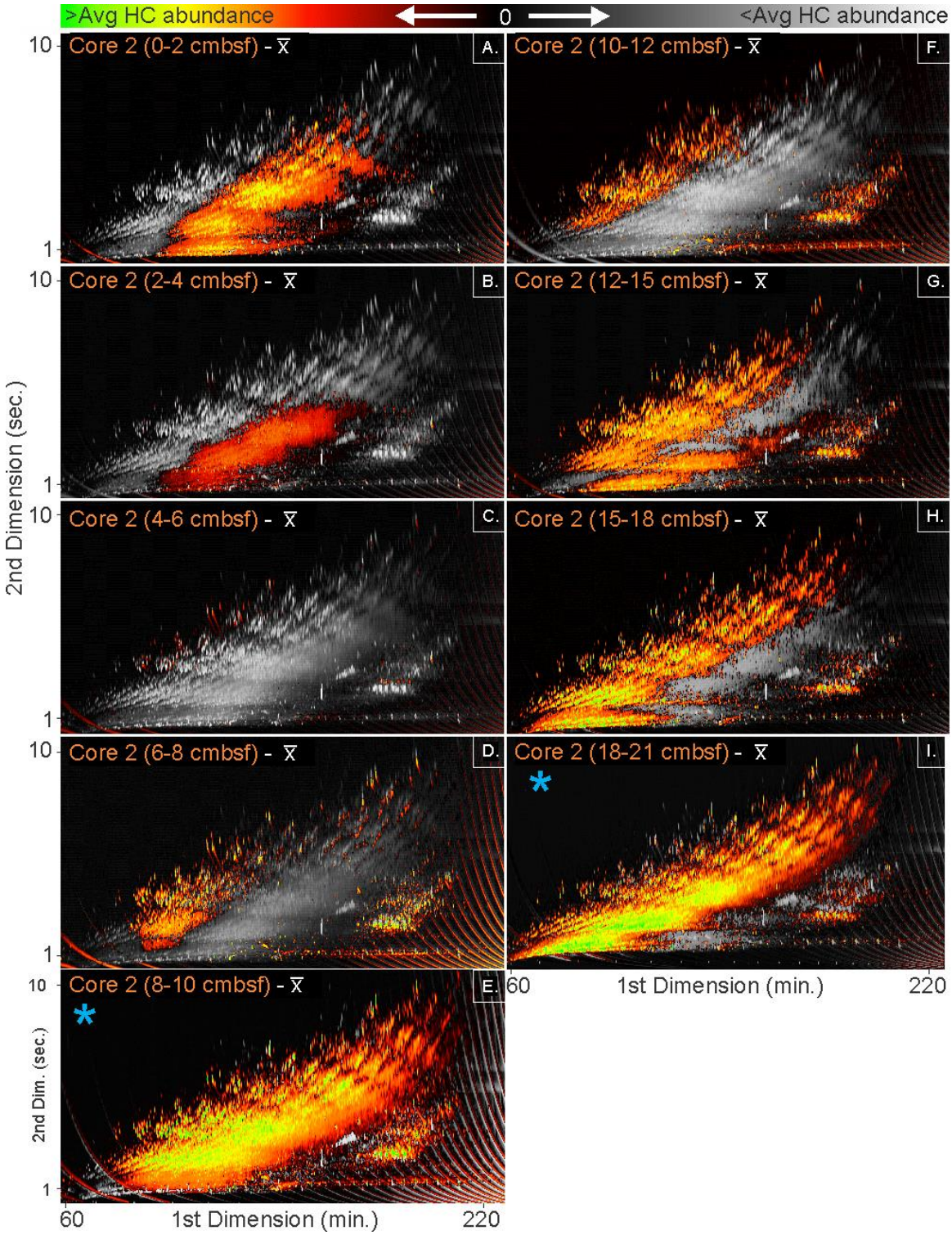


Fig. 11. Core 2 subtracted GC×GC FID chromatograms assessing the molecular differences between samples and the average hydrocarbon profile (x̄, n=27) from the chemometric

dataset. Warm colours indicate the hydrocarbon is at above-average concentration of a compound compared to its equivalent value in the transect mean (grey colors). Blue asterisk indicates subtracted sample was removal from the chemometric model.

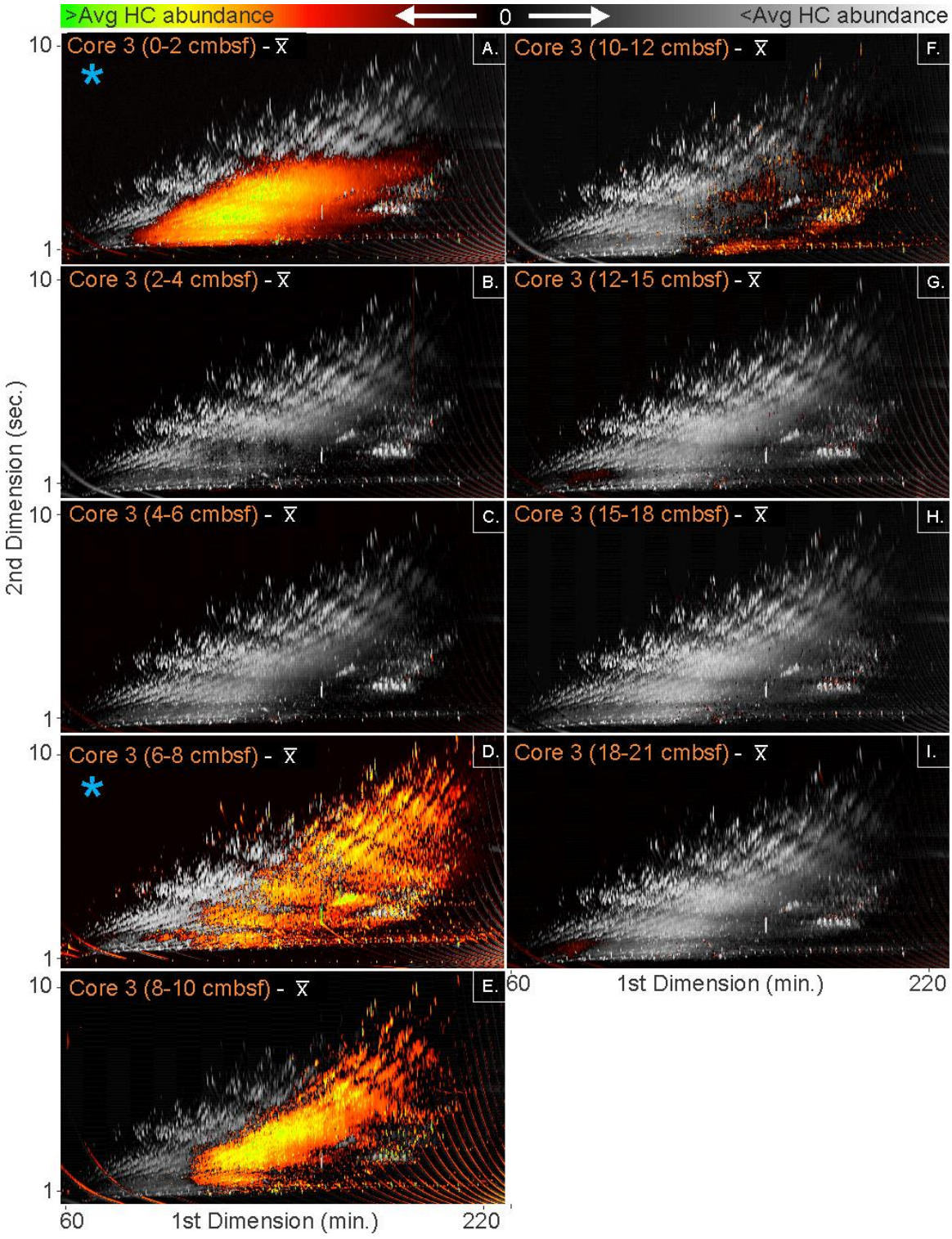


Fig. 12. Core 3 subtracted GC×GC FID chromatograms assessing the molecular differences between samples and the average hydrocarbon profile ($\bar{x}$, n=27) from the chemometric

dataset. Warm colours indicate the hydrocarbon is at above-average concentration of a compound compared to its equivalent value in the transect mean (grey colors). Blue asterisk indicates removal from the chemometric model. Blue asterisk indicates subtracted sample was removal from the chemometric model.

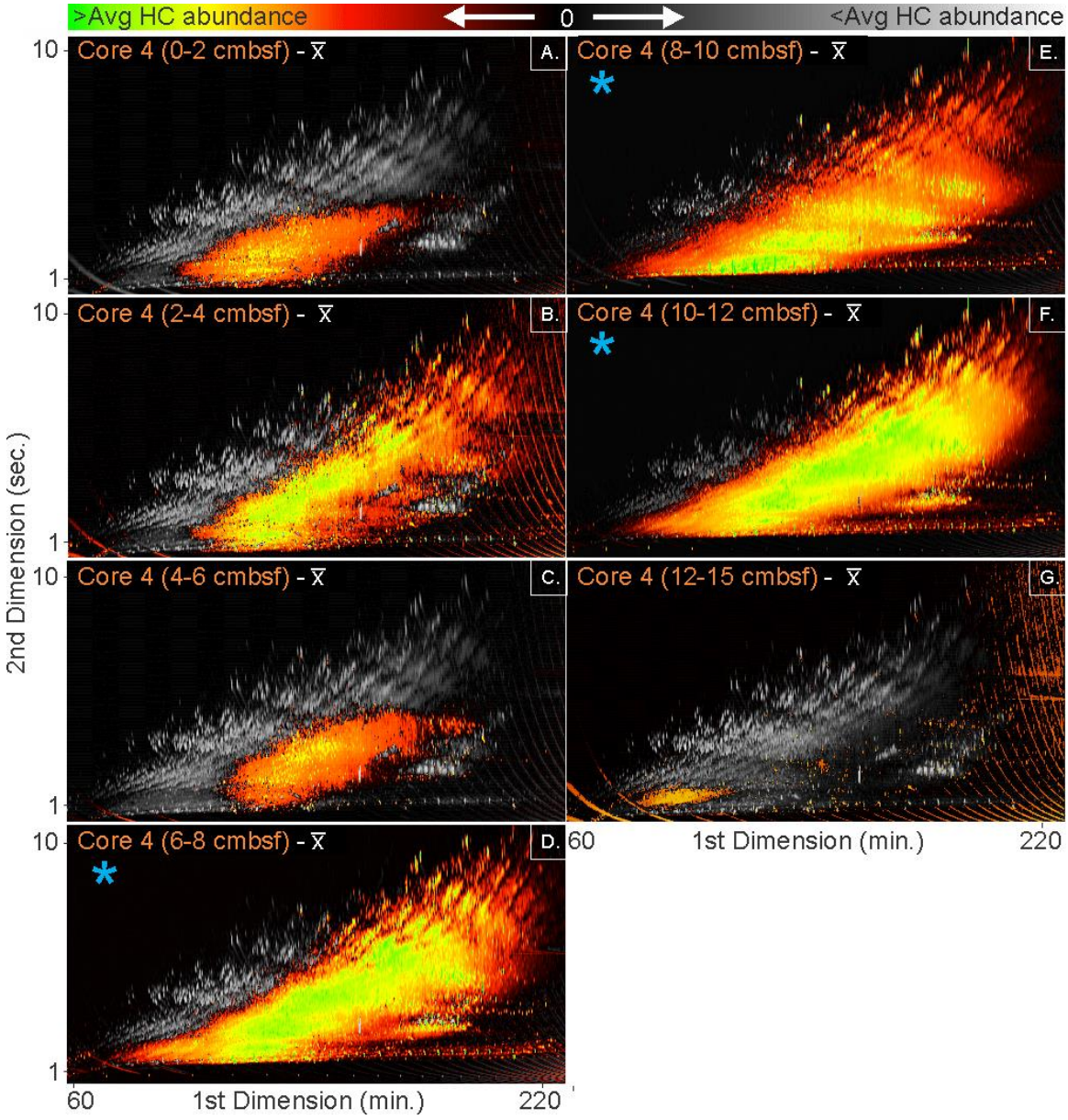


Fig. 13. Core 4 subtracted GC×GC FID chromatograms assessing the molecular differences between samples and the average hydrocarbon profile (x̄, n=27) from the chemometric dataset. Warm colours indicate the hydrocarbon is at above-average concentration of a compound compared to its equivalent value in the transect mean (grey colors). Blue asterisk indicates subtracted sample was removal from the chemometric model.

3.4.6.2 Molecular complexity of transect

The molecular complexity of an oil is a function of source, maturation, migration and physical, chemical, and biological weathering processes. For hydrothermal environments, where the generation of petroleum occurs within unconsolidated sediments of shallow burial, the migration, biodegradation, water washing, or pyrolytic activity strongly affects hydrocarbon compositions (e.g. Simoneit and Lonsdale, 1982; Simoneit, 1984, 1985, 1993; Kawka and Simoneit, 1987, 1990, 1994; Kawka, 1990; McCollom et al., 1999). To further evaluate molecular compositional changes across the transect, the sum of detected peaks in each sediment extract was interpolated across the transect area (Fig. 14). As this method is insensitive to hydrocarbon overloading, sample outliers originally removed from the chemometric models were included (n=34). The resulting heat map displays a complex spatial pattern to the hydrocarbon distribution in the shallow subsurface. Regions of high hydrocarbon diversity occur at the bottom of core 2 and as a narrow (~4 cm thick) band extending ~10 cmbsf across the transect (Fig. 14). The hydrocarbon fingerprint of this band is resolved in subtracted chromatograms (Fig. 10-13) as being composed of a broad array of alkylated aromatic hydrocarbons spanning from alkylbenzenes and indenes (compound class II) to compound class VIII (Fig. 2). From these, it was determined that the fluid compositions from each end of the flank are similar (Fig. 10-13) suggesting the layer may be formed by differential precipitation and fractionation, outward dispersion of oil sourced at hotter portions of the vent, or the merger of more than one plume advecting upwards from deeper within the basin. These hydrocarbon hotspots may also be evidence of in situ generation (see section 4.7 below).

Either way, the presence of the band appears to be an overprint to the localized hydrocarbon generation profile within these shallow sediments.

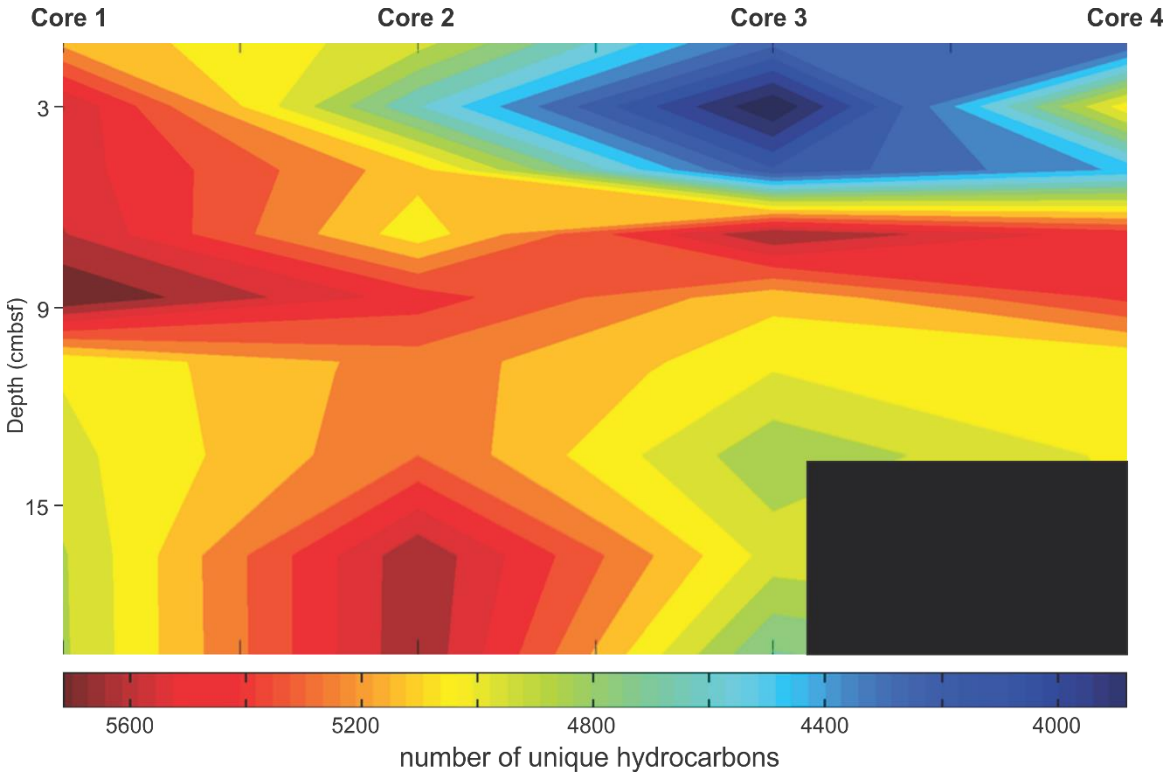


Fig. 14. Molecular complexity heat map of the Cathedral Hill transect. Black region marks area of no data (n=34).

3.4.6.3 Spearman's rank correlation coefficient and validation of the chemometric model

Spearman's rank correlation coefficient (SRCC) is a robust, nonparametric statistic that measures the strength of monotonic relationships between two sets of variables (Daniel, 1978; Saltelli, 1990 Gautheir, 2001). To this end, the relationships between

geothermal gradients as well as the coupled impacts resulting from upper and lower sedimentation rate dependent exposure times (see section 4.2) was compared to the multi-molecular changes recorded by MPCA scores (see section 4.4). The SRCCs were also calculated to determine the degree of correlation between MPCA scores and the direct tabulations of molecular hydrocarbon complexity (see section 4.6.2), and various hydrocarbon maturation parameters (Tables 4 and 5; SI-Fig. 5). Two of these parameters specifically target the dealkylation reactions effecting oils as measured by the methylphenanthrene index (MPI-1 and MPI-2; Table 4 and %; SI-Fig. 5; Radke et al., 1982a,b; Garrigues et al., 1986; Radke et al., 1986, 1991). With regards to MPI, the vent fluid chemistry of the hosted SOM can affect the alkylation of the hydrothermally derived petroleum products (i.e. Kawka and Simoneit, 1994; Ventura et al., 2012; 2020). In this regard, water/rock interactions can produce a wide variety of radicals, which act as potential reducing agents that can influence catagenesis (Siskin and Katritzky, 1991, 1999, 2001; McCollom and Seewald, 2001; Seewald, 2001). For both MPI parameters, increasing values are linked to increasing porewater temperatures (SI-Fig. 5).

For the aromatic region, all but the MPI (both -1 and -2) and moretane/hopane ratios produced statistically significant correlations at the 95% confidence interval on PC-1. For the biomarker region, all of the parameters produce statistically significant correlations at the 95% confidence interval on PC-1. Collectively, this suggests the MPCA scores are tracking first-order substrate composition and maturation trends (Table 4 and 5). However, for PC-2, only the hopane thermal maturity parameters for the aromatic region and the moretane/hopane and C_{27}-C_{29} sterene/sterane ratio for the biomarker region resulted in a

statistically significant positive correlation. For the aromatic region, PC-3 statistically significant negative correlations were obtained on porewater temperature, most of the kinetic modeled outputs, and three aromatic maturity parameters (MPR-1, 2 and MPI-2). The negative correlation is consistent with the maturation trends predicted by both proxies. In the biomarker region, a single positive correlation was obtained for the scores comparison with molecular complexity (Table 5). Collectively, the difference chromatograms, molecular complexity, and SRCCs align with trends described by the scores and factor loadings plots. In this regard, the chemometric approaches, resolve a complex tapestry of hydrocarbon generation, migration, and thermochemical oxidation for the shallow sediments of the Cathedral Hill vent site.

Table 4. *Cathedral Hill transect maturation parameters*[†]

Sample location and name	TMN [a]	MPR-1 [b]	MPR-2 [c]	MPI-1 [d]	MPI-2 [e]	Ts/Tm [f]	M/H [g]	Steroid ratio [h]
Microbial mat closest to vent								
Core 1 0-2 cmbsf	0.38	0.50	0.34	0.63	0.68	0.30	4.24	0.63
Core 1 2-4 cmbsf	0.40	0.88	0.47	0.31	0.36	0.19	5.79	0.71
Core 1 4-6 cmbsf	0.49	1.04	0.51	0.53	0.62	0.14	6.20	0.79
Core 1 6-8 cmbsf	0.54	1.20	0.54	0.75	0.86	0.15	2.54	0.82
Core 1 8-10 cmbsf	0.57	1.51	0.60	1.01	1.13	0.34	2.79	0.88
Core 1 10-12 cmbsf	0.59	1.69	0.63	1.05	1.16	0.44	0.60	0.92
Core 1 12-15 cmbsf	0.57	1.46	0.59	1.08	1.20	0.44	0.99	0.91
Core 1 15-18 cmbsf	0.59	1.63	0.62	1.08	1.19	0.56	0.69	0.87
Core 1 18-21 cmbsf	0.57	1.56	0.61	1.05	1.15	0.65	0.67	0.83
Microbial mat further from vent								
Core 2 0-2 cmbsf	0.57	0.75	0.43	0.76	0.84	0.31	4.67	0.67
Core 2 2-4 cmbsf	0.48	1.26	0.56	0.24	0.29	0.27	4.46	0.76
Core 2 4-6 cmbsf	0.49	1.28	0.56	0.57	0.66	0.29	2.01	0.78
Core 2 6-8 cmbsf	0.53	1.45	0.59	0.69	0.79	0.36	1.43	0.87
Core 2 8-10 cmbsf	0.54	1.33	0.57	0.80	0.91	0.37	2.46	0.91
Core 2 10-12 cmbsf	0.53	1.50	0.60	0.83	0.93	0.34	2.59	0.91
Core 2 12-15 cmbsf	0.56	1.36	0.58	1.03	1.14	0.38	1.67	0.89
Core 2 15-18 cmbsf	0.60	1.64	0.62	1.00	1.11	0.52	1.73	0.82
Core 2 18-21 cmbsf	0.62	1.65	0.62	1.05	1.17	0.53	1.11	0.69
Just inside microbial mat perimeter								
Core 3 0-2 cmbsf	0.66	1.88	0.65	1.21	1.30	0.19	6.58	0.68
Core 3 2-4 cmbsf	0.61	1.28	0.56	0.90	1.01	0.09	5.23	0.63
Core 3 4-6 cmbsf	0.49	1.10	0.52	0.87	0.94	0.08	3.50	0.73
Core 3 6-8 cmbsf	0.39	0.79	0.44	0.18	0.21	0.17	0.40	0.67

Core 3 8-10 cmbsf	0.34	0.72	0.42	0.46	0.53	0.11	3.05	0.79
Core 3 10-12 cmbsf	0.35	0.80	0.44	0.39	0.45	0.10	2.04	0.80
Core 3 12-15 cmbsf	0.45	0.89	0.47	0.35	0.41	0.06	2.14	0.88
Core 3 15-18 cmbsf	0.47	0.89	0.47	0.48	0.56	0.11	2.45	0.91
Core 3 18-21 cmbsf	0.40	0.46	0.31	0.40	0.45	0.11	2.82	0.90
Just outside microbial mat perimeter								
Core 4 0-2 cmbsf	0.52	2.31	0.70	1.00	1.12	0.20	7.04	0.60
Core 4 2-4 cmbsf	0.46	0.88	0.47	0.73	0.78	0.19	3.58	0.55
Core 4 4-6 cmbsf	0.40	1.23	0.55	0.93	0.99	0.23	4.34	0.50
Core 4 6-8 cmbsf	0.47	0.91	0.48	0.78	0.87	0.36	3.89	0.31
Core 4 8-10 cmbsf	0.53	0.94	0.48	0.78	0.86	0.44	1.58	0.36
Core 4 10-12 cmbsf	0.46	1.05	0.51	0.80	0.88	0.33	2.71	0.52
Core 4 12-15 cmbsf	0.47	0.94	0.48	0.66	0.72	0.13	1.57	0.63

[†] See SI-Fig. 5 for depth profile plots.

All compounds quantified in their respective GC×GC–FID sample chromatograms.

[a] Trimethylnaphthalene ratio (TMN) = 1,3,7-trimethylnaphthalene /(1,3,7-trimethylnaphthalene + 1,2,5-trimethylnaphthalene), identified via monitoring of the *m/z* 170 mass chromatograms.

[b] Methylphenanthrene ratio 1 (MPR-1) = 2-methylphenanthrene / 1-methylphenanthrene identified via monitoring of the *m/z* 191 mass chromatograms and quantified in their respective GC×GC–FID sample chromatograms.

[c] Methylphenanthrene ratio 2 (MPR-2) = 2-methylphenanthrene / (1-methylphenanthrene + 2-methylphenanthrene), identified via monitoring of the *m/z* 192 mass chromatograms.

[d] Methylphenanthrene Index 1 (MPI-1) = (3-methylphenanthrene + 2-methylphenanthrene)/(9-methylphenanthrene + 1-methylphenanthrene) (Radke et al., 1982a). Monomethylphenanthrenes, identified via monitoring of the *m/z* 192 mass chromatograms.

[e] Methylphenanthrene Index 1 (MPI-1) = 1.5 × (3-methylphenanthrene + 2-methylphenanthrene)/(phenanthrene + 9-methylphenanthrene + 1-methylphenanthrene) (Radke et al., 1982a). Monomethylphenanthrenes, identified via monitoring of the *m/z* 192 mass chromatograms.

[f] Ts/Tm = C_{27} 17α(H)-22,29,30-trisnorhopane /(C_{27} 17α(H)-22,29,30-trisnorhopane + 18α(H)-22,29,30-trisnorneohopane) (Table 3 compounds 1 and 2) identified via monitoring of the *m/z* 191 mass chromatograms.

[g] M/H = βα-hopane/ αβ-hopane (Table 3 compounds 11 and 7), identified via monitoring of the *m/z* 191 mass chromatograms.

[h] Steroid ratio = ΣC_{27}-C_{29} sterenes / (ΣC_{27}-C_{29} steranes + ΣC_{27}-C_{29} sterenes) marking compounds 3-6, 9-12, and 15-18 from Table 3. Compounds were identified via monitoring of the *m/z* 217 and 215 mass chromatograms.

Table 5. Spearman's rank correlation coefficient between PC scores and temperature, vitrinite reflectance equivalent, molecular complexity, and various maturation parameters (n =27).

Spearman's rank correlation coefficient	Aromatic Region			Biomarker Region		
	PC-1	PC-2	PC-3	PC-1	PC-2	PC-3
Temperature [a]	**0.68**	-0.29	**-0.44**	**0.83**	0.35	0.08
$\%R_e$ -15 °C [b]	**0.62**	-0.20	-0.35	**0.72**	0.20	0.00
$\%R_e$ 2 mm/yr [b]	**0.61**	-0.20	**-0.43**	**0.77**	0.24	-0.02
$\%R_e$ 1 mm/yr [b]	**0.61**	-0.20	**-0.43**	**0.77**	0.24	-0.02
$\%R_e$ +15 °C [b]	**0.61**	-0.24	**-0.44**	**0.79**	0.25	-0.01
Molecular complexity [c]	**0.76**	-0.17	0.31	**0.39**	0.24	**0.47**
TMN [d]	**0.44**	-0.26	-0.34	**0.48**	0.19	-0.30
MPR-1 [d]	**0.38**	-0.24	**-0.43**	**0.46**	0.11	-0.22
MPR-2 [d]	**0.38**	-0.25	**-0.41**	**0.44**	0.08	-0.21
MPI-1 [d]	0.31	-0.27	-0.37	**0.38**	0.10	-0.30
MPI-2 [d]	0.33	-0.27	**-0.39**	**0.41**	0.10	-0.29
Ts/Tm [d]	**0.65**	**-0.67**	-0.15	**0.55**	0.23	-0.26
M/H [d]	-0.30	**0.39**	**0.64**	**-0.70**	**-0.41**	0.16
Steroid ratio [d]	**0.46**	-0.28	**-0.72**	**0.89**	**0.54**	0.13
Statistically significant SRCCs	21			17		

Bold numbers indicate correlations significant at the 95% confidence interval ($p < 0.05$).
[a] Interpolated pore water temperatures (see section 4.1).
[b] Vitrinite reflectance equivalent (see section 4.2).
[c] Direct measurement of the number of unique hydrocarbons per sample (see section 4.6.2).
[d] See Table 4 for details.

3.4.7 Coherence with kinetic maturation model and oil generation

For Cathedral Hill, evidence of generation is complicated by the ubiquitous presence of oil staining in all of the transect sediments. Nonetheless, considerable heterogeneity between the type and concentrations of hydrocarbons exists across the transect. Kinetic modeled vitrinite reflectance predicts the onset of the oil window at 15-18 cmbsf in core 1 and potentially close to the bottom of core 2 (18-21 cmbsf) corresponding to porewater temperatures of 145 °C and 125 °C, respectively (Table 2; Fig. 3). Although this measure marks the combined effects of exposure with time and heat, the bulk and

molecular data suggest generation may be occurring at shallower depths. Elavated polar and apolar recoveries, that are independent of TLE-proxied SOM diagenetic depth trends (Fig. 4), record an elevated narrow band at 6-12 cmbsf that progressively broadens outwards from the vent center (Table 1; Fig. 4). Evidence of thermochemical oxidative weathering at greater depths (>6-8 cmbsf) by means of elevated reductions in alkylated PAHs, perhydroPAHs, and benzothiophenes along side increased concentration of C_1-C_4 substituted products is provided by systematic increases to MPI values (Tables 4 and 5; SI-Fig. 5), the visual display of substrate level modifications of oils via the down core subtracted chromatogram series (Fig. 10), and a change in PC-1 and -3 scores to progressively lower values (Fig. 7). With these trends is also the systematic increase in aromatic and biomarker thermal maturity parameters that crosses the oil window (Table 4; SI-Fig. 5). Collectively, these data point to in situ oil generation beginning at the shallow depth of 6-10 cmbsf with enhanced thermochemical oxidation of more deeply buried sediments.

Closer alignment to the kinetic model of generation is observed in core 2. For the aromatic region, PC-1 scores peak within the deepest core sediments (Fig. 7) matching the increased abundance of alkylated aromatic compounds observed in the down core subtracted chromatogram series (Fig. 11). This trend is similarly paralleled by a systematic increase in aromatic and biomarker thermal maturity parameters (Table 4; SI-Fig. 5).

The rapid formation of hydrothermal petroleum has been documented at various hydrothermal vent sites such as the NE Pacific: Escanaba Trough, off the Gorda Ridge, and Middle Valley, off axis of the Juan de Fuca Ridge (Simoneit, 1993; Simoneit et al., 1997;

Rushdi and Simoneit, 2002; Gieskes et al., 2002) as well as the Bransfield Strait of Antarctica (Brault and Simoneit, 1989), and Lake Chapala in the Citala Rift, Mexico (Zárate-del Valle and Simoneit, 2005; Zárate-del Valle et al., 2006). At Middle Valley hydrothermal petroleum samples produced 29,000 yrs radiocarbon age dates (Simoneit et al., 1992a). Several sites at Guaymas Basin, have reported radiocarbon ages as young at 6,000 yrs (Simoneit and Kvenvolden, 1993) with oils being predicted to form within 10-30 m below the seafloor (Peter et al. 1991). Even more rapid generation, on the order of ~100 yrs, was estimated for the Escanaba Trough due to intense heating at 300 and 350 °C (Kvenvolden et al., 1988).

Interestingly, the two locations of putative hydrocarbon generation in the Cathedral Hill transect are nearly isothermal at 105-125 ˚C (Table 1; Fig. 15). This may indicate SOM residence time is of marginal importance relative to vent fluid temperature. It may also indicate the metal and sulfide concentrations in the vent fluids (Von Damm, et al., 1985; Edmond and Von Damm, 1983; Simoneit, 1985; Bazylinski, et al., 1988; McDermott et al., 2015) play a critical role in catalyzing hydrocarbon cleavage (Lewan, 1998; Seewald, 2001, 2003). The reduced vent fluids commonly contain high abundances and diversities of metals (Fouquet et al., 1998; Zierenberg and Miller, 2000) that can further affect the formation and alteration of petroleum products. Alternatively, the recorded thermal environment at Cathedral Hill may not reflect its conditions over larger spans of time as hydrothermal vents can experience highly variable temperature fluctuations over relatively short temporal intervals (Johnson et al., 1988). Irrespective of the thermal stability, such shallow burial requires generation to have occurred in only 40-80 yrs. If so, then reported

radiocarbon age dates of bulk hydrothermal petroleums from Guaymas Basin (Peter et al., 1991; Simoneit and Kvenvolden, 1993) hosted at higher temperature vent systems are likely time averaged, mixed hydrocarbon signatures that in some cases overestimates the time needed for the natural hydrothermal catagenesis of SOM hosted in unconsolidated sediments.

3.4.8. Thermochemical oxidation and migration

The aromatization and degree of alkylation in thermally generated hydrocarbons is a function of exposure temperature (Giger and Blumer, 1974; Youngblood and Blumer, 1975; Kawka and Simoneit, 1987, 1990, 1994). High temperature pyrolyzates display a predominance of unsubstituted PAH products (Kawka and Simoneit 1990, 1994; Fetzer et al., 1996; McCollom et al., 1999;) not typically found in lower temperature generated crude oils (60–150 °C; Hunt, 1984; Tissot and Welte, 1984). Based on the molecular changes identified by MPCA, the subtracted chromatograms, and MPI ratios (Figs. 7, 10-12, SI-Fig. 5; Table 4 and 5) increasing thermochemical oxidation in the form of aromatization and dealkylation reactions becomes increasingly prominent within the hottest sections of the transect (Fig. 15). Although the potential exists for PAH distributions to be controlled/derived from continential sources (i.e. forest fires, anthropogenic inputs, etc; Venkatesan, 1988; Bomboi et al., 1990; Benlahcen et al., 1997), the subtracted chromatograms (Fig. 10-13) reveal no systematic cross-core fluctuations that would suggest a linkage to sedimentary flux (as with Chapter 2 Fig. 5). Thus, it is likley that these

alternative sources play a minor role, if any, affecting the overall distribution of PAHs in the subsurface.

Additionally, elevated fluid flow and high thermal gradients promotes partial migration of petroleum products within oceanic hydrothermal systems (Simoneit, 1984; Simoneit et al., 1992b; Simoneit and Fetzer, 1996; Ventura et al., 2012; 2020). Some of these migrating oils can become trapped at the seafloor (Simoneit, 1984; Simoneit and Fetzer, 1996) or in the subsurface near the vent site (Simoneit et al., 1992b; Ishibashi et al., 2002) via quenching of the high-temperature hydrothermal fluids.

A similar situation occurs at Cathedral Hill (Fig. 15). The accumulation of migrated oil is evident as a band of oil that overprints the existing hydrocarbon laden sediments (Fig. 14 and 15). This band, spanning the transect at ~10 cmbsf, is compositionally distinct from the surrounding sediments (see section 3.4.6.2). Additionally, evidence of trapped oil at the sediment–water interface comes from anomalously high Ts/Tm, MPI, TMN and MPR ratios (Table 4; SI-Fig. 5) that are out of alignment with porewater temperature depth trends, which becomes increasingly pronounced further away from the vent center (Table 1; Fig. 3). Selective removal of components is likely occurring by advection from the circulating hydrothermal fluids, with soluble components being further water-washed at the seabed (Kawka, and Simoneit, 1987; Kawka, 1990). Finally, although, this study did not focus on the effects of biodegradation, these oils are frequently hosted in sediments that exhibit high rates of heterotrophy (Edgcomb et al., 2002; Biddle et al., 2012; Liu et al., 2016).

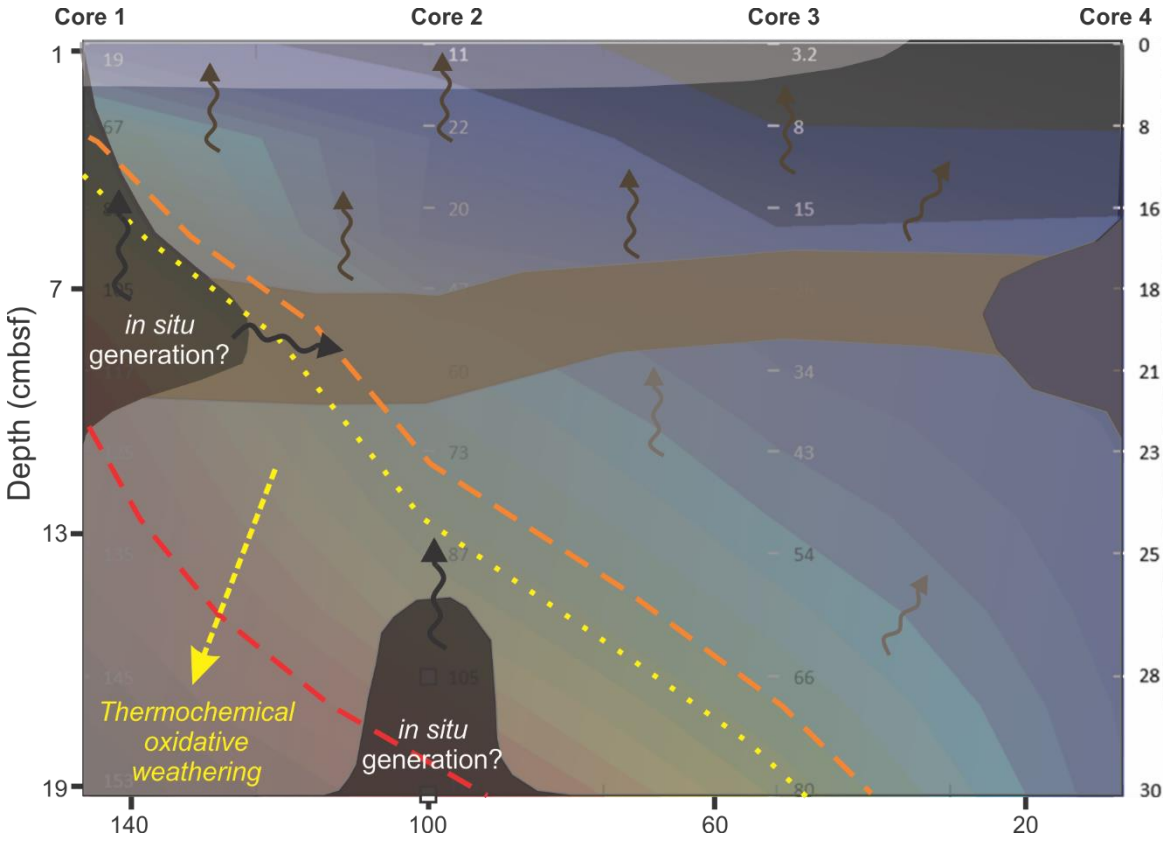


Fig. 15. Model of subsurface activity within the transect. Wavy arrows represent pervasive fluid flow in the subsurface. Orange, yellow, and red dotted lines indicate the onset of oil window for a conventional petroleum system, the thermal limit to biodegradation, and the kinetic modelled onset of oil generation by $\%R_e$ (section 4.2), respectively.

3.5 Conclusion

This study tests the resolving power of GC×GC-ToFMS and GC×GC-FID coupled with MPCA and HCA to identify and track multi-molecular changes between complex sets of oils sourced from the Cathedral Hill vent site at Guaymas Basin. Modeled chemometric outcomes were verified using difference chromatograms, summations of total unique hydrocarbons, and various aromatic and biomarker-based ratios. These data were then

compared to modeled vitrinite reflectance equivalent (%R_e) to further constrain zones of hydrocarbon generation within the SOM of a push core transect. Sensible statistically significant corrections were obtained for most PC-1 score comparisons for the thermal maturation parameters (pore water temperatures, %R_e, aromatic, hydrocarbon complexity and biomarker maturation parameters) suggesting that MPCA models can reliably track complex mixtures as they become modified by thermochemical oxidative reductive reactions, or become overprinted by migration events.

For regions close to the vent center at Cathedral Hill some level of hydrothermal petroleum is likely being generated at very shallow depths corresponding to 6-8 cmbsf reflecting SOM exposed to 105-125 ˚C vent temperatures with burial of 40-80 yrs. Below this depth thermochemical oxidative weathering of the oil becomes more pronounced with greater aromatization and dealkylation reactions taking place at depth. Additional advection as either a film or as plumes (blobs) appear to overprint the sediments with a band of immature oil forming a narrow layer ~10 cmbsf that extends out from the vent center. Together these data indicate the petroleum system of Cathedral Hill is highly complex at small spatial and temporal scales.

3.6 Acknowledgements

We thank the officers, crew, and pilots of the R/V Atlantis and HOV Alvin for their expert help at sea and their outstanding efforts acquiring the samples for this study. Funding for this study through NSERC Canadian Research Chair, Canada Foundation for Innovation (CFI) JELF-CRC, NSERC Discovery Grant (Application Number: RGPIN-2017-05822),

WHOI Deep Ocean Exploration Initiative 2008, Pengrowth Energy-Innovation Grant, and

NSF grant MCB-**0702677** (to JSS and SMS).

3.7 References

Aldenderfer M. and Blashfield, R. (1984) *Cluster analysis*. Sage Publications, California.

Arey J. S., Nelson R. K., Xu L. and Reddy C. M. (2005) Using comprehensive two dimensional gas chromatography retention indices to estimate environmental partitioning properties for a complete set of diesel fuel hydrocarbons. *Anal. Chem*. **77**, 7172–7182.

Arey J. S., Nelson R. K. and Reddy C. M. (2007) Disentangling oil weathering using GC GC. 1. Chromatogram analysis. *Environ. Sci. Technol*. **41**, 5738–5746.

Barbeira P. J. S., Pereira R. C. C. and Corgozinho C. N. C. (2007) Identification of gasoline origin by physical and chemical properties and multivariate analysis. *Energ. Fuel* **21**(4), 2212–2215.

Barman B. and Membrado V. (2000) Chromatographic techniques for petroleum and related products. *Crit. Rev. Anal. Chem.* **30**(2-3), 75–120.

Bazylinski A., Farrington W. and Jannasch W. (1988) Hydrocarbons in surface sediments from a Guaymas Basin hydrothermal vent site. *Org. Geochem.***12**, 547–558.

Benlahcen K. T., Chaoui A., Budzinski H., Bellocq J. and Garrigues, P. (1997) Distribution and sources of polycyclic aromatic hydrocarbons in some Mediterranean coastal sediments. *Mar. Pollut. Bull.* **34**(5), 298–305

Biddle J. F., Cardman Z., Mendlovitz H., Albert D.B., Lloyd K.G., Boetius A. and Teske A. (2012) Anaerobic oxidation of methane at different temperature regimes in Guaymas Basin hydrothermal sediments. *ISME J.* **6**(5), 1018–1031.

Boehm P. D., Douglas G. S., Burns W. A., Mankiewicz P. J., Page, D. S. and Bence A. E. (1997) Application of petroleum hydrocarbon chemical fingerprinting and allocation techniques after the Exxon Valdez oil spill. *Mar. Pollut. Bull.* **34**(8), 599–613.

Bomboi M. T., Hernández A., Mariño F. and Hontoria E. (1990) Distribution of hydrocarbon concentrations from urban runoff. *Sci. Total Environ.* **93**, 465–480.

Brault M. and Simoneit B. R. T. (1988) Steroid and triterpenoid distributions in Bransfield Strait sediments: Hydrothermally-enhanced diagenetic transformations. *Org. Geochem.* **13**, 697-705.

Brault M. and Simoneit B. R. T. (1989) Trace petroliferous organic matter associated with hydrothermal minerals from the Mid-Atlantic Ridge at the trans-Atlantic geotraverse 26°N site. *J. Geophys. Res.* **94**(C7), 9791.

Brooks, B. T., Bood C. E., Kurtz Jr. S. S. and Schmerling L. (1957) *The Chemistry of Petroleum Hydrocarbons,* Vol. 1. Reinhold Publishing, New York.

Burnham A.K. (2019) Kinetic models of vitrinite, kerogen, and bitumen reflectance. *Org. Geochem.* **131**, 50-59

Calvert S. E. (1966) Accumulation of diatomaceous silica in the sediments of the Gulf of California. *Geol. Soc. Am. Bull.* **77**, 569–596.

Castillo P. R., Hawkins J. W., Lonsdale P. F., Hilton D. R., Shaw A. M. and Glascock M. D. (2002) Petrology of Alarcon Rise lavas, Gulf of California: Nascent intracontinental ocean crust. *J. Geophys. Res.* **107**, 1-15

Christensen J. H. and Tomasi G. (2016) A multivariate approach to oil hydrocarbon fingerprinting and spill source identification. In *Standard Handbook Oil Spill Environmental Forensics: Fingerprinting and Source Identification, Second Edition* (eds. S. A. Stout and Z. Wang). Academic Press, Massachusetts, pp. 747–788.

Connolly J. F. (1966). Solubility of Hydrocarbons in Water Near the Critical Solution Temperatures. *J. Chem. Eng.* **11**(1), 13–16.

Corliss J. B., Dymond J., Gordon L. I., Edmond J. M., von Herzen R. P., Ballard R. D., Green K., Williams D., Bainbridge A., Crane K. and van Andel T. H. (1979) Submarine Thermal Springs on the Galápagos Rift. *Science* **203**(4385), 1073–1083.

Curray J., Moore D., and the DSDP Scientific Party. (1982) Guaymas Basin: Sites 477, 478, and 481. In *Initial reports of the Deep Sea Drilling Project-LXIV.* U.S. Govt. Printing Office, Washington DC.

Daniel W. W. (1978) *Applied nonparametric statistics*. Houghton Mifflin Harcourt, Massachusetts.

Didyk M., and Simoneit B. R. T. (1989). Hydrothermal oil of Guaymas Basin and implications for petroleum formation mechanisms. *Nature* **342**(6245), 65–69.

Didyk M. and Simoneit B. R. T. (1990) Petroleum characteristics of the oil in a Guaymas Basin hydrothermal chimney. *Appl. Geochem*. **5**, 29–40.

Dombrowski N., Teske A. P. and Baker B. J. (2018) Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* **9**(1), 1–13.

Edmond J. M. and Von Damm K. (1983) Hot springs on the ocean floor. *Sci. Am.* **248**, 78–93.

Eglinton T. I. and Douglas A. G. (1988) Quantitative study of biomarker hydrocarbons released from kerogens during hydrous pyrolysis. *Energ. Fuel.* **2**(1), 81–88.

Fetzer J. C., Simoneit B. R. T., Budzinski H. and Garrigues, P. (1996). Identification of Large PAHs in Bitumens from Deep-Sea Hydrothermal Vents. Polycycl. Aromat. Compd. **9**(1-4), 109–120.

Fisher A. T. and Becker K. (1991) Heat flow, hydrothermal circulation and basalt intrusions in the Guaymas Basin, Gulf of California. *Earth Planet. Sci. Lett.* **103**(1– 4), 84–99.

Fouquet Y., Zierenberg R. A., Miller D. J., Bahr J. M., Baker P. A., Bjerkgården T., Brunner C.A., Duckworth R.C., Gable R., Gieskes J., Goodfellow W. D., Gröschel-Becker H. M., Guèrin G., Ishibashi J., Iturrino G., James R. H., Lakschewitz K.S., Marquez L. L., Nehlig P., Peter J. M., Rigsby C. A., Simoneit B. R. T., Schultheiss P., Shanks W. C., Summit M., Teagle D. A. H., Urbat M. and Zuffa G. G. (1998) *Proceedings of the Ocean Drilling Program, Initial Reports,* Vol. 169. US Government Printing Office, Washington DC.

Frysinger G., Gaines R. and Reddy C. (2002) GC×GC—A new analytical tool for environmental forensics. *Environ. Forensics* **3**(1), 27–34.

Einsele G., Gieskes J., Curray J., Moore D., Aguayo E., Aubry M., Fornari D., Guerrero J., Kastner M., Kelts K., Lyle M., Matoba Y., Molina-Cruz A., Niemitz J., Rueda J., Saunders A., Schrader H., Simoneit B. and Vacquier V. (1980) Intrusion of basaltic sills into highly porous sediments and resulting hydrothermal activity. *Nature* **283**, 441–445.

Garrigues P., Saptorahardjo A., Gonzalez C., Wehrung P., Albrecht P., Saliot A. and Ewald M. (1986) Biogeochemical aromatic markers in the sediments from Mahakam Delta (Indonesia). *Org. Geochem.* **10**, 959–964.

Gieskes J., Simoneit B., Shanks W., Goodfellow W., James R., Baker P. and Ishibashi J. (2002) Geochemistry of fluid phases and sediments: relevance to hydrothermal circulation in Middle Valley, ODP Legs 139and 169. *Appl. Geochem*. **17**, 1381–1399.

Giger W. and Blumer M. (1974) Polycyclic aromatic hydrocarbons in the environment. Isolation and characterization by chromatography, visible, ultraviolet, and mass spectrometry. *Anal. Chem.* **46**(12), 1663–1671.

Gautheir, T. D. (2001) Detecting Trends Using Spearman's Rank Correlation Coefficient, *Environ. Forensics*, **2**(4), 359–362.

Haiping Lu, Plataniotis K.N. and Venetsanopoulos A. N. (2008) MPCA: Multilinear Principal Component Analysis of Tensor Objects. *IEEE T. Neur. Net. Lear*. **19**(1), 18–39

Hashemi-Nasab F. and Parastar H. (2019) Pattern recognition analysis of gas chromatographic and infrared spectroscopic fingerprints of crude oil for source identification. *Microchem. J.* **111**, 67–81.

Hoering T. C. (1968) Reactions of the organic matter in a recent marine sediment. *Carnegie Inst. Wash. Yearbook* **67**, 199–201.

Horsfield B. (1984) Pyrolysis studies and petroleum exploration. In *Advances in Petroleum Geochemistry* (eds. J. Brooks and D. H. Welte). Academic Press, London, pp. 247–292.

Hunt J. M. (1979) *Petroleum Geochemistry and Geology*. W.H. Freeman and Co., California.

Hunt J. M., Lewan M. D. and Hennet R .J. C. (1991) Modeling Oil Generation with Time-Temperature Index Graphs Based on the Arrhenius Equation 1. *AAPG Bull.* **75**, 795–807.

Ishibashi J. I., Sato M., Sano Y., Wakita H., Gamo T. and Shanks W. C. (2002) Helium and carbon gas geochemistry of pore fluids from the sediment-rich hydrothermal system in Escanaba Trough. *Appl. Geochem.* **17**, 1457–1466.

Ishiwatari R., Ishiwatari M., Rohrback B. and Kaplan I. (1977) Thermal alteration experiments on organic matter from recent marine sediments in relation to petroleum genesis. *Geochim. Cosmochim. Acta*, **41**(6), 815–828.

Johnson S. C. (1967) Hierarchical clustering schemes. *Psychometrika* **32**, 241–254.

Johnson K., Childress J. and Beehler C. (1988) Short-term temperature variability in the Rose Garden hydrothermal vent field: an unstable deep-sea environment. Deep Sea Research Part A. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **35**(10-11), 1711–1721.

Kawka E. (1990) Hydrothermal alteration of sedimentary organic matter in Guaymas Basin, Gulf of California. Ph.D. Thesis, Oregon State Univ. Corvallis.

Kawka E. and Simoneit B. (1987) Survey of hydrothermally-generated petroleums from the Guaymas Basin spreading center. *Org. Geochem.* **11**, 3 1 l–328.

Kawka E., Simoneit, B. (1990) Polycyclic aromatic hydrocarbons in hydrothermal petroleums from the Guaymas Basin spreading center. *Appl. Geochem*. **5**, 17–27.

Kawka E. and Simoneit B. (1994) Hydrothermal pyrolysis of organic matter in Guaymas Basin: I. Comparison of hydrocarbon distributions in subsurface sediments and seabed petroleums. *Org. Geochem.* **22**(6), 947–978.

Kelts K., Curray J. and Moore, D. (1982) Introduction and explanatory notes, DSDP leg 64 (California Basins). In *Initial reports of the Deep Sea Drilling Project-LXIV*. U.S. Govt. Printing Office, Washington DC.

Kvenvolden K. A., Rapp J. B., Hostettler F. D., Morton J. L., King J. D. and Claypool G. E. (1986) Petroleum Associated with Polymetallic Sulfide in Sediment from Gorda Ridge. *Science* **234**(4781), 1231–1234.

Kvenvolden K. A., Rapp J. B., Hostettler F. D., David King J. and Claypool G.E. (1988) Organic geothermometry of petroleum from Escanaba Trough, offshore northern California. *Org. Geochem*. **13**(1-3), 351–355.

Kvenvolden K. and Simoneit B. (1990) Hydrothermally derived petroleum: Examples from Guaymas Basin, Gulf of California and Escanaba Trough, Northeast Pacific. *A.A.P.G. Bull.* **74**, 223–237.

Lance G. N. and Williams W. T. (1967) A general theory of classificatory sorting strategies: 1. hierarchical systems. *Comput. J.* **9**(4), 373–380.

LaRowe DE, Arndt S., Bradley J. A., Estes E. R., Hoarfrost A., Lang, S. Q., Zhao, R. (2020) The fate of organic carbon in marine sediments - New insights from recent data and analysis. *Earth-Sci. Rev.* **204**, 103-146.

Larson R. L., Menard H. W. and Smith S. M. (1968) Gulf of California: A Result of Ocean-Floor Spreading and Transform Faulting. *Science* **161**(3843), 781–784.

Laso-Pérez R., Wegener G. and Knittel K. et al. (2016) Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396–401.

Leif R. N. and Simoneit, B. R. T. (1995) Confined-pyrolysis as an experimental method for hydrothermal organic synthesis. *Origins Life Evol. B.* **25**(5), 417–429.

Lewan M. D. (1985) Evaluation of petroleum generation by hydrous pyrolysis experimentation. *Phil Trans. Roy. Soc. London, Series A* **315**, 123-134.

Lewan M. (1998) Sulfur-radical control on petroleum formation rates. *Nature* **391**, 164–166.

Lewan M. D., Bjorøy M. and Dolcater D. L. (1986) Effects of thermal maturation on steroid hydrocarbons as determined by hydrous pyrolysis of Phosphoria Retort Shale. *Geochim. Cosmochim. Acta* **50**, 1977–1987.

Lonsdale P. (1977) Clustering of suspension feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep Sea Res.* **24**, 857–863.

Lonsdale P. and Becker K. (1985) Hydrothermal plumes, hot springs, and conductive heat flow in the southern trough of Guaymas Basin. *Earth Planet. Sci. Lett*. **73**, 211–225.

Lonsdale P. and Lawver L. (1980) Immature plate boundary zones studied with a submersible in the Gulf of California. *Geol. Soc*. **91**(9), 555.

Liu X-. K., Birgel D., Elling F. J., Sutton P. A., Lipp J. S., Zhu R., Zhang C., Könneke M., Peckmann J., Rowland S. J., Summons R. E. and Hinrichs K-. U. (2016) From ether to acid: a plausible degradation pathway of glycerol dialkyl glycerol tetraethers. *Geochim. Cosmochim. Acta* **183**,138–152.

Manly F. F. J. (1994). *Multivariate Statistical Methods: A Primer*. CRC Press, USA.

Marshall A. G., Hendrickson C. L. and Jackson G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17**(1), 1–35.

Martens C. S. (1990) Generation of short chain organic acid anions in hydrothermally altered sediments of the Guaymas Basin, Gulf of California. *Appl. Geochem*. **5**, 71–76.

McCollom T. M. and Seewald J. S. (2001) A reassessment of the potential for reduction of dissolved CO_2 to hydrocarbons during serpentinization of olivine. *Geochim. Cosmochim. Acta* **65**, 3769–3778.

McCollom T., Simoneit B. and Shock E. (1999) Hydrous pyrolysis of polycyclic aromatic hydrocarbons and implications for the origin of PAH in hydrothermal petroleum. *Energ. Fuel.* **13** (2), 401–410.

McDermott J. M., Ono S., Tivey M. K., Seewald J. S., Shanks W. C. and Solow A. R. (2015) Identification of sulfur sources and isotopic equilibria in submarine hot-springs using multiple sulfur isotopes. *Geochim. Cosmochim. Acta* **160**, 169–187.

McKay L., MacGregor B., Biddle J., Albert D., Mendlovitz H., Hoer D., Lipp J., Lloyd K. and Teske A. (2012) Spatial heterogeneity and underlying geochemistry of phylogenetically diverse orange and white Beggiatoa mats in Guaymas Basin hydrothermal sediments. *Deep-Sea Res.* **67**, 21–31.

Mesle M., Dromart G. and Oger P. (2013) Microbial methanogenesis in subsurface oil and coal. *Microbiol. Res.* **164**(9), 959–972.

Mogollón N., Ribeiro F., Poppi R., Quintana A., Chávez J., Agualongo D., Aleme H. and Augusto F. (2017) Exploratory analysis of biodiesel by combining comprehensive two-dimensional gas chromatography and multiway principal component analysis. *J. Braz. Chem. Soc.* **28**(5), 740–746.

Moore D. (1973) Plate-edge deformation and crustal growth, Gulf of California Structural Province. *Geol. Soc. Am. Bull.* **84**(6), 1883–1905.

Mullins O. C., Rodgers R. P., Weinheber P., Klein G. C., Venkataramanan L., Andrews A. B. and Marshall A. G. (2006) Oil reservoir characterization via crude oil analysis by downhole fluid analysis in oil wells with visible−Near-Infrared Spectroscopy and by Laboratory Analysis with electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Energ. Fuel.* **20**(6), 2448–2456.

Nelson R., Aeppli C., Arey J., Chen H., de Oliveira A., Eiserbeck C., Frysinger G., Gaines R., Grice K., Gros J., Hall G., Koolen H., Lemkau K., McKenna A., Reddy C., Rodgers R., Swarthout R., Valentine D. and White H. (2016) 8 -

Applications of comprehensive two-dimensional gas chromatography (GC × GC) in studying the source, transport, and fate of petroleum hydrocarbons in the environment. In *Standard Handbook Oil Spill Environmental Forensics* (2nd Edition). Academic Press, Massachusetts.

Nelson R. K., Kile B. M., Plata D. L., Sylva S. P., Xu L., Reddy C. M., Gaines R.B., Frysinger G.S. and Reichenbach S. E. (2006) Tracking the weathering of an oil spill with comprehensive two-dimensional gas chromatography. *Environ. Forensics*. **7**(1), 33–44.

Pavón J. L. P., Peña A. G., Pinto C. G. and Cordero B. M. (2006) Differentiation of types of crude oils in polluted soil samples by headspace-fast gas chromatography–mass spectrometry. *J. Chromatogr. A* **1137**(1), 101–109.

Pearson A., Seewald J. S. and Eglinton, T. I. (2005) Bacterial incorporation of relict carbon in the hydrothermal environment of Guaymas Basin. *Geochim. Cosmochim. Acta*, **69**(23), 5477–5486.

Peter J. M., Simoneit B. R. T., Kawka E. and Scott S.D. (1990) Liquid hydrocarbon-bearing inclusions in modern hydrothermal chimneys and mounds from the southern trough of Guaymas Basin. *Appl Geochem.* **5**, 51–63.

Peters K. E. and Moldowan J.M. (1991) Effects of source, thermal maturity and biodegradation on the distribution and isomerization of homohopanes in petroleum. Org. Geochem. **17**, 47–61.

Peter J., Peltonen P., Scott S., Simoneit B. and Kawka E. (1991) Carbon-14 ages of hydrothermal petroleum and carbonate in Guaymas Basin, Gulf of California Implications for oil generation, expulsion, and migration. *Geology* **19**, 253–256.

Peters K. E., Walters C. C. and Moldowan, J. M. (2005) *The Biomarker Guide. Biomarkers and Isotopes in the Environment and Human History,* Vol. 1. Cambridge University Press, UK.

Peters K. E., Walters C. C. and Moldowan J. M. (2007) *The Biomarker Guide Biomarkers and Isotopes in Petroleum Systems and Earth History*, Vol. 2. Cambridge University Press, UK.

Peters K. E. and Fowler M.G. (2002) Applications of petroleum geochemistry to exploration and reservoir management. *Org. Geochem*. **33**(1), 5–36.

Peters K. E., Scott Ramos L., Zumberge J. E., Valin Z. C., Scotese C. R. and Gautier D. L. (2007) Circum-Arctic petroleum systems identified using decision-tree chemometrics. *AAPG Bull.* **91**(6), 877–913.

Prata P. S., Alexandrino G. L., Mogollón N. G. S. and Augusto F. (2016) Discriminating Brazilian crude oils using comprehensive two-dimensional gas chromatography–mass spectrometry and multiway principal component analysis. *J. Chromatogr. A* **1472**, 99–106.

Radke M., Welte D.H., Willsch H., (1982a). Geochemical study of a well in the Western Canada Basin: relation of the aromatic distribution pattern to maturity of organic matter. *Geochim. Cosmochim. Acta* **46**, 1–10.

Radke M., Willsch H., Leythaeuser D. and Teichmüller M. (1982b) Aromatic components of coal: relation of distribution pattern to rank. *Geochim. Cosmochim. Acta* **46**, 1831–1848.

Radke M., Welte D. H. and Willsch H. (1986) Maturity parameters based on aromatic hydrocarbons: influence of the organic matter type. *Org. Geochem.* **10**, 51–63.

Radke M., Welte D.H. and Willsch H. (1991) Distribution of alkylated aromatic hydrocarbons and dibenzothiophenes in rocks of the Upper Rhine Graben. *Chem. Geol.* **93**, 325–341.

Reddy C., Eglinton T., Hounshell A., White H., Xu L., Gaines R. and Frysinger G. (2002) The West Falmouth oil spill after thirty years: the persistence of petroleum hydrocarbons in marsh sediments. *Environ. Sci. Technol.* **36**(22), 4754–4760.

Reddy C., Nelson R., Sylva S., Xu L., Peacock E., Raghuraman B. and Mullins O. (2007) Identification and quantification of alkene-based drilling fluids in crude oils by comprehensive two-dimensional gas chromatography with flame ionization detection. *J. Chromatogr. A* **1148**, 100–107.

Reeves E. (2010) Laboratory and Field-Based Investigations of Subsurface Geochemical Processes in Seafloor Hydrothermal Systems. Ph. D. thesis, Massachusetts Institute of Technology.

Rencher A.C. (2003) *Methods of multivariate analysis* (2nd ed.). John Wiley and Sons, New York, pp. 451-481.

Rushdi A. I. and Simoneit B. R. T. (2002a) Hydrothermal alteration of organic matter in sediments of the northeastern Pacific Ocean: Part 1. Middle Valley, Juan de Fuca Ridge. *Appl. Geochem.* **17**, 1401–1428.

Rushdi A. I. and Simoneit B. R. T. (2002b) Hydrothermal alteration of organic matter in sediments of the northeastern Pacific Ocean: Part 2. Escanaba Trough, Gorda Ridge. *Appl. Geochem*. **17**, 1467–1494.

Rusnak G., Fisher R. and Shepard F. (1964) Bathymetry and Faults of Gulf of California. in, Andel, Tj. H. Van, And Shor, G. G., eds., Marine geology of the Gulf of California: *Am. Assoc. Petroleum Geologists Mem*. **3**, 59–75.

Saltelli A. (1990) Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. *Reliab. Eng. Syst. Safe.* **28**, 229–253.

Sanders N. D. (1986) Visual observation of the solubility of heavy hydrocarbons in near-critical water. *Ind. Eng. Chem. Fund* **25,** 169-171.

Seewald J. S. (2001) Aqueous geochemistry of low molecular weight hydrocarbons at elevated temperatures and pressures: constraints from mineral buffered laboratory experiments. *Geochim. Cosmochim. Acta* **65**, 1641–1664.

Selley R. (1997) *Elements of Petroleum Geology*. Academic Press, Massachusetts.

Simoneit B. R. T. (1984) Hydrothermal effects on organic matter-high versus low temperature components. *Org. Geochem.* **6** 857–864.

Simoneit B. R. T. (1985) Hydrothermal petroleum: genesis, migration, and deposition in Guaymas Basin, Gulf of California. *Can. J. Earth Sci*. **22**, 1919–1929.

Simoneit B. R. T. (1990) Petroleum generation, an easy and widespread process in hydrothermal systems: an overview. *Appl. Geochem*. **5**(1-2), 3–15.

Simoneit B. R. T. (1992) Natural hydrous pyrolysis: Petroleum generation in submarine hydrothermal systems. In *Organic Matter: Productivity, Accumulation, and Preservation in Recent and Ancient Sediments* (eds. J. Whelan and J. W. Farrington). Columbia University Press, New York, pp. 368-403

Simoneit B. R. T. (1993) Hydrothermal activity and its effects on sedimentary organic matter. In *Bitumens in Ore Deposits* (eds. J. Parnell, H. Kucha and P. Landais) Springer-Verlag, Berlin, pp. 81-96

Simoneit B. R. T. and Kvenvolden, K. (1993) Comparison of ^{14}C ages of hydrothermal petroleums. *Org. Geochem.* **21**, 525–529.

Simoneit B. R. T., Goodfellow W. D. and Franklin J. M. (1992a) Hydrothermal petroleum at the seafloor and organic matter alteration in sediments of Middle Valley, northern Juan de Fuca Ridge. *Appl. Geochem.* **7**, 257-264.

Simoneit B.R.T., Leif R.N., Sturz A.A., Sturdivant A.E. and Gieskes J.M. (1992). Geochemistry of shallow sediments in Guaymas Basin, Gulf of California: hydrothermal gas and oil migration and effects of mineralogy. *Org. Geochem.* **18**, 765–784.

Simoneit B. R. T. and Lonsdale P. (1982) Hydrothermal petroleum in mineralized mounds at the seabed of Guaymas Basin. *Nature* **295**, 198–202.

Simoneit B. R. T. and Fetzer J. C. (1996.) High molecular weight polycyclic aromatic hydrocarbons in hydrothermal petroleums from the Gulf of California and northeast Pacific Ocean. *Org. Geochem.* **24**, 1065–1077.

Simoneit B. R. T, Schoell M. and Kvenvolden K. (1997) Carbon isotope systematics of individual hydrocarbons in hydrothermal petroleum from Escanaba Trough, Northeastern Pacific Ocean. *Org. Geochem.* **26**, 511–515.

Siskin M. and Katritzky A. R. (1991) Reactivity of organic compounds in hot water: geochemical and technological implications. *Science* **254**, 231–237.

Siskin M. and Katritzky A. R. (1999) Overview of the reactivity of organics in superheated water: geochemical and technology implications. In *Preprints Symposium – American Chemical Society, Division of Fuel Chemistry,* Vol. 44. ACS Publications, USA, pp. 324–330.

Siskin M. and Katritzky A. R. (2001) Reactivity of organic compounds in superheated water: general background. *Chem. Rev*. **101**, 825–835.

Teske A., de Beer D., McKay L., Tivey M., Biddle J., Hoer D., Lloyd K., Lever M., Roy H., Albert D., Mendlovitz H. and MacGregor B. (2016) The Guaymas Basin hiking guide to hydrothermal mounds, chimneys, and microbial mats: complex seafloor expressions of subsurface hydrothermal circulation. *Front. Microbiol.* **7**, 75–98.

Teske A., Hinrichs K.-U., Edgcomb V. P., de Vera Gomez A., Kysela D., Sylva S. P., Sogin M. and Jannasch H. (2002) Microbial diversity in hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl. Environ. Microbiol.* **68**, 1994–2007.

Tissot B. and Espitalie J. (1975) L'evolution thermique de la matière organique des sédiments : applications d'une simulation mathématique. Potentiel pétrolier des bassins sédimentaires de reconstitution de l'histoire thermique des sédiments. *Rev. Inst. Francais Pétrole* **30**(5), 743-778

Tissot B. and Welte D. (1984) *Petroleum Formation and Occurrence* (2nd edition). Springer-Verlag, Berlin.

Tran T., Logan G., Grosjean E., Ryan D. and Marriott P. (2010) Use of comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for the characterization of biodegradation and unresolved complex mixtures in petroleum. *Geochim. Cosmochim. Acta* **74**, 6468–6484.

Tran C. and Marriott P. (2007) Characterization of incense smoke by solid phase microextraction—Comprehensive two-dimensional gas chromatography (GCxGC). *Atmos. Environ.* **41**, 5756–5768.

Van Andel Tj. H. (1964) Recent marine sediments in the Gulf of California. In *Marine Geology of the Gulf of California: a symposium* (eds. Tj. H. van Andel and G. G. Shor Jr.) American Association of Petroleum Geologists, Oklahoma. pp. 216-311.

Van Graas G., de Leeuw J., Schenck P. and Haverkamp J. (1981) Kerogen of Toarcian shales of the Paris Basin. A study of its maturation by flash pyrolysis techniques. *Geochim. Cosmochim. Acta* **45**(12), 2465–2474.

Venkatesan M. I. (1988) Occurrence and possible sources of perylene in marine sediments-a review. *Mar. Chem.* **25**(1), 1–27.

Ventura G. T., Hall G. J., Nelson R. K., Frysinger G. S., Raghuraman B., Pomerantz A. E., Mullins O. C. and Reddy C. M. (2011) Analysis of petroleum compositional similarity using multiway principal components analysis (MPCA) with comprehensive two-dimensional gas chromatographic data. *J. Chromatogr. A* **1218**(18), 2584–2592.

Ventura G. T., Kenig F., Reddy, C. M., Schieber J., Frysinger G. S., Nelson R. K., Dinel E., Gaines R. B. and Schaeffer P. (2007) Molecular evidence of Late Archean archaea and the presence of a subsurface hydrothermal biosphere. *P.N.A.S.* **104**(36), 14260–14265.

Ventura G. T., Kenig F., Reddy C. M., Frysinger G. S., Nelson R. K., Mooy B. V., and Gaines R. B. (2008) Analysis of unresolved complex mixtures of hydrocarbons extracted from Late Archean sediments by comprehensive two-dimensional gas chromatography (GC×GC). *Org. Geochem.* **39**(7), 846–867.

Ventura G. T., Rossel P. E., Simoneit B .R. T. and Dittmar T. (2020) Analysis of thermochemical alteration of hydrothermally generated maltenes in sediments from the Escanaba Trough, northeast Pacific using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). Journal: *Org. Geochem*. Accepted with revision.

Ventura G. T., Simoneit B. R. T., Nelson R. K. and Reddy C. M. (2012) The composition, origin and fate of complex mixtures in the maltene fractions of hydrothermal petroleum assessed by comprehensive two-dimensional gas chromatography. *Org. Geochem.* **45**, 48–65.

Von Damm K. L., Edmond J. M., Measures C. I. and Grant B. (1985) Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochim. Cosmochim. Acta* **49**(11), 2221–2237.

Wang, Z. and Stout S. (2007) *Oil Spill Environmental Forensics. Fingerprinting and Source Identification.* Elsevier, Amsterdam.

Wang Y.-P., Zou Y.-R., Shi J.-T. and Shi J. (2018) Review of the chemometrics application in oil-oil and oil-source rock correlations. *J. Nat. Gas. Geo*. **3**, 217–232.

Wardlaw G. D., Arey J. S., Reddy C. M., Nelson R. K., Ventura G. T. and Valentine D. L. (2008) Disentangling Oil Weathering at a Marine Seep Using GC×GC: Broad Metabolic Specificity Accompanies Subsurface Petroleum Biodegradation. Environmental Science and Technology **42**(19), 7166–7173.

Weber A. and Jørgensen B. (2002) Bacterial sulfate reduction in hydrothermal sediments of the Guaymas Basin, Gulf of California, Mexico. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **49**(5), 827–841.

Whelan J. K., Simoneit B. R. T. and Tarafa M. E. (1988) C1-C8 hydrocarbons in sediments from Guaymas Basin, Gulf of California – Comparison to Peru Margin, Japan Trench, and California Borderlands. *Org. Geochem*. **12**, 171–194.

Williams D., Becker K., Lawver L. and Von Herzen R. (1979) Heat flow at the spreading centers of the Guaymas Basin, Gulf of California. *J. Geophys. Res.* **84**, 6757–6796.

Wise B., Gallagher N., Bro R., Shaver J., Windig W., and Koch. R. (2006) *PLS Toolbox 4.0 for use with MATLAB™*. Eigenvector Research, Inc., Massachusetts.

Ward J. H. (1963) Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**(301), 236.

Wold S., Geladi P., Esbensen K., and Öhman J. (1987) Multi-way principal components-and PLS-analysis. *J. Chemom.* **1**(1), 41–56.

Youngblood W. W. and Blumer M. (1975) Polycyclic aromatic hydrocarbons in the environment: homologous series in soils and recent marine sediments. *Geochim. Cosmochim. Acta* **39**(9), 1303–1314.

Zárate-del Valle P. F., Rushdi A. I. and Simoneit B. R. T. (2006) Hydrothermal petroleum of Lake Chapala, Citala Rift, western Mexico: Bitumen compositions from source sediments and application of hydrous pyrolysis. *Appl. Geochem.* **21**(4), 701–712.

Zárate-del Valle P.F. and Simoneit B. R. T. (2005) Hydrothermal bitumen generated from sedimentary organic matter of rift lakes – Lake Chapala, Citala Rift, western Mexico. *Appl. Geochem.* **20**(12), 2343–2350.

Zierenberg R. A., Miller D. J. (2000) Overview of ocean drilling program leg 169: Sedimented ridges II. In: *Proceedings of the Ocean Drilling Program, Scientific Results,* Vol. 169. ODP, Texas.

Zhan Z.-W., Lin X.-H., Zou Y.-R., Li Z., Wang D., Liu C. and Peng P. (2019) Chemometric differentiation of crude oil families in the southern Dongying Depression, Bohai Bay Basin, China. *Org. Geochem.* **127**, 37–49.

Zhuang G.-C., Montgomery A., Samarkin V. A., Song M., Liu J., Schubotz F., Teske A., Hinrichs K.-U. and Joye S. B. (2019) Generation and utilization of volatile fatty acids and alcohols in hydrothermally altered sediments in the Guaymas Basin, Gulf of California. *Geophys. Res. Lett.* **46**(5), 2637–2646.

Chapter 3 Supplementary: Resolution of multi-molecular hydrocarbon transformation in petroleum-bearing sediments from the Cathedral Hill hydrothermal vent complex at Guaymas Basin, Gulf of California by comprehensive two-dimensional gas chromatography and chemometric analyses

Connor J. Dalzell[1], G. Todd Ventura[1*], Robert K., Nelson[2], Clifford C. Walters[3], Christopher M. Reddy[2], Jeffrey S. Seewald[2], Stefan M. Sievert[2]

Affiliations:

[1] Saint Mary's University, Canada.
[2] Woods Hole Oceanographic Institution, USA.
[3] ExxonMobil Research and Engineering, USA.
[*] Corresponding author: todd.ventura@smu.ca

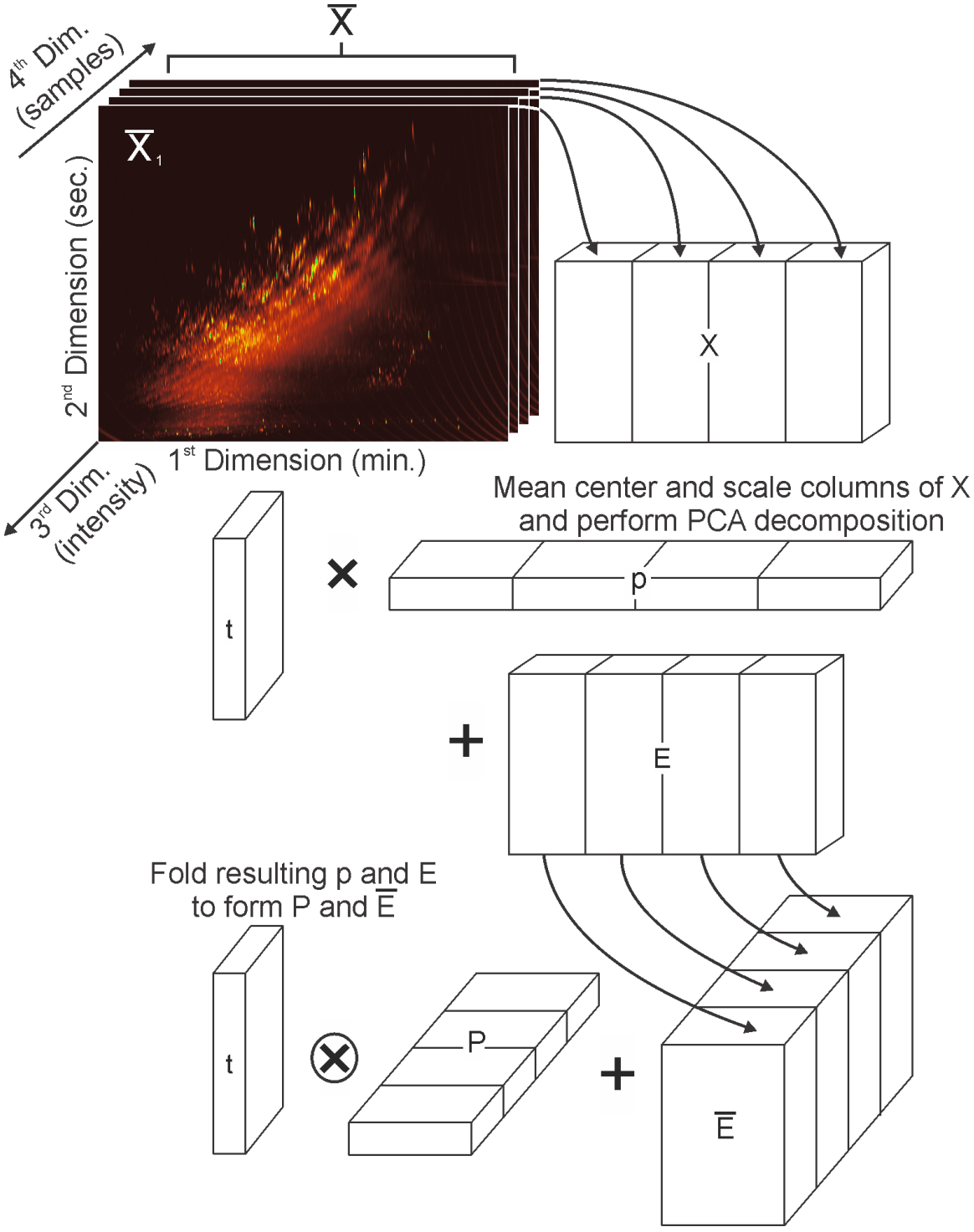
Supplemental Information

SI Text:

SI Tables:

SI Figures:

SI References

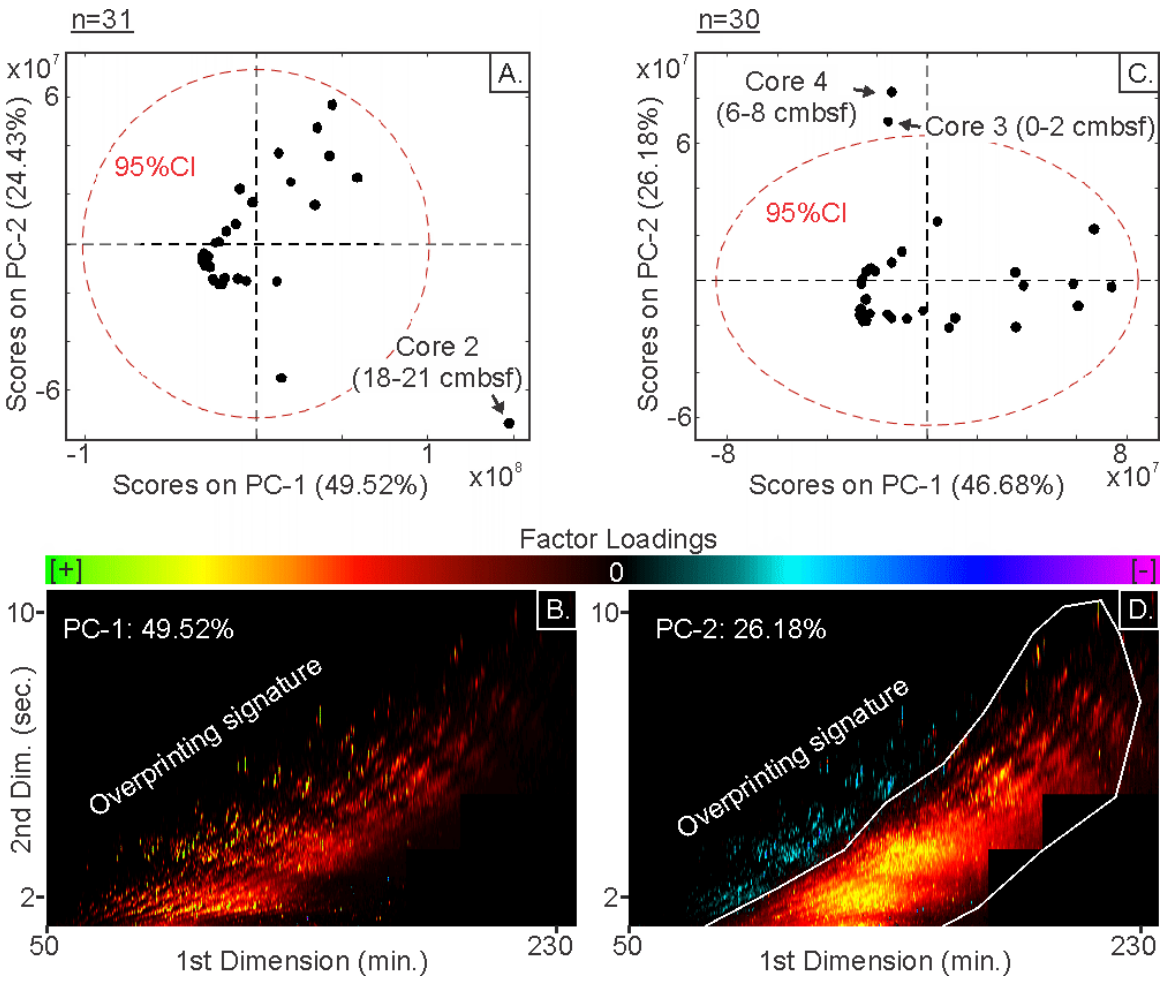


SI Fig. 1. MPCA algorithm as applied to stacked GC×GC chromatograms (modified from Wise et al., 2006).

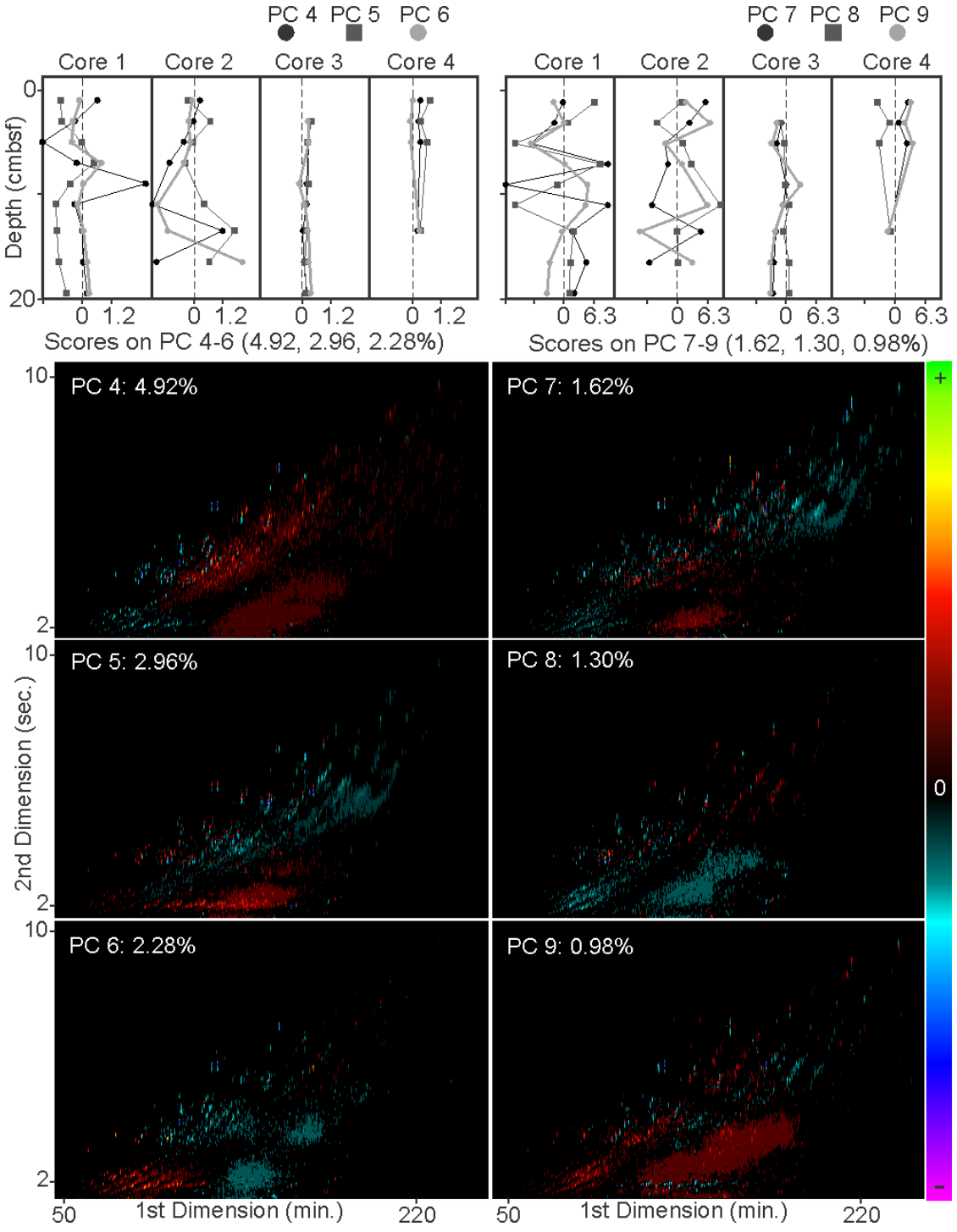
1.0 Detection of sample outliers and the range of calculated PCs

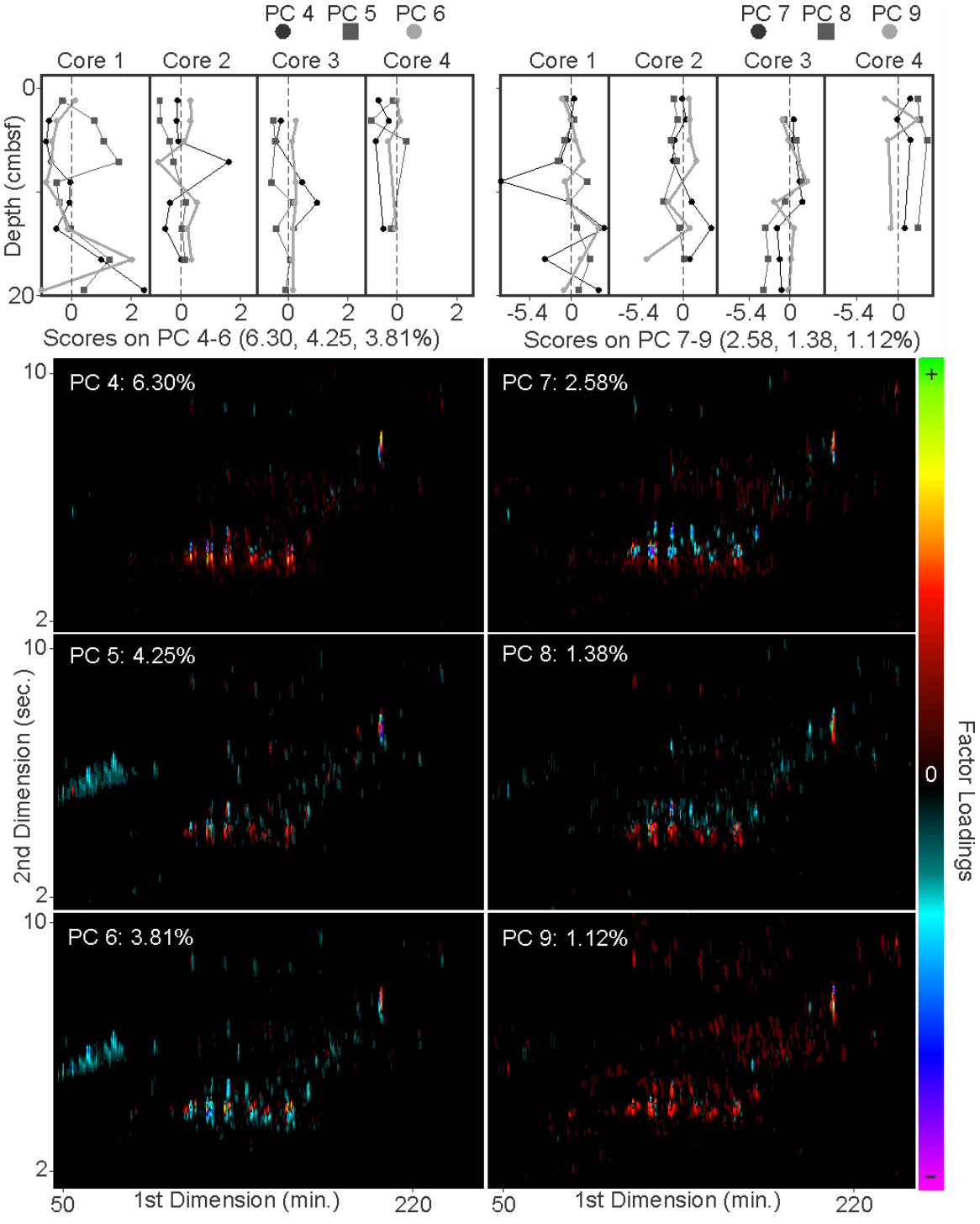
Compositionally distinct outliers were identified on PCs -1 and -2 for the Cathedral Hill sample chromatograms (n=34) (see section 4.4.1). Several outliers, not apparent in Fig. 5, only became observable once the most extreme outliers were removed (SI Fig. 2). An additional outlier, core 2 sample 8-10 cmbsf, was removed via a similar iterative process. The final data set for MPCA and HCA analyses was composed of 27 samples.

Additionally, more than 3 PCs were run to determine if high dimension data reductions would yield meaningful information. These PCs were of a low total summed variance (<7%) and produced mostly near-zero scores with few systematic down-core trends. From these experiments it was determined that PCs >3 did not resolve meaningful, or easily interpretable trends nor did they elucidate new biomarker patterns that had not already emerged in the lower-order PCs (SI Fig. 4).



SI Fig. 2. Scores and factor loadings plots of aromatics region indicating the presence of sample outliers. Plots A and B are the MPCA model results derived after removal of 3 outliers identified from Fig. 5 (core 4 sample 8-10 and 10-12 cmbsf, core 3 sample 6-8 cmbsf; n=31). The removal of these outliers exposed an additional outlier (core 2 sample 18-21 cmbsf), which was removed in plots C and D to reveal 2 additional outliers (core 4 sample 6-8 and core 3 sample 0-2 cmbsf).


PC 4 PC 5 PC 6
Core 1 Core 2 Core 3 Core 4
Depth (cmbsf)
0
20
0 1.2
Scores on PC 4-6 (4.92, 2.96, 2.28%)
PC 7 PC 8 PC 9
0 6.3
Scores on PC 7-9 (1.62, 1.30, 0.98%)
PC 4: 4.92%
PC 7: 1.62%
PC 5: 2.96%
PC 8: 1.30%
PC 6: 2.28%
PC 9: 0.98%
10
2
2nd Dimension (sec.)
50
1st Dimension (min.)
220
+
0
−
Factor Loadings



SI Fig. 3. Scores and factor loadings plots of the aromatics and biomarker region for higher PCs -4 to -9 (n=27). For the factor loadings plots, warm colours indicate positively weighted hydrocarbons (positive loadings) and cool colours indicate the opposite (negative

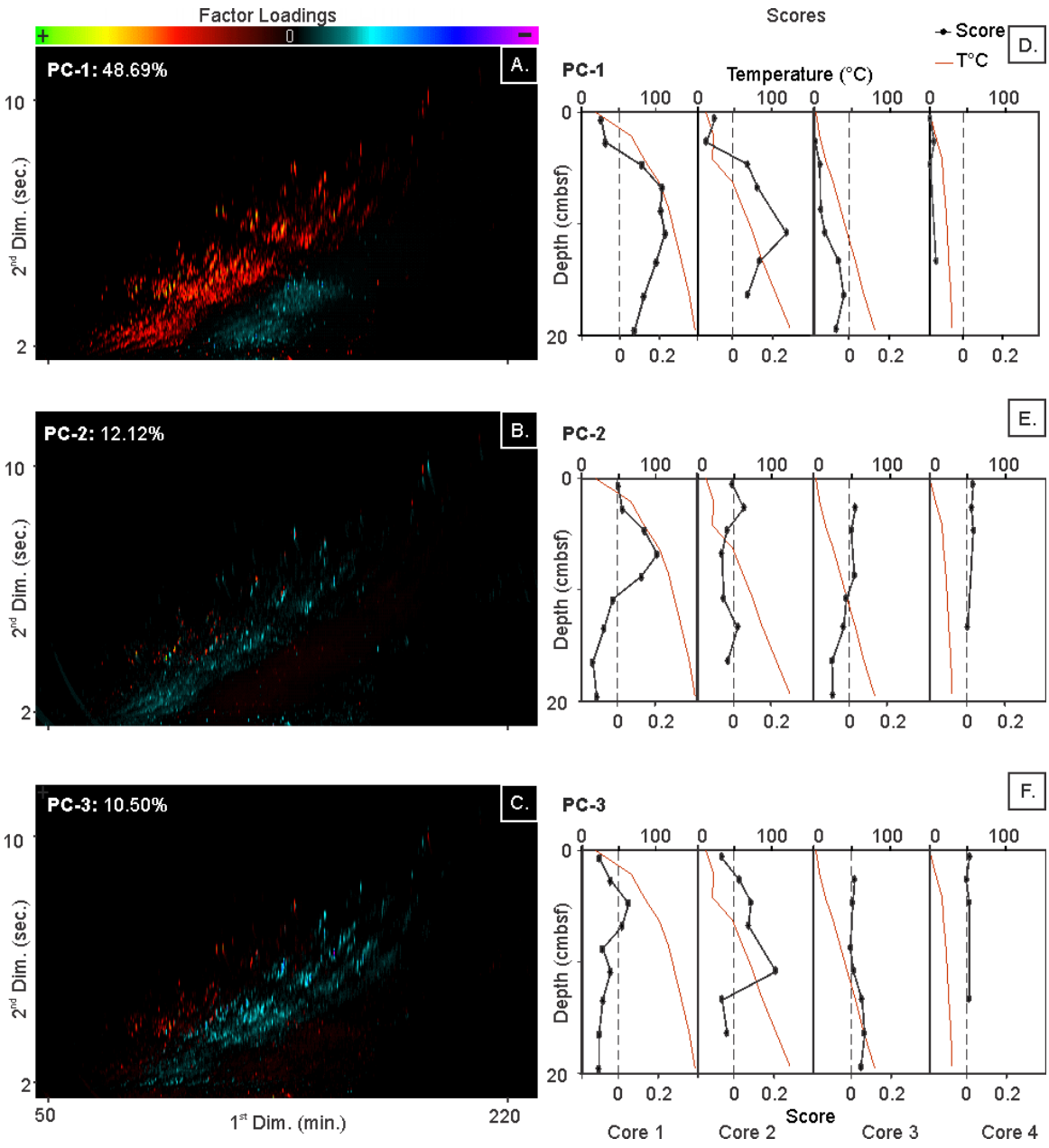
loadings). Sample scores replotted core depth. The calculation of higher-order PCs (beyond PC-3) do not illuminate systematic trends with respect to geochemical or geophysical conditions in the sediments or meaningful degrees of variance within new biomarkers.

2.0 Normalization

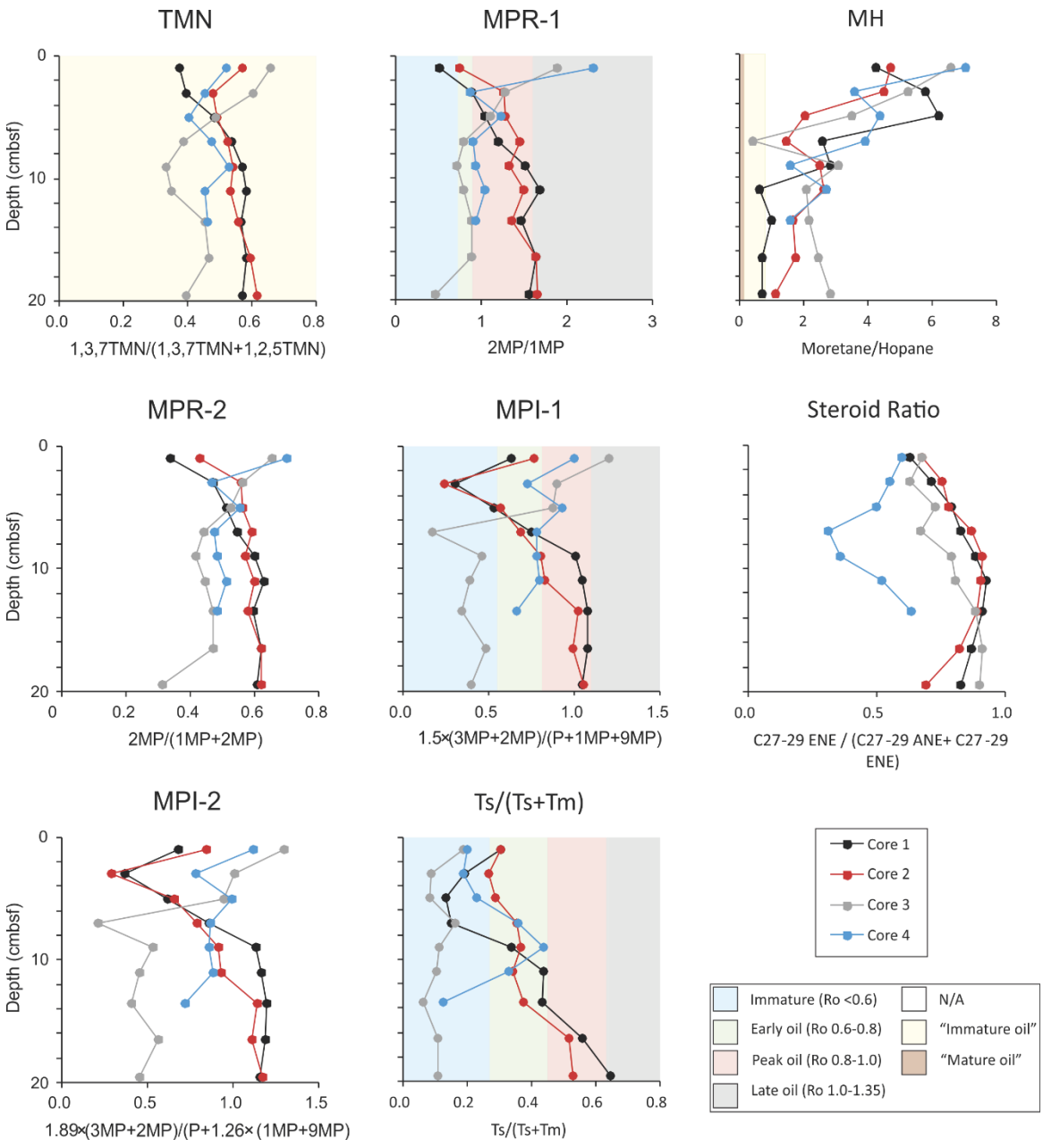
Two normalization techniques were considered to compensate for differences in injection concentration between GC×GC-FID chromatograms. The first technique was internal standard normalization (ISN). This technique would ideally be most viable as all matrix elements would be scaled to a non-sample controlled value. However, the chromatographic runs were not initially made with an internal standard and re-running the samples proved not to be an option. An alternative strategy would be to select a recalcitrant compound present in all samples and scale the matrix elements to that compound. However, it was not clear what compound could be suitable. All of the Cathedral Hill samples have variable loadings of material with different source and weathering profiles. Scaling to any compound that systematically changed would falsely imbed structure to the matrix producing false positive statistical correlations.

The second technique was a total sum normalization (TSN; SI Fig. 2). For TSN each sediment extract sample was normalized to the averaged molecular profile of the entire transect. With TSN, scores and factor loadings produced similar trends for PCs -1 and -2 when compared to the non-normalized model (Fig. 7 and SI Fig. 3). However, PC-3 factor loadings appear similar to PC-2 with somewhat comparable scores trends. This represents a loss of information compared to the non-normalized model (Fig. 7), which had a unique PC-3 profile that showed systematic changes with depth. Additionally, MPCA results

appear to be affected by normalization induced artifacts that translated into fewer statistically significant SRCCs when compared to that of the unmodified data set (SI Table 2). For these reasons, an unmodified data set was selected for the MPCA and HCA study.



SI Fig. 4. Scores and factor loadings plots of the aromatics region after TSN. For the factor loadings plots (A-C), warm colours indicate positively weighted hydrocarbons (positive loadings) and cool colours indicate negatively weighted hydrocarbons (negative loadings). Sample scores are replotted by core sediment depth alongside the thermal profile (red line: D-F).



SI Fig. 5. Thermal maturity parameters versus depth (see Table 4 for details).

SI Table 1. Spearman's rank correlation coefficient between PC scores and temperature, vitrinite reflectance equivalent (%R_e), molecular complexity, and various maturation parameters (n =27).

Spearman's rank correlation coefficient	TSN Aromatic Region			TSN Biomarker Region			Unmodified Aromatic Region			Unmodified Biomarker Region		
	PC-1	PC-2	PC-3	PC-1	PC-2	PC-3	PC-1	PC-2	PC-3	PC-1	PC-2	PC-3
Temperature [a]	**0.83**	-0.36	-0.32	**-0.85**	0.17	0.03	**0.68**	-0.29	**-0.44**	**0.83**	0.35	0.08
%R_e -15 °C [b]	**0.72**	-0.33	**-0.43**	**-0.73**	0.00	-0.05	**0.62**	-0.20	-0.35	**0.72**	0.20	0.00
%R_e 2 mm/yr [b]	**0.76**	-0.29	-0.40	**-0.81**	0.05	-0.04	**0.61**	-0.20	**-0.43**	**0.77**	0.24	-0.02
%R_e 1 mm/yr [b]	**0.76**	-0.29	-0.40	**-0.81**	0.05	-0.04	**0.61**	-0.20	**-0.43**	**0.77**	0.24	-0.02
%R_e +15 °C [b]	**0.77**	-0.33	-0.41	**-0.84**	0.05	-0.02	**0.61**	-0.24	**-0.44**	**0.79**	0.25	-0.01
Molecular complexity [c]	**0.57**	0.20	-0.13	-0.32	-0.33	0.09	**0.76**	-0.17	0.31	**0.39**	0.24	**0.47**
TMN [d]	**0.48**	-0.16	-0.30	**-0.46**	-0.14	-0.20	**0.44**	-0.26	-0.34	**0.48**	0.19	-0.30
MPR-1[e]	**0.43**	-0.12	-0.16	**-0.42**	-0.04	-0.33	**0.38**	-0.24	-0.43	**0.46**	0.11	-0.22
MPR-2 [f]	**0.42**	-0.10	-0.17	**-0.42**	-0.07	-0.34	**0.38**	-0.25	**-0.41**	**0.44**	0.08	-0.21
MPI-1[g]	0.32	-0.16	**-0.58**	**-0.41**	-0.07	**-0.38**	0.31	-0.27	-0.37	**0.38**	0.10	-0.30
MPI-2 [h]	0.35	-0.16	**-0.56**	**-0.42**	-0.06	**-0.40**	0.33	-0.27	**-0.39**	**0.41**	0.10	-0.29
Ts/Tm [i]	**0.57**	-0.31	**-0.53**	**-0.69**	-0.29	-0.03	**0.65**	**-0.67**	-0.15	**0.55**	0.23	-0.26
M/H [j]	**-0.59**	**0.68**	0.19	**0.60**	**-0.46**	0.08	-0.30	**0.39**	**0.64**	**-0.70**	**-0.41**	0.16
Steroid ratio [k]	**0.80**	**-0.55**	0.06	**-0.83**	**0.51**	-0.06	**0.46**	-0.28	**-0.72**	**0.89**	**0.54**	0.13
Statistically significant SRCCs	18			17			21			17		

All compounds quantified in their respective GC×GC–FID sample chromatograms.

Bold numbers indicate correlations significant at the 95% confidence interval ($p < 0.05$; n=27).

[a] Interpolated pore water temperatures (see section 4.1).

[b] Easy%R_o kinetic model (Burnham 2019) (see section 4.2).

[c] Direct measurement of the number of unique hydrocarbons per sample (see section 4.6.2).

[d] Trimethylnaphthalene ratio (TMN) = 1,3,7-trimethylnaphthalene /(1,3,7-trimethylnaphthalene + 1,2,5-trimethylnaphthalene), identified via monitoring of the *m/z* 170 mass chromatograms.

[e] Methylphenanthrene ratio 1 (MPR-1) = 2-methylphenanthrene / 1-methylphenanthrene identified via monitoring of the *m/z* 191 mass chromatograms and quantified in their respective GC×GC–FID sample chromatograms.

[f] Methylphenanthrene ratio 2 (MPR-2) = 2-methylphenanthrene / (1-methylphenanthrene + 2-methylphenanthrene), identified via monitoring of the *m/z* 192 mass chromatograms.
[g] Methylphenanthrene Index 1 (MPI-1) = 1.5 × (3-methylphenanthrene + 2-methylphenanthrene)/(phenanthrene + 9-methylphenanthrene + 1-methylphenanthrene) (Radke et al., 1982a). Monomethylphenanthrenes, identified via monitoring of the *m/z* 192 mass chromatograms.
[h] Methylphenanthrene Index 2 (MPI-2) = 1.89× (3-methylphenanthrene + 2-methylphenanthrene)/(phenanthrene + 1.26 ×(9-methylphenanthrene + 1-methylphenanthrene), identified via monitoring of the *m/z* 192 mass chromatograms.
[i] Ts/Tm = C_{27} 17α(H)-22,29,30-trisnorhopane /(C_{27} 17α(H)-22,29,30-trisnorhopane + 18α(H)-22,29,30-trisnorneohopane) (Table 3 compounds 1 and 2) identified via monitoring of the *m/z* 191 mass chromatograms.
[j] M/H = βα-hopane/ αβ-hopane (Table 3 compounds 11 and 7), identified via monitoring of the *m/z* 191 mass chromatograms.
[k] Steroid ratio = ΣC_{27}-C_{29} sterenes / (ΣC_{27}-C_{29} steranes + ΣC_{27}-C_{29} sterenes) marking compounds 3-6, 9-12, and 15-18 from Table 3. Compounds were identified via monitoring of the *m/z* 217 and 215 mass chromatograms.

3.0 SI References

Wise B., Gallagher N., Bro R., Shaver J., Windig W., and Koch. R. (2006) *PLS Toolbox 4.0 for use with MATLAB™*. Eigenvector Research, Inc., Massachusetts.

Chapter 4: Conclusions and future work

In this study, the degree to which the Cathedral Hill transect sediments are affected by in situ oil production, microbial activity, and migration was established using conventional biomarker analysis and multi-molecular chemometric analyses.

The biomarker results (**Chapter 2**) reveal increasing levels of thermal maturity with depth and proximity to the hydrothermal vent edifice. However, there is no substantiated, direct evidence for in situ oil production. Although, cracking/onset of the oil window is approximated within ~8-10 cmbsf in cores 1 and 2, there is significant conflict between the levels of maturity predicted by the biomarker proxies. Evidence of migration and the potential for temperature flux additionally cast doubt on any definitive areas of generation. The biomarker results showed ubiquitous microbial activity across the transect, excluded only in areas past the hydrocarbon biodegradation threshold. An important aspect was the discovery that the source of C_{30} hopane and C_{30} hopene were decoupled, which, unlike the steroid biomarkers, are therefore not diagenetically or catagenetically altered. In this regard, the hopene, diplotene, appears to function as a biomarkers derived directly from living bacteria, effectively outlining the zone of habitation for this system.

The chemometric results (**Chapter 3**), along with a survey of subtracted GC×GC chromatograms, measurements of hydrocarbon diversity, and statistical interpretation of the Chapter 2 thermal maturation parameters, indicate that whole-partial in situ pyrolysis likely begins at 8-10 cmbsf and 15-18 cmbsf in cores 1 and 2, at porewater temperatures exceeding ~110 °C. Compared to the EasyR_o Kinetic model, these depths are too shallow for in situ production, suggesting variable temperature flux or increased cracking efficiency in the subsurface. Ubiquitous staining of oil in the subsurface is also apparent, which shows

systematic, temperature controlled structural and stereochemical alteration across the transect. These culminate to decreasing diversity and abundance of alkylated compounds the production of dealkylated "parent" PAHs, steroids, and hopanoid biomarkers in progressively being hosted in more deeply buried, hotter sediments, as resolved by MPCA and other validation techniques.

Together, the results show marked spatial heterogeneity in the Cathedral Hill shallow sediments, along with thermally- and biologically-driven transformations within these sediments. Although no novel degradation pathways were found, the results support previously documented trends regarding the thermal maturity and oxidative weathering processes common to hydrothermally influenced, organic-rich sediments.

For future work, the chemometrics aspect of this study may be expanded upon and refined in a method-development sense. In the context of a more controlled environment (such as a conventional pyrolysis experiment), the various stages of pyrolysis/sediment maturation and resulting gradual changes to the multi-molecular hydrocarbon profile, may be captured without interference from oil staining or microbial activity. This may allow a chemometric model to be calibrated to these type of data, and demonstrably show clustering patterns and provide reference hydrocarbon profiles associated with increasing thermal maturity and in situ pyrolysis. An MPCA model with these controlled samples, alongside a subset of environmental oils with differing levels of maturation, would demonstrate the ability of this method to unambiguously bin samples based on hydrocarbon content/maturity. Also, if materials science provides a way to sample even hotter and deeper pushcore sediment (without risk of melting the pushcore core barrel), future studies could focus on providing broader insight into the subsurface environment of these systems, potentially capturing the stages of in situ catagenesis and metagenesis.

Appendix

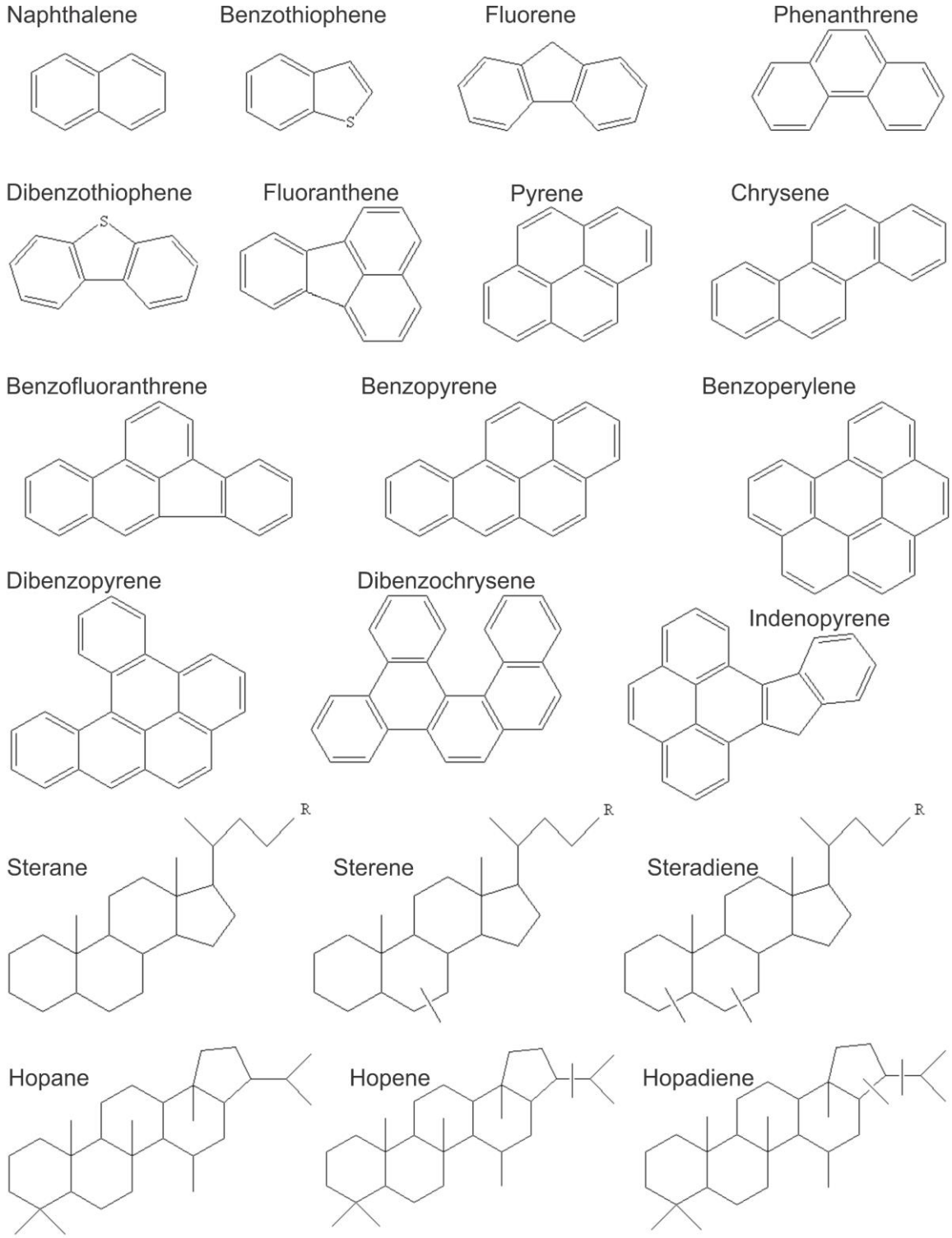


Fig A1: Chemical structures of aromatic hydrocarbons and biomarkers