1 Introduction

Lipid biomarkers are fossil biochemicals possessing recalcitrant organic skeletons that retain a significant proportion of the major structural features inherited from their biolipid or pigment precursors (Treibs, 1936; Eglinton et al., 1964; Peters et al., 2005; Kvenvolden, 2006). Depending on the thermal history of the host rocks, hopanes, steranes, and other hydrocarbon biomarkers can remain well preserved after hundreds of millions of years of burial in sedimentary rocks and in their associated petroleum products. Lipid biomarker studies have yielded valuable information pertaining to the balance of source organismal contributions and their metabolisms, the thermal maturity of the preserved organic matter (influenced by both the subsurface burial time and the temperature regime), and the paleoenvironmental conditions (of redox, salinity, etc.) in the water column and at the sediment surface that prevailed during sediment deposition (for reviews, see Summons & Lincoln, 2012; Briggs & Summons, 2014). Biomarker assemblages are still often useful even if the taxonomic affinities of all biomarker constituents cannot be precisely constrained. Lipid biomarkers were originally utilized by petroleum geochemists for molecular fingerprinting to facilitate oil-oil and oil-source rock correlations used in hydrocarbon exploration (Tissot & Welte, 1984; Peters et al., 2005). The temporal patterns can also provide unique insights into the evolution and diversification of life and modes of carbon cycling through geological time. Such molecular fossils are a valuable addition to the growing analytical repertoire used in geobiology.

Importantly, care and caution have to be exercised during the full procedures associated with molecular analysis, since biomarker studies may be compromised by high thermal maturity of some Precambrian organic matter and the possibility of organic contamination from petroleum-derived fluids and/or modern additives incorporated during sample processing and storage (e.g., Grosjean & Logan, 2007; Brocks, 2011; French et al., 2015). As such, judicious rock sample selection and proven analytical methodologies are critical for recovering genuine molecular signals and for recognizing contamination inputs.

2 Background

2.1 Sample Selection and Analytical Methodology: Checks and Pitfalls

Proper care must be taken in selecting appropriate sedimentary rock samples and cross-checking standard analytical protocols used to extract and detect biomarkers from ancient sedimentary rocks and oils. Self-consistency checks

for verifying biomarkers as genuine signals, rather than being compromised by contamination, need to be strictly implemented and practised. Methodologies for rock preparation, and biomarker recovery and analysis, have been continually assessed and refined through the years (e.g., French et al., 2015).

The single most important criterion when selecting ancient sedimentary rocks for organic geochemical analysis is undoubtedly the level of thermal maturity of the preserved organic matter within the host strata. Primary biomarker structures are not well preserved in overmature rocks (i.e., late "oil-window" maturity and beyond) due to thermal destruction and aromatization of the parent molecules. Thus, with a strong thermal overprint, we cannot accurately evaluate which biomarkers were ever original molecular constituents of the host sedimentary organic matter. Such an absence of evidence is an important distinction when interpreting a lack of biomarker signal in overmature rocks. Thus, the finding of no detectable polycyclic alkane biomarkers, such as hopanes or steranes, in Archean rocks (French et al., 2015) was expected and is primarily due to the high thermal maturity associated with their protracted burial history at high temperatures (>200°C). Similarly, late Neoproterozoic rocks from South China have provided an important fossil repository of paleoenvironmental and geobiological information for paleontology and biogeochemistry; but these are usually too thermally mature to yield reliable polycyclic biomarker profiles (Duda et al., 2014a; Ai et al., 2020; as supported by clay mineralogical assessment in Derkowski et al., 2013). Highly mature rocks should be generally avoided for detailed biomarker work (French et al., 2015). To date, thermally well-preserved rocks from the 1.64 Ga Barney Creek Formation (Brocks et al., 2005) and the more mature 1.73 Ga Wollogorang Formation (Vinnichenko et al., 2020), both from McArthur Basin in northern Australia, represent the oldest rock targets found that allow for acceptable preservation of the structural features of indigenous extractable alkane and aromatic hydrocarbon biomarkers. Maturity of the host rocks should be assessed regardless of their geologic age, prior to commencing analytical work, and reported for scrutiny in any publications. High maturity is not just a problem restricted to Precambrian rocks.

Established rock screening techniques have been used for decades in the field of petroleum geochemistry to constrain thermal maturity, including elemental analysis (to calculate atomic H/C ratios) or programmed pyrolysis parameters (e.g., Rock-Eval™ and HAWK). Such necessary screening measures are not always, unfortunately, implemented or reported in ancient biomarker investigations, and issues of contamination are particularly acute for Precambrian rocks. Programmed pyrolysis provides a standard petroleum source rock assessment of the thermal maturity of sedimentary organic matter from the Hydrogen Index

(HI), Production Index (PI), and Tmax parameters. Pyrolysis data can usually be applied to select the best candidates ("oil-window" mature samples) for detailed biomarker investigations (Peters et al., 2005). The extent of isomerization of any detectable hopane and sterane constituents, measured from abundance ratios as reliable molecular maturity parameters, should be self-consistent with the overall maturity of the host rocks gauged from these independent methods if the biomarkers are to be considered genuine endogenous constituents. Note, though, that for rocks with very low total organic carbon (TOC) contents ($\ll$1 wt.%) and/or for high maturity samples, the Tmax estimate can sometimes generate erroneous data due to a poorly defined (S2 peak) signal response for kerogen breakdown products. It is recommended to use at least two independent methods to gauge thermal maturity for such samples.

2.2 Lipid Biomarker Approaches: Bitumen and Kerogen

Even when the stage of thermal maturity of rocks and petroleum is ideal for biomarker preservation, i.e., prior to or within the oil window, contamination artifacts are still a concern. Conventional approaches for studying ancient biomarkers in thermally well-preserved rock targets make extensive use of the hydrocarbons (alkanes and aromatic compounds) and simple functionalized compounds (alcohols, carboxylic acids, etc. preserved at low thermal maturity only) in the solvent-extractable material obtained from rocks (termed bitumen). Most organic geochemical investigations of Proterozoic rocks have generally relied on careful solvent extraction, preferably from large blocks of pre-cleaned outcrop samples or drill core samples. The most common organic contaminants are usually mobile and extractable compounds residing in outer portions of rock samples from outcrop exposure or drilling fluids, i.e., already in the rock specimens. Overall, biomarker assemblages should be consistent with the source rock lithology, geological age, thermal maturity, and depositional setting of the host strata. Checking for any obvious evidence of suspicious concentration and/or compositional gradients between exterior and interior rock samples (reported as E/I ratios for different biomarker compounds) can help identify contamination problems in bitumen analysis (Brocks et al., 2011). Potentially deceptive E/I ratios are possible, though (Duda et al., 2014b), as mobile contaminants could have pervaded and homogenized within bulk rock specimens via fractures and pores, especially for thin and fissile shale beds. Sedimentary rocks with low TOC contents and high maturity samples are particularly prone to such effects, which is why the relevant bulk sample information (e.g., TOC contents and programed pyrolysis parameters) should be provided in publications.

Some workers have also used the solvent extracts released after acid demineralization of pre-extracted rock powders, usually termed *Bitumen II,* for direct comparison with the biomarker profiles generated by prior solvent extraction (*Bitumen I*, the most commonly used organic matter phase). This can be a useful extra analytical step to include for ancient biomarker analysis (Sherman et al., 2007; Nabbefeld et al., 2010; Duda et al., 2016), especially to highlight any suspicious biomarker compound features that may be attributable to contamination inputs arising from mobile organic species pervading into the host strata. If *Bitumen I* and *Bitumen II* show similar compositions, this strongly implies that they derive from the same pool of syngenetic organic matter. Although more strongly associated with the host rock than *Bitumen I*, the physicochemical location of where *Bitumen II* resides (within the pore structure of the kerogen matrix versus adsorbed/trapped within the mineral phase) cannot always be precisely constrained and the organic constituents are not as immobile as those linked into the kerogen phase.

Usually ignored are the significant quantities of well-preserved biomarker structures that are covalently bound within the insoluble macromolecular network (termed kerogen). Kerogen comprises the bulk of sedimentary organic matter (typically over 90% w/w) in ancient sedimentary rocks (see Figure 1), making it by far the largest pool of organic carbon found on Earth (Tissot & Welte, 1984; Peters et al., 2005). The insoluble and polymeric nature of kerogen, however, means that the bound biomarker pool can only be released by fragmentation in the laboratory using chemical or thermal degradation techniques. Since organic geochemical studies usually do not consider the kerogen phase, routine biomarker interpretations are commonly reliant on only the bitumen phase, which often represents a small fraction of the overall biomarker pool even from the earliest stages of diagenesis (Michaelis & Albrecht, 1979; Van Graas, 1986; Eglinton & Douglas, 1988; Love et al., 1995; Bishop et al., 1998; Koopmans et al., 1998; Murray et al., 1998; Farrimond et al., 2003; Lee et al., 2019). Previous studies (Love et al., 1998; Bowden et al., 2006; Reinhardt et al., 2018) have shown that significant compositional differences can be found between the *free* (extractable) hydrocarbon biomarkers and those within the *bound* phase linked within macromolecular fractions, consisting of kerogen, asphaltenes, and resins. Diagenetic processes can result in certain classes of lipids being partitioned between the bitumen and kerogen phases of source rocks in drastically different relative and absolute amounts (Rullkötter & Michaelis, 1990; Kohnen et al., 1991). Thus, the potential for problematic biomarker signal bias exists when only considering the extractable hydrocarbon fractions in the rock bitumen, especially for lower maturity rock samples. Additionally, the lipids protected by binding

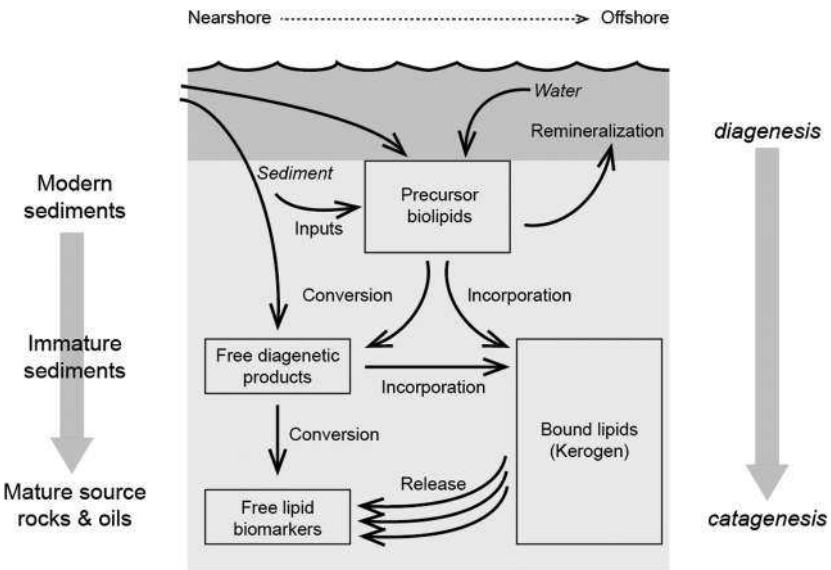

Figure 1 Simplified schematic representation of the free (extractable) and macromolecule-bound lipid biomarker pools that coexist in the geosphere in sedimentary organic matter. For ancient rocks, the insoluble kerogen is the main polymeric phase and typically comprises ca. 90% of total sedimentary organic matter, with the remainder termed the bitumen (extractable organic matter). A substantial portion of the total biomarker pool is sequestered by binding into kerogen, although kerogen is not routinely analyzed by organic geochemists since it is a challenging organic substrate to structurally characterize. Kerogen represents an abundant and informative repository of biogeochemical information and deserves more attention. The protracted catagenetic release of the bound biomarker pool, through covalent bond cleavage, persists through the oil window and is only complete at high thermal maturity.

within geomacromolecules have been shown to be significantly more resistant to oxic degradation during early diagenesis (Hoefs et al., 2002) and much less susceptible to long-term biodegradation (Cassani & Eglinton, 1986; Jones et al., 1988) than free lipids.

The development of continuous-flow catalytic hydropyrolysis (HyPy, using high pressure hydrogen gas and a molybdenum catalyst as a reaction medium) has provided an attractive route for rapid and reproducible recovery of biomarker compounds covalently linked into kerogen (Love et al., 1995, 2009; Bishop et al., 1998; Duda et al., 2016; Zumberge et al., 2019). Note, HyPy is a different analytical technique to hydrous pyrolysis (which involves heating kerogens or whole rocks with liquid water in a closed vessel over several days to

mimic petroleum generation; see Lewan, 1985), although both approaches can generate significant quantities of soluble products from ancient kerogens. Because they are chemically bound to a solid substrate, these kerogen-linked biomarker structures are immobile and therefore most assuredly genuine. Bound biomarkers consist of intact lipids plus diagenetic alteration products that retained reactive functional group(s) to enable covalent binding into the host matrix during proto-kerogen formation. Sequestration by covalent binding is advanced from the earliest stages of sedimentary diagenesis (Farrimond et al., 2003; Lee et al., 2019), and this bound biomarker pool is better thermally preserved and much less susceptible to contamination effects than the conventionally analyzed free hydrocarbons found in rock bitumens (solvent-extractable organic matter). The parallel analyses of free and bound biomarker distributions help to assess syngeneity of compounds and to ascribe these to a well-constrained stratigraphic position (Love et al., 2008, 2009; Duda et al., 2016; Zumberge et al., 2019). The combination of high product yields and excellent retention of original structural and stereochemical features in biomarker hydrocarbon products means that we can scrutinize the biomarker distributions contained within kerogen (Love et al., 1995, 1997, 2005, 2009; Bishop et al., 1998; Murray et al., 1998) with a high degree of confidence. The ability to exhaustively remove soluble organic contaminants from sediments using solvent treatment while leaving no problematic residue has been demonstrated previously for ancient rock cores contaminated with oil-based drilling muds (Murray et al., 1998).

3 Key Topics

3.1 The Protracted Ecological Expansion of Proterozoic Marine Eukaryotes

A transition to a world with elevated surface oxygenation and nutrients, capable of supporting a more complex and productive marine biosphere, likely occurred during the Neoproterozoic Era (1000–541 Ma). By the late Neoproterozoic, it is generally agreed that eukaryotes were an ecologically significant component as major primary producers in marine ecosystems (Knoll et al., 2007). Biomarker investigations have sought to better constrain the timing of diversification and expansion of eukaryotes typically using steroid records; tracking when eukaryotic organisms became widespread and abundant in the marine realm, including the initial rise to prominence of algae as major primary producers (e.g., Summons et al., 1988; Moldowan et al., 1990; Summons and Walter, 1990; Logan et al., 1995; Dutkiewicz et al., 2003; Brocks et al., 2005, 2016, 2017; McKirdy et al., 2006; Grosjean et al., 2009; Love et al., 2009; Kelly et al., 2011;

Blumenberg et al., 2012; Dutta et al., 2013; Pawlowska et al., 2013; Flannery and George, 2014; Luo et al., 2015; Duda et al., 2016, 2020; Hoshino et al., 2017; Stolper et al., 2017; Isson et al., 2018; Goryl et al., 2018; Pehr et al., 2018; Zumberge et al., 2018, 2019; Nguyen et al., 2019; van Meldegem et al., 2019a).

An outcome of these biomarker studies has been the recognition of an apparently long time lag between the first appearance of robust Proterozoic eukaryotic microfossils (ca. 1.8–1.6 Ga; see Lamb et al., 2009; Miao et al., 2019) and the ecological expansion of marine eukaryotes as inferred by the first occurrences and ubiquity of regular (4-desmethyl) steranes in marine rocks of ca. 820 Ma and younger, prior to the Sturtian glaciation (Brocks et al., 2017; Isson et al., 2018; van Maldegem, 2019a; Zumberge et al., 2019). Such a significant temporal change in biomarker patterns is consistent with the concept of a fundamental global marine ecological upheaval as Proterozoic oceans transitioned from supporting bacterially dominated ecosystems to communities rich in eukaryotic primary producers and heterotrophic protists. Pawlowska et al. (2013) suggested that low abundances of pre-820 Ma steranes could, alternatively, be attributed to a molecular taphonomic bias associated with efficient destruction of eukaryotic steroids in the surface layers of Proterozoic microbial benthic mats. This idea was termed the "mat-seal" hypothesis. Yet, lipid analyses of modern microbial mats typically show that abundant eukaryotic steroids survive diagenetic processes within the active mat surface layers and within their underlying sediments (Blumenberg et al., 2015; Shen et al., 2018; Lee et al., 2019). Thus, the mat-seal hypothesis seems an unlikely explanation for the striking dearth of regular steranes observed for most mid-Proterozoic rocks and oils (Figure 2). The current consensus then for the lack of regular sterane signal is a largely faithful representation of low eukaryotic abundance relative to bacteria in pre-820 Ma oceans.

3.1.1 Sterane versus Hopane Abundances: A Biomarker Proxy for Eukaryotic versus Bacterial Source Contributions to Sedimentary Organic Matter

The relative abundance of the major biomarker hydrocarbon constituents in rocks and oils permits assessment of major source biota inputs, environmental conditions at the time of deposition, and overall degree of thermal maturity due to the burial maturation history of the host strata. Molecular geochemical analyses of Proterozoic rocks and oils that have experienced a reasonably mild thermal history often contain a wide variety of linear, branched, and cyclic alkane compounds derived from diverse biological sources. A comprehensive

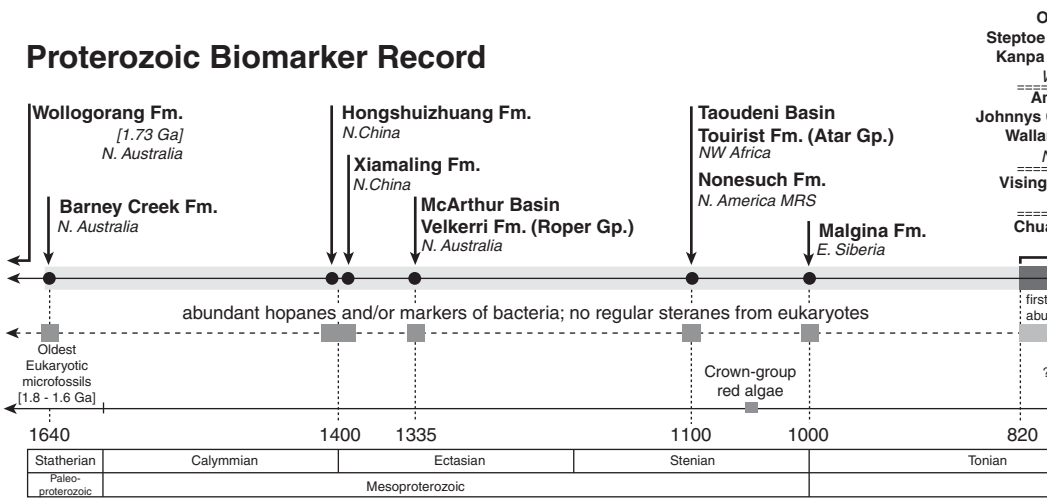

Figure 2 A revised biomarker timeline record compiled using the free biomarker hydrocarbons fr
matter (bitumen) of ancient marine mudstones and siltstones. Bacterial signatures generally do
Mesoproterozoic and other pre-820 Ma rocks and suggest very low eukaryotic contributions to
ecological expansion of eukaryotes in the marine realm was protracted until into the Neoproterozoic
the abundance patterns of regular (4-desmethyl) steranes. Regular steranes first appeared in the ro
(Brocks et al., 2017; Isson et al., 2018; Zumberge et al., 2019), with their increasing abundance relat
dominance suggestive of source inputs from eukaryotes, including red algae and unicellular hete
marine algae probably commenced during the Tonian Period (Zumberge et al., 2019), rather than the
A fundamental shift in marine sterane carbon number patterns favoring a C_{29} (stigmastane) domin
occurred during the interglacial Cryogenian Period (Love et al., 2009; Hoshino et al., 2017) and th
Period and through most of the Paleozoic Era (Grantham & Wakefield, 1988; Schwark & Empt,
Zumberge et al. (2019) and used with permission.

analysis of all the major compound peaks is a challenging task, and the molecular profiles can also be maturity dependent, so selected molecular ratio proxies are often used to measure key characteristics for the overall assemblage distributions. Biomarker records often compare the relative abundances of hopane and sterane biomarkers as a major proxy for tracking the respective contributions from bacteria and eukaryotes to the preserved organic matter. Such a measurement is also often inferred to broadly reflect the organismal balance of biota residing in the water column and the benthos, including the dominant primary producers at the base of the food chain (Grantham and Wakefield, 1988; Summons and Walter, 1990; Schwark and Empt, 2006; Knoll et al., 2007). Hopanoid and steroid lipids are synthesized in abundance by a broad and diverse range of bacteria and eukaryotes, respectively (Summons et al., 2006; Kodner et al., 2008).

No single proxy is ideal for providing such a broad measure of source inputs but there are sound practical reasons for measuring the relative abundances of steranes/hopanes (or, alternatively, hopanes/steranes). Petroleum geochemists have been carefully measuring these ratios for decades and have built up a comprehensive record of the average baseline values and ranges of this ratio through geological time. These records are much more complete for the Phanerozoic Eon, not surprisingly, and encompass a wide range of aquatic redox chemistry, depositional settings and maturities (Peters et al., 2005). These two lipid compound classes are usually amongst the most abundant polycyclic alkane biomarker constituents of ancient rocks and oils. Additionally, due to their stable polycyclic structures they are amongst the most recalcitrant of all lipid types, as reflected by their similar resistance to petroleum biodegradation. So, hopanes and steranes are thought to possess similar long-term preservation potential under various environmental conditions.

The diagenetic and catagenetic pathway of hopanoids and steroid transformation involves a complex and incompletely understood multistep process (Figure 1), proceeding via many intermediates and end products (Peters et al., 2005), rather than representing a simple precursor-product relationship. Still, the sterane/hopane (S/H) ratio is a convenient and useful way of obtaining a semi-quantitative measure of the relative abundance of eukaryotic to bacterial contributions to preserved sedimentary organic matter. It is not ideal, as some groups of bacteria do not biosynthesize hopanoids, but it appears to be an insightful ratio that can be readily and accurately measured. Alternative molecular proxies for gauging early eukaryotic abundance are associated with more problematic limitations. For example, gammacerane has been proposed as a eukaryotic marker for Neoproterozoic ciliates and other eukaryotic protists (van Maldegem et al.,

2019a), but it is usually only present in very low sedimentary abundance and can also be synthesized by various groups of bacteria (Banta et al., 2015, plus see references cited in Duda et al., 2020). S/H ratios also provide important context for appraising differences in biomarker assemblages between Proterozoic and younger rocks. Hopanoid-synthesizing cyanobacteria are important nitrogen-fixers in the modern ocean (Talbot et al., 2008; Saenz et al., 2012) and were also likely major ancient marine primary producers. In much the same way as for cyanobacteria, green algae were once the most abundant Neoproterozoic and Paleozoic eukaryotic marine phytoplankton (Grantham and Wakefield, 1988; Schwark and Empt, 2006) but became marginalized in the ocean by competition from derived red algal clades (dinoflagellates, diatoms, coccolithophores, etc.,) during the Mesozoic Era and now exhibit greater abundance and diversity within modern fresh waters.

Given the large and systematic shifts in the sedimentary S/H ratio from the mid-Proterozoic through to Ediacaran-age samples, this ratio has proved useful for tracking the ecological expansion of eukaryotes through the Proterozoic Eon. The abundance ratio of the major (C_{26}–C_{30}) steranes to (C_{27}–C_{35}) hopanes provides a direct and informative measure of the relative contributions of eukaryotes versus bacteria to sedimentary organic matter. The S/H ratios shift from values of approximately zero (<0.00) throughout the Paleoproterozoic and Mesoproterozoic (e.g., Brocks et al., 2005, 2016, 2017; Blumenberg, et al., 2012; Flannery and George, 2014; Luo et al., 2015; Isson et al., 2018; Nguyen et al., 2019) before extractable regular steranes become ubiquitous and abundant, to values reaching one-to-two orders of magnitude higher (up to ca. 0.4 for the bitumen and 1.3 for the kerogen-bound pool) in the mid-Neoproterozoic from ca. 820 Ma and younger (Brocks et al., 2017; Isson et al., 2018; van Maldegem, 2019a; Zumberge et al., 2019), prior to the Sturtian glaciation event.

3.1.2 Sterane Carbon Number Patterns: An Indicator of Eukaryotic Source Organisms

Eukaryotic sterane biomarkers first become detectable, abundant, and commonly found (Figure 2) in the late Tonian (ca. 820 Ma and younger) rock record (Brocks et al., 2017), as was verified recently from ca. 750–720 Ma Chuar and Visingsö Group biomarker assemblages for both the bitumen and the kerogen organic phases (Zumberge et al., 2019). The scarcity of sterane biomarkers in strata older than ca. 820 Ma is generally attributed to a low overall ecological abundance of eukaryotes (Brocks et al., 2017; Isson et al., 2018; Nguyen et al., 2019) in comparison with bacterial sources rather than a complete lack of these organisms in the marine realm. It is also possible though that low-oxygen-adapted