Lost world of complex life and the late rise of the eukaryotic crown

Jochen J Brocks, Benjamin J Nettersheim, Pierre Adam, Philippe Schaeffer, Amber J M Jarrett, Nur Güneli, Tharika Liyanage, Lennart M van Maldegem, Christian Hallmann, Janet M Hope

Lost world of complex life and the late rise of the eukaryotic crown

Jochen J. Brocks[1*], Benjamin J. Nettersheim[1,2*], Pierre Adam[3], Philippe Schaeffer[3], Amber J. M. Jarrett[1,4], Nur Güneli[1], Tharika Liyanage[1], Lennart M. van Maldegem[1], Christian Hallmann[5], Janet M. Hope[1]

[1] Research School of Earth Sciences, The Australian National University, Canberra, ACT 2601, Australia.

[2] MARUM – Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, 28359 Bremen, Germany

[3] Université de Strasbourg, CNRS, Institut de Chimie de Strasbourg UMR 7177, F-67000 Strasbourg, France

[4] Northern Territory Geological Survey, GPO Box 4550, Darwin NT 0801, Australia

[5] GFZ German Research Center for Geosciences, 14473 Potsdam, Germany

[*]These authors contributed equally to this work. e-mail: jochen.brocks@anu.edu.au; bnettersheim@marum.de

Keywords: stem-group eukaryotes, cycloartenol, lanostane, protosterols, eukaryote evolution, Proterozoic, Tonian, ursterols

Eukaryotic life appears to have flourished surprisingly late in the history of our planet. This view is based on the low diversity of diagnostic eukaryotic fossils in marine sediments of mid-Proterozoic age (~1,600 to 800 million years (Ma) ago) and an absence of steranes, the molecular fossils of eukaryotic membrane sterols[1,2]. This scarcity of eukaryotic remains is difficult to reconcile with molecular clocks that suggest that the last eukaryotic common ancestor (LECA) already emerged ~1200 to >1,800 Ma ago. LECA, in turn, must have been

preceded by stem-group eukaryotic forms by several hundred million years[3]. We report the discovery of abundant protosteroids in sedimentary rocks of mid-Proterozoic age. These primordial compounds remained unnoticed because their structures represent early intermediates of the modern sterol biosynthetic pathway, as famously predicted by Konrad Bloch[4]. The protosteroids reveal an ecologically prominent 'Protosterol Biota' that was widespread and abundant in aquatic environments from at least 1,640 to ~800 Ma ago and that likely comprised ancient protosterol-producing bacteria and deep-branching stem-group eukaryotes. Modern eukaryotes started to rise in the Tonian period (1,000 to 720 Ma), fuelled by the proliferation of red algae (rhodophytes) by ~800 Ma. This 'Tonian Transformation' emerges as one of the most profound ecological turning points in our planet's history.

MAIN TEXT

All living eukaryotes evolved from a last common ancestor (LECA) that lived ~1,200 to >1,800 Ma ago[3,5-7]. LECA and all its descendants form the crown of the eukaryotic tree including algae, plants, fungi, animals and all extant, unicellular protists. Yet, the domain Eukarya has a much deeper prehistory. The ancestral line leading towards LECA, and all its extinct side-branches, are stem-group Eukarya (Fig. 1b). The genome and cell structure of living descendants only provide limited insights into the evolution of LECA's ancestors[7], and almost nothing is known about their abundance, ecology and habitats. To study the hundreds of millions of years of hidden eukaryote evolution and ecology, we have to search for fossil and chemical remains directly in the geological record[8].

Body fossils with diagnostic eukaryotic features, such as processes, protrusions and cell-wall ornamentation, emerge in the rock record ~1,650 Ma ago and are found throughout the Mesoproterozoic (1,600 to 1,000 Ma) (Fig. 1c, specimens f-k). Yet, these fossils lack characteristics that would reliably place them within the eukaryotic crown, and so it is possible that they belonged to the stem[1,8,9]. The oldest fossils that can be assigned to the eukaryotic crown with confidence are the 1,050 Ma multicellular rhodophyte (red) alga *Bangiomorpha*[10] and 1,000 Ma chlorophyte (green) alga *Proterocladus*[11] (Fig. 1c, specimens c,d). However, body fossils of the crown only diversified and became abundant around 900 Ma with the emergence of fungi[12] followed by possible rhizarians and testate amoebozoans[13] (Fig. 1c, specimens a,b,e).

The current molecular fossil record

Biomarkers, the hydrocarbon fossils of biological lipids, provide complementary evidence on eukaryote evolution. Almost all eukaryotes produce tetracyclic sterols that regulate membrane physiology, while a wide range of bacteria synthesize the structurally related pentacyclic hopanepolyols. An increasing number of bacteria is also known to possess genes for sterol biosynthesis[14]. Sterols and hopanepolyols are relatively stable against degradation and may become preserved in sedimentary rocks as saturated steranes and hopanes as well as aromatic steroid and hopanoid derivatives (Fig. 2). The relative proportion of fossilized sterols over hopanepolyols is regarded as a first order estimate of the relative fluxes of eukaryotic and bacterial biomass to bottom sediments in ancient aquatic ecosystems[15,16]. The most common sterols possess cholestane (plotted in red in Fig. 1a), ergostane (blue) and stigmastane (green) carbon skeletons distinguished by the degree of alkylation of the side chain (Fig. 2). Such ‘crown-sterols’ occur across all major lineages of crown-group Eukarya, and LECA was presumably able to produce

sterols with the cholestane, ergostane, and possibly stigmastane skeleton[17]. The trajectory of the relative abundances of the three saturated crown-steranes in the geological record reflects trends of eukaryote evolution[18]. For example, increasing abundances of ergostane from the Mesozoic to Caenozoic reflect the rise of secondary endosymbiotic algae in the oceans[18], and the predominance of stigmastane in the Ediacaran and Paleozoic suggests that chlorophyte algae were the prevalent eukaryotic phototrophs[18,19]. But, stepping back in time into the Cryogenian and Tonian periods, the abundance of steranes relative to hopanes drops steeply, and steranes almost exclusively comprise cholestane, with traces of ergostane but no stigmastane (Fig. 3a).

The combined record of saturated steranes and body fossils of the past 800 Ma presents a coherent picture, with an initial increase in modern eukaryotic forms in the Tonian period, and a massive expansion in the Ediacaran. However, deeper in time the records become contradictory. While eukaryotic body fossils are found throughout the mid-Proterozoic interval, crown-steranes persistently remain below detection limits (Fig. 3a). The absence of these steranes has been explained by systematic preservation bias[20], prevalence of anaerobic eukaryotes that do not produce sterols[21-23], or ecosystems pervasively dominated by prokaryotes[2,22,24-27]. Yet, current ecological hypotheses struggle to explain the apparent scarcity of aerobic eukaryotes over a time interval of more than half a billion years and across all environments where biomarkers are detectable, including apparently nutrient-rich and oxygenated habitats (Supplementary Text)[26]. One hypothesis remains untested, the prospect that mid-Proterozoic ecosystems were dominated by eukaryotic stem-group forms that did not yet possess a full sterol biosynthetic pathway.

Protosterols and ursterols

Konrad Bloch, who deciphered the biosynthetic pathway of cholesterol, suggested that primordial eukaryotes may not have possessed the full sterol biosynthetic pathway but instead generated ancestral 'ursterols'[4]. The modern sterol biosynthetic pathway commences with the cyclization of oxidosqualene to one of two protosterols, lanosterol or cycloartenol (Fig. 2a). In energy and oxygen intensive downstream modifications, three methyl groups are then removed, double bonds modified, and new alkyl moieties added to the side chain. According to Bloch, each intermediate in this pathway, called an ursterol, was hundreds of millions of years ago a fully adapted end-product and an evolutionary improvement over its precursor. As LECA already possessed the capacity to produce crown-sterols, the ursterol pathway must have been assembled before LECA emerged. Cycloartenol in particular is a fully functional membrane molecule[28,29], and it may have been the sole sterol in deep-branching stem-group eukaryotes. Yet, the molecular fossils of protosterols and ursterols—saturated and aromatic protosteroids and ursteroids—remain largely unrecorded or unknown. In this study we apply new molecular search templates to detect these compounds in the deep geological record.

A new Proterozoic biomarker record

We obtained identification standards for fossil protosteroids and ursteroids through synthesis and by subjecting lanosterol and mixtures of cycloartenol and 24-methylene cycloartenol, to artificial maturation through pyrolysis. We then identified and quantified saturated and aromatic protosteroids in bitumens and oils of Caenozoic to late Palaeoproterozoic age using gas chromatography-mass spectrometry. All analyses were performed under strict exclusion of contamination (Methods). The oldest biomarker assemblage from the 1,640 Ma Barney Creek Formation revealed nearly 100 different protosterol derivatives, including lanostane, mono- and

diaromatic lanosteroids, and previously observed[30] triaromatic steroids. Also present in most bitumens and oils—albeit not in the pyrolysis products—was pentacyclic 'cyclosterane', tentatively interpreted as a geological transformation product of cycloartenol. Demethylated derivatives of some di- and triaromatic steroids may, in principle, also stem from cholesteroids or other downstream products of the sterol biosynthetic pathway. However, absence of detectable saturated counterparts such as cholestane and 4-methylcholestane, and absence of monoaromatic cholesteroids, makes such sources unlikely (Supplementary Text 2.8). Moreover, the presence of saturated lanostane shows that the observed assemblage is unlikely to result from uncommon diagenetic conditions that led to a quantitative aromatization of steroidal precursors. Thus, biologically demethylated sterols, if they existed, could have only been a minor component of the assemblage. Theoretically, the detected aromatic compounds may also derive from precursors of the euphol/tirucallol family, distinguished from protosterols by 13α,14β stereochemistry. However, the absence of 13α,14β-euphanes and presence of 13β,14α-lanostanes in saturated fractions excludes such sources (Extended Data Fig. 3b). Our laboratory experiments thus show that all fossilized steroids in the mid-Proterozoic record can be explained as degradation products of protosterols and 24-methyl ursterols (Fig. 2, Extended Data Fig. 1–9). Through the mid-Proterozoic interval, these protosteroids were detected in deep and relatively shallow water environments, microbial mats and pelagic habitats, shales and carbonates, as well as marine and likely lacustrine basins (Supplementary Text).

In all investigated bitumens, proto- and ursterols are overwhelmingly preserved as aromatic products, and Figure 3b summarizes the relative abundances of aromatic steroids and hopanoids through time. This aromatic biomarker record provides a complementary view to the trends of saturated steranes and hopanes described above (Fig. 3a), revealing unknown aspects of eukaryote

136 evolution. In the Palaeoproterozoic, the high relative abundances of aromatic protosteroids (shades
137 of purple) relative to hopanoids is striking, with proportions of steroids approaching Phanerozoic
138 levels. High proportions of protosteroids are observed to the end of the Mesoproterozoic, with the
139 additional appearance of side-chain methylated triaromatic ursteroids at ~1,300 Ma (cyan in Fig.
140 1a, 3b).

141 The new biomarker record thus reveals a 'Protosterol Biota' that was widespread and highly
142 abundant in the mid-Proterozoic (Fig. 1). Despite this profusion of protosteroids, the same mid-
143 Proterozoic formations failed to yield even traces of cholestane, ergostane and stigmastane,
144 bringing the absence of crown-steranes before 800 Ma into sharp focus (Fig. 3a). Our results show
145 that the lack of crown-steranes can no longer be dismissed as preservation artefact[20] but must
146 reflect extreme scarcity or absence of crown-group eukaryotes in open water habitat throughout
147 this interval.

148 The first crown-steroids emerge ~800 Ma ago (Fig. 1a). The record of saturated steranes in the
149 Tonian, Cryogenian and earliest Ediacaran is characterized by dominant cholestane, with sub-
150 percentage levels of ergostane in some samples, but a lack of stigmastane (Fig. 3a). However, in
151 contrast to earlier studies that exclusively focused on this saturated sterane record[15,16,19], the
152 aromatic hydrocarbon compilation reveals that steroids are far more abundant in the Tonian period
153 than previously appreciated (Fig. 3b). Moreover, ergosteroids constitute a one to two order of
154 magnitude larger proportion of total steroids, comprising up to 21% of monoaromatic steroids
155 (mean ± stdv: 5.9 ± 6.5%; n=16) and 33% of triaromatic steroids (6.7% ± 11.6%, n=26) (Fig. 1a).
156 In modern eukaryotes, this steroid homolog abundance pattern is found in rhodophytes (red algae)
157 that have a cholesteroid predominance of up to 100%, but typically contain ~5% ergosteroids[15].

158 The data thus substantiate previous suggestions[16,31,32] of an early expansion of rhodophytes and
159 suggest that algal primary production already started to rise in the Tonian period (Fig 3c). The
160 emergence of stigmasteroids after 635 Ma then heralds a second revolution at the base of the food
161 web, the rise of chlorophyte algae[15,19]. In the Phanerozoic, the average abundance of aromatic
162 protosteroids relative to aromatic crown-steroids drops to ~3.8 ± 4.0% (n=36), a proportion that
163 can be attributed to biosynthetic intermediates produced by crown-group eukaryotes plus unknown
164 contributions from bacteria (Supplementary Text).

165 **The origins of ancient protosteroids**

166 Apart from Eukarya, 94 bacterial taxa across twelve phyla are currently known to possess the genes
167 SQMO (squalene monoxygenase) and OSC (oxidosqualene cyclase) required for protosterol
168 biosynthesis[33]. The bacterial and eukaryotic genes have a common ancestry in the
169 Palaeoproterozoic and were horizontally transferred between the domains. By the mid-Proterozoic,
170 stem-group eukaryotes and bacteria both likely had the capacity to produce sterols[34].

171 The importance of sterol producing bacteria in the mid-Proterozoic is difficult to assess. Based on
172 the shallow clade depths of SQMO and OSC and their sporadic phylogenetic distribution in extant
173 Bacteria[34,35], most taxa must have acquired sterol biosynthesis genes relatively recently via
174 horizontal gene transfer (Supplementary Text 8.2, Supplementary Fig. 1 and 2). The roots of
175 SQMO and OSC are deepest among the orders Methylococcales and Myxococcales[33], and these
176 groups are thus the most likely sources of bacterial sterols in the Proterozoic. Yet, most extant
177 Methylococcales and deep-branching (Group I) Myxococcales mainly produce 4-
178 methylcholesteroids and cholesteroids, respectively[35]. Since the corresponding saturated
179 hydrocarbons, 4-methylcholestane and cholestane, were below detection limits in all samples >800

Ma, crown-group Methylococcales and Group I Myxococcales were not likely major components of the studied environments. However, contributions to the Protosterol Biota may have come from stem-representatives of these two bacterial orders, or from unknown or extinct bacterial groups. The importance of such contributions depends on the general abundance of sterol-producing bacteria in aquatic environments. While bacterial SQMO and OSC are detected in a wide variety of modern habitats[35], bacterial sterols themselves have only been reported from less common environments such as hydrothermal springs and methane seeps (Supplementary Text). The mid-Proterozoic likely witnessed elevated methane generation in marine sediments compared to the present, providing potential substrate for Methylococcales[30]. However, the carbon isotopic signature of triaromatic steroids from the 1,640 Ma Barney Creek Fm ($\delta^{13}C$ = -19 to -22‰)[36] is inconsistent with a methanotrophic origin and suggests heterotrophic[36] or phototrophic sources. Generally, however, bacterial sterol contributions to mid-Proterozoic environments are plausible. A test for such contributions does not exist at present, and future research needs to evaluate the importance and antiquity of sterol producing bacteria in marine environments. A solely bacterial origin of the protosteroids would open the fascinating proposition that aerobic crown- as well as stem-group eukaryotes remained extremely scarce for hundreds of millions of years after their emergence.

Stem-group eukaryotic origins of protosteroids, on the other hand, can reconcile the continuous record of eukaryote-like fossils with the scarcity of crown-steroids throughout the mid-Proterozoic interval and may explain why all but two Mesoproterozoic species lack diagnostic crown-group features. Early eukaryotic fossils already possessed complex cell-wall ornamentation and symmetrically- or asymmetrically-arranged, cylindrical and occasionally branched processes, structures that probably required a flexible endomembrane system[37] with active curvature control,

a capability aided by sterols in extant Eukarya[38]. Moreover, large but simple fossils interpreted as possibly eukaryotic, such as leiosphaerids, are widespread in mid-Proterozoic sediments[37] and may have been a prominent source of protosteroids. A stem-group eukaryotic origin of protosteroids is also supported by the detection of 24-methylated triaromatic steroids in Mesoproterozoic strata that may derive from ursterols such as 24-methylenecycloartenol, obtusifoliol or 4-methylfecosterol. Some bacteria possess SMT (sterol methyltransferase), a gene imparting the hypothetical ability to methylate the side chain at C-24. However, most SMT gene trees[33,39] do not support a deep origin of side-chain methylation in bacteria, and side-chain methylated sterols have never been detected in bacterial lipid extracts[35]. By contrast, SMT was already present in LECA[17,33,40]. If a eukaryotic interpretation is correct, then the side-chain methylated ursteroids may represent an evolutionary intermediate in the long biosynthetic pathway from cycloartenol to ergosterol, assembled in the stem-lineage toward LECA (Fig. 1b, 2a). In ~1,300 Ma old sediments these side-chain methylated steroids comprise ~20% of total steroids, pointing to notable abundances of such intermediate organisms. In summary, we posit that the Protosterol Biota partially or predominantly represents early stages of eukaryote evolution that did not yet possess a complete sterol biosynthetic pathway.

The late rise of the eukaryotic crown

The paucity of fossils of crown-group eukaryotes before ~1,050 Ma and of crown-steranes before ~800 Ma remains a conundrum in early-life research. Crown-group eukaryotes likely emerged between ~1,200 to >1,800 Ma[3,5,6] and rapidly diverged into major branches[7]. Based on all but the youngest molecular clock estimates, the crown thus reached considerable morphological and ecological diversity before 1,000 Ma, including sexually reproducing, multicellular algae[10], and

protists capable of endocytosis, phagocytosis and locomotion[41]. Based on the predicted taxonomic and ecological diversity of Mesoproterozoic crown-group eukaryotes, the scarcity of diagnostic fossils and biomarkers seems difficult to explain.

The discovery of the Protosterol Biota now suggests that the ecological success of early crown-group eukaryotes may have been shaped by competition with early-diverging stem-group relatives over several hundred million years of their co-existence (Fig. 1b and ref[23]). As all stem-group eukaryotes are now extinct, we remain oblivious to differences between crown-group and stem-group eukaryotes that co-existed in the mid-Proterozoic. However, the fossil protosteroids indicate that the composition of cell membranes was one of the physiological distinctions. The biosynthesis of crown-sterols comes at immense metabolic cost and must have conferred distinct advantages to the first eukaryotes adopting this pathway. Based on experiments on yeast, specific unsaturation patterns of crown-sterols may protect cells against osmotic shock during desiccation and rehydration[42]. Similarly, addition of an ethyl group to the sterol side chain adapts plant cells to cold shock and heat as the ethyl group appears to enhance cohesion with other lipids, providing better membrane ordering over large temperature scales[29,43]. Moreover, through branch-points in the sterol biosynthetic pathway, crown-group eukaryotes can regulate the proportion of different sterols. This capacity further extends the temperature range of membrane-associated processes[29,43], may play a role in membrane repair during UV-B exposure[43] and promotes asymmetric cell growth[29]. Potentially important but more speculative is the role of down-stream modified sterols in the protection of cell membranes against attack by membranolytic toxins[44-46] and for the response of cells to changing O_2 levels[43,47]. These sterol-based mechanisms to adapt to physical and chemical extremes were absent in organisms that solely produce protosterols. We thus posit that crown-sterols enabled LECA's ancestors to radiate into marginal niches such as agitated

248 shoreline facies, periodically exposed mudflats, braided rivers and subaerial habitat, environments
249 that experience cycles of desiccation and rehydration, diurnal extremes of high and low
250 temperatures, and elevated UV-radiation. If correct, then early crown-group representatives were
251 eukaryotic extremophiles of their time.

252 Biomarkers are rarely preserved in sediments deposited in well-oxygenated, high-energy
253 environments, and an origin of LECA's ancestors in such habitats may explain why crown-steranes
254 have remained elusive in mid-Proterozoic sediments. This view is supported by the discovery of
255 the oldest accepted crown-group fossils, rhodophytes and chlorophytes, in shallow water
256 deposits[10,11] (Fig. 1c).

257 The new biomarker record suggests that crown-sterol producers were not able to break into niches
258 inhabited by the Protosterol Biota in benthic and pelagic open-water settings. During the mid-
259 Proterozoic, atmospheric oxygen levels were still one to three orders of magnitude below the
260 present[48,49] and surface waters may have become anoxic during the dark hours or experienced
261 incursions of sulphidic waters, toxic to crown-group Eukarya[24]. Stem-group eukaryotes may have
262 possessed traits that allowed them to thrive in such conditions, and protosterol biosynthesis may
263 have provided an advantage due to its lower oxygen and energy requirements compared to
264 downstream sterols. Overall, the biomarker record is consistent with a facultatively anaerobic
265 origin of eukaryotes where the stem lineage moved from anoxic to increasingly oxygenated
266 environments in the course of the mid-Proterozoic[50,51].

267 The Tonian period eventually saw a rise in atmospheric oxygen levels, increased nutrient supply
268 to the oceans and changing abundances of bioessential trace elements[48,52]. The ecological, physical
269 and chemical changes that caused, or permitted, the expansion of crown-group eukaryotes into

open marine environments remain unknown, but a preference for more oxygenated habitats and their crown-sterol supported membranes may have provided crucial pre-adaptations for the changing conditions of the Tonian. Exactly when the last branches of the eukaryotic stem became extinct remains unknown, but competition from expanding crown-group eukaryotes may have contributed to their demise. Their protosterol containing membranes may also have proved maladapted to the cold of the Cryogenian Snowball Earth glaciations and the extreme global heat in the aftermath[53].

Konrad Bloch's Ursterols

In the essay 'Evolutionary Perfection of a Small Molecule'[4], Konrad Bloch suggested that the discovery of fossil 'ursterols' would strengthen the idea of molecular evolution, the notion that modern intermediates were once fully functional end products in ancient biosynthetic pathways. Bloch was sceptical that fossilized proto- and ursterols could be discovered, yet we now find that they are abundantly preserved in the oldest biomarker-bearing rocks. The detected compounds represent at least three enzymatic steps (OSC, SQMO, SMT) in the biosynthetic pathway towards modern eukaryotic ergosteroids. We posit that these fossil proto- and ursteroids are witnesses of a lost world of ancient stem-group eukaryotes that were widespread and possibly abundant during Earth's middle age. Our crown-group eukaryotic ancestors may have co-existed with these primordial organisms for several hundred million years, unable to occupy their niches. With rising atmospheric oxygen levels and changing marine chemistry, crown-group Eukarya started to expand 1,000 to 800 Ma ago. This 'Tonian Transformation' is recorded in the body and molecular fossil records and may be one of the most profound ecological transitions in the evolution of complex life.

MAIN REFERENCES

1 Butterfield, N. J. Early evolution of the Eukaryota. *Palaeontology* **58**, 5-17, doi:10.1111/pala.12139 (2015).

2 Gueneli, N. *et al.* 1.1-billion-year-old porphyrins establish a marine ecosystem dominated by bacterial primary producers. *Proceedings of the National Academy of Sciences* **115**, E6978-E6986, doi:10.1073/pnas.1803866115 (2018).

3 Betts, H. C. *et al.* Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nature Ecology & Evolution* **2** 1556-1562, doi:10.1038/s41559-018-0644-x (2018).

4 Bloch, K. in *Blondes in Venetian Paintings, the Nine-Banded Armadillo, and Other Essays in Biochemistry*, pages 14-36 (Yale University Press, 1994).

5 Eme, L., Sharpe, S. C., Brown, M. W. & Roger, A. J. On the age of eukaryotes: Evaluating evidence from fossils and molecular clocks. *Cold Spring Harbor Perspectives in Biology* **6**, doi:10.1101/cshperspect.a016139 (2014).

6 Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences* **108**, 13624-13629, doi:10.1073/pnas.1110633108 (2011).

7 Chernikova, D., Motamedi, S., Csuros, M., Koonin, E. & Rogozin, I. A late origin of the extant eukaryotic diversity: divergence time estimates using rare genomic changes. *Biology Direct* **6**, 26 (2011).

8 Knoll, A. H. Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb Perspect Biol* **6**, a016121 (2014).

9 Javaux, E. & Knoll, A. Micropaleontology of the lower Mesoproterozoic Roper Group, Australia, and implications for early eukaryotic evolution. *Journal of Palaeontology* **91**, 199-229, doi:10.1017/jpa.2016.124 (2017).

10 Butterfield, N. J. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386-404 (2000).

11 Tang, Q., Pang, K., Yuan, X. & Xiao, S. A one-billion-year-old multicellular chlorophyte. *Nature Ecology & Evolution* **4**, 543-549, doi:10.1038/s41559-020-1122-9 (2020).

12 Loron, C. C. *et al.* Early fungi from the Proterozoic era in Arctic Canada. *Nature* **570**, 232-235, doi:10.1038/s41586-019-1217-0 (2019).

13 Porter, S. M. & Knoll, H. Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. *Paleobiology* **26**, 360-385 (2000).

14 Welander, P. V. Deciphering the evolutionary history of microbial cyclic triterpenoids. *Free Radical Biology and Medicine* **140**, 270-278, doi:doi.org/10.1016/j.freeradbiomed.2019.05.002 (2019).

15 Brocks, J. J. *et al.* The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578–581, doi:10.1038/nature23457 (2017).

16 Zumberge, J. A., Rocher, D. & Love, G. D. Free and kerogen-bound biomarkers from late Tonian sedimentary rocks record abundant eukaryotes in mid-Neoproterozoic marine communities. *Geobiology* **n/a**, doi:10.1111/gbi.12378 (2019).

17 Desmond, E. & Gribaldo, S. Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. *Genome Biology and Evolution* **1**, 364-381, doi:10.1093/gbe/evp036 (2009).

18 Grantham, P. J. & Wakefield, L. L. Variations in the sterane carbon number distributions of marine source rock derived crude oils through geological time. *Organic Geochemistry* **12**, 61-73 (1988).

19 Hoshino, Y. *et al.* Cryogenian evolution of stigmasteroid biosynthesis. *Science Advances* **3**, doi:10.1126/sciadv.1700887 (2017).

20 Pawlowska, M. M., Butterfield, N. J. & Brocks, J. J. Lipid taphonomy in the Proterozoic and the effect of microbial mats on biomarker preservation. *Geology* **41**, 103-106 (2013).

21 Porter, S. M., Agić, H. & Riedman, L. A. Anoxic ecosystems and early eukaryotes. *Emerging Topics in Life Sciences* **2**, 299-309, doi:10.1042/etls20170162 (2018).

22 Nguyen, K. *et al.* Absence of biomarker evidence for early eukaryotic life from the Mesoproterozoic Roper Group: Searching across a marine redox gradient in mid-Proterozoic habitability. *Geobiology* **17**, 247-260, doi:10.1111/gbi.12329 (2019).
23 Porter, S. M. Insights into eukaryogenesis from the fossil record. *Interface Focus* **10**, 20190105, doi:doi:10.1098/rsfs.2019.0105 (2020).
24 Anbar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* **297**, 1137-1142 (2002).
25 Butterfield, N. J. Oxygen, animals and oceanic ventilation: an alternative view *Geobiology* **7**, 1-7 (2009).
26 Brocks, J. J. The transition from a cyanobacterial to algal world and the emergence of animals. *Emerging Topics in Life Sciences* **2**, 181-190, doi:10.1042/etls20180039 (2018).
27 Jarrett, A. J. M. *et al.* Microbial assemblage and paleoenvironmental reconstruction of the 1.3 Ga Velkerri Formation, McArthur Basin, northern Australia. *Geobiology* **17**, 360-380, doi:10.1111/gbi.12331 (2019).
28 Bloch, K. E. Sterol structure and membrane function. *CRC Crit. Rev. Biochem.* **14**, 47–92 (1983).
29 Dufourc, E. J. Sterols and membrane dynamics. *Journal of Chemical Biology* **1**, 63-77, doi:10.1007/s12154-008-0010-6 (2008).
30 Brocks, J. J. *et al.* Biomarker evidence for green and purple sulphur bacteria in a stratified Paleoproterozoic sea. *Nature* **437**, 866-870 (2005).
31 Summons, R. E. *et al.* Distinctive hydrocarbon biomarkers from fossiliferous sediments of the Late Proterozoic Walcott Member, Chuar Group, Grand Canyon, Arizona. *Geochimica et Cosmochimica Acta* **52**, 2625-2637 (1988).
32 van Maldegem, L. M. *et al.* Geological alteration of Precambrian steroids mimics early animal signatures. *Nature Ecology & Evolution* **5**, 169-173, doi:10.1038/s41559-020-01336-5 (2021).
33 Hoshino, Y. & Gaucher, E. A. Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. *Proceedings of the National Academy of Sciences* **118**, doi:doi.org/10.1073/pnas.2101276118 (2021).

34 Gold, D. A., Caron, A., Fournier, G. P. & Summons, R. E. Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* **543**, 420-423, doi:10.1038/nature21412 (2017).

35 Wei, J. H., Yin, X. & Welander, P. V. Sterol synthesis in diverse bacteria. *Front Microbiol* **7**, 990-990, doi:10.3389/fmicb.2016.00990 (2016).

36 Zhang, X., Paoletti, M., Izon, G., Fournier, G. & Summons, R. Isotopic evidence of photoheterotrophy in Palaeoproterozoic Chlorobi. *PREPRINT (Version 1) available at Research Square*, doi:doi.org/10.21203/rs.3.rs-2444442/v1 (21 February 2023).

37 Knoll, A. H., Javaux, E., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361** (2006).

38 Anderson, R. H. *et al.* Sterols lower energetic barriers of membrane bending and fission necessary for efficient clathrin-mediated endocytosis. *Cell Reports* **37**, 110008, doi:doi.org/10.1016/j.celrep.2021.110008 (2021).

39 Michellod, D. *et al. De novo* phytosterol synthesis in animals. *Science*, in press (2023).

40 Gold, D. A. The slow rise of complex life as revealed through biomarker genetics. *Emerging Topics in Life Sciences* **2**, 191-199, doi:10.1042/etls20170150 (2018).

41 Koumandou, V. L. *et al.* Molecular paleontology and complexity in the last eukaryotic common ancestor. *Critical Reviews in Biochemistry and Molecular Biology* **48**, 373-396 (2013).

42 Dupont, S., Beney, L., Ferreira, T. & Gervais, P. Nature of sterols affects plasma membrane behavior and yeast survival during dehydration. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 1520-1528, doi:dx.doi.org/10.1016/j.bbamem.2010.11.012 (2011).

43 Rogowska, A. & Szakiel, A. The role of sterols in plant response to abiotic stress. *Phytochemistry Reviews*, doi:10.1007/s11101-020-09708-2 (2020).

44 Santalova, E. A. *et al.* Sterols from six marine sponges. *Biochemical Systematics and Ecology* **32**, 153 (2004).

45 Tillmann, U. Kill and eat your predator: a winning strategy of the planktonic flagellate *Prymnesium parvum*. *Aquatic Microbial Ecology* **32**, 73-84 (2003).

46 Brocks, J. J. *et al.* Early sponges and toxic protists: possible sources of cryostane, an age diagnostic biomarker antedating Sturtian Snowball Earth. *Geobiology* **14**, 129-149, doi:10.1111/gbi.12165 (2016).

47 Galea, A. M. & Brown, A. J. Special relationship between sterols and oxygen: were sterols an adaptation to aerobic life? *Free Radical Biology and Medicine* **47**, 880 (2009).

48 Canfield, D. E. *Oxygen - A four billion year history*. (Princeton University Press, 2014).

49 Planavsky, N. J. *et al.* Low Mid-Proterozoic atmospheric oxygen levels and the delayed rise of animals. *Science* **346**, 635-638, doi:10.1126/science.1258410 (2014).

50 Mentel, M. & Martin, W. Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 2717-2729, doi:doi:10.1098/rstb.2008.0031 (2008).

51 Mills, D. B. *et al.* Eukaryogenesis and oxygen in Earth history. *Nature Ecology & Evolution* **6**, 520–532, doi:doi.org/10.1038/s41559-022-01733-y (2022).

52 Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth/'s early ocean and atmosphere. *Nature* **506**, 307-315, doi:10.1038/nature13068 (2014).

53 Hoffman, P. F. *et al.* Snowball Earth climate dynamics and Cryogenian geology-geobiology. *Science Advances* **3**, doi:10.1126/sciadv.1600983 (2017).

FIGURE LEGENDS

Fig. 1 | Geological time chart (in billion years, Ga) comparing the molecular fossil, microfossil and phylogenetic records of early eukaryote evolution. a, Relative abundances of aromatic protosteroids (purple and cyan tones) and crown-steroids (reds, blues and greens), highlighting the transition from a Protosterol Biota to a 'Crown-sterol Biota' in the Neoproterozoic. Each horizontal colour bar represents one sample, and grey triangles assign data bundles to geological units 1 to 11 (key provided in Methods). prot = protosteroids, meprot = 24-methyl protosteroids (ursteroids), chol = cholesteroids, erg = ergosteroids, stig = stigmasteroids. Key to colours see panel Figure 2 and Supplementary Table 1. For details on data assembly and geological formations (1–11) see Methods. **b,** Phylogenetic tree of the domain Eukarya with white, red and green highlighting crown-group branches. Stem-group branches (purple) are hypothetical only, illustrating the notion that mid-Proterozoic ecosystems may have been dominated by extinct stem forms that did not produce crown-sterols. LECA, the last common ancestor of all extant eukaryotes, may have emerged between 1.2 and > 2.0 Ga (see text). **c,** Microfossils of early eukaryotes. (*a*–*e*) Likely crown-group Eukarya (~1.1 to 0.7 Ga), (f–k) microfossils that are possibly or certainly eukaryotic but lack diagnostic crown-group characteristics. See Methods for detailed information and image credits. Tn, Cr, Ed are the Tonian, Cryogenian and Ediacaran periods; P, M, C are the Palaeo-, Meso- and Caenozoic eras.

Fig. 2 | Juxtaposition of biological and geological hopanoids and steroids, and the evolution of the sterol biosynthetic pathway through time. a, The pathway from squalene epoxide to protosterols, ursterols and crown-sterols was likely assembled during the Proterozoic. '2x' indicates number of successive enzymatic steps. **b,** Fossil steroids detected in the rock record that

correspond to intermediates and end-products in the evolving biosynthetic pathway. TAS = triaromatic steroids, MAL = monoaromatic lanosteroids, DAL = diaromatic lanosteroids, 'x,4-Me_2TAS' signifies a TAS methylated at C-4 and at an unknown position x. '?' indicates that cyclosterane likely contains a bridge at an unknown position.

Fig. 3 | The succession of fossil steroids and bacterial hopanoids through time. a, The relative abundances of saturated steranes (red, blue and green) and bacterial hopanes (hop, yellow). Note absence of detectable steranes before 800 Ma, highlighting the paucity of crown-group eukaryotes. **b,** A new view of biomarker evolution based on the relative abundances of aromatic protosteroids (purple), 24-methyl protosteroids (ursteroids, cyan), aromatic crown-steroids (reds, blues and greens) and aromatic hopanoids (yellow). The data reveal the existence of a protosterol-producing biota that was ecological dominant in the Proterozoic. For (a) and (b), each horizontal colour bar represents one sample. Key to colours, formations and abbreviations see Fig. 1. **c,** The geological succession of dominant primary producers (PP) based on molecular fossils.

METHODS

Assembly of contamination-free biomarker data. In biomarker research, identification of any traces of hydrocarbon contamination is essential. The problem of contamination in Precambrian biomarker research was discussed previously[15,54]. In the present study, tests for hydrocarbon syngeneity comprised cumulative laboratory system blanks and the quantitative comparison of hydrocarbon contaminants on rock surfaces and rock interiors. The removal and quantification of contaminants is based on well-established protocols[55-58]. Briefly, the laboratory system blanks captured all procedures from rock crushing and powder preparation, extraction of bitumens with solvents, column chromatography and instrumental analyses. Biomarker data was generally discarded if blank levels of individual compounds reached 10% of the concentration in sample extracts. Contaminants introduced into rock samples before analysis, for instance during drilling and storage, were identified and eliminated using Exterior/Interior (E/I) experiments. In these experiments, younger contaminant hydrocarbons are recognized by removing exterior rock surfaces[55], followed by separate quantification of individual compounds in the exterior portions of the rock ('E') and its interior ('I'). Indigenous hydrocarbons commonly yield E/I ~ 1, and surficial contaminants are recognized by E/I >> 1 and by E/I values that increase with rising carbon number within homologous series, an effect caused by contaminant diffusion from rock surfaces into the interior[55,56,58,59]. For Precambrian biomarker data used to assemble Figures 1 and 3 and summarized in Supplementary Tables 1 and 2, biomarker syngeneity was established using criteria given in the references above, whereas samples that fulfilled the contamination criteria were not included in our study.

Rock sample preparation. Provenance and lithology of all pre-Ediacaran rock samples are summarized in Extended Data Table 1, and for all other samples in Supplementary Table 3. In

Exterior/Interior experiments (see above), rock surfaces were either trimmed using a solvent-cleaned diamond wafering saw[56,59] (Buehler Isomet™ 1000; Illinois, U.S.A) or, for fissile mudstones and shales, surface ablation using a modified KG-1 Mini-Sonic tumbler[55] (Diamond Pacific, USA). Exterior and interior rock fragments were ground to a < 240 mesh powder using a Rocklabs iron puck mill. The mill was cleaned between samples by rinsing with methanol (MeOH) and dichloromethane (DCM), and by grinding annealed (600°C, 9 h) quartz sand.

Extraction and fractionation of bitumens. Bitumens were extracted from rock powders using a Dionex Accelerated Solvent Extractor (ASE 200) using 100% DCM or a mixture of 90% DCM and 10% MeOH. Solvent was reduced to ~100 µl under a stream of purified nitrogen gas. Rock extracts and petroleum samples were fractionated into saturated, aromatic, and polar fractions using micro column chromatography over annealed and dry packed silica gel. Saturate hydrocarbons were eluted with 0.5 dead volumes (DV) of *n*-hexane, aromatic hydrocarbons with 2 DV of *n*-hexane : DCM (1:1 or 4:1 v/v) and polars with 3 DV DCM : MeOH (1:1 v/v). As internal standard for full scan and selected iron recording (SIR) experiments, 1 to 2 µg 18-MEAME (18-methyl-eicosanoic acid methyl-ester; Chiron Laboratories AS) were added to the saturated and aromatic fractions and/or 50 ng D10 (Pyrene-d_{10}, Chiron) to the aromatic fraction. For metastable reaction monitoring (MRM) experiments, 25 to 50 ng of D4 (d4-C_{29}-ααα-ethylcholestane; Chiron Laboratories AS) were added to the saturated hydrocarbon fraction.

Characterization of protosteroids. Saturated and aromatic protosteroids were identified by their mass spectra, and co-injection and co-elution experiments with authentic standards and/or pyrolysis products of lanosterol, cycloartenol and 24-methylenecycloartenol (Supplementary Methods). Structures, mass chromatograms and mass spectra are given in Extended Data Figures

509 1 to 9. ^{1}H and ^{13}C-NMR chemical shifts of synthetic monoaromatic lanosteroid (20R MAL) are
510 provided in Supplementary Table 4.

511 **Gas chromatography-mass spectrometry (GC-MS).** Steroids and hopanoids were evaluated and
512 quantified using GC-MS analyses on an Agilent 6890 GC coupled to a Micromass Autospec
513 Premier double sector MS (Waters Corporation, Milford, MA, USA). The GC was equipped with
514 a 60 m DB-5 capillary column (0.25 mm i.d., 0.25 μm film thickness; Agilent J&W Scientific,
515 Agilent Technologies, Santa Clara, CA, USA), and helium was used as the carrier gas at a constant
516 flow of 1 ml/min. Samples were injected in splitless mode into a Gerstel PTV injector at 60°C
517 (held for 0.1 min) and heated at 260°C min^{-1} to 300°C. The MS source was operated at 260˚C in
518 EI mode at 70 eV ionization energy and 8000 V acceleration voltage. All samples were injected in
519 *n*-hexane to avoid deterioration of chromatographic signals by $FeCl_2$ build-up in the MS ion source
520 through use of halogenated solvents[60]. The GC oven was programmed from 60°C (held for 4 min)
521 to 315°C at 4°C min^{-1}, with a total run time of 100 min. Saturated steranes and hopanes were
522 quantified using metastable reaction monitoring (MRM), and precursor → product transitions and
523 MS parameters are given in Supplementary Tables 5 and 6. Aromatic (proto)steroids and
524 hopanoids were quantified using selected ion recording (SIR) of base or molecular ions under
525 magnet control. Recorded ions and MS parameters are given in Supplementary Table 7. All ratios
526 and abundance proportions are reported uncorrected for differences in MS-response. Mass spectra
527 were collected using full scan experiments at 1,000 resolution and with a total cycle time of 1.28
528 sec. The identity of 24-methylated triaromatic steroids was confirmed by elution time experiments
529 on a VF-200ms GC capillary column (60 m, 0.25 mm i.d., 0.25 μm film thickness; Agilent J&W
530 Scientific, Agilent Technologies, Santa Clara, CA, USA) using the same parameters as above.

Assembly of Figure 3a. Figure 3a shows the evolution of the relative abundances of saturated hopanes (hop, in yellow) and saturated steranes (cholestanes, chol in red; ergostanes, erg in blue; stigmastanes, stig in green) through geological time. Each horizontal bar represents one bitumen or oil. The proportion of each colour in the horizontals bars is proportional to the relative abundance of the respective compound class (expressed in %). For example, 100% yellow in the Palaeo- and Mesoproterozoic indicates that these samples contain hopanes but no detectable steranes; and predominant yellow, variable amounts red and nearly invisible traces of blue in the Tonian indicate that hopanes predominate and that steranes are overwhelmingly dominated by cholestanes with sub-percent traces of ergostanes but no detectable stigmastanes. A key to the colours and all molecules included in each compound class is supplied in Supplementary Table 2. In Figure 3a, samples are bundled into packages that represent broad geological time units or geological formations. For the Neoproterozoic to present, samples are bundled into the Tonian, Cryogenian, earliest Ediacaran, late Ediacaran, Paleozoic, Mesozoic and Caenozoic. The exact age of individual samples is given in Supplementary Table 3. Palaeo- and Mesoproterozoic bitumens were bundled into individual geological formations or groups labelled (1) to (11) in the time bar of Figure 1 and 3. Information about each geological unit is given in Methods 'Key to geological formations in Figures 1 and 3'. Grey shaded triangles to the left of Figure 1a and 3a assign these geological units to sample bundles. Care thus needs to be taken not to misread the colour charts as a continuous record of biomarker data through time. Biomarker bearing units are rare in the Proterozoic, and the record is temporally sparse and fragmented. Major temporal gaps in the record are represented by grey background.

Assembly of Figures 1a and 3b. Figure 3b shows the evolution of the relative abundances of aromatic hopanoids (yellow), aromatic protosteroids (shades of purple), aromatic 24-methyl

protosteroids or ‘ursteroids’ (cyan and light blue-green), and aromatic crown-steroids including aromatic cholesteroids (shades of red), ergosteroids (shades of blue) and stigmasteroids (shades of green). Figure 1a shows the same data but excludes hopanoids. Different colour shades represent mono-, di- and triaromatic steroids as well as different degrees of ring-system methylation. A key to the colours of all molecules included in each compound class is supplied in Supplementary Table 1, and the corresponding mass chromatograms, mass spectra and chemical structures are given in Extended Data Figures 1 to 9. As described in Methods ‘Assembly of Figure 3a‘, individual samples are bundled into geological units, formations or groups, and the figure should not be read as a continuous time record. See Supplementary Text for information about how individual mass chromatographic signals were assigned to protosteroid, ursteroid or crown-steroid sources. How different diagenetic triaromatic steroids (TAS) that lost methyl groups were assigned to protosterol or crown-sterol sources is described in detail in Supplementary Text 2.8.

Key to geological formations in Figures 1 and 3. ***1*** = 1,640 Ma Barney Creek Fm, McArthur Basin. ***2*** = 1,460 Ma Hongshuizhuang, ***3*** = 1,439 Ma Tieling, ***4*** = 1,392 Ma Xiamaling Fms, North China Craton. ***5*** = 1,308 Ma Velkerri and Kyalla fms, McArthur Basin. ***6*** = 1,100 Ma Nonesuch Fm, Keweenawan Rift. ***7*** = 1,100 Ma En Nesoar and Touirist fms, Taoudeni Basin. ***8*** = 1,000 Ma Ui and Lakhanda groups, Sette-Daban fold belt. ***9*** = 750 Ma Visingsö Gr, Lake Vättern Basin; 729 Ma Walcott Mm, Grand Canyon; 777–725 Ma Hussar Fm, 734 Ma Kanpa Fm, 717–725 Ma Steptoe Fm, Officer Basin; < 820 Ma Loves Creek and Johnnys Creek fms, > 750 Ma Wallara Fm, Amadeus Basin. ***10*** = 659 Ma Aralka Fm, Amadeus Basin. ***11*** = 635 Ma Guia and Mirassol d'Oeste fms, Araras Group.

Key to fossils in Figure 1c. (a) *Melicerion poikilon*, a possible testate rhizarian with an age of ~742 Ma[61]. Image courtesy S. Porter, Fig. 15 (panel 3) in ref[61]. (b) *Bonniea dacruchares,* a testate amoebozoan with an age of ~742 Ma[61]. Image courtesy S. Porter, Fig. 9 (panel 9) in ref[61]. (c) *Bangiomorpha pubescens*, a likely bangiacean rhodophyte alga[62] with an age of 1,047 +13/-17 Ma[63]. Image courtesy N. Butterfield, Fig. 9A in ref[64]. (d) *Proterocladus antiquus,* a likely multicellular, benthic, siphonocladalean chlorophyte alga with an age of 1,056 to 947.8 Ma[11]. Image courtesy Q. Tang, Fig. 2g in ref[11]. (e) *Ourasphaira giraldae,* a likely fungus with an age of 1,010 to 890 Ma[12]. Image courtesy C. Loron, Fig. 1b in ref[12]. (f) *Trachyhystrichosphaera aimika*, a microfossil with diagnostic eukaryotic features from the 1,100 Ma En Nesoar Formation. Image courtesy J. Beghin, Fig. 4a in ref[65]. (g) *Leiosphaeridia jacutica*, a microfossil of possible eukaryotic origin from the 1,100 Ma En Nesoar Formation. Image courtesy J. Beghin, Fig. 2e in ref[65]. (h) *Satka favosa*, a microfossil with diagnostic eukaryotic features from the ~1,300 Ma Roper Group[9]. Image courtesy E. Javaux, Fig. 5(9) in ref[9]. (i) *Valeria lophostriata*, a microfossil with diagnostic eukaryotic features from the ~1,300 Ma Roper Group[9]. Image courtesy E. Javaux, Fig. 7(2) in ref[9]. (j) *Tappania plana*, a microfossil with diagnostic eukaryotic features from the ~1,600 Ma Ruyang Formation[1]. Image courtesy Y. Leiming, Fig. 4A in ref[1]. (k) *Shuiyousphaeridium macroreticulatum*, a microfossil with diagnostic eukaryotic features from the 1,600 Ma Ruyang Formation[1]. Image courtesy Y. Leiming, Fig. 4B in ref[1]. [Permissions pending]

Data availability. All processed data generated during this study are included in this published article and its supplementary information files. Raw data is available from the corresponding author on reasonable request.

METHOD REFERENCES

54 French, K. L. *et al.* Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proceedings of the National Academy of Sciences* **112**, 5915-5920 (2015).

55 Jarrett, A., Schinteie, R., Hope, J. M. & Brocks, J. J. Micro-ablation, a new technique to remove drilling fluids and other contaminants from fragmented and fissile rock material. *Organic Geochemistry* **61**, 57-65 (2013).

56 Brocks, J. J. Millimeter-scale concentration gradients of hydrocarbons in Archean shales: live-oil escape or fingerprint of contamination? *Geochimica et Cosmochimica Acta* **75**, 3196-3213 (2011).

57 Schinteie, R. *et al.* Impact of drill core contamination on compound-specific carbon and hydrogen isotopic signatures. *Organic Geochemistry* **128**, 161-171, doi:10.1016/j.orggeochem.2019.01.003 (2019).

58 Schinteie, R. & Brocks, J. J. Evidence for ancient halophiles? Testing biomarker syngeneity of evaporites from Neoproterozoic and Cambrian strata. *Organic Geochemistry* **72**, 46-58, doi:10.1016/j.orggeochem.2014.04.009 (2014).

59 Brocks, J. J., Grosjean, E. & Logan, G. A. Assessing biomarker syngeneity using branched alkanes with quaternary carbon (BAQCs) and other plastic contaminants. *Geochimica et Cosmochimica Acta* **72**, 871-888 (2008).

60 Brocks, J. J. & Hope, J. M. Tailing of chromatographic peaks in GC–MS caused by interaction of halogenated solvents with the ion source. *Journal of Chromatographic Science* **52**, 471-475 (2014).

61 Porter, S. M., Meisterfeld, R. & Knoll, A. H. Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: a classification guided by modern testate amoebae. *Journal of Paleontology* **77**, 409-429 (2003).

62 Butterfield, N. J., Knoll, A. H. & Swett, K. A bangiophyte red alga from the Proterozoic of arctic Canada. *Science* **250**, 104-107 (1990).

63 Gibson, T. M. *et al.* Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology* **46**, 135-138, doi:10.1130/g39829.1 (2017).

64 Butterfield, N. J. Proterozoic photosynthesis – a critical review. *Palaeontology* **58**, 953-972, doi:10.1111/pala.12211 (2015).

65 Beghin, J. *et al.* Microfossils from the late Mesoproterozoic – early Neoproterozoic Atar/El Mreïti Group, Taoudeni Basin, Mauritania, northwestern Africa. *Precambrian Research* **291**, 63-82, doi:10.1016/j.precamres.2017.01.009 (2017).

66 Wang, X. *et al.* Oxygen, climate and the chemical evolution of a 1400 million year old tropical marine setting. *American Journal of Science* **317**, 861-900, doi:10.2475/08.2017.01 (2017).

67 Zhang, S. *et al.* Sufficient oxygen for animal respiration 1,400 million years ago. *Proceedings of the National Academy of Sciences* **113**, 1731-1736, doi:10.1073/pnas.1523449113 (2016).

68 Holba, A. G. *et al.* Application of tetracyclic polyprenoids as indicators of input from fresh-brackish water environments. *Organic Geochemistry* **34**, 441-469 (2003).

69 Peters, K. E., Walters, C. C. & Moldowan, J. M. *The Biomarker Guide, Volume 2*. 2 edn, (Cambridge University Press, 2004).

Acknowledgements

J.J.B. acknowledges funding support from Australian Research Council grants DP160100607, DP170100556 and DP200100004, and P.A. and P.S. from the Centre National de la Recherche Scientifique and the Université de Strasbourg. B.J.N. acknowledges doctoral and postdoctoral fellowships by the Australian National University, CSIRO Office of the Chief Executive, and the Central Research and Development Fund of the University of Bremen. A.J.M.J. publishes with the permission of the Executive Director, Northern Territory Geological Survey. We thank Dianne Edwards and the organic geochemistry team at Geoscience Australia (GA) for oil samples from the National Collection. For rock specimens and extracts, we thank the Geological Survey of Western Australia (GSWA), the Northern Territory Geological Survey (NTGS), Martin Brasier, Nicholas Butterfield, Junhong Chen, Timothy Gallagher, Emmanuelle Grosjean, Kalle Kirsimae, Aivo Lapland, Małgorzata Moczydłowska, Susannah Porter, Nathan Sheldon, Erik Sperling and Shuizhang Zhang.

Author Contributions

J.J.B. and B.J.N. interpreted the data and wrote the paper with contributions from C.H. and all co-authors. B.J.N., A.J.M.J., N.G., T.L. L.M.v.M, J.H. and J.J.B. conducted the biomarker analyses. B.J.N. and J.H. conducted pyrolysis and ring opening experiments. P.A. and P.S. synthesized standards, conducted RuO_4 oxidation experiments and assisted with compound identification. J.J.B. conceived the project and compiled data and figures.

Competing interests statement

The authors declare no competing interests.

Additional Information

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to J.J.B. and B.J.N.

Reprints and permissions information is available at www.nature.com/reprints.

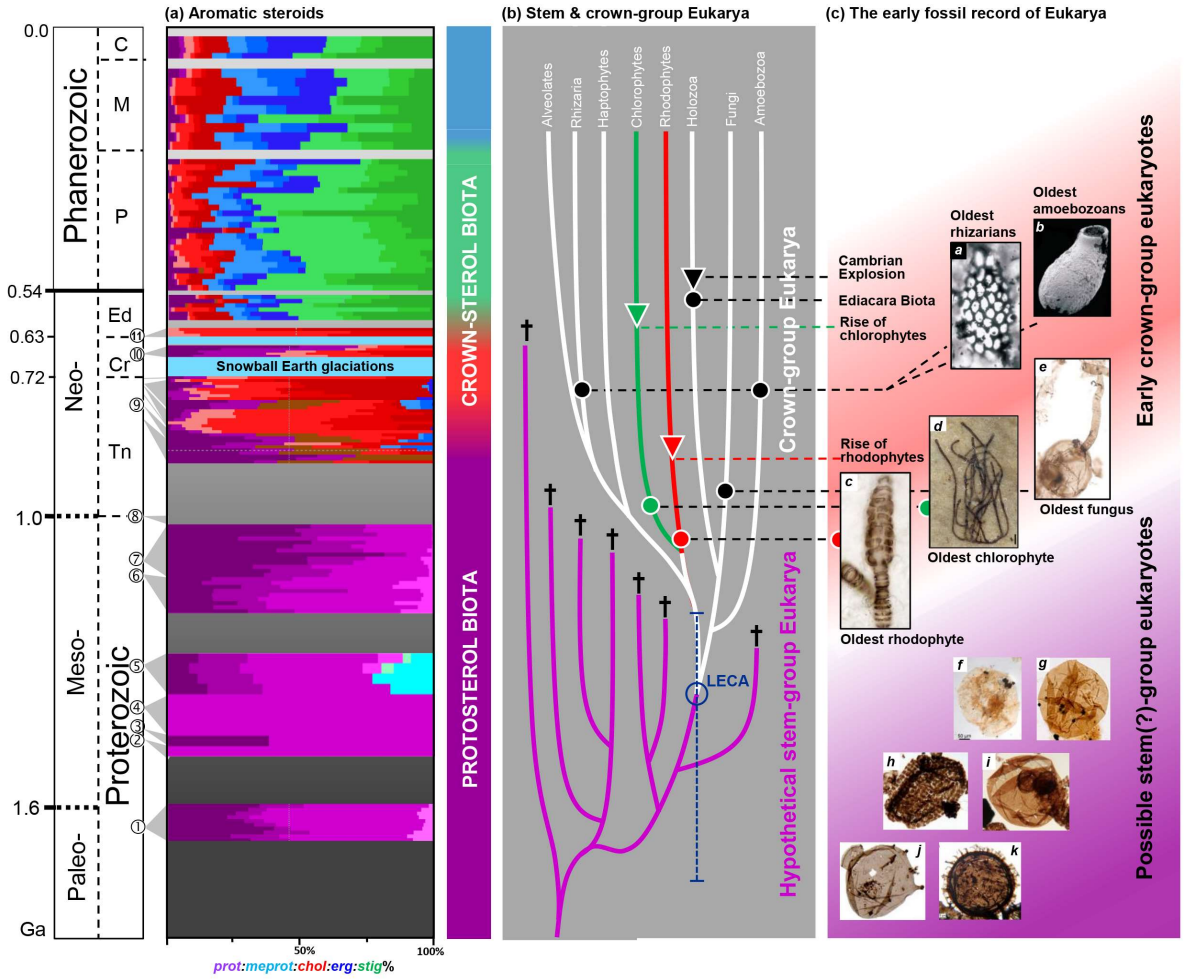


Fig. 1 | Geological time chart (in billion years, Ga) comparing the molecular fossil, microfossil and phylogenetic records of early eukaryote evolution. a, Relative abundances of aromatic protosteroids (purple and cyan tones) and crown-steroids (reds, blues and greens), highlighting the transition from a Protosterol Biota to a 'Crown-sterol Biota' in the Neoproterozoic. Each horizontal colour bar represents one sample, and grey triangles assign data bundles to geological units 1 to 11 (key provided in Methods). prot = protosteroids, meprot = 24-methyl protosteroids (ursteroids), chol = cholesteroids, erg = ergosteroids, stig = stigmasteroids. Key to colours see panel Figure 2 and Supplementary Table 1. For details on data assembly and geological formations (1–11) see Methods. **b,** Phylogenetic tree of the domain Eukarya with white, red and green highlighting crown-group branches. Stem-group branches (purple) are hypothetical only, illustrating the notion that mid-Proterozoic ecosystems may have been dominated by extinct stem forms that did not produce crown-sterols. LECA, the last common ancestor of all extant eukaryotes, may have emerged between 1.2 and > 2.0 Ga (see text). **c,** Microfossils of early eukaryotes. (*a–e*) Likely crown-group Eukarya (~1.1 to 0.7 Ga), (f–k) microfossils that are possibly or certainly eukaryotic but lack diagnostic crown-group characteristics. See Methods for detailed information and image credits. Tn, Cr, Ed are the Tonian, Cryogenian and Ediacaran periods; P, M, C are the Palaeo-, Meso- and Caenozoic eras.

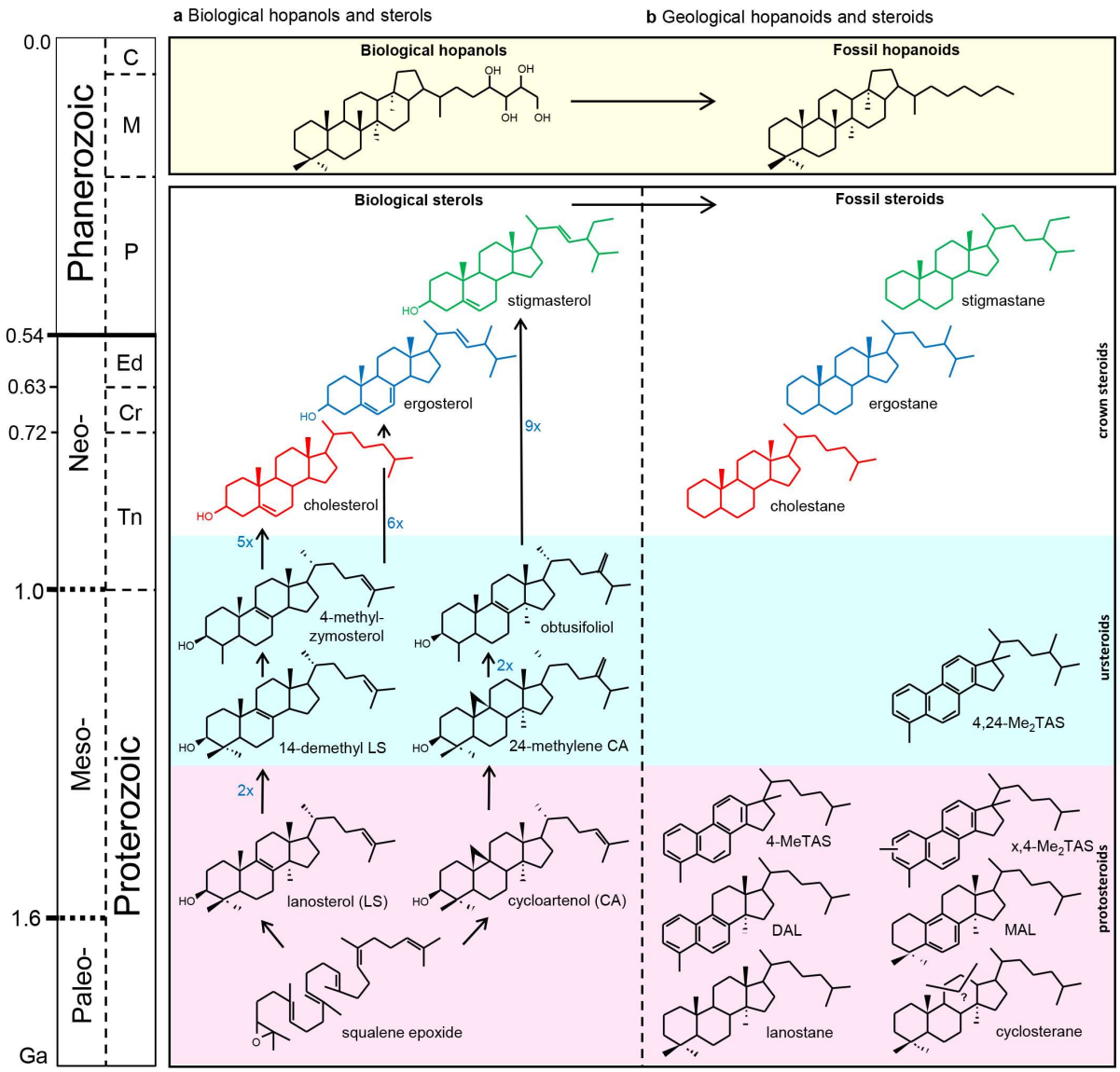


Fig. 2 | Juxtaposition of biological and geological hopanoids and steroids, and the evolution of the sterol biosynthetic pathway through time. a, The pathway from squalene epoxide to protosterols, ursterols and crown-sterols was likely assembled during the Proterozoic. '2x' indicates number of successive enzymatic steps. **b,** Fossil steroids detected in the rock record that correspond to intermediates and end-products in the evolving biosynthetic pathway. TAS = triaromatic steroids, MAL = monoaromatic lanosteroids, DAL = diaromatic lanosteroids, 'x,4- Me2TAS' signifies a TAS methylated at C-4 and at an unknown position x. '?' indicates that cyclosterane likely contains a bridge at an unknown position.

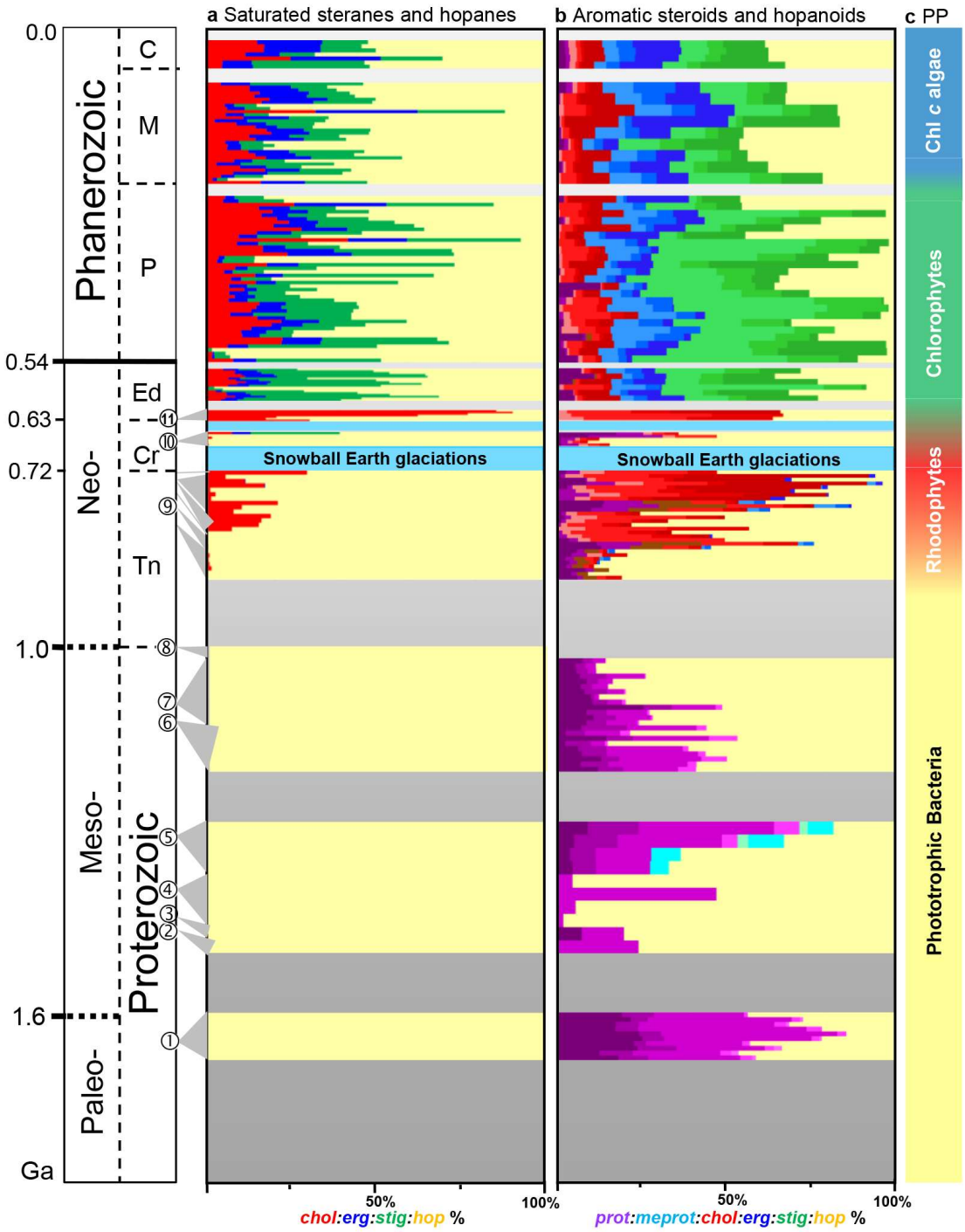


Fig. 3 | The succession of fossil steroids and bacterial hopanoids through time. a, The relative abundances of saturated steranes (red, blue and green) and bacterial hopanes (hop, yellow). Note absence of detectable steranes before 800 Ma, highlighting the paucity of crown-group eukaryotes. **b,** A new view of biomarker evolution based on the relative abundances of aromatic protosteroids (purple), 24-methyl protosteroids (ursteroids, cyan), aromatic crown-steroids (reds, blues and greens) and aromatic hopanoids (yellow). The data reveal the existence of a protosterol-producing biota that was ecological dominant in the Proterozoic. For (a) and (b), each horizontal colour bar represents one sample. Key to colours, formations and abbreviations see Fig. 1. **c,** The geological succession of dominant primary producers (PP) based on molecular fossils.

EXTENDED DATA TABLE & FIGURE LEGENDS

Extended Data Table 1. Information on Cryogenian to Palaeoproterozoic geological samples.

* Alternate sample names 14B311 = (#1-14); 14B316 = (#3-14); 14B313 = (#2-16) in refs[66,67].

Extended Data Fig. 1 | Diagenetic and pyrolysis products of ur- and protosterols. Chemical structures in purple and cyan are biogenic precursors. Structures in black are fossil lipids detected in mid-Proterozoic sedimentary rocks and generated in pyrolysis experiments of the respective biolipids. Green arrows signify biosynthetic reactions, black dashed arrows point to products of diagenesis (and laboratory pyrolysis). Note that the dashed arrows do not imply direct product-precursor relationships as diagenetic reactions commonly involve numerous intermediates and complex reaction networks. All structures in black, apart from **X**, were found in mid-Proterozoic bitumens. TAS = triaromatic steroids, MAL = monoaromatic lanosteroids, DAL = diaromatic lanosteroids. 'x,4,24-Me_3TAS' signifies a TAS methylated at C-4, C-24 and at an unknown position x.

Extended Data Fig. 2 | Mass spectra and elution behaviour of cyclosterane and its cleavage products in a severely biodegraded migrabitumen (12Z083, drill core MY4, 103.3 m). a, *M/z* 412 to 300 partial ion chromatograms of cyclosterane and successive side-chain cleavage products. 'x' indicate absence or low concentration of pseudohomologs indicative of side-chain branching positions. R is the pentacyclic core of cyclosterane, '*' denotes a chiral centre. **b,** Total ion current (TIC) of the saturated hydrocarbon fraction of the biodegraded oil, highlighting the biodegradation resistance of the two cyclosterane isomers *k1* and *k2*. '~' marks truncated signal of internal standard. **c,** Mass spectra of the cycloartane side-chain cleavage products. See text for explanation

of red and blue dashed lines. **d,** Juxtaposition of the mass spectra of cyclosterane isomer *k2* (upper panel) and cycloartane from the NIST 95 library (lower panel) (the mass spectrum of *k1* is not shown as it is nearly identical to k*2*). **e,** Suggested major MS fragmentation of hypothetical cyclosterane structure **I** and cycloartane.

Extended Data Fig. 3 | Mass chromatograms and spectra of protosteranes. a, *M/z* 259 partial mass chromatograms of a co-injection experiment of Eocene bitumen Y2 containing 8β(H),9α(H)-lanostane *l4* (see text) and 1,640 Ma Barney Creek Fm (B03178). **b,** MRM 414 → 259 showing the elution positions of lanostane isomers *l1* to *l4*, and the results of pyrolysis experiments on cycloartenol, lanosterol and euphenol (3β,13α,14β,17α-lanost-8-en-3-ol). **c,** Mass spectra of signal m3 of euphenol pyrolysis, *l4* in BCF sample B03178 and lanostane in the Y2 standard. **d, e,** MRM chromatograms highlighting the relative elution positions of stigmastanes, lanostane, cyclosteranes and C_{27} hopanes in (d) the 1,440 Ma Tieling Fm (sample 17B101) and (e) an Ordovician oil from Australia (GA#299). TTP1 and 2 are Tetracyclic Terpane isomers[68]. Note that 8β(H),9α(H)-lanostane *l4* co-elutes with TTP2, but that its presence can be recognized by an elevated TTP2 peak or a shoulder trailing the peak. The chromatograms are identified by MRM precursor → product transitions and relative signal heights in % relative to the highest signal.

Extended Data Fig. 4 | Mass chromatograms and spectra of monoaromatic lanosteroids (MAL). a, *M/z* 379 mass chromatogram identifying the 20S and 20R isomers of C_{29} MAL in sample B03163 from the 1,640 Ma Barney Creek Fm (black) and coinjection experiment with an authentic MAL standard on a DB-5MS capillary column (red). **b,** Authentic C_{29} MAL standard. **c,** C_{28} MAL of sample 14B211 from the 725 Ma Kanpa Fm showing an immature isomer distribution (S << R). **d,** C_{28} MAL generated through pyrolysis of lanosterol (sample BEX20150624). **e,** C_{28}

MAL of sample B03163 from the 1,640 Ma Barney Creek Fm showing a mature isomer distribution (S ≈ R). seco-hop = monoaromatic 8,14-secohopanoids (see Extended Data Fig. 8a). **f,** Mass spectra of signals labelled in (a) to (e). ***a1*** to ***a4*** are chromatographic signal identifiers for geological bitumens, and ***A1*** to ***A4*** identifiers for corresponding authentic standards and pyrolysis products. Ancient bitumen chromatograms are in black, standard chromatograms in blue.

Extended Data Fig. 5 | Mass chromatograms of diaromatic lanosteroids (DAL). a, d, g, *M/z* 404, 379 and 361 mass chromatograms identifying isomers of C_{28}, C_{29} and C_{30} DAL generated through pyrolysis of cycloartenol. **b, e, h,** C_{28}, C_{29} and C_{30} DAL of sample 14B211 from the 725 Ma Kanpa Fm showing an immature isomer distribution (20S << 20R). **c, f, i,** C_{28}, C_{29} and C_{30} DAL of sample B03162c from the 1,640 Ma Barney Creek Fm with a mature isomer distribution (20S ≈ 20R). ***b1*** to ***b10*** are chromatographic signal identifiers for geological bitumens, and ***B1*** to ***B10*** identifiers for pyrolysis products of authentic standards in blue.

Extended Data Fig. 6 | Mass spectra of DAL given in Extended Data Figure 5. ***b1*** to ***b10*** are chromatographic signal identifiers for geological bitumens, and ***B1*** to ***B10*** identifiers for pyrolysis products of cycloartenol. Vertical labels '13 Aug 18 35' are unique identifiers for individual GC-MS experiments. Signals marked 'x' are from coeluting compounds.

Extended Data Fig. 7 | Mass chromatograms of triaromatic steroids (TAS). a, *M/z* 259 mass chromatogram of TAS with two methylations in the ring system (Me_2TAS) in sample 13N14 from the 1,100 Ma El Mreïti Group, and **b,** in the pyrolysate of an enrichment mixture of cycloartenol (CA) and 24-methylene cycloartenol (MCA). **c,** *M/z* 245 chromatogram of A-ring methylated TAS (MeTAS) of the pyrolyzed CA and MCA mixture, **d,** of sample 10J093 of the ~750 Ma Chuar Group, **e** the El Mreïti Gr (as above), and **f,** the Phanerozoic-based AGSOSTD oil reference

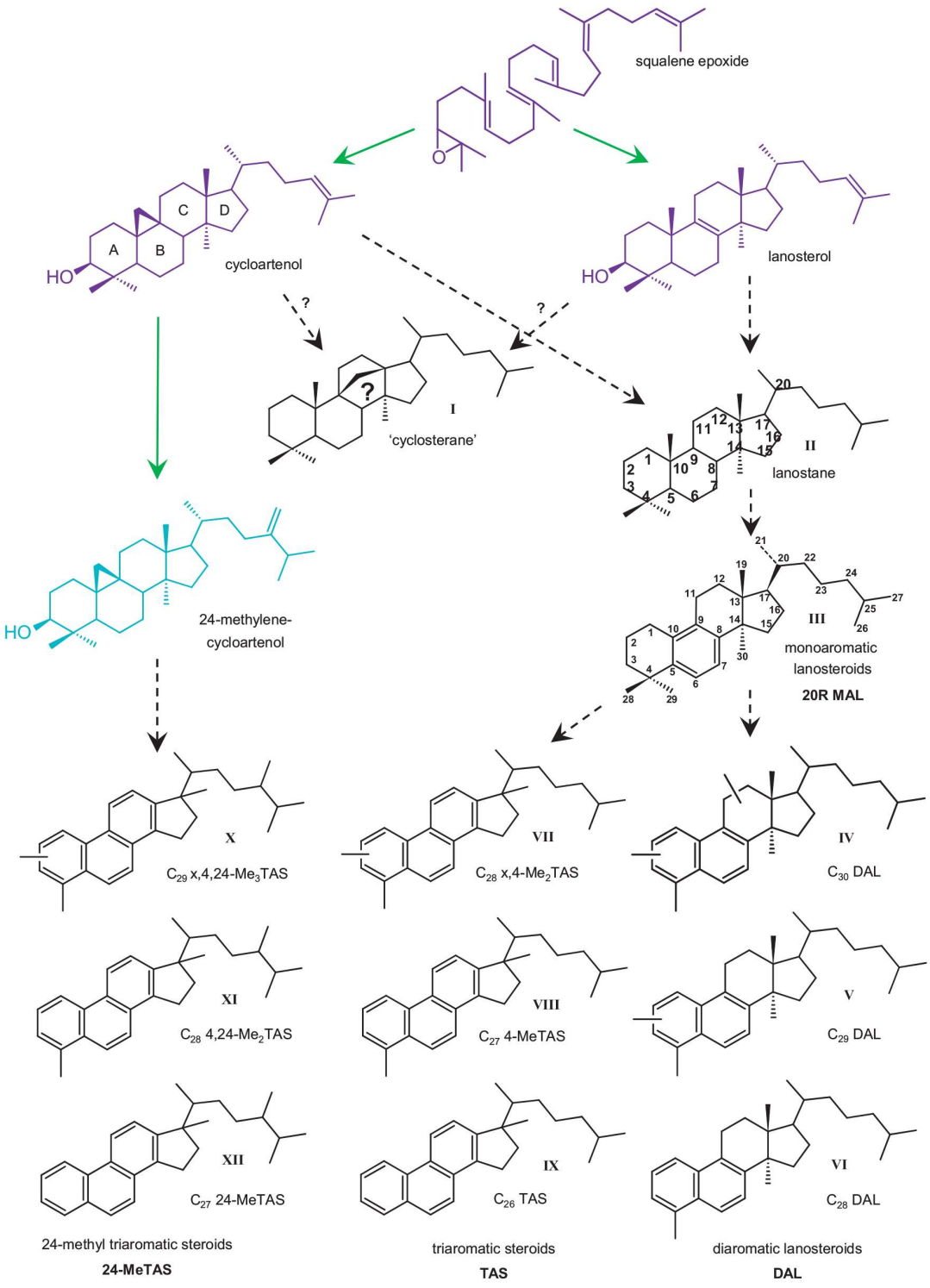
736 standard. TAS dinosteroids in grey. **g,** *M/z* 231 mass chromatogram of TAS without methylations
737 in the ring system of the Chuar Gr, **h,** the El Mreïti Gr and **i,** the AGSOSTD (samples as above).
738 **j,** *M/z* 253 chromatogram of monoaromatic steroids (MAS) of sample 14B212 of the ~725 Ma
739 Kanpa Fm and **k,** the AGSOSTD. Colour coding as in panel (e). MAS compound identification
740 and roman numeral nomenclature follows Figure 13.107 and 13.108 in ref[69].

741 **Extended Data Fig. 8 | Mass chromatograms of aromatic hopanoids in the 1,640 Ma Barney**
742 **Creek Fm (12B117), and thermal maturity evaluation of the sample set.** Mass chromatograms
743 of **a,** *m/z* 365 identifying regular aromatic 8,14-secohopanoids, **b,** *m/z* 351 regular 28-nor-8,14-
744 secohopanoids, **c,** *m/z* 414 8,14-secohopanoids with a fluorene moiety, **d,** *m/z* 416 and **e,** *m/z* 402
745 8,14-secohopanoids with an acenaphthene moiety, and **f,** *m/z* 191 benzohopanoids. **g, h,** Evaluation
746 of possible bias in the aromatic steroid and hopanoid record caused by thermal maturity based on
747 cross plots of thermal maturity indicator Rc(MPR) against (g) sample age, and (h) the relative
748 abundance of aromatic hopanoids and steroids (H/(H+S)). Rc(MPR) = computed vitrinite
749 reflectance (Rc) based on the Methyl Phenanthrene Ratio (MPR, see Supplementary Table 1).

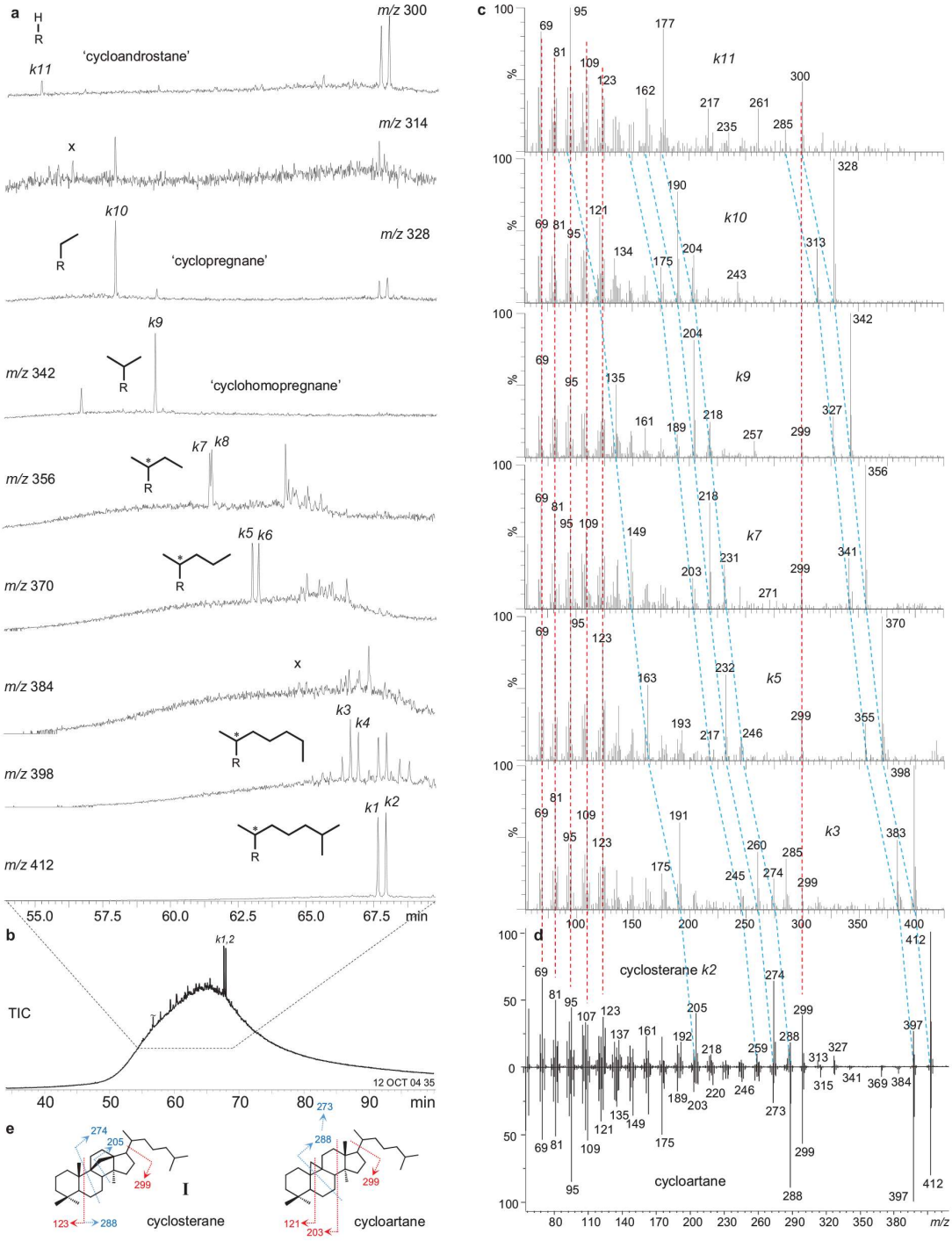
750 **Extended Data Fig. 9 | Confirmation of the elution position of S and R 4,24-dimethyl**
751 **triaromatic cholesteroids in SIR *m/z* 245 traces on DB-5MS (left) and VZ-200 (right) gas**
752 **chromatographic columns. a, c,** The AGSOSTD standard, **b, e,** the ~1,300 Ma Velkerri
753 Formation, Roper Group (sample 230692, drill core Altree-2, 410.55 m). Note that the 20S-isomer
754 of the 4,24-dimethyl triaromatic cholesteroid coelutes with the 20R-isomer of 4-methyl triaromatic
755 cholesteroid on the DB-5MS capillary column, but is resolved on VZ-200. **d,** A-ring methylated
756 24-methly triaromatic cholesteroid standard formed by pyrolysis of ergosterol (Supplementary

757 Methods). 4,24-dimethyl triaromatic cholesteroids were also generated through pyrolysis of the

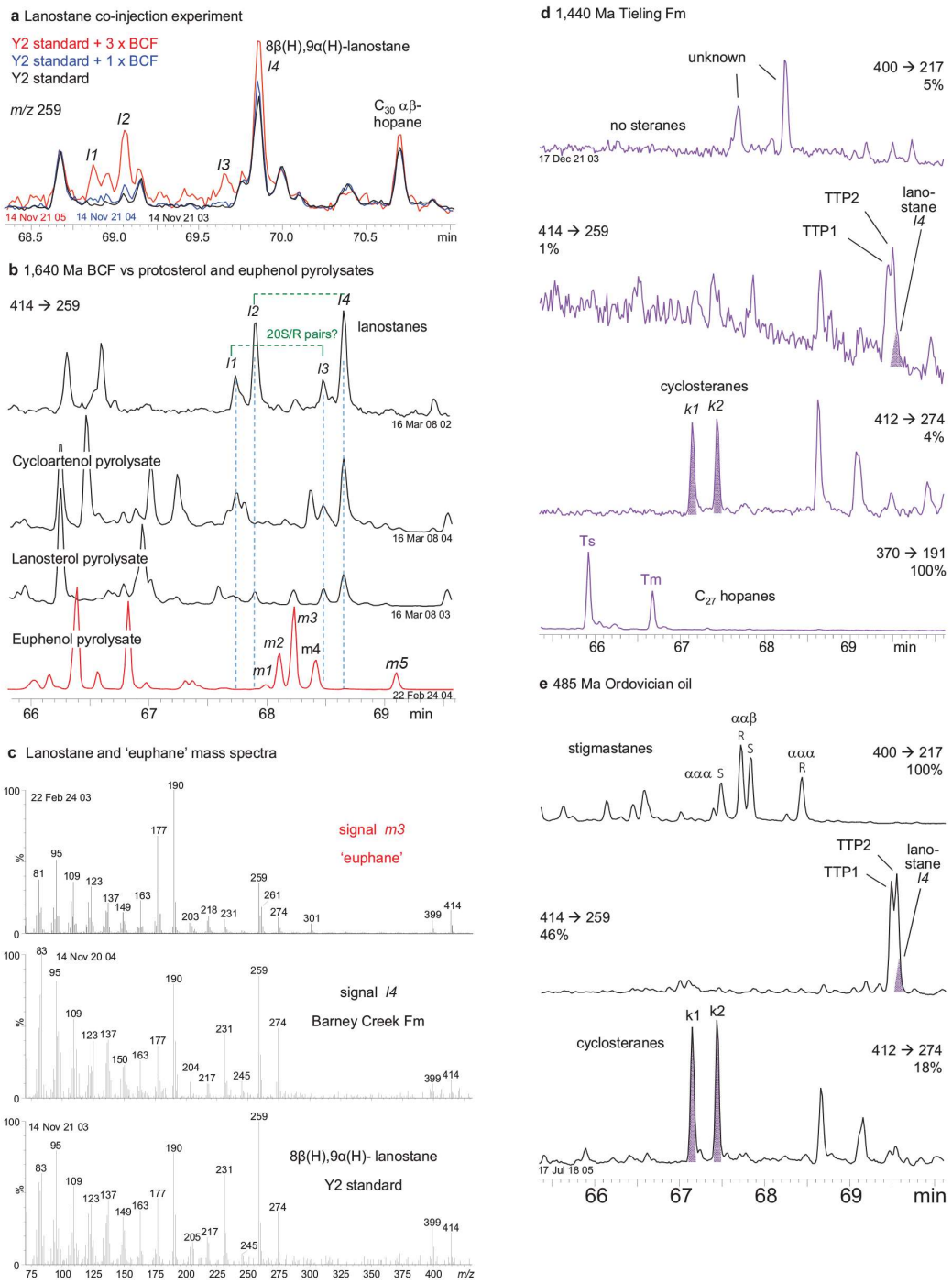
758 C_{31} ursterol 24-methylene cycloartenol (Extended Data Fig. 7c).



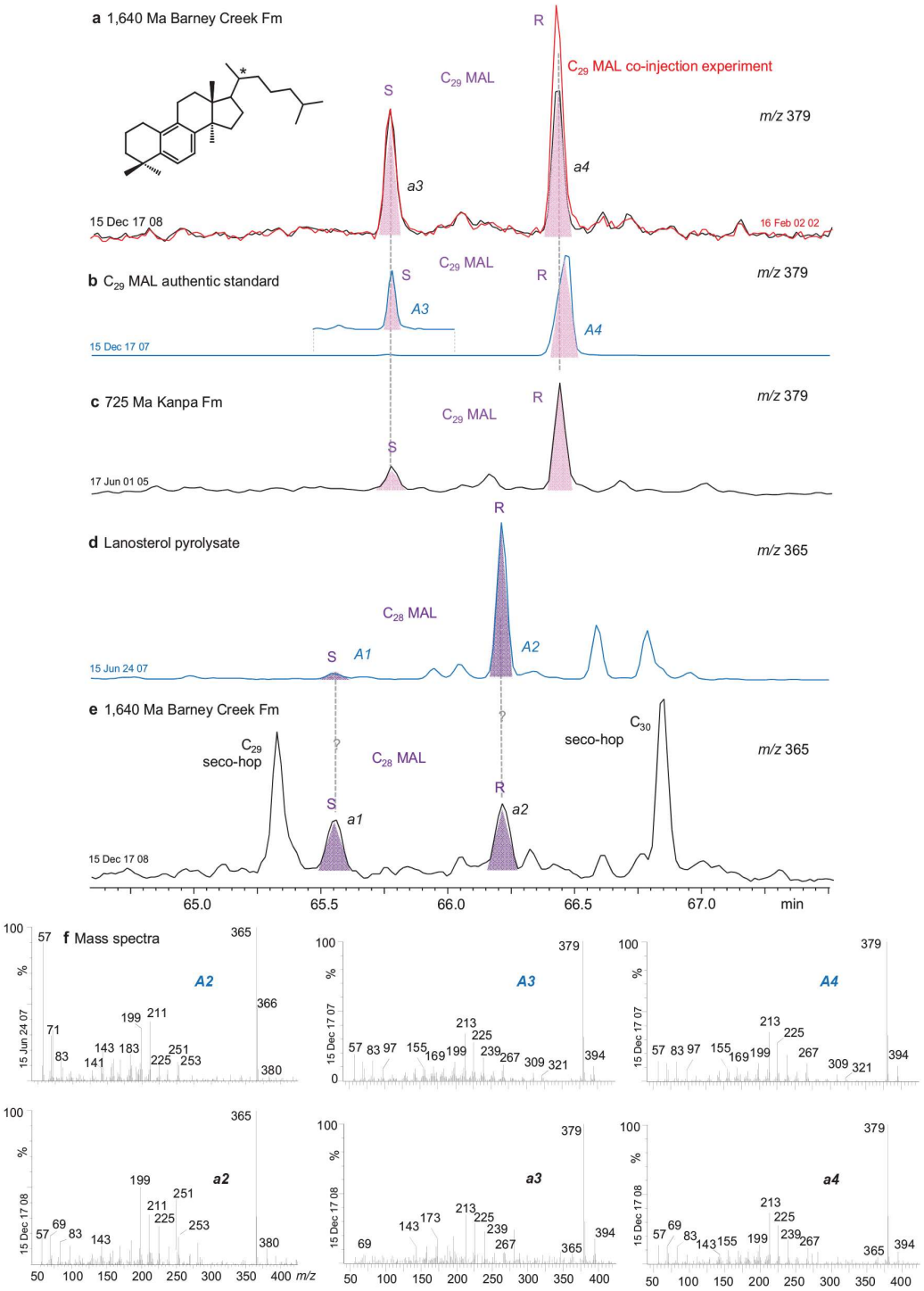
Extended Data Fig. 1 | Diagenetic and pyrolysis products of ur- and protosterols. Chemical structures in purple and cyan are biogenic precursors. Structures in black are fossil lipids detected in mid-Proterozoic sedimentary rocks and generated in pyrolysis experiments of the respective biolipids. Green arrows signify biosynthetic reactions, black dashed arrows point to products of diagenesis (and laboratory pyrolysis). Note that the dashed arrows do not imply direct product precursor relationships as diagenetic reactions commonly involve numerous intermediates and complex reaction networks. All structures in black, apart from **X**, were found in mid-Proterozoic bitumens. TAS = triaromatic steroids, MAL = monoaromatic lanosteroids, DAL = diaromatic lanosteroids. 'x,4,24-Me$_3$TAS' signifies a TAS methylated at C-4, C-24 and at an unknown position x.



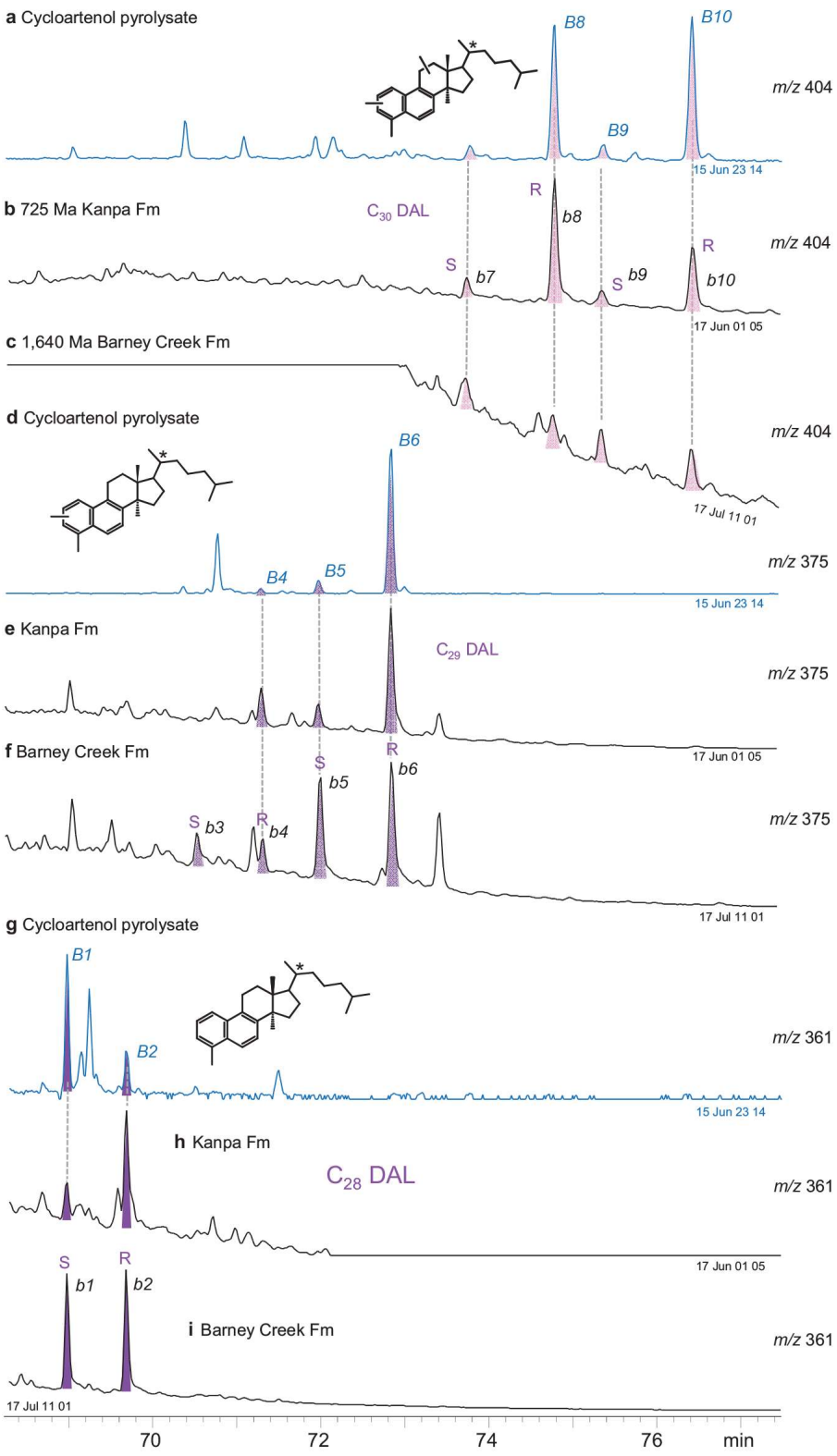
Extended Data Fig. 2 | Mass spectra and elution behaviour of cyclosterane and its cleavage products in a severely biodegraded migrabitumen (12Z083, drill core MY4, 103.3 m). a, *M/z* 412 to 300 partial ion chromatograms of cyclosterane and successive side-chain cleavage products. 'x' indicate absence or low concentration of pseudohomologs indicative of side-chain branching positions. R is the pentacyclic core of cyclosterane, '*' denotes a chiral centre. **b,** Total ion current (TIC) of the saturated hydrocarbon fraction of the biodegraded oil, highlighting the biodegradation resistance of the two cyclosterane isomers *k1* and *k2*. '~' marks truncated signal of internal standard. **c,** Mass spectra of the cycloartane side-chain cleavage products. See text for explanation of red and blue dashed lines. **d,** Juxtaposition of the mass spectra of cyclosterane isomer *k2* (upper panel) and cycloartane from the NIST 95 library (lower panel) (the mass spectrum of *k1* is not shown as it is nearly identical to k*2*). **e,** Suggested major MS fragmentation of hypothetical cyclosterane structure **I** and cycloartane.



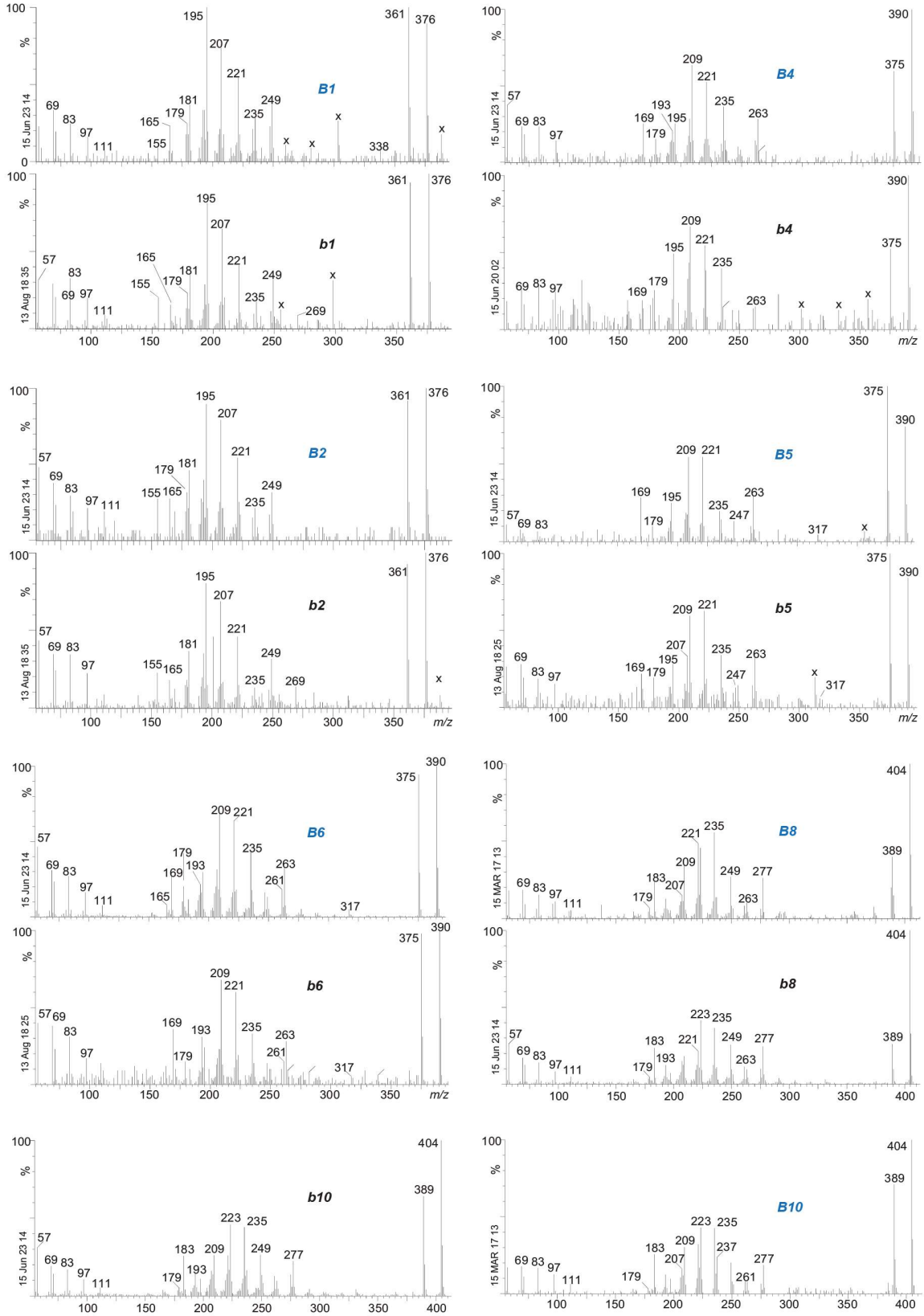
Extended Data Fig. 3 | Mass chromatograms and spectra of protosteranes. a, *M/z* 259 partial mass chromatograms of a co-injection experiment of Eocene bitumen Y2 containing 8β(H),9α(H)-lanostane *l4* (see text) and 1,640 Ma Barney Creek Fm (B03178). **b,** MRM 414 → 259 showing the elution positions of lanostane isomers *l1* to *l4*, and the results of pyrolysis experiments on cycloartenol, lanosterol and euphenol (3β,13α,14β,17α-lanost-8-en-3-ol). **c,** Mass spectra of signal m3 of euphenol pyrolysis, *l4* in BCF sample B03178 and lanostane in the Y2 standard. **d, e,** MRM chromatograms highlighting the relative elution positions of stigmastanes, lanostane, cyclosteranes and C_{27} hopanes in (d) the 1,440 Ma Tieling Fm (sample 17B101) and (e) an Ordovician oil from Australia (GA#299). TTP1 and 2 are Tetracyclic Terpane isomers68. Note that 8β(H),9α(H)-lanostane *l4* co-elutes with TTP2, but that its presence can be recognized by an elevated TTP2 peak or a shoulder trailing the peak. The chromatograms are identified by MRM precursor → product transitions and relative signal heights in % relative to the highest signal.



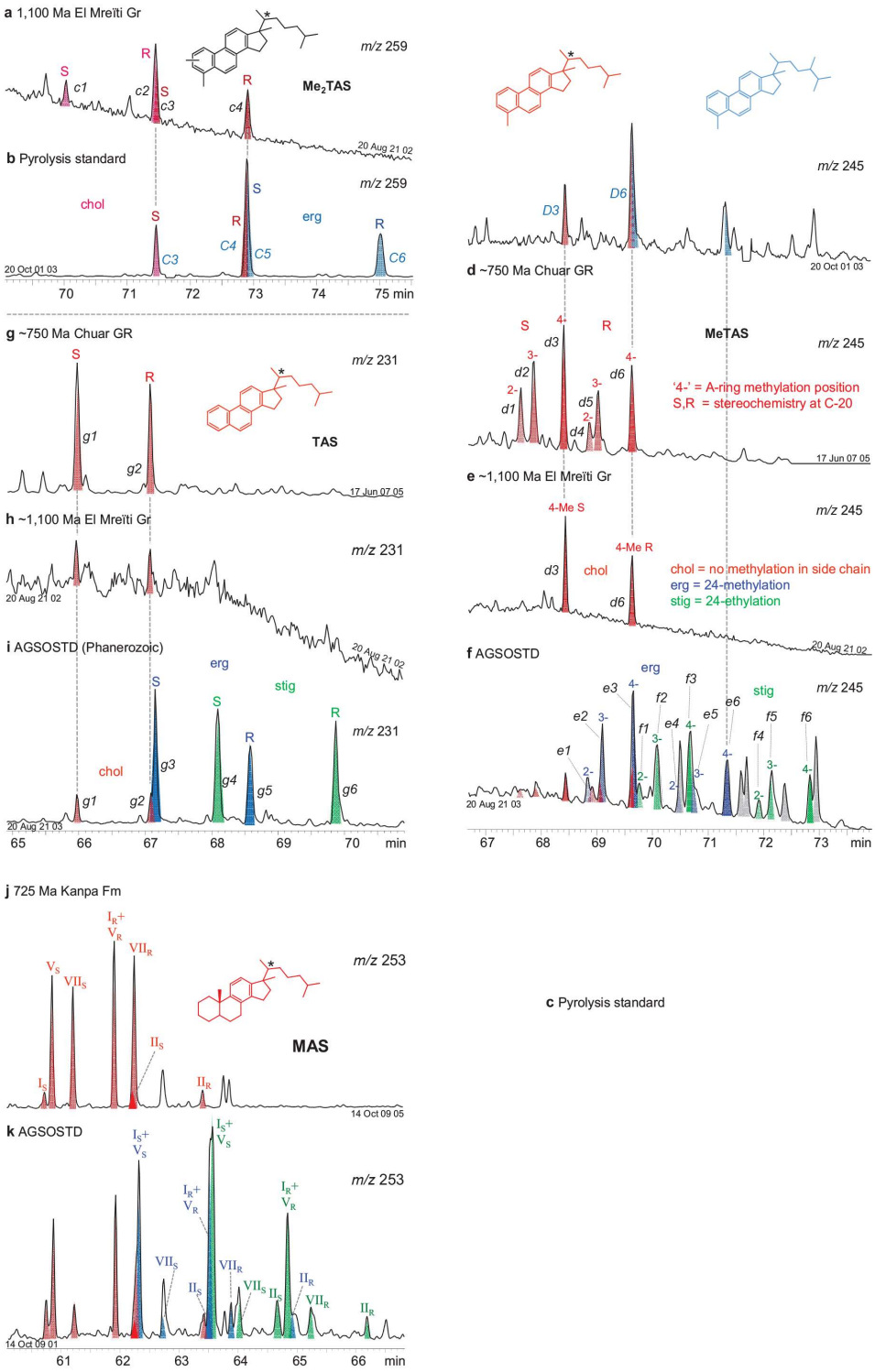
Extended Data Fig. 4 | Mass chromatograms and spectra of monoaromatic lanosteroids (MAL). **a,** *M/z* 379 mass chromatogram identifying the 20S and 20R isomers of C_{29} MAL in sample B03163 from the 1,640 Ma Barney Creek Fm (black) and coinjection experiment with an authentic MAL standard on a DB-5MS capillary column (red). **b,** Authentic C_{29} MAL standard. **c,** C_{28} MAL of sample 14B211 from the 725 Ma Kanpa Fm showing an immature isomer distribution (S << R). **d,** C_{28} MAL generated through pyrolysis of lanosterol (sample BEX20150624). **e,** C_{28} MAL of sample B03163 from the 1,640 Ma Barney Creek Fm showing a mature isomer distribution (S H R). seco-hop = monoaromatic 8,14-secohopanoids (see Extended Data Fig. 8a). **f,** Mass spectra of signals labelled in (a) to (e). ***a1*** to ***a4*** are chromatographic signal identifiers for geological bitumens, and ***A1*** to ***A4*** identifiers for corresponding authentic standards and pyrolysis products. Ancient bitumen chromatograms are in black, standard chromatograms in blue.



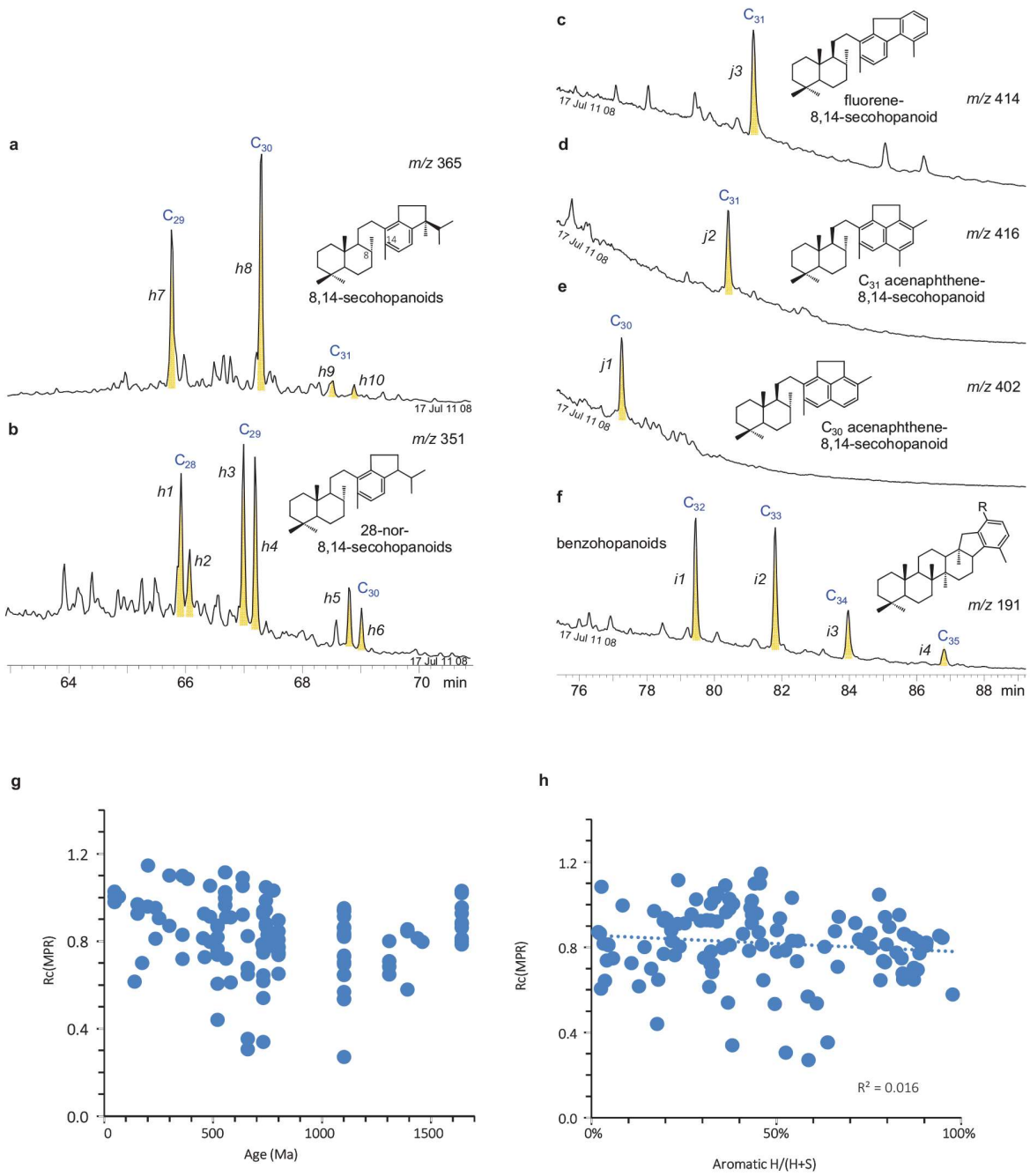
Extended Data Fig. 5 | Mass chromatograms of diaromatic lanosteroids (DAL). a, d, g, *M/z* 404, 379 and 361 mass chromatograms identifying isomers of C_{28}, C_{29} and C_{30} DAL generated through pyrolysis of cycloartenol. **b, e, h,** C_{28}, C_{29} and C_{30} DAL of sample 14B211 from the 725 Ma Kanpa Fm showing an immature isomer distribution (20S << 20R). **c, f, i,** C_{28}, C_{29} and C_{30} DAL of sample B03162c from the 1,640 Ma Barney Creek Fm with a mature isomer distribution (20S H 20R). ***b1*** to ***b10*** are chromatographic signal identifiers for geological bitumens, and ***B1*** to ***B10*** identifiers for pyrolysis products of authentic standards in blue.



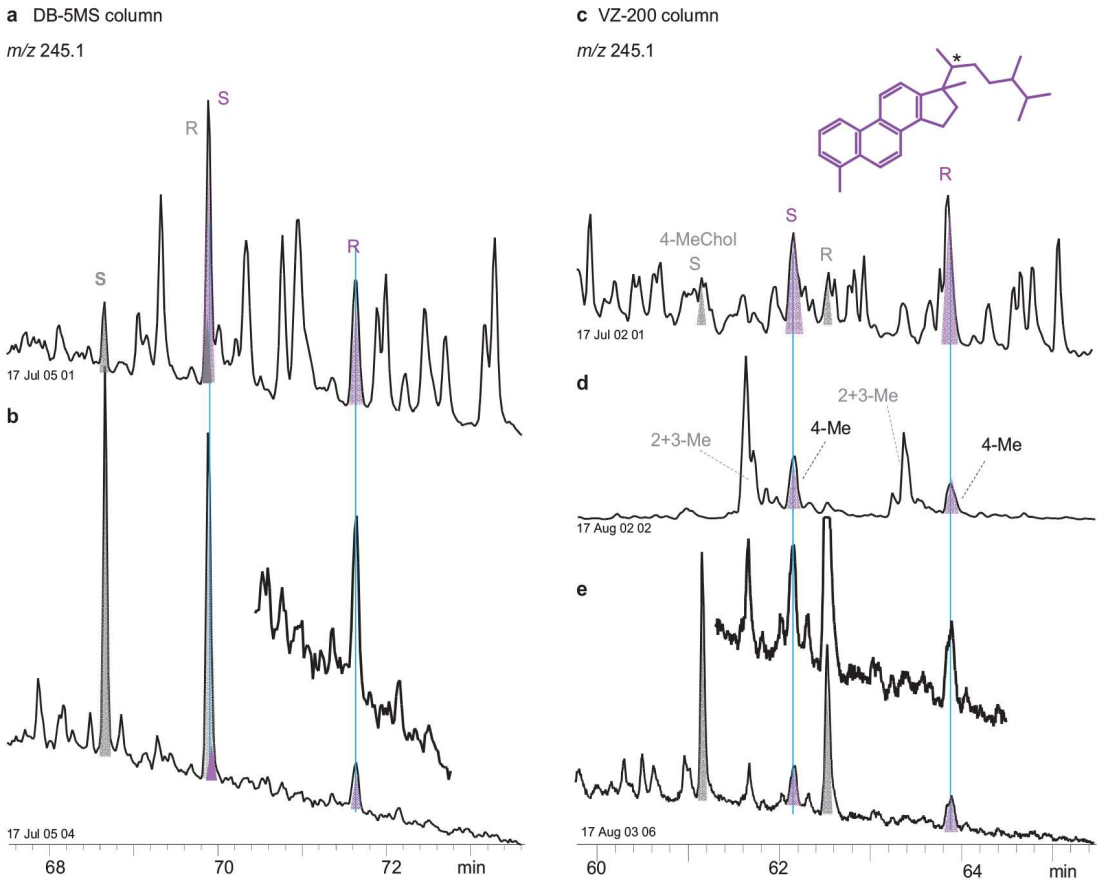
Extended Data Fig. 6 | Mass spectra of DAL given in Extended Data Figure 5. ***b1*** to ***b10*** are chromatographic signal identifiers for geological bitumens, and ***B1*** to ***B10*** identifiers for pyrolysis products of cycloartenol. Vertical labels '13 Aug 18 35' are unique identifiers for individual GC-MS experiments. Signals marked 'x' are from coeluting compounds.



Extended Data Fig. 7 | Mass chromatograms of triaromatic steroids (TAS). a, *M/z* 259 mass chromatogram of TAS with two methylations in the ring system (Me_2TAS) in sample 13N14 from the 1,100 Ma El Mreïti Group, and **b,** in the pyrolysate of an enrichment mixture of cycloartenol (CA) and 24-methylene cycloartenol (MCA). **c,** *M/z* 245 chromatogram of A-ring methylated TAS (MeTAS) of the pyrolyzed CA and MCA mixture, **d,** of sample 10J093 of the ~750 Ma Chuar Group, **e** the El Mreïti Gr (as above), and **f,** the Phanerozoic-based AGSOSTD oil reference standard. TAS dinosteroids in grey. **g,** *M/z* 231 mass chromatogram of TAS without methylations in the ring system of the Chuar Gr, **h,** the El Mreïti Gr and **i,** the AGSOSTD (samples as above). **j,** *M/z* 253 chromatogram of monoaromatic steroids (MAS) of sample 14B212 of the ~725 Ma Kanpa Fm and **k,** the AGSOSTD. Colour coding as in panel (e). MAS compound identification and roman numeral nomenclature follows Figure 13.107 and 13.108 in ref 69.



Extended Data Fig. 8 | Mass chromatograms of aromatic hopanoids in the 1,640 Ma Barney Creek Fm (12B117), and thermal maturity evaluation of the sample set. Mass chromatograms of **a,** *m/z* 365 identifying regular aromatic 8,14-secohopanoids, **b,** *m/z* 351 regular 28-nor-8,14- secohopanoids, **c,** *m/z* 414 8,14-secohopanoids with a fluorene moiety, **d,** *m/z* 416 and **e,** *m/z* 402 8,14-secohopanoids with an acenaphthene moiety, and **f,** *m/z* 191 benzohopanoids. **g, h,** Evaluation of possible bias in the aromatic steroid and hopanoid record caused by thermal maturity based on cross plots of thermal maturity indicator Rc(MPR) against (g) sample age, and (h) the relative abundance of aromatic hopanoids and steroids (H/(H+S)). Rc(MPR) = computed vitrinite reflectance (Rc) based on the Methyl Phenanthrene Ratio (MPR, see Supplementary Table 1).



Extended Data Fig. 9 | Confirmation of the elution position of S and R 4,24-dimethyl triaromatic cholesteroids in SIR *m/z* 245 traces on DB-5MS (left) and VZ-200 (right) gas chromatographic columns. a, c, The AGSOSTD standard, **b, e,** the ~1,300 Ma Velkerri Formation, Roper Group (sample 230692, drill core Altree-2, 410.55 m). Note that the 20S-isomer of the 4,24-dimethyl triaromatic cholesteroid coelutes with the 20R-isomer of 4-methyl triaromatic cholesteroid on the DB-5MS capillary column, but is resolved on VZ-200. **d,** A-ring methylated 24-methly triaromatic cholesteroid standard formed by pyrolysis of ergosterol (Supplementary Methods). 4,24-dimethyl triaromatic cholesteroids were also generated through pyrolysis of the C_{31} ursterol 24-methylene cycloartenol (Extended Data Fig. 7c).

Extended Data Table 1 : Information on Cryogenian to Palaeoproterozoic geological samples

Sample ID	Age [Ma]	Basin or region	Geological unit	Drill Core	Depth [m]	Sample lithology and characteristics
Ediacaran						
Araras Group, Brazil – Terconi Outcrop (~635 Ma)						
LvM 00311	635	Araras Group	Mirassol d'Oeste Fm	Outcrop	11.7	Grey Dolomite
LvM 00313	635	Araras Group	Guia Fm.	Outcrop	16.4	Dark Carbonate
LvM 00315	635	Araras Group	Guia Fm.	Outcrop	31.2	Dark Carbonate
Cryogenian (635 – 717 Ma)						
Amadeus Basin – Central Australia						
Aralka Formation (659 Ma)						
11J005 i	659	Amadeus Basin	Aralka Formation	BR05	242.20	Laminated grey mudstone
13J005 i	659	Amadeus Basin	Aralka Formation	BR05	454.00	Dark grey siltstone
12J019 i2	659	Amadeus Basin	Aralka Formation	BR05	471.50	Black laminated shale
13J0902 i	659	Amadeus Basin	Aralka Formation	BR05	476.20	Medium grey laminated shale
13J0903 i	659	Amadeus Basin	Aralka Formation	BR05	480.45	Medium grey laminated shale
Tonian (717 – 1000 Ma)						
Wallara Formation (aka Finke Beds) (>750 Ma)						
11J021 i	>750	Amadeus Basin	Wallara Formation	BR05	573.06	Laminated grey carbonate
13J717 i	>750	Amadeus Basin	Wallara Formation	BR05	585.2	Brown micritic mudstone with cross cutting molar tooth structures
Johnnys Creek Formation, Bitter Springs Group (<820 Ma)						
11J038 i	<820	Amadeus Basin	Johnnys Creek Formation	BR05	762.18	Laminated grey micritic limestone
13J718 i	<820	Amadeus Basin	Johnnys Creek Formation	BR05	808.2	Laminated dolomitic mudstone with molar tooth structures
MG006 i	<820	Amadeus Basin	Johnnys Creek Formation	BR05	813.6	Dark brown microbial mat in a light grey carbonate matrix
11J024 i	<820	Amadeus Basin	Johnnys Creek Formation	BR05	816.26	Dark brown laminated carbonate with light grey to white microbial mat
MG013	<821	Amadeus Basin	Johnnys Creek Formation	BR05	931.33	Laminated dolomitic mudstone with stromatolitic features
MG007 i	<820	Amadeus Basin	Johnnys Creek Formation	BR05	924.62	Microbialite with micrite/sparite lamination
11J050 I	<820	Amadeus Basin	Johnnys Creek Formation	BR05	942.74	Laminated dark grey carbonate
13J036 i	<820	Amadeus Basin	Johnnys Creek Formation	BR05	956.25	Laminated dark grey carbonate
13J107 i	<820	Amadeus Basin	Johnnys Creek Formation	BR05	961.91	Laminated grey carbonate
Loves Creek Formation, Bitter Springs Group (<820 Ma)						
11J063 i	<820	Amadeus Basin	Loves Creek Formation	BR05	1224.69	Grey dolomicrite cross-cut by thin white molar tooth structures
Officer Basin – Western Australia						
Steptoe Formation, Buldya Group (717-725 Ma)						
14B214 i	717-25	Officer Basin	Steptoe Formation	Empress 1A	509.14	Light and dark grey laminated mudstone
14B213 i	717-25	Officer Basin	Steptoe Formation	Empress 1A	588.21	Black laminated mudstone with evaporite nodules
Kanpa Formation, Buldya Group (725 ± 11 Ma)						
14B212 i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	629.6	Light grey dolomite, with soft sediment deformation
14B211 i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	737.39	Dark grey to dark brown mudstone
14B21 i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	757	Light grey siltstone with black mudstone laminae
14B209 i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	764.64	Dark grey finely laminated mudstone
14B208a i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	768.25	Stromatolite-like structure with wrinkled laminae, base and top
14B208b i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	768.25	Medium grey laminated mudstone
14B206 i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	829.42	Dark grey to black mudstone
14B204 i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	830.35	Medium grey extremely soft mudstone, some wavy bedding
14B203 i	725 ± 11	Officer Basin	Kanpa Formation	Empress 1A	839.23	Mudstone
Hussar Formation, Buldya Group (777-725 Ma)						
14B202 i	777-725	Officer Basin	Hussar Formation	Empress 1A	1072.85	Finely laminated (stromatolitic?) light grey dolomite with anhydrite nodules
14B201-b i	777-725	Officer Basin	Hussar Formation	Empress 1A	1074.15	Light and dark grey planar laminae, with anhydrite nodule
14B201-a i	777-725	Officer Basin	Hussar Formation	Empress 1A	1074.15	Light and dark grey planar laminae, with anhydrite nodule
Grand Canyon - Arizona - Outcrop at Sixtymile Canyon and Nankoweap Butte (depths are meters above base)						
Walcott Member, Kwagunt Formation, Chuar Group (729 ± 0.9 Ma)						
10J090	729 ± 0.9	Grand Canyon	Walcott Member	outcrop	1609	Dolomite nodule
10J093	729 ± 0.9	Grand Canyon	Walcott Member	outcrop	1544	Black shale
10J092	729 ± 0.9	Grand Canyon	Walcott Member	outcrop	1494	Massive black shale
10J091	729 ± 0.9	Grand Canyon	Walcott Member	outcrop	1480	Black shale
10J089	729 ± 0.9	Grand Canyon	Walcott Member	outcrop	1365	Black shale
B04026 - i	729 ± 0.9	Grand Canyon	Walcott Member	outcrop	1343	Black carbonaceous siltstone
B04029 - i	729 ± 0.9	Grand Canyon	Walcott Member	outcrop	1340	Black carbonaceous sandstone
Lake Vättern Basin - Sweden- Outcrop at Girabäcken locality						
Visingsö Group (> 740 Ma)						
VG-20-03-i	~750	Lake Vättern Basin	Visingsö Group	outcrop		Black mudstone
Mesoproterozoic (1000 – 1600 Ma)						
Sette– Daban fold belt, Siberian Craton, Russia						
08m001 (LK-4)	1000	Sette-Daban fold belt	Lakhanda Group, Neryuen Fm	outcrop		black-grey, sometimes discontinuously laminated shale, associated sediments yield mat-forming filamentous microfossils
08m009 (LK-81)	1000	Sette-Daban fold belt	Ui Group, Kandyk Fm	outcrop		variegated green-grey shale assoc. with ripples & mudcracks
Taoudeni Basin, El Mreïti Group, Mauritania						
11N048	1100	Taoudeni Basin	Touirist Fm	S2	139.6	Shale with grey and brown sub-mm laminae, carbonate lenses, mass-movement
13N06	1100	Taoudeni Basin	Touirist Fm	S2	140.25	Shale with grey and brown wrinkly sub-mm laminae
11N052	1100	Taoudeni Basin	Touirist Fm	S2	140.55	Shale with brownish and grey wrinkly sub-mm laminae
12N020	1100	Taoudeni Basin	Touirist Fm	S2	141.9	Shale with grey and black sub-mm laminae
11N049	1100	Taoudeni Basin	Touirist Fm	S2	142.2	Shale with brownish and grey wrinkly sub-mm laminae
11N050	1100	Taoudeni Basin	Touirist Fm	S2	151.95	Shale with brown and grey sub-mm laminae
11N054	1100	Taoudeni Basin	En Nesoar Fm	S2	185.4	Shale with dark grey and black sub-mm laminae
11N045b	1100	Taoudeni Basin	En Nesoar Fm	S2	187.3	Shale with grey and black sub-mm laminae
13N13	1100	Taoudeni Basin	En Nesoar Fm	S2	188.1	Dark-grey laminated calcitic shale/siltstone, oxic facies based on iron spec data
13N14	1100	Taoudeni Basin	En Nesoar Fm	S2	188.6	Calcareous green-grey shale, anoxic facies based on iron speciation data
11N055	1100	Taoudeni Basin	En Nesoar Fm	S2	200.01	Shale with dark and medium grey sub-mm laminae
11N051	1100	Taoudeni Basin	En Nesoar Fm	S2	203.1	Shale with very dark and black sub-mm laminae, ~2mm wide pyrite layer
11N056	1100	Taoudeni Basin	En Nesoar Fm	S2	206.08	Shale with dark and light grey sub-mm laminae, sign of mass-movement
Keweenawan Rift, Michigan Basin, USA						
13N36	1100	Keweenawan Rift	Nonesuch Fm	PI-2	90.10	shale with sub-mm dark and mid-grey laminae interlayered with carbonates
13N41	1100	Keweenawan Rift	Nonesuch Fm	PI-2	226.20	Calcific shale with light to dark grey slightly wrinkly layers, salty crust on surface
13N31	1100	Keweenawan Rift	Nonesuch Fm	D06	391.90	Shale with black and dark grey sub mm laminae
13N37	1100	Keweenawan Rift	Nonesuch Fm	D06	417.50	Shale with mid and light grey sub mm laminae
13N34	1100	Keweenawan Rift	Nonesuch Fm	WPB-3	204.70	Light grey shale interlayered with mid-grey silty clay and carbonate
B04005	1100	Keweenawan Rift	Nonesuch Fm	WPB-3	237.74	black shale with sub-mm laminae
13N45	1100	Keweenawan Rift	Nonesuch Fm	WPB-3	266.40	Shale with dark to light grey sub-mm laminae
13N42	1100	Keweenawan Rift	Nonesuch Fm	WPB-4	63.30	Shale with mid and light grey sub mm laminae
B04007	1100	Keweenawan Rift	Nonesuch Fm	WPB-4	140.67	Black shale with sub-mm Laminae
McArthur Basin, Northern Australia						
11N043	1313 ± 47	McArthur Basin	Kyalla Fm	Jamison-1	1649.81	Black shale
GA 20150106	1308 ± 41	McArthur Basin	Velkerri Fm	Altree-2	410.55	Dark brown to dark grey planar laminated siltstone
GA 20150115	1308 ± 41	McArthur Basin	Velkerri Fm	Altree-2	452.64	Dark grey to black siltstone, flaser-like cross-bedding
GA 20150118	1308 ± 41	McArthur Basin	Velkerri Fm	Altree-2	471.24	Dark grey laminated siltstone, silts are surrounded by 2 cm thick sand layers (couplets) with oil shows
GA 20150129	1308 ± 41	McArthur Basin	Velkerri Fm	Altree-2	527.37	Alternating layers of laminated light and dark grey siltstone surrounded by 2 cm thick sand layers (couplets)
Man 1	1308 ± 41	McArthur Basin	Velkerri Fm	McManus1	1150.00	finely laminated black shale
North China Craton						
14B311*	1392	North China Craton	Xiamaling Fm	Core 1 [1]	53.48	Black shale (courtesy S. Zhang, Petrochina)
14B316*	1392	North China Craton	Xiamaling Fm	Core 3 [1]	222.35	Black silty mudstone (courtesy S. Zhang, Petrochina)
14B313*	1392	North China Craton	Xiamaling Fm	Core 2 [1]	285.90	Black shale (courtesy S. Zhang, Petrochina)
17B100	1392	North China Craton	Xiamaling Fm	YJ2	73.60	Black Shale (courtesy S. Zhang, Petrochina)
17B101	1439	North China Craton	Tieling Fm	YJ2	117.40	Black Shale (courtesy S. Zhang, Petrochina)
17B103	1460	North China Craton	Hongshuizhuang Fm	YJ2	463.40	Black silty mudstone (courtesy S. Zhang, Petrochina)
Palaeoproterozoic (1600 – 2500 Ma)						
McArthur Basin, Northern Australia						
B03162c	1640	McArthur Basin	Barney Creek Fm	GR-7	45.35	dolomitic siltstonewith dark laminae
B03167	1640	McArthur Basin	Barney Creek Fm	GR-7	71.65	Dark dolomitic siltstone
B03168	1640	McArthur Basin	Barney Creek Fm	GR-7	82.95	Dark dolomitic siltstone
B03169	1640	McArthur Basin	Barney Creek Fm	GR-7	90.30	Dark dolomitic siltstone
B03175	1640	McArthur Basin	Barney Creek Fm	GR-7	162.85	Dark dolomitic siltstone
B03178	1640	McArthur Basin	Barney Creek Fm	GR-7	199.08	Dark dolomitic siltstone overlain by breccia
12B117	1640	McArthur Basin	Barney Creek Fm	LV09	382.22	dark and light grey laminated dolomite
12B118	1640	McArthur Basin	Barney Creek Fm	LV09	423.10	finely laminated mudstone
12Z167	1640	McArthur Basin	Barney Creek Fm	LV09	455.52	finely laminated mudstone
12B119	1640	McArthur Basin	Barney Creek Fm	LV09	492.98	dark and light grey finely laminated dolomite

Lost world of complex life and the late rise of the eukaryotic crown

SUPPLEMENTARY INFORMATION

Jochen J. Brocks[1*], Benjamin J. Nettersheim[1,2*], Pierre Adam[3], Philippe Schaeffer[3], Amber J. M. Jarrett[1,4], Nur Güneli[1], Tharika Liyanage[1], Lennart M. van Maldegem[1], Christian Hallmann[5], Janet M. Hope[1]

[1] Research School of Earth Sciences, The Australian National University, Canberra, ACT 2601, Australia.
[2] MARUM – Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, 28359 Bremen, Germany
[3] Université de Strasbourg, CNRS, Institut de Chimie de Strasbourg UMR 7177, F-67000 Strasbourg, France
[4] Northern Territory Geological Survey, GPO Box 4550, Darwin NT 0801, Australia
[5] GFZ German Research Center for Geosciences, 14473 Potsdam, Germany
[*]These authors contributed equally to this work. e-mail: jochen.brocks@anu.edu.au; bnettersheim@marum.de

Supplementary Methods and Supplementary Discussion

1. Geology and samples

Extended Data Table 1 and Supplementary Table 3 summarize information about the bitumens (organic rock extracts) and oils used to assemble the saturated and aromatic steroid and hopanoid records in Figure 1 and 3. Ediacaran and Phanerozoic biomarker data are included to provide a baseline to evaluate the older Proterozoic record. The geological and environmental context for Ediacaran and younger samples is available in the literature and not further described here. Cryogenian and Tonian bitumens come from a transect across palaeocontinent Rodina, including the Amadeus Basin in central Australia, the Officer Basin in Western Australia, the Lake Vättern Basin in Sweden, and from the Grand Canyon, USA. Geological units include the 659 million years old (Ma) Aralka Formation, > 750 Ma Wallara Formation, < 820 Ma Johnnys Creek and Loves Creek formations (both from the Bitter Springs Group), the 717–725 Ma Steptoe, 725 ± 11 Ma Kanpa and 777–725 Hussar formations (all from the Buldya Group), the 729 ± 0.9 Ma[70] Walcott Member of the Kwagunt Formation, the Chuar Group, and the ≤ 886 and 740 Ma Visingsö Group[71]. Age constraints for these formations and their geological and environmental context are summarized in the Supplementary Discussion of Brocks et al. (2017)[15]. The geology of Meso- and Palaeoproterozoic units are outlined below.

1.1. Late Mesoproterozoic El Mreïti Group, Taoudeni Basin, Mauritania (~1,100 Ma)

The epicratonic Taoudeni Basin covers a total area > 1,750,000 km^2 on the West African Craton, extending from Mauritania to northern Mali and Western Algeria[65,72]. Up to 1,300 m of gently dipping (< 0.5°) Mesoproterozoic to Palaeozoic sediments unconformably overlie an Archean/Palaeoproterozoic basement and are covered by thin Mesozoic/Caenozoic strata. The formations analysed in this study belong to the Hodh, the first of four Megasequences, resting directly on the basement. In the north-central part of the basin, the Hodh comprises, from oldest to youngest, the Douik, El Mreïti and Cheikhia groups.

The El Mreïti Group is subdivided into the Khatt, En Nesoar, Touirist, Aguelt el Mabha and Gouamîr formations. Re-Os geochronology on black shales yielded an age of 1107 ± 12 Ma for the Touirist and 1109 ± 22 Ma for the En Nesoar Formation[73]. The sediments of the El Mreïti Group are undeformed and unmetamorphosed and comprise marine stromatolitic carbonates and shales. In the north-west of the basin, the sediments were deposited in an open, relatively shallow epeiric sea[74], characterized by sea-level fluctuations in a general shallowing-upward trend[75-77].

The eleven samples analysed in this study comprise black shales from the En Nesoar and Touirist formations from drill core S2 that intersects the El Mreïti Group in the north-western part of the basin[78]. Five finely laminated black shales of the En Nesoar Formation have total organic carbon (TOC) contents of ~2 to 15%. In drill core S2, the black shales occur as decimetre-thick layers interbedded with dominant grey, green and brown shales and clayey siltstones with occasional ripples, gutter casts and wavy bedding[78,79]. These sediments were deposited in a relatively shallow, subtidal environment during a transgressive phase[77]. The Touirist Formation was deposited during maximal flooding of the basin beneath fair-weather wave base. The lower Touirist Formation comprises coniform stromatolites that thrived within the photic zone. The stromatolitic facies is overlain by thin-beds of light- to dark-grey clayey dolomite and limestone interbedded with decimetre to meter-thick layers of extremely carbon-rich black shales[77,78]. The six black shales analysed in this study (Extended Data Table 1) are finely planar- or wavy-laminated and have TOC contents of 10 to 32%.

The black shales of both formations formed in quiet sub-wave base environments that were possibly protected by off-shore stromatolite reefs[77] and, based on stratigraphic context, presumably not very deep. The black shales intersected by drill hole S2 were deposited beneath (transitory) ferruginous and occasionally sulphidic waters[78] and comprise micrometre-thin, plane to wrinkly, continuous organic laminae with occasional framboidal pyrite and pyritized filamentous sheaths[78]. The organic laminae are interpreted as benthic microbial mats that either represent heterotrophic and chemosynthetic microbial communities thriving beneath anoxic ferruginous to euxinic waters or communities of anoxygenic and possibly oxygenic phototrophic bacteria[78,2] persisting beneath

(transitory) anoxic waters. Fluorescence microscopy on thin-sections also revealed irregularly shaped, discrete or bedding-parallel accumulations of organic particles interpreted as planktonic debris[2].

Micropalaeontological work on the El Mreïti Group, including sediments from drill core S2, yielded an exquisitely preserved assemblage of organic-walled microfossils[65]. The assemblage included 8 taxa of ornamented and 3 taxa of process-bearing vesicles interpreted as unambiguously eukaryotic, 9 possibly eukaryotic taxa, 6 probable prokaryotes (diverse leiospheroids), and 22 other prokaryotic or eukaryotic taxa. The eleven, generally rare taxa classified as 'diagnostic eukaryotic' were absent from all samples of the black shale facies, but six of nine taxa classified as 'possibly eukaryotic' were present and partly abundant in this facies[78]. In the basin, the abundance of fossils with diagnostic eukaryotic features decreases from shallow to deeper environments, and this was interpreted as the possible ecological preference of these taxa, although preservational bias cannot be ruled out[78].

The Touirist Formation in drill core S2 did not contain any diagnostic eukaryotic taxa but yielded numerous leiospheroids classified as 'possibly eukaryotic'. Such leiospheroids were present in all samples[78] that yielded protosteroids (Extended Data Table 1, Supplementary Table 1). Likewise, all samples of the En Nesoar Formation that yielded protosteroids contained abundant leiospheroids[78].

The En Nesoar Formation of drill core S2 also yielded the diagnostic eukaryotic fossils *Simia annulare*, *Jacutianema solubila*, *Pterospermopsimorpha insolita*, *Trachyhystrichosphaera aimika*, *Trachyhystrichosphaera botula*, *Valeria lophostriata* and *Vidaloppala*[78]. Such fossils with typical eukaryotic features were also found in two samples from the En Nesoar Formation that yielded protosteroids but lacked crown-steroids: 13N13 contained *Jacutianema solubila*, *Trachyhystrichosphaera aimika* and *Simia annulare*; and 13N14 yielded *Simia annulare*, *Trachyhystrichosphaera aimika*, *Trachyhystrichosphaera botula*, and *Vidaloppala*[78]. These observations strengthen the hypothesis that the possibly eukaryotic and clearly eukaryotic fossils of the El Mreïti Group have a stem-group origin.

1.2. Late Mesoproterozoic Nonesuch Formation, Keweenawan Rift, USA (~1,100 Ma)

The Nonesuch Formation is part of the Oronto Group, in the upper Keweenaw Peninsula, and is recorded in several cores drilled around White Pine, Michigan, USA. The Oronto Group sits within the failed Midcontinent Rift System. The grey to black shales (50 to 200 m) of the Nonesuch formation overlie the Copper Harbor Conglomerate and are overlain by the Freda Sandstone. The age of the Nonesuch Fm was determined by Re/Os dating to be 1078 (±24) Ma[80].

The nine samples of the Nonesuch investigated in this study were collected from drill cores PI-2, DO-6, WPB-3 and WPB-4, all drilled near the White Pine Mine, Michigan, USA (Extended Data Table 1). Seven samples contain sub-mm dark grey to black laminae, one sample is a light grey shale interlayered with grey silty shale and carbonate (#13N34), and one sample is a calcific shale with wrinkly grey laminae (#13N41).

The Oronto Group was originally described as a transgressive-regressive alluvial fan-lacustrine-fluvial system[81]. The Nonesuch Formation contains several indicators that the basin was shallowing upwards, including microbial induced sedimentary structures[82], raindrop impressions, desiccation polygons[83] and mud cracks[84]. A stromatolitic horizon near the top of the Copper Harbor Conglomerate was interpreted as the onset of lacustrine deposition[85]. A lacustrine interpretation for the depositional environment is supported by mineralogy and total iron over aluminium (FeT/Al) systematics, with values reflective of authigenic enrichments of FeT/Al representative of iron flux similar to modern palaeo-lake settings within a volcanic province[86]. Furthermore, the presence of pigmented hematite titanite, leucoxene and authigenic titanium-oxide grains indicate the depositional basin witnessed in-place oxidation, supporting the hypothesis of an oxygenated water column during deposition. However, others suggest the basin was influenced by marine waters during deposition of the Nonesuch Fm causing brackish conditions. This interpretation is based on an enrichment of redox-sensitive elements U, Mo, Co, Zn and V, sulfur isotopes and slightly elevated Sr/Ba ratio[87].

Over the past decade, the Nonesuch lagerstätte has gathered a lot of interests as the organic rich shales contain a wide range of unicellular and simple multicellular microfossils[88,89], providing a window into eukaryotic diversification during the Late Mesoproterozoic. The formation yields a large variety of simple, presumably prokaryotic sphaeromorphs and filaments, but also eukaryotic microfossils such as *Valeria lophostriata,* linear multicellular clusters comparable with *Gloeodiniopis*, and the multicellular *Proterocladus* that is now recognized as a possible chlorophyte[11,83].

1.3. Mid-Mesoproterozoic Roper Group, McArthur Basin, Australia (~1,300 Ma)

The geology of the Roper Group was previously summarized[90-93]. The Roper Group is a component of the Wilton Package and was deposited in the McArthur Basin of northern Australia over an area of > 180,000 km^2. It consists of a 1 to 3 km thick package of siliciclastic rocks characterised by alternating mudrock and cross-bedded sandstones. They are thickest in the Beetaloo Sub-basin and gradually decreases in thickness away from the depocentre. Based on two competing models, the Roper Group was either deposited in a shallow-marine to shelf environment on an epicontinental platform or on an intracratonic ramp.

The Velkerri Fm, a shoaling sequence deposited between two sandstone units in the middle of the Roper Group, dominantly comprises deep-water, basinal facies with abundant black shale and organic-rich siltstones, and is sub-divided into the Kalala, Amungee and Wyworrie members[94].

The age of the Velkerri Formation is constrained using maximum and minimum depositional ages as well as direct shale measurements. The maximum deposition age is constrained by detrital zircon U-Pb ages of 1386 ± 13 Ma and 1385 ± 116 Ma in the underlying Bessie Creek Sandstone [93,95] and a detrital zircon age of 1308 ± 41 Ma in the Wyworrie Member[95]. The minimum age of the Velkerri Formation is constrained by the intrusive Derim Derim Dolerite that has a TIMS U-Pb baddeleyite age of 1312.9 ± 0.7 Ma[96]. The Amungee Member of the Velkerri Fm yielded Re-Os ages of 1361 ± 21 Ma (C shale) and 1417 ± 29 Ma (A shale)[97].

In the present study, four thermally immature[27] dark grey to black siltstones from the upper Velkerri Fm from drill core Altree-2 yielded indigenous protosteroids. Altree-2 was collared in the Beetaloo Sub-basin, the main depocentre of the Roper Seaway. During deposition of the siltstones and shales of the upper Velkerri Fm in Altree-2, the water column was presumably suboxic to anoxic and ferruginous at depth but not sulphidic[92]. Moreover, anoxic waters did not, or rarely, rise into the euphotic zone of the water column[22,27].

The Roper Group is well known for well preserved assemblages of early eukaryotic fossils (Fig. 1c)[98]. Diagnostic eukaryotic species such as *Satka favosa*, *Satka squamifera* and *Valeria lophostriata* occur largely in marginal marine to inner shelf facies, while the process and protrusion-bearing eukaryote *Tappania plana* is restricted to deeper, distal shelf facies. In basinal facies, such as those of the Velkerri Fm, fossils with diagnostic eukaryotic characteristics are rare, although possibly eukaryotic leiospheroids occur[98]. Moreover, the probable eukaryote *Blastanosphaera kokkoda* is abundant in basinal shales of the Mainoru Fm (a deep-water facies towards the base of the Roper Group), forming local monospecific assemblages. Importantly, rare specimens of *Dictyosphaera macroreticulata*, a large acritarch with convincing eukaryotic features, is restricted to basinal facies in the Velkerri Fm[9].

1.4. Early Mesoproterozoic formations of the North China Craton (1,392 to 1,460 Ma)

The Yanlioa Basin in the North China Craton was formed during the rifting of the Nuna (Columbia) Supercontinent and contains a ~10 km thick Mesoproterozoic and Neoproterozoic sedimentary sequence. The black shales and silty mudstones in the present study come from the Mesoproterozoic Jixian and Huailai groups, deposited between 1,685 and 1,320 Ma[99]). The lower Jixian Group is separated—listed from oldest to youngest—into the Gaoyuzhuang, Yangzhuang, Wumishan, Honghuizhuang and Tieling formations. The overlying Huailai Group only contains

the Xiamaling Fm. In this study we collected six samples from four drill cores covering the Xiamaling, Tieling and Hongshuizhuang formations (Extended Data Table 1).

The age of the Hongshuizhuang Fm is constrained by the ~1,485 Ma underlying Wumishan Fm[100] and the ~1,439 Ma overlying Tieling Fm (with an average age of 1,460 Ma given in Extended Data Table 1). The ~1,439 Ma age of the Tieling Fm is based on U/Pb analyses on zircons recovered from a tuff layer in the middle part of the formation[100,101]. The Tieling and overlying Xiamaling formations are separated by a ~14 million year unconformity linked to the Qinyu uplift (1,432–1,418 Ma[99]). A tuffaceous layer in Member 3 of the Xiamaling Fm is dated at ~1,392 Ma[100,102] while the upper limit of the top of the Xiamaling Fm is indicated to be ~1,350 Ma[103].

The Hongshuizhuang Fm consists of dolomitic carbonate interlayered with black to dark grey mudstones. Redox elemental analyses (Mo, Cr, U) revealed ferruginous to intermittently euxinic bottom waters during the deposition of the formation[104]. The overlying Tieling Fm starts with silty dolostone and shale deposits with occasional manganese dolostone in the basal Daizhuangzhi. The deposits of the overlying Laohuding member comprise limestone deposits with increasing stromatolitic features up-section, suggesting a shallowing basin[105].

The Xiamaling Fm displays a transgressive-regressive system with the lowest member (Member 1) recording ferruginous sandstone, sandy mudstones, and mudstones. Member 2 mainly contains green and dark grey mudstones with marly concretions. Member 3 is dominated by black shale, while the highest member (Member 4) consists of dark mudstones interlayered with stromatolitic carbonate units[106]. (Note that ref[67] divides the Xiamaling Fm into 6 units with a different numbering system). The organic rich shale horizons of the Xiamaling Fm, with TOC values up to 20%[67], were presumably deposited in a restricted basin with varying degrees of connection to the open ocean[107]. During the deposition of the organic rich horizons, the basin was severely restricted, while the geochemical characteristics of surrounding shale and silty mudstone units are interpreted as a marine system with persistent euxinic waters[106,107]. Others suggest that the Xiamaling Fm was predominantly deposited under deep oxygenated waters[67,108], although the interpretation of the exact redox conditions remains debated[109].

The fossil assemblages of the Xiamaling contain a large variety of palynomorphs, both of bacterial and eukaryotic origin. The majority of microfossils are found in Member 1 of the Xiamaling Formation preserved within dark grey and greyish-green shale, and silty shale interbedded with thin-bedded iron-rich siltstone horizons[110]. Preserved microfossils of possibly to certainly eukaryotic origin in the Xiamaling Fm include *Leiosphaeridia*, *Pterospermospimorpha insolita*, *Simia annulare*, *Satka favosa, Dictyosphaera macroreticulata*, *Germinosphaera bispinosa* and *Valeria lophostriata*[110].

1.5. Late Palaeoproterozoic Barney Creek Formation (BCF), McArthur Basin, Australia (1,640 Ma)

The McArthur Basin is a > 180,000 km^2 multiphase basin in northern Australia. Late Paleoproterozoic sediments of the Glyde Package were deposited across the McArthur Basin during north-south trending extension and active subsidence of the Northern Australian Craton (NAC). As such, the basin is characterised by paleohighs and structurally controlled sub-basins bounded by NNE-striking faults and south-dipping growth faults[111-113].

The geology of the Barney Creek Fm was summarized by Jackson et al. (1987)[90] and Ahmad et al. (2013)[114] and a recent facies analysis and sequence stratigraphy was provided by Kunzmann et al. (2019)[111]. The Barney Creek Fm, as part of the Umbolooga Subgroup of the McArthur Group, is a regionally extensive but poorly outcropping dolomitic siltstone-dominated sedimentary rock unit. The Barney Creek Fm comprises mainly dolomitic, carbonaceous and pyritic siltstones and shales, locally with abundant tuff beds, breccias and graded units[111,114]. Three tuff beds from the lower Barney Creek Fm (HYC Pyritic Shale Member) yielded U-Pb ages of 1,638 ± 7 Ma, 1,639 ± 3 and 1,640 ± 3 Ma[115]. Thickness of the Barney Creek Fm varies from ~10 m on paleohighs, to ~900 m in local subbasins such as Glyde River (drill core GRNT-79-7 = GR7), or ~160 m at the Leviathan anomaly (drill core LV09001 = LV09) to the east of the Emu Fault. Major non-carbonate components of the Barney Creek Fm were reported as microcrystalline K-feldspar, terrigenous silt and clay, as wells as locally abundant sulphides, carbonaceous and bituminous material[114,116].

Deposition is inferred as having occurred mostly under subtidal to slope conditions with local shoaling to shallow subtidal environments on paleohighs. The Barney Creek Fm comprises three members, the W-Fold shale, HYC pyritic Shale and Cooley Dolostone (only along fault scarps) that are overlain by the undifferentiated upper part of the formation. Jointly with the conformably overlying Reward Dolostone, the Barney Creek Fm comprises two 3rd-order transgressive-regressive stratigraphic sequences. A maximum flooding surface is located in the base metal prospective HYC Pyritic Shale Member that constitutes the most pyritic and organic matter-rich interval deposited during basin deepening, while another maximum flooding surface is located in the upper undifferentiated Barney Creek Fm[111]. Early interpretations of depositional environments ranged from shallow marine to lacustrine or lagoonal[111], but the HYC Pyritic Shale Member is generally regarded to have formed under deep subtidal conditions[117,118]. Nevertheless, similarities in carotenoid biomarker signatures between the undifferentiated Barney Creek Fm[30] and the saline lacustrine Eocene Green River Fm, were used to infer lacustrine depositional conditions[119]. However, redox stratification, generally low sulphate concentrations in Palaeoproterozoic marine basins[120] and the intracratonic setting of the BCF could also explain similarities in carotenoid

biomarker patterns. Fe-speciation indicates dominantly anoxic ferruginous redox conditions[121,122], even though sulfidic excursion may have been common with possible temporary injection of oxic waters[122].

Abundant fossil pigments of green (Chlorobiaceae) and purple (Chromatiaceae) sulfur bacteria indicate a shallow chemocline and an important ecological role of anoxygenic phototrophic primary producers, while indigenous algal steranes were not discovered[30,122]. Despite the lack of saturated steranes, the aromatic fractions of Barney Creek Fm rock extracts were previously found to contain high abundances of triaromatic steroids with an enigmatic dominance (> 90%) of steroids methylated at the C-4 position[30]. As a dominance of 4-methylated steroids without side-chain alkylation was mainly known from certain methanotrophic bacteria, these were tentatively proposed as the most likely source organisms, although the lack of saturated sterane equivalents hampered reconstructions of biological precursors and source organisms in previous studies. Most putative microfossils described from the Barney Creek Fm appear to be bacterial, although some ambiguous, larger and more complex structures have been interpreted as putative eukaryotic fossils[123].

After the underlying, 1,730 Ma Wollogorang Fm[124], the Barney Creek Fm contains the probably oldest-known, clearly indigenous biomarkers[125] and may constitute the oldest active petroleum system[126]. Samples analysed in the present study come from the thermally least mature sections in the basin, the upper 200 m of drill core GR7 and drill core LV09. The thermal maturity falls into the early oil window-range, with Tmax ranging from ~435 to ≤440°C in the upper 200 m of GR7[127], whereas biomarkers show an even better thermal preservation throughout core LV09 where Tmax values range from 423 to 440°C[128]. One additional sample (12Z083, 103.3 m depth) used to evaluate the identity of cyclosterane, contains strongly biodegraded migrabitumen filling a calcite vug in the Coxco Dolomite of drill core MY4 collared in the Myrtle area in the southern McArthur Basin, but this bitumen was likely sourced from the Barney Creek Fm and was subsequently biodegraded in the permeable carbonate reservoir.

2. Supplementary Methods

2.1. Generation of protosterol, ursterol and euphol pyrolysis products

2.1.1. *Pyrolysis of cycloartenol and lanosterol*

Identification standards for saturated and aromatic hydrocarbon derivatives of protosterols were generated by pyrolysis of commercially available cycloartenol (by GC-MS full scan: 95% cycloartenol, 5% 24-methylene cycloartenol, traces of cycloartenol isomers, no detectable 4- or 14-demethylated products, no detectable crown sterols; abcr GmbH & Co. KG, cat AB 165530) and lanosteroid mixture (by GC-MS full scan: isomers of lanosterol, lanostenol and lanostadienols,

impurities of lanosteroid oxidation products, no detectable 4- or 14-demethylated products, no detectable crown sterols; abcr GmbH & Co. KG, cat AB 137660). Similar to activated carbon pyrolysis procedures described in the literature[129,130], ~10 mg of activated carbon and ~1 mg of standard were inserted into a ~3 mm i.d. glass tube and sealed under vacuum. The tubes were heated at 300°C or 330°C for 18 h and after cooling to room temperature, scored and opened. Pyrolysis products and active carbon were transferred with DCM into a glass wool plugged silica pipette and extracted with ~4 mL *n*-hexane and DCM. After removal of the solvent under a stream of purified nitrogen gas, products were re-dissolved in *n*-hexane for GC-MS analyses. Half of the pyrolysates were hydrogenated by stirring for 1 h in the presence of PtO_2 catalyst under a continuous flow of H_2 gas, before capping and stirring for another 3 h.

2.1.2. *Pyrolysis of a cycloartenol/24-methylenecycloartenol mixture*

To obtain elution positions and mass spectra, 24-methylated protosteroids were generated by pyrolysis of rice bran powder (Lotus Foods, Australia) containing subequal amounts of cycloartenol and 24-methylenecycloartenol (among other sterols). 150 mg of rice bran powder were placed into baked glass tubes (200 mm long, 6.3 mm OD and 4.1 mm ID) sealed at one end. A glass fibre plug was placed on top of the bran to prevent dispersion of the powder. To remove air, the tubes were evacuated and purged with nitrogen gas three times and then sealed under vacuum. The 160 mm long pyrolysis tubes were inserted into a custom-built heating block so that the upper 70 mm remained at room temperature outside the heating zone to allow condensation of water and pyrolysis products. The heating block was set to an initial temperature of 300°C and then rapidly ramped to 500°C. Tubes were removed after 5 to 15 minutes, cooled, scored and opened. Pyrolysis products were extracted with dichloromethane (1 mL, 3 times). After removal of the solvent under a stream of purified nitrogen gas, the pyrolysis products were taken up in *n*-hexane (2 mL) and hydrogenated with H_2 gas on ~20 mg PtO_2/C under constant stirring for 12 hours. *n*-Alkanes were removed from the rice bran pyrolysate by filtration over silicalite and elution with pentane[131].

2.1.3. *Pyrolysis of euphenol (24(25)-dihydroeuphol)*

Euphanes were generated by pyrolysis of ~1 mg euphenol (24(25)-dihydroeuphol, purity > 95% by GC), obtained from the natural products reference collection of late Prof Ourisson, University of Strasbourg, France, following the protocol in Section 2.1.2. Saturated+unsaturated hydrocarbons were separated from aromatic and polar pyrolysis products over silica gel using *n*-hexane as eluent, followed by hydrogenation of the saturated+unsaturated hydrocarbon fraction on PtO_2/C as described above. Pyrolysis products were characterized by GC-MS in full scan and MRM mode on a DB5-MS column (Extended Data Fig. 3b, c).

2.2. Synthesis of authentic C_{29} monoaromatic lanosteroid (MAL)

The monoaromatic lanosteroid was synthesized from lanost-8-ene following an established method[132] for the synthesis of a monoaromatic fernene derivative starting from fern-9(11)-ene. Lanost-8-ene was first heated with elemental sulfur (200°C, 5 h). The crude mixture obtained after removal of the elemental sulfur using activated copper was sequentially treated with Raney nickel and hydrogenated over Pd/C under H_2 atmosphere. The monoaromatic lanosteroid was purified from the crude mixture by column chromatography on a silica gel column using *n*-hexane as eluent.

GC-MS (EI, 70 eV): *m*/*z* (relative intensity) 394 (15%), 379 (100), 267 (17), 253 (12), 239 (20), 225 (33), 213 (55). 199 (19), 169 (11), 155 (11), 83 (7).

The structure of the synthesized compound (Extended Data Fig. 1, 20R MAL) was confirmed by mass spectrometry and NMR studies comprising 1D (^{1}H, ^{13}C, DEPT) and 2D homo (^{1}H-^{1}H COSY) and heteronuclear (^{1}H-^{13}C-HSQC, and ^{1}H-^{13}C-HMBC) experiments. NMR analyses were performed on a Bruker Avance I – 500 MHz spectrometer operating at an observation frequency of 500 MHz (^{1}H) and 125 MHz (^{13}C) (^{1}H, ^{13}C, DEPT,^{1}H-^{13}C-HSQC, and ^{1}H-^{13}C-HMBC) and on a Bruker Avance III – 600 MHz spectrometer at an observation frequency of 600 MHz (^{1}H) (^{1}H, ^{1}H-^{1}H COSY). The chemical shifts (Supplementary Table 4) are reported in ppm relative to tetramethylsilane with the solvent used as internal standard (CD_2Cl_2: δ^1H 5.32 ppm; δ^{13}C 53.5 ppm).

Supplementary Table 4. ^{1}H and ^{13}C-NMR chemical shifts of synthetic monoaromatic lanosteroid (20R MAL) (CD_2Cl_2).

Position	δ_C	δ_H	δ_H	Position	δ_C	δ_H	δ_H
1	27.4	2.51	2.51	16	28.3	2.10	1.45
2	19.5	1.81	1.81	17	50.6	1.65	-
3	38.7	1.61	1.61	18	-	-	-
4	33.6	-	-	19	15.6	0.62	-
5	142.3	-		20	36.4	1.45	-
6	123.4	7.12		21	18.6	0.96	
7	122.2	6.83		22	36.4	1.40	1.05
8	144.2	-		23	24.1[c]	1.40	1.19
9	133.1	-		24	39.5	1.16	1.16
10	133.8	-		25	28.0	1.51	-
11	24.0[c]	2.62	2.62	26	22.6[a]	0.880[b]	-
12	30.9	1.98	1.98	27	22.3[a]	0.876[b]	-
13	43.9	-	-	28	31.8	1.26	-
14	50.1	-	-	29	31.8	1.26	-
15	32.1	1.82	1.70	30	27.9	1.02	-

a, b, c: can be exchanged

2.3. Elucidation of the possibility of double bonds and cyclopropyl moieties in cyclosterane

Aliquots of the saturate fraction of the biodegraded oil from drill core MY4 (section 2.6.1) were subjected to experiments in order to evaluate the possible presence of a double bond or cyclopropyl moiety in cyclosterane (*k1* and *k2*, Extended Data Fig. 2b). In the first experiment, an aliquot of the bitumen was hydrogenated by stirring for 3 h in the presence of PtO_2 catalyst and a constant stream of H_2 gas, followed by extraction of products with 1 mL *n*-hexane and 3 mL DCM. Subsequently, the hydrogenation products on ~7 mg active carbon were transferred into a glass tube that was evacuated and sealed, followed by pyrolysis at 300°C for 18h. After extracting the active carbon twice with 2 mL *n*-hexane and 2 mL DCM, the pyrolysis products were hydrogenated by stirring for 30 min in the presence of PtO_2 catalyst and a constant stream of H_2 gas. In the second experiment, an aliquot of bitumen was subjected to acidic conditions for 30 minutes by stirring it in 2 mL DCM acidified with 10 µL concentrated HCl dissolved in 50 µL methanol, followed by removal of the solvent under a stream of N_2 gas. The products were taken up in *n*-hexane and hydrogenated first for 4 h with H_2 gas on PtO_2, and then on Pt/C for 21 h. In the third experiment, an aliquot of bitumen was hydrogenated under acetic acid conditions under a constant stream of H_2 gas by stirring for 3 h in 50 µL acetic acid and 2 mL DCM in the presence of PtO_2 catalyst, before the vial was capped and stirred for another 69 hours. The same experiment was also conducted on cycloartenol. All hydrogenation products were extracted by ultrasonication (5 min) in DCM, followed by extraction with 1.5 mL *n*-hexane and 1.5 mL DCM. In a fourth experiment, HCl gas was reacted with cycloartenol, cycloartane (BOC Sciences, NY, USA, CAS# 511-64-8) and an aliquot of the saturate fraction of the MY4 biodegraded bitumen. HCl gas was let through a capillary into a 4 mL vial filled with the organic substrates dissolved in 2 mL DCM. After being flushed several times with HCl gas, the reactant vial was capped and stirred overnight. To remove HCl, the DCM was washed with water (4 times) and dried over $NaSO_4$.

2.4. RuO_4 degradation of a fraction containing k1 *and* k2

A small fraction containing *k1* and *k2* (neat) was treated with a few drops of a solution of RuO_4 freshly prepared according to Piatak et al. (1969)[133], but replacing acetone by CCl_4. After 10 min, a mixture of CCl_4 and isopropanol (1 ml) was added in order to destroy the excess of reagent. The solvents were removed under a flow of argon, the crude mixture was filtered on a short silica gel column using dichloromethane as eluent and analysed using GC-MS.

2.5. Mass spectrometer parameters for the detection of (proto)steroids and hopanoids

Metastable Reaction Monitoring (MRM) precursor → product transitions for the measurement of saturated steranes and hopanes is given in Supplementary Table 5 (total cycle time 1362.1 ms), MRM precursor → product transitions for the measurement of saturated protosteranes and ursteranes in Supplementary Table 6 (total cycle time 962.9 ms), and Selected Ion Recording (SIR)

mass to charge (*m/z*) ratios for the measurement of aromatic protosteroids, steroids and hopanoids in Supplementary Table 7 (total cycle time 1280.1 ms).

Supplementary Table 5. MRM parameters for saturated steranes and hopanes.

Transition	Precursor	Product	Dwell [ms]	Delay [ms]
1	358.35	217.20	10.0	40.0
2	358.35	217.20	25.0	40.0
3	414.42	231.21	25.0	21.6
4	372.38	217.20	20.0	21.4
5	386.39	217.20	20.0	22.8
6	404.41	221.22	15.0	20.7
7	400.41	217.20	20.0	22.0
8	414.42	217.20	25.0	22.5
9	426.42	205.19	20.0	29.8
10	370.36	191.18	20.0	20.0
11	440.44	205.19	20.0	22.1
12	384.37	191.18	25.0	20.3
13	454.45	205.19	20.0	21.7
14	398.39	191.18	20.0	20.6
15	468.47	205.19	25.0	21.3
16	412.41	191.18	20.0	20.9
17	482.48	205.19	100.0	21.0
18	426.42	191.18	20.0	21.1
19	496.50	205.19	150.0	20.7
20	440.44	191.18	20.0	21.3
21	454.45	191.18	20.0	21.9
22	398.39	177.16	20.0	21.2
23	468.47	191.18	20.0	20.6
24	482.48	191.18	20.0	21.7

Supplementary Table 6. MRM parameters for protosteranes and ursteranes.

Transition	Precursor	Product	Dwell [ms]	Delay [ms]
1	412.40	288.29	10.0	40.0
2	412.41	288.29	40.0	20.0
3	426.40	288.29	40.0	23.1
4	412.41	274.27	60.0	26.1
5	414.40	259.25	40.0	30.3
6	400.40	245.24	30.0	26.4
7	414.40	245.25	30.0	22.8
8	386.38	231.21	30.0	23.8
9	400.40	231.21	30.0	22.8
10	414.42	231.21	30.0	22.6
11	372.37	217.20	10.0	21.4
12	386.39	217.20	10.0	22.8
13	404.41	221.22	10.0	20.7
14	400.40	217.20	10.0	22.0
15	414.41	217.20	30.0	22.5
16	370.40	191.20	10.0	50.0

Supplementary Table 7. SIR parameters for aromatic (proto)steroids and hopanoids

Transition	*m/z*	Dwell [ms]	Delay [ms]
1	416.34	20.0	40.0
2	416.34	20.0	20.0
3	414.33	20.0	20.7
4	404.34	20.0	23.3
5	402.33	20.0	20.7
6	393.35	20.0	23.0
7	379.34	20.0	24.8
8	375.31	20.0	21.4
9	365.32	20.0	23.5
10	361.29	20.0	21.4
11	351.31	20.0	23.5
12	347.27	20.0	21.5
13	340.33	20.0	22.5
14	267.21	20.0	140.0
15	367.21	20.0	20.0
16	359.15	20.0	23.3
17	253.20	20.0	22.5
18	245.13	20.0	23.4
19	231.12	20.0	26.1
20	217.20	20.0	20.0
21	212.14	10.0	40.0
22	191.18	10.0	40.0
23	191.18	20.0	20.0

2.6. Characterization of saturated protosteranes

The GC elution behaviour, mass spectra and structures of saturated protosterol derivatives were investigated on thermally well preserved bitumens from the Barney Creek Fm from the upper ~200 m of drill core GR-7 (Extended Data Table 1, and additional samples B03163, GR7, 47.55 m; 12Z083, MY4, 103.3 m) and on strongly biodegraded migrabitumen filling a calcite vug in the Coxco Dolomite of drill core MY4 collared in the Myrtle area in the McArthur Basin (12Z083). In drill core MY4, the Coxco Dolomite directly underlies the BCF, and the solid migrabitumen filling the vug was likely sourced from the BCF or deeper units.

2.6.1. *Cyclosterane (I)*

Extended Data Figure 2b shows the total ion chromatogram of the saturated fraction of the migrabitumen filling the calcite vug in drill core MY4. The bitumen is strongly biodegraded with a very prominent unresolved complex mixture (UCM) and absence of *n*-alkanes and acyclic isoprenoids pristane and phytane. Hopanes only occur in traces. The aromatic fraction (not shown) is also dominated by a UCM. Triaromatic steroids (TAS), particularly abundant in most BCF

bitumens, were below detection limits. TAS are generally among the most biodegradation-resistant compounds, and their absence places the migrabitumen into biodegradation category 10, the highest rank[134]. Despite severe biodegradation, the saturated fraction shows two dominant, well-resolved chromatographic signals (*k1* and *k2*) with nearly identical mass spectra containing molecular ion [$M^{+\circ}$] at *m/z* 412 and base ion *m/z* 274 (Extended Data Fig. 2d). *k1* and *k2* were subsequently also detected in lower relative concentrations in all thermally well preserved BCF bitumens, suggesting that the compounds were enriched from the original migrated oil through biodegradation and that they possess unusual structural features placing them among the most biodegradation resistant hydrocarbons known to date. The molecular mass of 412 Da points to C_{30} compounds with five double bond equivalents (DBE), i.e. with a total of five rings or double bonds (compare also with peaks P in ref[125]).

Due to the unusual enrichment of the unknown compounds in the biodegraded migrabitumen, it was possible to detect a pseudo-homologous series of side-chain cleavage products *k3* to k*11* with 22 to 29 carbon atoms, corresponding to the sequential loss of up to 8 carbon atoms from the hydrocarbon core (Extended Data Fig. 2a). However, two members of the pseudo-homologous series with molecular ions [$M^{+\circ}$] at *m/z* 384 (C_{28}) and 314 (C_{23}) are close to detection limits (marked 'x'), pointing to two branching positions in the side chain. In total, the pseudo-homologous series reveals a 6-methylhept-2-yl side chain, the typical C_8 side chain of sterols such as cholesterol, lanosterol or cycloartenol. Moreover, all pseudohomologs from *k1* to *k8* show two isomers with nearly identical mass spectra, while *k10* and *k11* are single peaks. This likely indicates that the signal pairs represent S and R isomers with a chiral centre in the side chain marked '*' in Extended Data Figure 2a, another characteristic of protosterol side chains.

Extended Data Figure 2c, d shows the mass spectra of chromatographic signals *k1* to *k11*. The dashed lines follow particular fragments through the pseudo-homologous series, whereby constant mass throughout the pseudo-homologous series (red) indicates that the fragment belongs to the 'left-hand side' of the compounds (i.e. the polycyclic core), whereas successive shifts of *m/z* 14 (blue) indicate fragments belonging to the right-hand side (side-chain including fragments). The main side-chain including fragments of the inferred parent compounds *k1* and *k2* ([$M^{+\circ}$] 412) are *m/z* 397, 288, 274, 259 and 205. By contrast, the *m/z* 299 fragment is constant in *k1* to *k10* and is represented by *m/z* 300 in the last member of the series *k11*. The gain of one mass unit in *k11* indicates that no further methyl group is lost, i.e. substitution with hydrogen instead of an alkyl-group. This confirms that the *m/z* 299 fragment corresponds to a polycyclic core with 22 carbon atoms and five DBE attached to a typical saturated C_8 sterol side chain as, for instance, found in cycloartane, the hydrocarbon equivalent of cycloartenol. A comparison of the spectra of *k1* and *k2* with the NIST MS database yielded a 63% agreement with cycloartane **(**Extended Data Fig. 2d), albeit with several slightly shifted masses and different relative intensities. *k1* also elutes ~1.5

minutes earlier than commercially available cycloartane (CAS# 511-64-8, BOC Sciences, NY, USA) under the described chromatographic conditions (Methods). The similarity of the mass spectra suggests that *k1* and *k2* are closely related structural isomers of cycloartane.

In comparison to typical steranes that possess a core with four rings, compounds *k1* to *k11* possess an additional DBE in the ring system. The fifth DBE may represent a double bond (as in lanosterol), a cyclopropyl ring (as in cycloartenol) or a bridge across an existing ring (as found in some geological oleanane-type triterpenoids[135]). Non-aromatic double bonds are not particularly stable in sedimentary environments and commonly do not persist over geologic periods of time in systems as mature as the BCF. Moreover, unsaturated hydrocarbons are commonly more prone to biodegradation than saturated counterparts, contradicting the resistance of *k1* to *k11* to biodegradation. However, to test for the presence of a sterically highly protected double bond, the biodegraded migrabitumen was subjected to prolonged catalytic hydrogenation (SI Methods). We found that hydrogenation on platinum catalyst for several hours is commonly sufficient to even hydrogenate many aromatic hydrocarbons. However, *k1* and *k2* survived such catalytic treatment even under the harshest conditions, hydrogenation on PtO_2 and Pt/C for more than 20 h and under acidification with concentrated HCl (Section 2.3). The fact that *k1* and *k2* survived this treatment argues against the presence of a double bond. The potential presence of a sterically hindered double bond was further tested under yet harsher conditions using RuO_4 (Section 2.4). *k1* and *k2* survived this treatment, strongly arguing against the presence of a double bond, which would otherwise have been oxidized. The combined experiments thus identify *k1* to k*11* as saturated pentacyclic hydrocarbons.

Cyclopropyl rings are not stable under geological conditions and have, to our knowledge, never been observed in mature sedimentary rocks. However, to test for the presence of a cyclopropyl moiety, the biodegraded oil was subjected to strong heating, acidified PtO_2 hydrogenation and treatment with HCl gas (Section 2.3). Pyrolysis of cycloartenol at 300°C for 18 hours in an evacuated silica tube resulted in quantitative conversion, indicating that cyclopropyl-rings are efficiently opened under these conditions. However, *k1* and *k2* survived this treatment. In a second experiment, cycloartenol and *k1* and *k2* in the biodegraded bitumen were subjected to 70 hours of acidified (concentrated acetic acid in DCM) PtO_2-catalyzed hydrogenation. Under this treatment, the cyclopropyl moiety of cycloartenol was quantitatively opened and the final product was lanostanol, whereas *k1* and *k2* again survived the treatment. In a further test, HCl gas was reacted with cycloartenol, cycloartane and *k1* and *k2* in the biodegraded oil. The reaction of cycloartenol with HCl gas quantitatively opened the three-ring, leading to the formation of compounds with lanostenone-like mass spectra. The experiment was repeated using commercially available cycloartane, resulting in quantitative conversion to lanostenes, thus confirming that HCl gas

treatment also opens cyclopropyl moieties of otherwise unfunctionalized hydrocarbons. However, *k1* and *k2* in the biodegraded oil again survived the same HCl gas treatment.

The combined experiments suggest that *k1* and *k2* do not possess a double bond or cyclopropyl ring, and the only remaining plausible option for the fifth DBE is a bridge across one of the four rings, a bicyclo-structure. This assignment is supported by the observation of bicyclo-structures in oleanane-type triterpanes that presumably formed through geological rearrangement processes[135]. We thus suggest the informal name '**cyclosterane**' for *k1* and *k2*. The bridge structure may explain the unusual resistance of these compounds to biodegradation.

The position of the bridge in cyclosterane remains unknown, but the mass spectra of *k1* to *k11* (Extended Data Fig. 2c, d) are most consistent with a bridge across ring C, such as structure **I** (Extended Data Fig. 1, 2b). The mass spectrum of cyclosterane is similar to cycloartane, suggesting that the compounds are structurally related and that the differences in spectra may provide clues to structural distinctions. Notable differences in the mass spectra of *k1* and *k2* versus cycloartane are shifts from *m/z* 121 in cycloartane to *m/z* 123 in cyclosterane, from *m/z* 135 to *m/z* 137 and from *m/z* 203 to *m/z* 205 (Extended Data Fig. 2d). The *m/z* 121 fragment of cycloartane represents a C_9H_{13} left-hand fragment with 3 DBE that encompasses the A-ring and is generated by cleavage through the B-ring as well as the cyclopropyl ring (Extended Data Fig. 2e). *M/z* 121 is thus indicative of the attachment of the cyclopropyl ring to the A and B rings. By contrast, the *m/z* 123 fragment of *k1* to *k11* represents a C_9H_{15} left-hand fragment with 2 DBE that encompasses the A-ring and is generated by cleavage through the B-ring, but indicating that there is no bridge across ring A or a ring or bridge attached to any of the carbon atoms of ring A. According to the degradation series in Extended Data Figure 2d, the *m/z* 205 fragment of cyclosterane represents a $C_{15}H_{25}$ right-hand fragment with 3 DBE that includes the D-ring and is generated by cleavage through the C-ring and one additional DBE that tentatively points to a bridge across ring C that is attached to ring D, for example structure **I** (although other bridge configurations and substitution patterns are possible). Extended Data Figure 2e shows possible MS fragmentation patterns that we would expect from hypothetical structure **I**, although other structural isomers may also explain the data.

In mid-Proterozoic bitumens, lanostane and cyclosterane are the only detected saturated hydrocarbon products of sterols. However, cyclosterane was also detected in low percentage levels (relative to total steranes) in virtually all Phanerozoic bitumens and oils that were investigated using MRM 412 → 274 transitions (Supplementary Table 6) and yielded conventional steranes. Cyclosterane is thus the diagenetic product of a biological compound that persisted through geological time but is far less abundant than crown-steranes. Based on the inferred structure of cyclosterane, likely precursors are sterols with the cycloartane skeleton such as cycloartenol.

Cycloartenol occurs in low percentage levels as a biosynthetic intermediate in a large proportion of extant eukaryotes, including all plants and algae (Supplementary Text). While cyclosterane was not observed in our pyrolysis experiments on cycloartenol and cycloartane, low-temperature diagenetic opening of the cyclopropyl-ring may plausibly lead to the stabilization of the resulting carbocation or radical by intramolecular formation of a carbon bridge. The fact that lower temperatures are sufficient, and possibly even required, for bridge formation in geological triterpenoids is illustrated by the presence of methylene-bridge containing oleanane derivatives in notable quantities in immature (R_o 0.36–0.41%) sedimentary samples[135]. In this case, the bridged oleanane compound most likely derives from a specific triterpenic precursor functionalized at the C-27 methyl group. Formation of the additional ring can be explained by early diagenetic processes likely to occur under low temperature conditions and leading first to a hexacyclic functionalized oleanoid that is further diagenetically reduced into an oleanane saturated hydrocarbon with a methylene-bridge. Similarly, it can also be envisaged that the cyclosterane skeleton is formed from functionalized cycloartenol-related triterpenes at the earliest stage of diagenesis. This process would first lead to a functionalized compound with the cyclosterane skeleton which is further reduced into the geological cyclosterane hydrocarbons. Since cyclosteranes are not necessarily formed by diagenetic rearrangement of a triterpenoid saturated hydrocarbon already having five rings, the possibility that the cyclosterane skeleton resulted from early diagenetic transformation of lanosterol-related precursors rather than from cycloartenol cannot be excluded. In this process, the double bond of lanosterol would aid formation of the additional ring.

Synthesis of authentic standards, purification and NMR experiments (if sufficient quantities of cyclosterane-containing oils or bitumens can be obtained in the future) and/or diagenesis experiments on cycloartenol and related compounds are needed to confirm the exact structure and biological precursors of cyclosterane **I**.

2.6.2. *Lanostane (II)*

In 1989, Chen and colleagues[136] described the first geologic record of lanostane **II** in Eocene sample Y2 (2036 m) from the Biyang Basin, China. They confirmed the structure through a co-elution experiment with authentic 8β(H),9α(H)-lanostane (95% purity, Chiron). An aliquot of original sample Y2 (courtesy J.H. Chen, Geoscience Australia) was used to evaluate the presence of lanostane in the BCF. A co-injection experiment of Y2 on a 60 m DB-5 MS capillary column with a BCF sample (B03178, GR7, 199.08 m depth) and comparison of mass spectra confirmed the presence of 8β(H),9α(H)-lanostane in the BCF (Extended Data Fig. 3a, signal *l4*). However, many BCF bitumens show three additional chromatographic signals *l1* to *l3* with mass spectra similar to *l4* (Fig. 9-26 in Nettersheim, 2017[122]), suggesting existence of further stereo or structural isomers (Extended Data Fig. 3b). To evaluate the origin of these compounds, we performed

artificial maturation experiments on lanosterol and cycloartenol through pyrolysis at 300ºC on active carbon followed by hydrogenation with H_2 gas on PtO_2. These experiments yielded a wide array of compounds with the molecular mass of lanostane (414 Dalton) and mass spectra similar to lanostane (Fig. 9-30 in Nettersheim, 2017[122]), including signals corresponding to the elution positions of *l1* to *l4* (Extended Data Fig. 3b). The pyrolysis experiments show that lanosterol as well as cycloartenol are suitable biogenic precursors of 8β(H),9α(H)-lanostane *l4* and potentially some of the other BCF isomers with lanostane-type MS fragmentation patterns *l1* to *l3*, although the exact identity and origin of these additional compounds in the BCF has not yet been unambiguously established.

2.7. Characterization of aromatic protosteroids and ursteroids

Previous studies reporting aromatic steroids nearly exclusively focused on derivatives of crown-sterols such as cholesteroids, ergosteroids, stigmasteroids and dinosteroids. However, in contrast to the Phanerozoic where such compounds are ubiquitous and abundant, these crown-sterol derivatives are below detection limits in mid-Proterozoic bitumens, with two exceptions. Firstly, mid-Proterozoic bitumens that contain abundant protosterol degradation products also commonly yield traces of C_{26} triaromatic steroid **IX**. These compounds may form as aromatization products of C_{27} crown-sterols. However, as the corresponding saturated cholestanes and monoaromatic cholesteroids always remain below detection limits, it is likely that **IX** in mid-Proterozoic sediments formed by aromatization and demethylation of protosteroid precursors (Extended Data Fig. 1). Secondly, solvent extracts from 1,300 Ma old sediments of the Roper Group yielded traces of triaromatic ergosteroid **XII**. However, as with **IX**, the corresponding saturated and monoaromatic ergosteroids remain below detection limits. Moreover, the 4-methylated counterpart **XI** is more abundant than **XII**, suggesting that the precursor was a 4,24-dimethylated ursterol such as 24-methylenecycloartenol or a 4- and/or 14-demethylated downstream products (see Section 2.7.2.1).

In this study, we identified a large variety of additional aromatic steroids that are likely derived from ursterols and/or protosterols. They were identified based on their elution behaviour and mass spectra in comparison to the literature, authentic standards, reference oils and/or laboratory pyrolysis products of protosterols.

2.7.1. *Triaromatic steroids (TAS)*

In this study we identified triaromatic steroids without ring methylation (**IX**), with one methyl group at ring-A (almost exclusively at position C-4, **VIII**, among the most abundant compounds in some mid-Proterozoic bitumens) and TAS with two methyl-groups attached to the aromatic core (**VII**). Electron impact (EI+) mass spectra of TAS are strongly dominated by a single mass

fragment comprising the aromatized polycyclic ring system and are best identified and quantified using base ions *m/z* 231, 245 and 259 respectively.

2.7.2. *TAS without methyl groups at the ring system (m/z 231)*

TAS in *m/z* 231 mass chromatograms were identified through comparison with the well-characterized Phanerozoic-based AGSOSTD oil reference standard[46] (Extended Data Fig. 7i). Extended Data Figure 7g shows abundant S and R isomers of **IX** in the ~750 Ma Chuar Group where they presumably derive from C_{27} crown-sterols as they co-occur with abundant saturated cholestane and MAS cholesteroids[46]. By contrast, in the mid-Proterozoic they only occur in traces, as highlighted for the 1,100 Ma El Mreïti Gr in Extended Data Figure 7h, and are interpreted as demethylation products of 4-methyl TAS **VIII**, which are abundant in such samples (Extended Data Fig. 7e). TAS cholesteroids **IX** were also generated in traces in protosterol pyrolysis experiments (not shown). In the mid-Proterozoic, 24-methylated TAS **XII** were exclusively detected in traces in the 1,300 Ma Roper Group (Supplementary Table 2).

2.7.2.1. A-ring methylated TAS (m/z 245)

A-ring methylated TAS were identified in *m/z* 245 mass chromatograms through comparison with AGSOSTD[46]. In the AGSOSTD, the distribution of cholesteroid-, ergosteroid- and stigmasteroid-type TAS with methylation at positions C-2, C-3 or C-4 of the aromatic ring system, as well as 4,23,24-trimethylated TAS (dinosteroids) is shown in Extended Data Fig. 7f.

The ~750 Ma Chuar Group contains abundant 2-, 3- and 4-methylated TAS cholesteroids, but no detectable ergosteroids or stigmasteroids (Extended Data Fig. 7d), a pattern that mirrors the saturated and monoaromatic steroid fractions and is typical for the Tonian period[15,46]. These Tonian TAS are interpreted as A-ring methylation products of C_{27} crown-sterols. By contrast, mid-Proterozoic bitumens only contain the 4-methylated TAS isomer **VIII**, exemplified by the 1,100 Ma El Mreïti Gr in Extended Data Figure 7e. The only exception is some bitumens of the 1,640 Ma Barney Creek Fm where traces of 2- and 3-methlyated isomers were also detected (Supplementary Table 2), interpreted as isomerization products of the 4-methyl isomers, which are one to two orders of magnitude more abundant (Supplementary Table 2). Pyrolysis experiments on lanosterol and cycloartenol confirm that 4-methylated TAS **VIII** (but not 2- or 3-methylated TAS) are generated by thermal degradation of protosterols (red-coloured signals in Extended Data Fig. 7c).

In the mid-Proterozoic, 4,24-dimethylated TAS **XI** were detected in the 1,300 Ma Roper Group (Supplementary Table 2). The identity of these compounds was verified by comparing elution positions in *m/z* 245 mass chromatograms on two GC-columns with different polarities (DB-5MS

and VZ-200) relative to AGSOSTD (Extended Data Fig. 9). On the DB-5MS column, the 20R isomer of 4-methylated TAS cholesteroid **VIII** coelutes with the 20S isomer of the 4,24-dimethylated TAS ergosteroid **XI**, making confirmation of the presence of the 20S/R isomer pair uncertain (Extended Data Fig. 9a, b). However, on the VZ-200 column, all isomers are chromatographically fully separated, confirming the presence of these compounds. The identity of **XI** is further supported by the unusual and broad chromatographic peak shape that is presumably caused by the presence of closely eluting 24S and R isomers. In the absence of 2- and 3-methylated isomers and of 24-methylated ergosteroids in the saturated and monoaromatic steroid fractions, **XI** are interpreted as the degradation products of 24-methylated ursterols such as 24-methylenecycloartenol, obtusifoliol or 4-methylfecosterol. Pyrolysis of 24-methylenecycloartenol confirms that 4,24-dimethylated TAS **XI** (but not 2,24- or 3,24-dimethylated isomers) can be generated during degradation of 24-methylated ursterols (blue-coloured signals in Extended Data Fig. 7c).

2.7.2.2. TAS with double-methylated ring-system (m/z 259)

Virtually all mid-Proterozoic bitumens in the data set (Supplementary Table 1) that yielded 4-methylated TAS **VIII** of the *m/z* 245 series also yielded four signals (*c1* to *c4*, Extended Data Fig. 7a) in *m/z* 259 ion chromatograms, indicative of additional methylation in the aromatic ring system. (Signals c2 and c3 coelute on DB5-MS chromatographic columns but are separated on DB1-MS columns). TAS of the *m/z* 259 series are highly abundant in mid-Proterozoic bitumens, on average 25% of the sum of all TAS. However, relative concentrations drop to 15% in the Tonian period and to less than 1% in the Ediacaran and Phanerozoic (Supplementary Table 1).

The identity of compounds *c1* to *c4* as protosteroid aromatization products was confirmed in laboratory pyrolysis experiments on a cycloartenol + 24-methylene cycloartenol enrichment mixture (Extended Data Fig. 7b) as well as on pure cycloartenol and lanosterol. The mass spectra of the pyrolysis products *C3* and *C4* showed close similarity to the mass spectra of c2/c3 and c4 in Barney Creek Fm bitumens, revealing the dominant base ion *m/z* 259 as well as a minor molecular ion [$M^{+\circ}$] at *m/z* 372 and minor fragments *m/z* 244 and 229, corroborating the presence of two methyl groups in the triaromatic ring system (Fig. 10-10 in Nettersheim, 2017[122]). By comparison with signals of the *m/z* 231 and *m/z* 245 series, and by comparison of bitumens with different thermal maturities, *c1* and *c3* are likely 20S isomers while *c2* and *c4* possess the biological 20R configuration. The precise location of the two methyl groups in *c1* to *c4* remains unknown, but it is plausible that all isomers retain one methyl group at C-4. We thus suggest the tentative structure **VII** (Extended Data Fig. 1)

2.7.3. *Diaromatic lanosteroids (DAL)*

Diaromatic steroids are rarely reported in the literature. Among the few reports are a series of 14-methyl steroids with aromatic A- and B-rings that constitute the major hydrocarbon constituents in biodegraded oil seeps in Pakistan, tentatively attributed to 14-methyl sterol precursors in green algae[137]. A second report tentatively describes A/B-ring diaromatic steroids methylated at C-4, C-13 and C-14 from sediments of Lake Caçó (Brazil). As the lake sediments were rich in aromatized higher plant-derived triterpenoids, the diaromatic lanosteroids were tentatively attributed to higher plant sources. The mass spectral interpretation of these structures was further substantiated by the presence of a methyl ether, a ketone and an alcohol derivative of lanosteroids[138]. Here, we describe the detection of diaromatic lanosteroids in Proterozoic sediments based on mass spectral characteristics and comparison to protosterol pyrolysates.

2.7.3.1. C_{28} diaromatic lanosteroids (C_{28} DAL)

In bitumens of the 1,640 Ma Barney Creek Formation, two molecules (*b1* and *b2*) with a molecular mass of 376 Da and base ion *m/z* 361 are among the most abundant aromatic compounds (Extended Data Fig. 5i). The chromatographic signals *b1* and *b2* are commonly well resolved in total ion chromatograms where they elute between the two characteristic 4-MeTAS peaks *d3* and *d6*. Signals *b1* and *b2* exhibit the same main mass spectral fragments (*m/z* 376, 361,195, 207, 221, 181, 249, 235) as the 4,13,14-trimethylated A/B-ring diaromatic lanosteroids previously identified in Lake Caçó (compare Extended Data Fig. 6 with Fig. 4 in ref[138]). In the present work, the lanosteroid structure was confirmed by pyrolysis experiments on cycloartenol, which yielded chromatographic signals *B1* and *B2* with identical mass spectra and elution positions to *b1* and *b2* (Extended Data Fig. 5g-i, 6). Therefore, *b1* and *b2* are tentatively identified as DAL structure **VI**. The high abundance of these diaromatic steroids supports the 13,14-methylation pattern as these methyl groups would hamper progressive aromatization of the C-ring[137], which is also observed for B-ring aromatized isoarborinol and fernane-derivatives[139]. It is therefore probably no coincidence that the sparse reports of diaromatic steroids are restricted to 13,14-dimethylated derivatives[137,138], as the slow pace of B-ring aromatization in crown-sterols combined with the considerable pace of A-ring and C-ring aromatization[140] usually results in the formation of either mono- or triaromatic steroid derivatives.

2.7.3.2. C_{29} diaromatic lanosteroids (C_{29} DAL)

The aromatic fraction of most mid-Proterozoic bitumens also contained a series of later eluting compounds (*b3* to *b6*) with a molecular mass of 390 Da and base ion *m/z* 375 (Extended Data Fig. 5e, f and 6). These compounds possess mass spectra with major fragments at *m/z* 390, 375, 209,

221, 195, 179, 263 and 235. In analogy to the very similar fragmentation patterns of the C_{28} DAL series, but with main fragments shifted by 14 Da to higher masses (e.g. 209 instead of 195), these compounds likely constitute a series of four C_{29} DAL with additional methylation in the ring system. Cycloartenol pyrolysis yielded chromatographic signals *B4* to *B6* with matching elution positions and mass spectra to *b4* to *b6* (Extended Data Fig. 5d and 6), confirming a diaromatic lanosteroid structure (**V**).

2.7.3.3. C_{30} diaromatic lanosteroids (C_{30} DAL)

Several mid-Proterozoic bitumens yielded a series of yet later eluting compounds *b7* to *b10* with molecular masses of 404 Da and mass spectra pointing to a C_{30} homolog of the C_{28} and C_{29} DAL series with an additional methylation in the ring system (Extended Data Fig. 5a-c and 6). Cycloartenol pyrolysis yielded signals *B7* to *B10* with identical mass spectra and elution positions to *b7* to *b10* (Extended Data Fig. 5a and 6), confirming a lanosteroid structure (**IV)**. The positions of the methyl groups remain to be investigated.

2.7.4. *B-ring monoaromatic lanosteroids (MAL)*

Previously, B-ring monoaromatic lanosteroids were only known from Oligocene evaporites in France[141]. In the present study, the identity of C_{29} B-ring monoaromatic lanosteroids was confirmed by co-injection with an authentic standard.

2.7.4.1. C_{29} B-ring monoaromatic lanosteroids (C_{29} MAL)

Almost all Proterozoic bitumens analysed in this study yielded two mass chromatographic peaks *a3* and *a4* with a molecular mass of 394 Da and base ion *m/z* 379 (Extended Data Figure 4). Compounds *a3* and *a4* have mass spectral fragmentation patterns similar to C_{29} DAL signals *b4* to *b6* (cmp. Extended Data Fig. 4f and 6), but all main fragments are shifted four mass units to heavier values indicative of one instead of two aromatic rings. The identity of *a4* was confirmed by comparison of mass spectra and coelution experiments of 1,640 Ma Barney Creek Formation sample B03163 (GR7, 47.55 m depth) with an authentic standard (section 2.2) represented by signal *A4* (20R C_{29} MAL) in Extended Data Figure 4a, b, f. Compound *a3* with identical mass spectrum was identified as the 20S isomer of *a4*, a hypothesis supported by the co-occurrence of this compound in small amounts with the 20R C_{29} MAL standard (signal *A3* in Extended Data Fig. 4b; compound with same mass spectrum and retention time as *a3*). While protosterol pyrolysates yielded several peaks with very similar mass spectra, compounds with the exact elution positions of *a3* and *a4* were not generated in detectable quantities.

2.7.4.2. C_{28} B-ring monoaromatic lanosteroids (C_{28} MAL)

Some samples also yielded two chromatographic signals *a1* and *a2* in *m/z* 365 chromatograms eluting between aromatic C_{29} and C_{30} secohopanoids (Extended Data Fig. 4e). The mass spectra of the M^+ 380 compounds are very similar to those of the C_{29} MAL with main fragments shifted 14 Da towards lighter masses, suggestive of methyl-loss from the polycyclic core (Extended Data Fig. 4f). Elution positions (Extended Data Fig. 4d) also match those of compounds produced during the laboratory maturation of lanosterol (*A1* and *A2* in Extended Data Fig. 4f) and cycloartenol. However, due to poor signal to noise ratios, it remains unclear whether the mass spectra are identical (compare *a2* and *A2* in Extended Data Fig. 4f). Authentic standards are required for unambiguous structural identification.

2.7.4.3. Characterization of aromatic hopanoids and benzohopanoids

In addition to the commonly analysed C_{32}-C_{35} benzohopanoids that encompass side-chain cyclisation and aromatization, we included the aromatic hopanoid derivatives C_{30} and C_{31} acenaphthene-8,14-secohopanoids, fluorene-8,14-secohopanoids, 8,14-secohopanoids and their 28-nor derivatives to calculate aromatic hopanoid abundances (Extended Data Fig. 8).

C_{32}-C_{35} benzohopanoids *i1*-*i4* (Extended Data Fig. 8f) were identified relative to compounds in AGSOSTD reference oil (Fig. 8-14 in Nettersheim, 2017[122]) and mass spectral characteristics[142]. Eluting between the C_{32} and C_{33} benzohopanoids under our chromatographic conditions, the chromatographic signal *j3* (Extended Data Fig. 8c) is more abundant and clearly resolvable in the TIC of many BCF samples. It yields a molecular ion [$M^{+\circ}$] at *m/z* 414 and was identified as a C_{31} 8,14-secohopanoid with a fluorene moiety based on comparison with published mass spectral characteristics[143] (see also Fig. 8-15 in Nettersheim, 2017[122]). Eluting before *j3*, the chromatographic signal *j2* (Extended Data Fig. 8d)—following Carillo-Hernández et al. (2001)[144]—was tentatively identified as a C_{31} acenaphthene-8,14-secohopanoid based on its mass spectral characteristics of dominant [$M^{+\circ}$] at *m/z* 416 and a similarly abundant *m/z* 209 fragment attributed to cleavage through the former C ring that is facilitated by the 8,14-seco (cleavage) structure. Similarly, chromatographic signal *j1*, eluting before C_{32} benzohopanoid *i1* (Extended Data Fig. 8e) was tentatively identified as the corresponding C_{30} acenaphthene-8,14-secohopanoid based on dominant [$M^{+\circ}$] at *m/z* 402 and *m/z* 195 fragments[144]. Chromatographic signals *h7* to *h10* in *m/z* 365 ion chromatograms (Extended Data Fig. 8a) were identified as a series of regular aromatic C_{29}-C_{31} 8,14-secohopanoids by comparison with published elution patterns and mass spectra[145] (see also Fig. 8-19 and 8-20 in Nettersheim, 2017[122]). These compounds are characterized by a *m/z* 365 base ion comprising the polycyclic core after side-chain cleavage, a *m/z* 159 fragment comprising the part of the polycyclic core that contains the aromatic moiety after cleavage of the C-11 to C-12 bond in the 8,14-cleaved (seco) former C-ring; as well as the equivalent fragment after 11,12-cleavage that retains the side chain (i.e. *m/z* 187 in the C_{29}

derivative h7). A similar series is found in *m/z* 351 (Extended Data Fig. 8b) and *m/z* 145 ion chromatograms, comprising double peaks *h1–h6* with [$M^{+\circ}$] of 380, 394 and 408 corresponding to aromatic C_{28}-C_{30} 8,14-secohopanoids that lost one of the methyl groups at the aromatic or pentacyclic ring[146].

2.8. Assignment of steroids to stem or crown sources for Figures 1a and 3b

As described in Methods, Figure 1a and 3b summarize the evolution of the relative abundances of aromatic (proto)steroids and hopanoids from the late Palaeoproterozoic to the Phanerozoic, assigning colours to different compound classes. Supplementary Table 1 provides a key to the chromatographic signals that contribute to each compound class or colour. Shown in shades of purple to pink are aromatic protosteroids interpreted as degradation products of lanosterol and/or cycloartenol, and in cyan and light blue-green aromatic derivatives of 24-methylated protosterols such as cyclolaudenol or 24-methyllanosterol, and in shades of red, blue and green the aromatic derivatives of cholesteroids, ergosteroids and stigmasteroids respectively. Several groups of aromatic steroids in the data set (Supplementary Table 1) may be the diagenetic products of stem as well as crown-sterol precursors: triaromatic cholesteroids and ergosteroids without ring-system methylation (TAS 231 Chol and TAS 245 Erg, where TAS stands for 'triaromatic steroid', and 231 denotes the base ion *m/z* 231 used to quantify the compounds), and 2-, 3- and 4-methylated TAS Chol and Erg (TAS 245 MeChol and MeErg). For instance, TAS 231 Chol may be the aromatisation product of cholesterol but may also form through aromatization and demethylation of cycloartenol. TAS 245 4-MeChol may be generated through diagenetic methylation of cholesteroid precursors but also through demethylation of cycloartenol precursors. However, it is possible to distinguish the dominant precursors by investigating the absence or presence of other protosterol products. Supplementary Table 1 assigns TAS 245 4-MeChol to a protosterol source (entry 3 in SI Table 1) if the corresponding saturated bitumen fraction lacks cholestanes AND the aromatic fraction contains no, or almost no, 2- and 3-methyated TAS AND the aromatic fraction lacks monoaromatic cholesteroids (MAS). Moreover, bitumens of the 1,640 Ma Barney Creek Fm contain traces (usually close to detection limits) of 2- and 3-methylated TAS, and these are interpreted as rearrangement products of 4-methylated protosteroid precursors (entry 5 in SI Table 1). Conversely, if indigenous cholestanes are present in the saturated fraction OR 2- and 3-methylated TAS are present in more than traces, then SI Table 1 assigns methylated TAS to a crown-sterol source (entries 7 and 9 in SI Table 1). The same criteria apply to TAS 231 Erg and TAS 245 4-MeErg.

Several samples in the age range 800 to 520 Ma possess a mixed signal where major proportions of TAS 245 Chol appear to have a protosterol as well as C_{27} sterol origin. This is typified by the presence of indigenous cholestane in the saturated fraction in combination with a very high

proportion of TAS 4-methylated cholesteroids over 2- and 3-methylated isomers. In these cases, it is currently not possible to quantify the proportion of stem- and crown- contributions. However, to acknowledge this effect, SI Table 1 arbitrarily assigns TAS 245 4MeChol to a 'mixed' source (entry 6 in SI Table 1) if all off the following is true: (a) cholestane is present in the saturated fraction, (b) [TAS 245 4MeChol] > 0.5 * [TAS 231 Chol], and (c) [TAS 245 4MeChol] > [TAS 245 2+3MeChol]. These mixed sources of 4-methylated TAS are plotted in a red-brown hue, optically between the purple tones of stem sources and red tones of crown-sterol sources.

3. Evaluation of euphol/tirucallol series precursors for Proterozoic tetracyclic terpenoids

The structures of tetracyclic triterpenoids of the euphol/tirucallol series resemble those from protosterols but have a different stereochemistry at positions C-13, C-14 and C-17[147]. As these compounds are not true protosterols, it is crucial to test whether they are a potential source of tetracyclic terpenoids found in Proterozoic sedimentary rocks.

Tetracyclic triterpenoids of the euphol/tirucallol series are found in the latex of some higher plants such as *Euphorbia* (spurges)[147], and in traces in bacteria as accidental side products of hopanoid biosynthesis[148,149]. Higher plants can evidently be ruled out as sources for such compounds in the Precambrian, and the very high concentrations of MAL, DAL and TAS defy an origin as accidental side products. However, it cannot be ruled out that unknown or extinct microorganisms may have produced such compounds as major natural products in the past.

The stereochemistry of protosterol and euphol/tirucallol series compounds is determined during enzymatic folding of squalene epoxide. The biosynthesis of lanosterol and cycloartenol involves the intermediate formation of the protosteryl cation that has a chair-boat-chair conformation, leading to products with a 13β(Me),14α(Me) configuration. By contrast, euphol (3β,13α,14β,17α,20R-lanost-8,24-dien-3-ol) and tirucallol (the 20S isomer of euphol) are examples of triterpenes biosynthesized by cyclization of squalene epoxide in a chair-chair-chair conformation, with a dammarenyl cation as intermediate, leading to products with a 13β(Me),14α(Me) configuration[150]. Thus, euphols and tirucallols are strictly not sterols.

The recognition of diagenetic products of the euphol/tirucallol series among protosteroids is not trivial. TAS retain their original stereochemistry at C-17 and MAL and DAL at C-13, C-14 and C-17 (Extended Data Fig. 1). However, on achiral GC columns, these compounds cannot be distinguished from their enantiomers (i.e. stereochemical mirror images) that may derive from the euphol/tirucallol series. A similar situation was encountered in the case of enantiomeric monoaromatic derivatives of isoarborinol and fernenol (or fernene), which could only be separated using chiral phase HPLC and measurement of optical activities[132]. However, distinguishing TAS, MAL and DAL from their enantiomers without authentic GC injection standards and suitable

chiral GC columns is not feasible at present. By contrast, the saturated hydrocarbon products of protosterols and euphols/tirucallols should possess unique GC elution times and mass spectra. To test this, we subjected lanosterol and euphenol (3β,13α,14β,17α,20R-lanost-8-en-3-ol) to pyrolysis followed by hydrogenation on PtO_2 (see Section 2.1.3) and comparison of the products with the saturated hydrocarbon fraction of the Proterozoic BCF using GC-MS.

Lanosterol pyrolysis yielded compounds *l1* to *l4* in the BCF, including unambiguously identified 8β(H),9α(H)-lanostane *l4*. Euphenol pyrolysis yielded five compounds *m1* to *m5* in the broad elution region of lanostanes, but none corresponds to a compound present in the BCF (Extended Data Fig. 3b). Moreover, while the mass spectra of 'euphanes' *m1* to *m5* are nearly identical, they differ notably from those of lanostanes (Extended Data Fig. 3c). Major differences include the much higher relative intensity of *m/z* 177 and 190 fragment ions, the presence of notable *m/z* 161 and 301 fragments, and much lower *m/z* 83, 231, 259 and 274 ions. Compounds with such mass spectral characteristics were not detected in the BCF. We thus conclude that euphane isomers generated by pyrolysis of euphenol are not detectable in the BCF, while lanostanes are unambiguously identified. By extension, compounds labelled MAL, DAL and TAS most likely belong to the lanosterol/cycloartenol series and are not the corresponding enantiomers of the euphol/tirucallol series.

4. Maturity dependencies of aromatic steroid parameters

Systematic changes of biomarker distributions over geological time, such as the relative abundances of hopanoids and steroids, might be influenced by thermal maturity trends. The average relative abundance of thermally more labile compounds may decrease with age, particularly towards the Palaeo- and Mesoproterozoic where the number of thermally well preserved sedimentary successions decreases. To test for such biases, we determined the Methyl Phenanthrene Ratio[151] (MPR = 2-MP/1-MP; MP = methylphenanthrene) for samples in the data base and computed the corresponding vitrinite reflectance equivalent R_c(MPR) = 1.1*$\log_{10}$(MPR)+0.95. R_c(MPR) was chosen as maturity parameter as it is minimally influenced by kinetic and source effects[152] and was available for the majority of samples (Table S2).

Extended Data Figure 8g plots R_c(MPR) against geological age, revealing that there is no systematic trend. The average R_c for the Phanerozoic is 0.87 ± 0.16% (n = 33), for the Neoproterozoic 0.79 ± 0.18% (n = 39) and for the Meso- and Palaeoproterozoic 0.78 ± 0.15% (n = 38). Furthermore, Extended Data Figure 8h shows that the relative abundance of aromatic hopanoids and steroids shows no systematic trend with calculated vitrinite reflectance (r^2 = 0.016, n = 110). We also did not find a trend between thermal maturity and the abundance of aromatic stem-steroids relative to all aromatic steroids and hopanoids (r^2 = 0.052), the ratio of mono- to

triaromatic steroids (i.e. MAS in *m/z* 253 over TAS in *m/z* 231, $r^2 = 0.003$), and the percentage of triaromatic stigmasteroids among all triaromatic steroids in the *m/z* 231 trace ($r^2 = 0.001$) (not shown). We thus conclude that the trends presented in Figures 1 and 3 are not significantly affected by thermal maturity.

5. Phanerozoic and Recent Protosteroids

The following synopsis provides a summary of the most relevant information of Appendix A, which provides an encyclopaedic review on protosterol synthesis genes in bacteria, protosterol production in extant organisms, and protosteroids detected in modern environments and ancient sediments.

1. All major groups of extant eukaryotes produce sterols and LECA had the biosynthetic capacity to produce crown-sterols. Crown-sterols can be found in virtually all modern sedimentary environments. Likewise, fossil crown-steroids are found in virtually all Tonian to Phanerozoic bitumens and oils that are diagenetically and thermally well preserved.
2. In our dataset, fossil protosteroids are present in all bitumens and oils that yield crown-steroids, without exception. However, in the Phanerozoic the proportion of protosteroids is usually low. Aromatic protosteroids account for 3.8 ± 4.0% (n = 36) of the total aromatic steroid pool, and the approximate average percentage of saturated protosteranes among total steranes is 0.9 ± 1.4% (n = 15, see Section 10.1). These low relative abundances may largely represent eukaryotic biosynthetic intermediates (Section 8).
3. Based on metagenomic surveys, sterol producing bacteria exist in a wide variety of modern environments (Section 8.2.7). Yet, the bacterial contribution of protosterols to such sedimentary environments remains unexplored (Section 9), with a few notable exceptions. Strongly elevated levels of sterols of likely bacterial origin have only been found at hydrothermal vents, methane seeps and beneath the surface of some cyanobacterial mats (Section 9).
4. There are only few modern and Phanerozoic examples of common sedimentary environments with strongly elevated protosterol levels (Section 9). In some instances, such elevated abundances are likely derived from higher plants (Section 8.1.2).
5. Stem-group eukaryotes likely acquired the ability to produce protosterols before 1,700 Ma (Section 8.2). These organisms must have started to contribute sterols to the sedimentary record by that time at the latest. Likewise, at least one lineage of bacteria likely possessed the ability to produce sterols by 1,700 Ma.
6. The first two genes for sterol biosynthesis, SQMO and OSC, have been detected in at least twelve phyla of bacteria, and actual sterol production has been demonstrated for three of

them. However, all bacterial clusters that possess SQMO and OSC, except Methylococcales and Myxococcales, have relatively shallow clade depths, suggesting that most acquired the genes via horizontal gene transfer (HGT) relatively recently. It is currently not possible to state with certainty that any of the currently know groups of sterol producing bacteria existed in the early mid-Proterozoic or, in fact, predated LECA (Section 8.2).

7. The suggestion that stem-group eukaryotes acquired cholesteroid biosynthesis genes as a full syntenic set from the Myxococcales suborder Cystobacterineae[33] is not supported by biomarker data, and is not supported by molecular clocks that point to a Phanerozoic origin of these bacteria. The direction and timing of any gene transfer between Eukarya and Myxococcales remains unresolved (Section 8.2.3).

6. Biological origins of mid-Proterozoic steroids

6.1. The biological origin of side-chain methylated ursteroids

The 1,300 Ma Roper Group contains indigenous triaromatic steroids that are methylated at Ring-A at position C-4 and in the side chain at position C-24. The precursor lipid may have been the C_{31} sterol 24-methylenecycloartenol, which is the second intermediate in the biosynthetic pathway to phytosterols[17]. However, more advanced downstream sterols demethylated at C-14 and/or demethylated once at C-4 are plausible precursors as well. Saturated steranes were below detection limits in the Roper Group bitumens, so the exact precursor carbon skeletons cannot be confirmed. The 24-methylated ursteroids occur in high relative abundances of 12 to 24% of total steroids (18 ± 4%, n = 4), suggesting that they are biogenic and not diagenetic methylation products. Functionalized sterol side chains may generally become methylated during diagenesis, and this has also been observed in pyrolysis experiments[32,153]. However, these diagenetic methylation products form in relatively low abundances (usually low single digit percentage levels), and abiogenic methylation has thus far only been observed at penultimate and terminal carbon positions in the sterol side chain. For example, pyrolysis of the C_{27} sterol desmosterol, which possesses an unsaturation at C-24, resulted in products alkylated dominantly at the terminal C-26 position. Likewise, sedimentary rocks of Tonian and Cryogenian age contain 26-methyl cholestanes (cryostanes) at a relative abundance of 1.7 ± 3.1% (n = 47, range is 0 to 7.8% plus an outlier at 15%)[15,46] that are demonstrably the product of diagenetic C_{27} sterol methylation[32]. By contrast, among the 47 Tonian and Cryogenian samples, only two contained detectable 24-methyl cholestanes (ergostanes) at relative abundances of 0.9 and 1.4%, and these may be biogenic. These values suggest that diagenetic 24-methylation does not occur or is uncommon even in sediments where diagenetic 26-methylation is notable. Moreover, outside the Roper Group, not one mid-

Proterozoic bitumen yielded 24-methylated steroids, highlighting that diagenetic 24-methylation, if it occurs at all, is not a common sedimentary process. Thus, there is currently no indication that 24-methylated steroids can form through abiogenic processes, particularly at the high relative abundances observed in the Roper Group.

Sterol methyltransferase (SMT), the gene responsible for the alkylation of sterols in the side chain at C-24, is found in eukaryotes and some bacteria. The C-24 methylated steroids of the Roper Group may thus, in principle, have a bacterial and/or stem-group eukaryotic origin. Although none of the bacteria screened for sterols to date yielded side-chain alkylated sterols[35], SMT genes of bacteria that live in symbiosis with sponges, when expressed *in vitro*, methylate C_{27} sterols to form 24-methylated or even 24-ethylated cholesteroids[154,155]. Some SMTs even yield 24-isopropenylcholesterol and related C_{30} sterols that had been regarded as diagnostic for sponges. Similar functionalities were found for bacterial SMTs identified in metagenomes of marine, freshwater and other environments. These experiments demonstrate that bacterial SMTs are functional. The study also found bacterial SMTs that occur in sterol biosynthesis gene clusters, suggesting that some bacteria have the capacity to produce side-chain alkylated sterols *de novo*[155]. Biosynthesis of side-chain methylated sterols may thus not be a strictly diagnostic eukaryotic trait. Future analyses have to determine whether and which groups of SMT-bearing bacteria produce sterols *de novo* or, conversely, alkylate sterols taken from the environment (as shown for Chlamydiae)[155]. It will also be crucial to assess whether sterol alkylation by bacteria has a detectable impact on sedimentary sterol profiles.

For the interpretation of the 1,300 Ma old 24-methylated ursteroids, the most important question is the antiquity of SMT genes in bacteria. While this question cannot be answered with certainty at present, SMT trees provided in refs[33,39,155] suggest that bacterial SMTs have relatively shallow clade depths. Moreover, in the trees provided by refs[39,155] all analysed bacterial SMTs branch within extant Eukarya, suggesting that they were acquired after the emergence of LECA. While these observations do not exclude bacterial side-chain methylation by 1,300 Ma, it makes a stem-group eukaryotic origin for these ursteroids more likely. Generally, 24-alkylation is a common characteristic of Eukarya[35,156]. It occurs across all major eukaryotic groups and it was probably already present in LECA[17]. The few eukaryotic clades that do not alkylate the side chain, such as most eumetazoans, lost the ability later. The presence of 24-methylation in LECA demonstrates that SMT was acquired in the eukaryotic stem. Thus, the 24-methylated ursteroids extracted from 1,300 Ma old sedimentary rocks are consistent with a eukaryotic origin and, in combination with a lack of crown-steroids in these sediments, with a stem-group eukaryotic origin.

6.2. Anaerobic stem and crown-group eukaryotes

Porter et al. (2018)[21] suggested that anaerobic eukaryotes may have been common in the mid-Proterozoic interval. As anaerobic eukaryotes cannot produce sterols, this hypothesis elegantly reconciled the existence of diagnostic eukaryotic fossils with the paucity of crown-group steranes. However, the new protosteroid record demonstrates that steroids were in fact abundant throughout the interval. Thus, although anaerobic stem and crown-group eukaryotes likely existed, they are not needed to explain the record.

6.3. Microaerophilic crown-group eukaryotes

It is conceivable that mid-Proterozoic protosteroids derive from crown-group eukaryotes that secondarily reverted to a microaerophilic lifestyle, conserving scarce oxygen resources by truncating the sterol biosynthetic pathway after the first step. Yet, there is no independent fossil evidence for the existence of microaerophilic crown-group eukaryotes in the mid-Proterozoic, nor is there any evidence that microaerophilic crown-group eukaryotes revert to protosterol biosynthesis. Extant eukaryotes living under extreme oxygen deficiency either take up sterols from their environment, produce sterol surrogates such as tetrahymanol or entirely lack polycyclic membrane modifers[157-159]. Moreover, crown-group eukaryotes only require molecular oxygen at very low levels to complete the full sterol biosynthetic pathway[160], and crown-sterol biosynthesis may have been possible in the mixed zone of most open water habitats, in cyanobacterial mats and ventilated deep basins[67] throughout the mid-Proterozoic interval. Thus, considering the paucity of downstream modified steranes in the mid-Proterozoic biomarker record, a notable contribution of protosterols by crown-group eukaryotes is an unlikely scenario.

7. The nature of mid-Proterozoic eukaryotic fossils in light of the steroid record

The mid-Proterozoic record of microfossils that bear diagnostic eukaryotic characteristics is sparse. The oldest fossils confidently assigned to crown-group Eukarya emerge about one billion years ago, the 1,050 Ma multicellular rhodophyte alga *Bangiomorpha*[10] and ~1,000 Ma chlorophyte alga *Proterocladus*[11]. Other fossils with diagnostic eukaryotic features, such as processes, protrusions and cell-wall ornamentation, occur throughout the Mesoproterozoic and date back to ~1,650 Ma (Fig. 1c). However, these fossils lack characteristics of extant clades, and it remains unknown whether they belong to the eukaryotic stem or crown[1,8,9]. By contrast, fossil steroids extracted from 1,640 to 1,000 Ma old sedimentary rocks have a likely stem-group eukaryotic origin, while clearly indigenous and generally accepted traces of crown-steroids do not emerge until ~800 Ma (Fig. 3; note that there is no data for the critical interval < 1,000 to > 800 Ma).

To interpret these combined fossil and biomarker records, it is important to consider the different taphonomic conditions responsible for microfossil and molecular fossil preservation. Biomarkers are preserved in sedimentary environments where a fraction of organic matter escapes mineralization and is protected by reducing conditions. This commonly occurs in sediments deposited in low energy environments characterized by small mineral grains such as clay particles that protect organic matter through surface adsorption and limited diffusion of oxygen and other oxidants. Biomarkers are thus not commonly preserved in coarse sedimentary rocks such as sandstones, or in environments with persistently well-oxygenated bottom waters such as many near-shore environments, although they may be found in oxygenated, tide-influenced environments where microbial mats limit oxygen exchange between sediment and water column[161]. Moreover, organic matter from shallow oxygenated waters (and in the post-Silurian also from terrestrial environments) may be transported to, and preserved in, deeper depositional settings. So, while a part of the biomarker signal likely represents contributions from well oxygenated, high-energy environments, the record is clearly biased towards a limited number of specific depositional environments favouring organic preservation. However, in those environments where biomarkers are well preserved, they afford semi-quantitative information about relative abundances of broad taxonomic groups.

In contrast to biomarkers, mid-Proterozoic microfossils have very different modes of preservation, are found in a far wider range of habitats, and may provide morphological information to the species level. However, they lack information about ecological importance and abundance. Soft bodied, single cellular organisms lacking preservable structures such as resistant cell wall material or mineral tests, will only leave a fossil record under exceptional circumstances. Such organisms may entirely evade detection in the body fossil record. Therefore, based on body fossils, we may remain oblivious of the most abundant and ecologically dominant organisms of the mid-Proterozoic interval. It is possible that we do not possess a single fossil of an organism that contributed detectable abundances of protosteroids to sediments. Conversely, organisms that only constitute a small percentage of total biomass but have outstanding potential to fossilize, will be overrepresented in the body fossil record, possibly by orders of magnitude. Thus, organisms leaving a copious fossil record in specific sediments may never have been abundant and may fail to contribute detectable quantities of biomarkers.

In summary, while molecular fossils are biased towards oxygen-starved environments, they can provide information about the relative abundances of broad organism groups. Body fossils are taxonomically far more specific, are found in a wider range of depositional environments, but rarely provide data on ecological importance. Thus, while the two records are complementary, it is currently not possible to link individual mid-Proterozoic microfossils to specific biomarkers and vice versa. While the lack of crown-steroids and the immense abundance of protosteroids in mid-

Proterozoic sediments suggests that most fossils are stem-group representatives, it is not possible to make statements about individual specimens. The 1,050 Ma old red alga *Bangiomorpha* and chlorophyte alga *Proterocladus* are cases in point, and other fossils may also belong to the crown-group.

APPENDIX

APPENDIX A

This Appendix provides a literature review and collection of information on protosterol synthesis genes in bacteria, protosterol production in extant organisms, and protosteroids detected in modern environments and ancient sediments. The Appendix is intended to be encyclopaedic and does not have to be read in any specific order. A synopsis of the most relevant information is provided in Section 5.

8. Protosterols in extant organisms

The abundance of protosteroids is generally low in Phanerozoic bitumens and oils relative to other steroids, but individual samples display strongly elevated proportions. In Supplementary Table 1, aromatic protosteroids in the Phanerozoic account for 3.8 ± 4.0% (n = 36), and the approximate average percentage of protosteranes among total steranes is 0.9 ± 1.4% (n = 15, see Section 10.1). Examples for elevated abundances include a Cretaceous lacustrine oil (GA224) with 7.1% aromatic protosteroids, a Jurassic coal-resinitic oil (GA19999642) containing 6.8%, Ordovician Tarim oil (HD_1 #5458) with 20.5%, and bitumen from a Cambrian marine limestone (15B602) at 13.4%. The following section discusses potential biogenic sources that may explain typical background abundances but also elevated concentrations of protosteroids in modern and Phanerozoic depositional environments.

8.1. Eukaryotic sources

8.1.1. *Protosterols as biosynthetic intermediates*

Lanosterol or cycloartenol are the first polycyclic biosynthetic intermediates in the biosynthesis of sterols in virtually all living eukaryotes and thus occur at least in low abundances in all species that produce sterols *de novo*. Some species contain major concentrations of protosterols, albeit largely for unknown reasons, and these exceptions are reviewed in Sections 8.1.2 and 8.1.3.

Information on the abundance of protosterols as biosynthetic intermediates is surprisingly sparse. In early work, Goad and Goodwin (1967)[162] purified sterols from larch leaves (*Larix decidua*) and found a mass content of ~1% cycloartenol and 13% methylenecycloartenol relative to total sterol content (computed from data provided in the text). A sterol survey[163] on basal, flagellated fungi found a lanosterol proportion of 2 to 7% of total sterols (average 3.6 ± 2.2%, n = 5; data excludes two samples where the lanosterol content was not reported and one outlier where lanosterol was the major sterol, see Section 8.1.3). Organisms of the phylum Hyphochytriomycota are heterokont protists with rhizoidal or hypha-like vegetative system that, in the past, were grouped with fungi. Two analysed species had a lanosterol content of 2% relative to total sterols[163]. In a detailed sterol

survey[164] of 106 diatom species, 104 did not contain detectable concentrations of protosterols (detection limit presumably < 0.5% of total sterols). The authors concluded that diatoms are generally unlikely to be major sources of lanosterol or cycloartenol in natural environments. However, in two species, protosterols were exceptionally high, and this is further discussed in Section 8.1.3. Among higher plants, the *Arabidopsis* wildtype contains 3% cycloartenol and 4% 24-methylenecycloartanol[165]. In summary, while the data are sparse, the abundance of protosterols as biosynthetic intermediates in eukaryotes presumably falls into the low single digit percentage range relative to total sterols.

8.1.2. *Elevated protosterol abundances in higher plants*

Many higher plants contain terpenoids that may, for instance, function as a defence against herbivores and pathogens. In some plants, compounds with protosteroid skeletons are part of this terpenoid pool, which can result in significantly elevated protosterol levels in plant organic matter relative to crown-sterols. Since protosterols are commonly isolated together with triterpenoids rather than crown-sterols, most studies only report their abundances relative to other triterpenoids such as oleanoids. Here we briefly review some of the reported occurrence of protosterols in a variety of higher plants, and where possible provide information about the abundance of protosterols relative to ursterols and crown-sterols.

Vegetable oils may contain substantial quantities of protosterols[166]. The triterpene alcohol fraction of 19 vegetable oils were strongly dominated by cycloartenol and 24-methylene cycloartenol. Based on data in refs[166,167], calculated protosterol ratios (protosterols/[protosterols + crown-sterols]*100) for the oils ranged between ~1% and 37%, with an average of ~15%. For commercial Shea butter, the non-glyceride saponifiable fraction (6 wt% of Shea butter; compared to 1 wt% free triterpene alcohols and sterols) yielded 1 wt% parkeol and 1.2 wt% 24-methylene-lanost-9(11)en-3-ol, compared to a crown-sterol content of 1.5% α-spinasterol and 1.5% stigmastenol, yielding a protosterol ratio of ~42%[168]. In addition to the more common cycloartenol or cyclolaudenol, some plants also contain protosterol derivatives with various types of functional groups[169].

The bark of some trees is particularly rich in cycloartanol-derivatives. A petroleum extract (1% of wood weight) of birch wood (*Betula verrucosa*) contained 2% cycloartenol, 4% 24-methylene cycloartenol, 5% 4-methylstigmastadienol and 6% β-sitosterol derivatives in addition to 4% non-esterified β-sitosterol[170], corresponding to a protosterol ratio of ~37% ([cycloartenol + methylene cycloartenol]/ [cycloartenol + methylene cycloartenol + β-sitosterol]*100). Abietospiran, a cycloartanol-derived ether with poly-functionalized side chain is highly abundant in the bark of white fir (*Abies alba*), comprising 0.14 g/g bark[171]. Similar cycloartanol-derivatives were also

isolated in bark of other firs (e.g. *Abies grandis*[172]). The stem and branches of the angiosperm *Neolitsea aciculata* (Magnoliids) contain not only cycloartenol and parkeol, but also a large variety of 24-methylated and 24-ethylated lano- and cycloartanol-type triterpenoids and even some with dimethylation at C-24 and trimethylation at C-25. The triterpene alcohol fraction comprises near-exclusively protosterol-derivatives, except for 12.4% lupeol and 6.4% unidentified compounds. The 4-methyl sterol fraction comprises 24-methyl and 24-ethyl 31-norlanosteroids as well as citrostadienol, a 24-ethylated dinorlanosteroid lacking one of the C-4 and the C-14 methyl groups[173].

Latex may also contain high absolute abundances and proportions of protosterols and their derivatives. The tetracyclic and pentacyclic latex constituents of five different varieties of *Euphorbia pulcherrima* contain 2.0 to 3.2% sitosterol and 0.4 to 1.2% cycloartenol by dry weight[174], corresponding to protosterol ratios (cycloartenol/[cycloartenol + sitosterol]*100) of ~12 to 37% (average = 26%; n =5). The latex of *Syngonium auritum* contains ~1.85 mg sterols/mL, and cycloartenol constitutes ~32% of this sterol fraction. Likewise, *Xanthosoma sagittifolium* contains 6.2 mg sterols/mL latex, including ~27% cycloartenol, and *Xantholosoma violaceum* contains 8.4 mg sterols/mL latex with ~23% cycloartenol[175]. Since the other major sterols were not identified in the study (and might comprise additional protosterol derivatives), the cycloartenol ratios of 23 to 32% can be considered as minimum protosterol ratios for the latex of these three New World Araceae species.

Some grasses (Gramineae or Poaceae) also contain substantial amounts of protosteroids. For example, non-saponifiable fractions of 2-3 weeks-old maize shoot extracts contained ~7.3 mg 4,4-dimethyl and 187.6 mg 4-desmethyl sterols per gram of shoots[176]. The 4-desmethyl sterols were found to contain ~75% β-sitosterol, ~14% stigmasterol, ~9% campesterol and ~2% cholesterol, while the 4,4-dimethyl sterol fraction was reported to contain ~65% cycloartenol, ~13% 24-methylenecycloartenol and ~22% were comprised of two unidentified compounds[176]. Based on the ≥78% cycloartenol + methylene cycloartenol content of the 4,4-dimethyl-fraction and the ~100% crown-sterol content of the 4-desmethyl fraction, we calculate a minimum protosterol ratio of ~2.9%. Similarly, pea leaves have also been shown to contain cycloartenol and minor 24-methylene cycloartenol. Assuming that the 4,4-dimethyl triterpene fraction comprises ~90% protosterols (cf. Fig. 5a in ref[177]), from the 4,4-dimethyl (1.5 mg) and 4-desmethyl sterol fraction contents (71.6 mg)[177], we estimated a protosterol ratio of ~1.9%.

Plant waxes may also contain substantial amounts of protosteroids. They were reported for example from the epicuticular wax of spurges (Euphorbia)[178], maple leaves[179] and tomatoes[180]. In *Euphorbia esula*, cyclolaudenol was found to comprise between 5 and 31% (average = 21.2%) of total triterpenoids in five *E. esula* biotypes, corresponding to 0.5 to 2.5% (average = 1.35%) of the

total wax (calculated from data reported in ref[178]). While sterol contents were unfortunately not reported, the high proportion of protosterols of total wax content, in combination with the observation that plant waxes often form a prominent component of many sedimentary biomarker signatures, we infer that plant waxes may contribute a substantial proportion of the protosterol inventory of near-shore post-Silurian strata.

Bryophytes, non-vascular plants such as mosses and liverworts, often contain notable amounts of cycloartane-type triterpenoids[181]. *Sphagnum* peat mosses are particularly noteworthy as they dominate many bogs and might thus be important protosteroid sources in peats, lignite coals and lake deposits. In the green part (capitulum) of the rusty peat moss *Sphagnum fuscum*, cycloartenol and methylene cycloartenol mainly occurred in esterified form and constituted ~0.02 wt% of the entire biomass of the capitulum[182]. In the 21-24 cm segment below the green top of a *S. fuscum* shoot, where first signs of decay at the microscopic level were observed, protosterols constituted ~0.004 wt% of the dry weight of the plant material, yielding a calculated global (free and esterified) protosterol ratio of ~15% (calculated from data in ref[182]). Protosteroids also occur in a variety of other bryophytes. A review of the chemical constituents of bryophytes[181] lists (in their Table IIIa) 15 different species of bryophytes from three different classes (Sphagnopsida, Polytrichopsida and Bryopsida) for which cycloartenol-, cyclolaudenol, cycloeucalenol and/or obtusifoliol-type triterpenoids have been reported. In the moss *Pseudoscleropodium purum* (Bryopsida), cyclolaudenol comprises 0.008% (dry wt.) and 31-norcyclolaudenol 0.006% compared to 0.015% campesterol, 0.01% stigmasterol and 0.015% sitosterol[183], corresponding to a protosterol ratio of ~26%. This shows that some of the earliest land plants may already have contributed notable amounts of protosteroids to the Phanerozoic rock record.

8.1.3. *Elevated protosterol abundances in other eukaryotes*

Sterol production is an energy and oxygen-intensive process. The conversion of the isoprenoid precursor squalene to a protosterol requires only one O_2 molecule to introduce the characteristic 3-hydroxylation of sterols via the epoxidation of squalene prior to enzymatic folding. The subsequent enzymatic down-stream modifications that result in the formation of typical crown-sterol such as ergosterol, cholesterol or sitosterol requires an additional 10 (cholesterol, sitosterol) to 11 (ergosterol) O_2 molecules[156]. It is thus plausible that eukaryotes with a very limited energy or oxygen supply may abbreviate the pathway to accumulate protosterols as major sterols unless this is detrimental to membrane and metabolic function. It is also plausible that some eukaryotes other than higher plants (section 8.1.2) accumulate protosterols as a deterred against grazers or predators.

Reports of elevated protosterol abundances in eukaryotes (other than higher plants) are scarce, although this may be strongly biased because protosterols are not routinely screened in lipid surveys, whereas they typically elute in triterpenoid fractions included in many plant studies. An exception among algae are diatoms. The diatom *Haslea sp.* was shown to biosynthesize 20% lanosterol and 5% 4-norlanosterol as part of its sterol mix[164]. Similarly, the diatom *Stauroneis simulans* contained 11% 24-methylcycloartenol, 2% lanosterol and 82% 24-methyl-14-norcycloartenol. The function of the elevated protosterol content in these species is unknown. However, elevated protosterol abundances in diatoms are generally an exception as protosterols were below detection limits (presumably < 0.5%) in a further 104 species. The authors concluded that diatoms are generally unlikely to be major sources of protosterols in natural environments[164]. Yet, blooms of specific diatom species with high protosterol abundances may well have contributed to elevated protosteroid occurrences in Phanerozoic sediments.

A relatively high protosterol content was observed in the excavate protist *Naegleria gruberi* that accumulates cycloartenol in substantial amounts (500 µg/g dry weight)[184]. Based on the reported high abundance of 4-methyl sterols (3-4 mg/g dry weight) and the reportedly similar abundances of 4-desmethyl sterols, we determine a protosterol ratio of ~6-8%.

In animals, lanosterol-derivatives are highly enriched in lanolin (wool wax). The unsaponifiable fraction of wool wax consists of ~30% steroids (mostly cholesterol, but also cholestanol and 7-ketocholesterol) and 25% triterpenes (predominantly lanosterol, smaller amounts of lanostatrienol, and the respective dihydro derivatives)[185], corresponding to a protosterol ratio of ~46% (percentage of protosterols of total identified sterols).

Among fungi, *Catenaria anguillulae*, a member of the Chytridiomycota and an endoparasite of nematodes, yields lanosterol (45%) as the main sterol[163]. While two more of the ten fungi (*Chytridium confervae* and *Rhizophlyctis rosea*) analysed by ref[163] also accumulated appreciable amounts of lanosterol (7% of total sterols), the other fungi in that study only contained minor amounts (typically ~2%). The function of elevated lanosterol levels in above fungi is unknown, but there is evidence that fungi can down-regulate crown-sterol biosynthesis under conditions of O_2 limitation and intermittently accumulate 24-methylene lanosterol, albeit presumably only as a signalling molecule, not as a critical membrane component[186].

8.2. Bacterial sources

8.2.1. *Notes on gene and enzyme nomenclature*

For the nomenclature of genes and enzymes, different combinations of capital, small, italic and roman fonts are recommended to distinguish enzymes and genes, and also genes of bacteria,

eukaryotes and humans. In the context of the following discussion, these distinctions are irrelevant, and we follow Hoshino and Gaucher (2021)[33] using capitalized roman letters for all genes and enzymes, e.g. SQMO and OSC.

8.2.2. *The phylogeny of sterol producing Bacteria*

Based on genomic and metagenomic surveys, knowledge about the distribution of sterol biosynthesis genes in the domain Bacteria has increased markedly in the past years[17,33-35,187,188]. Most studies focused on SQMO, squalene monoxygenase, catalysing the oxygenation of squalene to oxidosqualene, and OSC, oxidosqualene cyclase, catalysing the cyclization of oxidosqualene to protosterols. A comprehensive survey of all genes involved in sterol biosynthesis was provided by Hoshino and Gaucher (2021)[33].

In a pioneering combined genomic and lipid analysis, Pearson et al. (2003)[187] detected SQMO and OSC in the planctomycete bacterium *Gemmata obscuriglobus* and subsequently identified the protosterol lanosterol and its double-bond isomer parkeol in saponified cell extracts. Subsequent genome surveys[17,33-35,188] provided a comprehensive overview of the presence of sterol biosynthesis genes in bacterial genomes that were available in the public domain at that time. Wei et al. (2016)[35] performed a combined genomic and lipid analysis and searched for homologs of OSC through BLASTP in all bacterial genomes (31,237) available in 2016 in the JGI database. The search yielded 34 bacterial OSC homologs in five different phyla, including the bacterial groups where sterols had been detected in earlier lipid studies[187,189,190] plus several new occurrences. Similarly, in a search for SQMO and OSC protein sequences in the National Center for Biotechnology Information (NCBI) protein database, Gold et al. (2017)[34] found 27 bacterial taxa across 6 phyla, 9 classes and 9 orders, confirming the findings of Wei et al. (2016)[35]. In an update on sterol biosynthesis genes in the NCBI data base, Hoshino and Gaucher (2021)[33] detected OSC in 94 bacterial taxa across twelve phyla (based on their Table S1).

The combined studies detected the genetic potential for protosterol biosynthesis in the following major bacterial groups:

- Phylum **Bacteroidetes:** Genera *Fluviivola* (Crocinitomicaceae), *Eudoraea* (Flavobacteriaceae) and an unnamed species in Cytophagales
- Phylum **Actinobacteria**: Eight named and six unnamed species, including the genera *Streptomyces*, *Streptosporangium*, *Nocardia*, *Lentzea*, *Nonomuraea* and *Actinomadura*
- Phylum **Cyanobacteria:** Six genera from three different orders (Oscillatoriales, Synechococcales, Nostocales)
- Phylum **Planctomycetes:** two species of the genus *Gemmata*
- Phylum **Verrucomicrobia:** Eleven unnamed species

- Phylum **Chloroflexi:** three unnamed species
- Phylum **Nitrospirae:** two species of the genus *Nitrospira*
- Phylum **Rokubacteria:** three unnamed species
- Phylum **Dadabacteria:** one unnamed species
- Phylum **Gemmatimonadetes:** two unnamed species
- Phylum **Acidobacteria:** one unnamed species in the Holophagales
- Class **Alphaproteobacteria:** One unnamed species each in Rhodobacterales and Rhodospirillales, and the methanotrophs *Methyloceanibacter caenitepidi* and *Methylobacterium methanicus* belonging to two different families in the order Hyphomicrobiales (=Rhizobiales)
- Class **Deltaproteobacteria:** Twelve genera of the order Myxococcales in the suborders Cystobacterineae (5), Nannocystineae (3) and Sorangiineae(4), ten more unnamed Myxococcales isolates, and one isolate not further characterized.
- Class **Gammaproteobacteria:** Nine genera of aerobic methanotrophic bacteria in the order Methylococcales (four Type 1a and five Type 1b), the genus *Glaciecola* in Alteromonadales, two unnamed species in Chromatiales, and seven isolates not further characterized

Wei et al. (2016)[35] also detected 176 OSC sequences in 2707 environmental metagenomes, of which at least 37 clustered among bacterial OSC. Many of the detected genes clustered closely with known bacterial OSC, while others may belong to additional classes and phyla. However, all OSC from the metagenomic data set clustered within the known major Groups 1 and 2a-d of Wei et al. (2016)[35], and addition of metagenomic data only caused relatively minor changes to the topology of their OSC phylogenetic tree.

The OSC phylogenetic trees presented in different studies have the same major characteristics and some relatively minor differences. The OSC tree of Desmond and Gribaldo (2009)[17] only included four bacterial species. In that study, *Plesiocystis* (Myxococcales), *Methylococcus* (Type 1b Methylococcales) and *Gemmata* (Planctomycetes) formed an outgroup to eukaryotic OSC while a second genus within Myxococcales (*Stigmatella*) branched within Eukarya, invoking a relatively recent horizontal gene transfer (HGT). In the OSC tree of Villanueva et al. (2014)[188], which is based on twelve bacterial OSC sequences, *Stigmatella* falls outside eukaryotic OSC, but *Fluviicola* (Bacteriodetes) and two further genera of Methylococcales of Type 1a (*Methylobacter*, *Methylomicrobium*) branch within Eukarya, distributing the Methylococcales between two distinct major clusters. The larger data set of Wei et al. (2016)[35] largely echoes the topology of Villanueva et al. (2014)[188], with a major cluster 'Group 1' outside Eukarya that includes the Myxococcales suborders Nannocystineae and Sorangiineae ('Myxococcales Group I'), the three cyanobacterial

orders, *Gemmata* and two genera of Type 1b Methylococcales. Three smaller clusters, Groups 2a-c, branch close to Eukarya, including the Myxococcales suborder Cystobacterineae ('Myxococcales Group I'). Group 2d falls within Eukarya, including three genera of Type 1a Methylococcales, *Fluviicola* and a representative of phylum Verrucomicrobia. Notable is the distribution of different genera of Myxococcales, Methylococcales and Bacteriodetes across major different groupings. The OSC tree topologies of Gold et al. (2017)[34] and Hoshino and Gaucher (2021)[33] (their Fig. S13) show the same general features, with Bacteriodetes and Type 1a Methylococcales branching within Eukarya, and similar splits of Myxococcales, Methylococcales and Bacteriodetes across different clusters.

SQMO and OSC phylogenetic trees offer several insights. Firstly, the topologies of the SQMO and OSC trees are complex and do not follow rDNA based evolutionary patterns of bacteria. Secondly, apart from Myxococcales and Methylococcales, the occurrence of OSC is phylogenetically shallow, largely occurring in individual genera and species within larger groupings that otherwise lack SQMO and OSC. For instance, in Cyanobacteria, OSC only occurs in six genera and these are distributed across three different orders. As the number of published cyanobacterial genomes and metagenomes is large, it is unlikely that OSC is widespread within the phylum but remains undetected. In the words of Gold et al. (2017)[34], "the shallow clade depth and sparse phylogenetic distribution of bacterial genes—combined with the relatively long branches leading to eukaryal representatives—are most consistent with these genes being transferred one or more times from stem eukaryotes to bacteria". We agree with this assessment, but note that the data is currently not robust enough to determine the sources, direction and timing of HGT events in most cases. This is exemplified by the most comprehensive study of Hoshino and Gaucher (2021)[33], discussed Section 8.2.3.

8.2.3. *Myxococcales as the possible origin of sterol biosynthesis*

Hoshino and Gaucher (2021)[33] performed a comprehensive analysis of all genes involved in sterol biosynthesis, including 'Stage 1' enzymes involved in the biosynthesis of squalene, and 'Stage 2' enzymes that generate and modify the polycyclic core of sterols, in particular demethylation at positions C-4 and C14, plus further downstream enzymes involved in the modification of unsaturations and side-chain alkylation. The authors found that Stage 1 genes in Myxococcales (and one species of Dadabacteria) were more similar to the eukaryotic pathway than to other bacteria. Moreover, they found that Stage 2 genes involved in the removal of methyl groups from the protosterol core to generate C_{27} cholesteroids are present in the two Myxococcales suborders Nannocystineae and Sorangiineae (Myxococcales Group I), but not in suborder Cystobacterineae (Myxococcales Group II) or any other group of bacteria. The core Stage 2 genes in bacteria were

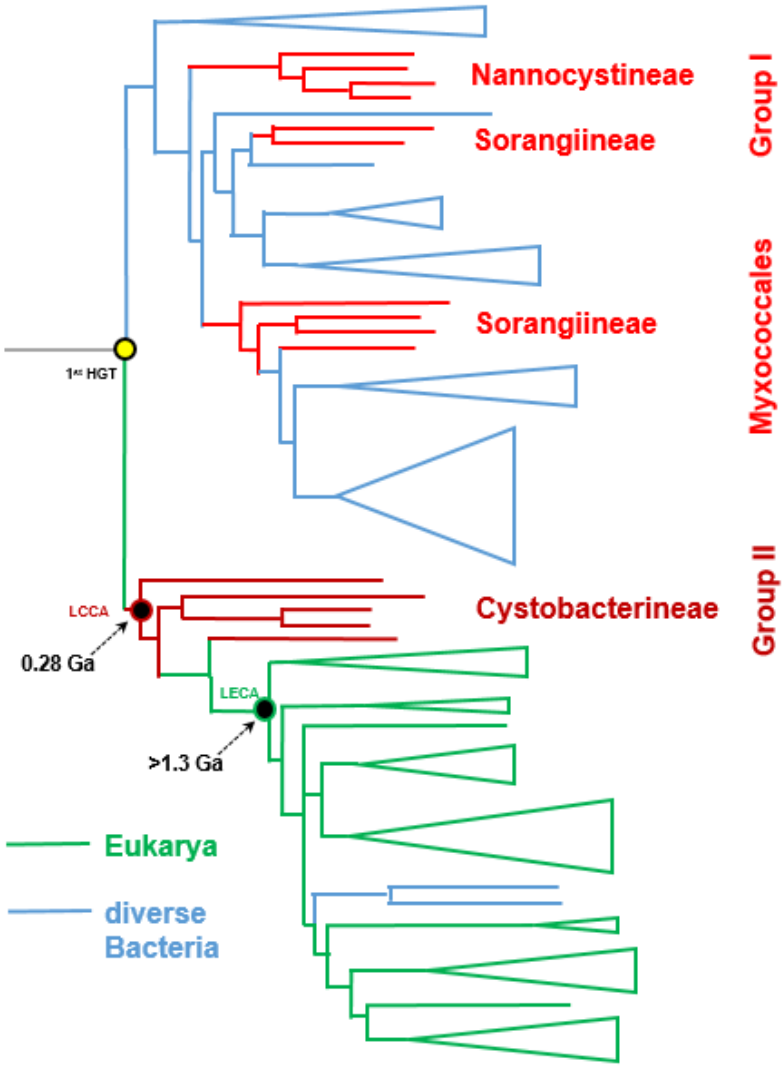
mostly in a syntenic relationship with each other, i.e. they were physically co-localized with other sterol biosynthesis genes.

To create a possibly more robust phylogenetic tree of sterol biosynthesis genes, Hoshino and Gaucher (2021)[33] used concatenated sequences of Stage 2 genes involved in the biosynthesis of protosterols (SQMO, OSC), the opening of the three-ring of cycloartenol (CPI1) and the subsequent removal of methyl groups from C-14 (CYP51, ERG24) and C-4 (ERG25, ERG26 and ERG27-analog). A simplified version of the tree is reproduced in Supplementary Figure 1. In the concatenated tree, but also in the individual gene trees, crown-group Eukarya always branches within Bacteria. Moreover, in the OSC, SQMO and concatenated tree, Eukarya branch within Myxococcales Group II (suborder Cystobacterineae) (Supplementary Fig. 1). The branching pattern of Myxococcales in these trees broadly follows phylogeny, i.e. the sterol genes often (but not always) fall into the correct suborders. Based on these observations, Hoshino and Gaucher (2021)[33] suggested that sterol biosynthesis, including demethylation of the sterol core, evolved within Bacteria and was transferred from Myxococcales Group II (suborder Cystobacterineae) to stem-group eukaryotes. Moreover, sterol biosynthesis was probably acquired by eukaryotes from Myxococcales Group II as a full syntenic set. As Myxococcales Group II does not possess any genes for downstream modification, the authors suggest that all genes apart from SQMO and OSC were later lost from these Myxococcales.

The model of Hoshino and Gaucher (2021)[33] makes strong predictions about the timing of this HGT event. As LECA presumably already possessed all eight genes of the concatenated set, the transfer must have occurred broadly before ~1,300 Ma ago. However, the authors further suggest that sterols were likely needed for endocytosis and the acquisition of the mitochondrion, presumably moving the HGT event considerably further back in time. Based on molecular clock estimates of Gold et al. (2017)[34], the maximum probability for the horizontal transfer of the eight Stage 2 sterol biosynthesis genes from Myxococcales Group II to eukaryotes would then be ~2,300 Ma, with a minimum age of ~1,700 Ma.

The timing of these events and the direction of gene transfer critically depend on correct tree topology. However, the phylogeny generated by the concatenation of the eight Stage 2 sterol biosynthesis genes (Supplementary Fig. 1), as well as the individual gene trees, are presumably inconsistent with the evolution of Myxococcales. In the concatenated gene tree, the Cystobacterineae cluster includes the species *Archangium gephyra*, *Cystobacter fuscus*, *Corallococcus coralloides*, *Stigmatella aurantiaca* and an unnamed species. The four named species are terrestrial chitin degraders that acquired chitinase of the GH18 family via HGT from Bacilli (Firmicutes), which in turn received it via HGT from crown-group eukaryotic fungi[191]. Based on well calibrated molecular clocks, Gruen (2019)[191] estimated the date of the HGT of

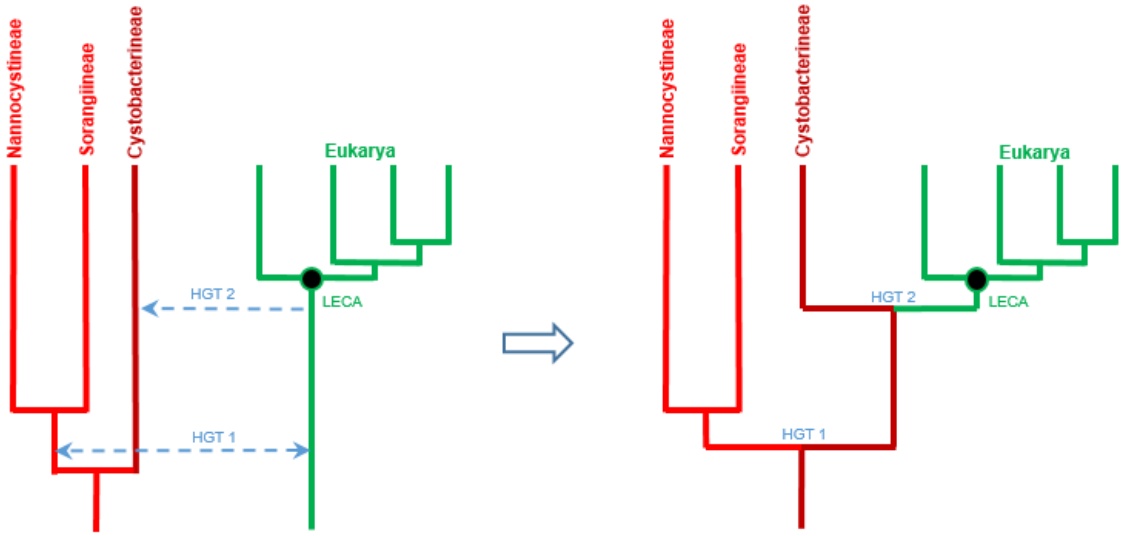
chitinase from fungi to Bacilli at 408 Ma (323 to 498 Ma, 95% confidence interval) and from Bacilli to Cystobacterineae at 275 Ma ago (174 to 389 Ma, 95% confidence interval) (see age annotation in Supplementary Fig. 1). Based on these age estimates, Cystobacterineae is a very young bacterial clade, and the origin of chitinase in these organisms may postdate LECA by over one billion years. Even if these age estimates were widely wrong, an ancestor of the group *Melittangium* + *Archangium* + *Cystobacter* + *Corallococcus* + *Stigmatella*, which comprises most of the clade depth of named Cystobacterineae species, did receive a fungal gene in a single HGT event[191]. Fungi are descendants of LECA. Thus, it is not possible that an ancestor of LECA acquired sterol biosynthesis from these Cystobacterineae. This means that the gene trees of Hoshino and Gaucher (2021)[33] apparently failed to produce the correct topology around the Eukarya-Cystobacterineae branching point.



Supplementary Fig. 1 | Phylogenetic tree modified from Figure 3 in Hoshino and Gaucher (2021)[33]. The tree is based on concatenated sequences of SQMO, OSC, CYP51, ERG24, ERG25, ERG26, ERG27-analog and CPI1 (i.e. genes generating and modifying the polycyclic core of sterols). Eukaryotes (green) branch within Bacteria among Cystobacterinea (dark red), a suborder of Myxococcales (red). The two nodes (black filled circles) are annotated with molecular clock ages for the last eukaryotic common ancestor (LECA) and the common ancestor of the Cystobacterinea species *Archangium gephyra*, *Cystobacter fuscus*, *Corallococcus coralloides*, *Stigmatella aurantiaca* (LCCA). For discussion of contradictory ages see text. The yellow node indicates the first hypothetical horizontal gene transfer (HGT) between stem-group Eukarya and Bacteria.

While the correct evolutionary relationship of sterol biosynthesis genes between Myxococcales and Eukarya remains unresolved, it is clear that crown-group eukaryotes should not branch within

Myxococcales Group II and, thus, sterol biosynthesis genes of Myxococcales Group II are either a close outgroup to crown-group Eukarya or branche within crown-group Eukarya. Both scenarios significantly reduce the probability that stem-group eukaryotes received a full syntenic set of sterol biosynthesis genes from Myxococcales. We thus suggest that the gene trees are best explained by two HGT events, as shown in Supplementary Figure 2: The first HGT event is the transfer of sterol biosynthesis genes (possibly in a syntenic set) between stem-group Eukarya and Myxococcales Group I (Nannocystineae + Sorangiineae). However, this transfer may have occurred in either direction. Second, a more recent HGT event of SQMO and OSC from Eukarya to Myxococcales Group II (Cystobacterineae). The second transfer may have occurred as recently as the Phanerozoic from crown-group eukaryotes to Cystobacterineae, or in the Proterozoic from late stem-group eukaryotes to Cystobacterineae (Supplementary Fig. 2a). Such a second gene transfer



Supplementary Fig. 2 | a, Two independent horizontal gene transfer (HGT) events of sterol biosynthesis genes from stem-group Eukarya (green) to Myxococcales (red) will cause a tree topology **b,** where eukaryotic sterol biosynthesis genes branch within Myxococcales, creating the possibly wrong impression that these genes were transferred from Cystobacterineae to Eukarya.

will create a tree topology where Eukarya branch within Myxococcales, creating the possibly wrong impression that sterol biosynthesis genes were transferred from Cystobacterineae to Eukarya. (Supplementary Fig. 2b). This alternative scenario also explains the absence of downstream Stage 2 genes in Myxococcales Group II without requiring the loss of these genes from Cystobacterineae.

Are two sterol gene transfers likely between eukaryotes and Myxococcales? Generally, HGT of sterol biosynthesis genes has occurred many times based on the large number of isolated bacterial groups that have acquired these genes independently, although these genes may have been largely transferred among bacterial clades. However, several groups of bacteria also possess sterol biosynthesis genes derived from crown-group eukaryotes (e.g. SQMO and OSC in Bacteroidetes, Methylococcales, and Alteromonadales within the Deltaproteobacteria), demonstrating that the transfer crown-group Eukarya → Bacteria has occurred. It is not unlikely that similar transfers occurred from the long stem of Eukarya to bacterial groups, and this may account for some of the occurrences of SQMO and OSC in bacteria that branch outside crown-group Eukarya. Many Myxococcales feed on insoluble eukaryotic substrates such as starch, cellulose, and fungal and arthropod chitin. The close physical association with eukaryotic biomass may increase the probability of gene transfer.

Hoshino and Gaucher (2021)[33] suggest that the synteny of sterol synthesis genes in Myxococcales Group I speaks for a simultaneous lateral transfer of all Stage 2 sterol biosynthesis genes as a set to stem-group Eukarya. If correct, then Myxococcales as well as the first sterol synthesizing eukaryotes had the ability to make C_{27} cholesteroids as early as 2,300 Ma ago, and by 1,700 Ma ago at the latest[34]. This implies that Myxococcales and eukaryotes were capable of C_{27} sterol production through the entire mid-Proterozoic interval. Yet, despite the abundance of protosteroids, not even traces of indigenous, saturated C_{27} steranes have been detected in sedimentary rocks during that time interval. In our opinion it is implausible that cholesteroid biosynthesis existed in eukaryotes and bacteria by 2,300 Ma ago but remained undetectable in the rock record until 800 Ma ago. It is more likely that a very ancient transfer of an entire Stage 2 syntenic set of sterol biosynthesis genes between Eukarya and Myxococcales did not occur and that downstream modification of the protosterol core evolved later.

In Summary, Hoshino and Gaucher (2021)[33] provided the most comprehensive and thorough analyses of the phylogenetic distribution and evolution of sterol biosynthesis in bacteria and eukaryotes to date. However, the conclusion that stem-group eukaryotes acquired sterol biosynthesis from Myxococcales Group 2 is poorly supported due to the very young age of that group. Future studies need to evaluate whether Myxococcales Group II branches within Eukarya or just outside Eukarya, and whether this group acquired its sterol biosynthetic genes from eukaryotes via HGT or via vertical inheritance from stem-group Myxococcales. The most difficult unresolved question is whether Myxococcales Group I obtained sterol biosynthesis from stem-group Eukarya or transferred it to Eukarya, and whether syntenic relationships were inherited via HGT or were achieved independently.

8.2.4. *The geological age of sterol producing bacteria*

To interpret protosteroids in the mid-Proterozoic biomarker record, it is important to assess which groups of extant bacteria had the capacity to produce protosterols between 1,640 and 1,000 Ma ago. With the exception of Myxococcales and Methylococcales, known bacterial SQMO and OSC have very shallow clade depths, and some even branch within crown-group Eukarya[33-35]. Thus, sterol production in most bacteria likely has quite recent origins via HGT from eukaryotes and among bacteria. Yet, it is not possible at present to determine precisely when various lineages of bacteria acquired sterol biosynthesis. The scarcity of calibration points in the bacterial fossil record, and hundred million to billion-year error bars associated with molecular clocks, make it difficult in most cases to determine which groups of extant sterol producers existed in the mid-Proterozoic interval. However, existing data provides some estimates for the two deepest branching lineages, Myxococcales and Methylococcales.

8.2.4.1. Myxococcales ('slime bacteria', Deltaproteobacteria)

The age of the last common ancestor of Myxococcales was roughly estimated at ~900 Ma, and the divergence of Myxococcales from other bacteria at ~1,000 Ma[192]. These estimates are broadly consistent with more refined molecular clock calculations based on the horizontal transfer of a fungal chitinase gene to Myxococcales calibrated with well dated fossils of fungi[191]. The clock places the base of suborder Cystobacterineae at 275 Ma (with 95% confidence interval from 174 to 389 Ma). Although we are not aware of molecular clock estimates for the ages of suborders Nannocystineae and Sorangiineae, a comparison of the data of Gruen et al. (2019)[191] with published phylogenetic trees[193-195] suggest that an origin of Myxococcales of around 1,000 to 1,200 Ma is reasonable. If these estimates are broadly correct, then it is unlikely that crown-group representatives of the order Myxococcales go back to 1,640 Ma, the age of the oldest known protosteroids.

8.2.4.2. Methylococcales (aerobic methanotrophs, Gammaproteobacteria)

The literature offers widely diverging views on the age of Methylococcales. Based on one study[196], Methylococcales and Thiotrichaceae (sulfide oxidizing bacteria, SOB) radiate near the base of the Gammaproteobacteria-Betaproteobacteria division (but also see ref[197] where the radiation of gammaproteobacterial orders is spread across nearly 1 billion years). Using molecular clocks calibrated on SOB symbionts in bivalves, the pioneering work of Canfield and Teske (1996)[198] suggested that the radiation of Thiotrichaceae occurred in the Neoproterozoic 760 ± 320 to 620 ± 280 Ma ago (and with an unlikely absolute age maximum of 1,400 Ma). Based on this data, the order Methylococcales may have a similar age and did not likely exist early in the mid-Proterozoic

interval. By contrast, a second molecular clock study places the origin of Methylococcales at ~1,850 to 2,200 Ma[197]. The authors applied the carotenoid biomarker okenane as a calibration point for Chromatiaceae, which is a sister group of Methylococcales. Okenane occurs in the 1,640 Ma Barney Creek Formation together with protosteroids and is regarded as a marker for purple sulfur bacteria[30,199]. However, the calibration age of 1,640 Ma was used as the minimum age for the last common ancestor of all Chromatiaceae while it would have been more correct to assign it to total-group Chromatiaceae, which would move the age for Methylococcales a few hundred million years to the present. In any case, if okenane is correctly interpreted as a biomarker for purple sulfur bacteria, then total-group Methylococcales should also have existed by 1,640 Ma. It is thus crucial to determine when sterol biosynthesis evolved *within* Methylococcales.

SQMO and OSC are observed in nine named genera of the order Methylococcales that fall in two different groupings. *Methylocaldum, Methylococcus, Methylospira, Methyloterricola* and *Methyloumidiphilus* belong to Group 1 of Wei et al. (2016)[35] and Hoshino and Gaucher (2021)[33], which branches outside crown-group eukaryotic sterol biosynthesis genes. These genera belong to the phylogenetic cluster of so-called Type 1b methanotrophs as defined by Knief (2015)[200]. Conversely, *Methylobacter*, *Methylomicrobium, Methylosarcina* and *Methyloprofundus* belong to a phylogenetic cluster within a second branch of Methylococcales, the Type Ia methanotrophs. Interestingly, this cluster possesses distinct protosterol biosynthesis genes that were obtained from crown-group eukaryotes via HGT[33-35]. As the cluster of sterol-producing Type 1a methanotrophs demonstrably obtained SQMO and OSC from crown-group eukaryotes, it is not unlikely that Type 1b methanotrophs obtained them from stem-group eukaryotes or from bacteria such as the Myxococcales.

It is difficult to ascertain when Type 1b methanotrophs acquired sterol biosynthesis genes. If we apply the most extreme molecular clock estimate where Methylococcales emerge at 2,200 Ma[197], then the last common ancestor of Type 1b may well have existed by 1,640 Ma. However, based on other estimates discussed above, methanotrophs likely acquired sterol biosynthesis later and are not a likely source of protosterols by 1,640 Ma.

Irrespective of the age of origin of sterol biosynthesis in methanotrophic bacteria, one additional observation speaks against a major contribution of steroids from Methylococcales to the mid-Proterozoic biomarker pool: all extant sterol-producing Type 1a and 1b Methylococcales, and thus presumably their respective last common ancestors, synthesize 4-methyl and 4,4-dimethyl cholesteroids as their major sterols. Yet, the corresponding saturated hydrocarbon biomarkers, 4-methylcholestane and 4,4-dimethylcholestane, have not been detected in mid-Proterozoic sediments of any age despite the fact that 4,4,14-trimethylated cholesteroids, including lanostane, are abundant. This observation suggests that crown-group representatives of Type 1a and 1b

Methylococcales were not a major component of the Protosterol Biota irrespective of their age of origin. This also means that an earlier interpretation that attributed triaromatic steroids of the 1,640 Ma Barney Creek Formation to methanotrophic bacteria[30] is no longer supported. However, the structure of these steroids is, in principle, still consistent with a stem-group methylococcalean origin.

Additional strong evidence against crown- or stem-group Methylococcales sources comes from the carbon isotopic composition of individual biomarker from the Barney Creek Fm. Methane generation in mid-Proterozoic marine sediments was likely generally higher than today, potentially providing substrate for substantial activity of methane oxidizing bacteria[30]. As methane consumed by Methylococcales is usually isotopically strongly depleted in ^{13}C, their lipids possess diagnostically low $\delta^{13}C$ values ($\delta^{13}C < -60‰$). This is incompatible with $\delta^{13}C$ values of 4-methylated triaromatic steroids from the Barney Creek Fm that are isotopically strongly enriched ($\delta^{13}C = -19$ to $-22‰$) relative to bulk biomass ($\delta^{13}C_{kerogen} = -30.4$ to $-32.1‰$) and cyanobacterial lipids ($\delta^{13}C_{\beta\text{-carotane}} = -37.3 \pm 0.9‰$)[36]. The carbon isotopic values of the steroids are consistent with a heterotrophic[36] or phototrophic source, but exclude Methylococcales as major sterol producers.

8.2.4.3. Summary

Based on current age estimates, none of the crown-groups of sterol-producing bacteria are likely sources of protosteroids going as far back as 1,640 Ma. Bacterial steroid contributions to mid-Proterozoic sediments are thus more likely derived from unknown or extinct bacterial lineages or from stem-group representatives of Myxococcales and Type 1b Methylococcales. However, this assessment is tentative as molecular clock estimates and phylogenetic tree topologies may change with emerging sequence data.

8.2.5. *Did bacteria or eukaryotes produce sterols in the Palaeoproterozoic at all?*

Based on molecular clocks, Gold et al. (2017)[34] estimate that the maximum marginal probability for the divergence of bacterial and eukaryotic protosterol biosynthesis genes is ~2,300 Ma, and the minimum age ~1,700 Ma. The authors cautiously and correctly did not claim that this age range corresponds to a HGT event of SQMO and OSC between bacteria and eukaryotes. This is because the computed divergence age might correspond to one of two alternate scenarios: (1) the age may reflect the bifurcation of two stem-group eukaryotic lineages, one of which later transferred SQMO and OSC to a bacterial group and then went extinct, while the second lineage gave rise to LECA. (2) The age may reflect the bifurcation of two bacterial lineages, one of which later transferred SQMO and OSC to stem-group eukaryotes and the other to an extant bacterial lineages (e.g.

Myxococcales), and then both went extinct. Scenario (1) is not unlikely given the generally high frequency of HGT from eukaryotes to bacteria, the fact that all stem lineages went extinct (except the one giving rise to LECA), and the knowledge that transfers of SQMO and OSC from eukaryotes to bacteria do in fact occur[34,35].

There are additional hypotheses about sterol origins that are currently intractable. For instance, stem-group eukaryotes and extant bacterial lineages may have received protosterol biosynthesis from an extinct bacterial group, or from a bacterial group that has since lost these genes or remains to be discovered.

8.2.6. *The sterols produced by bacteria*

For many bacterial species that possess SQMO and OSC, it remains unknown whether these genes are expressed and, if so, which sterols are produced. In particular, sterol biosynthesis has not been confirmed through lipid analysis in any representative of Cyanobacteria, Actinobacteria and Verrucomicrobia[35]. However, for 18 species, sterol biosynthesis has been confirmed through lipid surveys. *Gemmata obscuriglobus* (Planctomycetes) produces lanosterol and its double-bond isomer parkeol[187]. The sterols are essential to growth and function as membrane molecules[201]. *Fluviicola taffensis* (Bacteriodetes) and the aerobic methanotroph *Methyloceanibacter caenitepidi* (Alphaproteobacteria) exclusively produce cycloartenol, while aerobic methanotrophs of the Methylococcales (Gammaproteobacteria) have the capacity to demethylate protosterols in the C-14 position and remove one methyl group from the C-4 position. The final sterol products are 4-methyl and 4,4-dimethyl cholesteroids. Methylococcales do not appear to accumulate notable amounts of C_{30} protosterols[35,189,202]. Myxococcales (Deltaproteobacteria) of the suborders Nannocystineae and Sorangiineae (part of Group 1 and branching deeply outside Eukarya[35]) have the distinguishing capacity to remove all methyl groups from positions C-4 and C-14 to produce the C_{27} sterol zymosterol as the main product. *Enhygromyxa salina* (Nannocystineae) biosynthesizes cholesterol that is bound to a lipid or protein moiety but does not occur as a free membrane component[203]. By contrast, representatives of the suborder Cystobacterineae, which form a separate cluster in the OSC tree closer to eukaryotes, either did not contain detectable sterols or produce cycloartenol and/or lanosterol[35,190]. An unusual case is *Eudoraea adriatica* (Bacteroidetes) that only yielded traces of lanosterol while the main products of the OSC-like cyclase were two pentacyclic terpenols with an arborane skeleton, eudoraenol and adriaticol[204].

8.2.7. *Environments inhabited by sterol producing Bacteria*

Bacteria with genes for protosterol biosynthesis are found in a wide variety of habitats. OSC protein sequences were found in metagenomes from soil, in marine and freshwater environments, estuarine microbial mats, hydrothermal vent fluids, wastewater and in the deep subsurface. Most

sterol producing Myxococcales were acquired from soil, but at least two come from marine environments. Sterol producing Methylococcales were enriched from sewage, marine pelagic environments, hot springs, freshwater lake sediments, and soda lake sediments. The alphaproteobacterial methanotroph *Methyloceanibacter caenitepidi* (Rhizobiales) was isolated from a marine hydrothermal vent, *Eudoraea adriatica* came from a coastal sediment, and *Fluviicola taffensis* from river sediment[35]. *Gemmata obscuriglobus* was isolated from the littoral zone of a freshwater lake. Based on genomic data, no particular environment stands out as a preferred habitat for bacterial sterol producers.

8.2.8. *Summary*

Phylogenetic studies on sterol biosynthesis provide six notable insights: (1) SQMO and OSC of extant bacteria and crown-group eukaryotes most likely diverged ~2,300 Ma ago and probably before 1,700 Ma ago[34]. (2) After this first major divergence of SQMO and OSC, at least one transfer of these genes occurred between stem-group eukaryotes and bacteria, but the direction of transfer remains unknown (see ref[33] for a contrasting view). (3) At least one transfer of SQMO and OSC occurred from crown-group eukaryotes to bacteria, and these eukaryotic genes are now found in representatives of Methylococcales, Alteromonadales and Bacteriodetes. (4) The distribution of sterol biosynthesis genes in bacteria is sporadic and complex and presumably affected by a combination of frequent HGT events and gene losses, some vertical inheritance (chiefly in Myxococcales and Methylococcales) and minor convergent evolution[205] (see also existence of SQMO and alternative-SQMO in eukaryotes[206]). (5) The only bacterial clusters where sterol biosynthesis has very deep roots is Myxococcales, but it is currently not possible to state with certainty when this group emerged and whether it predates LECA.

8.3. Comparison of sterol abundances in eukaryotic and bacterial cells

To assess whether bacteria can make a significant contribution to the sterol content of sediments and sedimentary rocks, we need to compare the absolute sterol content of bacterial and eukaryotic cells. All sterol contents in the following discussion are reported in percent of dry weight.

The sterol content of plant matter can range from very low to very high depending on wood and carbohydrate contents and presence of sterol-rich oils. Lipid surveys measured sterol contents of 0.02-0.4% in vegetables, 0.05-0.29% for fruit[207], 0.04-0.22% in berries, 0.02-0.1% in the kernels of grain crops, 0.1-0.3% in seeds of oil crops, and 0.1-0.2% in legumes[208]. These values are comparable to a 0.1% sterol content in the polypore mushroom *Polystictus versicolor*[189]. The sterol content of algae is generally higher than of land plants with an average content of 0.31 ± 0.18% (range 0.06-0.65%) in 18 species of macroalgae[209] and an average of 0.51 ± 0.71% (range 0.02-2.6%) in ten species of microalgae[210]. Yet higher is the average sterol content of single celled

protists and fungi. The free living amoeba *Acanthamoeba polyphaga* contains 0.4-0.5% sterols[211] and the fungal mould *Aspergillus nidulans* yielded 0.6-0.9%[189]. A sterol survey[163] on basal flagellated fungi found high sterol contents of 0.3-3.6% (average 1.7 ± 1.1%, n = 8). Similarly high sterol contents of 1.4% and 2.2% were found in two heterkonts of the phylum Hyphochytriomycota, protists which possess a fungi-like rhizoidal or hypha-like vegetative system[163].

Bloch (1983)[28] suggested that the few bacteria that appear to produce sterols only contain traces. Although there is still only limited reliable information on the abundance of sterols in bacteria, we now know that sterol concentrations per dry weight are not significantly different from eukaryotes. Reports of very low concentration are most likely based on culture contamination[156] such as traces of cholesterol (0.0035%) in *Streptomyces olivaceus*[212], 0.01% of a mixture of six sterols (including ergosterol) in *Azotobacter chroococcum*[213] and traces in some other bacteria[214]. Reliable data exists for *Gemmata obscuriglobus* with 2.0%[187], *Nannocystis exedens* with 0.4%[215] and *Methylococcus capsulatus* with 0.22%[189]. Extremely low concentrations of 0.00002% lanosta-8,22,24-triene were detected in *Methylobacterium organophilum* (Alphaproteobacteria, Rhizobiales)[216]. This latter report seems credible as this unusual sterol is not a likely contaminant and the related bacterium *Methyloceanibacter caenitepidi* is known to produce sterols. Yet, it cannot be excluded that the sterol was contributed by other, minor methanotrophic bacteria present in the medium, so the existence of such low sterol concentrations needs to be re-evaluated in pure culture.

Based on this limited data, the sterol content of bacteria per dry weight is in the same range as in eukaryotes, supporting the hypothesis that sterols can be a notable component of bacterial membranes. Bacterial sterols may thus comprise a significant fraction of sedimentary lipids in those environments where sterol-producing bacteria are abundant (Section 9).

9. Modern environments and sediments with elevated protosterol levels

All organisms that produce sterols also contain protosterols as biosynthetic intermediates, commonly at low percentage levels. Thus, all modern environments and sediments that contain sterols should yield protosterols. Yet, there are very few studies that report these compounds, presumably because the small chromatographic signals are overlooked or not reported. By contrast, protosterol abundances far above typical biosynthetic background levels are usually conspicuous in chromatograms and difficult to overlook. Yet, despite the vast number of sterol surveys of recent environments and sediments, elevated protosterol abundances are very rarely observed, and the few instances in the literature invariably point to uncommon and unusual environments such as methane seeps and microbial mats. It thus appears that protosterols rarely

exceed background levels in almost all modern marine, lacustrine and terrestrial environments. Known exceptions are discussed below.

9.1. Cycloartenol in the Lost City hydrothermal system

Lost City is an ocean-floor hydrothermal field located at a depth of 750 to 900 meters at the mid-Atlantic ridge. At Lost City, hot reducing fluids rich in hydrogen and methane and with temperatures of < 40 to 90ºC, emanate from large carbonate chimneys. A lipid survey[217] at Lost City detected cycloartenol in concentrations approaching 500 ng per gram of dry rock. Cycloartenol was the most abundant sterol in some samples. In one case, cycloartenol occurred at an abundance of 100 ng/g, while the only other sterol, cholesterol, merely yielded 40 ng/g rock. Moreover, in that sample, cycloartenol was enriched in ^{13}C by more than 12‰ relative to cholesterol. Based on the high relative abundance and diverging carbon isotopes, the authors concluded that the protosterol at Lost City is a final product, not a biosynthetic intermediate[217]. The authors excluded Myxococcales as a source of cycloartenol as these were not detected in rRNA gene surveys. The authors further concluded that "*the most probable source of abundant cycloartenol is one or more of the protists inhabiting the carbonate chimneys*", noting that "*a diverse population of protists has been detected in Lost City carbonates. Ciliates are the most dominant, and other alveolates, fungi, heterokonts, radiolaria and other cercozoa, and heterolobosea have also been detected* (citing ref[218])". However, the authors also cautioned that the sterol biosynthetic pathway of most protists remains unknown[217]. As protosterol biosynthesis requires ten times less oxygen than crown-sterol biosynthesis, they further speculated that (quote) "*cycloartenol might be a favored sterol from an amphiaerobic eukaryote inhabiting vent fluids with only sporadic access to oxygenated seawater"*. Based on the carbon isotopic enrichment of the Lost City cycloartenol, the authors further suggest that it may derive from protists that inhabit the carbonate chimneys and feed on prokaryotes that are part of the methanogenic community. We agree with the assessment that the protosterol in Lost City may, in principle, derive from a crown-group eukaryote adapted to microaerobic conditions. However, more recent studies reveal additional groups of bacteria, apart from Myxobacteria, that produce cycloartenol as their final sterol product, including one representative of the phylum Bacteriodetes and, notably, the aerobic methanotroph *Methyloceanibacter* (Alphaproteobacteria) that inhabits marine sediments at hydrothermal vents[35]. A bacterial source is thus likely.

Crown-sterols were also surprisingly abundant at Lost City. Detected were cholesterol at concentrations of up to 580 ng per gram of dry carbonate, attributed to animals, C_{28} and C_{29} sterols likely derived from phytoplankton, and ergosterol ascribed to fungi that had been detected at the site. The abundance of crown-sterols in carbonates from a deep-water hydrothermal vent is notable.

9.2. Cycloartenol in benthic mats of the Black Sea

Protosterols were also detected in a study on a 1 to 5 cm thick microbial mat covering several m^2 of sea floor at 180 m depth in the Black Sea[219]. The mat covered the interior of a 10 m wide caldera-like structure (pockmark), presumably the remnant of an old gas outburst. The rim of the pockmark consisted of carbonate crust generated by ANME (anaerobic methanotroph) activity. The microbial mat was floating on the seafloor, trapping methane rich fluids emanating from below. At a depth of 180 m, the mat was located just below the anoxic/oxic interface in the water column. ANME consortia were active in the mat, while aerobic methane oxidizing bacteria were detected at the top of the mat, possibly becoming active upon sporadic oxygen exposure. Alternatively, the aerobic methanotrophs are of planktonic origin from the overlying water column.

Solvent extracts of the mat yielded cycloartanol, minor cycloartenol and 4,4-dimethyl-cholestenol[219]. The authors refer to these sterols as "major" with up to 3.6 µg/g dry weight for cycloartanol, but no information is provided about the presence and abundance of other sterols, such as phytosterols, so it is difficult to place these concentrations into context. The carbon isotopic composition of cycloartanol ranged from -74.4‰ to -69.8‰, and 4,4-dimethylcholestenol from -80.9‰ to -64.3‰. However, in individual extracts, $\delta^{13}C$ of the two sterols diverged considerably, with offsets ranging from +8.6‰ to -7.9‰. The carbon isotopic compositions suggest that the source organisms directly or indirectly consumed methane-derived carbon, and the isotopic offsets imply that the two sterols have different biogenic sources[219]. However, the bulk exopolymeric substance (EPS) of the mat was also carbon isotopically strongly depleted (-73‰), suggesting that the bulk of the mat community is fuelled by methane.

The authors of the study suggest that the 4,4-dimethylcholestenol is either derived from methylococcalean bacteria consuming methane at the top of the mat or from "a microbe thriving in the water column above"[219]. We agree that Methylococcales are a plausible source but consider the suggested mat origin more likely than planktonic sources based on the high concentration of the sterol of up to 15 µg/g dry weight and based on the observations that this sterol is not commonly observed at such concentrations elsewhere in Black Sea sediments. For cycloartenol and cycloartanol, the authors suggested eukaryotic ciliates as a source, feeding on isotopically depleted biomass in the mat or in the water column above. They noted that ciliates are generally an important component of the Black Sea food chain and point to a study[220] that reported cycloartenol in ciliates. However, based on the cited work, cycloartenol is only a minor sterol in ciliates and presumably of dietary origin[220,221]. In fact, based on current knowledge, ciliates do not appear to produce or contain protosterols as major membrane lipids[157,159]. Yet, we agree in principle that cycloartanol in Black Sea mats may derive from eukaryotes thriving under oxygen limiting conditions. This possibility is supported by PCR data revealing eukaryotic ribosomal genes in the

mats[219], although no further information about the nature of these organisms is provided. However, it is now known that the methane consuming alphaproteobacterium *Methyloceanibacter* produces cycloartenol (although not cycloartanol)[35], and we suggest that these or related methanotrophs are the most plausible source of the protosterols in the Black Sea mat.

9.3. 4-Methylated cholesteroids in a methane-emitting mud volcano

A study of steroids and hopanoids in microbial mats associated with a methane-emitting mud volcano at the Norwegian margin of the Barents Sea yielded abundant 4-methylated cholesteroids and traces of 4,4-dimethyl cholesteroids that were carbon isotopically depleted by -68 to -77‰, pointing to organisms directly or indirectly consuming methane carbon[222]. A 16S rRNA survey confirmed the presence of aerobic methanotrophic bacteria of the order Methylococcales (Gammaproteobacteria) that are known to produce 4-methylated cholesteroids. However, the protosterols lanosterol and cycloartenol were not detected at the site, consistent with the observation that all known sterol-producing Methylococcales remove carbon C-14 from lanosterol to produce downstream sterols. The crown-sterols cholesterol and brassicasterol were detected in concentrations similar to the 4-methylated sterols. Their carbon isotopic composition was around -30‰, indicating a eukaryotic planktonic origin.

9.4. Cycloartenol in cyanobacterial mats

Protosteroids and protosteroid derivatives have also been observed in modern cyanobacterial mats alongside crown-group sterols, occasionally in substantial abundances. In extracts of different microbial mat layers at Solar Lake, North-East Sinai, Edmund and Eglington (1984)[223] observed a varied and diverse distribution of sterols up to a depth of 70 cm. Among the sterols were cycloartanol, oxidation products such as cycloartanone, and cycloartanol derivatives demethylated at C-4 once or twice. Cycloartanol and cycloartanone were found in all layers ranging from 2% at the surface, 0.8% at a depth of ~1 cm, and rising sharply to a maximum of 19% at 27 cm (percentages relative to total sterols). Cycloartanoids that lost one or two methyl groups at C-4 were even more abundant with 11% at the surface and a maximum of 21% at 27 cm. The maximum abundance of all cycloartanoid derivatives combined was 37% at a depth of 27 cm beneath the surface of the mat.

The biological origin of the protosteroids in Solar Lake mats remains unknown. In particular, the observation of cycloartanoids demethylated at C-4 once or even twice, leading to 4,14-dimethylated and 14-methylated steroids, is unusual as these sterols have not been recorded in bacteria and are also not typical biosynthetic intermediates of crown-sterol biosynthesis. While 4,14-dimethylated sterols occur as biosynthetic intermediates between cycloartenol and phytosterols in plants, these compounds are also always methylated at C-24, which is missing in

the Solar Lake sterols. However, the excavate protist *Naegleria gruberi* contains ~6-8% cycloartenol and ~7% 4,14-dimethylcholesterol derivatives relative to total sterols (computed from data in ref[184]). Generally, the sterol content of eukaryotic protists is poorly explored, making it likely that *Naegleria* is not the only genus with such unusual sterols. Hence, the cycloartanol and the 4-desmethyl cycloartanoids in Solar Lake mats may well have a eukaryotic origin. It is also currently impossible to exclude the hypothesis that these sterols and sterones became enriched by heterotrophic processes, e.g. through selective degradation of crown-group sterols, and that the 4-demethylated cycloartanoids are diagenetic products without direct biogenic precursors.

In cyanobacteria-dominated microbial mats from Abu Dhabi, the relative proportion of cycloartenol among sterols is yet higher at ~60% of total sterols[224], but still one order of magnitude lower than hopanols. A similar study of lipids preserved in flat-laminated cyanobacterial mats from the Guerrero Negro saltworks found that the surface layer of the mat mainly contained crown-group sterols (C_{27}, C_{28} and C_{29}) with minor contributions of cycloartenol and several 4-methyl sterols[225]. However, in an apparently anoxic layer at a depth of 0.4 to 1.0 cm, the concentration of protosterol derivatives "increased dramatically" (quote ref[225]). However, absolute and relative abundances of the protosterols were not reported and it remains unclear whether they are as abundant as in the mats from Abu Dhabi and Solar Lake. Furthermore, no comments were made about possible origins of the cycloartenol in the Guerrero Negro saltworks. Yet, contrary to the older studies, a genomic survey assessed the activity of eukaryotes in the mat. The survey reported 15 species of eukaryotes, more than half of which were nematodes, one insect, one crustacean, a stramenopile alga and, intriguingly, a eukaryote cluster with no known kingdom level affiliation. Thus, bacterial as well eukaryotic sources are possible for the protosterols, and an enrichment of these compounds by selective heterotrophic processes cannot be ruled out.

9.5. Cyclolaudenol in SOB mats beneath the Peru upwelling zone

McCaffrey et al. (1989)[226] analysed the lipid composition of microbial mats covering surface sediments beneath dysaerobic (< 0.1 mL/L) bottom waters in the coastal upwelling zone off the coast of Peru at a depth of 73 to 100 meters. The mats were predominantly (80%) composed of thick, filamentous sulfide oxidizing bacteria (SOB) of the genus *Thioploca* (Gammaproteobacteria). For lipid analysis, *Thioploca* filaments from two mats collected 17 km apart were cleaned with distilled water and inspected with a binocular microscope for contaminants. The enriched *Thioploca* filaments yielded ~4% lipids by dry weight, mostly comprising fatty acids and 0.008% cyclolaudenol per dry weight total biomass, but no hydrocarbons or hopanoids. The mat extracts also included traces of unspecified sterols, but cyclolaudenol constituted 95% of total sterols. By contrast, sediment directly beneath a *Thioploca* mat only contained traces of cyclolaudenol (1.3 μg/g dry weight) but much higher concentrations

of other sterols of presumably planktonic animal and algal origins (29 µg/g cholesterol; 22 µg/g dinosterol). This distribution suggests that the cyclolaudenol was produced within the mat and is not derived from planktonic sources or from the sediment below. The distribution also demonstrates that cyclolaudenol is the final sterol product and not a biosynthetic intermediate.

Cyclolaudenol is a rare C_{31} sterol, a cycloartanoid methylated at the side chain in position C-24 (24-methylcycloart-25-enol). Its detection in a microaerobic mat is interesting as it possesses the same carbon skeleton as 24-methylenecycloartenol, the second sterol intermediate in the biosynthesis of phytosterols in land plants[17], and it is a potential diagenetic precursor for the 4,24-dimethylated triaromatic steroids in the ~1,300 old Roper Group. It is thus important to understand whether cyclolaudenol is biosynthesized by *Thioploca*, whether SOB generally have the capacity to produce cycloartanoids, and whether SOB may be the source for 24-methylated steroids in the mid-Proterozoic.

Due to the low sterol content of only 0.008% per dry weight of mat, it is not possible to assign cyclolaudenol to *Thioploca* with certainty. KOH treatment of the lipid extract and alkaline and acid hydrolyses of the solid residue did not release any further sterols[226], confirming the low concentration. Cyclolaudenol in the *Thioploca* mats is thus ~30 to 250 times less abundant than confirmed sterols in bacteria (Section 8.3). SOB mats may host numerous other species including prokaryotes and microeukaryotes, and smaller bacteria growing on filament surfaces can theoretically make up 60% of the total lipid mass[227]. For example, 40 to 46% of the sterols detected in an SOB mat of the sister genus *Beggiatoa* were contributed by eukaryotes that consume biomass within the mat[227]. It is thus plausible that cyclolaudenol is not produced by *Thioploca* but derived from other prokaryotes or microeukaryotes inhabiting SOB filaments. Some bacteria possess the SMT gene required to methylate the side chain[33]. Such organisms would have remained invisible during the binocular microscopic inspection by McCaffrey et al. (1989)[226].

While cyclolaudenol cannot be assigned to a specific organism at present, *Thioploca* cannot be ruled out as the source. Although the sterol content in dried *Thioploca* mats is very low, the mass of the *Thioploca* sheaths may supresses the content considerably. We thus have to investigate whether cyclolaudenol has been observed in other SOB mats and examine genomic evidence for or against sterol biosynthesis in the group.

There are several studies that provide lipid analyses of SOB mats, or of sediments associated with SOB mats, that explicitly report sterol contents, however cyclolaudenol or protosterols were not detected in any of these analyses. One study[228] analysed surface sediments in the upwelling region off Namibia, Peru and Chile with dense populations of SOB *Thiomargarita*, *Beggiatoa*, and *Thioploca*. *Thioploca* was dominant in the Chilean sediments. Lipid analysis yielded typical, previously observed[226] SOB fatty acid patterns. The study also included a detailed survey of

polyterpenoids where diverse algal sterols and hopanoids were observed. Yet, cyclolaudenol was not detected in any of the SOB communities. A second study[222] provided a detailed study of sterols in a *Beggiatoa* mat at an active marine mud volcano at the Norwegian margin of the Barents Sea. The survey detected unusual 4-methylated cholesteroids that were assigned to methylococcalean bacteria based on their carbon isotopic composition. However, cyclolaudenol or protosterols were not detected. A third study[227] purified sterols extracted from *Beggiatoa* mats from a hydrothermal vent in the Guaymas Basin, reporting C_{27} to C_{29} sterols attributed to eukaryotes, but no cyclolaudenol or protosterols. Likewise, a fourth study[229] analysed lipids from the *Beggiatoa alba* layer of an intertidal microbial mat in Qatar, detecting abundant algal sterols but no unusual sterols. A fifth lipid survey[230] reported lipids from a well-preserved fossil *Beggiatoa*-like microbial mat consisting of filamentous fossils in Miocene methane-seep limestone from the Romagna Apennine (Pietralunga, Italy). They study reported an unusual pseudohomologous series of C-4 methylated C_{29}, C_{30}, and C_{31} steranes. The C_{30} compound was identified as lanostane. The steranes were strongly depleted in ^{13}C, suggesting a methanotrophic bacterial source. The C_{31} compound is described as a derivative of lanostane "methylated in ring A or ring B" (quote ref[230]), and can thus not be a cyclolaudanoid or other side-chain methylated protosteroid. Based on a chromatogram given in Figure 4 of ref[230], the concentration of lanostane was substantially higher than of other steranes (which are not visible in the chromatogram) and also higher than C_{30} hopene, the only hopanoid in the chromatogram.

In a comprehensive lipid analysis of three species of *Bathymodiolus* mussels from the Pacific Antarctic Ridge hosting methane-oxidizing and/or SOB, 4-methyl cholesteroids, 4,4-dimethyl cholesteroids and lanosteroids were found in bivalves with methanotrophic symbionts, an origin that was confirmed by the carbon isotopic composition of the sterols[231]. However, bivalves harbouring SOB symbionts, but lacking methanotrophic bacteria, explicitly lacked unusual sterols. Similarly, analysis of sterols of the bivalve *Solemya velum* harbouring SOB symbionts yielded abundant sterols but no cyclolaudenol or protosterols. Likewise, analysis of the lipids of the SOB *Thiomicrospira crunogena* failed to detect any sterols[232]. Based on a statement in the acknowledgements, the authors of that study received advice on lipid analysis from the first author of the original study that had detected cyclolaudenol in *Thioploca* mats[226]. It is thus likely that cyclolaudenol was absent rather than overlooked. There are also numerous studies that investigated lipids in SOB mats but do not mention whether sterols were monitored or whether cyclolaudenol or protosterols were present[233-242]. Yet, many of these studies cite the work that had reported cyclolaudenol in *Thioploca* mats[226], so we surmise that at least some of these studies would have detected cyclolaudenol if it was present. Based on lipid analyses, it is thus clear that sterol biosynthesis is not widespread among SOB, updating the statement of Peckmann et al. (2004)[230]

1801 that “sterol biosynthesis has so far been corroborated neither for *Thioploca* nor for any other
1802 sulfide-oxidizing bacteria.”

1803 Biosynthesis of sterols by SOB is also not supported by genomic analyses. There are now hundreds
1804 of genomes available in the Joint Genome Institute Integrated Microbial Genomes database of the
1805 order Thiotrichales, and dozens for Thiotrichaceae, the family comprising SOB, including finished
1806 genomes of the genera *Beggiatoa*, *Thiomicrorhabdus*, *Thiomicrospira*, and *Thioploca,* and draft
1807 genomes of *Achromatium*, *Thiomargarita* and *Thiothrix*. For the genus *Thioploca*, genomes are
1808 available for *T. ingrica* and *T. araucae*[243]. A surveys of homologs of OSC through BLASTP in all
1809 bacterial genomes (31,237) available at that time in the JGI database yielded 34 bacterial OSC
1810 homologs in five different phyla, but no Thiotrichales[35]. Similarly, a search for SQMO and OSC
1811 protein sequences in the National Center for Biotechnology Information (NCBI) protein database
1812 found 27 bacterial taxa, across 6 phyla, 9 classes and 9 orders, but again did not observe any hits
1813 among Thiotrichales[34]. Likewise, the most recent and vastly expanded survey of sterol
1814 biosynthesis genes did not yield any hits for Thiotrichales[33]

1815 Based on the combined lipid and genomic data, we conclude that cyclolaudenol biosynthesis in
1816 SOB, if it occurs at all, would be restricted to a specific *Thioploca* strain found in the upwelling
1817 region of Peru[226]. Based on molecular clock estimates, this strain would have gained the required
1818 genes via horizontal gene transfer relatively recently. SOB (Thiotrichaceae) radiate near the base
1819 of the Gammaproteobacteria-Betaproteobacteria division[196]. Using molecular clocks calibrated on
1820 SOB symbionts in bivalves, Canfield and Teske (1996)[198] estimated that this radiation of
1821 Thiotrichaceae occurred in the Neoproterozoic 760 ± 320 to 620 ± 280 Ma ago (and with an
1822 unlikely absolute ago maximum of 1,400 Ma). Even if these divergence estimates are wrong and
1823 true ages are substantially greater, any particular *Thioploca* strain must have acquired sterol
1824 biosynthesis relatively recently. It is thus unlikely that SOB contributed to the mid-Proterozoic
1825 steroid pool, and exceedingly unlikely that *Thioploca* is the source for 4,24-dimethylated steroids
1826 in the ~1,300 Ma Roper Group.

1827 ### *9.6. Lack of bacterial sterols in dysoxic water columns*

1828 In contrast to microaerobic benthic environments, we were not able to find reports on notable
1829 abundances of protosterols, or bacterial 4-methylated cholesteroids, within water bodies. Even in
1830 euxinic water bodies that may serve as closest analogues for Proterozoic basins, protosterols
1831 remain inconspicuous. For example, one study[244] provided a detailed survey of sterols in the
1832 particulate organic matter of the water column and in bottom sediments of the Black Sea, with
1833 particular attention to the chemocline. While crown-sterols were highly abundant throughout the
1834 water column and in the sediment, protosterols and 4,4-dimethylcholesteroids were not detected,

and minor abundances of 4-methyl cholesteroids were attributed to dinoflagellates. This is corroborated by a second survey[245] of lipids of aerobic methanotrophic bacteria in the chemocline of the Black Sea that failed to detect any unusual sterols.

10. The geological record of protosteroids

10.1. Average protosteroid abundances through time (this study)

There are only few published studies reporting fossil protosteroids. As with modern protosterols, the likely reason is that these biomarkers are rarely conspicuous in total ion chromatograms and only detectable using targeted MS experiments. However, with appropriate elution standards, identification of these biomarkers is straightforward. In the present study, all thermally well preserved Palaeoproterozoic to Phanerozoic bitumens and oils yielded, without exception, fossil protosteroids. However, the proportion of protosteroids to crow-group steroids changes notably through geological time. Supplementary Table 1 shows that the average proportion of aromatic protosteroids relative to total aromatic steroids for the Palaeo- and Mesoproterozoic is 100 ± 0% (n = 42; average ± stdev), for the Tonian and Cryogenian 33 ± 24% (n = 34), for the Ediacaran 5.8 ± 6.3% (n = 9), and for the Phanerozoic 3.8 ± 4.0% (n = 36). We were not able to compute reliable corresponding values for saturated protosteranes because the major lanostane isomer coelutes with the tetracyclic triterpane 'TPP', and we also lack data for most samples of the Geoscience Australia Phanerozoic oil collection (Supplementary Table 3). However, the approximate percentages of cyclosterane over steranes for the Palaeo- to Mesoproterozoic is 100 ± 0% (n = 34; average ± stdev), for the Tonian and Cryogenian 14 ± 19% (n = 22), for the Ediacaran 0.5 ± 0.5% (n = 8), and for the Phanerozoic 0.9 ± 1.4% (n = 15).

10.2. Fossil protosteroids in the literature

10.2.1. *Protosteroids in sub-Recent sediments of Lake Caçó*

Lake Caçó is a small oligotrophic lake in north-eastern Brazil. Jacob et al. (2005, 2007)[246,247] analysed the aromatic and polar triterpenoids extracted from (sub)recent sediments in a 6 m long Lake Caçó core. The sediments had been deposited during the last 20,000 years under different environmental conditions during wet and dry seasons[246] and, according to previous chemo-taxonomic studies, grasses (Gramineae) colonized the savannas of Northern Brazil at the time of deposition[247]. In particular the upper 3 m of core are thought to be influenced by influx of material from the development of a belt of the sedge-plant spike-rush[246].

The lake sediment extracts yielded dominantly plant-derived aromatic biomarkers with only few saturate triterpanes such as onocerane and some hopanoids. Strikingly, aromatic triterpenoid

fingerprints from two samples (1.6 m and 4.5 m depth) resembled those described for a lignite and sub-bituminous coal respectively[248,249], comprising a series of pentacyclic monoaromatic compounds with various structures and unsaturations, triaromatic pentacyclic compounds, and a series of supposedly diaromatic tetracyclic triterpene derivatives dominated by a *m/z* 195 peak—mass spectral characteristics also observed in unidentified triterpenoids reported in the lignite and coal[248,249]. Based on mass spectral interpretations, Jacob et al. (2007)[246] provided the first tentative identification of these biomarkers as diaromatic tetracyclic triterpenes with a lanostane/euphane structure. These compounds show the same mass spectra (M^+ shifted downwards by 2 Da for those isomers still containing a double bond in the side chain), as the diaromatic lanosteroids produced in our protosterol pyrolysis experiments, supporting the structural assignment of the authors. In total ion chromatogram (TIC), the lanosteroids are amongst the most abundant aromatic triterpenoids (Fig. 3 in ref[246]), attesting to exceptionally high protosteroid vs crown-steroid and hopanoid ratios in some of the plant/terrestrial-influenced lacustrine sediments.

The aromatization in Lake Caçó is extremely early and may have occurred in the watershed prior to entering the lake waters and sediments. The lake sediments contain a variety of typical plant-derived triterpenoids such as oleanane-derivatives. Accordingly, Jacob et al. (2007)[246] interpret the aromatic triterpenoids to largely derive from higher plants, and the authors also consider these as the most likely sources of lanosteroids. As grasses (Gramineae) are thought to be amongst the dominant sources of triterpenoids in Lake Caçó[246,247], grasses are also likely sources of the lanosteroids. This inference is consistent with the tentative identification of a lanosteroid methyl ether in Lake Caçó[246,247] as protosterol methyl ethers do occur in several Gramineae species[250]. However, organic matter from trees, mosses or other plants, as well as washed-in plant waxes are also plausible sources for the elevated lanosteroid content (see Section 8.1.2).

10.2.2. *Lanostanes from an Eocene saline lake*

Chen et al. (1989)[136] were the first to discover lanostanes in the sedimentary record. They detected saturated lanostane, 24-methyllanostane and 24-ethyllanostane in Eocene sediments from a saline lake of the Biyang Basin, China. The lanostanes only occurred in higher salinity samples and were not detected in sediments deposited under lower salinities. The abundances of C_{30} lanostane and C_{31} 24-methylated lanostane were similar, while the ethylated homolog was about three times lower. Yet higher homologs were not detected. Information about the abundance of the lanostanes relative to other steranes was unfortunately not provided. However, based on an MRM 414 → 259 chromatogram in Figure 7 of ref[251], lanostane is less abundant than 3β-propylcholestane, a diagenetic propylation product that commonly only occurs in traces in bitumens and oils. It is thus probable that the abundance of lanostanes relative to other steranes was very low (sub percentage levels). These low-concentrated lanostanes may have a plant origin. The bitumens extracted from

the Eocene lake sediments yielded odd-over-even predominant *n*-alkane patterns, the plant terpanoid oleanane and a dominance of stigmastane among other steranes, all pointing to influx of plant matter into the lake[251]. Moreover, side-chain methylation and ethylation of protosterols is currently unknown among bacteria but common in plants, several of which producing side-chain methylated and ethylated protosteroids in high relative abundances (Section 8.1.2).

C_{30} to C_{32} lanostanes were also observed in Eocene lime-mudstones deposited under saline stratified waters of the Bohai Bay basin, China[252]. The bitumens contained highly abundant oleanane, so a higher plant source for the side-chain methylated and ethylated lanostanes is plausible here as well.

10.2.3. *Other protosteroid occurrences*

Solvent extracts of Miocene Pietralunga methane seep limestone yielded lanostane with a carbon isotopic composition $\delta^{13}C$ = -70‰ and 4-desmethyl lanostane with $\delta^{13}C$ = -80‰[253] . The authors attributed the steranes to methanotrophic bacteria (see also Section 9.5).

A highly sulphidic Paleocene crude oil from Jianghan Basin, China, yielded cyclic sulfide derivatives of C_{30} lanostane and C_{31} 24-methyl lanostane[254]. While hopanes and steranes were abundant in all fractions of the oil, the lanostanes were exclusively detected in the sulfide fraction[254], pointing to a specific unsaturation pattern that promoted sulfurization, or to an unusual biogenic origin that exposed these sterols preferentially to a sulfur rich environment. Unfortunately, the study[254] did not provide information about the abundance of the lanostanes relative to other steranes, so it is currently not possible to assess whether these compounds are low-concentrated biosynthetic intermediates or have a more unusual biogenic origin.

Another study[255] detected two lanostane isomers and four isomers of 4-desmethyl lanostane in limestones of the Lower Cambrian Sinyaya Formation on the eastern Siberian Platform. Based on peak heights taken from chromatograms in Figures 3 and 6 of ref[255], we estimated that lanostane constituted ~9%, and 4-desmethyl lanostanes ~31%, of total steranes in the richest samples (where 'total steranes' refers to the sum of C_{27} to C_{29} regular and diasteranes, lanostanes and 4-desmethyl lanostanes). The origin of the lanostanes remains unknown.

C_{30} to C_{32} lanostanes were also detected in migrated petroleum presumably sourced from the carbonate facies of the latest Ediacaran Bilara Fm, Marwar Supergroup from the shallow marine Bikaner-Nagaur Basin, north-west India[256]. Mass spectra were not provided for the two higher homologs, so their structures remain unclear. The authors speculate that the C_{31} and C_{32} compounds formed in the sediment through diagenetic methylation. However, given the apparent high relative signal intensities of the C_{31} and C_{32} homologs provided in GCxGC-TOFMS spectra,

this interpretation appears unlikely. No information was provided about the abundances of the lanostanes relative to other biomarkers.

Triterpenoids with diaromatic lanosteroid (DAL) fragmentation patterns have been reported from a Tertiary angiospermous lignite[249] and a sub-bituminous coal form the East Elbian coal basin[248], and these occurrences can be plausibly attributed to plants. A plant origin is also likely for 24-methyl-28-norlanosteroids and 24-methyl-19,28-bisnorlanosteroids extracted from a 2 million year old fossil plant of the family Lauraceae[257].

10.3. The origins of low and high protosteroid abundances in the Phanerozoic

Most Phanerozoic bitumens and oils contain low single-digit percentages of protosteroids relative to total steroids (Section 10.1). These background levels are presumable largely derived from biosynthetic intermediates found in crown-group eukaryotes plus an unknown contribution from bacteria. However, as mention in Section 8, some geological samples yield elevated abundances, such as Cretaceous lacustrine oil (GA224) with 7.1% aromatic protosteroids, a Jurassic coal-resinitic oil (GA19999642) containing 6.8%, Ordovician Tarim oil (HD_1 #5458) with 20.5%, bitumen from a Cambrian marine limestone from Australia (15B602) at 13.4% and from Siberia with 9% lanostane and ~31%, 4-desmethyl lanostanes (Section 10.2.3). In samples with terrestrial influence (lacustrine oils, coals), higher plant origins are likely (Section 8.1.2), but for pre-Devonian organic matter, the origin of the elevated abundances is unclear. (The predominantly Ordovician Tarim oil also contains plant biomarkers, pointing to a mixed age source[258], so the protosteroids in HD_1 #5458 could be plant derived).

Biodegradation is a likely cause for some elevated protosteroid occurrences. Chen et al. (1989)[136] observed lanostanes in biodegraded tar sands but failed to detect these compounds in the non-degraded Eocene source rocks. The authors suggested that lanostanes may be particularly biodegradation resistant, causing enrichment relative to regular steranes and hopanes in the degraded tars. This interpretation is supported by our observations on biodegraded migrated bitumens from the 1,640 Ma Barney Creek Formation. GC-MS analysis of solid bitumen recovered from a calcite vug within dolomite of drill core LV09 revealed severely biodegraded hydrocarbons consisting of a smooth unresolved complex mixture lacking *n*-alkanes, methyl alkanes and isoprenoids, and only containing traces of hopanes close to detection limits. Yet, the chromatogram also showed outstanding signals of cyclosteranes **I** (Extended Data Fig. 2), attesting to their extreme resistance to biodegradation. The resistance of cyclosterane to biodegradation is presumably caused by the inferred bicyclo/bridge structure and by the relatively large proportion of quaternary carbon atoms. For comparison, while cholestane possesses two quaternary carbon centres, lanostane has four and cyclosterane presumably five. Likewise, mono- and diaromatic

lanosteroids possess a quaternary carbon atom at C-4 that is absent in aromatic crown-steroids, possibly increasing their resistance to biodegradation. Although the effects of biodegradation on protosteroids requires systematic evaluation, it is plausible that this effect may cause protosteroid enrichment.

Although entirely speculative at present, the high number of quaternary carbon centres in functionalized protosterols may also increase their resistance to heterotrophic reworking in modern environments. Increased resistance of quaternary carbon centres to heterotrophic reworking may lead to a general enrichment of protosterols in sediments and soils relative to other sterols. Such an effect may generally contribute to an elevated average abundance of protosteroids in Phanerozoic bitumens and oils when compared to protosterol fractions in living eukaryotes (Section 8.1.1).

Supplementary References

70 Rooney, A. D. *et al.* Coupled Re-Os and U-Pb geochronology of the Tonian Chuar Group, Grand Canyon. *GSA Bulletin* **130**, 1085-1098, doi:10.1130/b31768.1 (2017).

71 Moczydłowska, M., Pease, V., Willman, S., Wickström, L. & AgiĆ, H. A Tonian age for the Visingsö Group in Sweden constrained by detrital zircon dating and biochronology: implications for evolutionary events. *Geological Magazine* **155**, 1175-1189, doi:10.1017/S0016756817000085 (2018).

72 Bronner, G., Roussel, J., Trompette, R. & Clauer, N. in *Dynamics of Plate Interiors* (1980).

73 Rooney, A. D., Selby, D., Houzay, J.-P. & Renne, P. R. Re–Os geochronology of a Mesoproterozoic sedimentary succession, Taoudeni basin, Mauritania: Implications for basin-wide correlations and Re–Os organic-rich sediments systematics. *Earth and Planetary Science Letters* **289**, 486-496, doi:10.1016/j.epsl.2009.11.039 (2010).

74 Gilleaudeau, G. J. & Kah, L. C. Heterogeneous redox conditions and a shallow chemocline in the Mesoproterozoic ocean: evidence from carbon–sulfur–iron relationships. *Precambrian Research* **257**, 94-108 (2015).

75 Benan, C. A. A. & Deynoux, M. Facies analysis and sequence stratigraphy of Neoproterozoic platform deposits in Adrar of Mauritania, Taoudéni Basin, West Africa. *Geologische Rundschau* **87**, 283-302 (1998).

76 Bertrand-Sarfati, J. & Moussine-Pouchkine, A. Evolution and environmental conditions of *Conophyton-jacutophyton* associations in the Atar Dolomite (upper Proterozoic, Mauritania). *Precambrian Research* **29**, 207-234 (1985).

77 Kah, L. C., Bartley, J. K. & Teal, D. A. Chemostratigraphy of the Late Mesoproterozoic Atar Group, Taoudeni Basin, Mauritania: Muted isotopic variability, facies correlation, and global isotopic trends. *Precambrian Research* **200–203**, 82-103, doi:10.1016/j.precamres.2012.01.011 (2012).

78 Beghin, J. *et al.* A palaeoecological model for the late Mesoproterozoic – early Neoproterozoic Atar/El Mreïti Group, Taoudeni Basin, Mauritania, northwestern Africa. *Precambrian Research* **299**, 1-14 (2017).
79 Trompette, R. *Le Précambrien supérieur et le Paléozoïque inférieur de l'Adrar de Mauritanie (bordure occidentale du bassin de Taoudeni, Afrique de l'Ouest), un exemple de sédimentation de craton. Étude stratigraphique et sédimentologique-TOMES 1 (Série 1) et 3 (Annexes)* PhD thesis, Université de Provence-Aix-Marseille I, (1973).
80 Cumming, V. M., Poulton, S. W., Rooney, A. D. & Selby, D. Anoxia in the terrestrial environment during the late Mesoproterozoic. *Geology* **41**, 583-586 (2013).
81 Elmore, R. D., Milavec, G. J., Imbus, S. W. & Engel, M. H. The Precambrian Nonesuch Formation of the North American mid-continent rift, sedimentology and organic geochemical aspects of lacustrine deposition. *Precambrian Research* **43**, 191-213 (1989).
82 Sheldon, N. D., Noffke, N. & Chafetz, H. Microbially induced sedimentary structures in the ca. 1100 Ma terrestrial midcontinent rift of North America. *Microbial Mats in Siliciclastic Depositional Systems Through Time: SEPM Special Publication* **11** (2012).
83 Strother, P. K. & Wellman, C. H. The Nonesuch Formation Lagerstätte: a rare window into freshwater life one billion years ago. *Journal of the Geological Society* **178** (2021).
84 Jones, S. *et al.* A marine origin for the late Mesoproterozoic Copper Harbor and Nonesuch Formations of the Midcontinent Rift of Laurentia. *Precambrian Research* **336**, 105510 (2020).
85 Fedorchuk, N. D. *et al.* Early non-marine life: evaluating the biogenicity of Mesoproterozoic fluvial-lacustrine stromatolites. *Precambrian Research* **275**, 105-118 (2016).
86 Slotznick, S. P., Swanson-Hysell, N. L. & Sperling, E. A. Oxygenated Mesoproterozoic lake revealed through magnetic mineralogy. *Proceedings of the National Academy of Sciences* **115**, 12938-12943 (2018).
87 Stüeken, E. E. *et al.* Geochemical fingerprints of seawater in the Late Mesoproterozoic Midcontinent Rift, North America: life at the marine-land divide. *Chemical Geology* **553**, 119812 (2020).
88 Strother, P. K. *et al.* A possible billion-year-old holozoan with differentiated multicellularity. *Current Biology* (2021).
89 Strother, P. K. & Wellman, C. H. Palaeoecology of a billion-year-old non-marine cyanobacterium from the Torridon Group and Nonesuch Formation. *Palaeontology* **59**, 89-108 (2016).
90 Jackson, M. J., Muir, M. D. & Plumb, K. A. Geology of the southern McArthur Basin, Northern Territory. *Bureau of Mineral Resources Bulletin, Geology and Geophysics* **220.** (1987).
91 Abbott, S. T. & Sweet, I. P. Tectonic control on third-order sequences in a siliciclastic ramp-style basin: An example from the Roper Superbasin (Mesoproterozoic), northern Australia. *Australian Journal of Earth Sciences* **47**, 637-657, doi:10.1046/j.1440-0952.2000.00795.x (2000).
92 Cox, G. M. *et al.* Basin redox and primary productivity within the Mesoproterozoic Roper Seaway. *Chemical Geology* **440**, 101-114, doi:10.1016/j.chemgeo.2016.06.025 (2016).

93 Munson, T. J. Sedimentary characterisation of the Wilton package, greater McArthur Basin, Northern Territory. *NTGS Record* **2016-003** (2016).
94 Munson, T. J. & Revie, D. Stratigraphic subdivision of the Velkerri Formation, Roper Group, McArthur Basin, Northern Territory. *NTGS Record* **2018-006** (2018).
95 Yang, B. *et al.* Spatial and temporal variation in detrital zircon age provenance of the hydrocarbon-bearing upper Roper Group, Beetaloo Sub-basin, Northern Territory, Australia. *Precambrian Research* **304**, 140-155, doi:10.1016/j.precamres.2017.10.025 (2018).
96 Yang, B. *et al.* Using Mesoproterozoic sedimentary geochemistry to reconstruct basin tectonic geography and link organic carbon productivity to nutrient flux from a Northern Australian large igneous Province. *Basin Research* **32**, 1734-1750, doi:doi.org/10.1111/bre.12450 (2020).
97 Kendall, B., Creaser, R. A., Gordon, G. W. & Anbar, A. D. Re-Os and Mo isotope systematics of black shales from the Middle Proterozoic Velkerri and Wollogorang Formations, McArthur Basin, northern Australia. *Geochimica et Cosmochimica Acta* **73**, 2534 (2009).
98 Javaux, E., Knoll, A. H. & Walter, M. R. Morphological and ecological complexity in early eukaryotic ecosystems. *Nature* **412**, 66-69 (2001).
99 Lyu, D. *et al.* Using cyclostratigraphic evidence to define the unconformity caused by the Mesoproterozoic Qinyu Uplift in the North China Craton. *Journal of Asian Earth Sciences* **206**, 104608 (2021).
100 Li, H. K. *et al.* The first precise age constraints on the Jixian System of the Meso-to Neoproterozoic Standard Section of China: SHRIMP zircon U-Pb dating of bentonites from the Wumishan and Tieling formations in the Jixian Section, North China Craton. *Acta Petrologica Sinica* **30**, 2999-3012 (2014).
101 Guo, W. *et al.* Zircon U-Pb dating and Hf isotopes of K-bentonites from the Tieling Formation in a new exposure of the Jixian Section, Tianjin, North China Craton. *Acta Petrologica Sinica* **35**, 2433-2454 (2019).
102 Zhang, S. *et al.* Paleoenvironmental proxies and what the Xiamaling Formation tells us about the mid-Proterozoic ocean. *Geobiology* **17**, 225-246 (2019).
103 Tang, D., Shi, X., Jiang, G., Zhou, X. & Shi, Q. Ferruginous seawater facilitates the transformation of glauconite to chamosite: An example from the Mesoproterozoic Xiamaling Formation of North China. *American Mineralogist: Journal of Earth and Planetary Materials* **102**, 2317-2332 (2017).
104 Wang, X., Ye, Y., Wang, H. & Zhang, S. Decoupled Cr, Mo, and U records of the Hongshuizhuang Formation, North China: Constraints on the Mesoproterozoic ocean redox. *Marine and Petroleum Geology* **132**, 105243 (2021).
105 Tosti, F. & Riding, R. Fine-grained agglutinated elongate columnar stromatolites: Tieling Formation, ca 1420 Ma, North China. *Sedimentology* **64**, 871-902 (2017).
106 Luo, G., Hallmann, C., Xie, S., Ruan, X. & Summons, R. E. Comparative microbial diversity and redox environments of black shale and stromatolite facies in the Mesoproterozoic Xiamaling Formation. *Geochimica et Cosmochimica Acta* **151**, 150-167, doi:10.1016/j.gca.2014.12.022 (2015).

107 Diamond, C. W., Planavsky, N. J., Wang, C. & Lyons, T. W. What the ~1.4 Ga Xiamaling Formation can and cannot tell us about the mid-Proterozoic ocean. *Geobiology* **16**, 219-236, doi:doi:10.1111/gbi.12282 (2018).
108 Wang, X. *et al.* The aerobic diagenesis of Mesoproterozoic organic matter. *Scientific Reports* **8:133248**, doi:DOI:10.1038/s41598-018-31378-6 (2018).
109 Planavsky, N. J. *et al.* No evidence for high atmospheric oxygen levels 1,400 million years ago. *Proceedings of the National Academy of Sciences* **113**, E2550-E2551 (2016).
110 Miao, L., Moczydłowska, M. & Zhu, M. A diverse organic-walled microfossil assemblage from the Mesoproterozoic Xiamaling Formation, North China. *Precambrian Research* **360**, 106235 (2021).
111 Kunzmann, M., Schmid, S., Blaikie, T. N. & Halverson, G. P. Facies analysis, sequence stratigraphy, and carbon isotope chemostratigraphy of a classic Zn-Pb host succession: The Proterozoic middle McArthur Group, McArthur Basin, Australia. *Ore Geology Reviews* **106**, 150-175, doi:10.1016/j.oregeorev.2019.01.011 (2019).
112 Rawlings, D. J. *et al.* The 2002 Southern McArthur Basin Seismic Reflection Survey. *Geoscience Australia Record 2004/17*, pp. 87 (2004).
113 Blaikie, T. N. & Kunzmann, M. Geophysical interpretation and tectonic synthesis of the Proterozoic southern McArthur Basin, northern Australia. *Precambrian Research* **343**, 105728, doi:doi.org/10.1016/j.precamres.2020.105728 (2020).
114 Ahmad, Dunster & Munson. Geology and mineral resources of the Northern Territory. *Northern Territory Geological Survey Special Publication* **5**, 15:11–15:72 (2013).
115 Page, R. W. & Sweet, I. P. Geochronology of basin phases in the western Mt Isa Inlier, and correlation with the McArthur Basin. *Australian Journal of Earth Sciences* **45**, 219-232 (1998).
116 Brown, M., Claxton, C. & Plumb, K. The Proterozoic Barney Creek Formation and some associated units of the McArthur Group. *Northern Territory, Australia: Bureau of Mineral Resources Record* **145**, 59 (1969).
117 Bull, S. W. Sedimentology of the Palaeoproterozoic Barney Creek Formation in DDH BMR McArthur 2, southern McArthur Basin, Northern Territory. *Australian Journal of Earth Sciences* **45**, 21-31 (1998).
118 Winefield, P. R. *Sedimentology and Diagenesis of late Palaeoproterozoic Carbonates, southern McArthur Basin, Northern Australia*, University of Tasmania, (1999).
119 French, K. L., Birdwell, J. E. & Vanden Berg, M. D. Biomarker similarities between the saline lacustrine Eocene Green River and the Paleoproterozoic Barney Creek Formations. *Geochimica et Cosmochimica Acta* **274**, 228-245, doi:10.1016/j.gca.2020.01.053 (2020).
120 Kah, L. C., Lyons, T. W. & Frank, T. D. Low marine sulphate and protracted oxygenation of the Proterozoic biosphere. *Nature* **431**, 834-838 (2004).
121 Planavsky, N. J. *et al.* Widespread iron-rich conditions in the mid-Proterozoic ocean. *Nature* **477**, 448 (2011).
122 Nettersheim, B. J. *Reconstructing earth's alien ancient ecology – a multiproxy study of the 1.64 billion-year-old Barney Creek Formation, Northern Australia* PhD thesis, The Australian National University, (2017).
123 Oehler, J. H. Microflora of the H.Y.C. Pyritic Shale Member of the Barney Creek Formation (McArthur Group), middle Proterozoic of northern Australia. *Alcheringa* **1**, 315-349 (1977).

124 Vinnichenko, G., Jarrett, A. J. M., Hope, J. M. & Brocks, J. J. Discovery of the oldest known biomarkers provides evidence for phototrophic bacteria in the 1.73 Ga Wollogorang Formation, Australia. *Geobiology* **18**, doi:10.1111/gbi.12390 (2020).
125 Summons, R. E., Powell, T. G. & Boreham, C. J. Petroleum geology and geochemistry of the Middle Proterozoic McArthur Basin, northern Australia: III. Composition of extractable hydrocarbons. *Geochimica et Cosmochimica Acta* **52**, 1747-1763 (1988).
126 Jarrett, A. *et al.* in *AGES 2019 Proceedings, NT Geological Survey* 92-105 (2019).
127 Crick, I. H., Boreham, C. J., Cook, A. C. & Powell, T. G. Petroleum geology and geochemistry of Middle Proterozoic McArthur Basin, northern Australia II: assessment of source rock potential. *AAPG Bulletin* **72**, 1495-1514 (1988).
128 Abballe, P., Hall, T. & Kennedy, M. Geologic controls of black shale deposition in the Palaeo-Proterozoic of the McArthur Basin. *Northern Terriitory Geological Survey - Destructive Analysis Report*, 13 pages (2014).
129 Alexander, R., Berwick, L. & Pierce, K. Single carbon surface reactions of 1-octadecene and 2,3,6-trimethylphenol on activated carbon: Implications for methane formation in sediments. *Organic Geochemistry* **42**, 540 (2011).
130 Berwick, L., Alexander, R. & Pierce, K. Formation and reactions of alkyl adamantanes in sediments: Carbon surface reactions. *Organic Geochemistry* **42**, 752 (2011).
131 West, N., Alexander, R. & Kagi, R. I. The use of silicalite for rapid isolation of branched and cyclic alkane fractions of petroleum. *Org. Geochem.* **15**, 499-501 (1990).
132 Hauke, V. *et al.* Isoarborinol through geological times: Evidence for its presence in the Permian and Triassic. *Organic Geochemistry* **23**, 91-93, doi:doi.org/10.1016/0146-6380(95)00002-V (1995).
133 Piatak, D. M., Bhat, H. B. & Caspi, E. Oxidation of steroidal ketones. VII. Cleavage of steroidal conjugated ketones with ruthenium tetroxide. *The Journal of Organic Chemistry* **34** 112-116. (1969).
134 Peters, K. E. & Moldowan, J. M. *The Biomarker Guide*. (Prentice Hall, 1993).
135 Nytoft, H. P., Kildahl-Andersen, G. & Rise, F. Unusual hexacyclic oleananes in Late Cretaceous/Tertiary terrigenous oils: NMR characterisation of the major hexacyclic oleanane in Niger Delta oil. *Organic Geochemistry* **101**, 196-206, doi:10.1016/j.orggeochem.2016.08.016 (2016).
136 Chen, J. H., Philp, R. P., Fu, F. M. & Sheng, G. Y. The occurrence and identification of C_{30}-C_{32} lanostanes: a novel series of tetracyclic triterpenoid hydrocarbons. *Geochimica et Cosmochimica Acta* **53**, 2775-2779 (1989).
137 Lemoine, S., Adam, P., Albrecht, P. & Connan, J. Novel series of diaromatic 14-methyl steroids occurring in petroleum. *Tetrahedron Letters* **37** 2837-2840 (1996).
138 Jacob, J. *et al.* Contrasted distributions of triterpene derivatives in the sediments of Lake Caçó reflect paleoenvironmental changes during the last 20,000 yrs in NE Brazil. *Organic Geochemistry* **38**, 180-197 (2007).
139 Hauke, V. *et al.* Novel triterpene derived hydrocarbons of arborane/fernane series in sediments. Part I. *Tetrahedron* **48** 3915-3924 (1992).
140 Yang, C. *et al.* Aromatic steroids in crude oils and petroleum products and their applications in forensic oil spill identification. *Environmental Forensics* **14**, 278-293 (2013).

141 Adam, P. *Nouvelles structures organo soufrées d intérêt géochimique: Aspects moléculaires et macromoléculaires* PhD thesis, Université Louis Pasteur, (1991).
142 Hussler, G. *et al.* Benzohopanes, a novel family of hexacyclic geomarkers in sediments and petroleums. Part I. *Tetrahedron* **48**, 3915-3924 (1984).
143 Sinninghe Damsté, J. S., Schouten, S., van Vliet, N. H. & Geenevasen, J. A. J. A sedimentary fluorene derivative of bacteriohopanepolyols. *Tetrahedron Letters* **39**, 3021-3024 (1998).
144 Carrillo Hernández, T., Schaeffer, P. & Albrecht, P. Acenaphthenic hopanoids, a novel series of aromatised triterpenoids occurring in crude oil *Chemical Communications* 1976-1977 (2001).
145 Hussler, G., Connan, J. & Albrecht, P. Novel families of tetra- and hexacyclic aromatic hopanoids predominant in carbonate rocks and crude oils. *Organic Geochemistry*, 39-49 (1984).
146 Killops, S. D. Novel aromatic hydrocarbons of probable bacterial origin in a Jurassic lacustrine sequence. *Organic Geochemistry* **17**, 25-36 (1991).
147 Forestier, E. *et al.* Distinct triterpene synthases in the laticifers of Euphorbia lathyris. *Scientific Reports (Nature Publisher Group)* **9**, doi:dx.doi.org/10.1038/s41598-019-40905-y (2019).
148 Pale-Grosdemange, C., Feil, C., Rohmer, M. & Poralla, K. Occurrence of Cationic Intermediates and Deficient Control during the Enzymatic Cyclization of Squalene to Hopanoids. *Angewandte Chemie International Edition* **37**, 2237-2240, doi:doi.org/10.1002/(SICI)1521-3773(19980904)37:16<2237::AID-ANIE2237>3.0.CO;2-9 (1998).
149 Douka, E., Koukkou, A.-I., Drainas, C., Grosdemange-Billiard, C. & Rohmer, M. Structural diversity of the triterpenic hydrocarbons from the bacterium Zymomonas mobilis: the signature of defective squalene cyclization by the squalene/hopene cyclase. *FEMS Microbiology Letters* **199**, 247-251, doi:10.1111/j.1574-6968.2001.tb10682.x (2001).
150 Hess Jr., B. A. & Smentek, L. The Concerted Nature of the Cyclization of Squalene Oxide to the Protosterol Cation. *Angewandte Chemie International Edition* **52**, 11029-11033, doi:doi.org/10.1002/anie.201302886 (2013).
151 Radke, M., Welte, D. H. & Willsch, H. Maturity parameters based on aromatic hydrocarbons: Influence of the organic matter type. *Org. Geochem.* **10**, 51-63 (1986).
152 Szczerba, M. & Rospondek, M. J. Controls on distributions of methylphenanthrenes in sedimentary rock extracts: Critical evaluation of existing geochemical data from molecular modelling. *Organic Geochemistry* **41**, 1297-1311, doi:10.1016/j.orggeochem.2010.09.009 (2010).
153 Bobrovskiy, I. *et al.* Algal origin of sponge sterane biomarkers negates the oldest evidence for animals in the rock record. *Nature Ecology & Evolution* **5**, 165-168, doi:10.1038/s41559-020-01334-7 (2021).
154 Brown, M., McShea, H., Olagunju, B., Giner, J. & Welander, P. Testing the Sponge Biomarker Hypothesis Through Identification of 24-Isopropenylcholesterol Biosynthesis Enzymes. **2021**, 1-2, doi:doi.org/10.3997/2214-4609.202134237 (2021).

155 Brown, M. O., Olagunju, B. O., Giner, J.-L. & Welander, P. V. Sterol methyltransferases in uncultured bacteria complicate eukaryotic biomarker interpretations. *Nature Communications* **14**, 1859, doi:doi.org/10.1038/s41467-023-37552-3 (2023).
156 Summons, R. E., Bradley, A. S., Jahnke, L. L. & Waldbauer, J. R. Steroids, triterpenoids and molecular oxygen. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 951 (2006).
157 Takishita, K. *et al.* Microbial Eukaryotes that Lack Sterols. *Journal of Eukaryotic Microbiology* **64**, 897-900, doi:10.1111/jeu.12426 (2017).
158 Tomazic, M. L., Poklepovich, T. J., Nudel, C. B. & Nusblat, A. D. Incomplete sterols and hopanoids pathways in ciliates: Gene loss and acquisition during evolution as a source of biosynthetic genes. *Molecular Phylogenetics and Evolution* **74**, 122-134, doi:10.1016/j.ympev.2014.01.026 (2014).
159 Takishita, K. *et al.* Lateral transfer of tetrahymanol-synthesizing genes has allowed multiple diverse eukaryote lineages to independently adapt to environments without oxygen. *Biology Direct* **7**, 5 (2012).
160 Waldbauer, J. R., Newman, D. K. & Summons, R. E. Microaerobic steroid biosynthesis and the molecular fossil record of Archean life. *Proceedings of the National Academy of Sciences* **108**, 13409-13414, doi:10.1073/pnas.1104160108 (2011).
161 Bobrovskiy, I., Hope, J. M., Krasnova, A., Ivantsov, A. & Brocks, J. J. Molecular fossils from organically preserved Ediacara biota reveal cyanobacterial origin for Beltanelliformis. *Nature Ecology & Evolution* **2**, 437–440, doi:10.1038/s41559-017-0438-6 (2018).
162 Goad, L. J. & Goodwin, T. W. Studies on phytosterol biosynthesis: The sterols of *Larix decidua* leaves. *European Journal of Biochemistry* **1**, 357 -362 (1967).
163 Weete, J. D., Fuller, M. S., Huang, M. Q. & Gandhi, S. Fatty acids and sterols of selected hyphochytriomycetes and chytridiomycetes. *Experimental Mycology* **13**, 183-195, doi:10.1016/0147-5975(89)90023-6 (1989).
164 Rampen, S. W., Abbas, B. A., Schouten, S. & Sinninghe Damste, J. S. A comprehensive study of sterols in marine diatoms (Bacillariophyta): Implications for their use as tracers for diatom productivity. *Limnology and Oceanography* **55**, 91-105, doi:10.4319/lo.2010.55.1.0091 (2010).
165 Schaller, H. The role of sterols in plant growth and development. *Progress in Lipid Research* **42**, 163-175, doi:10.1016/S0163-7827(02)00047-4 (2003).
166 Itoh, T., Tamura, T. & Matsumoto, T. Sterol composition of 19 vegetable oils. *Journal of the American Oil Chemists Society* **50**, 122-125 (1973).
167 Itoh, T., Tamura, T. & Matsumoto, T. Methylsterol compositions of 19 vegetable oils. *Journal of the American Oil Chemists Society* **50**, 300-303 (1973).
168 Peers, K. E. The non-glyceride saponifiables of shea butter. *Journal of the Science of Food and Agriculture* **28**, 1000-1009, doi:10.1002/jsfa.2740281109 (1977).
169 Khan, M. T. H., Khan, S. B. & Ather, A. Tyrosinase inhibitory cycloartane type triterpenoids from the methanol extract of the whole plant of Amberboa ramosa Jafri and their structure–activity relationship. *Bioorganic & Medicinal Chemistry* **14**, 938-943, doi:10.1016/j.bmc.2005.09.010 (2006).
170 Bergman, J., Lindgren, B. O. & Svahn, C. M. Triterpenes and 4-a-methylsterols in birch wood. *Acta chem. scand* **19**, 10 (1965).

171 Steglich, W., Klaar, M., Zechlin, L. & Hecht, H. J. Abietospiran, das Triterpen der Weißtannenrinde (Abies alba). *Angewandte Chemie* **91**, 751-751 (1979).
172 Allen, F., Kutney, J. P., Trotter, J. & Westcott, N. D. The structures and absolute stereochemistry of cyclograndisolide and epicyclograndisolide, novel triterpene lactones from Abies grandis. *Tetrahedron Letters* **12**, 283-286 (1971).
173 Yano, K., Akihisa, T., Tamura, T. & Matsumoto, T. Four 4α-methylsterols and triterpene alcohols from Neolitsea aciculata. *Phytochemistry* **31**, 2093-2098, doi:10.1016/0031-9422(92)80369-P (1992).
174 Biesboer, D. D., D'Amour, P., Wilson, S. R. & Mahlberg, P. Sterols and triterpenols in latex and cultured tissues of Euphorbia pulcherrima. *Phytochemistry* **21**, 1115-1118, doi:10.1016/S0031-9422(00)82427-9 (1982).
175 Fox, M. G. & French, J. C. Systematic Occurrence of Sterols in Latex of Araceae: Subfamily Colocasioideae. *American Journal of Botany* **75**, 132-137, doi:10.2307/2443911 (1988).
176 Kemp, R. J., Hammam, A. S. A., Goad, L. J. & Goodwin, T. W. Studies on phytosterol biosynthesis: Observations on the esterified sterols of higher plants. *Phytochemistry* **7**, 447-450, doi:10.1016/S0031-9422(00)90885-9 (1968).
177 Goad, L. & Goodwin, T. The biosynthesis of sterols in higher plants. *Biochemical Journal* **99**, 735-746 (1966).
178 Manners, G. D. & Davis, D. G. Epicuticular wax constituents of North American and European Euphorbia esula biotypes. *Phytochemistry* **23**, 1059-1062 (1984).
179 Prasad, R. B. N. & Gülz, P.-G. Epicuticular Waxes from Leaves of Maple (Acer pseudoplatanus L.). *Zeitschrift für Naturforschung C* **45**, 599-601, doi:10.1515/znc-1990-0606 (1990).
180 Bauer, S., Schulte, E. & Thier, H.-P. Composition of the surface wax from tomatoes. *European Food Research and Technology* **219**, 223-228 (2004).
181 Asakawa, Y. in *Progress in the Chemistry of Organic Natural Products* (ed Y. Asakawa) 1-562 (Springer Vienna, 1995).
182 Karunen, P. & Ekman, R. Senescence-related Changes in the Composition of Free and Esterified Sterols and Alcohols in Sphagnum fuscum. *Zeitschrift für Pflanzenphysiologie* **104**, 319-330, doi:10.1016/S0044-328X(81)80071-2 (1981).
183 Marsili, A., Morelli, I. & Iori, A. M. 21-Hopene and some other constituents of Pseudoscleropodium purum. *Phytochemistry* **10**, 432-433, doi:10.1016/S0031-9422(00)94064-0 (1971).
184 Raederstorff, D. & Rohmer, M. Sterol biosynthesis via cycloartenol and other biochemical features related to photosynthetic phyla in the amoebae *Naegleria lovaniensis* and *Naegleria gruberi*. *European Journal of Biochemistry* **164**, 427-434, doi:10.1111/j.1432-1033.1987.tb11075.x (1987).
185 Luddy, F. E., Turner, A. & Scanlan, J. T. Spectrophotometric Determination of Cholesterol and Triterpene Alcohols in Wool Wax. *Analytical Chemistry* **25**, 1497-1499, doi:10.1021/ac60082a023 (1953).
186 Hughes, A. L., Lee, C.-Y. S., Bien, C. M. & Espenshade, P. J. 4-Methyl Sterols Regulate Fission Yeast SREBP-Scap under Low Oxygen and Cell Stress*. *Journal of Biological Chemistry* **282**, 24388-24396, doi:10.1074/jbc.M701326200 (2007).

187 Pearson, A., Budin, M. & Brocks, J. J. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proc. Natl. Acad. Sci.* **100**, 15352–15357 (2003).
188 Villanueva, L., Rijpstra, W. I. C., Schouten, S. & Damsté, J. S. S. Genetic biomarkers of the sterol-biosynthetic pathway in microalgae. *Environmental Microbiology Reports* **6**, 35-44, doi:10.1111/1758-2229.12106 (2014).
189 Bird, C. W. *et al.* Steroids and squalene in *Methylococcus capsulatus* grown on methane. *Nature* **230**, 473-474 (1971).
190 Bode, H. B. *et al.* Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Molecular Microbiology* **47**, 471–481 (2003).
191 Gruen, D. S., Wolfe, J. M. & Fournier, G. P. Paleozoic diversification of terrestrial chitin-degrading bacterial lineages. *BMC Evolutionary Biology* **19**, 34, doi:10.1186/s12862-019-1357-8 (2019).
192 Shimkets, L. J. Social and developmental biology of the myxobacteria. *Microbiol Rev* **54**, 473-501 (1990).
193 Albataineh, H. & Stevens, D. C. Marine Myxobacteria: A Few Good Halophiles. *Mar Drugs* **16**, 209, doi:10.3390/md16060209 (2018).
194 Brinkhoff, T. *et al.* Biogeography and phylogenetic diversity of a cluster of exclusively marine myxobacteria. *The ISME Journal* **6**, 1260-1272, doi:10.1038/ismej.2011.190 (2012).
195 Thomas, S. H. *et al.* The Mosaic Genome of Anaeromyxobacter dehalogenans Strain 2CP-C Suggests an Aerobic Common Ancestor to the Delta-Proteobacteria. *PLoS One* **3**, doi:10.1371/journal.pone.0002103 (2008).
196 Williams, K. P. *et al.* Phylogeny of Gammaproteobacteria. *Journal of Bacteriology* **192**, 2305-2314, doi:10.1128/jb.01480-09 (2010).
197 Hugoson, E., Ammunét, T. & Guy, L. Host-adaptation in Legionellales is 2.4 Ga, coincident with eukaryogenesis. *bioRxiv*, 852004, doi:10.1101/852004 (2020).
198 Canfield, D. E. & Teske, A. Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulfur-isotope studies. *Nature* **382**, 127-132 (1996).
199 Brocks, J. J. & Schaeffer, P. Okenane, a biomarker for purple sulfur bacteria (Chromatiaceae), and other new carotenoid derivatives from the 1,640 Ma Barney Creek Formation. *Geochimica et Cosmochimica Acta* **72**, 1396-1414 (2008).
200 Knief, C. Diversity and Habitat Preferences of Cultivated and Uncultivated Aerobic Methanotrophic Bacteria Evaluated Based on pmoA as Molecular Marker. *Front Microbiol* **6**, 1346-1346, doi:10.3389/fmicb.2015.01346 (2015).
201 Rivas-Marin, E. *et al.* Essentiality of sterol synthesis genes in the planctomycete bacterium Gemmata obscuriglobus. *Nature Communications* **10**, 2916, doi:10.1038/s41467-019-10983-7 (2019).
202 Bouvier, P., Rohmer, M., Benveniste, P. & Ourisson, G. $\Delta^{8(14)}$-Steroids in the bacterium *Methylococcus capsulatus*. *Biochem. J.* **159**, 267-271 (1976).
203 Welander, P. & Lee, A. Cholesterol Production by the Bacterium *Enhygromyxa salina*. *EarthDoc* **2021**, 1-2, doi:doi.org/10.3997/2214-4609.202134239 (2021).

204 Banta, A. B., Wei, J. H., Gill, C. C. C., Giner, J.-L. & Welander, P. V. Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. *Proceedings of the National Academy of Sciences* **114**, 245-250, doi:10.1073/pnas.1617231114 (2017).
205 Lee, A. K. *et al.* C-4 sterol demethylation enzymes distinguish bacterial and eukaryotic sterol synthesis. *Proc Natl Acad Sci U S A* **115**, 5884-5889, doi:10.1073/pnas.1802930115 (2018).
206 Pollier, J. *et al.* A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis. *Nature Microbiology* **4**, 226-233, doi:10.1038/s41564-018-0305-5 (2019).
207 Piironen, V., Toivo, J., Puupponen-Pimiä, R. & Lampi, A.-M. Plant sterols in vegetables, fruits and berries. *Journal of the Science of Food and Agriculture* **83**, 330-337, doi:10.1002/jsfa.1316 (2003).
208 Zhang, X., Lin, K. & Li, Y. Highlights to phytosterols accumulation and equilibrium in plants: Biosynthetic pathway and feedback regulation. *Plant Physiology and Biochemistry* **155**, 637-649, doi:10.1016/j.plaphy.2020.08.021 (2020).
209 Lopes, G. *et al.* Sterol profiles in 18 macroalgae of the Portuguese coast. *Journal of Phycology* **47**, 1210-1218, doi:10.1111/j.1529-8817.2011.01028.x (2011).
210 Ahmed, F., Zhou, W. & Schenk, P. M. Pavlova lutheri is a high-level producer of phytosterols. *Algal Research* **10**, 210-217, doi:10.1016/j.algal.2015.05.013 (2015).
211 Raederstorff, D. & Rohmer, M. Sterol biosynthesis de nova via cycloartenol by the soil amoeba Acanthamoeba polyphaga. *Biochemical Journal* **231**, 609-615, doi:10.1042/bj2310609 (1985).
212 Schubert, K., Rose, G. & Hörhold, C. Cholesterin in Streptomyces olivaceus. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **137**, 168-171, doi:10.1016/0005-2760(67)90020-3 (1967).
213 Schubert, K., Rose, G., Wachtel, H., Hörhold, C. & Ikekawa, N. Zum Vorkommen von Sterinen in Bakterien. *European Journal of Biochemistry* **5**, 246-251, doi:10.1111/j.1432-1033.1968.tb00364.x (1968).
214 de Souza, N. J. & Nes, W. R. Sterols: isolation from a blue-green alga. *Science* **162**, 363-364 (1968).
215 Kohl, W., Gloe, A. & Reichenbach, H. Steroids from the myxobacterium *Nannocystis exedens*. *J. Gen. Microbiol.* **129**, 1629-1635 (1983).
216 Patt, T. E. & Hanson, R. S. Intracytoplasmic membrane, phospholipid, and sterol content of Methylobacterium organophilum cells grown under different conditions. *Journal of Bacteriology* **134**, 636-644 (1978).
217 Bradley, A. S., Hayes, J. M. & Summons, R. E. Extraordinary 13C enrichment of diether lipids at the Lost City Hydrothermal Field indicates a carbon-limited ecosystem. *Geochimica et Cosmochimica Acta* **73**, 102-118, doi:10.1016/j.gca.2008.10.005 (2009).
218 López-García, P., Vereshchaka, A. & Moreira, D. Eukaryotic diversity associated with carbonates and fluid–seawater interface in Lost City hydrothermal field. *Environmental Microbiology* **9**, 546-554, doi:10.1111/j.1462-2920.2006.01158.x (2007).
219 Krüger, M. *et al.* A novel, multi-layered methanotrophic microbial mat system growing on the sediment of the Black Sea. *Environmental Microbiology* **10**, 1934-1947, doi:10.1111/j.1462-2920.2008.01607.x (2008).

220 Boëchat, I. G., Krüger, A. & Adrian, R. Sterol Composition of Freshwater Algivorous Ciliates Does Not Resemble Dietary Composition. *Microbial Ecology* **53**, 74-81, doi:10.1007/s00248-006-9014-3 (2007).
221 Boëchat, I. G. *Biochemical Composition of Protists: Dependency on Diet and Trophic Mode and Consequences for their Nutritional Quality* PhD thesis, Humboldt-Universität zu Berlin, (2005).
222 Elvert, M. & Niemann, H. Occurrence of unusual steroids and hopanoids derived from aerobic methanotrophs at an active marine mud volcano. *Organic Geochemistry* **39**, 167-177, doi:10.1016/j.orggeochem.2007.11.006 (2008).
223 Edmunds, K. L. H. & Eglington, G. in *Microbial Mats: Stromatolites* (eds Y. Cohen, R.W. Castenholz, & H.O. Halvorson) 343-389 (Alan R. Liss, 1984).
224 Cardoso, J. N. *et al.* A biogeochemical study of the Abu Dhabi algal mats; a simplified ecosystem. *Chemical Geology* **23**, 273-291 (1978).
225 Jahnke, L. L. & Des Marais, D. J. in *Astrobiology Science Conference 2017.* LPI Contrib. No. 1965.
226 McCaffrey, M. A., Farrington, J. W. & Repeta, D. J. Geochemical implications of the lipid composition of Thioploca spp. from the Peru upwelling region—15°S. *Organic Geochemistry* **14**, 61-68, doi:10.1016/0146-6380(89)90019-3 (1989).
227 Pearson, A., Seewald, J. S. & Eglinton, T. I. Bacterial incorporation of relict carbon in the hydrothermal environment of Guaymas Basin. *Geochimica et Cosmochimica Acta* **69**, 5477-5486, doi:10.1016/j.gca.2005.07.007 (2005).
228 Arning, E. T. *et al.* Lipid Biomarker Patterns of Phosphogenic Sediments from Upwelling Regions. *Geomicrobiology Journal* **25**, 69-82, doi:10.1080/01490450801934854 (2008).
229 Słowakiewicz, M. *et al.* Biogeochemistry of intertidal microbial mats from Qatar: New insights from organic matter characterisation. *Organic Geochemistry* **102**, 14-29, doi:10.1016/j.orggeochem.2016.09.006 (2016).
230 Peckmann, J. *et al.* A Microbial Mat of a Large Sulfur Bacterium Preserved in a Miocene Methane-Seep Limestone. *Geomicrobiology Journal* **21**, 247-255, doi:10.1080/01490450490438757 (2004).
231 Kellermann, M. Y. *et al.* Symbiont–host relationships in chemosynthetic mussels: A comprehensive lipid biomarker study. *Organic Geochemistry* **43**, 112-124, doi:10.1016/j.orggeochem.2011.10.005 (2012).
232 Conway, N. & McDowell Capuzzo, J. Incorporation and utilization of bacterial lipids in theSolemya velum symbiosis. *Marine Biology* **108**, 277-291, doi:10.1007/BF01344343 (1991).
233 Zhang, C. L. *et al.* Lipid Biomarkers and Carbon Isotope Signatures of a Microbial (Beggiatoa) Mat Associated with Gas Hydrates in the Gulf of Mexico. *Applied and Environmental Microbiology* **71**, 2106-2112, doi:10.1128/aem.71.4.2106-2112.2005 (2005).
234 Jacq, E. *et al.* Microscopic examination and fatty acid characterization of filamentous bacteria colonizing substrata around subtidal hydrothermal vents. *Archives of Microbiology* **152**, 64-71, doi:10.1007/BF00447013 (1989).
235 Guezennec, J., Ortega-Morales, O., Raguenes, G. & Geesey, G. Bacterial colonization of artificial substrate in the vicinity of deep-sea hydrothermal vents. *FEMS Microbiology Ecology* **26**, 89-99, doi:10.1111/j.1574-6941.1998.tb00495.x (1998).

236 Guezennec, J. & Fiala-Medioni, A. Bacterial abundance and diversity in the Barbados Trench determined by phospholipid analysis. *FEMS Microbiology Ecology* **19**, 83-93, doi:10.1111/j.1574-6941.1996.tb00201.x (1996).
237 Elvert, M., Boetius, A., Knittel, K. & Jørgensen, B. B. Characterization of Specific Membrane Fatty Acids as Chemotaxonomic Markers for Sulfate-Reducing Bacteria Involved in Anaerobic Oxidation of Methane. *Geomicrobiology Journal* **20**, 403-419, doi:10.1080/01490450303894 (2003).
238 Rossel, P. E. *et al.* Intact polar lipids of anaerobic methanotrophic archaea and associated bacteria. *Organic Geochemistry* **39**, 992-999, doi:10.1016/j.orggeochem.2008.02.021 (2008).
239 Knittel, K. *et al.* Thiomicrospira arctica sp. nov. and Thiomicrospira psychrophila sp. nov., psychrophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacteria isolated from marine Arctic sediments. *International journal of systematic and evolutionary microbiology* **55 Pt 2**, 781-786 (2005).
240 Jean, M. R. N. *et al.* Two New Beggiatoa Species Inhabiting Marine Mangrove Sediments in the Caribbean. *PLoS One* **10**, doi:10.1371/journal.pone.0117832 (2015).
241 Li, J. *et al.* Abundance and distribution of fatty acids within the walls of an active deep-sea sulfide chimney. *Journal of Sea Research* **65**, 333-339, doi:10.1016/j.seares.2011.01.005 (2011).
242 Li, Y.-L., Peacock, A. D., White, D. C., Geyer, R. & Zhang, C. L. Spatial patterns of bacterial signature biomarkers in marine sediments of the Gulf of Mexico. *Chemical Geology* **238**, 168-179, doi:10.1016/j.chemgeo.2006.11.007 (2007).
243 Kojima, H. *et al.* Ecophysiology of Thioploca ingrica as revealed by the complete genome sequence supplemented with proteomic evidence. *The ISME Journal* **9**, 1166-1176, doi:10.1038/ismej.2014.209 (2015).
244 Wakeham, S. G. & Beier, J. A. Fatty acid and sterol biomarkers as indicators of particulate matter source and alteration processes in the Black Sea. *Deep Sea Research Part A. Oceanographic Research Papers* **38**, S943-S968, doi:10.1016/S0198-0149(10)80018-4 (1991).
245 Blumenberg, M., Seifert, R. & Michaelis, W. Aerobic methanotrophy in the oxic–anoxic transition zone of the Black Sea water column. *Organic Geochemistry* **38**, 84-91, doi:10.1016/j.orggeochem.2006.08.011 (2007).
246 Jacob, J. *et al.* Contrasted distributions of triterpene derivatives in the sediments of Lake Caçó reflect paleoenvironmental changes during the last 20,000 yrs in NE Brazil. *Organic Geochemistry* **38**, 180-197 (2007).
247 Jacob, J. *et al.* Pentacyclic triterpene methyl ethers in recent lacustrine sediments (Lagoa do Caçó, Brazil). *Organic Geochemistry* **36**, 449-461, doi:10.1016/j.orggeochem.2004.09.005 (2005).
248 Hazai, I., Alexander, G., Szekely, T., Essiger, B. & Radek, D. Investigation of hydrocarbon constituents of a young sub-bituminous coal by gas chromatography-mass spectrometry. *Journal of Chromatography A* **367**, 117-133 (1986).
249 Stout, S. A. Aliphatic and aromatic triterpenoid hydrocarbons in a Tertiary angiospermous lignite. *Organic Geochemistry* **18**, 51-66 (1992).

250 Russell, G. B., Connor, H. E. & Purdie, A. W. Triterpene methyl ethers of Chionochloa (Gramineae). *Phytochemistry* **15**, 1933-1935, doi:10.1016/S0031-9422(00)88849-4 (1976).
251 Chen, J. & Summons, R. E. Complex patterns of steroidal biomarkers in Tertiary lacustrine sediments of the Biyang Basin, China. *Org. Geochem.* **32**, 115-126 (2001).
252 Bao, J. & Li, M. Unprecedented occurrence of novel C26–C28 21-norcholestanes and related triaromatic series in evaporitic lacustrine sediments. *Organic Geochemistry* **32**, 1031-1036, doi:10.1016/S0146-6380(01)00060-2 (2001).
253 Birgel, D. & Peckmann, J. Aerobic methanotrophy at ancient marine methane seeps: A synthesis. *Organic Geochemistry* **39**, 1659-1667, doi:10.1016/j.orggeochem.2008.01.023 (2008).
254 Peng, P. *et al.* Lanostane sulfides in an immature crude oil. *Organic Geochemistry* **28**, 125-134, doi:10.1016/S0146-6380(97)00112-5 (1998).
255 Parfenova, T. M. Hydrocarbons of the lanostane homologous series in the Phanerozoic organic matter and their probable biologic sources. *Russian Geology and Geophysics* **52**, 773-780, doi:10.1016/j.rgg.2011.07.003 (2011).
256 Bhattacharya, S., Dutta, S. & Kumar, S. Identification of lanostanes, A-ring methylated steranes and secosteranes in late Neoproterozoic crude oils by GC×GC-TOFMS: New insights into molecular taphonomy of steroids. *Geobios*, doi:10.1016/j.geobios.2021.04.003 (2021).
257 Murae, T., Naora, M., Hosokawa, K., Tsuyuki, T. & Takahashi, T. The occurrence of 19,28-bisnorlanostane derivatives in a plant fossil: A novel geochemical degradation process of triterpenoids. *Geochimica et Cosmochimica Acta* **54**, 3253-3257, doi:10.1016/0016-7037(90)90143-9 (1990).
258 Killops, S. D., Zhang, S. & Lichtfouse, E. Triaromatic dinosteroids – Isomeric distributions and their geochemical significance. *Organic Geochemistry* **162**, 104300, doi:doi.org/10.1016/j.orggeochem.2021.104300 (2021).