OPINION

Unravelling ancient microbial history with community proteogenomics and lipid geochemistry

Jochen J. Brocks and Jillian Banfield

Abstract | Our window into the Earth's ancient microbial past is narrow and obscured by missing data. However, we can glean information about ancient microbial ecosystems using fossil lipids (biomarkers) that are extracted from billion-year-old sedimentary rocks. In this Opinion article, we describe how environmental genomics and related methodologies will give molecular fossil research a boost, by increasing our knowledge about how evolutionary innovations in microorganisms have changed the surface of planet Earth.

The first appearance of metabolic pathways that allowed microorganisms to exploit new and previously untapped energy sources, such as methanogenesis, sulphate reduction and photosynthesis, must have had profound effects on the chemical composition of the atmosphere and oceans. One of the main goals of geobiology is to determine when such physiologies first appeared in the Earth's history and to correlate these events with the associated chemical and isotopic changes in the rock record (FIG. 1).

The focus of most microbiological investigations is on living cells. However, even in death, the remains of microorganisms from all three domains of life can provide important information about how natural ecosystems work. In lakes and oceans, most cellular polymers, such as DNA, proteins and carbohydrates, are degraded rapidly in the water column and, subsequently, in the bottom sediments. Under exceptional circumstances, ancient bacterial DNA can persist in sediments for several hundred thousand years[1], an observation that could revolutionize the study of relatively recent microbial ecosystems. However, more recalcitrant biomolecules, such as lipids, can permanently escape the recycling process. In fine-grained sediments, where water circulation is restricted, remineralization of dead biomass by oxygen-consuming heterotrophs and sulphate reducers often leads to anoxic and sulphidic pore waters. Under these reducing conditions, lipids lose their functionalities; for example, their double bonds and hydroxy groups. The exact mechanisms of the biological and chemical sedimentary reactions that turn natural products in recent ecosystems into the corresponding population of molecular fossils (diagenesis) are poorly understood and present a formidable challenge for biomarker interpretations. During diagenesis, molecules of dominant community members could be completely lost, whereas lipids of less abundant organisms could become enriched. Therefore, it is currently difficult to make quantitative statements about community composition based on biomarker abundance, and virtually impossible to draw conclusions about the absence of specific organisms. As a result, studies of diagenetic transformations under different chemical and geological conditions will be a priority of biomarker geochemistry in the coming years. A recent breakthrough in this area was the recognition that non-biological addition of hydrogen sulphide to double bonds, followed by reductive desulphurization, is an important pathway of lipid preservation[2,3] (FIG. 2).

Although most taxonomic information is lost in the transition from biomolecule to defunctionalized geomolecule, the generated hydrocarbon skeletons can often still be linked to precursors that are diagnostic for specific groups of organisms[4] (TABLE 1). These molecules are also highly resistant against further degradation. Encased in lithified sediments, they can persist as recognizable molecular skeletons for billions of years.

These molecular fossils can be extracted from rocks (the antiquity of which can be established by radiometric isotopic techniques) and identified using mass spectrometry. They can provide intriguing glimpses into the species composition of the Earth's ancient microbial ecosystems. A fascinating example is the abundance of glycerol dialkyl glycerol tetraethers (GDGTs) in black shales, deposited during the mid-Cretaceous oceanic anoxic event 1b, approximately 112 million years (Myr) ago. These GDGTs point to a massive expansion of pelagic, chemoautotrophic Crenarchaeota into the world's oceans[5].

In this Opinion article we describe how genomics, proteomics and metabolomics of natural microbial consortia could be combined to produce a tree of life in which each lineage is annotated with key biomarkers and the base sequence of all genes that are required for their biosynthesis. Ideally, some nodes in the tree would be roughly dated, and the topology would yield data about the frequency of horizontal gene transfer (HGT) and convergent evolution in lipid biosynthetic pathways.

Biomarkers and microbial detection

Biomarkers can be used to measure geochemical parameters, such as salinity, temperature and redox state. For example, the incorporation of cyclopentane rings into the biphytanyl chains of GDGTs increases the thermal transition point of cell membranes and is a mechanism for temperature adaptation in archaea. The number of incorporated rings in GDGTs extracted from marine sediments has a linear correlation with annual mean sea surface temperature[6]. As a classical example of the application of this correlation, GDGTs yielded a high-resolution record of sea surface temperatures across the Palaeocene–Eocene thermal maximum (PETM), which occurred ~55 Myr ago and was one of the most rapid global

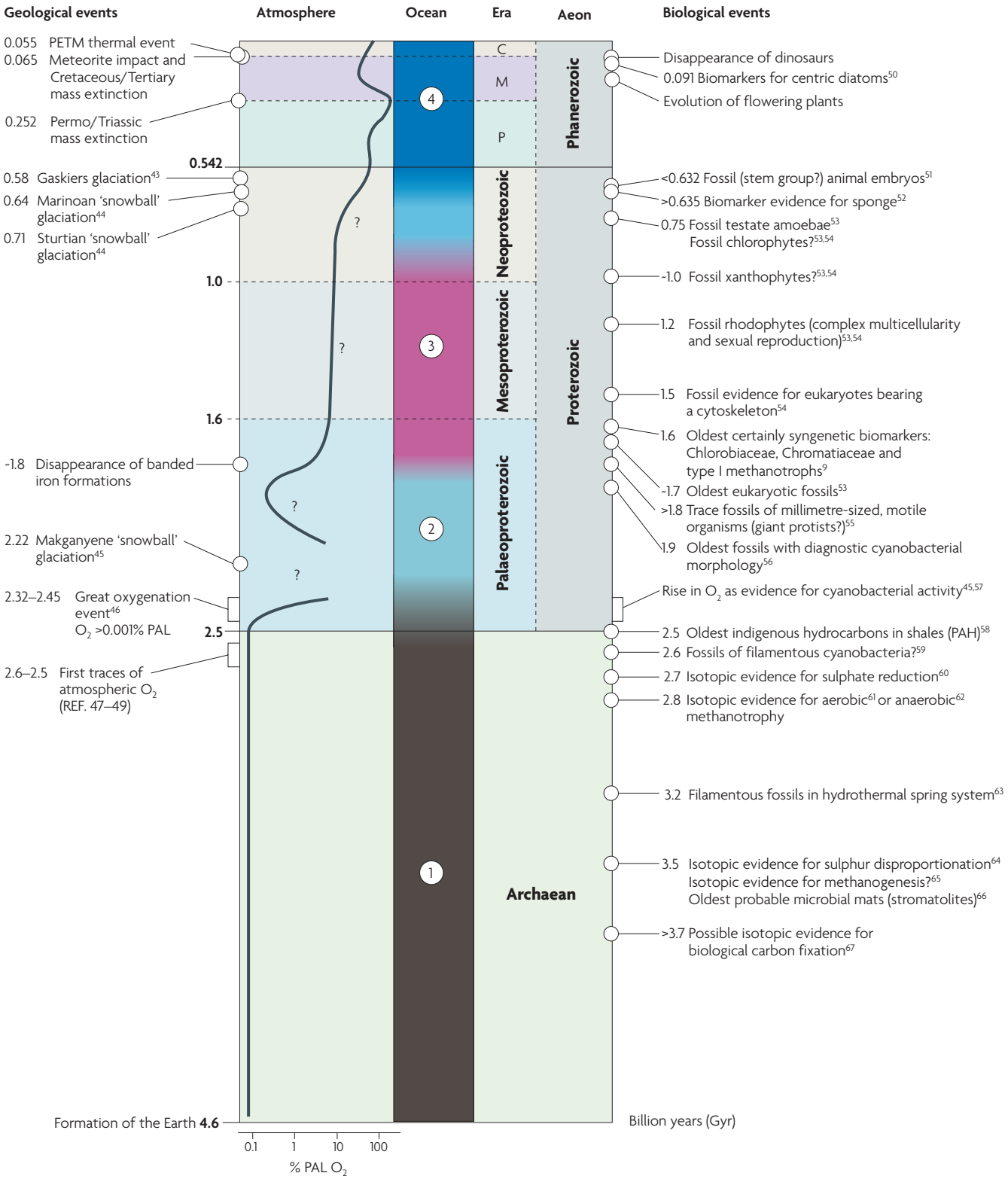

Figure 1 | **Timeline (in billions of years (Gyr) before the present) of major environmental and biological events in the Earth's history.** The left panel shows the evolution of atmospheric O_2 concentrations through time as a percentage of the present atmospheric level (PAL)[42]. The question marks indicate that the trajectory of the curve is poorly constrained. The middle panel shows the evolution of the redox state of the oceans (excluding temporary conditions such as oceanic anoxic events), including: anoxic and ferrous deep oceans and anoxic surface waters (step 1); surface waters that became increasingly oxygenated (step 2); anoxic and possibly sulphidic deep ocean waters and oxygenated surface waters (step 3); and predominantly oxygenated deep oceans (step 4). C, Cainozoic; M, Mesozoic; P, Palaeozoic; PAH, polyaromatic hydrocarbon; PETM, Palaeocene–Eocene thermal maximum.

heating events in the Earth's history. The GDGT data, in combination with other geochemical proxies, showed that the sharp heating event was the cause, not the consequence, of a rapid injection of the greenhouse gas methane into the atmosphere–ocean system[7].

A second, unique application of biomarkers is the reconstruction of extremely ancient microbial ecosystems. The oldest-known indigenous biomarkers[8] were found in 1,640-Myr-old sedimentary rocks that were deposited in northern Australia, in a deep rift basin that was probably connected to the ocean (FIG. 1). The biomarkers identified an entirely bacterial and archaeal ecosystem that was distinct from communities found in modern marine environments[9]. Extracts of the oil-prone rocks yielded high concentrations of the aromatic carotenoids okenane, indicating that there was prolific growth of planktonic purple sulphur bacteria (Chromatiaceae), and chlorobactane and isorenieratane, indicating that green and brown pigmented green sulphur bacteria (Chlorobiaceae) were also present. The sea was stratified and sulphidic up into the photic zone of the water column, consistent with models that predict vastly euxinic oceans between ~1,800 and 800 Myr ago (FIG. 1). A lack of steroids suggests that eukaryotic algae did not yet have an important role in this ancient sea.

Okenane is unique to the 1,640-Myr-old rocks from northern Australia and has never been discovered in younger rocks. In the modern biosphere, the only known precursor with an okenane skeleton is okenone (FIG. 2), a purple-coloured carotenoid found exclusively in Chromatiaceae. The presence of okenane in these rocks was therefore interpreted as evidence that the family Chromatiaceae must have evolved before the end of the Palaeoproterozoic period[9] (FIG. 1). But is this interpretation robust? Could other, as-yet-uncharacterized modern lineages produce okenone? Could Chromatiaceae have acquired the ability to synthesize okenone more recently, through HGT? If acquisition was through HGT, the biomarker should match the gene donor and not the recipient. An alternative explanation is that the okenane in ancient rocks was derived from a group of organisms that is now extinct. It is clear that the interpretation of molecular fingerprints preserved in rocks is limited by our knowledge of the biosynthetic capacities of organisms that existed when the rocks were formed.

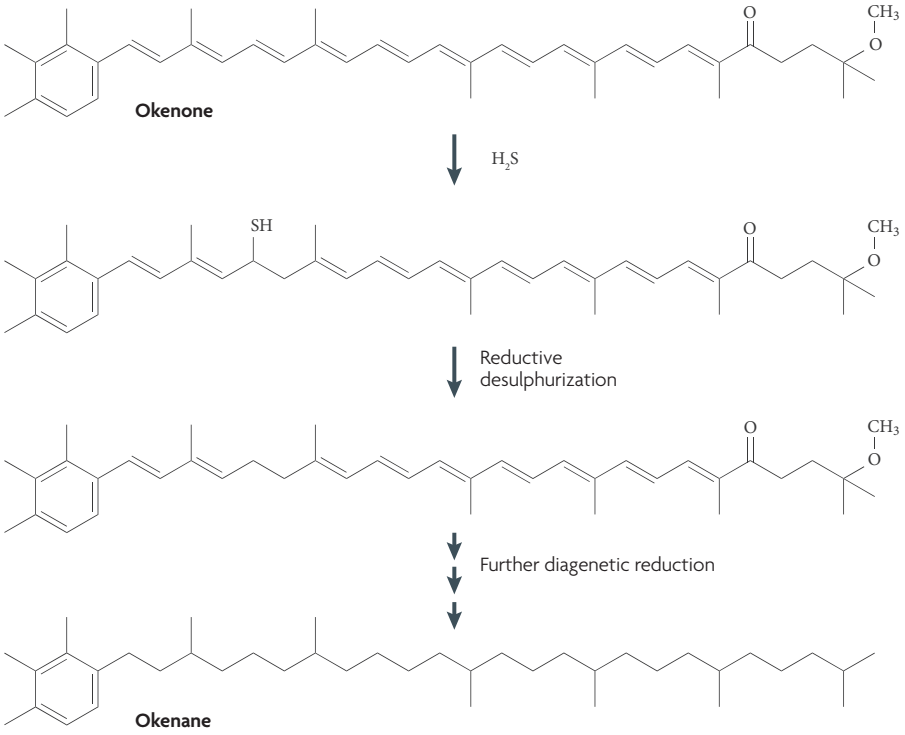

Figure 2 | **The diagenetic transformation of okenone, a carotenoid found exclusively in planktonic purple sulphur bacteria (Chromatiaceae), to the molecular fossil okenane.** Okenane retains the diagnostic carbon skeleton of its precursor and can be extracted from billion-year-old sedimentary rocks. It is regarded as a biomarker for Chromatiaceae.

Currently, biomarker interpretations are based primarily on lipids detected in extant organisms that have been grown in the laboratory. However, most bacteria and archaea have not yet been isolated[10]. Of 105 known bacterial and archaeal phyla, 73 do not have a single cultured representative, and we do not have a single complete genome for 76 phyla[11]. Even when organisms can be grown in isolation, their lipid products might differ from those produced in nature. Consequently, our knowledge about the distribution of biomarkers in nature is not sufficient to assess the reliability of biomarkers.

'Orphan biomarkers' are hydrocarbons that have been detected in sedimentary rocks and crude oils, but have as-yet-unknown biological sources[12]. The most prominent orphan biomarkers are cheilanthanes, which are tricyclic terpanes that occur in most sedimentary rocks from the Precambrian period to the present day. These biomarkers are as abundant and widely distributed as eukaryotic steranes and bacterial hopanes, yet we have not recognized a single living representative. Evidently, we are still unaware of the natural products of some of the most common organisms in the ocean.

Lipids produced by bacteria and archaea

Improved cultivation and isolation techniques[10] would clearly be an important step towards linking newly described bacteria and archaea to their lipid products. However, the diversity of bacteria and archaea is probably vast, and the challenges of isolating specific organisms of interest are so considerable that most of these analyses are likely to be impossible in the foreseeable future. Cultivation-independent techniques could offer a promising alternative way to fill the gaps in our knowledge of lipid biosynthetic pathways, and such techniques have already considerably expanded our understanding of microbial biodiversity[13]. The generation of 16S rRNA gene clone libraries and group-specific fluorescence *in situ* hybridization (FISH) has helped to describe which organisms are present in natural environments and which microbial groups are quantitatively important[14]. Cultivation-independent sequencing of functional genes has provided information about the number and diversity of organisms that can produce a certain class of lipids in a particular environment. For example, in a seminal study[15], Pearson and colleagues used degenerate PCR primers to study the domain-wide genetic diversity of the squalene-hopene cyclase

Table 1 | **Selected biomarkers, their biological origins and environmental interpretation**

Biomarker	Predominant biological origins and environmental interpretation
Archaea	
Head-to-head linked acyclic isoprenoids (C_{31} to C_{40}), archaeol, other isoprenoidal glycerol ethers and isoprenoidal glycerol dialkyl glycerol tetraethers	Archaea in general[4,12]; the number of rings incorporated into the isoprenoid chains of glycerol dialkyl glycerol tetraethers is a proxy for annual mean sea surface temperature (TEX_{86})[6]
2,6,15,19-Tetramethylicosane	Marine pelagic Crenarchaeota; exclusively associated with a mid-Cretaceous oceanic anoxic event[5,72]
Bacterioruberane	A predicted biomarker for Haloarchaea; postulated to be indicative of salt lake environments[12]
Crocetane	Archaeal anaerobic methane oxidizers; indicative of sub-marine gas seeps, gas hydrates or mud volcanoes[73] (also found as a degradation product of diaromatic carotenoids)
2,6,10,15,19-Pentamethylicosane	Methanogenic and methanotrophic archaea[74]
Crenarchaeol	Marine pelagic Crenarchaeota; found from the Cretaceous period to the present[5]
Bacteria	
C_{31} to C_{35} hopanes	Bacteria in general[75]
2α-Methylhopanes	Cyanobacteria[17], Pseudomonadales and Rhizobiales[76]
3β-Methylhopanes	Type I methanotrophic bacteria (Methylococcaceae), other methylotrophs and acetic acid bacteria[77,78]
4-Methylcholestanes and 4,4-dimethylcholestanes	Type I methanotrophic bacteria (Methylococcaceae)[9,79], dinoflagellates and biosynthetic intermediates in all eukaryotes
Chlorobactane	Green pigmented Chlorobiaceae; indicates anoxic and sulphidic conditions in the water column in the presence of light (photic zone euxinia) or in microbial mats[80]
Isorenieratane	Carbon isotopically heavy isorenieratane is diagnostic for brown pigmented Chlorobiaceae and is indicative of photic zone anoxia (in the water column or in microbial mats)[81]; isotopically lighter isorenieratane may be derived from sponges and actinomycetes[82]
Okenane	Purple sulphur bacteria (Chromatiaceae); indicative of photic zone euxinia and planktonic conditions (currently only associated with the late Palaeoproterozoic period[9,82])
Eukaryotes	
24-Norcholestane (C_{26} steroids)	Diatoms; found from the Cretaceous period to the present[83]
Cholestanes (C_{27} steroids)	Eukaryotes in general[84] with possible minor contributions from myxobacteria[85]; probably predominantly derived from red algae in the Proterozoic period
Ergostanes (C_{28} steroids)	Diagnostic for eukaryotes[84]; relative concentrations increased through the Phanerozoic period, when they were predominantly derived from secondary symbiotic algae such as dinoflagellates, diatoms and coccolithophorids[86]
Stigmastane (C_{29} steroids)	Diagnostic for eukaryotes[84]; the main sources in the Neoproterozoic period were probably green algae and in the post-Devonian period, probably terrestrial plants
24-*n*-Propylcholestanes (C_{30} steroids)	Pelagophyte algae ('brown tides' and Sarcinochrysidales); almost always representative of marine environments[87]
24-Isopropylcholestanes (C_{30} steroids)	Demosponges; found from the Neoproterozoic period to the present and represent the oldest (>635 million years (Myr) old) credible evidence for animal life in the geological record[52,88,89]
Dinosteranes (C_{30} steroids)	Predominantly dinoflagellates (rare in diatoms); found from the Mesozoic to Cenozoic periods (dinosteranes in older rocks are either from younger contamination or are derived from 'protodinoflagellates')[90]
C_{30} highly branched isoprenoid alkanes	Centric diatoms (*Rhizosolenia*); found in marine environments from 91.5 Myr ago (Upper Turonian period) to the present[50,91]
C_{37} to C_{39} di- to tetra-unsaturated alkenones	Prymnesiophyte algae (marine); the number of double bonds is used to estimate past temperatures in the photic zone of the water column (U^{37}_{k} proxy)[92]
Gammacerane	Probably predominantly bacterivorous ciliates (possibly also fungi, ferns and alphaproteobacteria)[93,94]
Oleanane, cadinanes and bicadinanes[95]	Angiosperms; oleanane in pre-Cretaceous rocks predates the oldest angiosperm fossils[96]
Phyllocladanes, beyerane, kaurane and atisane	Conifers; found from the Devonian period to the present[4,97]

(*sqhC*) gene in marine and freshwater bacterioplankton. Functionalized hopanoids are preserved in large quantities in modern aquatic sediments, and hopanes, their fossil analogues, are probably the most abundant polycyclic hydrocarbons in crude oil of all ages[16]. Cyanobacteria were cited as a major source[17]. However, Pearson's environmental study yielded 79 novel *sqhC* sequences that averaged only 60% translated amino acid identity to their closest relatives in the databases. Surprisingly, cyanobacteria were not among them. This study highlights how the biological sources of even the more common biomarkers remain largely unknown[15].

Although the recovery of single targeted genes can yield insights into the lipid biosynthetic capacity of bacteria and archaea in natural environments, we posit that it is possible to determine an almost-complete lipid complement produced by individual bacteria and archaea in an environment through cultivation-independent genomic, proteomic and organic geochemical analysis of natural consortia. Using the approach outlined in FIG. 3, near-complete genomes of all dominant (>5% relative abundance) bacteria and archaea can be recovered from an environmental sample. This was first achieved for simple consortia that are dominated by a small number of organisms for which deep genomic sampling of dominant populations is possible. The approach has since been deployed across a wide range of environment types (reviewed in REF. 18). Community genomic data from an acidophilic biofilm yielded near-complete (>95%) genomes of the five dominant archaea and bacteria[19]. Only one of these species had been cultured previously. With improved methods for assigning reconstructed genome fragments to the organisms from which they were derived (binning), extensive genomic reconstruction of the same community has been achieved for nine archaea and five bacteria, some of which constitute only a few percent of the community (G.J. Dick, personal communication).

Correlating genes with lipid products. Once a near-complete gene complement has been recovered for an uncultivated organism, it should be possible to predict all the lipids that the organism has the ability to produce (FIG. 3). In practice, however, association of genes with lipid products can be difficult, owing to the fact that generic proteins, such as cytochromes or demethylases, are implicated in the structural modification and functional ornamentation of lipids. Even slight changes to the amino acid sequence of enzymes can cause widely different lipid products. For example, a single mutation in the protosteryl cation stabilizing region of cycloartenol synthase enables the production of additional sterols that have a different carbon skeleton (lanosterol and parkeol)[20]. In many cases, it will therefore be difficult to deduce the carbon skeleton of a lipid product based on genomic information alone, and even more challenging to predict the number, position and stereochemistry of functional groups. However, in the case of well-studied proteins for which the three dimensional structures of the reactive centres and substrate pockets are known, it is possible to predict exact lipid products. For example, the position of bulky amino acid residues protruding into the product channel of prenyltransferases can be used to predict where the elongation of the growing polyprenyl chain will be terminated and whether the major products will have 20, 25, 30 or more carbon atoms[21].

For many lipid classes, such as sterols, carotenoids, acyclic isoprenoids and phospholipids, identification of complete biosynthetic pathways and exact lipid products based on genomic information is already possible. However, it remains a challenge to recognize genes that belong to new or incompletely characterized pathways. Most genes involved in lipid biosynthesis remain unknown, and environmental samples continuously yield new lipid structures and even new, and unusual, lipid classes, such as ladderanes[22]. Assigning lipids with uncharacterized biosynthetic pathways to bacteria and archaea that persistently defy isolation requires unconventional methods. Lipids with unusual carbon isotopic compositions, such as those of methanotrophs or green sulphur bacteria, could be linked to specific source organisms by comparison with the carbon isotopic composition of diagnostic DNA fragments[23] or by micron-scale ^{13}C measurements of fluorescently labelled cell aggregates using secondary ion mass spectrometry (SIMS)[24]. Moreover, with time-of-flight (TOF)-SIMS it is possible to map the distribution of lipids in mounted environmental samples, such as sections through microbial mats[25]. At a spatial resolution of 5 μm, TOF-SIMS can detect attomolar quantities of intact lipids with a sufficient mass resolution to identify a wide range of molecules. At lower mass resolution, TOF-SIMS can achieve a spatial resolution in the submicron range, at the size of individual microorganisms. If sensitivity and mass accuracy can be further improved at high spatial resolution, TOF-SIMS, combined with FISH, will allow characterization of the lipids in a single cell.

Analysing parallel variations of abundance of lipids and potential source organisms across multiple samples can also give clues about which source produces what. The genomes of potential source organisms can then be inspected, using bioinformatics, for distant homologues of enzymes known to be involved in the production of similar lipids. However, it can be difficult to recognize homologues from highly divergent lineages. This limitation will be overcome once comprehensive databases of genomic sequences and lipid products become available. At that time it might be possible to determine the function of a gene involved in lipid biosynthesis by comparison with three-dimensional enzyme structures from potentially distantly related organisms[26]. Corresponding genes could also be transferred to, and expressed in, organisms in culture and then be subjected to functional screening, an approach sometimes referred to as functional genomics[27,28]. For example, a metagenomic library of river sediments surprisingly yielded a putative bacteriochlorophyll *a* synthase (*bchG*) gene associated with a 16S rRNA gene from the uncultivated miscellaneous crenarchaeotal group. The putative *bchG* was cloned and heterologously expressed in *Escherichia coli*; the resulting protein was capable of synthesizing bacteriochlorophyll *a*, presenting the first evidence for a functional enzyme for bacteriochlorophyll biosynthesis in archaea[29].

Community genomics also help to define the conditions that might bring recalcitrant bacteria and archaea into culture[30], facilitating the study of unknown lipid biosynthetic pathways using genetic and molecular biological techniques. New information about the conditions that can allow microorganisms to be cultured will also come from comparisons of the proteomes of cultivated whole communities with those of environmental communities. Anomalies detected in the culture, such as overproduction of oxidative stress response proteins and osmoprotectants, will reveal how the media could be revised to improve the growth of stressed organisms, ultimately leading to their isolation and the study of their lipid content.

Environmental regulation of lipid biosynthesis. Some lipids, and their structural variants, are produced in response to changing environmental conditions; for example, the GDGTs discussed above incorporate an increasing number of rings into the lipid chain in response to rising temperatures. Although community genomic data sets contain information about which

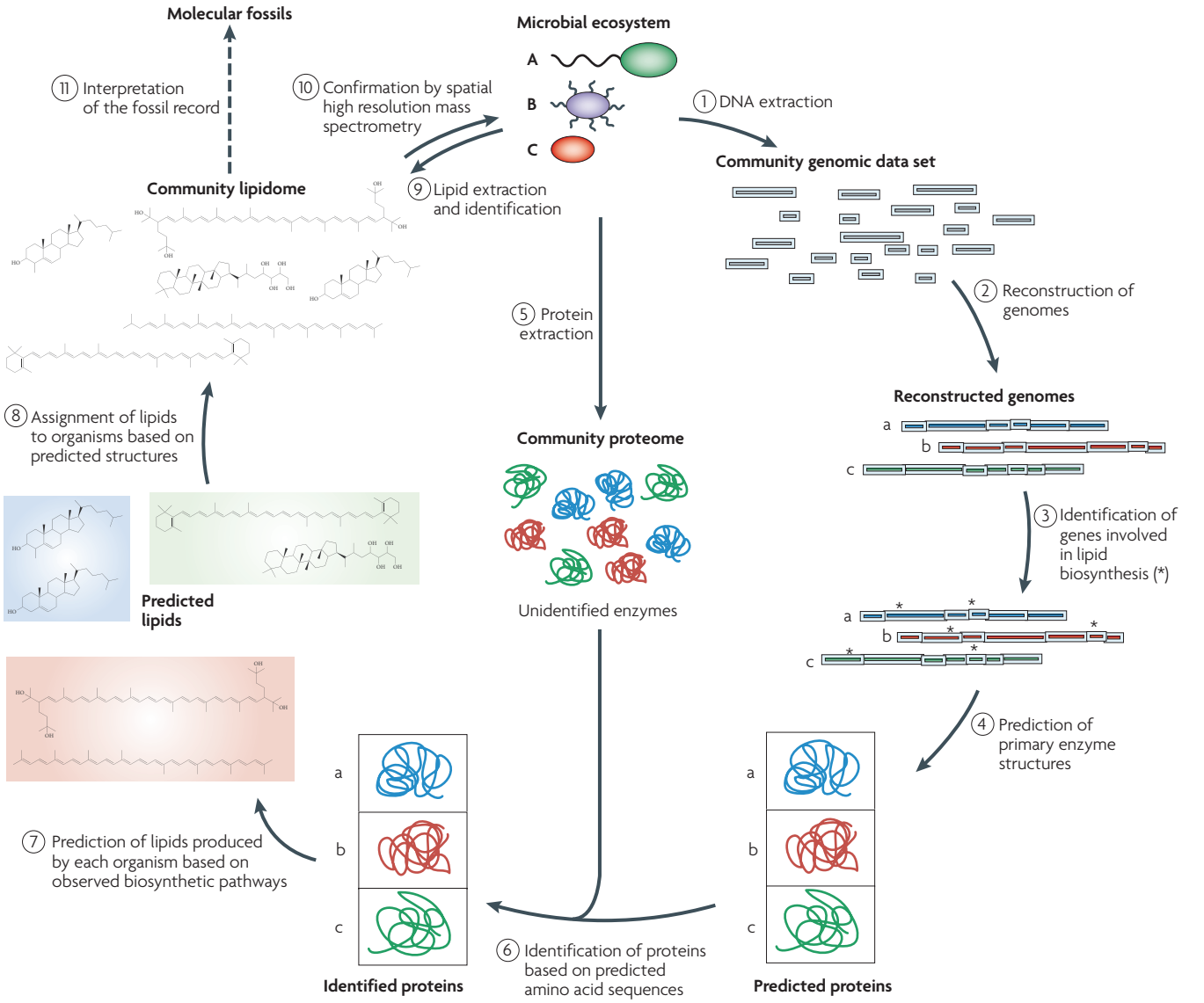

Figure 3 | **Community genomics and proteomics for prediction of lipid products.** With nearly completely reconstructed genomes and knowledge of which genes were expressed at the time when the sample was taken, it should, in principle, be possible to assign each lipid detected in a community lipidome to an organism or group of organisms in a consortium, even if these organisms cannot be isolated and grown in culture. In step 1, DNA is extracted from natural samples, fragments are sequenced and sequences are assembled to reconstruct larger stretches of genome. The reconstructed DNA segments are assigned to populations of organisms by co-assembly into larger fragments that carry phylogenetically informative genes or are binned based on metrics such as depth of sampling, GC percentage[19] or through sequence signature analysis[68–70], following methods pioneered by Karlin and colleagues[71]. In this way, nearly complete genomes of all dominant (>5% relative abundance) bacteria and archaea can be recovered (step 2). Genes involved in lipid biosynthesis are identified (step 3) and primary protein structures are predicted according to the base sequences of these genes (step 4). Proteins involved in lipid biosynthesis that were extracted from the same environmental sample (step 5) can then be identified on the basis of the predicted primary protein structures and assigned to individual organisms (step 6). Using information about the enzymes that are part of well-studied lipid biosynthetic pathways and that were detected in the environmental sample, it should be possible to predict major lipid end products (step 7) and thereby assign actual lipids in the environment to individual organisms or a group of organisms (steps 8–10). The identification of lipids of organisms that cannot be isolated could improve our understanding of the sources and environmental importance of fossil lipids in the geological record (step 11).

microorganism can potentially produce what type of lipid, we cannot determine which lipid biosynthetic enzymes are actually expressed under specific environmental conditions, or at a particular point in time, using this information alone. This question of 'who makes what, when' can be answered by community proteomics, the study of the presence and abundance of proteins in natural microbial environments (FIG. 3; steps 4, 5 and 6).

Without genomic reference information, the study of proteins in environmental samples is restricted to a small fraction of the community proteome[31]. However, using information from predicted amino acid sequences derived from community genomic data sets (FIG. 3; step 4), identification of several thousand protein structures is possible[32]. In these analyses, proteins are extracted from environmental samples (FIG. 3; step 5) and enzymatically fragmented at specific amino acids to produce peptides. The mass to charge ratio is determined for

both the whole peptides and their fragmentation products. To identify peptides, the mass spectra are then compared with theoretical mass to charge ratios for peptides and their fragmentation products that are predicted from genome sequences (FIG. 3; step 6). The experimentally identified peptides are mapped back onto predicted proteins to determine the source organism. This methodology was used to identify 2,033 proteins from the 5 dominant species that formed an acidic biofilm[32]. For these 5 species, over 50% of all predicted proteins were identified (with a <1% false positive rate[33]). A probable function was recognized for 69% of these proteins, including enzymes involved in lipid metabolism. The community proteome reveals which enzymes involved in lipid biosynthesis are expressed, and roughly in what concentration, at the time of sampling. In this way, it should become possible to assign each lipid detected in the community lipidome to an organism, or group of organisms, in the consortium (FIG. 3; steps 7–10). Community proteomic methods offer a snapshot of the lipid biosynthetic activity of a microbial community under defined environmental conditions and can be applied to environmental perturbation or time series experiments.

Proteogenomic studies on environmental samples that contain uncultivated organisms have, as yet, only been applied to abundant members of microbial communities[32]. Proteogenomics of fractionated samples, such as cells separated according to size or fluorescence-activated cell sorting[14], could be used to capture genomes, proteomes and lipids of organisms that can play a crucial part in an ecosystem but that have low relative abundances. Studies of the genetic diversity and proteomes of complex microbial ecosystems such as soil and marine water will also require the collection of extensive reference genomes, and DNA amplification from single cells will probably become the method of choice to generate this data[34–36].

With recent advances in peptide mass spectrometry and sample fractionation, it is possible to detect proteins with five orders of dynamic range[37]. As the different proteins of microorganisms have a concentration range of about six orders of magnitude, it will be possible to identify all but the least-abundant proteins in sorted cell concentrates. This offers the potential to study subtle adjustments of the lipid biosynthetic apparatus to changing physical and chemical conditions without disturbing the system. It will also be possible to detect adjustments in lipid metabolism in microbial communities by sequencing cDNA products from environmental mRNA transcripts[38]. Using the techniques described above, changes in lipid biosynthesis and other metabolic activities could be followed simultaneously in all major and some minor organisms of a community. This, in turn, will help to identify the functions of those lipids, and their structural variants, that are important in the geological record, such as hopanoids, methylhopanoids, and sterols and their alkylated counterparts.

Some secondary metabolites carry specific carbon skeletons that can, in principle, be preserved as hydrocarbon fossils. However, their concentrations will be minute, and their detection in the complex matrix of bitumens and oils will be a formidable challenge. Coupling two-dimensional gas chromatography techniques with fast multidimensional MS (such as TOF–TOF) could disentangle complex hydrocarbon mixtures and make even the minutest traces visible. In this way, fossil siderophores could yield information about the scarcity of specific ions in ancient oceans and fossil antibiotics could tell us when the first bacterivorous predators arose.

Extrapolating into the past

Specific lipids can be present in different lineages through common inheritance, convergent evolution or HGT. HGT is particularly problematic for biomarker interpretations because too little is known about its prevalence. For example, 1.6-Gyr (billion years)-old rocks from northern Australia contain steroids that are methylated at C4. The biosynthetic capacity to produce these steroids is found in all eukaryotes but also occurs in at least three unrelated groups of bacteria[39]. Phylogenetic trees of oxidosqualene cyclase, the first enzyme in the sterol biosynthetic pathway, show that sterol biosynthesis was laterally transferred between eukaryotes and the three bacterial groups[39]. The timing and direction of transfer are currently unknown. However, information about the age of the HGT events, obtained from molecular clocks, could help to unravel which of the four steroid-producing groups is the most likely source of the 1.6-Gyr-old steroids. For example, if the bacterial groups acquired oxidosqualene cyclase from eukaryotes less than 1.6 Gyr ago, then Eukarya are the most likely source. Even though such molecular clock estimates will commonly have uncertainties in the order of hundreds of millions of years, the information can be sufficient to discriminate between alternative biological sources of particularly ancient biomarkers.

One of the greatest uncertainties in the interpretation of ancient biomarkers is the possibility that suitable precursor lipids were also present in groups of organisms that are now extinct. Although this problem seems to be intractable, it should be possible to assess the probability that extinct groups acquired complete or partial biosynthetic pathways in the past, based on information about the frequency of HGT and convergent evolution in extant organisms. For example, we could estimate the probability that extinct archaea acquired steroid biosynthesis from bacteria by looking at the frequency of similar events between archaea and bacteria in the past. The frequency of HGT is probably dependent on the function of the metabolic product, the position of the newly acquired pathway in the existing metabolic network, and the classes and number of genes involved in the transfer. The frequency of HGT between two groups will also strongly depend on the evolutionary distance between them. On the basis of experimental[40] and community genomic data[41], it is now well-established that the frequency of genetic exchange by homologous recombination shows a log-linear dependence on sequence identity. In other words, homologous recombination events become increasingly rare as organisms become more distantly related. But can relationships be defined for other lateral transfer mechanisms? For example, although inter-domain transfers of lipid biosynthetic pathways have been documented, they seem to be rare. Massive, cultivation-independent sequence data sets can provide the required rates of HGT for different metabolic products at all hierarchical levels in the tree of life. They will also yield insights into the mechanisms of HGT, particularly the role of viruses and other mobile elements in gene transfer. Other questions that could be answered include: how commonly do viral genomes encode lipid pathways and how wide is the host range of modern viruses? Although uncertainties will remain, such as the size of the microbial biosphere in the past and changes in frequency of HGT over time, the parameters discussed above can be combined to estimate the probability of transfer events and convergent evolution in the past and to make statements about the reliability of biomarker interpretations.

Outlook

Bacteria and archaea evolved early in the Earth's history (FIG. 1). Yet there is no diagnostic morphological evidence for the domain Bacteria in the rock record until 2.1 Gyr ago, and no fossilized cell of any age

has been recognized as an archaeon. The reason is clear: even in the exceptionally rare event that soft-bodied organisms are preserved in rock, the simple shapes and forms of almost all archaea and bacteria make it difficult to assign cellular fossils to any particular clade. Molecular fossils fill some of the vast gaps in the geological record. Diagnostic biomarkers exist for organisms that otherwise leave no traces, such as green and purple sulphur bacteria, crenarchaeotes, and bacterial and archaeal methanotrophs. These biomarkers, accompanied by isotopic signatures of biogenic materials, rare diagnostic body fossils, and deductions from molecular clocks and ancestral state reconstructions, are our window into the Earth's microbial past (FIG. 1). Biomarkers have helped to track the response of microorganisms to cataclysmic events in the Earth's history, such as major perturbations of the carbon cycle, anoxic oceans, major glaciations, global heating events and resulting catastrophic losses in biodiversity. As many organisms and their lipids are sensitive to environmental conditions, biomarkers have also yielded records of marine temperatures, eutrophication, anoxic and sulphidic conditions and hypersalinity.

The study of ancient microbial ecosystems using hydrocarbon biomarkers is a young field and most applications remain to be uncovered. As we fill in the gaps in our knowledge about the biosynthetic capacities of extant organisms, and link gene inventories to lipid profiles from environmental samples, we will discover new biomarkers and their meaning. This will help decipher how the Earth and its biosphere have evolved together. In particularly ancient rocks, we could discover molecules that resemble extant lipids but that are the products of an ancestral, abbreviated biosynthetic pathway. The sequences of such molecules would yield information about the timing of the evolution of lipid biosynthetic pathways directly from the geological record. We might also finally be able to adopt orphan biomarkers and use them to probe the appearance of currently unknown organisms and link their metabolisms to events in the Earth's history.

Jochen J. Brocks is at the Research School of Earth Sciences, and Centre for Macroevolution and Macroecology, The Australian National University, Canberra, ACT 0200, Australia.

Jillian Banfield is at the Department of Earth and Planetary Science, University of California, Berkeley, CA 94720-4767, USA.

Correspondence to J.J.B.
e-mail: jochen.brocks@anu.edu.au

doi:10.1038/nrmicro2167

1. Coolen, M. J. L. & Overmann, J. 217,000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. *Environ. Microbiol.* **9**, 238–249 (2007).
2. Werne, J. P. *et al.* Investigating pathways of diagenetic organic matter sulfurization using compound-specific sulfur isotope analysis. *Geochim. Cosmochim. Acta* **72**, 3489–3502 (2008).
3. Hebting, Y. *et al.* Biomarker evidence for a major preservation pathway of sedimentary organic carbon. *Science* **312**, 1627–1631 (2006).
4. Peters, K. E., Walters, C. C. & Moldowan, J. M. *The Biomarker Guide* (Cambridge Univ. Press, UK, 2004).
5. Kuypers, M. M. M. *et al.* Massive expansion of marine archaea during a mid-Cretaceous oceanic anoxic event. *Science* **293**, 92–94 (2001).
6. Schouten, S., Hopmans, E. C., Schefuss, E. & Sinninghe Damsté, J. S. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* **204**, 265–274 (2002).
7. Sluijs, A. *et al.* Environmental precursors to rapid light carbon injection at the Palaeocene/Eocene boundary. *Nature* **450**, 1218–1221 (2007).
8. 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. *Geochim. Cosmochim. Acta* **52**, 1747–1763 (1988).
9. Brocks, J. J. *et al.* Biomarker evidence for green and purple sulphur bacteria in a stratified Paleoproterozoic sea. *Nature* **437**, 866–870 (2005).
10. Giovannoni, S. & Stingl, U. The importance of culturing bacterioplankton in the 'omics' age. *Nature Rev. Microbiol.* **5**, 820–826 (2007).
11. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial species. *Nature Rev. Microbiol.* **6**, 431–440 (2008).
12. Brocks, J. J. & Summons, R. E. in *Treatise on Geochemistry* Vol. 8 (ed. Schlesinger, W. H.) 63–115 (Elsevier–Pergamon, Oxford, 2004).
13. Medini, D. *et al.* Microbiology in the post-genomic era. *Nature. Rev. Microbiol.* **6**, 419–430 (2008).
14. Amann, R. & Fuchs, B. M. Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nature Rev. Microbiol.* **6**, 339–348 (2008).
15. Pearson, A., Flood Page, S. R., Jorgenson, T. L., Fischer, W. W. & Higgins, M. B. Novel hopanoid cyclases from the environment. *Environ. Microbiol.* **9**, 2175–2188 (2007).
16. Ourisson, G. in *Early Life on Earth, Nobel Symposium* Vol. 84 (ed. Bengtson, S.) 259–269 (Columbia Univ. Press, New York, 1994).
17. Summons, R. E., Jahnke, L. L., Hope, J. M. & Logan, G. A. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400**, 554–557 (1999).
18. Wilmes, P., Simmons, S. L., Denef, V. J. & Banfield, J. F. The dynamic genetic repertoire of microbial communities. *FEMS Microbiol. Rev.* **33**, 109–132 (2009).
19. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
20. Hart, E. A. *et al.* Directed evolution to investigate steric control of enzymatic oxidosqualene cyclization. An isoleucine-to-valine mutation in cycloartenol synthase allows lanosterol and parkeol biosynthesis. *J. Am. Chem. Soc.* **121**, 9887–9888 (1999).
21. Ohnuma, S. *et al.* A pathway where polyprenyl diphosphate elongates in prenyltransferase. Insight into a common mechanism of chain length determination of prenyltransferases. *J. Biol. Chem.* **273**, 26705–26713 (1998).
22. Sinninghe Damsté, J. S. *et al.* Linearly concatenated cyclobutane lipids form a dense bacterial membrane. *Nature* **419**, 708–712 (2002).
23. Sessions, A. L., Sylva, S. P. & Hayes, J. M. A moving-wire device for carbon-isotopic analyses of nanogram quantities of nonvolatile organic carbon. *Anal. Chem.* **77**, 6519–6527 (2005).
24. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**, 484–487 (2001).
25. Thiel, V. *et al.* Biomarkers at the microscopic range: ToF-SIMS molecular imaging of Archaea-derived lipids in a microbial mat. *Geobiology* **5**, 413–421 (2007).
26. Shin, D. *et al.* Structure-based inference of molecular functions of proteins of unknown function from Berkeley Structural Genomics Center. *J. Struct. Funct. Genomics* **8**, 99–105 (2007).
27. Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A. & Handelsman, J. Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *ISME J.* **3**, 241–251 (2008).
28. Lim, H. K. *et al.* Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl. Environ. Microbiol.* **71**, 7768–7777 (2005).
29. Meng, J. *et al.* An uncultivated crenarchaeota contains functional bacteriochlorophyll *a* synthase. *ISME J.* **3**, 106–116 (2009).
30. Tyson, G. W. & Banfield, J. F. Cultivating the uncultivated: a community genomics perspective. *Trends Microbiol.* **13**, 411–415 (2005).
31. Denef, V. J., Shah, M. B., VerBerkmoes, N. C., Hettich, R. L. & Banfield, J. F. Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis. *J. Proteome Res.* **6**, 3152–3161 (2007).
32. Ram, R. J. *et al.* Community proteomics of a natural microbial biofilm. *Science* **308**, 1915–1920 (2005).
33. Lo, I. *et al.* Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446**, 537–541 (2007).
34. Walker, A. & Parkhill, J. Single-cell genomics. *Nature Rev. Microbiol.* **6**, 176–177 (2008).
35. Raes, J. & Bork, P. Molecular eco-systems biology: towards an understanding of community function. *Nature Rev. Microbiol.* **6**, 693–699 (2008).
36. Woyke, T. *et al.* Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4**, e5299 (2009).
37. VerBerkmoes, N. C., Denef, V. J., Hettich, R. L. & Banfield, J. F. Functional analysis of natural microbial consortia using community proteomics. *Nature Rev. Microbiol.* **7**, 196–205 (2009).
38. Frias-Lopez, J. *et al.* Microbial community gene expression in ocean surface waters. *Proc. Natl Acad. Sci. USA* **105**, 3805–3810 (2008).
39. Pearson, A., Budin, M. & Brocks, J. J. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proc. Natl Acad. Sci. USA* **100**, 15352–15357 (2003).
40. Roberts, M. S. & Cohan, F. M. The effect of DNA sequence divergence on sexual isolation in Bacillus. *Genetics* **134**, 401–408 (1993).
41. Eppley, J. M., Tyson, G. W., Getz, W. M. & Banfield, J. F. Genetic exchange across a species boundary in the archaeal genus Ferroplasma. *Genetics* **177**, 407–416 (2007).
42. Canfield, D. E. The early history of atmospheric oxygen: homage to Robert M. Garrels. *Annu. Rev. Earth Planet Sci.* **33**, 1–36 (2005).
43. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* **315**, 92–95 (2007).
44. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic snowball Earth. *Science* **281**, 1342–1346 (1998).
45. Kirschvink, J. L. & Kopp, R. E. Palaeoproterozoic ice houses and the evolution of oxygen-mediating enzymes: the case for a late origin of photosystem II. *Philos. Trans. R. Soc. Lond. B* **363**, 2755–2765 (2008).
46. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
47. Anbar, A. D. *et al.* A whiff of oxygen before the Great Oxidation Event? *Science* **317**, 1903–1906 (2007).
48. Kaufman, A. J. *et al.* Late Archean biospheric oxygenation and atmospheric evolution. *Science* **317**, 1900–1903 (2007).
49. Wille, M. *et al.* Evidence for a gradual rise of oxygen between 2.6 and 2.5 Ga from Mo isotopes and Re-PGE signatures in shales. *Geochim. Cosmochim. Acta* **71**, 2417–2435 (2007).
50. Sinninghe Damsté, J. S. *et al.* The rise of the rhizosolenid diatoms. *Science* **304**, 584–587 (2004).
51. Xiao, S., Zhang, Y. & Knoll, A. H. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* **391**, 553–558 (1998).
52. Love, G. D. *et al.* Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**, 718–721 (2009).
53. Knoll, A. H., Javaux, E., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Philos. Trans. R. Soc. Lond. B* **361**, 1023–1038 (2006).
54. Butterfield, N. J. A vaucheriacean alga from the middle Neoproterozoic of Spitsbergen: implications for the evolution of Proterozoic eukaryotes and the Cambrian explosion. *Paleobiology* **30**, 231–252 (2004).

55. Rasmussen, B., Bengtson, S., Fletcher, I. R. & McNaughton, N. J. Discoidal impressions and trace-like fossils more than 1200 million years old. *Science* **296**, 1112–1115 (2002).
56. Hofmann, H. J. Precambrian microflora, Belcher Islands, Canada: significance and systematics. *J. Paleontol.* **50**, 1040–1073 (1976).
57. Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. Reassessing the first appearance of eukaryotes and Cyanobacteria. *Nature* **455**, 1101–1104 (2008).
58. Brocks, J. J., Summons, R. E., Buick, R. & Logan, G. A. Origin and significance of aromatic hydrocarbons in giant iron ore deposits of the late Archean Hamersley Basin in Western Australia. *Org. Geochem.* **34**, 1161–1175 (2003).
59. Kazmierczak, J. & Altermann, W. Neoarchean biomineralization by benthic cyanobacteria. *Science* **298**, 2351 (2002).
60. Shen, Y., Buick, R. & Canfield, D. E. Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**, 77–81 (2001).
61. Hayes, J. M. in *Early Life on Earth* (ed. Bengtson, S.) 220–236 (Columbia Univ. Press, New York, 1994).
62. Hinrichs, K.-U. Microbial fixation of methane carbon at 2.7 Ga: was an anaerobic mechanism possible? *Geochem. Geophys. Geosyst.* **3**, 1042 (2002).
63. Rasmussen, B. Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. *Nature* **405**, 2676–2679 (2000).
64. Philippot, P. *et al.* Early Archaean microorganisms preferred elemental sulfur, not sulfate. *Science* **317**, 1534–1537 (2007).
65. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519 (2006).
66. Walter, M. R., Buick, R. & Dunlop, J. S. R. Stromatolites 3,400–3,500 Myr old from the North Pole area, Western Australia. *Nature* **284**, 443–445 (1980).
67. Rosing, M. T. ^{13}C-depleted carbon microparticles in > 3700-Ma sea-floor sedimentary rocks from West Greenland. *Science* **283**, 674–676 (1999).
68. Teeling, H., Meyerdierks, A., Bauer, M., Amann, R. & Glöckner, F. O. Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ. Microbiol.* **6**, 938–947 (2004).
69. Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S. & Ikemura, T. Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. *DNA Res.* **12**, 281–290 (2005).
70. Mavromatis, K. *et al.* Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nature Methods* **4**, 495–500 (2007).
71. Karlin, S., Mrázek, J. & Campbell, A. M. Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.* **179**, 3899–3913 (1997).
72. Vink, A., Schouten, S., Sephton, S. & Sinninghe Damsté, J. S. A newly discovered norisoprenoid, 2,6,15,19-tetramethylicosane, in Cretaceous black shales. *Geochim. Cosmochim. Acta* **62**, 965–970 (1998).
73. Thiel, V. *et al.* Highly isotopically depleted isoprenoids: molecular markers for ancient methane venting. *Geochim. Cosmochim. Acta* **63**, 3959–3966 (1999).
74. Schouten, S., van der Maarel, M. J., Huber, R. & Sinninghe Damsté, J. S. 2,6,10,15,19-pentamethylicosenes in *Methanolobus bombayensis*, a marine methanogenic archaeon, and in *Methanosarcina mazei*. *Org. Geochem.* **26**, 409–414 (1997).
75. Rohmer, M., Bouvier-Navé, P. & Ourisson, G. Distribution of hopanoid triterpenes in prokaryotes. *J. Gen. Microbiol.* **130**, 1137–1150 (1984).
76. Rashby, S. E., Sessions, A. L., Summons, R. E. & Newman, D. K. Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc. Natl Acad. Sci. USA* **104**, 15099–15104 (2007).
77. Zundel, M. & Rohmer, M. Prokaryotic triterpenoids 1. 3 β-methylhopanoids from *Acetobacter* species and *Methylococcus capsulatus*. *Eur. J. Biochem.* **150**, 23–27 (1985).
78. Summons, R. E. & Jahnke, L. L. in *Biological Markers in Sediments and Petroleum* (eds Moldowan, J. M., Albrecht, P. & Philp, R. P.) 182–200 (Prentice Hall, New Jersey, 1992).
79. Bird, C. W. *et al.* Steroids and squalene in *Methylococcus capsulatus* grown on methane. *Nature* **230**, 473–474 (1971).
80. Schaeffer, P., Adam, P., Wehrung, P. & Albrecht, P. Novel aromatic carotenoid derivatives from sulfur photosynthetic bacteria in sediments. *Tetrahedr. Lett.* **38**, 8413–8416 (1997).
81. Summons, R. E. & Powell, T. G. Chlorobiaceae in Paleozoic seas revealed by biological markers, isotopes and geology. *Nature* **319**, 763–765 (1986).
82. 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. *Geochim. Cosmochim. Acta* **72**, 1396–1414 (2008).
83. Holba, A. G. *et al.* 24-norcholestanes as age-sensitive molecular fossils. *Geology* **26**, 783–786 (1998).
84. Volkman, J. K. Sterols in microorganisms. *Appl. Microbiol. Biotechnol.* **60**, 496–506 (2003).
85. 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*. *Mol. Microbiol.* **47**, 471–481 (2003).
86. Grantham, P. J. & Wakefield, L. L. Variations in the sterane carbon number distribution of marine source rock derived oils through geological time. *Org. Geochem.* **12**, 61–73 (1988).
87. Moldowan, J. M. *et al.* Sedimentary 24-*n*-propylcholestanes, molecular fossils diagnostic of marine algae. *Science* **247**, 309–312 (1990).
88. McCaffrey, M. A. *et al.* Paleoenvironmental implications of novel C_{30} steranes in Precambrian to Cenozoic age petroleum and bitumen. *Geochim. Cosmochim. Acta* **58**, 529–532 (1994).
89. McCaffrey, A. M., Dahl, J. E. P., Sundararaman, P., Moldowan, J. M. & Schoell, M. Source rock quality determination from oil biomarkers II: a case study using tertiary-reservoired Beauford Sea oils. *Am. Assoc. Pet. Geol. Bull.* **78**, 1527–1540 (1994).
90. Moldowan, J. M. & Talyzina, N. M. Biogeochemical evidence for dinoflagellate ancestors in the Early Cambrian. *Science* **281**, 1168–1170 (1998).
91. Nichols, P. D., Volkman, J. K., Palmisano, A. C., Smith, G. A. & White, D. C. Occurrence of an isoprenoid C_{25} di-unsaturated alkene and high neutral lipid content in Antarctic sea-ice diatom communities. *J. Phycol.* **24**, 90–96 (1988).
92. Brassell, S. C., Eglinton, G., Marlowe, I. T., Pflaumann, U. & Sarnthein, M. Molecular stratigraphy: a new tool for climatic assessment. *Nature* **320**, 129–133 (1986).
93. Ten Haven, H. L., Rohmer, M., Rullkötter, J. & Bisseret, P. Tetrahymanol, the most likely precursor of gammacerane, occurs ubiquitously in marine sediments. *Geochim. Cosmochim. Acta* **53**, 3073–3079 (1989).
94. Sinninghe Damsté, J. S. *et al.* Evidence for gammacerane as an indicator of water column stratification. *Geochim. Cosmochim. Acta* **59**, 1895–1900 (1995).
95. van Aarssen, B. G. K., Hessels, J. K. C., Abbink, O. A. & de Leeuw, J. W. The occurrence of polycyclic sesqui-, tri-, and oligoterpenoids derived from a resinous polymeric cadinene in crude oils from southeast Asia. *Geochim. Cosmochim. Acta* **56**, 1231–1246 (1992).
96. Moldowan, J. M. *et al.* The molecular fossil record of oleanane and its relation to angiosperms. *Science* **265**, 768–771 (1994).
97. Noble, R. A., Alexander, R., Kagi, R. I. & Knox, J. Tetracyclic diterpenoid hydrocarbons in some Australian coals, sediments and crude oils. *Geochim. Cosmochim. Acta* **49**, 2141–2147 (1985).

Acknowledgements
We thank W. Fischer and B. Rasmussen for advice on figure 1.

FURTHER INFORMATION
Jochen J. Brocks' homepage: http://shrimp.anu.edu.au/people/jjb/jjb.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

SCIENCE AND SOCIETY

Communicable disease among displaced Afghans: refuge without shelter

Alefiyah Rajabali, Omer Moin, Amna S. Ansari, Mohammad R. Khanani and Syed H. Ali

Abstract | More than 23 years of warfare in Afghanistan has caused over 6 million Afghans to seek asylum in approximately 70 different countries, with most Afghan refugees settling in the developing countries of Pakistan and Iran. In a developing host country, poor sanitation and nutrition, overcrowding and inaccessibility to health care facilities act synergistically to influence morbidity and mortality from infectious disease in the refugee population. In this Science and Society article we discuss the prevalence of transmissible infection, modes of transmission, associated risk factors, and the state of health and health care in the displaced Afghan population.

Afghan refugees constitute the single largest refugee population in the world[1]. Although most live in Pakistan (1.9 million) and Iran (0.9 million) (FIG. 1a), a number of other countries also host Afghan refugees, including India, Uzbekistan, the United Kingdom, Germany, the Netherlands, Denmark, Sweden, Norway, Austria, Hungary, Egypt, Canada, the United States and Australia[2]. Three key periods in the history of Afghanistan have generated waves of migrant refugees (FIG. 1b). The first