Royal Netherlands Institute for Sea Research

This is a pre-copyedited, author-produced version of an article accepted for publication, following peer review.

Villanueva, L.; Coolen, M.J.L. (2022). Contributions of genomics to lipid biomarker research: From paleoclimatology to evolution. *Elements 18(2)*: 87-92. DOI: 10.2138/gselements.18.2.87

Published version: https://dx.doi.org/10.2138/gselements.18.2.87

NIOZ Repository: http://imis.nioz.nl/imis.php?module=ref&refid=359700

[Article begins on next page]

The NIOZ Repository gives free access to the digital collection of the work of the Royal Netherlands Institute for Sea Research. This archive is managed according to the principles of the Open Access Movement, and the Open Archive Initiative. Each publication should be cited to its original source - please use the reference as presented.
When using parts of, or whole publications in your own work, permission from the author(s) or copyright holder(s) is always needed.

Contributions of genomics to lipid biomarker research: from paleoclimatology to evolution

Laura Villanueva[1,2], Marco J.L. Coolen[3]

[1]Royal Netherlands Institute for Sea Research. Landsdiep 4 1797 SZ ‘t Horntje (The Netherlands)

[2]Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Princetonlaan 8a 3584 CB Utrecht (The Netherlands)

[3]WA-Organic and Isotope Geochemistry Centre (WA-OIGC), School of Earth and Planetary Sciences (EPS) and the Institute for Geoscience Research (TIGeR), Curtin University, Bentley, WA 6102, Australia

ABSTRACT

Lipid biomarkers can be preserved over long geological timescales and are widely used as taxonomic markers of past and present microbial communities and as parts of organic paleoclimate proxies. Nevertheless, questions remain regarding the precise biological sources and evolution of the acquisition of specific lipids, and why and how they are synthesized. In the last two decades, the use of DNA-based approaches has proven to be key in unraveling some of these questions. As methodological approaches improve, (paleo)genomics increasingly supports lipid biomarker research. Here, we provide an overview on the usefulness of DNA-based approaches over the years, including ancient sedimentary DNA (*sed* aDNA) research and phylogenomics, and a perspective on the upcoming challenges of this field.

KEYWORDS: Lipid biomarkers, ancient sedimentary DNA, paleomicrobiology, Lipid biosynthesis, phylogenomics

INTRODUCTION

Microbial lipids have been widely used as markers of the presence of specific microorganisms in both present and past ecosystems, and as proxies for the reconstruction of past environments and climates (Naeher et al., this issue). In some

cases, the biological source (i.e., precursor) of a microbial lipid biomarker is unknown, in which case they are referred to as 'orphan biomarkers'. The identification of a lipid biomarker directly from pure (only one microorganism) or enriched (with higher abundance of the target microorganism) laboratory culture is the most definite way to confirm that biomarker's biological source. Nevertheless, it is not possible to assure that other microorganisms, which have not been screened yet for their lipid biomarker composition, are also potential sources of a lipid biomarker of interest. These caveats bring a certain level of uncertainty to the application of lipid biomarkers in past and present environments. Ideally, given a lipid biomarker of interest, we would like to address the following questions: which organism(s) produced the biomarker? How, where, and when was the lipid biomarker produced? Since when has this lipid biomarker been synthesized and how was it acquired? Studies based on the current microbial ecology of the biological sources of specific lipid biomarkers have proven to be very useful to determine the distribution, diversity, seasonality, metabolic potential of lipid biomarker producers, and to correct interpretations made on the presence of those lipid biomarkers in the sedimentary record. This information has been obtained by using complementary approaches targeting microbial molecules other than lipids, i.e., nucleic acids, such as DNA (deoxyribonucleic acid) or RNA (ribonucleic acid).

GENOMIC BIOMARKERS

The complete set of DNA of a (micro)organism is called its genome. A genome is composed of genes that encode for structural transcripts such as ribosomal RNA (rRNA) that functions as structural components of ribosomes, as well as functional genes that encode for messenger RNA (mRNA). The latter transcripts are translated within the ribosomes into proteins, notably enzymes involved in a wide scale of cellular metabolic processes. Some of these genes are present in most microbial taxa and play an important role in cellular functioning. Based on evolutionary differences in the nucleotide positions, sequencing analysis of these taxonomic marker genes can

identify the various members that are present in complex environmental microbial populations and track their evolutionary history. Here, we define those as 'genomic markers'. The gene encoding for 16S rRNA, the structural component of the small subunit of the prokaryotic ribosome, functions as a "molecular clock" that can be used to track when and how different organisms diverged. This is also the taxonomic marker gene most frequently used to identify bacterial and archaeal community members in environmental samples and to reconstruct phylogenies (i.e., history of the evolution of a species or group). The counterpart for eukaryotic organisms is the 18S rRNA gene. Apart from ribosomal RNA coding genes, other genes can be used as 'functional genes' to indicate the potential roles that the microorganisms carrying these genes play in important biogeochemical cycling processes. The majority of 'functional genes' also have been proven to be excellent taxonomic markers, and thus also help elucidate the composition and diversity of environmental microbial communities.

COMBINING rRNA/FUNCTIONAL GENES & LIPID BIOMARKERS

Nucleic acids (i.e., DNA, RNA) have a shorter preservation potential than lipid biomarkers based on their lower inherent stability and mechanisms of degradation (FIG. 1). RNA molecules can degrade in a matter of minutes or hours after cell death while DNA can be preserved for longer time scales, especially under rapid burial of sediment, low temperature, or lack of oxygen (as detailed below).

A major advantage of the use of lipid molecules as biomarkers is their great preservation potential (see FIG. 1, and Naeher et al., this volume). Nevertheless, they are limited in taxonomic resolution and can be synthesized by multiple sources, making it difficult to assign a certain lipid molecular structure to a specific microbial producer. In contrast, despite their lower preservation potential, 'genomic markers' are excellent taxonomic and/or functional markers, allowing the identification of member species of complex microbial communities, and addresses the question 'who is there and what are they potentially doing?'.

The first studies combining lipid and DNA-based biomarkers (i.e., the 16S rRNA gene) were done to better understand the relationship between microbial communities and their modern environments ('microbial ecology'). The lipid biomarkers used in those studies were phospholipid-derived fatty acids (PLFAs), which are chemotaxonomic markers of bacteria and eukaryotes. PLFAs can be used as biomarkers of microbial physiological status, viability, and to a certain degree (limited by the low molecular diversity of PLFAs), to elucidate microbial diversity. The fact that PLFAs rapidly degrade after cell death makes them excellent as indicators of viable (alive) microorganisms with an intact cell membrane. Nevertheless, this characteristic invalidates them as lipid biomarkers for paleo interpretations. In order to compensate for the low taxonomic potential of lipid biomarkers to identify potential biological sources, microbial ecology studies based on PLFAs were soon complemented with DNA-based biomarkers, i.e., 16S rRNA (e.g., Stephen et al., 1999). Sequencing analysis of the 16S rRNA gene is still widely used in combination with other lipid biomarkers that have longer preservation lifetimes than PLFAs. This combination approach has proven to be very useful to identify or confirm potential biological sources of a specific lipid biomarker, which can aid in the paleo interpretation of those biomarkers in studies of ancient materials in which DNA is not preserved. Nonetheless, conclusions derived from these combined DNA-lipid biomarker analyses should be taken with caution, as even in dynamic microbial systems the turnover time of these molecules can be different. Namely, DNA was shown to be more persistent (FIG. 1) after cell death than phospholipids and represents both living and non-viable microorganisms.

A next generation of multi-proxy studies combined analysis of lipid biomarkers and 16S rRNA gene sequencing with profiling of functional genes, elucidating not only taxonomic diversity (who is there?) but also the metabolic potential of the source organisms (what are they potentially doing?). As an example, members of the Thaumarchaeota (formerly known as marine crenarchaeota) were identified and taxonomically placed in the tree of life as a novel marine archaeal group based on 16S

rRNA gene analysis. Later, it was confirmed that they had a metabolism based on carbon fixation (i.e., chemolithoautotrophs) (Wuchter et al., 2003) and oxidation of ammonia (e.g. Könneke et al., 2005). Incubation studies of samples enriched in Thaumarchaeota and spiked with isotopically labeled bicarbonate confirmed that this archaeal group synthesizes archaeal tetraether lipids, glycerol dibiphytanyl glycerol tetraethers, GDGTs, with zero to 4 cyclopentane rings, and the specific thaumarchaeotal lipid biomarker crenarchaeol, which has 4 cyclopentane and one cyclohexane ring in its structure. These discoveries paved the way for the development of the sea-surface temperature (SST) proxy, TEX_{86}, based on the distribution of the GDGTs of Thaumarchaeota. Later studies have also confirmed that Thaumarchaeota carry the functional marker gene, *amo*A, which encodes for the ammonia monooxygenase enzyme that is involved in the first step of ammonia oxidation. The detection of this 'functional gene' in microbial ecology studies (along with 16S rRNA gene and lipid biomarkers) has allowed not only the detection of this microbial group and its diversity, but also the indication of their potential role within the global biogeochemical cycle of nitrogen.

However, the combined analysis of lipid biomarkers and profiling of environmental 16S rRNA- and functional genes does have inherent limitations: (i) one can only link all three markers to the same source organism in environmental samples if there is prior knowledge about the taxonomic and functional diversity as well as the lipid biomarker composition of related taxa (e.g., from cultivation experiments), (ii) the analysis of single taxonomic marker genes does not give a complete view of the metabolic or functional potential of the targeted microorganism , (iii) differences in preservation potential complicate direct comparisons of the DNA and lipid data.

LIPID BIOSYNTHETIC PATHWAYS

Some of these issues can now be partly addressed thanks to recent technological advances that have drastically reduced the cost of DNA sequencing along with the availability of very extensive genomic databases for data comparison, and the

continuing development of advanced bioinformatics and biostatistical toolsets for processing and the interpretation and visualization of big sequencing datasets. For example, the random sequencing of all the genes of all microorganisms present in an environmental DNA (shotgun metagenomes) and subsequent reassembly of these fragments into full genomes (i.e., Metagenome Assembled Genomes; MAGs) using advanced bioinformatics tools, makes it possible to determine the genetic pathways involved in the biosynthesis of a specific lipid biomarkers, and to identify the source organism(s) without the need for cultivation. Nevertheless, in most cases information is still lacking regarding the genes encoding for the enzymes involved in a specific lipid biomarker biosynthetic pathway. When this approach is applied to an unknown lipid biosynthetic pathway, the genome of the relevant microorganism(s) must be determined before we can identify 'candidate' genes that are potentially involved in that pathway. Once those candidate genes have been identified, based on their position within the genome or its resemblance to previously described biosynthetic processes, the activity of those genes must be confirmed. The activity of those candidate lipid biosynthetic genes can be confirmed by removing or incapacitating the candidate gene in the targeted microorganism (processes referred to as deletion, mutation, or 'genetic knock-out') and by observing a loss of function (e.g. the inability to make the lipid biomarker). Alternatively, or in cases where this is not possible, one can artificially incorporate and express the candidate gene in another microorganism, i.e., *Escherichia coli,* that is more easily subject to genetic manipulation, and then observe a 'gain of function' (new capacity to make the lipid biomarker; this method is referred to as 'heterologous gene expression') (FIG. 2). These approaches also have been successfully applied to determine enzymatic steps leading to specific lipid biomarkers.

Targeting lipid biosynthetic pathways of well-known lipid biomarkers

One of the most remarkable cases of the use of genomics to clarify the origin and diversity of biological sources of lipid biomarkers is that of hopanoid lipids (i.e., polycyclic terpenoids located in bacterial membranes and thought to be functional

analogues to eukaryotic sterols, are involved in the maintenance of the membrane stability). Hopanoid lipids are preserved as hopanes in the sedimentary record. The biosynthesis of these biomarkers requires the isoprenoid precursor, squalene, which is cyclized by a squalene–hopene cyclase encoded by the *shc* gene (FIG. 3). Hopanoids have been detected in a minor proportion of bacterial groups and the squalene–hopene cyclase encoding gene needed for hopanoid production is present in only approximately 10% of all bacteria (Pearson et al., 2007).

Among hopanoid biomarkers, 2-methylhopanes were initially considered to be biosynthesized by cyanobacteria and their presence in sediments as old as 2,500 Myr was used as an indication of the advent of oxygenic photosynthesis (Summons et al., 1999). Nevertheless, the identification and confirmation of the gene required for hopanoid C-2 methylation, and the fact that it was not generally found in all cyanobacterial genomes and was present in other bacterial taxa, invalidated the use of 2-methyhopanes as biomarkers of the appearance of oxygenic photosynthesis on Earth (Welander et al., 2010).

A similar case is that of the 3-methylhopanoids, which were generally attributed to aerobic methanotrophic bacteria. But the gene involved in the hopanoid C-3 methylation is found in genomes of microorganisms other than aerobic methanotrophic bacteria, suggesting 3-methylhopanoids cannot be used as biomarkers of aerobic methanotrophy in the past (Welander et al., 2012). The application of a 'genomic' approach for this lipid biomarker question was also helpful to understand the physiological role of 3-methylhopanoids, since the 'removal' (deletion) of the gene responsible for C-3 methylation from the obligate methanotroph *Methylococcus capsulatus* demonstrated that 3-methylhopanoids are required for the maintenance of the membranes and cell survival in stationary conditions.

Other lipid biomarkers with well-characterized lipid biosynthetic pathways are sterols, which perform essential functions in the membranes of eukaryotes. Sterols preserve as steranes in the geological record and have been used as biomarkers for

the appearance of eukaryotes and of the occurrence of oxygen on Earth (Brocks et al., 1999). The enzyme oxidosqualene cyclase (OSC) mediates the first committed step of sterol biosynthesis requiring oxygen (FIG. 3). Sterols also have been found to be produced by some bacterial groups, either by screening of pure cultures or inferred from the presence of bacterial OSC homologs in their genomes. An aerobic marine heterotrophic bacterium of the Bacteroidetes, *Eudoraea adriatica*, was predicted to harbor a bacterial OSC in its genome (Banta et al., 2017). Lipid analyses of *E. adriatica* and gene expression of its bacterial OSC in *E. coli* (FIG 2) confirmed that this strain makes two isoarborinol-like lipids (eudoraenol and adriaticol) directly from an oxidosqualene precursor. This is particularly relevant since arboranes, degradation products of isoarborinol, have often been used as indicators of terrestrial input into aquatic environments, as these biomarkers were initially thought to be solely biosynthesized and to originate from specific angiosperm vegetation. Banta et al. (2017) confirmed the possibility of bacterial sources of arborane, which was already suspected due to the occurrence of arborane biomarkers in Permian and Triassic sediments, which predates the origin of angiosperms.

Unraveling the evolutionary acquisition of lipid biomarkers

By targeting the lipid biomarker biosynthetic pathways, it is also possible to determine when a specific lipid biomarker was acquired through the course of evolution and to determine if that capacity has been transferred to other microbial groups by, for example, horizontal gene transfer. There are several examples in which the study of lipid biomarker biosynthetic pathways has led to the resolution of some evolutionary mysteries by the application of phylogenomics (analysis involving genomic data and evolutionary reconstruction).

For example, squalene synthase mediates the first committed step in the biosynthesis of sterols within the isoprenoid pathway, catalyzing the reaction from farnesyl-PP (FPP) to squalene (FIG. 3). The evolutionary story of squalene synthase has allowed us to decipher the possible paths of squalene formation. The eukaryotic photosynthetic green algae *Botryococcus braunii* race B accumulates triterpenes,

predominantly botryococcene. It has received attention as it is considered an ancient algal species dating back at least 500 Mya and is one of the few organisms known to have directly contributed to oil and coal shale deposits. Due to their similar structure, botryococcene biosynthesis was expected to resemble that of squalene. The genome of *B. braunii* harbors a typical squalene synthase (SS) and three squalene synthase-like (SSL) encoding genes. Enzymatic studies proved that a combined activity of SSL-1+SSL-3 leads to botryococcene, and of SSL1+SSL-2 to squalene (SQ) biosynthesis (Niehaus et al., 2011). It has been estimated that SQ and the three SSLs genes separated from a common ancestor ~500 Mya in a process of gene duplication and specialization. This enzymatic specialization is believed to regulate the flux of FPP into sterol biosynthesis independent from the flux into the biosynthesis of botryococcene, which accumulates in high amounts into the membrane of *B. braunii* to provide flotation and maximize its exposure to sunlight.

Biosynthetic pathways of lipid biomarkers used as paleoclimate proxies

The TEX_{86} SST proxy presents an example of the use of genomics to better constrain a paleoclimate proxy based on microbial lipid biomarkers. There is accumulating evidence that temperature is not the only variable affecting the Thaumarchaeota GDGT distribution (on which the TEX_{86} proxy is based; e.g., Qin et al., 2015; Hurley et al., 2016). It is also known that TEX_{86} does not reflect SSTs but rather subsurface temperatures, and that archaea in the deeper water column may also contribute to the signal of GDGTs in sediments, potentially leading to a warm bias in TEX_{86} SST estimates. In this regard, some studies indicated that deeper waters are characterized by a higher abundance of GDGT-2 relative to GDGT-3 and an increase in the crenarchaeol isomer (Taylor et al., 2013). This observation was further supported by the detection of different marine Thaumarchaeota groups inhabiting shallow and deep waters based on the use of the *amo*A ‘functional gene marker’ and of a gene coding for the enzyme involved in the first ether bond formation of the GDGT molecule (i.e., GGGP synthase; Villanueva et al., 2015). In addition, the TEX_{86} signal may also be affected by input from other archaea, which could potentially synthesize

GDGTs used in TEX_{86}. Previous studies have suggested a potential contribution to the GDGT signal by the marine euryarchaeota group II (MGII), which is present in oceanic surface waters. Because members of the MGII have not yet been cultivated, it is currently unknown which archaeal lipids they synthesize. Regardless, the lack of a homolog of the recently described GDGT cyclase encoding genes (Zeng et al., 2019; FIG. 3) in the MGII genomes suggests that this archaeal group does not synthesize GDGTs with cyclopentane rings, which are involved in the TEX_{86} paleotemperature proxy. Further analysis of lipid biosynthetic genes of MGII genomes also suggested this archaeal group is potentially unable to synthesize 'classical' archaeal membrane lipids (Villanueva et al., 2017), leaving unanswered questions about the diversity and evolution of membrane lipid acquisition in Archaea.

SEDIMENTARY ANCIENT DNA & PALEOMICROBIOLOGY

Lake and marine sediments are comprised of organic and inorganic deposits and represent rich archives of temporal changes in both planktonic (autochthonous) and terrestrial (allochthonous) ecosystem members. The majority of paleoecological studies involve the microscopic analysis of a relatively limited number of taxa that form or leave behind easily identifiable morphological remains in sediments. Lipid biomarkers can be preserved in the absence of morphological remains, but their taxonomic specificity is generally limited, as outlined above. A more recently developed approach targeting ancient DNA preserved in lake and marine sediments (i.e., sedimentary ancient DNA or *sed* aDNA) is rapidly increasing in popularity.

Despite the more limited preservation potential compared to lipid biomarkers and the need for stringent measures to prevent cross contamination with modern environmental DNA during sampling and sample processing in a dedicated clean lab facility, sequencing analysis of preserved *sed* aDNA markers offers unprecedented detail into the composition of past biota, from individuals to complex ecosystems, and across all domains of life, including taxa that do not leave behind any other diagnostic remains (e.g., Coolen et al., 2013). Paired with paleoenvironmental proxy data, *sed*

aDNA can be used to question how aquatic and terrestrial ecosystems have responded to natural vs. anthropogenic environmental perturbations (see Armbrecht et al. 2019, and references therein). Under ideal depositional and post-burial conditions (e.g. low temperatures, anoxia, and/or adsorption of extracellular DNA to clay minerals), past plankton communities have been identified at genus level from *sed*aDNA in ~270 Ka-old lake and ~ 1Ma-old marine sediments (e.g. Kirkpatrick et al., 2016).

Strategies used to study and authenticate sedimentary ancient DNA

Following the extraction and purification of sedimentary ancient DNA, the past planktonic community composition (e.g., chloroplast-containing algae) can be studied through sequencing analysis of PCR-amplified taxonomic marker genes (i.e., amplicon sequencing). Several bioinformatics 'pipelines' (data processing algorithms) have been developed to process the resulting sequence datasets to ultimately generate a matrix with the total counts of unique taxonomically assigned amplicon sequence variants (ASVs). These "species" x sample abundance matrices form the basis for biostatistical analysis to study past ecosystem responses to shifts in categorical (e.g., lithology type, climate- or environmental stages) or quantitative environmental changes (e.g., paleosalinity, paleotemperature, paleoprecipitation) given that suitable paleoenvironmental datasets have been generated in parallel (e.g., Coolen et al., 2013).

'Hybridization capture' is a method that offers great potential to selectively enrich for *sed* aDNA from, for example, planktonic taxa that can inform about past ecosystem responses to paleoenvironmental changes (Armbrecht et al., 2021). This approach involves the selective binding of marker genes to magnetic beads, followed by magnetic separation, clean-up, gene amplification, and sequencing.

Moreover, the metabolic properties of environmental microbial communities, such as the role they play in elemental cycling can be inferred from shotgun metagenomic sequencing as mentioned earlier. Mostly applied to modern environments, this

approach was recently used to reconstruct microbial populations and genes involved in nitrogen cycling during long-term expansion of the Oxygen Minimum Zone (OMZ) in the Arabian Sea over the last 43,000 years (Orsi et al., 2017), and to reconstruct microbial communities that were involved in the sulfur cycle in the Black Sea during the Holocene (More et al., 2019). Importantly, shotgun metagenomic datasets can be analyzed using sophisticated bioinformatics tools such as Mapdamage2 (Jónsson et al., 2013) to differentiate between altered, post-mortem ancient microbial DNA vs. intact DNA from sedimentary bacteria that are shaped by modern *in situ* environmental conditions (e.g., the availability of electron acceptors and donors as well as sediment porosity). By way of another example to authenticate ancient sedimentary microbial DNA, bioinformatic tools have recently been developed to calculate the growth rate of environmental bacteria by measuring genome replication rates from shotgun metagenomic data (Emiola et al., 2018).

Examples of combined stratigraphic analysis of lipid biomarkers and sed aDNA

The first study to combine lipid biomarker and *sed* aDNA data was performed on Holocene sediments of Lake Mahoney, Canada (Coolen and Overmann, 1998). The water column of this permanently stratified saline lake contains an oxygenated photic zone and higher density anoxic and sulfidic bottom water. A dense population of purple sulfur bacteria (PSB) occurs at the transition between both waters (i.e., within the chemocline) where both light and toxic hydrogen sulfide (H_2S) is available for PSB to perform anoxygenic photosynthesis. The concomitant presence of the specific carotenoid okenone as well as 16S rRNA genes of PSB in the sedimentary record of Lake Mahoney revealed when the lake was stratified and experienced periods of anoxia, and that *sed* aDNA can be preserved and analyzed to identify the source of lipid biomarkers covering at least the last 10,000 years (Coolen and Overmann, 1998).

Subsequently, the combined presence of *sed* aDNA and fossil carotenoids from past chemocline-dwelling green sulfur bacteria (i.e., chlorobactene and/or isorenieratene

instead of okenone) indicative of permanently stratified and anoxic conditions was used to reconstruct orbitally controlled changes in paleoproductivity that resulted in the deposition of organic-rich sediment intervals (i.e., sapropels) in the eastern Mediterranean Sea (Coolen and Overmann, 2007), to reconstruct environmental changes in the Black Sea associated with the postglacial reconnection with the Mediterranean Sea via the Bosporus Strait (Manske et al., 2008), vs. the postglacial isolation of the brackish Antarctic Ace Lake from the Southern Ocean (Coolen et al., 2004).

Over the last three decades, it has been shown that the level of unsaturation in long-chain alkenones produced by the calcifying haptophyte alga *Emiliania huxleyi*, a cosmopolitan oceanic plankton, is linearly related to the growth temperature. Downcore changes in the alkenone composition were used to develop the so-called UK' 37 proxy for past sea surface temperatures (SST; Brassell et al., 1986). This index was initially developed and calibrated for *E. huxleyi*. However, parallel analysis of *sed* aDNA and alkenones revealed novel groups of non-calcifying "naked" haptophytes as invisible sources of fossil alkenones in the sedimentary records of the brackish Black Sea and Ace Lake, where the original alkenone SST proxies that were calibrated for *E. huxleyi* revealed spurious results (Coolen et al., 2004; 2013). These studies then resulted in the development of novel calibrated alkenone-SST indices to accurately reconstruct paleo SST from terrestrial settings (e.g., Huang et al., 2021).

By way of a final example, dinoflagellates are important plankton members and are traditionally identified from the sedimentary record based on microscopic analysis of calcareous or organic-walled cysts, or from the presence of the lipid biomarker dinosterol. However, 80% of dinoflagellate species produce cysts and dinosterol is only known from a relatively small number of species. In Holocene sediments of Ellis Fjord near Ace Lake, *sed a*DNA analysis revealed late-Holocene changes in the relative abundance of the sea ice dinoflagellate *Polarella glacialis*, indicative of Holocene changes in the extent of sea ice formation in this East Antarctic region. Other proxies failed to reproduce this result since *P. glacialis* does not produce the lipid biomarker

dinosterol and its small organic-walled cysts escaped microscopic identification (Boere et al., 2009).

Taken together, the examples summarized above show that the parallel comparison of lipid biomarkers and *sed a*DNA profiles can be used to cross-validate the reliability of both approaches, to identify sources of lipid biomarkers, and to refine the interpretation of paleoenvironmental reconstructions.

FUTURE CHALLENGES

Here, we have reviewed the studies that prove the usefulness of combining DNA and lipid biomarker analysis to better understand the microbial sources of these compounds and their physiology, in both present and past environments. Nonetheless, the definite confirmation of microbial sources of lipid biomarker will depend on further methodological developments. For example, prior knowledge of physiological properties inferred from genome sequencing would aid in the targeted development of cultivation methods and facilitate the confirmation of the lipid composition of yet-uncultured microorganisms. Further physiological studies in the lab also would be useful to determine under which conditions those lipids are synthesized. The improvement of computational methods will be key to predict lipid biosynthetic genes, and combined with novel gene editing tools (i.e., CRISPR), would aid in the confirmation of the biosynthetic pathways that produce specific lipid biomarkers. Improvement of evolutionary reconstruction methods based on genomic data will also contribute to a better understanding of how and when certain lipid biomarkers were acquired. The study of *sed* aDNA will also benefit from the development of novel tools to lower the detection limit of *sed a*DNA and to explore the preservation potential of *sed a*DNA in sedimentary records that have been deposited under less favorable conditions, such as long-term exposure to warm temperatures in understudied tropical lakes. The stratigraphic analysis of *sed a*DNA has the potential to provide a unique opportunity to reconstruct ecosystem biodiversity across all domains of life and may provide a holistic overview of past

biotic interactions and ecosystem functioning and to reconstruct the role they have played in biogeochemical cycling processes within the overlying water column at the time of deposition. Studying the latter requires the development of experimental and/or bioinformatics approaches to accurately distinguish between members of the indigenous sedimentary microbiome, which are actively shaped by *in situ* environmental conditions, vs. ancient microbial sources (i.e., the paleomicrobiome) of sedimentary DNA that originate from the overlying water column and have been shaped by environmental conditions that prevailed at the time of deposition. A further challenge is to successfully recover ancient genomes and compare them with genomes from closely related modern taxa to reconstruct past evolutionary events that led to present-day taxonomic diversity and physiological properties of planktonic ecosystems.

REFERENCES

Armbrecht, L. H., Coolen, M. J. L., Lejzerowicz, F., George, S. C., Negandhi, K., Suzuki, Y., Young, J., Foster, N. R., Armand, L. K., Cooper, A., Ostrowski, M., Focardi, A., Stat, M., Moreau, J. W., and Weyrich, L. S., 2019, Ancient DNA from marine sediments: Precautions and considerations for seafloor coring, sample handling and data generation: Earth-Science Reviews, v. 196.

Armbrecht L, Hallegraeff G, Bolsch CJS, Woodward C, Cooper C (2021) Hybridization capture allows DNA damage analysis of ancient marine eukaryotes. Scientific Report 11(3230): doi.org/10.1038/s41598-021-82578-6.

Banta AB, Wei JH, Gill CC, Giner J, Welander PV (2017) Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. Proceedings of the National Academy of Sciences of the United States of America 114: 245-250

Boere AC, Abbas B, Rijpstra WIC, Versteegh GJM, Volkman JK, Sinninghe Damsté JS, Coolen MJL (2009) Late-Holocene succession of dinoflagellates in an Antarctic fjord using a multi-proxy approach: paleoenvironmental genomics, lipid biomarkers and palynomorphs. Geobiology 7: 265-281

Brocks JJ, Logan GA, Buick R, Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. Science 285: 1033-1036

Coolen MJL, Muyzer G, Rijpstra WIC, Schouten S, Volkman JK, Sinninghe Damsté JS (2004) Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic Lake. Earth and Planetary Science Letters 223: 1-2

Coolen MJL, Orsi WD, Balkema C, Quince C, Harris K, Sylva SP, Filipova-Marinova M, Giosan L (2013) Evolution of the plankton paleome in the Black Sea from the Deglacial to Anthropocene. Proceedings of the National Academy of Sciences of the United States of America 110: 8609-8614

Coolen MJL, Overmann J (1998) Analysis of subfossil molecular remains of purple sulfur bacteria in a lake sediment: Applied and Environmental Microbiology 64: 4513-4521

Emiola A, Oh J (2018) High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. Nature Communications 9: 4956

Hurley SJ, Elling FJ, Könneke M, Suchwald C, Wankel SC, Santoro A, Lipp JS, Hinrichs K-U, Pearson A (2016) Influence of ammonia oxidation rate on thaumarchaeal lipid composition and the TEX_{86} temperature proxy. Proceedings of the National Academy of Sciences of the United States of America 113: 7762–7767

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters: Bioinformatics 29: 1682-1684

Kirkpatrick, J.B.; Walsh, E.A.; D'Hondt, S. Fossil DNA Persistence and Decay in Marine Sediment over Hundred-Thousand-Year to Million-Year Time Scales. Geology 2016, 44, 615–618, doi:10.1130/G37933.1.

Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437: 543-546

Lincoln SA, Wai B, Eppley JM, Church MJ, Summons RE, DeLong EF (2014) Marine Group II planktonic Euryarchaeota are significant contributors to tetraether lipids in the ocean. Proceedings of the National Academy of Sciences of the United States of America 111: 9858-9863

Manske AK, Henssge U, Glaeser J, Overmann J (2008) Subfossil 16S rRNA gene sequences of green sulfur bacteria in the Black Sea and their implications for past photic zone anoxia. Applied and Environmental Microbiology 74: 624-632

More KD, Giosan L, Grice K, Coolen MJL (2019) Holocene paleodepositional changes reflected in the sedimentary microbiome of the Black Sea. Geobiology 00:1-13. doi.org/10.1111/gbi.12338.

Niehaus TD, Okada S, Devarenne TP, Watt DS, Sviripa V, Chappell J (2011) Identification of unique mechanisms for triterpene biosynthesis in Botryococcus braunii. Proceedings of the National Academy of Sciences of the United States of America 108:12260-12265.

Orsi WD, Coolen MJL, Wuchter C, He LJ, More KD, Irigoien X, Chust G, Johnson C, Hemingway JD, Lee M, Galy V, Giosan L (2017) Climate oscillations reflected within the microbiome of Arabian Sea sediments. Scientific Reports 7: 6040

Pearson A, Flood Page SR, Jorgenson TL, Fischer WW, Higgins MB (2007) Novel hopanoid cyclases from the environment. Environmental Microbiology 9: 2175-2188

Qin W, Carlson LT, Armbrust EV, Devol AH, Moffett JW, Stahl DA, Ingalls AE (2015) Confounding effects of oxygen and temperature on the TEX_{86} signature of marine Thaumarchaeota. Proceedings of the National Academy of Sciences of the United States of America 112: 10979-10984

Stephen JR, Chang YJ, Gan YD, Peacock A, Pfiffner SM, Barcelona MJ, White DC, MacNaughton SJ (1999) Microbial characterisation of a JP-4 fuel contaminated site using a combined lipid biomarker/PCR-DGGE based approach. Environmental Microbiology 1: 231-243

Summons RE, Jahnke LL, Hope JM, Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature 400: 554-557

Taylor KWR, Huber M, Hollis CJ, Hernandez-Sanchez MT, Pancost RD (2013) Re-evaluating modern and Paleogene GDGT distributions: Implications for SST reconstructions. Global Planetary Change 108: 158-174

Villanueva L, Schouten S, Sinninghe Damsté JS (2015) Depth-related distribution of a key gene of the tetraether lipid biosynthetic pathway in marine Thaumarchaeota. Environmental Microbiology 17: 3527-3539

Villanueva L, Schouten S, Sinninghe Damsté JS (2017) Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. Environmental Microbiology 19: 54-69

Welander PV, Coleman ML, Sessions AL, Summons RE, Newman DK (2010) Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. Proceedings of the National Academy of Sciences of the United States of America 107: 8537-8542

Welander PV, Summons RE (2012) Discovery, taxonomic distribution, and phenotypic characterization of a gene required for 3-methylhopanoid production. Proceedings of the National Academy of Sciences of the United States of America 109: 12905-12910

Wuchter C, Schouten S, Boschker HTS, Sinninghe Damste JS (2003) Bicarbonate uptake by marine Crenarchaeota: FEMS Microbiology Letters 219: 203-207

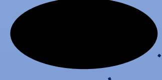
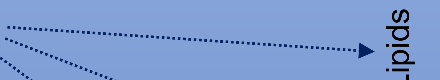
Zeng Z, Liu XL, Farley KR, Wei JH, Metcalf WW, Summons RE, Welander PV (2019) GDGT cyclization proteins identify the dominant archaeal sources of tetraether lipids in the ocean. Proceedings of the National Academy of Sciences of the United States of America 116: 22505-22511

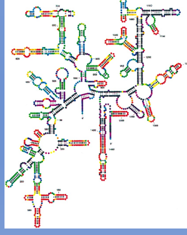
FIGURE LEGENDS

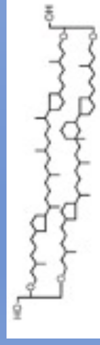
FIGURE 1. Preservation potential or turnover time of different microbial biosignatures or biomarkers including messenger RNA (mRNA), ribosomal RNA (rRNA), DNA and lipids (core or intact polar lipids). The yellow bars represent environmental lifetimes (deposition + diagenesis time) in a log-based time axis not drawn to scale.

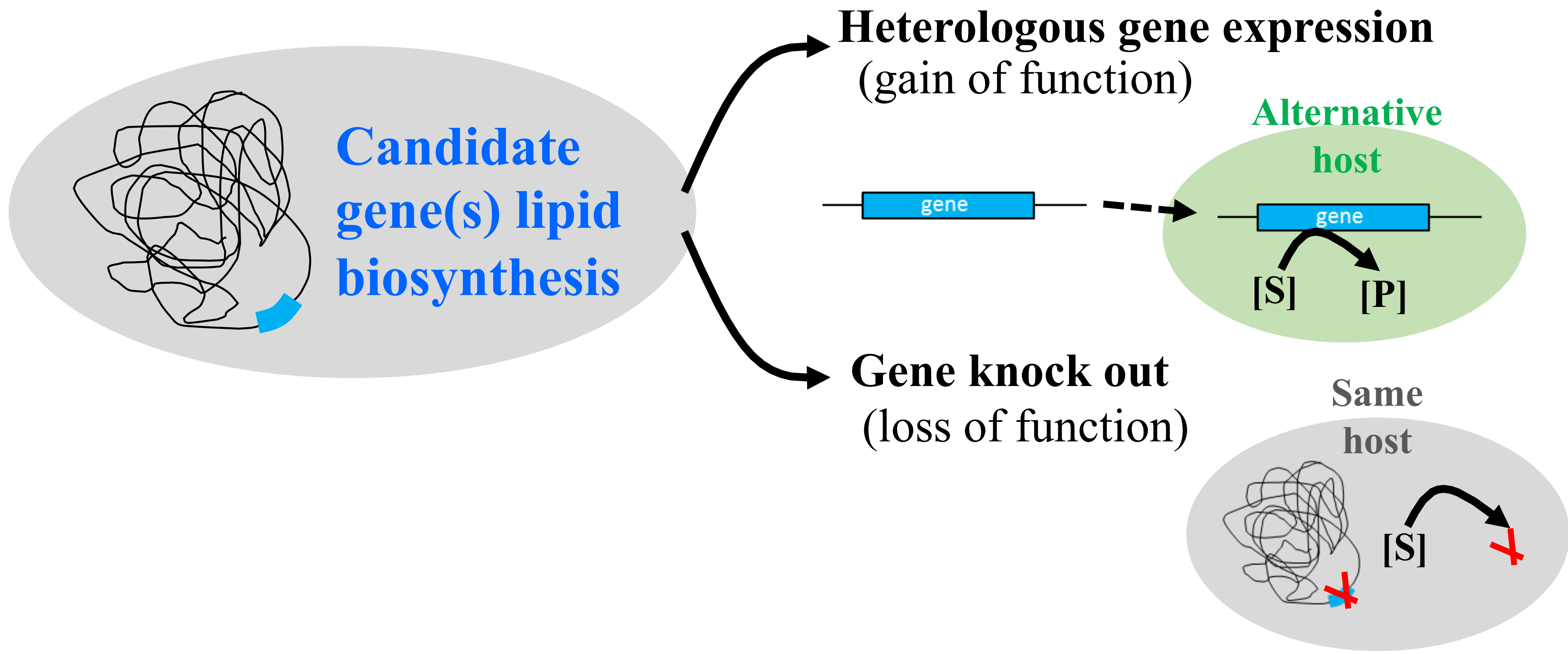
FIGURE 2. Methodological options for the confirmation of the activity of candidate genes involved in a specific lipid biosynthetic pathway. Heterologous gene expression involves the addition of a candidate gene to an alternative microbial host that is genetically more treatable (e.g. *E. coli*), in which the expression of the gene will lead to a gain of function. By performing a gene knock-out, the targeted gene is deleted from the original microorganism (same host) and loss of the function or formation of the lipid or intermediate should be lost. S, substrate; P, product of the enzymatic reaction.

FIGURE 3. Overview of the lipid biosynthetic pathways discussed in the text and key enzymes. DMAPP, dimethylallyl diphosphate GGPP, geranylgeranyl diphosphate; GGGP synthase geranylgeranylglyceryl phosphate synthase; DGGGP, digeranylgeranylglyceryl phosphate synthase, G1P, glycerol-1-phosphate; GDGT, glycerol dialkyl glycerol tetraether.

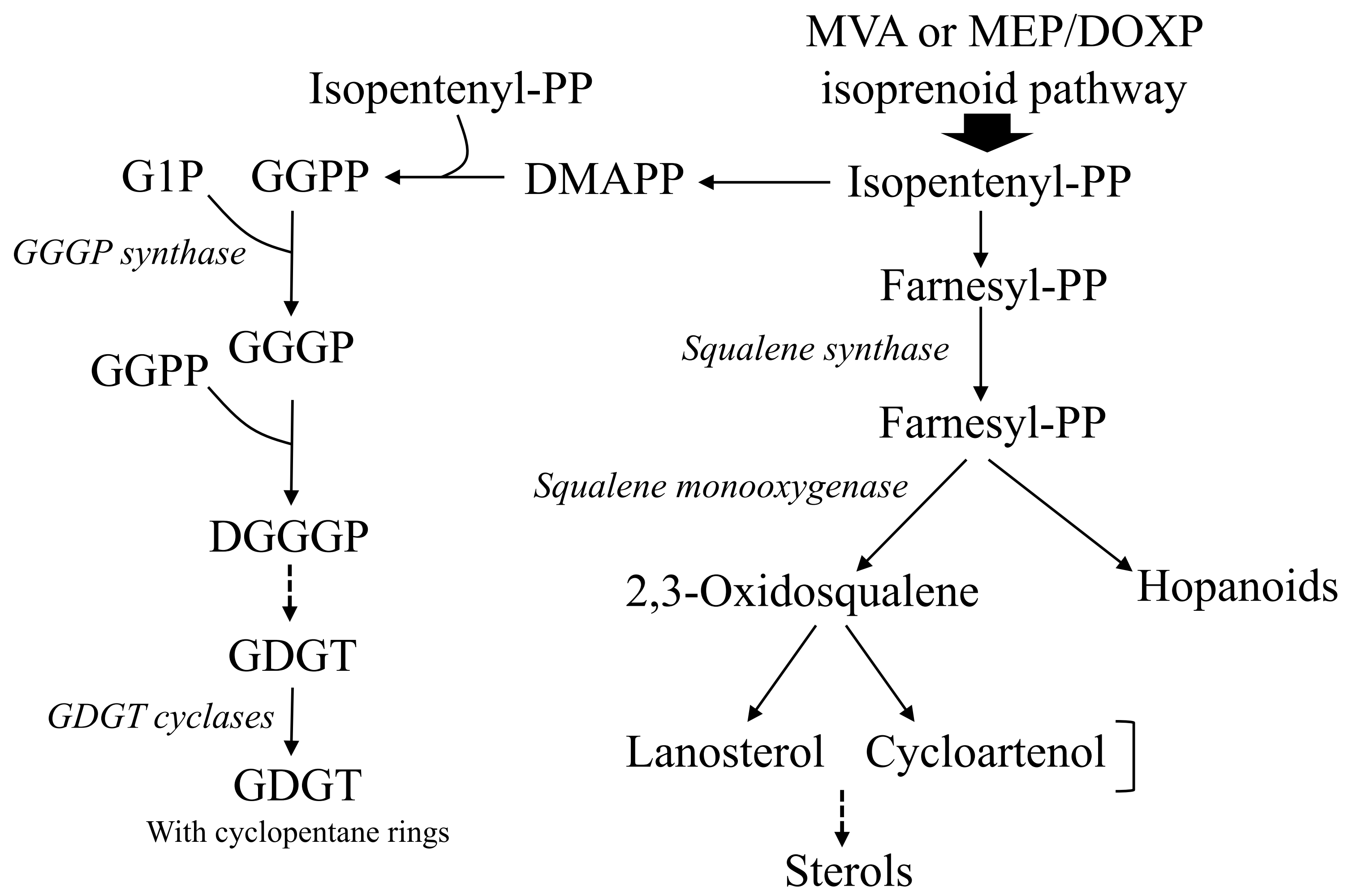






Microorganism
Lipids
mRNA
rRNA
DNA
BIOSIGNATURES
hours
Days?
Kyr?
Myr?
Time: deposition + diagenesis



Candidate gene(s) lipid biosynthesis
Heterologous gene expression
(gain of function)
Alternative host
gene
gene
[S]
[P]
Gene knock out
(loss of function)
Same host
[S]


MVA or MEP/DOXP
isoprenoid pathway
Isopentenyl-PP
Isopentenyl-PP
DMAPP
G1P
GGPP
GGGP synthase
GGGP
GGPP
DGGGP
GDGT
GDGT cyclases
GDGT
With cyclopentane rings
Farnesyl-PP
Squalene synthase
Farnesyl-PP
Squalene monooxygenase
2,3-Oxidosqualene
Hopanoids
Lanosterol
Cycloartenol
Sterols