# Hopanoid-producing bacteria in the Red Sea include the major marine nitrite-oxidizers

# Hopanoid-producing bacteria in the Red Sea include the major marine nitrite-oxidizers

Authors:

Kharbush, Jenan J.

Thompson, Luke R.

Haroon, Mohamed Fauzi.

Knight, Rob

Aluwihare, Lihini I.

## ABSTRACT

Hopanoids, including the extended side chain-containing bacteriohopanepolyols (BHPs), are bacterial lipids found abundantly in the geological record and across Earth's surface environments. However, the physiological roles of this biomarker remain uncertain, limiting interpretation of their presence in current and past environments. Recent work investigating the diversity and distribution of hopanoid producers in the marine environment implicated low-oxygen regions as important loci of hopanoid production, and data from marine oxygen minimum zones (OMZs) suggested that the dominant hopanoid producers in these environments are nitrite-utilizing organisms, revealing a potential connection between hopanoid production and the marine nitrogen cycle. Here we use metagenomic data from the Red Sea to investigate the ecology of hopanoid producers in an environmental setting that is biogeochemically distinct from those investigated previously. The distributions of hopanoid production and nitrite oxidation genes in the Red Sea are closely correlated, and the majority of hopanoid producers are taxonomically affiliated with the major marine nitrite oxidizers, *Nitrospinae* and *Nitrospirae*. These results suggest that the relationship between hopanoid production and nitrite

oxidation is conserved across varying biogeochemical conditions in dark ocean microbial ecosystems.

Keywords: hopanoids, microbial ecology, Red Sea, Nitrospina, nitrite oxidizing bacteria, metagenomics

## INTRODUCTION

Hopanoids are pentacyclic triterpenoid lipids produced by many bacteria as cell membrane components, and represent one of the most important lipid biomarker classes used to infer microbial activity in the fossil record (Ourisson and Albrecht 1992; Brocks *et al.* 1999). Interpreting the distribution of hopanoids to understand the current and past biosphere requires an understanding of hopanoid function in bacteria and the environmental conditions that stimulate hopanoid production. The physiological functions of hopanoids are hypothesized to be related to membrane stabilization, similar to the role of sterols in eukaryotes (Sáenz *et al.* 2012a). For example, hopanoids were shown to play a role in pH tolerance and membrane integrity (Welander *et al.* 2009), lipid raft formation (Sáenz 2010) and antibiotic resistance (Schmerk *et al.* 2011). Hopanoids are also important in several groups of marine nitrogen-fixing cyanobacteria (Sáenz *et al.* 2012b) and in nutrient stress responses of akinete-forming cyanobacteria (Ricci *et al.* 2017). However, the vast majority of marine hopanoid producers have not been cultured, and whether hopanoids have similar functions in these organisms to the cultured minority is unknown.

The distribution and structural diversity of the extended side chain-containing bacteriohopanepolyols (BHPs) has been studied in various marine environments to link the sedimentary inventory of hopanoids with their source organisms in the marine water column.

Several studies showed that BHP production is higher at low oxygen ($O_2$) concentrations. Specifically, BHP concentration and structural diversity both increase as $O_2$ concentration decreases, and some BHP structures are specific to bacteria in low-$O_2$ marine environments (Sáenz *et al.* 2011; Kharbush *et al.* 2013; Rush *et al.* 2014). Nevertheless, the seemingly cosmopolitan nature of hopanoid structures identified in water column samples provides little insight into the ecology and metabolic characteristics of hopanoid-producing bacteria.

More recently, insight into the potential environmental niches and metabolic strategies of hopanoid producers was obtained by examining the genetic diversity of squalene hopene cyclase (*sqhC* and SHC for the gene and protein, respectively), the gene primarily responsible for hopanoid biosynthesis. PCR-based surveys of marine environments revealed that *sqhC* is phylogenetically widespread (Pearson *et al.* 2007, 2009, Kharbush *et al.* 2013, 2015a) but uncommon, occurring in only 5-10% of bacterial species (Pearson and Rusch 2009). Furthermore, the phylogeny generated using both newly amplified and previously archived *sqhC* sequences implied that hopanoid production is associated with unusual metabolic strategies that are found in a limited number of modern environments and in many cases are associated with low-$O_2$ conditions (Kharbush *et al.* 2013). Interrogating metagenomic datasets from low-$O_2$ regions of the Eastern Pacific revealed that the relative abundance of *sqhC* was positively correlated with nitrite concentrations and that the majority of hopanoid producers in these environments are nitrite-utilizing organisms (Kharbush *et al.* 2015a).

The metabolic capabilities of the dominant hopanoid producers in the low-$O_2$ environments examined so far reveal previously unrecognized connections with suboxic processes of the marine nitrogen (N) cycle, and suggest a potential application of hopanoids as tracers of low-$O_2$ conditions in the water column. It remains to be determined whether similar patterns can be found in other regions of the modern ocean, including in environments where $O_2$

concentrations are higher than in the OMZs already studied. In this study, we utilized a recently generated metagenomic dataset from the Red Sea (Haroon *et al.* 2016; Thompson *et al.* 2017) to explore the diversity and ecology of hopanoid producers in an environment distinct from any examined previously. This dataset is particularly valuable because the metagenomic data are accompanied by a comprehensive collection of metadata that facilitated an examination of correlations between *sqhC* genes and environmental variables. The results illuminate the phylogenetic and physical distribution of hopanoid producers in the relatively well-oxygenated environment of this oligotrophic, semi-enclosed basin.

## MATERIALS AND METHODS

The metagenomic dataset was generated as part of the 2011 KAUST Red Sea Expedition (KRSE) in late summer (September-October) of 2011, as described previously (Thompson *et al.* 2017). Briefly, samples were collected from eight stations (locations shown in **Figure 1A**) from depths of 10, 25, 50, 100, 200, and 500 m, except in cases where the seafloor was shallower than 500 m, in which case the deepest sample was taken just above the seafloor. DNA was extracted from filters containing the 0.1-1.2 μm microbial size fraction, and whole-genome shotgun (WGS) libraries constructed, using Illumina HiSeq 2000 paired-end (2 x 100 bp) sequencing, yielding about 10 million reads per sample for the 45 samples.

The 45 metagenomes were analyzed for the relative abundance of gene families and biochemical pathways using HUMAnN (Abubucker *et al.* 2012) as described by Thompson *et al.* (2017). KEGG Orthology (KO) relative abundances—normalized to total mapped reads—were contained in the file RedSea_KORelativeAbundance_AllTaxa.csv downloaded from https://github.com/cuttlefishh/papers/tree/master/red-sea-spatial-series/data/humann1.

To explore depth-based clustering of samples and correlations between environmental metadata and gene abundances, the data matrix was visualized with canonical correspondence analysis (CCA) using the vegan 2.3-1 package in R, implemented according to Legendre and Legendre (2012). Vector loadings of key KOs of interest were plotted in CCA space.

For insight into the taxonomic affiliations of hopanoid producers, metagenomic hits corresponding to the *sqhC* gene were extracted using its unique KEGG identifier. However, previous studies have shown that very few *sqhC* sequences amplified from the marine environment can be reliably taxonomically classified using cultured organisms in databases like KEGG (Pearson *et al.* 2007, 2009; Kharbush *et al.* 2013). Therefore, to increase taxonomic resolution and to compare to previously-amplified marine *sqhC* sequences, metagenomic reads mapping to the *sqhC* KO (K06045) were assigned taxonomically by finding the best match in a second custom database containing these environmental *sqhC* sequences as well as those found in KEGG genomes. The search was done using usearch v7.0.1001 (http://drive5.com) with ublast and default parameters. Phylogenetic and/or metabolic groups were then inferred using a previously published phylogeny of *sqhC* (Kharbush *et al.* 2013). Genetic diversity patterns of *sqhC* metagenomic data were visualized using the Circos Table Viewer (http://mkweb.bcgsc.ca/tableviewer/visualize/) (Krzywinski *et al.* 2009).

To explore the phylogenetic context of *Nitrospin*a found in the Red Sea, trimmed metagenomic sequences were mapped against representative *Nitrospina* 16S rRNA gene sequences (*n*=482) downloaded from the NCBI database (accessed on Feb 2016), using BWA-MEM v0.7.12 (Li 2013). Mapped reads were extracted using Samtools v1.2 (Li *et al.* 2009) and subsequently assembled using Spades v3.8.0 with options "--meta --only_assembler" (Nurk *et al.* 2013). Assembled sequences were searched against the SILVA incremental aligner database online v1.2.11 (https://www.arb-silva.de/aligner/ accessed on June 2016, (Pruesse *et al.*, 2012).

All assembled *Nitrospina* hits ≥ 200 bp in length were used for downstream tree construction.

Near full-length (≥ 1400bp) *Nitrospina* 16S rRNA gene sequences (*n*=46) were aligned using SSU-ALIGN v0.1.1 (Nawrocki 2009) and a 16S RNA gene tree was constructed using FastTree v2.1.7, using default parameters with a generalized time-reversible model (Price *et al.* 2010). 100 resampled alignments were generated using the PHYLIP SEQBOOT module (Felsenstein 1989), and a script (CompareToBootstrap.pl) included in the FastTree package was used to compare the original tree to the resampled trees and generate bootstrap values. The representative tree was visualized in ARB, and the mapped *Nitrospina* 16S rRNA sequences assembled from the Red Sea metagenomes were placed into the tree by parsimony insertion in ARB (Ludwig *et al.* 2004). The final tree was exported and cleaned up using Adobe Illustrator.

## RESULTS

### Physicochemical parameters of the Red Sea water column

The 2011 KRSE cruise collected depth profile samples for metagenomics analysis from eight stations spanning from Station 12 near the Bab Al-Mandab strait (BAM) to Station 192 at the northern end of the sea (**Figure 1A**). The thermohaline circulation pattern of the Red Sea is readily apparent in the nitrate and $O_2$ concentrations along the cruise track (**Figure 1B**). The introduction of nutrients to the Red Sea is limited to a seasonal intrusion of Gulf of Aden Intermediate Water (GAIW) in the summer months of June-September (Souvermezoglou *et al.* 1989; Smeed 1997). This relatively less oxygenated, nitrate-rich GAIW can be seen around Stations 12 and 34, as it enters the Red Sea above the sill at BAM and mixes into the surface waters. The southward flow of saline Red Sea Water is visible as an even more nitrate-rich and oxygen-deficient water mass below 400 m depth, flowing back toward BAM. Notably,

however, $O_2$ levels do not fall below 0.6 mL/L (27 μM) and therefore do not even reach suboxia (1-20 μM, as defined by Wright et al., 2012).

**Distribution and abundance of *sqhC* gene family**

A search of the metagenomes for hits to the *sqhC* gene family (K06045, hereafter referred to as *sqhC*) demonstrated that very few hopanoid producers are found in samples taken from the upper 50 m (**Figure 2A**), in agreement with previous studies (Pearson and Rusch 2009; Kharbush *et al.* 2013, 2015a). Exceptions to this trend include two 50-m samples from Station 12 and 34 that are influenced by inflowing GAIW and therefore contain both lower $O_2$ concentrations and distinct taxonomic composition (Thompson *et al.* 2017) relative to samples from comparable depths at other stations. Two additional samples from Station 34 from depths of 100 and 258 m show a similar shift in $O_2$ and *sqhC* abundance (**Figure 2A**). At 258 m, however, it is unclear if this is a result of the GAIW mixing down into the water column, the return flow of Red Sea Water at depth, or exchange with sediment pore water (depth profiles shown in **Supplementary Figure S1**). Because GAIW intrusion is an annual occurrence that likely influences the biogeochemistry and microbial ecology of the Red Sea, we elected to retain GAIW-influenced samples in subsequent analyses.

Canonical correlation analysis (CCA) including all available metadata showed that *sqhC* abundance is correlated with depth and nutrients, and anti-correlated with oxygen. The abundance of *sqhC* increases with depth and with decreased $O_2$ concentrations, with the majority of *sqhC* hits found at depths beyond 100 m and below an $O_2$ concentration of approximately 3.75 mL/L (**Figure 2A**). A similar pattern was observed for the genes encoding the subunits of the enzyme responsible for nitrite oxidation, nitrite oxidoreductase (*nxrA* and *nxrB,* **Figure 2B,C**). This implies a connection between nitrite oxidation and hopanoid

production in this environment- in particular that these nitrite oxidizers and hopanoid producers could be the same organisms.

To validate this connection we initially looked at the assembled metagenomic contigs, but were unsuccessful in finding both *sqhC* and *nxr* on the same contigs. Therefore we next compared the phylogenetic associations of the metagenomic hits to each gene. Although depth and $O_2$ co-vary in the Red Sea, because previous work suggested a connection between the metabolic strategies of hopanoid producers and N cycling under low-$O_2$ conditions (Kharbush *et al.* 2013, 2015a), *sqhC* and *nxr* diversity was evaluated with respect to $O_2$ concentrations. Based on the distribution shown in Figure 2A, we separated the samples into two groups for analysis based on whether the $O_2$ concentration fell above or below a threshold of 3.75 mL/L, hereafter referred to as oxygenated and low-$O_2$ samples, respectively.

In total, the low-$O_2$ samples (including the GAIW samples) contained approximately three times as many *sqhC* sequences as the oxygenated samples. The taxonomic distribution of *sqhC* hits in low-$O_2$ samples, organized by sampling station in addition to taxonomic group, shows that the most common hopanoid-producing bacteria in the Red Sea are various groups of *Proteobacteria* and the two major marine nitrite-oxidizing bacteria (NOB), *Nitrospina* and *Nitrospira* (**Figure 3**). Other groups such as *Planctomycetes*, *Actinobacteria*, and *Bacillus*-like organisms contributed smaller numbers of *sqhC* hits, while the "Unassigned" category contains sequences that could not be reliably assigned to any group. In comparison, the taxonomic distribution of *sqhC* hits in the oxygenated samples shows higher numbers of *Proteobacteria*-like *sqhC,* increasing from about 30% of the total hopanoid-producing community in low-oxygen samples to over 50% in oxygenated samples **(**see **Supplementary Figure S2)**. It is important to note that because genome databases are limited by available genomes, some reads might match better to genomes not contained in current databases, and therefore be wrongly assigned in the current

analysis. However, overall the distribution reinforces earlier observations that implicated *Proteobacteria* and NOB as the major groups contributing to hopanoid production in marine environments, and suggests that *Proteobacteria*-like hopanoid producers are more prevalent in oxygenated samples compared to NOB hopanoid producers, which apparently dominate at lower $O_2$ concentrations in the Red Sea.

**Distribution of *nxr* gene family**

The taxonomic distribution of *nxrA* and *nxrB* shows that the majority of hits in low $O_2$ samples are affiliated with *Nitrospina* or *Nitrospira*, as well as a significant number of hits in the unassigned category (see **Supplementary Figure S3)**. The overlapping phylogenetic affiliations of *sqhC* and *nxr* thus provide indirect evidence that a large fraction of hopanoid producers in the dark, low $O_2$ regions of the Red Sea are also nitrite oxidizers.

Among NOB, *Nitrospinae* in particular is recognized as a widespread, ecologically significant phylum in marine environments (Beman *et al.* 2013; Levipan *et al.* 2014; Spieck *et al.* 2014), that consists of multiple genetically and ecologically distinct lineages (Ngugi *et al.* 2016). In fact, *Nitrospina* were recently recognized as major contributors to dark ocean chemoautotrophic carbon fixation- contributing up to 45% of total inorganic carbon fixed in the mesopelagic North Atlantic Ocean (Pachiadaki *et al.* 2017). If this group of organisms also contributes hopanoids to sediments, it could greatly influence the interpretation of hopanoids in the geological record. However, because *Nitrospinae* is such a large phylum it is also possible that not all members are hopanoid-producers.

To compare the *Nitrospina*-like hopanoid producers found in the Red Sea water column with *Nitrospina* from other regions, a tree was constructed using full-length and partial *Nitrospina* 16S sequences (**Figure 4**). With a few exceptions, the resulting phylogeny is very similar to that

published by Ngugi *et. al*. However, unlike the single-cell-amplified genomes (SAGs) from the Atlantis II deep BSI ("Ngugi AAA799-A02 and "Ngugi AAA799-C22", **Figure 4**), the Red Sea water column sequences fall within the uncultured Clade 2, rather than Clade 1, the recently proposed *Candidatus* Nitromaritima lineage (Ngugi *et al.* 2016). Also included in Clade 2 are most of the partial 16S rRNA sequences obtained from OMZs in the ETSP (Levipan *et al.* 2014) and ETNP (Beman *et al.* 2013). In contrast, all sequences from the NESAP OMZ fall within Clade 1.

## DISCUSSION

This study used metagenomics data to identify the phylogenetic associations and potential ecology of hopanoid producers in the Red Sea. It is important to note, however, that the presence of a gene alone does not predict activity. In the case of *sqhC*, the majority of cultured and sequenced bacteria that contain a functional *sqhC* gene do produce hopanoids (Pearson *et al.* 2007), whereas organisms that do not produce hopanoids also lack *sqhC*. Therefore we have assumed that the presence of *sqhC* likely also implies production of hopanoids. However, any quantitative conclusions were made with caution in the absence of biomarker or transcriptomic data.

The link between hopanoid production and nitrite oxidation, identified through metagenomic surveys, provides new insight into a specific metabolism associated with hopanoid production in marine environments. Thus, it has the potential to improve the interpretation of patterns of hopanoid distribution in marine sediments for both current and past ocean environments. Importantly, although lipid biomarkers have been used to constrain N cycling processes in past ocean environments (reviewed in Rush and Sinninghe Damste, 2017), biomarkers for N cycling

processes that occur under suboxic or anoxic conditions are still limited. A better understanding of the role hopanoid producers play in the modern marine N cycle may improve our ability to track these processes in past and present marine environments.

For example, hopanoids in ancient sediments are often used to indicate the presence of aerobic heterotrophic bacteria and by extension the microbial degradation of organic matter (e.g. van Dongen *et al.* 2006). Recent work, however, suggests that anaerobic or microaerophilic bacteria in low oxygen regions of the water column, and in sediments, are likely important contributors to hopanoid sediment inventories (Sáenz *et al.* 2011; Kharbush *et al.* 2013; Berndmeyer *et al.* 2014). Furthermore, several studies, including this one, now suggest that many of these low oxygen-adapted bacteria are chemoautotrophic organisms, specifically anammox Planctomycetes and NOB (Sinninghe Damsté *et al.* 2004; Rush *et al.* 2014; Kharbush *et al.* 2015b), with the NOB (specifically, *Nitrospina*) being responsible for a large fraction of deep ocean DIC fixation (Pachiadaki *et al.* 2017). As such, variability in hopanoid/BHP abundances in the sediment record could identify past shifts in both water column suboxia/anoxia and the importance of nitrite oxidation relative to other N cycling processes throughout Earth's history or even in the more recent past. Because *Nitrospina* sp., for example, are found in the water column and both coastal (Rani *et al.* 2017) and abyssal marine sediments (Durbin and Teske 2011; Nguyen *et al.* 2017), hopanoids in sediments will include both water column and *in situ* sources. Further study of both the ecology and BHP structures produced by *Nitrospina* and other important NOB will be needed before sediment hopanoid abundances can be interpreted in the context of N-cycle processes and environmental conditions.

The unique biogeochemical conditions of the Red Sea distinguish this environment from those previously examined for the presence and diversity of hopanoid producers. Yet despite its warm waters, high salinity, and relatively well-oxygenated water column, the Red Sea harbors a broad

phylogenetic distribution of *sqhC* sequences. The greater presence of *Proteobacteria*-affiliated *sqhC* sequences in oxygenated waters of the Red Sea (**Supplementary Figure S2**) suggests that these hopanoid producers are adapted to higher $O_2$ concentrations, consistent with previous observations that identified mostly *Proteobacteria sqhC* sequences in Global Ocean Survey (GOS) metagenomic samples, which included exclusively aerobic environments in ocean surface waters (Pearson and Rusch 2009). Therefore the importance of *Proteobacteria*, including *Alphaproteobacteria*, seems to be greater in more oxygenated, shallower waters versus within OMZ environments.

Even under the higher $O_2$ and low nitrite concentrations found throughout the water column of the Red Sea, a relatively large number of *sqhC* sequences are affiliated with the NOB *Nitrospina* and *Nitrospira.* This corroborates previous assertions that these two groups of NOB are globally distributed organisms important in both hopanoid production and N cycling. These NOB are frequently found in OMZs (Wright *et al.* 2012; Hawley *et al.* 2014), so their presence in the Red Sea is somewhat surprising given that $O_2$ levels do not reach the suboxic range. However, both *Nitrospina* and *Nitrospira* have also been detected in well-oxygenated waters in trenches (Nunoura *et al.* 2015), although the specific environmental parameters influencing their distribution and potential niche separation are still being elucidated. In addition, *Nitrospina* and *Nitrospira* both contain a periplasmically-oriented membrane-bound nitrite oxidoreductase (NXR) enzyme rather than the cytoplasmically-oriented version of *Nitrobacter* and *Nitrococcus* sp. (Lücker *et al.* 2010, 2013). This type of NXR likely allows growth at extremely low nitrite concentrations (Sorokin *et al.* 2012; Lücker *et al.* 2013), like those seen in much of the Red Sea (**Supplementary Figure S3)**.

The presence of NOB in the Red Sea may be a result of inherent metabolic plasticity; perhaps NOB introduced to the Red Sea through the annual influx of GAIW are able to survive

despite drastically different $O_2$ levels. Alternatively, they may represent distinct species that are adapted to occupy specific niches in the Red Sea. The major OTUs of *Nitrospina* identified in the ETNP OMZ displayed different depth distributions, demonstrating that various species or ecotypes can be adapted to particular water column conditions (Beman *et al.* 2013). The phylogenetic distance between the *sqhC* sequences of *Nitrospina* sp. found in the ETSP versus the ETNP OMZs provides additional evidence in support of this hypothesis (Kharbush *et al.* 2015a), as does the existence of multiple distinct Nitrospinae lineages in a recently generated phylogeny of phylum Nitrospinae (Ngugi *et al.* 2016).

The Red Sea water column *Nitrospina* sp. were placed into phylogenetic context using this recent phylogeny (Ngugi *et al.* 2016), which included 16S rRNA sequences from various environments including the water column of the NESAP OMZ, seafloor sediments, and the brine-seawater interfaces at the bottom of the Red Sea. Because sequence lengths varied, preventing construction of a *de novo* phylogenetic tree, parsimony insertion was used to add partial 16S rRNA sequences from the ETNP and ETSP OMZs and the Red Sea metagenomes to this initial phylogeny. The results show that the Red Sea water column *Nitrospina* sp. are most closely related to ETSP and ETNP *Nitrospina*, and fall within Clade 2 rather than the newly defined *Candidatus* Nitromaritima clade (Clade 1), which includes all available sequences from the NESAP OMZ (**Figure 4**). This apparent phylogenetic separation between *Nitrospina* in the major OMZ regions is somewhat surprising, as is the separation between the *Nitrospina* SAGs identified in the brine-seawater interface (BSI) of the Atlantis II deep vs. those found in other deep Red Sea BSIs such as the Kebrit, Discovery, and Erba deeps. However, the deep brine pools of the Red Sea contain very different physicochemical conditions, including variations in temperature, $O_2$ concentration, and nutrient concentrations, which creates distinct environmental niches and results in diverse and complex microbial communities specific to each deep basin

(Guan *et al.* 2015). Similarly, there are important physicochemical differences between OMZ environments: for example, the NESAP OMZ is located much deeper in the water column than the ETSP and ETNP OMZs, and contains much higher $O_2$ concentrations at its core. Therefore it is possible that these two major marine clades of phylum Nitrospinae have distributions based on physiological adaptations rather than geographical proximity.

Overall, the emerging picture of hopanoid producers as also nitrite-utilizing organisms represents a significant shift in our understanding of the ecology of these organisms in marine environments. The data presented in this study show that even in non-OMZ environments, nitrite-oxidizing bacteria make up a significant fraction of the marine hopanoid-producing community. While early studies of the taxonomic affiliations of hopanoid producers identified significant metabolic diversity, the physiological roles of this biomarker have remained inconclusive. This study, along with previous work in the eastern North Pacific Ocean, identifies a specific group of organisms and a metabolism for which hopanoid production is important. As such, this work paves the way for targeted studies that can examine the role of hopanoid production in the physiology of an ecologically important and widely distributed group of marine nitrite oxidizers.

FUNDING

This work was supported by the National Science Foundation Graduate Research Fellowship Program (awarded to JJK). In addition, the KAUST Red Sea Expedition 2011 was funded by the King Abdullah University of Science and Technology (KAUST).

ACKNOWLEDGEMENTS

We thank the participants of the KAUST Red Sea Expedition 2011, including the P.I. Dr. Ulrich Stingl (currently at University of Florida) and those who helped to generate the data. We also thank Dr. Peter Girguis (Harvard University) for supporting MFH during the preparation of this manuscript.

REFERENCES

Abubucker S, Segata N, Goll J *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 2012;**8**, DOI: 10.1371/journal.pcbi.1002358.

Beman JM, Shih JL, Popp BN. Nitrite oxidation in the upper water column and oxygen minimum zone of the eastern tropical North Pacific Ocean. *ISME J* 2013;**7**:2192–205.

Berndmeyer C, Thiel V, Schmale O *et al.* Biomarkers in the stratified water column of the Landsort Deep (Baltic Sea). *Biogeosciences Discuss* 2014;**11**:9853–87.

Brocks JJ, Logan GA, Buick R *et al.* Archean molecular fossils and the early rise of eukaryotes. *Science (80- )* 1999;**285**:1033–6.

van Dongen BE, Talbot HM, Schouten S *et al.* Well preserved Palaeogene and Cretaceous biomarkers from the Kilwa area, Tanzania. *Org Geochem* 2006;**37**:539–57.

Durbin AM, Teske A. Microbial diversity and stratification of South Pacific abyssal marine sediments. *Environ Microbiol* 2011;**13**:3219–34.

Felsenstein J. *PHYLIP 3.2 Manual*. University of California Herbarium, Berkeley, 1989.

Guan Y, Hikmawan T, Antunes A *et al.* Diversity of methanogens and sulfate-reducing bacteria in the interfaces of five deep-sea anoxic brines of the Red Sea. *Res Microbiol* 2015;**166**:688–99.

Haroon MF, Thompson LR, Parks DH *et al.* A catalogue of 136 microbial draft genomes from Red Sea metagenomes. *Sci Data* 2016;**3**, DOI: 10.1038/sdata.2016.50.

Hawley AK, Brewer HM, Norbeck AD *et al.* Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc Natl Acad*

*Sci* 2014;**111**:11395–400.

Kharbush JJ, Kejriwal K, Aluwihare LI. Distribution and abundance of hopanoid producers in low oxygen environments of the Eastern Pacific Ocean. *Microb Ecol* 2015a;**71**:401–8.

Kharbush JJ, Kejriwal K, Aluwihare LI. Distribution and Abundance of Hopanoid Producers in Low-Oxygen Environments of the Eastern Pacific Ocean. *Microb Ecol* 2015b;**71**:401–8.

Kharbush JJ, Ugalde JA, Hogle SL *et al.* Composite bacterial hopanoids and their microbial producers across oxygen gradients in the water column of the California Current. *Appl Environ Microbiol* 2013;**79**:7491–501.

Krzywinski M, Schein J, Birol I *et al.* Circos- An information aesthetic for comparative genomics. *Genome Res* 2009.

Legendre P, Legendre LF. *Numerical Ecology*. Amsterdam: Elsevier, 2012.

Levipan HA, Molina V, Fernandez C. Nitrospina-like bacteria are the main drivers of nitrite oxidation in the seasonal upwelling area of the Eastern South Pacific (Central Chile ~36°S). *Environ Microbiol Rep* 2014;**6**:565–73.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Prepr* 2013;**0**:1–3.

Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9.

Lücker S, Nowka B, Rattei T *et al.* The Genome of Nitrospina gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Front Microbiol* 2013;**4**:27.

Lücker S, Wagner M, Maixner F *et al.* A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci U S A* 2010;**107**:13479–84.

Ludwig W, Strunk O, Westram R *et al.* ARB: A software environment for sequence data. *Nucleic Acids Res* 2004;**32**:1363–71.

Nawrocki EP. Structural RNA Homology Search and Alignment using Covariance Models. 2009.

Ngugi DK, Blom J, Stepanauskas R *et al.* Diversification and niche adaptations of Nitrospina-like bacteria in the polyextreme interfaces of Red Sea brines. *ISME J* 2016;**10**:1383–99.

Nguyen TT, Cochrane SKJ, Landfald B. Perturbation of seafloor bacterial community structure by drilling waste discharge. *Mar Pollut Bull* 2017, DOI: 10.1016/j.marpolbul.2017.10.039.

Nunoura T, Takaki Y, Hirai M *et al.* Hadal biosphere: Insight into the microbial ecosystem in the deepest ocean on Earth. *Proc Natl Acad Sci* 2015:E1230–6.

Nurk S, Bankevich A, Antipov D *et al.* Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. *Journal Comput Biol* 2013;**20**:714–37.

Ourisson G, Albrecht P. Hopanoids. 1. Geohopanoids: the most abundant natural products on Earth? *Acc Chem Res* 1992;**25**:398–402.

Pachiadaki MG, Sintes E, Bergauer K *et al.* Major role of nitrite-oxidizing bacteria in dark ocean carbon fixation. *Science (80- )* 2017;**358**:1046–51.

Pearson A, Flood Page SR, Jorgenson TL *et al.* Novel hopanoid cyclases from the environment. *Environ Microbiol* 2007;**9**:2175–88.

Pearson A, Leavitt WD, Sáenz JP *et al.* Diversity of hopanoids and squalene-hopene cyclases across a tropical land-sea gradient. *Environ Microbiol* 2009;**11**:1208–23.

Pearson A, Rusch DB. Distribution of microbial terpenoid lipid cyclases in the global ocean metagenome. *ISME J* 2009;**3**:352–63.

Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**, DOI: 10.1371/journal.pone.0009490.

Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 2012;**28**:1823–9.

Rani S, Koh H, Rhee S *et al.* Detection and diversity of the nitrite oxidoreductase alpha subunit (nxrA) gene of Nitrospina in marine sediments. *Microbiol Ecol* 2017;**73**:111–22.

Ricci JN, Morton R, Kulkarni G *et al.* Hopanoids play a role in stress tolerance and nutrient storage in the cyanobacterium *Nostoc punctiforme*. *Geobiology* 2017;**15**:173–83.

Rush D, Sinninghe Damste JS. Lipids as paleomarkers to constrain the marine nitrogen cycle. *Environ Microbiol* 2017;**19**:2119–32.

Rush D, Sinninghe Damsté JS, Poulton SW *et al.* Anaerobic ammonium-oxidising bacteria: A biological source of the bacteriohopanetetrol stereoisomer in marine sediments. *Geochim Cosmochim Acta* 2014;**140**:50–64.

Sáenz JP. Hopanoid enrichment in a detergent resistant membrane fraction of Crocosphaera watsonii: Implications for bacterial lipid raft formation. *Org Geochem* 2010;**41**:853–6.

Sáenz JP, Sezgin E, Schwille P *et al.* Functional convergence of hopanoids and sterols in membrane ordering. *PNAS* 2012a;**109**:14236–40.

Sáenz JP, Wakeham SG, Eglinton TI *et al.* New constraints on the provenance of hopanoids in the marine geologic record: Bacteriohopanepolyols in marine suboxic and anoxic environments. *Org Geochem* 2011;**42**:1351–62.

Sáenz JP, Waterbury JB, Eglinton TI *et al.* Hopanoids in marine cyanobacteria: probing their phylogenetic distribution and biological role. *Geobiology* 2012b;**10**:311–9.

Schmerk CL, Bernards MA, Valvano MA. Hopanoid production is required for low-pH

tolerance, antimicrobial resistance, and motility in Burkholderia cenocepacia. *J Bacteriol* 2011;**193**:6712–23.

Sinninghe Damsté JS, Rijpstra WIC, Schouten S *et al.* The occurrence of hopanoids in planctomycetes: implications for the sedimentary biomarker record. *Org Geochem* 2004;**35**:561–6.

Smeed D. Seasonal variation of the flow in the strait of Bab al Mandab. *Oceanol Acta* 1997;**20**:773–81.

Sorokin DY, Lücker S, Vejmelkova D *et al.* Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi. *ISME J* 2012;**6**:2245–56.

Souvermezoglou E, Metzl N, Poisson A. Red Sea budgets of salinity, nutrients and carbon calculated in the Strait of Bab-El-Mandab during the summer and winter seasons. *J Mar Res* 1989;**47**:441–56.

Spieck E, Keuter S, Wenzel T *et al.* Characterization of a new marine nitrite oxidizing bacterium, Nitrospina watsonii sp. nov., a member of the newly proposed phylum "Nitrospinae." *Syst Appl Microbiol* 2014;**37**:170–6.

Thompson LR, Williams GJ, Haroon MF *et al.* Metagenomic covariation along densely sampled environmental gradients in the Red Sea. *ISME J* 2017;**11**:138–51.

Welander P V, Hunter RC, Zhang L *et al.* Hopanoids play a role in membrane integrity and pH homeostasis in Rhodopseudomonas palustris TIE-1. *J Bacteriol* 2009;**191**:6145–56.

Wright JJ, Konwar KM, Hallam SJ. Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* 2012;**10**:381–94.

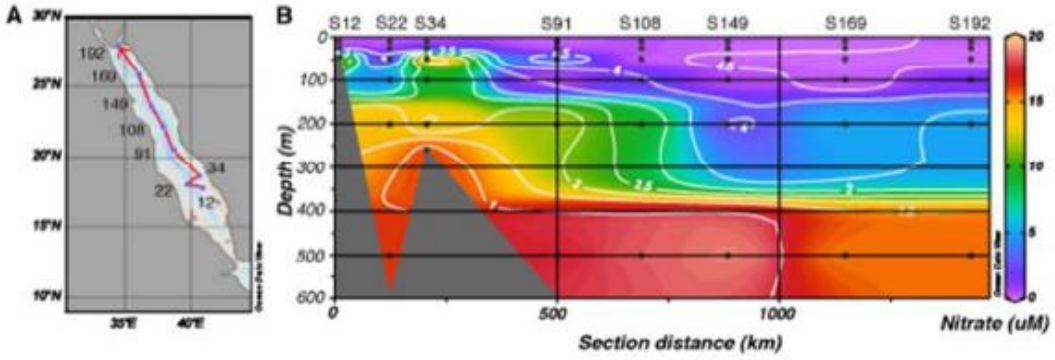


**Figure 1.** A map of the Red Sea and locations of the eight stations sampled along the 2011 KRSE cruise is shown in (**A).** The red arrow indicates the direction in which the stations are plotted in the accompanying section plot (**B**) showing nitrate in color and oxygen concentrations (in mL/L) as white contour lines, plotted against depth. Black dots represent the depths sampled at each station, which are labeled at the top of the plot (e.g. "S12" is "Station 12").

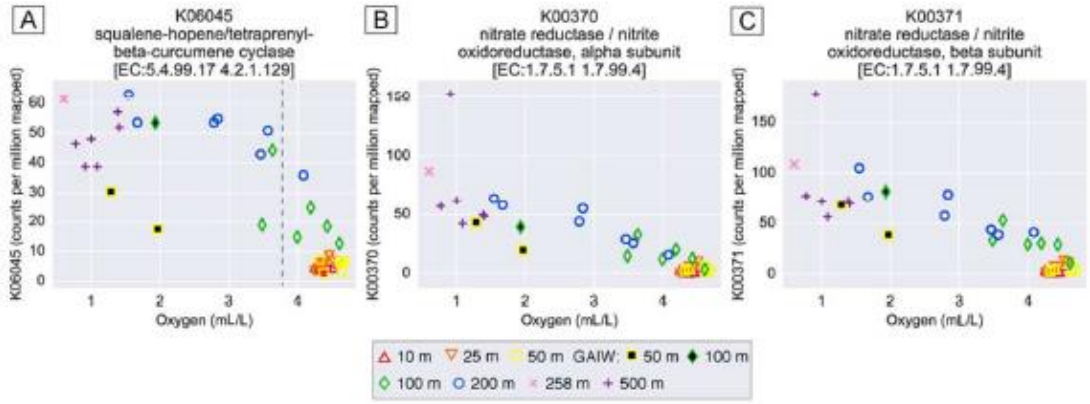


**Figure 2.** Covariation of relative abundance of *sqhC* gene family (K06045) and nitrite oxidoreductase (K00370/K00370) with oxygen and depth in the Red Sea water column. Gulf of Aden Intermediate Water (GAIW)- influenced samples are separately designated by symbols with solid centers. (A) Scatter plot of *sqhC* relative abundance versus oxygen concentration. Dashed line represents the chosen threshold of 3.75 mL/L used to separate samples of lower and higher oxygen content in later analyses. (B) Plot of *nxrA* relative abundance versus oxygen. (C) Plot of *nxrB* relative abundance versus oxygen.

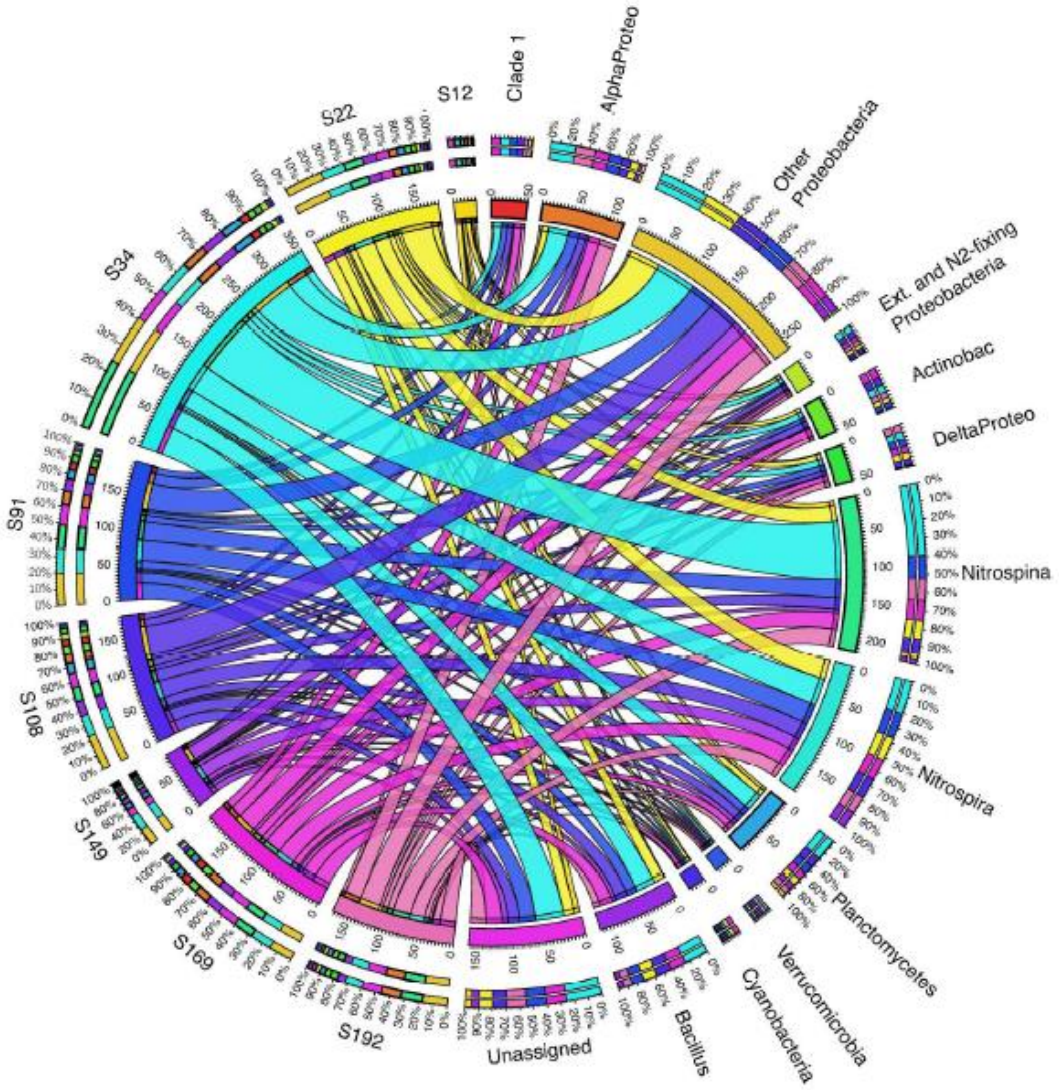


**Figure 3.** Circos plot showing the phylogenetic distribution of *sqhC* metagenomic hits from low-$O_2$ samples among stations. Categories of hopanoid-producing bacteria are shown along the right side of the plot, while stations are on the left. Colors of the bands correspond to each station. Width of the bands at each end is proportional to the number of sequences associated with the station and with the microbial group. Absolute numbers and percentages can be found in the inner and outer circles, respectively. For example, in the samples from Station 34, about 350 *sqhC* sequences were identified, and of those approximately 85 or 40% of the total were affiliated with *Nitrospina*. Bacterial categories were assigned based on a previously published phylogeny that incorporated *sqhC* from sequenced organisms as well as from uncultured organisms in environmental samples. The "unassigned" category therefore represents sequences that could not be reliably assigned to any group in the existing phylogeny. Abbreviations: "Actinobac" = Actinobacteria, "AlphaProteo"= Alphaproteobacteria, "DeltaProteo" =

Deltaproteobacteria. “Ext. and N2-fixing Proteobacteria” = Extremophilic and/or N2-fixing Proteobacteria. “C1” represents a clade of hopanoid producers investigated in a recent study (Kharbush *et al.* 2015). “Other Proteobacteria” contains sequences most closely related to cultured Beta- and Gammaproteobacteria, while “Ext. and N2- fixing Proteobacteria” contained sequences most closely related to organisms known to occupy extreme environments and/or fix nitrogen, such as *Teredinibacter turnerae*, a Gammaproteobacterium that lives symbiotically in the gut of shipworms. It is important to note that these classifications are purposefully broad due to the dissimilarity of environmental *sqhC* sequences compared to cultured organisms, and should be treated as an estimation of diversity rather than an absolute measure.

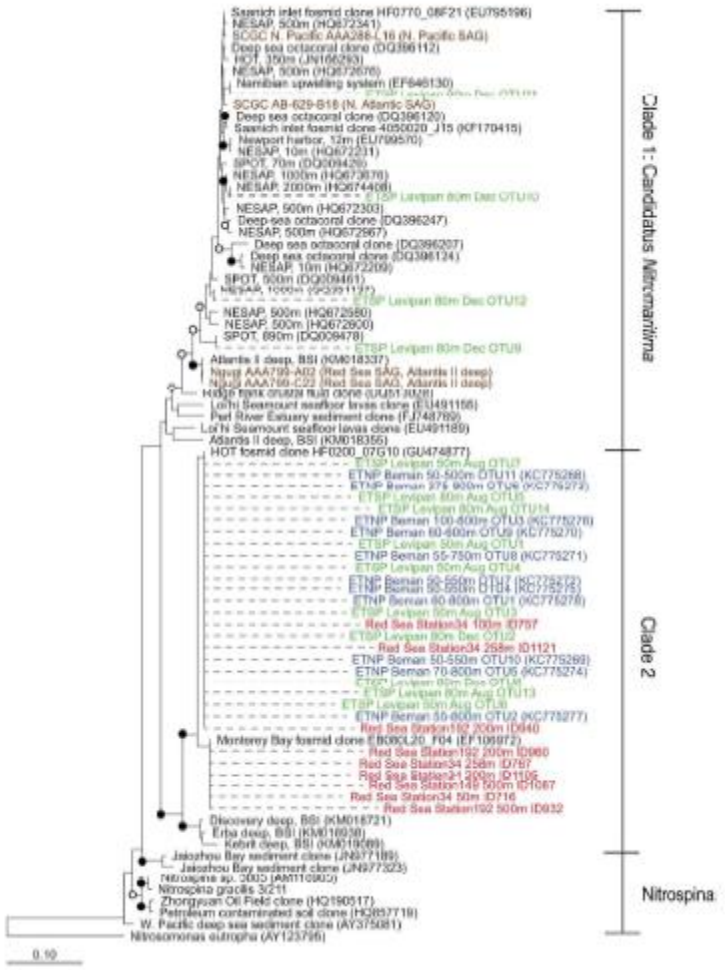


**Figure 4.** Phylogenetic tree of 16S sequences from phylum Nitrospinae, with *Nitrosomonas eutropha* as the outgroup. Dashed lines indicate shorter sequences that were inserted by parsimony. Solid circles indicate bootstrap support >90% while open circles indicate bootstrap support >70%. Colors: Black = Sequences previously included in the most recent Nitrospinae phylogeny (Ngugi *et al.* 2016), Brown = sequences identified from SAGs, Blue = ETNP PCR amplicons (Beman *et al.* 2013), Green = ETSP PCR amplicons (Levipan *et al.* 2014), Red = Red Sea water column 16S sequences (this study).