Microbial community functioning at hypoxic sediments revealed by targeted metagenomics and RNA stable isotope probing

Microbial community functioning at hypoxic sediments revealed by targeted metagenomics and RNA stable isotope probing

By Christina Pavloudi

Promotors: Prof. Dr. Marleen De Troch

Prof. Dr. Michael W. Friedrich

Dr. Christos Arvanitidis

Academic year 2017-2018

PhD predefence: 24/08/2017

PhD defence: 29/09/2017

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor in Science: Marine Sciences of the MARES Joint Doctoral Training Programme in Marine Ecosystem Health and Conservation. The doctoral studies were jointly supervised by the University of Bremen and Ghent University, according to a Cotutelle Agreement.

Members of the examination committee and reading committee

Prof. Dr. Marleen De Troch, Promotor

Ghent University, Ghent, Belgium

Prof. Dr. Michael W. Friedrich, Promotor

Bremen University, Bremen, Germany

Dr. Christos Arvanitidis, Promotor

Hellenic Centre for Marine Research, Heraklion, Greece

Dr. ir. Jo De Vrieze

Ghent University, Ghent, Belgium

Prof. Dr. Anne Willems

Ghent University, Ghent, Belgium

Prof. Dr. Dominique Adriaens (Chair)

Ghent University, Ghent, Belgium

Dr. Ulrike Braeckman (Secretary)

Ghent University, Ghent, Belgium

PD Dr. Holger Auel

Bremen University, Bremen, Germany

Prof Dr. Jens Harder

Max Planck Institute for Marine Microbiology, Bremen, Germany

Acknowledgements

And after four years, the time has come to finish this journey and start another one. But of course, this wouldn't have been possible without the help and support of many people.

First of all, I would like to thank my promotors for making this possible. Dr. Christos Arvanitidis who has been my supervisor for many years and has supported me when and where it was needed. From discussing scientific hypotheses, to sampling campaigns, laboratory work and scientific writing, he was eager to introduce me to all the stages of a scientific career and I am grateful for that. Prof. Dr. Michael W. Friedrich who accepted me at his lab and introduced me to RNA Stable Isotope Probing, which is such an interesting and promising method and who gave me the opportunity to participate in the M103/1 cruise so that I would be responsible for my samples from the beginning to the end. Prof. Dr. Marleen De Troch who also accepted me in the Marine Biology group in Ghent and supported me in the final stage of my PhD, when it was undoubtedly needed.

I would also like to thank the members of my examination committee who took the time and effort to thoroughly read my thesis and comment where it was necessary. The content has definitely improved since I followed their suggestions.

My special acknowledgements to the MARES initiative. It was great to be a part of this network which would not have been possible without emer. Prof. Dr Magda Vincx and Dr. Tim Deprez who established it and made it successful. I would like to additionally thank emer. Prof. Dr Magda Vincx for her support and assistance since she was my promotor at the beginning of my PhD.

Of course, my special thanks to members of the Microbial Ecophysiology in Bremen and in particular to Dr. Saar Sztejrenszus, Ajinkya Kulkarni and Yin Xiuran. Saar, I would like to thank you for taking the time to thoroughly teach me RNA Stable

Isotope Probing and for all our discussions, not only on science. Yin thank you for helping me when I needed it and for exchanging ideas and comments on the practical lab work. Ajinkya, thank you for being my office partner when I visited your lab; it was really fun to be around and may the force be with you!

I would also like to thank the members of the Marine Biology group in Ghent and especially Anna-Maria Vafeiadou, Qian Huang and Ee Zin Ong. You made it much easier and relaxing for me to be in Ghent and I am grateful for that. Good luck with your PhDs, I know you will do great work.

I would also really like to thank Dr. Umer Zeeshan Ijaz for providing access to the Orion IT cluster, where part of the analyses was conducted, and for developing several bioinformatics tutorials which really facilitated my work.

I am also very grateful to the members of the Genetics lab of the Hellenic Centre for Marine Research and especially to Dr. Jon Bent Kristoffersen. Jon, I am really grateful for agreeing to share your expertise and for all our discussions on the library preparation protocols, on MiSeq and all the practical laboratory techniques.

Finishing this PhD would not have been possible without the support of my friends and lab partners at the Biodiversity lab of the Hellenic Centre for Marine Research. I would never forget your support and love, all the great times we had drinking coffee (and not only), doing “group therapy”, gossiping, brainstorming, “smoking”, sampling and even inventing new yoga positions…! And there are so many pictures to show how great we are all together, in sickness and in health, for better or for worse! Eva, Niki, Giorgo, Sara you are all “mountains” με “αμβρακιά”!! Wanda you know that you will always be our lab queen, we have become a family after all these years…!

My special thanks and love to the other “mountain/κουραμπιέ/πουλέν/drinking buddy/cubicle partner”, my friend Katerina, for everything she did and for being there during these years. Also, to the other τσούχτρες, Vicky and Ioanna, for all the

craziness that kept me sane and for not being able to have a conversation without changing the subject every 10 minutes…!
A big thank you from the bottom of my heart to my friends Marina, Angeliki and Nadia for all their love and support. We may be living so far away but nothing comes between us.

To Panos, who believed in me even when I didn't and survived being with me during my Master and my PhD, and to all our animal-friends, I wouldn't have done it without you… Πες μου ποιο φόβο αγάπησες πάλι…;

This last word of acknowledgment is of course for my parents, my brother, my sister-in-law, who is more of a sister to me than my brother's wife, and my baby nephew. They were there throughout my mood swings and when I was feeling angry and unhappy, because of that and despite of that. I love you all and I thank you for everything.

Universität Bremen

Ort, Datum: 14/09/2017, Heraklion, Crete, Greece

Versicherung an Eides Statt

(affirmation in lieu of oath)

Ich, Christina Pavloudi (Address: Parodos Kapetanissis Marigos 25, 71201, Heraklion, Crete, Greece)

versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.
Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.
Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

14/09/2017
Heraklion, Crete, Greece

_____
Ort, Datum Unterschrift

Table of contents

Summary

Microorganisms are instrumental to the structure and functioning of marine ecosystems and to the chemistry of the ocean due to their essential part in the cycling of the elements and in the recycling of the organic matter. Two of the most critical ocean biogeochemical cycles are those of nitrogen and sulfur, since they can influence the synthesis of nucleic acids and proteins, primary productivity and microbial community structure.

Oxygen concentration in marine environments is one of the environmental variables that have been largely affected by anthropogenic activities; its decline induces hypoxic events which affect benthic organisms and fisheries. Hypoxia has been traditionally defined based on the level of oxygen below which most animal life cannot be sustained. Hypoxic conditions impact microbial composition and activity since anaerobic reactions and pathways are favoured, at the expense of the aerobic ones. Naturally occurring hypoxia can be found in areas where water circulation is restricted, such as coastal lagoons, and in areas where oxygen-depleted water is driven into the continental shelf, i.e. coastal upwelling regions.

Coastal lagoons are highly dynamic aquatic systems, particularly vulnerable to human activities and susceptible to changes induced by natural events. For the purpose of this PhD project, the lagoonal complex of Amvrakikos Gulf, one of the largest semi-enclosed gulfs in the Mediterranean Sea, was chosen as a study site.

Coastal upwelling regions are another type of environment limited in oxygen, where also formation of oxygen minimum zones (OMZs) has been reported. Sediment in upwelling regions is rich in organic matter and bottom water is often depleted of oxygen because of intense heterotrophic respiration. For the purpose of this PhD project, the chosen coastal upwelling system was the Benguela system off Namibia, situated along the coast of south western Africa.

The aim of this PhD project was to study the microbial community assemblages of hypoxic ecosystems and to identify a potential link between their identity and function, with a particular emphasis on the microorganisms involved in the nitrogen

and sulfur cycles. The methodology that was applied included targeted metagenomics and RNA stable isotope probing (SIP).

It has been shown that the microbial community diversity pattern can be differentiated based on habitat type, i.e. between riverine, lagoonal and marine environments. Moreover, the studied habitats were functionally distinctive. Apart from salinity, which was the abiotic variable best correlated with the microbial community pattern, oxygen concentration was highly correlated with the predicted metabolic pattern of the microbial communities. In addition, when the total number of Operational Taxonomic Units (OTUs) was taken into consideration, a negative linear relationship with salinity was identified (see Chapter 2).

Microbial community diversity patterns can also be differentiated based on the lagoon under study since each lagoon hosts a different sulfate-reducing microbial (SRM) community, again highly correlated with salinity. Moreover, the majority of environmental terms that characterized the SRM communities were classified to the marine biome, but terms belonging to the freshwater or brackish biomes were also found in stations were a freshwater effect was more evident (see Chapter 3).

Taxonomic groups that were expected to be thriving in the sediments of the Benguela coastal upwelling system were absent or present but in very low abundances. *Epsilonproteobacteria* dominated the anaerobic assimilation of acetate as confirmed by their isotopic enrichment in the SIP experiments. Enhancement of known sulfate-reducers was not achieved under sulfate addition, possibly due to competition for electron donors among nitrate-reducers and sulfate-reducers, to the inability of certain sulfate-reducing bacteria to use acetate as electron donor or to the short duration of the incubations (see Chapter 4).

Future research should focus more on the community functioning of such habitats; an increased understanding of the biogeochemical cycles that characterize these hypoxic ecosystems will perhaps allow for predictions regarding the intensity and direction of the cycling of elements, especially of nitrogen and sulfur given their biological importance. Regulation of hypoxic episodes will aid the end-users of these ecosystems to possibly achieve higher productivity, in terms of fish catches, which otherwise is largely compromised by the elevated hydrogen sulfide concentrations.

Περίληψη

Οι μικροοργανισμοί συμβάλλουν στη δομή και λειτουργία των θαλάσσιων οικοσυστημάτων και στη χημεία του ωκεανού λόγω του ουσιαστικού τους ρόλου στους κύκλους των χημικών στοιχείων και στην ανακύκλωση της οργανικής ύλης. Δύο από τους πιο σημαντικούς θαλάσσιους βιογεωχημικούς κύκλους είναι αυτοί του αζώτου και του θείου, καθώς μπορούν να επηρεάσουν τη σύνθεση των νουκλεϊκών οξέων και πρωτεϊνών, την πρωτογενή παραγωγικότητα και τη δομή της μικροβιακής κοινότητας.

Η συγκέντρωση του οξυγόνου στα θαλάσσια περιβάλλοντα είναι μία από τις περιβαλλοντικές μεταβλητές που έχουν επηρεαστεί σε μεγάλο βαθμό από τις ανθρωπογενείς δραστηριότητες. Η μείωση του οξυγόνου προκαλεί υποξικά φαινόμενα που επηρεάζουν τους βενθικούς οργανισμούς και την αλιεία. Η υποξία έχει παραδοσιακά οριστεί με βάση το επίπεδο οξυγόνου κάτω από το οποίο δε μπορεί να διατηρηθεί η ζωή των μακροπανιδικών οργανισμών. Οι υποξικές συνθήκες επηρεάζουν τη σύνθεση και δραστηριότητα της μικροβιακής κοινότητας, εφόσον ευνοούνται αναερόβιες αντιδράσεις και οδοί, σε βάρος των αερόβιων. Η φυσική υποξία μπορεί να παρατηρηθεί σε περιοχές όπου υπάρχει περιορισμός στην κυκλοφορία του νερού, όπως οι παράκτιες λιμνοθάλασσες, και σε περιοχές όπου υδάτινες μάζες αποστερημένες από οξυγόνο οδηγούνται στην ηπειρωτική υφαλοκρηπίδα, δηλαδή σε περιοχές παράκτιων ανοδικών αναβλύσεων.

Οι παράκτιες λιμνοθάλασσες είναι δυναμικά υδάτινα συστήματα, ιδιαίτερα ευάλωτα στις ανθρωπογενείς δραστηριότητες και ευαίσθητα στις αλλαγές που προκαλούνται από φυσικά φαινόμενα. Για το σκοπό της διδακτορικής διατριβής επιλέχθηκε ως περιοχή μελέτης το συγκρότημα λιμνοθαλασσών του Αμβρακικού κόλπου, ενός εκ των μεγαλύτερων ημί-κλειστων κόλπων της Μεσογείου.

Οι περιοχές παράκτιων ανοδικών αναβλύσεων είναι ένας άλλος τύπος περιβάλλοντος με περιορισμένη συγκέντρωση οξυγόνου, όπου έχουν επίσης αναφερθεί σχηματισμοί ζωνών ελαχίστου οξυγόνου. Τα ιζήματα στις περιοχές ανοδικών αναβλύσεων είναι πλούσια σε οργανική ύλη και το νερό του πυθμένα είναι συχνά αποστερημένο από οξυγόνο εξαιτίας της έντονης ετερότροφης

αναπνοής. Για το σκοπό της διδακτορικής διατριβής, η επιλεγμένη περιοχή παράκτιων ανοδικών αναβλύσεων ήταν το σύστημα της Μπενγκουέλα στη Ναμίμπια, που βρίσκεται κατά μήκος των ακτών της νοτιοδυτικής Αφρικής.

Σκοπός αυτής της διδακτορικής διατριβής ήταν η μελέτη των μικροβιακών συνευρέσεων των υποξικών οικοσυστημάτων και ο εντοπισμός μιας πιθανής σχέσης μεταξύ της ταυτότητας και της λειτουργίας τους, με ιδιαίτερη έμφαση στους μικροοργανισμούς που εμπλέκονται στους κύκλους του αζώτου και του θείου. Η μεθοδολογία που εφαρμόστηκε περιελάμβανε στοχευόμενη μεταγονιδιωματική και σήμανση του RNA με χρήση σταθερών ισοτόπων.

Έχει δειχθεί ότι το πρότυπο ποικιλότητας της μικροβιακής κοινότητας μπορεί να διαφοροποιηθεί με βάση τον τύπο του οικοτόπου, δηλαδή μεταξύ ποταμιού, λιμνοθάλασσας και θαλάσσιου περιβάλλοντος. Επιπλέον, τα ενδιαιτήματα που μελετήθηκαν ήταν λειτουργικά διακριτά. Εκτός από την αλατότητα, η οποία ήταν η αβιοτική μεταβλητή που συσχετίζεται καλύτερα με το πρότυπο της μικροβιακής κοινότητας, η συγκέντρωση του οξυγόνου ήταν σε μεγάλο βαθμό συσχετισμένη με το προβλεπόμενο μεταβολικό πρότυπο των μικροβιακών κοινοτήτων. Επιπλέον, όταν ελήφθη υπόψη ο συνολικός αριθμός των λειτουργικών ταξινομικών μονάδων, εντοπίστηκε μια αρνητική γραμμική σχέση με την αλατότητα (βλ. Κεφάλαιο 2).

Τα πρότυπα ποικιλότητας της μικροβιακής κοινότητας μπορούν επίσης να διαφοροποιηθούν με βάση τη λιμνοθάλασσα, δεδομένου ότι κάθε μία φιλοξενεί μια διαφορετική μικροβιακή ομάδα θειοαναγωγικών βακτηρίων, που επιπλέον συσχετίζεται σε μεγάλο βαθμό με την αλατότητα. Επιπλέον, οι περισσότεροι περιβαλλοντικοί όροι που χαρακτήρισαν τις κοινότητες των θειοαναγωγικών βακτηρίων ταξινομήθηκαν στο θαλάσσιο περιβάλλον, αλλά βρέθηκαν όροι που ανήκουν σε περιβάλλοντα γλυκού ή υφάλμυρου νερού στους σταθμούς όπου ήταν πιο εμφανής η επιρροή του γλυκού νερού (βλ. Κεφάλαιο 3).

Οι ταξινομικές ομάδες που αναμενόταν να προσδιοριστούν στα ιζήματα της παράκτιας ανοδικής ανάβλυσης της Μπενγκουέλα απουσίαζαν ή υπήρχαν αλλά σε πολύ χαμηλές αφθονίες. Τα *ε-πρωτεοβακτήρια* κυριαρχούσαν στα πειράματα σήμανσης σταθερών ισοτόπων και ειδικότερα σε αυτά που περιελάμβαναν αναερόβια αφομοίωση οξικού οξέος, όπως επιβεβαιώνεται από τον ισοτοπικό εμπλουτισμό τους. Η ενίσχυση των γνωστών θειοαναγωγικών βακτηρίων δεν ήταν

δυνατή υπό την προσθήκη θειικού άλατος, πιθανώς λόγω ανταγωνισμού για δότες ηλεκτρονίων μεταξύ θειοαναγωγικών και νιτροαναγωγικών βακτηρίων, στην αδυναμία ορισμένων θειοαναγωγικών βακτηρίων να χρησιμοποιούν οξικό οξύ ως δότη ηλεκτρονίων ή στη σύντομη διάρκεια των πειραμάτων (βλ. Κεφάλαιο 4).

Η μελλοντική έρευνα πρέπει να επικεντρωθεί περισσότερο στη λειτουργία των μικροβιακών κοινοτήτων τέτοιων οικοτόπων. Η ενισχυμένη κατανόηση των βιογεωχημικών κύκλων που χαρακτηρίζουν αυτά τα υποξικά οικοσυστήματα θα επιτρέψει πιθανώς προβλέψεις σχετικά με την ένταση και την κατεύθυνση των κύκλων των στοιχείων, ιδιαίτερα του αζώτου και του θείου, δεδομένης της βιολογικής τους σημασίας. Η ρύθμιση των υποξικών επεισοδίων θα βοηθήσει τους τελικούς χρήστες αυτών των οικοσυστημάτων να επιτύχουν πιθανότατα υψηλότερη παραγωγικότητα, όσον αφορά τα ιχθυαλιεύματα, τα οποία διαφορετικά βλάπτονται σε μεγάλο βαθμό από τις αυξημένες συγκεντρώσεις υδρόθειου.

Samenvatting

Micro-organismen zijn essentieel voor de structuur en het functioneren van mariene ecosystemen en voor de chemie van de oceaan door hun essentiële rol in de cycli van chemische elementen en bij het recycleren van organisch materiaal. Twee van de meest kritische biogeochemische cycli in de oceaan zijn de cycli van stikstof en zwavel, aangezien zij de synthese van nucleïnezuren en eiwitten, primaire productiviteit en microbiële gemeenschapsstructuur kunnen beïnvloeden.

De concentratie aan zuurstof in mariene systemen is één van de omgevingsvariabelen die grotendeels beïnvloed wordt door menselijke activiteiten; de afname van zuurstofconcentratie veroorzaakt hypoxische condities die benthische organismen en visserij beïnvloeden. Hypoxia wordt traditioneel gedefinieerd op basis van het zuurstofniveau waarbij het merendeel van het dierlijk leven niet kan worden behouden. Hypoxische omstandigheden beïnvloeden de microbiële samenstelling en hun activiteit omdat anaërobe reacties en chemische cycli worden bevorderd ten koste van de aërobe stoffen. Natuurlijk voorkomende hypoxia komt voor in gebieden met beperkte watercirculatie, zoals kustlagunes, en in gebieden waar zuurstofarm water in de continentale shelf wordt gedreven, namelijk in opwellingsgebieden aan de kust.

Kustlagunes zijn zeer dynamische aquatische systemen, maar zijn ook bijzonder kwetsbaar voor menselijke activiteiten en vatbaar voor veranderingen veroorzaakt door natuurlijke gebeurtenissen. Voor dit doctoraatsonderzoek werd het lagunecomplex van de Amvrakikos-Golf, één van de grootste semi-omsloten baaien in de Middellandse Zee, gekozen als studieplaats.

Opwelling regio's langs de kust zijn een andere situatie die tevens beperkt zijn in zuurstof en waar ook zuurstof minimale zones (OMZ's) gevormd kunnen worden. Sediment in opwellingsgebieden is rijk aan organisch materiaal en bodemwater is vaak zuurstofarm door intense heterotrofe respiratie. Voor dit doctoraatsproject werd het opwellingsysteem van het Benguela-systeem gelegen langs de kust van Zuidwest-Afrika ter hoogte van Namibië gekozen.

Het doel van dit doctoraatsonderzoek was om de microbiële gemeenschappen van hypoxische ecosystemen te bestuderen en om een potentiële link tussen hun identiteit en functie te identificeren, met bijzondere nadruk op de micro-organismen die betrokken zijn bij de stikstof- en zwavelcycli. De toegepaste methodologie omvatte gerichte metagenomics en RNA stabiele isotopenonderzoek (SIP).

de bekomen resultaten tonen aan dat het diversiteitspatroon van de microbiologische gemeenschap gedifferentieerd kan worden op basis van habitat type, d.w.z. tussen rivier-, lagune- en mariene systemen. Bovendien waren de bestudeerde habitats functioneel verschillend. Naast de saliniteit, de abiotische variabele die het beste correleerde met het microbiële gemeenschapspatroon, was zuurstofconcentratie sterk gecorreleerd met het voorspelde metabolische patroon van de microbiële gemeenschappen. Daarenboven, werd een negatieve lineaire relatie met zoutgehalte geïdentificeerd (zie hoofdstuk 2) wanneer het totale aantal operationele taxonomische eenheden (OTU's) in aanmerking werd genomen.

Microbiële gemeenschapsdiversiteitspatronen kunnen tevens gedifferentieerd worden op basis van de lagune die wordt onderzocht, aangezien elke lagune een andere sulfate-reducerende microbiële (SRM) gemeenschap herbergt, die opnieuw sterk gecorreleerd is met saliniteit. Bovendien werden de meeste milieuomstandigheden (ENVO termen) die de SRM-gemeenschappen kenmerkten ingedeeld in het mariene bioom, maar ook de omstandigheden die eigen zijn aan zoetwater of brakwater biomen werden gevonden in stations waar een effect van zoetwater duidelijker was (zie hoofdstuk 3).

Taxonomische groepen die verwacht werden om goed te gedeien in de sedimenten van het Benguela Coast opwellingssysteemwaren afwezig of werden gevonden in zeer geringe hoeveelheden. *Epsilonproteobacteriën* domineerden de anaërobe assimilatie van acetaat, zoals bevestigd door hun isotopische aanrijking in de SIP-experimenten. Verbetering van gekende sulfaat-reductoren werd niet bereikt bij de toevoeging van sulfaat, mogelijks door de concurrentie voor elektronendonoren onder nitraatreduceerder- en sulfaatreduceerders, door het onvermogen van bepaalde sulfaatreducerende bacteriën om acetaat als elektrondonor te gebruiken of door de korte duur van de incubaties (zie hoofdstuk 4).

Toekomstig onderzoek zou nog meer moeten richten zijn op het functioneren van de gemeenschap in dergelijke habitats; beter inzicht in de biogeochemische cycli die deze hypoxische ecosystemen karakteriseren kunnen voorspellingen toelaten over de intensiteit en de richting van de cycli van de elementen, in het bijzonder van stikstof en zwavel, gelet op hun biologisch belang. controle van hypoxische afleveringen zal de eindgebruikers van deze ecosystemen helpen om eventueel een hogere productiviteit te bereiken met name wat de visvangst betreft, die in grotere mate wordt gecompenseerd door de verhoogde concentraties aan waterstofsulfide.

Zusammenfassung

Mikroorganismen sind maßgeblich an der Struktur und dem Funktionieren der marinen Ökosysteme und der Chemie des Ozeans aufgrund ihres wesentlichen Anteils am Elementkreislauf und beim Recycling der organischen Substanz beteiligt. Zwei der kritischsten bioneochemischen Kreisläufe sind diejenigen von Stickstoff und Schwefel, da sie die Synthese von Nukleinsäuren und Proteinen beeinflussen können; primäre Produktivität und mikrobielle Gesellschaftsstruktur.

Die Sauerstoffkonzentration in marinen Umwelten ist eine der Umweltvariablen, die weitgehend von anthropogenen Aktivitäten betroffen ist; ihr Rückgang verursacht hypoxische Ereignisse, die benthische Organismen und Fischereigründe beeinflussen. Hypoxie wurde traditionell als der Sauerstoffgehalt definiert, unter dem nahezu kein tierisches Leben mehr möglich ist. Hypoxische Bedingungen beeinflussen die mikrobielle Zusammensetzung und Aktivität, da anaerobe Reaktionen und Stoffwechselwege begünstigt werden, auf Kosten der aeroben Organismen. Natürlich auftretende Hypoxie findet sich in Gebieten, in denen die Wasserzirkulation beschränkt ist, wie Küstenlagunen und jene Gebiete, in denen sauerstoffarmes Wasser Richtung Festlandsockel strömt, d.h. küstennahe Auftriebsgebiete.

Küstenlagunen sind hochdynamische aquatische Systeme, welche besonders anfällig für menschliche Aktivitäten sind und empfindlich auf Veränderungen reagieren, die durch Naturereignisse verursacht werden. Für dieses Promotionsvorhaben wurde der Lagunenkomplex des Amvrakikos-Golfs, eine der größten halbgeschlossenen Lagunen im Mittelmeer, als Untersuchungsgebiet gewählt.

Küstennahe Auftriebsgebiete sind spezielle Habitate, die sich durch Sauerstoff-Limitierung auszeichnen und bei denen auch über das Auftreten von Sauerstoff-Minimumzonen (engl. Oxygen Minimum Zones, OMZs) berichtet wurde. Das Sediment in Auftriebsgebieten ist reich an organischer Substanz und Bodenwasser ist oft Sauerstoff-arm wegen der intensiven Atmungstätigkeit heterotropher Organismen. Für diese Doktorarbeit wurde das Benguela Auftriebsgebiet von Namibia ausgewählt, das sich entlang der Küste des südwestlichen Afrikas befindet.

Ziel dieser Doktorarbeit war es, die mikrobiellen Zusammenhänge von hypoxischen Ökosystemen zu untersuchen und eine mögliche Verknüpfung zwischen Identität und Funktion zu identifizieren, mit besonderem Augenmerk auf die Mikroorganismen, die an den Stickstoff- und Schwefelkreisläufen beteiligt sind. Die angewandte Methodik beinhaltete gezielte Metagenomik und stabile Isotopenbeprobung von RNA (SIP; Stable Isotope Probing).

In dieser Arbeit konnte gezeigt werden, dass mikrobielle Diversitätsmuster auf der Grundlage des Lebensraums (Fluss-, Lagunen- und Meeresumgebungen) differenziert werden können. Darüber hinaus waren die untersuchten Lebensräume funktional einzigartig. Abgesehen vom Salzgehalt, welcher eine abiotische Variable war, die am besten mit dem mikrobiellen Diversitätsmuster korreliert war, war die Sauerstoffkonzentration hochgradig korreliert mit dem vorhergesagten Stoffwechselmuster der mikrobiellen Gemeinschaften. Darüber hinaus wurde bei der Berücksichtigung der Gesamtzahl der Operational Taxonomic Units (OTUs) eine nicht-lineare Beziehung zum Salzgehalt festgestellt (siehe Kapitel 2).

Mikrobielle Diversitätsmuster konnten auch auf der Grundlage der untersuchten Lagune unterschieden werden, da jede Lagune eine andere Sulfat-reduzierende mikrobielle (SRM) Gemeinschaft beherbergt, die wiederum mit dem Salzgehalt stark korreliert war. Darüber hinaus wurde die Mehrheit der Umweltbedingungen, die die SRM-Gemeinschaften charakterisierten, in das marine Biom eingeordnet, aber auch Bedingungen, die normalerweise für Süßwasser- oder Brackwasser-Biome gelten, wurden in Stationen gefunden, wo ein Süßwasser-Einfluss deutlicher war (siehe Kapitel 3).

Taxonomische Gruppen, von denen erwartet wurde, dass sie in den Sedimenten des Benguela-Küstenauftriebsgebiet vorkommen, waren abwesend oder vorhanden, aber in sehr niedrigen Häufigkeiten. *Epsilonproteobakterien* dominierten die anaerobe Assimilation von Acetat, was durch ihre isotopische Anreicherung in SIP-Experimenten bestätigt wurde. Die Zunahme der bekannten Sulfat-Reduzierer konnte nicht durch Sulfat-Zugabe erreicht werden. Gründe dafür könnten sein, a) die Konkurrenz um Elektronendonatoren zwischen Nitrat-Reduzierern und Sulfat-Reduzierern, b) die Unfähigkeit bestimmter sulfatreduzierender Bakterien Acetat als

Elektronendonator nutzen zu können und c) die vielleicht nicht ausreichende Inkubationsdauer (siehe Kapitel 4).

Die zukünftige Forschung sollte sich stärker auf die Funktionsweise solcher Lebensräume konzentrieren; ein besseres Verständnis der biogeochemischen Kreisläufe, die diese hypoxischen Ökosysteme charakterisieren, wird vielleicht Vorhersagen über die Bedeutung des Elementkreislaufs zulassen, insbesondere was die biologische Bedeutung des Stickstoff- und Schwefelkreislauf angeht. Die Regulierung hypoxischer Episoden wird den Endnutzern dieser Ökosysteme helfen, möglicherweise eine höhere Produktivität in Bezug auf Fischfänge zu erreichen, die ansonsten weitgehend durch die erhöhten Schwefelwasserstoffkonzentrationen gefährdet sind.

Chapter 1: General introduction

1.1 Marine Microbial Ecology: Biodiversity, Microbial Activity and Metabolism under low oxygen availability

Microorganisms are a phylogenetically diverse group of organisms that are classified mainly by size which can range from less than a μm up to a few cm; they include both simple unicellular forms, as well as multicellular forms. More often the term microorganisms is used to refer to prokaryotes, i.e. *Bacteria* and *Archaea*, and it will be used as such throughout this document. In a wider sense, it can also refer to eukaryotic organisms, such as Fungi, Protozoa, Algae, Rotifers etc., and viruses.

Based on the type of energy source (Figure 1.1A), on the carbon source (Figure 1.1B) and the electron source (Figure 1.1C), microorganisms can be divided in certain categories. Practically, all these terms can be combined; for example, chemolithoautotrophic bacteria, e.g. nitrifying and sulfur-oxidizing bacteria, as well as photoautotrophic bacteria, e.g. purple bacteria and Green sulfur bacteria, can be found in nature. Also, mixotrophy can be either photochemotrophic (e.g. SAR11, SAR86) or lithoorganotrophic (e.g. many marine *Alphaproteobacteria*).

Microorganisms are instrumental to the structure and functioning of marine ecosystems and to the chemistry of the ocean due to their essential part in the cycling of the elements and in the recycling of the organic matter (Falkowski *et al.*, 2008). They also play an essential role in trophic networks by perpetually interacting with other biological components (Glöckner *et al.*, 2012).

The metabolic pathways that are catalyzed by microbes include all the major elements that are the necessary prerequisites for the construction of biological macromolecules, i.e. carbon, hydrogen, oxygen, phosphorus, nitrogen and sulfur (Falkowski *et al.*, 2008). Despite the fact that, over the last decades, our knowledge on the specific enzymatic reactions that regulate these pathways has increased, there are still more to be discovered (Arrigo, 2005) especially in the light of projected climate changes (Singh *et al.,* 2010).

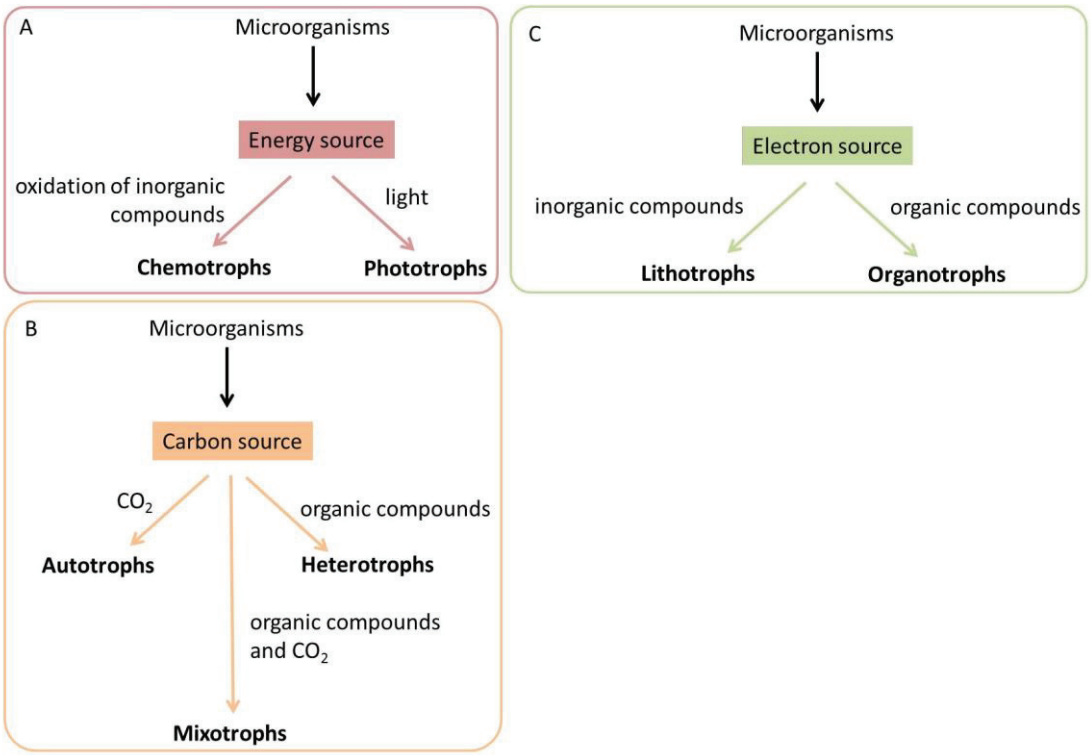

Figure 1.1: Overview of the metabolic categories of microorganisms. Division based on: A: the energy source, B: the carbon source, C: the electron source.

Two examples of biogeochemical cycles that interact and can act complementary are the nitrogen (Figure 1.2) and sulfur cycle (Figure 1.3). Both include aerobic as well as anaerobic metabolic reactions, with oxygen being the final electron acceptor in the former and nitrate or sulfate in the latter.

Sediment diagenetic reactions, i.e. processes that change a sediment subsequent to its deposition from water while excluding metamorphism and weathering processes (Berner, 1980), generate sources and sinks of elements, such as nitrogen and sulfur (Emerson and Hedges, 2003). Nitrate is produced during oxic and consumed during anoxic diagenesis (Emerson and Hedges, 2003). It has been estimated that in sediments the global denitrification rate, i.e. the consumption of nitrate during anoxic diagenesis, is ~18 x 10^{12} mol N yr^{-1} (Middelburg *et al.,* 1996). This is double the rate of water-column denitrification, thus demonstrating the crucial role of marine sediments as sink for fixed nitrogen (Emerson and Hedges, 2003). In addition, since oceans constitute the largest pool of sulfur (~1.3 x 10^{9} teragrams) (Schidlowski,

1989), it is indisputable that sedimentary-sulfur transformations are an important part of the sulfur cycle (Vairavamurthy *et al.,* 1995).

1.1.1 Nitrogen cycle in the ocean

Nitrogen is the fourth most abundant element in organic matter, after hydrogen, oxygen and carbon, but at the same time it is often limited in marine environments (Zehr and Kudela, 2011). Because of this, nitrogen cycling is one of the most critical ocean biogeochemical cycles since it can influence primarily the synthesis of nucleic acids and proteins, primary productivity (Vitousek *et al.*, 2002) and, subsequently, the structure of microbial communities (Ward and Jensen, 2014).

Nitrogen sources in the environment are mainly ammonia (NH_3), nitrate (NO_3^-) and nitrogen gas (N_2), with the latter being used only by nitrogen-fixing prokaryotes (Madigan *et al.*, 2003). Nitrogen undergoes several transformations, with the major ones being nitrogen fixation and ammonification (resulting in the production of ammonia (NH_3)), denitrification and anammox (resulting in the production of nitrogen (N_2)) and nitrification (resulting in the production of nitrate (NO_3^-)) (Figure 1.2) (Bernhard, 2010). Under nitrate limited conditions and in anoxic environments, dissimilative reduction of nitrate to ammonia (DRNA) can also contribute to nitrate and nitrite reduction (Madigan *et al.*, 2003). The above processes are catalyzed primarily by microbes (Zehr and Ward, 2002) which, undoubtedly, are vital for global nitrogen cycling (Bernhard, 2010).

During oxic conditions, nitrate (NO_3^-) is produced by nitrification in a two step process, with the formation of nitrite (NO_2^-) as an intermediate (Madigan *et al.*, 2003). In the anoxic respiration of nitrate (NO_3^-) to nitrite (NO_2^-), nitrous oxide (N_2O) and nitrogen (N_2), the redox potential is positive ($E_0' = +0.4$ V), thus the energy balance is favourable. This process is commonly referred to as denitrification because all the end products are gas and thus “escape” from the environment (Madigan *et al.*, 2003). In oxygen-deficient marine environments, organic matter can be oxidized by facultative anaerobic microorganisms which use nitrate as an electron acceptor. This process is referred to as heterotrophic denitrification, and results in

the release of N_2 (Kuypers *et al.*, 2005) and ammonium is expected to remain as such and not further oxidized due to the absence of oxygen (Devol, 2003). Anammox can also lead to the production of nitrogen (N_2), since ammonia (NH_3) is oxidized anaerobically, with nitrite (NO_2^-) being the electron acceptor (Madigan *et al.*, 2003).

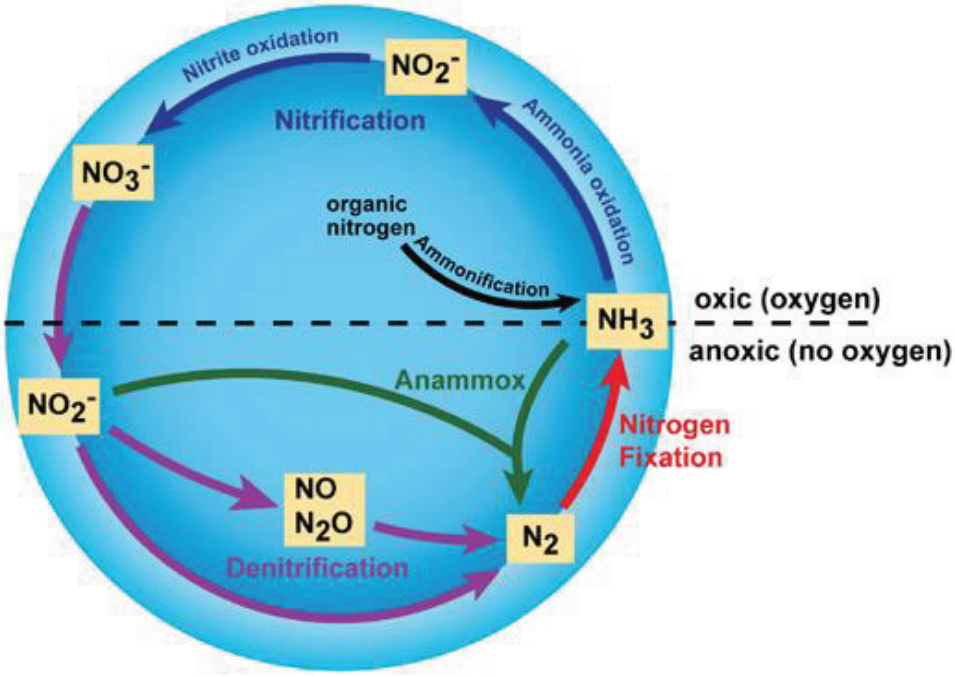

Figure 1.2: Overview of the nitrogen cycle (Bernhard, 2010; © 2010 Nature Education[1]).

Denitrification occurs in the presence of nitrate and low oxygen concentration (Zehr and Kudela, 2011), conditions that can be found in environments like OMZs and the oxic-anoxic interface of benthic sediments (Naqvi *et al.*, 2008). The balance of oceanic nitrogen budget is strongly dependent on the coupling between nitrogen fixation and denitrification, with the former representing the input of and the latter representing the output of nitrogen (Zehr and Kudela, 2011).

[1] Terms of Use: You may reproduce this material, without modifications, in print or electronic form for your personal, non-commercial purposes or for non-commercial use in an educational environment.

1.1.2 Sulfur cycle in the ocean

The largest quantities of sulfur on Earth are found in the form of sulfate and sedimentary minerals (e.g. pyrite), with oceans being its major reservoirs (Sievert *et al.*, 2007). Sulfur is also a vital element for life as it is required for protein synthesis, coenzymes etc. although it is rarely limited (Sievert *et al.*, 2007). Cells can incorporate several forms of sulfur, including sulfide (HS^-) and sulfate (SO_4^{2-}) (Madigan *et al.*, 2003).

Sulfur undergoes several transformations (Figure 1.3), with the most important ones being oxidation of elemental sulfur and inorganic sulfur compounds (e.g. hydrogen sulfide, sulfite, thiosulfate) to sulfate (SO_4^{2-}), catalyzed by chemoautotrophic and photosynthetic bacteria and reduction of sulfate, which can be either assimilatory or dissimilatory (Madigan *et al.*, 2003). The large number of oxidation states, along with the fact that certain sulfur transformations can also occur abiotically, render the microbial transformations of sulfur more complex than those of nitrogen (Madigan *et al.*, 2003).

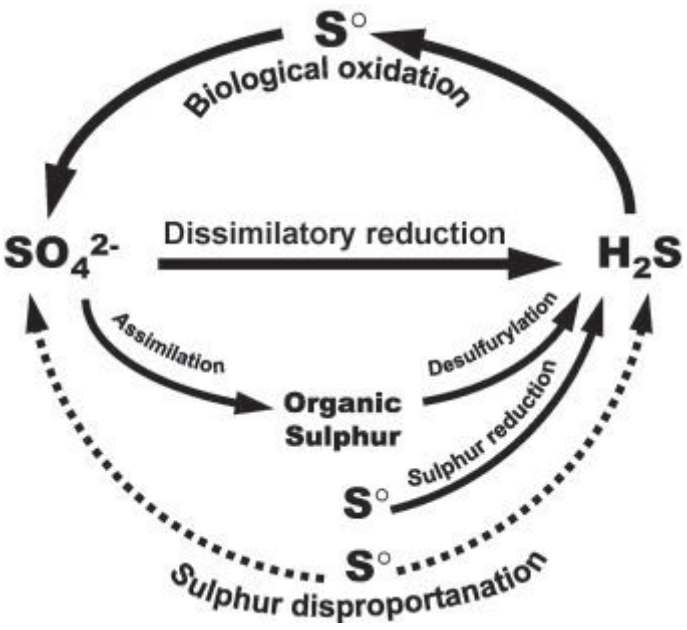

Figure 1.3: Simplified overview of the sulfur cycle (from Tang *et al.* (2009); reproduced with permission © Elsevier[2]).

[2] License Number: 4135990839081; Order date: Jun 25, 2017

Under oxic conditions, oxidation of sulfide and elemental sulfur (S^0) is catalyzed by sulfur-oxidizing chemolithotrophic bacteria; however, anoxic oxidation of sulfide can also occur by phototrophic purple and green sulfur bacteria (Madigan *et al.*, 2003). The anaerobic respiration of sulfate (SO_4^{2-}) to hydrogen sulfide (HS^-) is referred to as dissimilatory sulfate reduction; due to negative redox potential of this reaction ($E_0' = -0.22$ V), i.e. the less favourable energy balance, the growth of an organism using sulfate is slower than that of organisms using oxygen or nitrate (Madigan *et al.*, 2003). Severe hypoxia and total lack of oxygen (anoxia) can lead to the production of hydrogen sulfide, resulting from anaerobic mineralization of organic matter by sulfate-reducing bacteria (Grote *et al.*, 2012). However, chemolithotrophic oxidation of sulfide with nitrate can be observed in OMZs, thus leading to the detoxification of sulfidic waters (Lavik *et al.*, 2009).

1.2 Oxygen deficiency in marine environments

Oxygen concentration in marine environments is one of the environmental variables that have been largely affected by anthropogenic activities, such as eutrophication (Diaz, 2001). The decline of oxygen concentration has induced seasonal hypoxic events in areas where this was not reported previously, thus affecting benthic organisms and fisheries. These events are mainly driven by water column stratification and decomposition of organic matter in the sediment (Diaz, 2001).

Hypoxia has been traditionally defined based on the level of oxygen below which most animal life cannot be sustained (Diaz, 2001). As a general perception, hypoxia occurs where oxygen concentration is less than 2 mg/l (or 63 µM or 1.4 ml/l) (Middelburg and Levin, 2009), although the existence of a single universal threshold cannot sufficiently reflect the metabolic needs of all the different macroorganisms (Dimitriou *et al.*, 2015). Alternative states of oxygen concentration in the water column have also been used in the literature, such as the oxic, dysoxic (20–90 µmol/kg) and suboxic (1–20 µmol/kg) states (Wright *et al.*, 2012). It is clear that the thresholds used to define oxygen conditions are quite inconsistent; for example, the term suboxic has been proposed to be used in conditions where oxygen is no longer

present (Yakushev and Newton, 2012), although this state was traditionally referred to as anoxia (Diaz, 2001) and despite the fact that others proposed the term suboxic to be characterized as obsolete (Canfield and Thamdrup, 2009).

Hypoxia, apart from being human-induced, can be naturally occurring or it can be a result of the combination of natural and anthropogenic processes (Middelburg and Levin, 2009). Independently of its occurrence, modifications in hydrology and global warming will enhance hypoxic incidents, and their subsequent adverse effect on survival and functioning of organisms (Middelburg and Levin, 2009), since increase in water temperature results in decrease of oxygen solubility (Wright *et al.*, 2012). Except from the obvious impact on macroorganisms, the impact on microbial composition and activity cannot be neglected; anaerobic reactions and pathways are favoured, at the expense of the aerobic ones. Hypoxia alters the availability of electron acceptors in the water column, which subsequently affects the oxidative pathways of microbial metabolism in the sediments; an example of a group that is expected to be affected by hypoxic conditions includes the chemolithoautotrophic microbes (Lipsewers *et al.*, 2017).

Naturally occurring hypoxia can be found in areas where water circulation is restricted, such as fjords (Nordberg *et al.*, 2001), estuaries (Abell *et al.*, 2011) and coastal lagoons (Kristiansen *et al.*, 2002). Two of the most pronounced examples of hypoxic ecosystems with restricted water circulation are the Baltic Sea (Conley *et al.,* 2002) and the Black Sea (Kuypers *et al.,* 2003). Black Sea has been characterized as the largest anoxic basin of the world, with high ammonium concentration in the suboxic zone (Kuypers *et al.,* 2003). Baltic Sea on the other hand, is an example of a semi-enclosed basin where hypoxia has been enhanced by cultural eutrophication (Conley *et al.,* 2002). In addition, naturally occurring hypoxia can be developed in coastal upwelling regions where oxygen-depleted water is driven into the continental shelf (Monteiro *et al.*, 2008; Roegner *et al.*, 2011). Microbial communities inhabiting oxygen deficient environments have been found to have similar composition (Wright *et al.*, 2012), although each of these environments is characterized by distinct features.

The study sites that were chosen for the PhD project (coastal lagoons and coastal upwelling region) are naturally hypoxic environments of high economic value, since they host a variety of anthropogenic activities.

1.2.1 Coastal lagoons: biogeochemistry and diversity of microorganisms

Coastal lagoons are highly dynamic and extremely unpredictable aquatic systems due to the fluctuation of their environmental variables (Barnes, 1980), while being particularly vulnerable to human activities (Guelorget and Perthuisot, 1992; Benlloch *et al.*, 1995). This is due to confinement, i.e. the time of renewal of the elements of marine origin at any given point (Guelorget and Perthuisot, 1983), and their shallow depth (Bellan, 1972; Guelorget and Perthuisot, 1992). In addition, coastal lagoons are ecosystems of great economic value, because of their high productivity (Knoppers, 1994) which leads to their intensive exploitation through aquaculture and fisheries activities (Pérez-Ruzafa and Marcos, 2012). Furthermore, human induced eutrophication can occur in coastal lagoons due to the excessive use of fertilizers for agricultural purposes in their watersheds and drainage basins (Cloern, 2001).

Apart from human interference, lagoons are also susceptible to changes induced by natural events, such as storms and river flooding (Avramidis *et al.*, 2013). The structure and function of lagoonal ecosystems is largely affected by the input of organic matter from land, the marine environment and the atmosphere (Reizopoulou *et al.*, 1996; Arvanitidis *et al.*, 1999; Viaroli *et al.*, 2008; Tagliapietra *et al.*, 2009) and its subsequent decomposition and removal (Tagliapietra *et al.*, 2012).

Lagoons may suffer dystrophic crises due to the accumulation of excessive concentrations of organic matter and subsequent increase in microbial heterotrophic activities, leading to consumption of the dissolved oxygen (Guelorget and Perthuisot, 1992; Avramidis *et al.*, 2013). Hypoxia, or even anoxia in certain cases, induced by the increased rate of microbial respiration, leads to sulfate reduction and subsequent accumulation of hydrogen sulfide (Danovaro and Pusceddu, 2007).

Microbial communities in lagoonal environments are different compared to freshwater and marine environments (Ghai *et al.*, 2012); although they are located in an intermediate salinity regime with pronounced temporal variations which poses a significant stress in biota, which is added to the stress induced by hypoxia, coastal lagoons exhibit high microbial diversity (Danovaro and Pusceddu, 2007). Apart from salinity, nutrient availability (Segnini de B *et al.*, 2015) and level of eutrophication (de Wit, 2008) seem to influence microbial activity in lagoonal sediments.

For the purpose of this PhD project, the lagoonal complex of Amvrakikos Gulf was chosen as a study site, due to its characteristics and since previous knowledge on the ecosystem was available.

1.2.1.1 Amvrakikos Gulf

Amvrakikos Gulf is a semi-enclosed embayment located in the northwestern part of Greece (Ionian Sea, Western Greece) (Figure 1.4). It is about 35 km long and 15 km wide with a maximum depth of 65 m (Ferentinos *et al.*, 2010), and it is one of the largest semi-enclosed gulfs in the Mediterranean Sea. The wetlands of Amvrakikos Gulf, i.e. lagoons, reed-beds and marshes (Avramidis *et al.*, 2013), are protected by the international Ramsar convention and included in the Natura 2000 network (Vasileiadou *et al.*, 2016). The Gulf is connected with the Ionian Sea via a narrow channel, the Preveza (Aktio) Strait, of minimum 600 m width, 3 km length and 8.5 m depth (Kapsimalis *et al.*, 2005). Recent sedimentation patterns in the Gulf are related to the terrigenous inputs primarily from the Arachthos River, located in the northeast of the Gulf, and secondarily from the Louros River, located in the northwest (Poulos *et al.*, 1993, 1995; Kapsimalis *et al.*, 2005).

Oxygen depletion events are not unusual for Amvrakikos Gulf, as hypoxic and anoxic conditions have been recorded since the Holocene both due to natural geologic processes (Ferentinos *et al.*, 2010), as well as due to human activities (Avramidis *et al.*, 2013). In addition, H_2S and CH_4 gas charged sediments and eutrophication have been recorded in the Gulf (Papatheodorou *et al.*, 1993; Ferentinos *et al.*, 2010; Kountoura and Zacharias, 2011). About 20% of the total area of the Gulf is

characterized by dysoxic conditions, while about 30% is characterised by anoxic conditions (Ferentinos *et al.*, 2010).

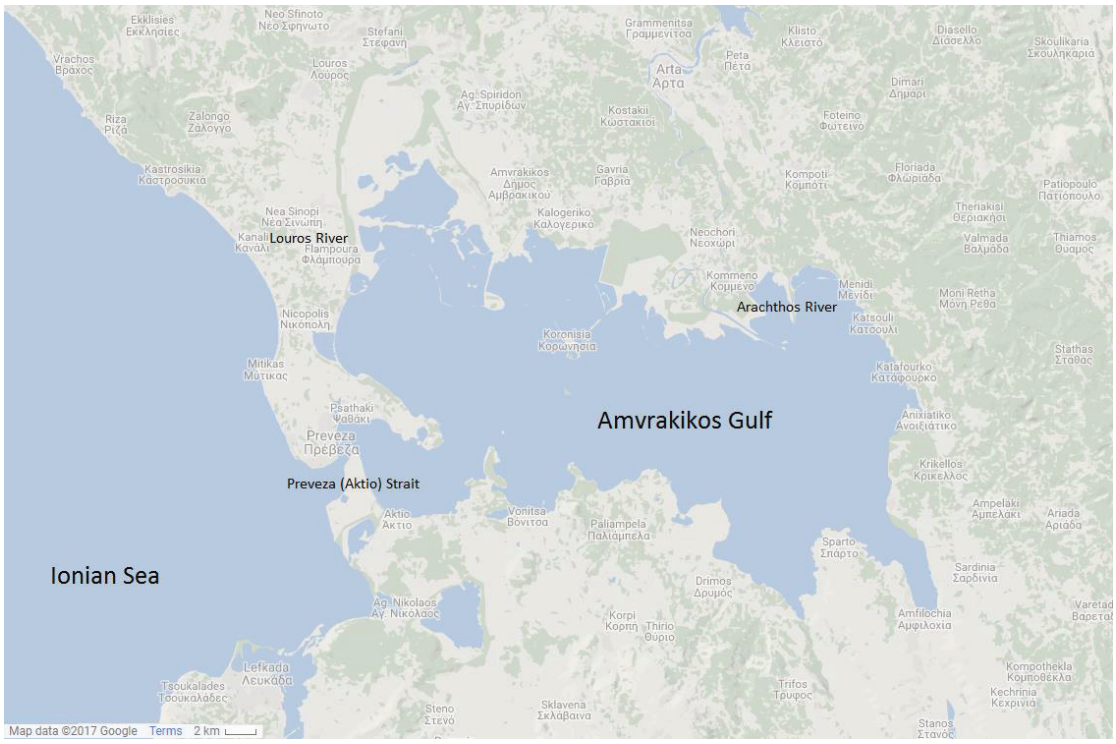

Figure 1.4: Map of Amvrakikos Gulf, showing the location of the two rivers and of the channel connecting the Gulf to the Ionian Sea (Map data: Google).

1.2.2 Oxygen minimum zones: biogeochemistry and diversity of microorganisms

Coastal upwelling regions are another type of environment limited in oxygen, where also formation of OMZs has been reported. Exceptional physicochemical conditions, such as reduced ventilation and stagnant circulation, lead to the reduction of bottom water oxygen concentration and to the formation of OMZs (Cassman *et al.*, 2012). OMZs are formed when the respiratory oxygen demand, during organic matter degradation, exceeds its availability (Wright *et al.*, 2012). Although there is no exact definition of an OMZ, the upper threshold of 20 µM of oxygen has been proposed; as such, the global ocean volume that can be characterized as OMZ is about 1% of the total ocean volume (Lam and Kuypers, 2011).

Sediment in upwelling regions is rich in organic matter and bottom water is often depleted in oxygen because of intense heterotrophic respiration, thus leading to high sulfide production rates (Ferdelman *et al.*, 1997). As the second-most favorable electron acceptor (after oxygen), nitrate may be used for the oxidation of sulfide, which results in a close coupling of the sulfur and the nitrogen cycles in such areas (Schulz *et al.*, 1999). As oxygen concentration decreases, other terminal electron acceptors are used sequentially, starting with manganese, iron, sulfate and ending with carbon dioxide (Wright *et al.*, 2012). OMZs occur in the Pacific Ocean, the Atlantic Ocean and the Arabian Sea and they are distinct between the different oceans; for example, they are more voluminous in the Pacific (Wright *et al.*, 2012), but in the other two oceans they are more oxygen deficient (Estrada and Marrasé, 1987).

Coastal upwelling systems react directly to changes in external (climatic) forcing (Lahajnar, 2011) and hydrodynamic conditions influence directly the supply of oxygen to the continental shelf (e.g. Emeis *et al.*, 2009; Leduc *et al.*, 2010). Fluctuating oxygen levels over the continental shelf have significant consequences for nutrient levels, for rates of exchange at the sediment-water interface, for gas exchange between the ocean and the atmosphere, and for biological production (e.g. Bakun *et al.*, 2010; Finney *et al.*, 2010).

It has been shown that taxonomic, phylogenetic and functional diversity of microorganisms declines with the reduction of oxygen concentration, i.e. from the oxygen-rich surface water to the OMZ (Bryant *et al.*, 2012). Microbial communities in OMZs are considered unique, since they use alternative electron donors and acceptors for respiration (Lam and Kuypers, 2011; Cassman *et al.*, 2012). Globally, OMZs and in particular heterotrophic denitrification occurring in these areas are responsible for 30–50% of the total nitrogen loss (Gruber and Sarmiento, 1997; Codispoti *et al.*, 2001). However, the extremely low concentration of ammonium indicate that anammox, i.e. anaerobic ammonium oxidation, can be responsible for massive losses of fixed nitrogen as gaseous N_2 from OMZ waters (Kuypers *et al.*, 2005). Physicochemical conditions prevailing in OMZs allow for the completion of nitrogen cycling, leading to subsequent nitrogen depletion (Lam and Kuypers, 2011). In addition, an active coupling between nitrogen and sulfur cycles has been

suggested to exist in nitrate-rich OMZs, with the latter being directly linked to anammox and other nitrogen cycling processes (Canfield *et al.*, 2010). Apart from the important role of OMZs in the global nitrogen cycle, the existence of microbial communities involved in methane production and oxidation in OMZs has been investigated, without any conclusive evidence for such an existence so far (Lüke *et al.*, 2016).

For the purpose of this PhD project, the chosen coastal upwelling system was the Benguela system off Namibia, one of the two main coastal upwelling systems of the southern hemisphere (Moloney *et al.*, 2005) and one of the five coastal upwelling regions globally (Summerhayes, 2015).

1.2.2.1 Benguela coastal upwelling system

The Benguela coastal upwelling system is situated along the coast of south western Africa, with its south boundary located east of the Cape of Good Hope and the north reaching Angola waters, thus it is encompassing the full extent of Namibia's marine environment (Shannon and O'Toole, 2003) (Figure 1.5).

The Benguela Current flows equatorward along the western coast of southern Africa forming the eastern boundary current segment of the subtropical gyre of the South Atlantic Ocean (Weeks *et al.*, 2004). The Benguela system off Namibia is the most productive of the subtropical eastern boundary current regions (Carr, 2001), i.e. Peru-Humboldt, California and Canary, although this productivity is not referring to the low fish catch yields (Lavik *et al.*, 2009) which have been attributed to the episodic occurrence of hydrogen sulfide gas (Brüchert *et al.*, 2009), a potent respiratory toxin to aquatic organisms (Bagarinao, 1992).

Upwelling occurs all along the coast of Namibia, but it is most intense, irrespective of the season, between 26^{o} and 25^{o} S (Brongersma-Sanders, 1983). The intensity decreases with latitude, particularly north of Walvis Bay (23^{o} S), where upwelling is only important in winter and early spring and the intensity of the process decreases even to insignificance in the warmer time of the year (Stander, 1964). The principal upwelling centre is situated near Lüderitz in southern Namibia and it creates a natural internal divide within the Benguela, with the domains to the north and south of it functioning rather differently (Shannon and O'Toole, 2003). High fluidity and instability of sediments is characterizing the upwelling locations (Schulz *et al.*, 1999).
In the sediments underneath the coastal upwelling system off central Namibia, bacterial sulfate reduction rates near the sediment surface exceed the capacity of the sediment to oxidize and precipitate sulfide; this leads to a net hydrogen sulfide flux across the sediment-water interface (Brüchert *et al.*, 2003). In addition, episodic advective transport of sulfide from the methanogenic zone has been suggested to contribute to the water column concentration of hydrogen sulfide (Brüchert *et al.*, 2003).

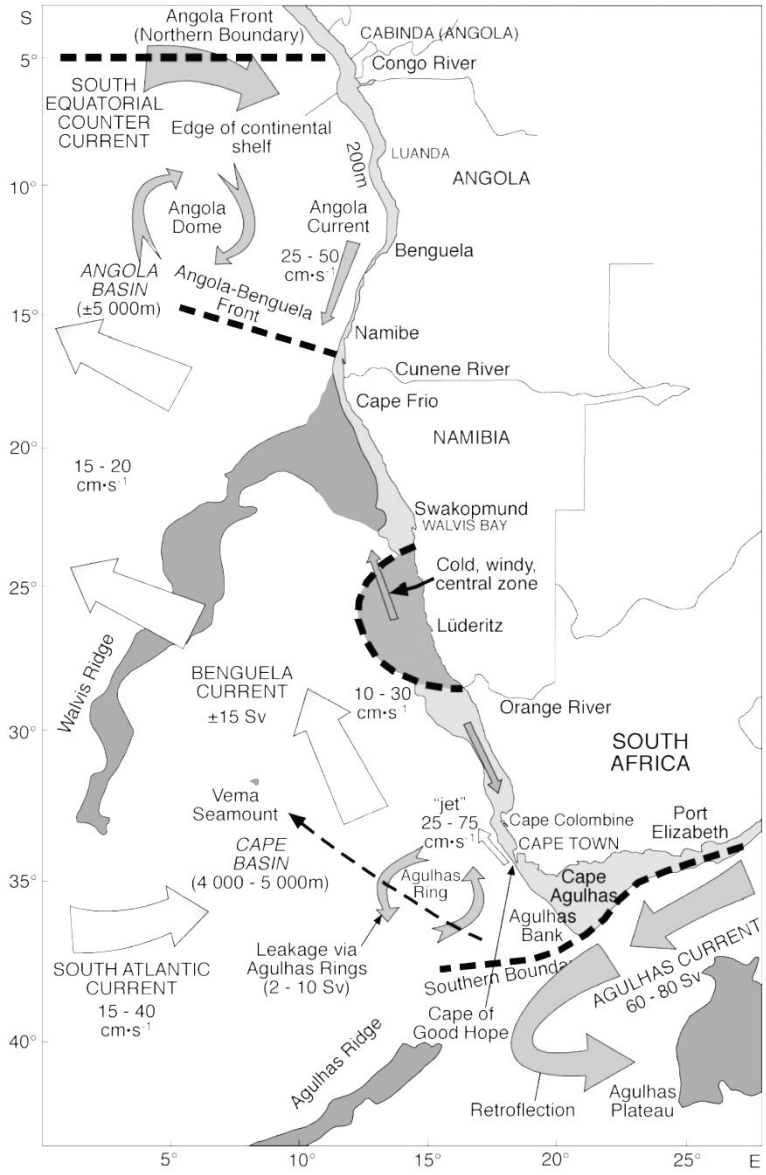

Figure 1.5: Boundaries of the Benguela Current Large Marine Ecosystem, bathymetric features and surface (upper layer) currents (adapted from Shannon and O'Toole (2003)). (1 Sv = 10^6 m^3 s^{-1}).

1.3 Identification and characterization of microbial communities

Traditionally, microbial ecology was based on enrichment cultures and isolation in monocultures (e.g. Gich *et al.*, 2005; Plante *et al.*, 2008). However, only the minority of microorganisms can be cultured using the media that have been developed so far (Glöckner and Joint, 2010). This led to the development of different methods based on molecular analysis of microbial communities (DeLong and Pace, 2001); the state-

of-the-art method for years was cloning and subsequent Sanger sequencing of the 16S rRNA gene (e.g. Polymenakou *et al.*, 2009). Similar methodologies were also being used complementary to cloning, such as denaturing gradient gel electrophoresis (DGGE) (Muyzer, 1999), terminal restriction fragment length polymorphisms (tRFLPs) (Liu *et al.*, 1997), automated ribosomal intergenic spacer analysis (ARISA) (e.g. Ranjard *et al.*, 2001) and others. However, the aforementioned methods were providing an overview of the microbial community profile in a non-qualitative way.

The "gap" in the molecular analysis techniques was filled in by the advent of pyrosequencing (Ronaghi *et al.*, 1998), which allowed high-throughput sequencing, initially of the 16S rRNA gene but also of protein coding genes, and characterization of microbial communities. Despite the power of pyrosequencing and its establishment at the state-of-the-art for the next few decades, this power came with a price which was the introduction of sequencing biases and the reduction of the sequencing length. In order to deal with the previously mentioned issues, bioinformatic pipelines were developed and standardized for the sequence processing (Xu, 2006; Scholz *et al.*, 2012).

Although the gain of information that emerged from pyrosequencing cannot be disputed, the rate of biodiscovery was not linearly correlated with that gain as pure cultures are a prerequisite for the advancement of marine biodiscovery (Joint *et al.*, 2010). The need for development of more elaborate methodologies that would permit the *in situ* assessment of microbial functioning, led to the emergence of stable isotope probing (SIP) techniques (Radajewski *et al.*, 2003). The SIP approach can provide a direct link between biogeochemical processes and the identity of microbial populations that catalyze those processes (Kreuzer-Martin, 2007). It is based on (i) the incorporation of a labelled substrate, e.g. ^{13}C, ^{15}N or ^{18}O, into microbial biomarkers, such as nucleic acids (Whiteley *et al.*, 2006), proteins (Jehmlich *et al.*, 2008) and phospholipid fatty acids (Murase *et al.*, 2011), (ii) the separation of the labelled and unlabelled biomarker by density gradient centrifugation, and (iii) the subsequent molecular characterization of microbial communities from the separated template (Lueders *et al.*, 2016).

SIP can characterize microbes based on their potential to respire and utilize an added substrate. One of the most commonly used labelled substrates, which is most difficult to ferment for thermodynamic reasons and therefore an ideal substrate for identifying anaerobic respiratory microbes, is acetate (CH_3COO^-) (Gutierrez-Zamora and Manefield, 2010; Hori *et al.*, 2010). Although acetate is one of the most important metabolites in anoxic environments, certain sulfate-reducing bacteria are not capable of acetate oxidation (e.g. *Desulfovibrio*, *Desulfobulbus*, *Desulfomicrobium* and others), thus they use lactic acid, pyruvate, ethanol and some fatty acids as electron donors. However, many sulfate-reducing bacteria can oxidize acetate (e.g. *Desulfobacter*, *Desulfobacterium*, *Desulfosarcina*, *Desulfococcus* and others) and, in fact, they can oxidize it completely to CO_2 (Madigan *et al.*, 2003). In addition, acetate has been widely used as a carbon source for the identification of denitrifying bacteria (Ginige *et al.*, 2005; Osaka *et al.*, 2006), and it has been compared with other carbon substrates such as propionate, butyrate (Paul *et al.*, 1989) and methanol (Hallin *et al.*, 1996).

1.4 Aims and objectives

The importance of hypoxia on the biogeochemical cycling of crucial elements, such as nitrogen and sulfur, has been presented in the previous sections. Therefore, it is evident that, since microbes mediate the majority of the transformations of these elements, understaning of microbial diversity and function in hypoxic ecosystems is crucial. Moreover, since sediments act as sinks for nitrogen and sulfur, undoubtedly the focus should be drawn in the processes occurring in hypoxic sediments.

The aim of this PhD project was to study the microbial community assemblages in sediments of hypoxic ecosystems and to identify a potential link between their identity and function, with a particular emphasis on the microorganisms involved in the nitrogen and sulfur cycles. The methodology that was applied included targeted metagenomics, i.e. 16S rRNA and *dsrB* sequencing, and RNA stable isotope probing.

In chapter 2, microbial communities of a gradient river-lagoon-open sea were studied and the relationship of those communities with salinity was investigated,

both in terms of diversity as well as in terms of function. The tested hypothesis was that microbial communities in the three different habitats would change according to the main environmental variable characterizing such a gradient, which is salinity.

In chapter 3, sulfate-reducing microorganismic communities were assessed in a Mediterranean lagoonal complex. The tested hypothesis was that sulfate-reducing microorganismic communities would be different across the complex since there are distinct environmental conditions prevailing in each of the tested lagoons.

In chapter 4, the existence and activity of sulfate-reducing and denitrifying microorganisms was investigated in the Namibian oxygen minimum zone. The tested hypothesis was that ^{13}C labelling would characterize the communities actively reducing nitrate and sulfate and that it would produce different microbial community patterns, based on the different sampling stations as well as the chosen terminal electron acceptors.

In chapter 5, an overview of tools developed for analyzing next-generation sequencing data derived from biodiversity studies was provided, along with some general recommendations on the employment of metagenomics. In the previous chapters, different pipelines and algorithms were used in each case. This chapter, on metagenomics methods and data analyses, builds in the different methodologies that were applied, while presenting and discussing the usage of other methods for analyses of metagenomic data.

Finally, chapter 6 includes the general discussion and conclusions that have derived from this research.

List of manuscripts and contribution of authors

Sediment microbial taxonomic and functional diversity in a natural salinity gradient challenge Remane's "species minimum" concept (Chapter 2)

Christina Pavloudi, Jon Bent Kristoffersen, Anastasis Oulas, Marleen De Troch, Christos Arvanitidis

Journal: *PeerJ*, Status: *Manuscript accepted for publication*

Author contributions: **CP** conceived the idea, designed and performed the experiment, produced and analyzed the data and wrote the manuscript. JBK assisted with the choice of the library preparation protocol. JBK, AO, MDT and CA made critical revisions and approved the final version of the manuscript.

Contribution of **CP**: Experimental concept and design: 100%; Experimental work: 90%; Data analysis and interpretation: 100%; Preparation of figures and tables: 100%; Writing of the manuscript: 90%.

Diversity and abundance of sulfate-reducing microorganisms in a Mediterranean lagoonal complex (Amvrakikos Gulf, Ionian Sea) derived from dsrB gene (Chapter 3)

Christina Pavloudi, Anastasis Oulas, Katerina Vasileiadou, Georgios Kotoulas, Marleen De Troch, Michael W. Friedrich, Christos Arvanitidis

Journal: *Aquatic Microbial Ecology*, Status: *Manuscript published (doi: 10.3354/ame01829)*

Author contributions: **CP** conceived the idea, designed and performed the experiment, produced and analyzed the data and wrote the manuscript. AO assisted with data analysis. KV assisted on the sampling and the experiment. AO, KV, GK, MDT, MWF and CA made critical revisions and approved the final version of the manuscript.

Contribution of **CP**: Experimental concept and design: 100%; Experimental work: 90%; Data analysis and interpretation: 90%; Preparation of figures and tables: 100%; Writing of the manuscript: 90%.

Identification of microbial communities by RNA stable isotope probing and 16S rRNA sequencing from the Benguela coastal upwelling system (Chapter 4)

Christina Pavloudi, Saar Y. Sztejrenszus, Jon Bent Kristoffersen, Niko Lahajnar, Marleen De Troch, Christos Arvanitidis, Michael W. Friedrich

Journal: *Environmental Microbiology*, Status: *Manuscript submitted (under review)*

Author contributions: **CP**, SYS and MWF conceived the idea and designed the experiment. **CP** performed the experiment, produced and analyzed the data and wrote the manuscript. JBK assisted with the choice of the library preparation protocol. NL assisted during the sampling cruise and provided the environmental data. SYS, JBK, NL, MDT, CA and MWF made critical revisions and approved the final version of the manuscript.

Contribution of **CP**: Experimental concept and design: 70%; Experimental work: 90%; Data analysis and interpretation: 90%; Preparation of figures and tables: 100%; Writing of the manuscript: 80%.

Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies (Chapter 5)

Anastasis Oulas, **Christina Pavloudi**, Paraskevi Polymenakou, Georgios A. Pavlopoulos, Nikolas Papanikolaou, Georgios Kotoulas, Christos Arvanitidis, Ioannis Iliopoulos

Journal: *Bioinformatics and Biology Insights*, Status: *Manuscript published (doi: 10.4137/BBI.S12462)*

Author contributions: AO, GAP and II conceived the idea of the manuscript. AO and **CP** wrote the first draft of the manuscript. **CP** was responsible for the sections referring to sequencing technologies and marker gene metagenomics. GAP, II, NP, PP, GK and CA made critical revisions and approved the final version of the manuscript.

Contribution of **CP**: Experimental concept and design: Not available; Experimental work: Not available; Data analysis and interpretation: Not available; Preparation of figures and tables: 50%; Writing of the manuscript: 50%.

Chapter 2: Sediment microbial taxonomic and functional diversity in a natural salinity gradient challenge Remane's "species minimum" concept

Citation: Pavloudi C, Kristoffersen JB, Oulas A, De Troch M, Arvanitidis C (2017) Sediment microbial taxonomic and functional diversity in a natural salinity gradient challenge Remane's "species minimum" concept. *PeerJ,* Accepted for publication[3]

Abstract

Several models have been developed for the description of diversity in estuaries and other brackish habitats, with the most recognized being Remane's Artenminimum ("species minimum") concept. It was developed for the Baltic Sea, one of the world's largest semi-enclosed brackish water bodies with a unique permanent salinity gradient, and it argues that taxonomic diversity of macrobenthic organisms is lowest within the horohalinicum (5 to 8 psu).

The aim of the present study was to investigate the relationship between salinity and sediment microbial diversity at a freshwater-marine transect in Amvrakikos Gulf (Ionian Sea, Western Greece) and assess whether species composition and community function follow a generalized concept such as Remane's.

DNA was extracted from sediment samples from six stations along the aforementioned transect and sequenced for the 16S rRNA gene using high-throughput sequencing. The metabolic functions of the OTUs were predicted and the most abundant metabolic pathways were extracted. Key abiotic variables, i.e. salinity, temperature, chlorophyll-a and oxygen concentration etc., were measured and their relation with diversity and functional patterns was explored.

Microbial communities were found to differ in the three habitats examined (river, lagoon and sea) with certain taxonomic groups being more abundant in the freshwater and less in the marine environment, and *vice versa*. Salinity was the

[3] Modified version of the accepted manuscript.

environmental factor with the highest correlation to the microbial community pattern, while oxygen concentration was highly correlated to the metabolic functional pattern. The total number of OTUs showed a negative relationship with increasing salinity, thus the sediment microbial OTUs in this study area do not follow Remane's concept.

2.1 Introduction

Salinity is considered as one the most influential environmental factors, not only for the distribution of benthic and pelagic organisms (Remane, 1934; Rolston and Dittmann, 2009; Palmer *et al.*, 2011; Darr *et al.*, 2014), but also for microbial community composition (e.g. Barcina *et al.,* 2006; Wu *et al.,* 2006; Lozupone and Knight, 2007; Logares *et al.,* 2009). Bacterial abundance, activity and growth can be affected by salinity (Ben-Dov *et al.*, 2007; Caporaso *et al.*, 2011) and, in certain cases salinity can induce mortality of bacteria, thus regulating bacterial abundance in some estuaries (Painchaud *et al.*, 1995). Specifically, salinity fluctuations, and their subsequent effect on aquatic biota, are more noticeable in estuaries and other brackish water bodies, as these habitats are characterized by a more or less pronounced salinity gradient (Telesh *et al.*, 2013).

It has been suggested that in brackish water ecosystems, taxonomic diversity of macrobenthic organisms is lowest within the horohalinicum, which occurs at salinity 5 to 8 psu, because the number of brackish specialist species does not compensate for the decline of the marine and freshwater species richness (Remane, 1934). This concept, referred to as the Remane's Artenminimum ('species minimum') concept originated from the Baltic Sea, one of the world's largest semi-enclosed, brackish water bodies with a unique permanent salinity gradient (Telesh *et al.*, 2011). Despite being developed for the Baltic Sea, Remane's concept became the recognized model for the description of diversity in estuaries and other brackish habitats (McLusky and Elliott, 2004). However, alternative models challenging Remane's concept have also been developed. In certain cases a reverse curve has been observed, with the peak of species occurring in the horohalinicum (Telesh *et al.*, 2011) while in others a linear

decrease (Attrill, 2002) or even no change (Herlemann *et al.*, 2011) in the number of species across the salinity gradient were observed.

Remane's concept can be projected in other aquatic bodies with similar evolution to the Baltic Sea, such as the Amvrakikos Gulf (Ionian Sea, Western Greece) (Ferentinos *et al.*, 2010). The Gulf was formed in the Middle Quaternary period (Anastasakis *et al.*, 2007); the marine transgression took place at approximately 11 ka BP and the Gulf attained its present shape at approximately 4 ka BP (Kapsimalis *et al.*, 2005). In addition, the low tidal range (on average 5 cm) and the low energy wave regime prevailing in Amvrakikos Gulf (Ferentinos *et al.*, 2010), render the latter as a Baltic Sea analogue in the Mediterranean Sea.

In the light of projected climate changes, and the subsequent sea level rise and saltwater intrusion that will occur, microbial populations in freshwater wetlands near the coast will be subjected to elevated salinities (Chambers *et al.*, 2011). Due to the long-term effect of sea level rise, saltwater intrusion can affect ecosystems on timescales of decades (Neubauer *et al.*, 2013). It is therefore crucial to explore the current status of microbial communities in wetlands in order to comprehend the impact of such acute changes in their diversity patterns, since they are involved in biogeochemical processes that are crucial for maintaining the planet in a habitable state (e.g. Falkowski *et al.,* 2008).

Currently, state-of-the-art studies of microbial communities through high throughput sequencing of the 16S rRNA gene, i.e. marker gene metagenomics (Oulas *et al.*, 2015), allow documentation of the high diversity of microbial communities in a variety of habitats, e.g. wetlands (Yu Wang *et al.*, 2012; Jiang *et al.*, 2013), estuaries (Bobrova *et al.*, 2016), lakes (Zhang *et al.*, 2015), rivers (Staley *et al.*, 2013) and coastal lagoons (Highton *et al.*, 2016). The substantial data derived from such techniques can be used to test ecological hypotheses, on which sound conclusions can be drawn; for example if microorganisms have a biogeography (Zinger *et al.*, 2011) or if their biodiversity is driven by changes in elevation (Fierer *et al.*, 2011).

The aim of the present study was to investigate the sediment prokaryotic diversity along a transect river-lagoon-open sea, i.e. from freshwater to marine, in Amvrakikos Gulf in order to test whether it follows a generalized concept, such as Remane's, both in terms of Operational Taxonomic Unit (OTU) composition as well as on

distribution of metabolic functions. If the applicability of the concept is confirmed in Amvrakikos Gulf, it would enhance its transferability to other brackish water bodies than the Baltic Sea. In addition, confirmation of Remane's concept for prokaryotes would mean that small-bodied, fast-developing and rapidly evolving microbes respond in a similar way as benthic macroorganisms in a salinity gradient.

2.2 Materials and Methods

2.2.1 Field sampling

The transect river-lagoon-open sea, i.e. from freshwater to marine, chosen for the present study is naturally occurring at Amvrakikos Gulf (Ionian Sea, Western Greece) (Figure 2.1). The Gulf is a semi-enclosed embayment connected with the Ionian Sea via a narrow channel, the Preveza (Aktio) Strait (Kapsimalis *et al.*, 2005) and is characterized by a fjord-like oceanographic regime (Ferentinos *et al.*, 2010). The wetlands of Amvrakikos Gulf are listed in both the Ramsar international convention and the Natura 2000 Network.

The northwest part of the Amvrakikos Gulf is formed by the rivers Arachthos and Louros (Poulos *et al.*, 1993, 1995; Vasileiadou *et al.*, 2012). Arachthos river is the main source of freshwater from November to April (Poulos *et al.*, 1993) and its flow is controlled by two hydroelectric dams: the first being an earthfill dam, of 107 m height, mainly built for flood control and the second being a concrete dam, of 42 m height, primarily having a regulatory role, that is to ensure the constant flow of water throughout the year.

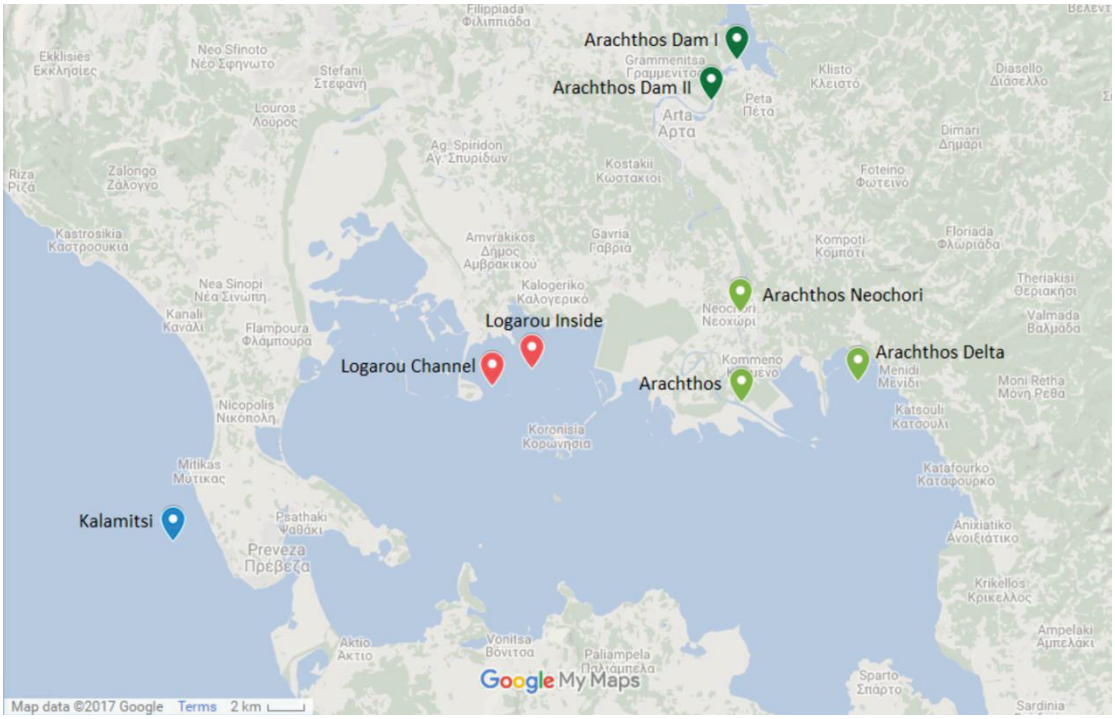

Figure 2.1: Map showing the location of the six sampling stations and the location of the river dams (Map data: Google).

Two stations were chosen at the Arachthos river, thus representing the freshwater conditions: one being close to its mouth (39.029690 N, 21.025880 E) and one in the upper limit of saltwater intrusion when the dams are closed, closed to the village of Neochori (39.070770 N, 21.025060 E). One more station was chosen at the Arachthos delta (39.039199 N, 21.094200 E), as an intermediate between the freshwater and the marine realm. In addition, two stations were chosen in the Logarou lagoon, due to its vicinity to Arachthos, representing the brackish stations: one at the inner part of the lagoon (39.045528 N, 20.902283 E), closer to the terrestrial end, and one near the channel connecting the lagoon to the Gulf (39.037458 N, 20.878626 E). Finally, Kalamitsi station (38.966250 N, 20.690783 E) was chosen from the marine realm, i.e. outside of the Amvrakikos Gulf.

Sampling was carried out in winter of 2014 (November - December), as described in Pavloudi et al. (2016). For the Logarou lagoon and Arachthos river stations, salinity, water temperature and dissolved oxygen concentration were measured in the water overlaying the sediments by means of a portable multi-parameter (WTW Multi 3420 SET G). For the Arachthos delta and Kalamitsi station, water abiotic variables

(fluorescence, salinity, water temperature and dissolved oxygen concentration) were recorded with a Sea-Bird Electronics 25 CTD probe. Fluorescence was regarded as a proxy for chlorophyll-a concentration.

Sediment samples from the Logarou lagoon and the Arachthos river were collected by means of a modified manually-operated box-corer, with sampling surface 156.25 square centimeters and sediment penetration depth of 25 centimeters, deployed from fishing boats specifically used at the lagoons. Samples from the Arachthos delta and Kalamitsi stations were collected with a Smith McIntyre Grab operated from the R/V Philia. The permission to conduct the field study was provided by the Amvrakikos Wetlands Management Body.

Cylindrical sampling corers (internal sampling surface 15.90 square centimeters) were placed inside the box-corer and the Smith McIntyre Grab in order to collect sub-samples of the sediment's upper layer (0-2 cm). Three replicate units, each retrieved from a different box-corer to avoid pseudoreplication, were taken for each analysis from each sampling station, in order to determine variability within and between stations. Samples for molecular analysis (about 15 cm^3 each), i.e. DNA extractions, were placed in 50 ml falcon tubes (Sarstedt, Nümbrecht, Germany) and were stored at -20 ^oC, until further processing in the laboratory.

Samples were also collected from cylindrical corers for the measurement of the Particulate Organic Carbon (POC), chloroplast pigments concentration (chlorophyll-a and phaeopigments) and sediment granulometry (for the latter, the sampling depth was four centimeters to allow comparison with previously published data on the study area). To quantify chloroplast pigments concentration in the water column, three replicate water samples were collected from the Logarou lagoon and the Arachthos river by means of Niskin bottles (5 lt). The aforementioned samples were processed at the Chemistry Lab of the IMBBC (HCMR), based on standard techniques (chloroplast pigments: Yentsch and Menzel, 1963; POC: Hedges and Stern, 1984; granulometry: Gray and Elliott, 2009). Sampling was conducted in winter (November-December) of 2014.

2.2.2 DNA extraction, PCR amplification and 16S rRNA sequencing

DNA was extracted using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA), as recommended by the manufacturer. About 0.5 (± 0.2) grams of wet sediment were used from each sample and the quality of the extracted DNA was evaluated by gel electrophoresis.

PCR amplification was performed targeting the V3-V4 region of the 16S rRNA gene using the bacterial primer pair S-DBact-0341-b-S-17 (or 341F) and S-D-Bact-0785-a-A-21-B (or 805RB), which has been referred to as the most promising primer pair (Herlemann *et al.*, 2011; Klindworth *et al.*, 2013), with a revision for detection of SAR11 bacterioplankton (Apprill *et al.*, 2015).

The Two-Step PCR Approach was used for this study. The first-step PCR was performed with the aforementioned primers containing a universal 5' tail as specified in the Nextera library protocol from Illumina. The amplification reaction mix of the first PCR contained 6 µl 5x KAPA HiFi Fidelity buffer, 0.9 µl BSA (2 µg/µl), 0.75 µl KAPA dNTP Mix (10 mM), 1.5 µl from each primer (10 µM), 0.75 µl KAPA HiFi HotStart DNA polymerase (1 U/µl) in a final volume of 30 µl per reaction. DNA template concentration was about 10 ng/µl. The first PCR protocol used was the following: 95 ^{o}C for 5 minutes; 25 cycles at 98 ^{o}C for 20 seconds, 57 ^{o}C for 2 minutes, 72 ^{o}C for 1 minute; 72 ^{o}C for 7 minutes.

The resulting PCR amplicons (~531 bp) were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA), quantified using Qubit fluorometric quantitation (Thermo Scientific Fisher, USA) and were used as templates for the second-step PCR in order to include the indexes (barcodes), as well as the Illumina adaptors. The amplification reaction mix of the second PCR contained 6 µl 5x KAPA HiFi Fidelity buffer, 0.75 µl KAPA dNTP Mix (10 mM), 3 µl from each primer (10 µM), 0.75 µl KAPA HiFi HotStart DNA polymerase (1 U/µl) in a final volume of 30 µl per reaction. DNA template concentration was about 20 ng/µl. The second PCR protocol used was the following: 95 ^{o}C for 3 minutes; 8 cycles at 98 ^{o}C for 20 seconds, 55 ^{o}C for 30 seconds, 72 ^{o}C for 30 seconds; 72 ^{o}C for 5 minutes. Amplifications were carried out using T100 Thermal Cycler (BIORAD). Again, the

resulting PCR amplicons (~600 bp) were purified and quantified as mentioned previously, mixed in equimolar amounts and sequenced using a MiSeq Reagent Kit v3 (2 x 300 cycles) at the IMBBC (HCMR).

All the raw sequence files of this study were submitted to the European Nucleotide Archive (ENA) (Leinonen *et al.*, 2011) with the study accession number PRJEB20211 (available at http://www.ebi.ac.uk/ena/data/view/PRJEB20211).

2.2.3 Data analyses

The raw sequence reads retrieved from all the sediment samples were quality trimmed using sickle (Joshi and Fass, 2011), to where the average quality score dropped below 20 (-q 20) as well as where read length was below 10bp (-l 10). SPAdes assembler (Bankevich *et al.*, 2012), that incorporates BayesHammer (Nikolenko *et al.*, 2013), was used for the creation of error-corrected paired-end reads, since this strategy along with overlapping paired-end reads reduces errors for MiSeq (Schirmer *et al.*, 2015). According to the pipeline options used for the SPAdes assembler, read error correction only was performed (--only-error-correction) along with reduction of the number of mismatches and short indels with MismatchCorrector, a post processing tool (--careful).

Afterwards, pandaseq (Masella *et al.*, 2012) was used to overlap the paired-end reads using a minimum overlap of 20 (-o 20). The overlapped sequences were combined and dereplicated. Then, using USEARCH (Edgar, 2010), reads were sorted based on abundance, singletons were discarded and OTU clustering and de novo chimera removal were performed. Following the relevant recommendation, a reference-based chimera filtering step was performed using UCHIME (Edgar *et al.*, 2011) using the "Gold" database as a reference.

Reads, including singletons, were then mapped back to OTUs, using the 97% similarity threshold level. Afterwards, they were aligned using MAFFT (Katoh *et al.*, 2005) and a phylogenetic tree was created using FastTree (Price *et al.*, 2010). Finally, taxonomic profiles of the OTUs were generated using RDP classifier (Wang *et al.*, 2007).

The metabolic function of the OTUs was predicted using the Tax4Fun package (Aßhauer *et al.*, 2015), which transforms the SILVA-based OTUs into a taxonomic profile of organisms which have a metabolic profile in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (fctProfiling = T), normalized by the 16S rRNA copy number (normCopyNo = T). The method for pre-computing the functional reference profiles was the ultrafast protein classification tool (Meinicke, 2015) (refProfile = UProC) and the functional reference profiles were computed based on 400 bp reads (shortReadMode = F). The 100 most abundant metabolic pathways were extracted.

The number of sequences assigned to an OTU represented its relative abundance at a given replicate sample of each sampling station. A matrix containing the microbial OTUs as variables and sampling stations as samples was constructed. A second matrix was constructed, with the predicted metabolic functions of the microbial OTUs as variables and sampling stations as samples. Both matrices were subsequently used to calculate triangular similarity matrices using the Bray-Curtis similarity coefficient (e.g. Clarke and Warwick, 1994). In order to test whether the microbial community pattern and the metabolic function pattern could be differentiated based on the sampled habitat, we performed non-metric multidimensional scaling (nMDS) (Clarke, 1993) and permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001). The design considered two factors: "habitat" and "location" (999 permutations), with the latter being nested in the former.

A third matrix was also constructed, with the respective abiotic parameters as variables and the sampling stations as samples, which was normalized (i.e. the values for each variable had their mean subtracted and were divided by their standard deviation) and used as an input for the BIO-ENV analysis (Clarke and Ainsworth, 1993) by employing the Spearman's rank coefficient. The analysis was performed to identify the subsets of abiotic parameters that were associated with the community (i.e. OTU) and the metabolic function matrices. In order to account for factor confounding, and estimate the amount of variation each factor might explain in the community as well the metabolic function similarity matrices, variation partitioning analysis was performed. The amount of variation in both matrices that was due to

salinity uniquely, while taking the other environmental factors into consideration, was estimated and its significance was tested using distance-based redundancy analysis (db-RDA) and Analysis of Variance (ANOVA).

A suite of diversity indices (Margalef's species richness, Pielou's evenness, Shannon-Wiener (Pielou, 1969), Chao-1, Abundance Coverage Estimator (ACE)) was calculated for each sampling station. The indices were subsequently tested for significant differences between the different locations and habitats using the Kruskal-Wallis test. The Mann-Whitney U test (Mann and Whitney, 1947) was used for the *post-hoc* pairwise comparisons; a Bonferroni-correction was applied and the level of significance for the results was lowered to 0.008, in the case of the locations, and to 0.017, in the case of the habitats.

Linear regression was used to examine the significance of relationships between OTU diversity, i.e. the average number of OTUs as well as the average values of diversity indices, and salinity. Shapiro-Wilk test was used to assess normality in the residuals (Shapiro and Wilk, 1965), while homoscedastistic residual variances were confirmed by examining plots of the standardized residuals (Draper and Smith, 1981).

The non-parametric Kruskal-Wallis test (Kruskal and Wallis, 1952) was used to determine which of the predicted metabolic pathways were statistically significant between the different habitats, i.e. "lagoon", "river" and "sea". The p value threshold was adjusted from the initial 0.05, using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). In addition, the relative abundance values of the main microbial taxa were compared among the locations and habitats of the sampling stations, using the Kruskal-Wallis test.

The *vegan* package (Oksanen *et al.*, 2008) was used for the nMDS (*metaMDS* function), PERMANOVA (*adonis* function), BIO-ENV (*bioenv* function), variation partitioning analysis (*varpart* function) and db-RDA (*capscale* function), for the calculation of diversity indices and for the generation of rarefaction curves. Groups in the nMDS plots were displayed using the *ordiellipse* and *veganCovEllipse* functions; in particular, *ordiellipse* function was used to retrieve the spread of points based on the chosen grouping (e.g. habitat) using standard deviations of points (kind = sd) and confidence limit for ellipses (conf = 0.95), while *veganCovEllipse* function calculated a covariance matrix of each group (e.g. habitat) to generate the ellipse

points[4]. Linear regressions, Shapiro-Wilk, Mann-Whitney, Kruskal-Wallis and ANOVA tests were conducted using *stats* package (R Core Team, 2015). Graphs were constructed using the *ggplot2* package (Wickham, 2009). The aforementioned analyses were performed using R version 3.2.1 (R Core Team, 2015).

2.3 Results

2.3.1 Microbial community composition

The results of the processing of the sequences are shown in Table S2.1. The 1,893,500 overlapped reads were dereplicated (1,578,523 remained) and singletons were removed (111,223 remained). After *de novo* chimera removal (7,088 OTUs) and chimera removal using the "Gold" database as a reference, the final number of OTUs was 7,050. The corresponding rarefaction curve is shown in Figure S2.1; all samples have reached a plateau, with the riverine samples having reached a higher OTU abundance followed by the lagoonal and the marine ones.

The nMDS of the microbial OTUs (Figure 2.2) showed that their spatial pattern differs both by habitat and location, which was also confirmed by the PERMANOVA results (habitat: $F.Model = 19.416$, $p < 0.01$; location: $F.Model = 10.647$, $p < 0.01$).

[4] More information can be found in:
http://userweb.eng.gla.ac.uk/umer.ijaz/bioinformatics/ecological.html

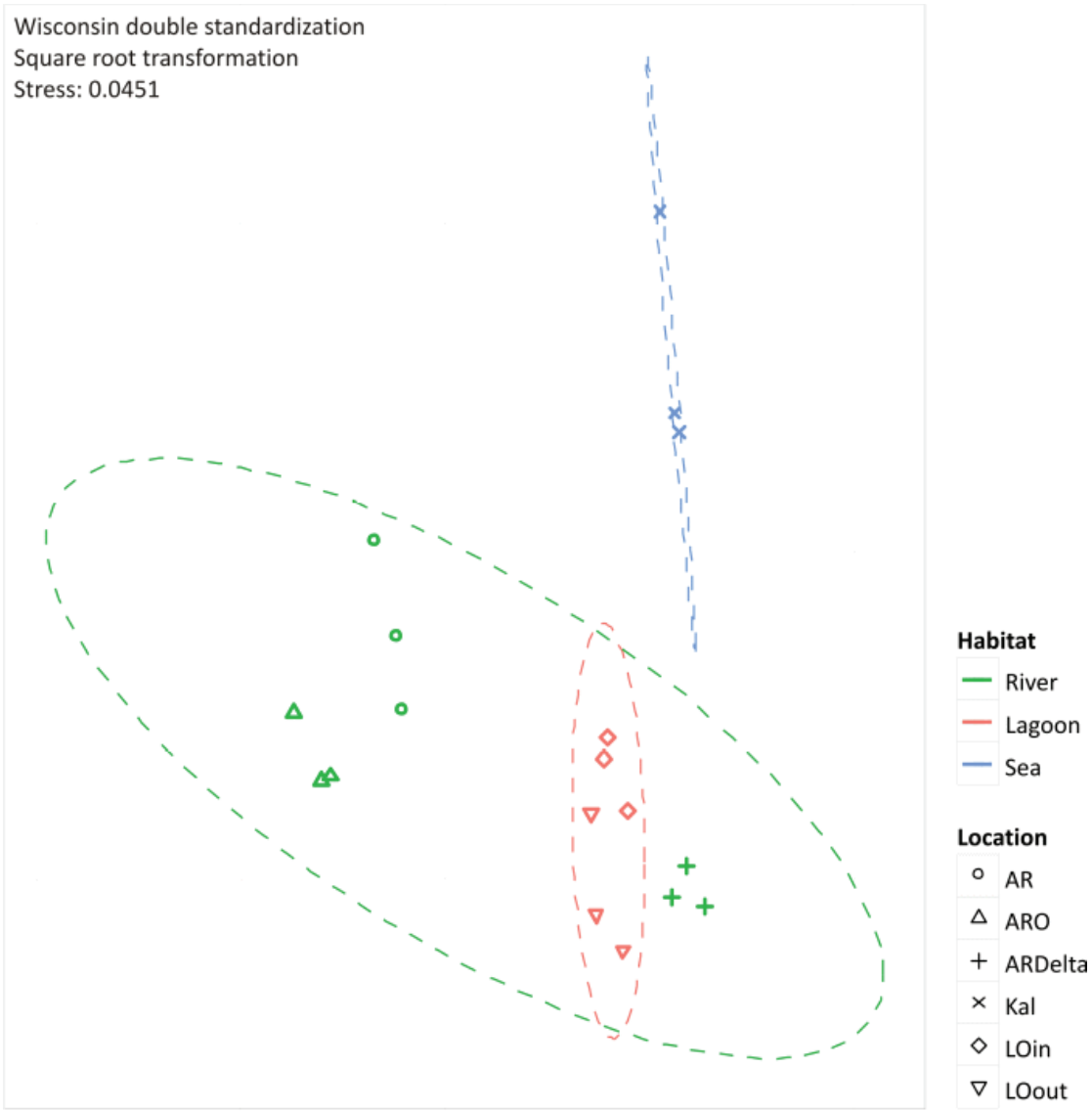

Figure 2.2: nMDS of the similarity matrix of the sampling stations based on the microbial OTUs abundances. Ellipses according to habitat, signs according to location. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

An alternative representation of the sampling stations, where the Arachthos Delta sampling station was grouped with the lagoonal samples, is shown in Figure S2.2; this representation was also statistically significant (PERMANOVA; habitat: F.Model = 9.071, $p < 0.01$; location: F.Model = 14.155, $p < 0.01$) but it was not chosen as the primary one. Arachthos Delta station was combined with the riverine samples (from

Arachthos and Arachthos Neochori) due to its vicinity with the river mouth and the subsequent continuous influence from the river outflow.

The relative abundance percentages of the microbial taxa, at the phylum level, did not show a significant differentiation between the different locations or habitats (Kruskal-Wallis: $p > 0.05$ for all cases). However, as shown in Figure 2.3 where the relative abundance percentages of each replicate sample have been averaged per sampling station, there are certain differences that can be observed. For example, the *Bacteroidetes* showed an increasing abundance when moving from the inner station of the river (~9%) to the more brackish stations (~28%), and decrease again in the marine station (~11%). A similar trend was observed for the *Proteobacteria*, but in this case the higher abundance was found in the marine environment (~49%). The abundance of *Archaea* was quite low in all the sampling stations; higher values were observed at the inner station (~17%) of the river and decrease towards the marine station. *Cyanobacteria/Chloroplasts* were more abundant in the marine station (~10%) and were also found in the Arachthos delta and inner station of Logarou lagoon at lower abundance (~5%).

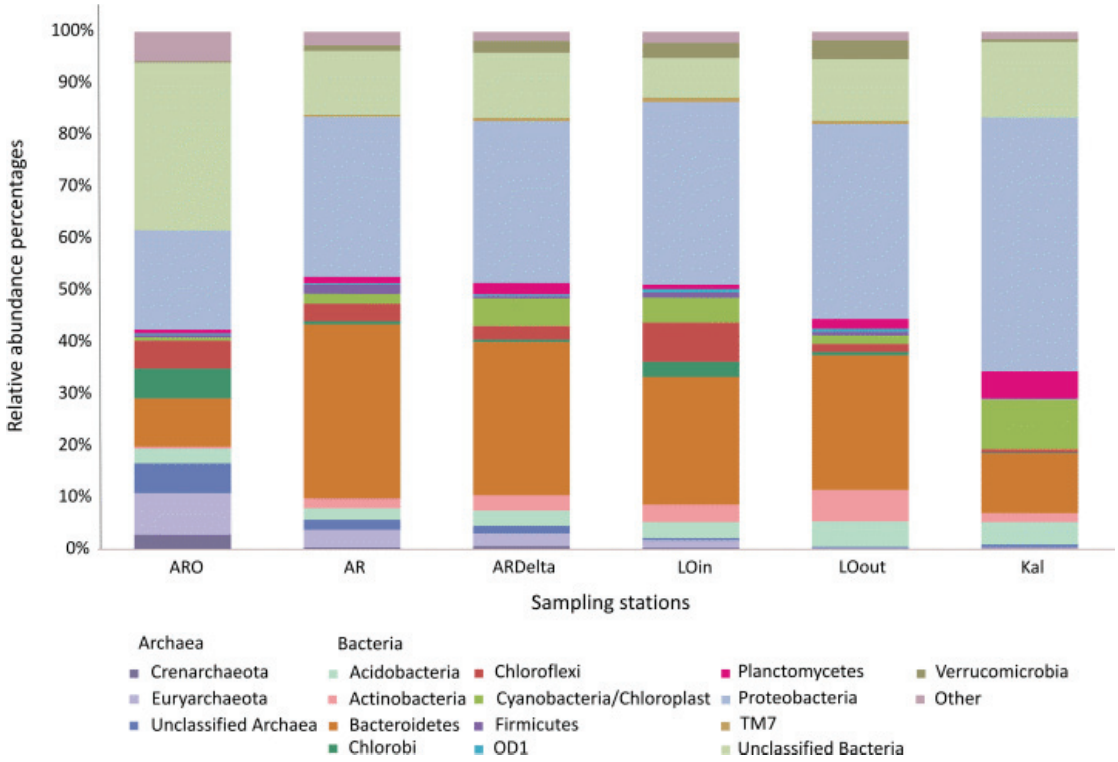

Figure 2.3: Bar chart showing the abundances of the main microbial taxa, at the phylum level, at the sampling stations. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

When the most abundant phyla are examined in more detail, again using the relative abundance percentages of each replicate sample averaged per sampling station, there are certain differences observed at the sampling stations, although non-significant (Kruskal-Wallis: $p > 0.05$ for all the cases). In the case of *Bacteroidetes* (Figure S2.3), *Flavobacteria* were less abundant at the inner station of the river (~17%); their abundance was higher in the other riverine stations (~37%) and in the lagoonal stations (~54%) while remaining stable at the marine station (~49%). *Sphingobacteria* exhibited the lower abundance in the Arachthos delta (~7%) and the highest in Kalamitsi (~28%).

Regarding the *Proteobacteria* phylum (Figure S2.4), *Alphaproteobacteria* exhibited higher abundances at the mouth of the Arachthos river (~49%) and at the marine station (~56%). *Betaproteobacteria* were almost exclusively present in the inner station and mouth of the Arachthos river (~30% and ~11% respectively).

Gammaproteobacteria were mostly abundant in the lagoonal stations (~54%) and *Deltaproteobacteria* in the Arachthos delta (~49%) which was the station with the lowest oxygen concentration in the water overlaying the sediment (0.24 mg/l; Table 2.2).

Diversity measured as Shannon-Wiener, Pielou's evenness and Margalef's species richness indices (Table 2.1) was significantly different between the locations and habitats (Kruskal-Wallis test, Table S2.2). In addition, the number of OTUs, Chao-1 and ACE were also significantly different between the different habitats (Kruskal-Wallis test, Table S2.2). The results of the values of the Mann-Whitney U tests for the post-hoc comparisons (Table S2.3), show that the main driver for the difference between the habitats is the difference of the riverine and the marine environment and secondly the difference of the former with the lagoonal environment.

2.3.2 Functional community composition

The retrieved KEGG metabolic profiles, and their abundance in each sample, are provided in Table S2.4. For certain microbial OTUs a KEGG profile could not be retrieved, thus these OTUs constitute the fraction of unexplained taxonomic units (FTU) (Aßhauer *et al.*, 2015), i.e. the amount of sequences assigned to a taxonomic unit and not transferable to KEGG reference organisms. As shown in Figure S2.5 and Table S2.5, this fraction was highest in the lagoonal samples (67.51 %) and lowest in the marine samples (38.91 %). From the riverine samples (56.85 %), the highest FTU was observed at the Arachthos Delta (71.32 %). Due to the aforementioned FTU values, the interpretation of the results should be done cautiously as there are many OTUs for which retrieval of a metabolic profile was not possible.

However, the metabolic profiles retrieved from the OTUs were significantly different between the three habitats (Kruskal-Wallis, $p < 0.05$). Specifically, as shown in Figure 2.4, there were 13 enzymes that could be linked to certain metabolic pathways, which were responsible for the between-habitat dissimilarity.

In the nMDS of the KEGG metabolic profiles (Figure 2.5), it is depicted that the different habitats were functionally distinctive. This was supported by the

PERMANOVA results (habitat: F.Model = 10.743, $p < 0.01$; location: F.Model = 11.206, $p < 0.01$).

Table 2.1: Diversity indices of the samples. OTUs: total number of OTUs. N: total microbial relative abundance values. H′: Shannon-Wiener. J′: Pielou's evenness. d: Margalef's species richness. ACE: Abundance Coverage Estimator. R: River. L: Lagoon. S: Sea. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi. A, B, C: replicate samples.

	OTUs	N	H′(ln)	J′	d	Chao-1	ACE
R_AR_A	2414	51617	6.64	0.85	222.36	2771.31	2736.48
R_AR_B	3070	74057	7	0.87	273.71	3450.95	3353.58
R_AR_C	2810	87640	6.77	0.85	246.82	3145.4	3109.94
R_ARO_A	2271	49793	6.66	0.86	209.88	2643.01	2570.08
R_ARO_B	2419	77520	6.66	0.86	214.78	2625.88	2582.98
R_ARO_C	2480	74292	6.59	0.84	221.03	2773.28	2717.73
R_ARDelta_A	2098	42365	6.66	0.87	196.83	2388.94	2319.51
R_ARDelta_B	2479	49442	6.61	0.85	229.26	2821.49	2737.81
R_ARDelta_C	2247	38209	6.57	0.85	212.87	2614.22	2565.73
L_LOin_A	2510	70514	6.39	0.82	224.75	2849.83	2770.58
L_LOin_B	2082	64400	5.9	0.77	187.94	2445.6	2399.07
L_LOin_C	2237	57953	6.16	0.8	203.88	2567.4	2532.17
L_LOout_A	1850	60250	5.66	0.75	168	2271.05	2200.13
L_LOout_B	2487	80942	6.38	0.82	219.97	2841.38	2766.75
L_LOout_C	2088	83935	5.95	0.78	184.07	2365.67	2323.44
S_Kal_A	1732	52975	5.9	0.79	159.13	1906.93	1903.36
S_Kal_B	2004	48280	6.40	0.84	185.72	2278.12	2237.63
S_Kal_C	1974	51129	6.34	0.84	181.98	2211.46	2182.35

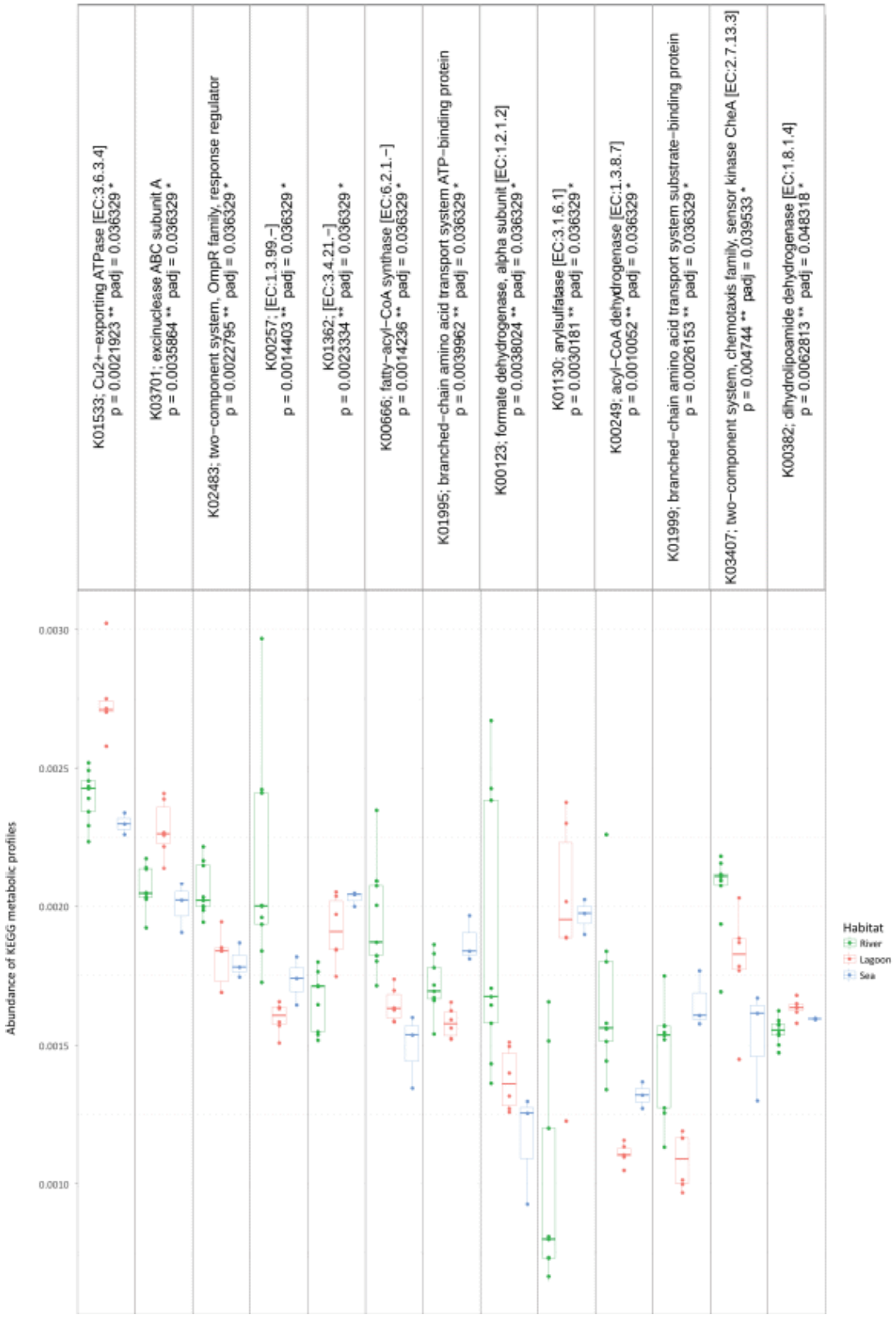

Figure 2.4: The enzymes that were significantly different between the three habitats.

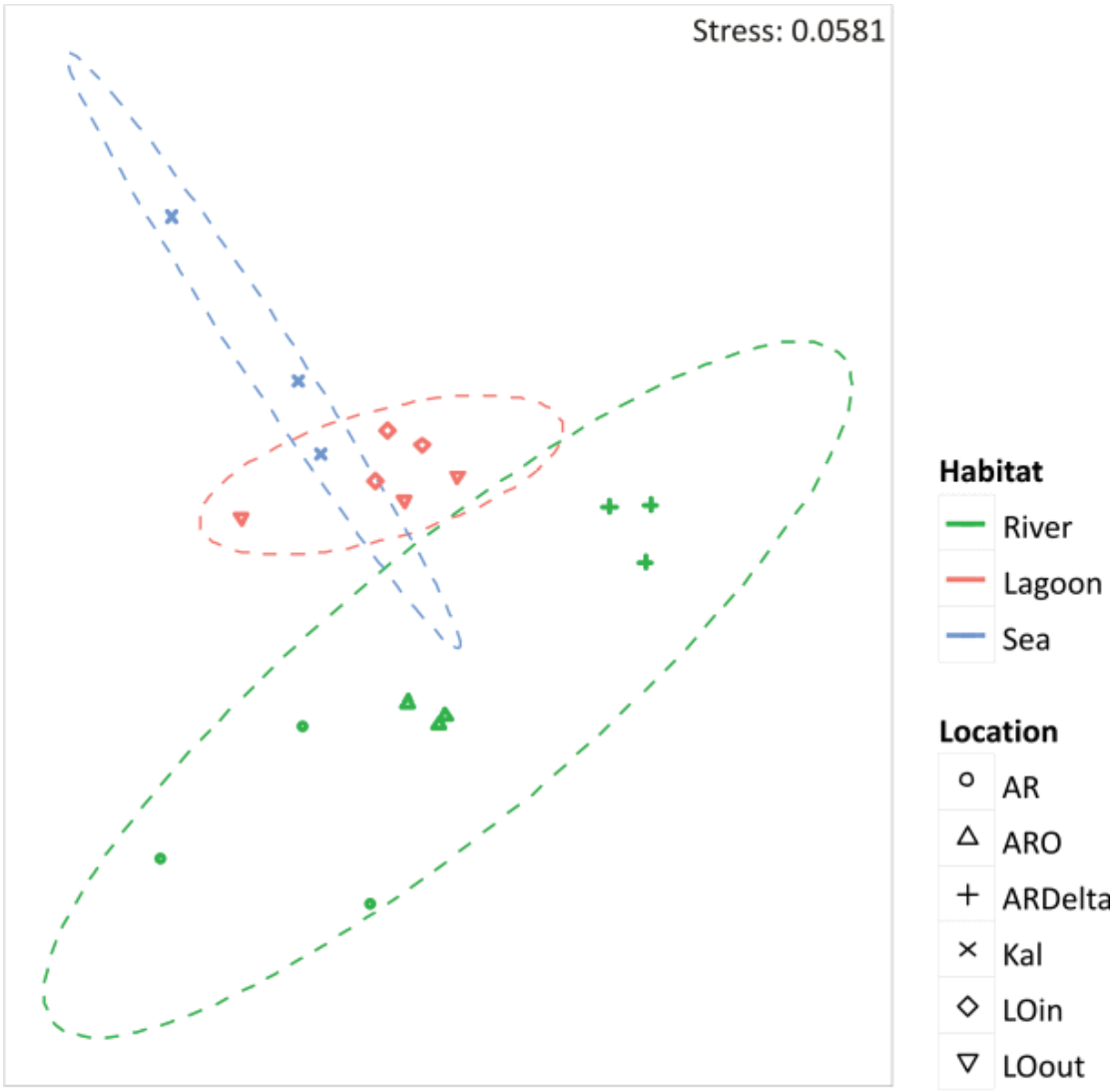

Figure 2.5: nMDS of the similarity matrix of the sampling stations based on the abundances of KEGG metabolic profiles. Ellipses according to habitat, signs according to location. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

2.3.3 Correlation with abiotic parameters

The average values of the physicochemical variables per sampling station are provided in Table 2.2.

Table 2.2: The average values of the physicochemical variables of the sampling stations. R: River. L: Lagoon. S: Sea. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi. Chl-a: Chlorophyll-a concentration (ug/l) and Fluorescence. O_2: Oxygen concentration. Chl-a sediment: Chlorophyll-a concentration in the sediment (ug/g). CPE: Chloroplastic pigment equivalents. POC: Particulate Organic Carbon. silt & clay: percentage of silt and clay. sand: percentage of sand.

		R_AR	R_ARO	R_ARDelta	L_LOin	L_LOout	S_Kal
Water	Chl-a	0.46	0.38	1.1 *	1.25	3.56	0.99 *
	Salinity (psu)	0.2	0.2	36.47	27.8	30.1	38.82
	O2 (mg/l)	10.24	9.99	0.24	8.42	8.42	6.81
	Temperature (oC)	14.1	14.9	19.06	11.1	14.2	19.77
Sediment	Chl-a sediment	1.12	3.35	2.11	8.22	24.02	2.83
	Phaeopigments (ug/g)	1.27	3.09	22.01	10.98	41.1	2
	CPE (ug/g)	2.38	6.44	24.12	19.2	65.12	4.83
	POC (ug/g)	10868.53	7217.63	22875.81	22797.08	30756.93	5100.95
	silt & clay	21.32	0.85	1.56	3.29	1.68	1.52
	sand	78.68	99.15	98.44	96.71	98.32	98.48
	Depth (m)	1.5	0.5	20	0,5	0.7	40

* Fluorescence values were regarded as a proxy for chlorophyll-a concentration

According to the results shown on Table 2.3, the abiotic variable that was best correlated with the microbial community pattern is salinity (ρ = 0.88). When the metabolic function pattern is considered, oxygen concentration was the variable showing the highest correlation with the former (ρ = 0.73), although the combination of salinity and oxygen concentration was also highly correlated with the metabolic pattern (ρ = 0.63). In addition, as shown by the variation partitioning analysis for all the combinations physicochemical variables of that resulted in models with adjusted R^2 of residuals less than 0.60 (Table 2.4), the combination of salinity, POC and temperature explained 56% of the total variation in the community similarity matrix. When the explanatory variables were regarded separately, salinity accounted for 27% of the variation, followed by POC (21%) and temperature (14%). In metabolic function pattern, 50% of the variation was explained by salinity and

oxygen concentration; salinity alone accounted for 12% of the variation while oxygen concentration for 33%.

Regarding the relationship between the number of OTUs and the salinity values, as shown in Figure 2.6, a linear decrease of the former was observed from the freshwater to the marine stations; salinity explained over 65% of the variation in the number of OTUs. Robust models were also evident when the data were divided into taxonomic groups (Table S2.6), with salinity explaining from 70% up to 98% of the variation of the OTUs diversity, in the case of Firmicutes and TM7 respectively. The residuals of all significant models presented in Table S2.6 showed no evidence of heteroscedasticity and were found to be normally distributed.

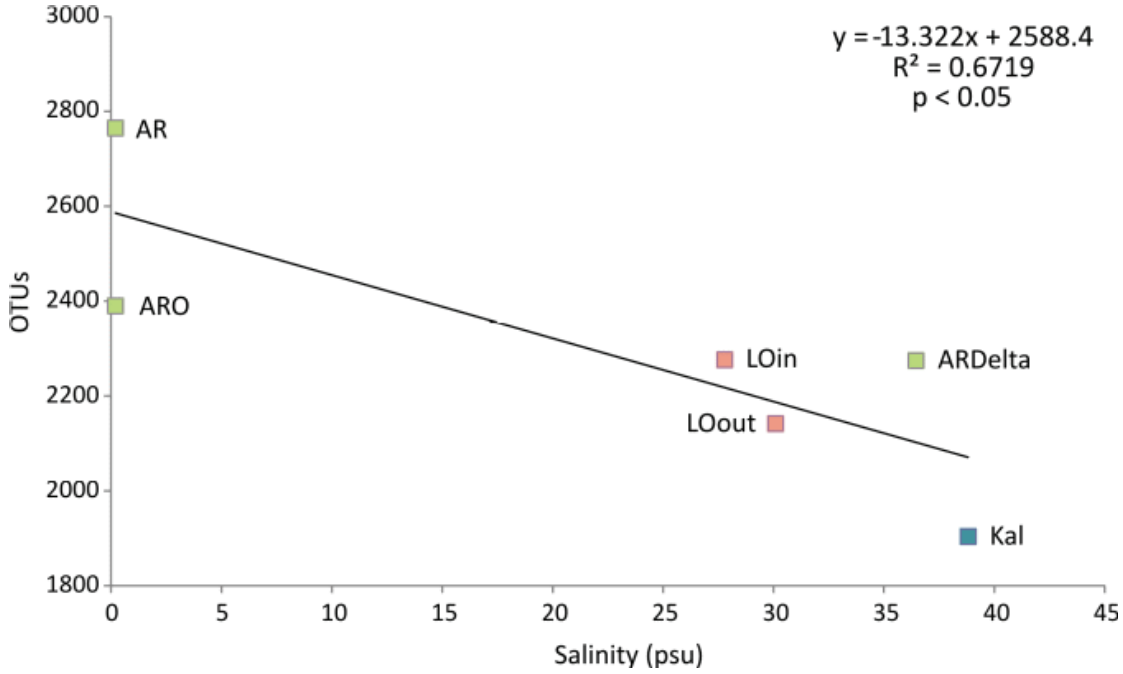

Figure 2.6: Linear regression between the number of OTUs (averaged per sampling station) and the salinity of the sampling stations. Colour according to habitat. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

Table 2.3: The environmental variables best correlated with the community and the metabolic function pattern, as provided by the BIO-ENV analysis. Each row represents the correlation of the microbial community pattern or the metabolic community pattern with a different set of environmental variables, indicated by the plus sign. ρ: Spearman rank correlation coefficient. POC: Particulate Organic Carbon. CPE: Chloroplastic pigment equivalents. Chl-a: Chlorophyll-a concentration (ug/l) and Fluorescence.

		Water				Sediment				
	ρ	Salinity (psu)	O_2 (mg/l)	Chl-a	Temperature (^oC)	POC (ug/g)	Sand (%)	Silt & Clay (%)	Phaeopigments (ug/g)	CPE (ug/g)
Community pattern	0.88	+								
	0.85	+				+				
	0.79	+	+			+				
	0.72	+	+			+		+		
	0.66	+	+			+		+		+
	0.64	+	+		+	+		+	+	
	0.58	+	+		+	+	+	+	+	
	0.54	+	+	+	+	+	+	+	+	
Metabolic function pattern	0.73		+							
	0.63	+	+							
	0.57		+		+			+		
	0.54	+	+				+	+		

Table 2.4: The percentage of variation explained (adjusted R^2) of each explanatory physicochemical variable, as well as their combinations. POC: Particulate Organic Carbon. *: p < 0.05. **: p < 0.01. ***: p < 0.001.

	Community pattern	Metabolic function pattern
Salinity	27% ***	12% *
O_2		33% **
Temperature	14% **	
POC	21% ***	
Salinity + O_2		50% *** (residuals: 50%)
Salinity + Temperature	47% *** (residuals: 53%)	
Salinity + POC	45% *** (residuals: 55%)	
POC + Temperature	36% ***	
Salinity + POC + Temperature	56% *** (residuals: 44%)	
Salinity with Temperature as condition variable	32% ***	
Salinity with POC as condition variable	24% ***	
Salinity with O_2 as condition variable		16% **
Salinity with Temperature and POC as condition variables	20% ***	
Temperature with Salinity as condition variable	20% ***	
Temperature with POC as condition variable	15% ***	
Temperature with Salinity and POC as condition variables	11% **	
POC with Salinity as condition variable	18% ***	
POC with Temperature as condition variable	22% ***	
POC with Salinity and Temperature as condition variables	10% **	
O_2 with Salinity as condition variable		37% ***

2.4 Discussion

2.4.1 Microbial community composition

Based on the results of the present study, it is suggested that the microbial community diversity pattern differs by habitat and location, thus indicating that, at least to some extent, each habitat hosts a different microbial community from the others, regarding both the number of OTUs and their composition, as has been also shown from similar studies in the Baltic Sea (Herlemann *et al.*, 2011). Furthermore, the abiotic variable that is best correlated with the microbial community pattern is salinity, as it has been shown from studies on bacterioplankton (e.g. Kirchman *et al.,* 2005; Nemergut *et al.,* 2011; Fortunato *et al.,* 2012), as well as sediment bacterial communities (Bolhuis and Stal, 2011; Severin *et al.,* 2012; Bolhuis *et al.,* 2013; Pavloudi *et al.,* 2016).

The majority of the observed OTUs was classified as *Bacteroidetes* and *Proteobacteria*. *Bacteroidetes* have been found to be omnipresent along estuarine gradients (e.g. Bouvier and del Giorgio, 2002) although in certain cases they dominated higher salinity communities (Campbell and Kirchman, 2013; Dupont *et al.*, 2014) which has been attributed to their ability for degradation of complex organic matter (Blümel *et al.*, 2007). In the present study, their abundance indeed increased in the brackish stations but decreased in the marine station, which could be due to the lower availability of organic matter in the latter, as reflected in the concentration of particulate organic carbon.

Alphaproteobacteria dominate marine communities (Edmonds *et al.*, 2009; Campbell and Kirchman, 2013; Herlemann *et al.*, 2016) but they are also ubiquitous in freshwater habitats (Zhang *et al.*, 2014), which justifies their high abundance in both the mouth of the Arachthos river and the Kalamitsi stations. *Betaproteobacteria* were almost exclusively present in the inner station and mouth of the Arachthos river; this complies with the general trend of *Betaproteobacteria* dominating freshwater habitats (Campbell and Kirchman, 2013; Zhang *et al.*, 2014) and declining with increasing salinity (Wu *et al.*, 2006). Their absence from the other sampling

stations can be explained by the fact that they are typical freshwater bacteria (e.g. Bie *et al.,* 2001), adapted to live in low salt concentrations and low osmotic pressure (Zhang *et al.*, 2014).

Gammaproteobacteria are generally more abundant in higher salinities (Wu *et al.*, 2006; Zhang *et al.*, 2014b; Herlemann *et al.*, 2016), which has been attributed to their opportunistic life strategies (Pinhassi and Berman, 2003) and the low salinity conditions being unfavorable for their growth (Zhang *et al.*, 2014b). However, *Gammaproteobacteria* can dominate brackish habitats (Edmonds *et al.*, 2009) as has also been shown from previous studies in the same sampling stations (Pavloudi *et al.*, 2016) and from the results of the present study.

Deltaproteobacteria were mostly abundant in the Arachthos delta; this can be attributed to the hypoxic conditions prevailing in this sampling station (Crump *et al.*, 2007) and to their generally documented abundance in brackish habitats (Edmonds *et al.*, 2009; Pavloudi *et al.*, 2016).

The low abundance of *Archaea* could be attributed to the bacterial primers used for the present study, which were not specific for amplification of the archaeal communities. In addition, the low abundances found can be traced to their inability to grow under estuarine environmental conditions since they are primarily of allochthonous origin (Bouvier and del Giorgio, 2002). However, the abundances of *Archaea* were decreasing with increasing salinity; in the marine environment, *Archaea* are generally limited to shallow or deep-sea anaerobic sediments and extreme environments (DeLong, 1992), which could explain their absence from the marine station in the present study.

It is evident that sea level rise, and the subsequent saltwater intrusion that it will cause, will impose an osmotic stress in microbial populations in the riverine stations (Chambers *et al.,* 2011). This may eventually shift the community in a more brackish state, although this will depend on the time-scale and intensity of saltwater intrusion. Microbial communities in fluctuating environments have been shown to be rather resilient, which causes variations in their expected functional response to change (Hawkes and Keitt, 2015). However, saltwater intrusion may induce an indirect effect in microbial communities, for example by causing changes in

vegetation (Nelson *et al.,* 2015) or by increasing sulfate concentrations, thus by stimulating its reduction (Chambers *et al.,* 2011).

As far as the relationship between the number of OTUs and salinity values is concerned, it has been shown that it is negative, i.e. diversity decreases with the increase of the salinity. Thus, the sediment microbial communities in the Amvrakikos Gulf salinity gradient do not appear to follow Remane's concept but rather the linear model proposed by Attrill (2002) with the species minimum at the point of maximum salinity range. This is in contrast with the constant relationship observed by Herlemann *et al.* (2011) for the Baltic Sea bacterioplankton along the salinity gradient. The divergence from the species-minimum concept could be attributed to the different life strategies of micro- and macroorganisms, as has been suggested by Telesh *et al.* (2013). In addition, microorganisms can be transported with water movement, apart from experiencing adaptation only at the molecular and cellular level (Telesh *et al.,* 2015). Thus, they experience salinity stress differently from benthic animals with reduced mobility (Skarlato and Telesh, 2017), which causes their deviation from the recognized models for macrobenthic organisms.

Although a decreasing trend of microbial diversity along gradients of increasing salinity has been observed (Rodriguez-Valera *et al.*, 1985; Benlloch *et al.*, 2002), the results of our study showed that this was only evident for the total number of OTUs; the other diversity indices did not show a statistically significant negative relationship with salinity (data not shown).

2.4.2 Functional community composition

The higher number of FTU observed in the lagoonal samples is indicative of the high microbial community diversity of this habitat. It also reflects the need for more extensive studies in order for an enhanced taxonomic resolution to be achieved for sequences found in lagoons. Similarly, the lower FTU was found in the marine samples, thus reflecting the number of studies that have been conducted in the marine environment so far which allowed for more OTUs to be transferred to KEGG reference organisms. However, despite the potential of the chosen method on the

metabolic profile retrieval of 16SrRNA sequences, its results should be regarded with caution since only through shotgun metagenomics the functional potential of a microbial community can be revealed. Nevertheless, in the next few years, as the software will become more elaborate and the queried databases will be enriched, better results will be derived from such methods and more trustworthy conclusions will be drawn.

Analogous studies in Chesapeake Bay, which is also a large water body with a pronounced salinity gradient, have also suggested that protein identification in environmental samples is rather challenging. Sequence databases have derived from cultured organisms and thus, it is unlikely that significant matches can be found between the bacterioplankton metaproteome and the queried databases (Kan *et al.,* 2005). Hence, more studies on isolation and cultivation of microorganisms from these habitats are needed in order to enrich the available information on KEGG and similar databases.

Although retrieval of a metabolic profile was not possible for the majority of the OTUs in certain cases, it is shown that different habitats were functionally distinctive and that salinity and oxygen concentration were highly correlated with the retrieved metabolic pattern. This is in accordance with studies suggesting that the actual patterns of composition and metabolism transition are strongly linked to hydrological conditions (Bouvier and del Giorgio, 2002). Furthermore, salinity has been found to be the main factor explaining almost all differences in the key metabolic capabilities of the Baltic Sea bacterial communities (Dupont *et al.*, 2014), so a similar relationship could be expected for the Mediterranean analogue of the Baltic Sea. In addition, salinity has been shown to influence proteome profiles from bacterioplankton communities of the Chesapeake Bay, since samples with higher salinity, i.e. closer to marine origin, are more similar and, at the same time distinct from inner Bay samples (Kan *et al.,* 2005). Similar results have been found for the San Francisco Bay, where salinity is one of the key variables influencing the abundance and activity of ammonia-oxidizing prokaryotes and denitrifiers (Mosier and Francis, 2008; 2010).

2.5 Conclusions

From the results of the present study, it can be concluded that the sediment microbial OTUs of the Amvrakikos Gulf salinity gradient do not follow the Remane's concept, i.e. there is no decrease in the intermediate salinities, but rather a negative trend. In addition, different taxonomic groups were more abundant in the freshwater stations while others were more abundant in the marine environment. Salinity was also found to influence the metabolic function patterns that were retrieved for the sampling stations. However, future studies are needed to decipher the metabolic capabilities of all the OTUs found at the habitats under study and investigate in depth the impact of salinity at their functional potential. Furthermore, experimental studies are needed in order to directly examine the effect of salinity on microbial community composition and investigate how the latter will respond when subjected to salinities varying from freshwater to marine in the light of climate change and sea level rise.

Chapter 3: Diversity and abundance of sulfate-reducing microorganisms in a Mediterranean lagoonal complex (Amvrakikos Gulf, Ionian Sea) derived from *dsrB* gene[5]

Citation: Pavloudi C, Oulas A, Vasileiadou K, Kotoulas G, De Troch M, Friedrich MW, Arvanitidis C (2017) Diversity and abundance of sulfate-reducing microorganisms in a Mediterranean lagoonal complex (Amvrakikos Gulf, Ionian Sea) derived from dsrB gene. *Aquatic Microbial Ecology,* 79:209-219, doi: 10.3354/ame01829. © Inter-Research 2017

Abstract

Sulfate-reducing microorganisms (SRMs) are a phylogenetically and physiologically diverse group of microorganisms, responsible for the dissimilatory reduction of sulfate. SRMs thrive under anaerobic conditions with high availability of organic matter. Such conditions characterize lagoonal ecosystems which experience regular dystrophic crises. The aim of the present study was to explore the biodiversity patterns of SRMs and to examine the extent to which these patterns are associated with biogeographic and environmental factors. Sediment samples were collected from 5 lagoons in the Amvrakikos Gulf (Ionian Sea, western Greece). DNA was extracted from the sediment and was further processed through pyrosequencing of a region of the dissimilatory sulfite reductase β-subunit (*dsrB*). The results of this exploratory study show that the majority of the observed operational taxonomic units (OTUs) belong to the *Deltaproteobacteria* supercluster and more specifically, to the *Desulfobacteraceae* family. Salinity and ammonium ions are the environmental factors that best correlated with the SRM community pattern. Furthermore, the SRM

[5] Reproduced with permission from Inter-Research. The copyright in the article is not transferred by this license and remains with. This license permits the article copyright to Inter-Research to be further reproduced and distributed in print or electronically as an embedded part of the thesis. Modified version of the published paper.

community of the brackish lagoons is differentiated from that of the brackish-marine lagoons and the studied lagoons have distinct SRM communities.

3.1 Introduction

Sulfate-reducing microorganisms (SRMs) are a taxonomically diverse group involved in the biogeochemical cycles of carbon, sulfur (Jørgensen, 1982) and mercury (Gilmour *et al.*, 1998). They are anaerobically respiring microorganisms, which couple the degradation of organic compounds to the reduction of sulfate as a terminal electron acceptor (Rabus *et al.*, 2006), resulting in the production of sulfide (Muyzer and Stams, 2008). SRMs have been found in a variety of ecosystems, including freshwater wetlands (Li *et al.*, 1999; Pester *et al.*, 2012), estuarine sediments (Jiang *et al.*, 2009) and extreme environments (Dhillon *et al.*, 2003; Fishbain *et al.*, 2003); they have also been found in high abundances in polluted sites (Pérez-Jiménez and Kerkhof, 2005) and associated with metals (Nakagawa *et al.*, 2002).

Six phylogenetic lineages constitute the cultured representatives of known SRMs, with 4 of them belonging to the bacterial (*Deltaproteobacteria*, *Nitrospirae*, *Firmicutes*, *Thermodesulfobacteria*) and 2 to the archaeal domain (*Euryarchaeota*, *Crenarchaeota*) (Muyzer and Stams, 2008; Müller *et al.*, 2015). Due to their polyphyletic nature, 16S rRNA gene based analysis cannot sufficiently describe the SRMs and functional gene markers should be used instead for the assessment of their abundance and diversity (Wagner *et al.*, 2005).

Sulfate reduction is a reaction found in sulfate-reducing prokaryotes and other organisms, such as plants, algae and fungi, and it can be further distinguished into assimilatory and dissimilatory sulfate reduction (Madigan *et al.*, 2012). The final step of the latter, namely the reduction of (bi)sulfite to sulfide, is catalyzed by the dissimilatory (bi)sulfite reductase (*dsr*) which is encoded by the *dsrAB* gene (Kondo and Butani, 2007; Liu *et al.*, 2009). Therefore, the *dsrAB* gene is considered a key functional marker for molecular analysis and detection of SRMs (Wagner *et al.*, 1998; Dar *et al.*, 2007) and its application has revealed a great diversity of organisms that

are not closely related to known and recognized SRMs. However, it should be interpreted with caution since *dsrAB*, in reverse, is also involved in the oxidative steps of the biogeochemical sulfur cycle (Müller *et al.*, 2015).

In the present study, the chosen technique for the identification and enumeration of SRMs was pyrosequencing (Ronaghi *et al.*, 1998) of the dissimilatory sulfite reductase β-subunit (*dsrB*) gene. Pyrosequencing is a molecular technique that has been widely applied in the field of microbial ecology, mostly targeting hypervariable regions of the 16S rRNA gene (e.g. Thompson *et al.*, 2011; Yu Wang *et al.*, 2012) instead of functional marker genes (e.g. Pelikan *et al.*, 2016).

The site under investigation was the lagoonal complex of the Amvrakikos Gulf (Ionian Sea, western Greece), one of the largest semi-enclosed embayments of the Mediterranean Sea, which is characterized by a fjord-like oceanographic regime (Ferentinos *et al.*, 2010). The structure and function of lagoonal ecosystems is largely determined by the input of organic matter from terrestrial and marine environments, as well as from the atmosphere (e.g. Viaroli *et al.*, 2008). The subsequent decomposition and removal of organic matter (Tagliapietra *et al.*, 2012) defines the structure and function of lagoonal ecosystems.

The study aimed to identify SRMs, and specifically to test (1) whether the SRM communities in the studied lagoons exhibit biogeographic patterns and (2) the extent to which these patterns are associated with environmental factors.

3.2 Materials and Methods

3.2.1 Location and general characteristics of the lagoons and sampling sites

The lagoons of the Amvrakikos Gulf (38° 59’ N, 20° 57’ E) are protected by the Ramsar convention and are listed in the Natura 2000 network. The Gulf is connected with the Ionian Sea via a narrow channel, the Preveza (Aktio) Strait (Kapsimalis *et al.*, 2005). The lagoonal complex at the northwest part of the Amvrakikos Gulf is formed by the rivers Arachthos and Louros (Poulos *et al.*, 1995).

Five lagoons of the Amvrakikos Gulf were sampled for the purposes of this study: Logarou, Rodia, Tsoukalio, Mazoma and Tsopeli. In each lagoon, 2 sampling stations were chosen at the extremes of the confinement gradient: the first station was located at the inner part of the lagoon and the second near the channel connecting the specific lagoon with the Gulf (Figure 3.1). Sampling was carried out in February 2011.

Figure 3.1: Amvrakikos Gulf in the Ionian Sea, indicating the location of lagoons (Logarou, Rodia, Tsoukalio, Mazoma and Tsopeli) and sampling stations (2 per lagoon) used for study of the diversity and abundance of sulfate-reducing microorganisms. Filled symbols are stations inside the lagoons; empty symbols are stations in the channel connecting each lagoon to the gulf.

3.2.2 Sampling methodology

Sediment samples were collected from all stations by means of a modified manually operated boxcorer, with a sampling surface of 156.25 cm^2 and a sediment penetration depth of 25 cm. Cylindrical sampling corers, with an internal sampling

surface of 15.9 cm^2, were placed inside the box-corer and sub-samples of the sediment's upper layer (0 to 0.2 cm) were collected from them. Three replicate units were taken from each sampling station, to determine variability within and among stations. Samples for molecular analysis (each consisting of about 15 cm^3 of sediment) were placed in 50 ml falcon tubes (Sarstedt) and were stored at −20 °C until further processing in the laboratory.

In addition, a variety of environmental variables were measured both in the sediment and in the water column (for a detailed description see Vasileiadou *et al.*, 2016; Pavloudi *et al.*, 2016).

3.2.3 DNA extraction, PCR amplification and pyrosequencing of the *dsrB* gene

DNA was extracted using the UltraClean® Soil DNA Isolation Kit (MO BIO Laboratories), according to the 'alternative protocol for maximum yields', as recommended by the manufacturer. About 0.5 g (± 0.2 g) of wet sediment from each sample were used.

PCR amplification was performed with newly designed primers, based on genomes of known SRMs and targeting part of the *dsrB* gene. Primer design was done in accordance with the recommendations of Roche (manufacturer of the GS FLX Titanium) and specifically guided by the advice that amplicons should cover the sequence of interest within the first 400 bp of sequencing. Primer coverage was tested *a posteriori* with the ARB Probe Match tool (Ludwig *et al.*, 2004) against the 1292 core nucleotide sequences from the reference database (Müller *et al.*, 2015), using perfect match and one weighted mismatch.

The primers used were 1595f (5'-YCA YGA RAT CCT BGA RCC-3') and 1905r (5'-CTG GGT RTG RAC GAT RTT G-3'). The primers were complemented with the 454 adapters (Lib-A Chemistry) and with sample-specific 5 bp barcodes (nucleotide 'keys') downloaded from VAMPS (Huse *et al.*, 2014). Six different primer pairs were used, each one bearing a specific barcode which enabled the separation of the different samples after the sequencing (demultiplexing).

The amplification reaction mix contained 6 µl 5X KAPAHiFi Fidelity buffer, 0.9 µl KAPA dNTP Mix (10 mM), 1.5 µl from each primer (10 µM), and 0.6 µl KAPAHiFi HotStart DNA polymerase (1 U/µl) in a final volume of 30 µl per reaction. DNA template concentration was about 50 ng/µl. The PCR protocol used as follows: 95°C for 5 min; 30 cycles at 98°C for 20 s, 54°C for 15 s, 72°C for 30 s; 72°C for 5 min. Amplifications were carried out using MyCycler (BIORAD) and DNA Engine DYAD (Peltier Thermal Cycler, MJ Research). Samples were purified using Agencourt AMPure XP (Becker Coulter). Amplicons were quantified with the Picogreen assay (Molecular Probes), mixed in equimolar amounts and sequenced using the 454 GS FLX Titanium Series (Roche) hosted at IMBBC (HCMR), in compliance with the recommendations of the manufacturer and using 4 lanes of the sequencing plate. Sequencing of one amplicon (sample M_01_A; see Table S3.1) was considered to be faulty, resulting in a high number of errors. Therefore this sample was not included in further analyses.

All raw sequence files of this study were submitted to the European Nucleotide Archive (ENA) (Leinonen *et al.*, 2011) with the study accession number PRJEB3370 (available at www.ebi.ac.uk/ena/data/view/PRJEB3370).

3.2.4 Sequence processing

The raw sequence reads retrieved from all sediment samples were processed with the AmpliconNoise algorithm for removal of 454 sequencing errors, PCR single base errors and chimeras (Quince *et al.*, 2011), as described in Pavloudi *et al.*, (2016).

In addition to filtering and denoising, a further level of error correction was employed by translating nucleotide sequences into amino acids. The clustered high quality nucleotide sequences, using the 90% nucleotide similarity cut-off (Kjeldsen *et al.*, 2007; Angermeyer *et al.*, 2016), were translated to protein using FrameBot (Qiong Wang *et al.*, 2013); detection and correction of frameshift errors in the reads were done using a set of known *dsrB* protein sequences.

As described in Pelikan *et al.* (2016), the translated amino acid sequences of the present study were aligned to the *dsrAB* reference sequence alignment in MEGA 6

(Tamura *et al.*, 2013) using MUSCLE (Edgar, 2004) and they were placed into the reference tree using the Evolutionary Placement Algorithm (EPA; Berger and Stamatakis, 2011) in RAxML (version 8.0.23) (Stamatakis, 2014) and the PROTGAMMA -AUTO option which automatically chooses the best protein model for the data provided. The EPA derived OTUs classification was compared to the classification of the core dsrAB dataset of Pelikan *et al.* (2016) for the taxonomic inference of the OTUs.

Furthermore, the nucleotide sequences of the OTUs that were translated into protein were used as an input to the Seqenv pipeline (Sinclair *et al.*, 2016), using the unique isolation option, in an attempt to link the 100 most abundant OTUs with descriptive environmental terms and determine in which environments they have been previously observed.

3.2.5 Statistical processing

Lagoons were assigned to salinity categories based on their salinity ranges following the Venice system (International Symposium on the Classification of Brackish Waters, 1958), with the (mixo-) mesohaline domain (salinity 5 to 18 psu) further divided as described by Pavloudi *et al.* (2016). Hence, the 3 categories comprised the (1) (mixo-) polyhaline (salinity 18-30 psu), (2) (mixo-) b-mesohaline (salinity 8 to 18 psu) and (3) (Mixo-) a-mesohaline (salinity 5 to 8 psu) domains.

The number of sequences belonging to each OTU was considered representative of OTU relative abundance. Subsequently, a matrix of the OTU abundance was constructed, with the microbial OTUs as variables and sampling stations as samples. Nucleotide sequences that failed to translate to amino acids were excluded from the matrix. The OTU abundance matrix was used for the calculation of the triangular similarity matrix using the Bray-Curtis similarity coefficient (e.g. Clarke and Warwick, 1994). In order to investigate the bacterial community pattern in the area under study, non-metric multidimensional scaling (nMDS) (Clarke, 1993) and permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001) were performed. The design considered 4 factors: ‘lagoon’, ‘location’, ‘lagoon and location’ and

'salinity category' (999 permutations). Due to data limitations that did not allow for a successful 4-factor design to be tested, each factor was tested separately (under a design of unrestricted permutation of raw data).

In order to test the second hypothesis, an abundance matrix was constructed with the sampling stations as samples and the 38 environmental terms found associated with the OTUs as variables, which was also treated as mentioned previously. In addition, BIO-ENV analysis was applied, thus permitting investigation of all potential correlations between the biotic and abiotic matrices, by employing the weighted Spearman rank coefficient ρ_w (Clarke and Ainsworth, 1993). Environmental variables that were highly correlated ($-0.9 > \rho_w > 0.9$, $p < 0.05$) were excluded from further analyses (Clarke and Ainsworth, 1993). The RELATE routine (Clarke and Gorley, 2006) was applied to test for the significance of the correlated patterns, as calculated by the BIOENV analysis. This was performed between the biotic similarity matrices and those produced by subsets of the environmental parameters, as identified by the BIO-ENV analysis in each case.

OTU richness was estimated via extrapolation using the Chao-1 (Chao, 1987; Chiu *et al.*, 2014) and the Abundance Coverage Estimator (ACE) (O'Hara, 2005). In addition, a suite of diversity indices (Margalef's species richness, Pielou's evenness, Shannon-Wiener; (Pielou, 1969)) was calculated. The diversity indices, as well as the relative abundance percentages of the SRM OTUs, were tested for significant differences between the different salinity categories and lagoons by means of the nonparametric analysis of variance Kruskal-Wallis test (Kruskal and Wallis, 1952). The nonparametric Mann-Whitney U-test (Mann and Whitney, 1947) was used for the *post hoc* pairwise comparisons; a Bonferroni-correction was applied and the level of significance for the results of the Mann-Whitney pairwise tests was lowered from 0.05 to 0.017 in the case of the salinity categories, and from 0.05 to 0.01 in the case of the lagoons. In addition, the Mann-Whitney U-test was used to test for significant differences between the locations.

The DIVERSE routine (Clarke and Warwick, 1994) of the PRIMER (v.6) package (Clarke and Gorley, 2006) was used for the calculation of diversity indices. The same software was used for the BIO-ENV analysis and the RELATE routine. nMDS and PERMANOVA were performed with the R virtual laboratory (RvLab) (Varsos *et al.*,

2016). Chao-1 and ACE estimator were calculated using the EstimateR function of the vegan package (Oksanen *et al.*, 2016). Mann-Whitney U and Kruskal-Wallis tests were conducted with the IBM SPSS Statistics for Windows (Version 22.0). The Venn diagrams were constructed using the jvenn JavaScript library (Bardou *et al.*, 2014).

3.3 Results

The average values of the environmental variables per sampling station are provided in Table 3.1. As it can be seen from this table, oxygen concentration levels in the sampling stations, as measured in the water overlying the sediments, were similar; thus, SRM community patterns should be expected to be influenced by other environmental variables.

The coverage of the primers is presented in Table 3.2. The forward primer did not have as many positive hits as the reverse one; however, a substantial percentage of the *dsrAB* core dataset could have been amplified when tested *in silico*. Therefore, one could make the assumption that the results of the present study are indicative of the SRM community in the study area, although representatives of the community may exist for which amplification was not successful. A set of recommended primers for the *dsrAB* has been recently proposed and *in silico* evaluated (Müller *et al.*, 2015); however, it was not available during our study.

Table 3.1: The average values of the environmental variables of the sampling stations. M: Mazoma lagoon, L: Logarou lagoon, S: Tsopeli lagoon, T: Tsoukalio lagoon, R: Rodia lagoon. 01: Station inside the lagoons, 02: Station in the channel connecting each lagoon to the gulf. (from Pavloudi *et al.* (2016); reproduced with permission © Elsevier[6]). NH_4^+: ammonium ion. PO_4^{3-}: phosphate ion. NO_3^-: nitrate ion. NO_2^-: nitrite ion. Chl-a: chlorophyll a. POC: particulate organic carbon. TRIS: Total Reduced Inorganic Sulfur. Eh: redox potential. MD: median particle diameter. σ1: sorting coefficient. Sk1: skewness.

		M_01	M_02	L_01	L_02	S_01	S_02	T_01	T_02	R_01	R_02
	Depth (m)	2	0.7	0.5	0.7	0.3	1.5	0.7	0.5	1.5	3
	NH_4^+ (uM)	1.89	0.19	2.79	5.1	8.65	10.69	10.28	6.95	4.7	10.59
	NO_3^- (uM)	4.88	2.64	1.53	1.33	4.94	4.38	7.96	15.76	3.85	6.41
	NO_2^- (uM)	0.35	0.2	0.2	0.34	0.96	0.98	0.69	0.73	0.35	0.74
	Chl-a (ug/l)	34.1	59.31	9.13	11.36	5.48	4.04	6.16	0.96	3.04	1.75
Water	Phaeopigments (ug/l)	7.11	7.89	4.93	11.67	1.09	1.28	3.84	1.16	2.61	1.13
	POC (ug/l)	4856	5780	2141	2233	1055.25	862	1286.5	607.5	687	967.5
	Temperature (°C)	11	11.8	12.2	12.4	12.1	12.8	12.4	11.9	11.5	11.6
	Salinity (psu)	14.6	15.2	16.9	22	14.6	14.9	7.2	5.7	6.6	7.3
	pH	8.58	8.73	8.2	7.95	8.2	8.25	8.08	7.98	8.22	8.2
	O_2 (mg/lt)	9.43	8.73	7.8	7	8.39	8.03	8	7.45	7.96	8.1
	TRIS (uM/gr)	208.33	180.74	149.52	165.42	177.96	173.36	161.83	174.27	162.43	154.07
	Chl-a (ug/g)	80.27	43.44	85.17	34.89	51.01	36.3	20.5	63	56.42	65.89
	Phaeopigments (ug/g)	115.3	50.74	71.63	26.27	58.51	80.54	22.16	68.11	50.39	17.32
	POC (ug/g)	48721.71	26092.65	26122.93	13646.28	23778.05	33568.62	28445.72	32908.14	34381.61	30504.07
	% labile Organic Matter	12.78	5.98	6.74	3.63	5.46	7	5.36	6.65	7.18	6.22
Sediment	Sediment temperature (°C)	12	12	13	13.01	13	13	13	12.5	12	12
	Eh (mV)	-61.33	-89	17.33	286.67	96.67	-87.67	-13	152.67	-89	258.33
	Median Diameter (MD)	0.73	0.66	0.54	0.67	0.41	0.41	0.36	0.51	0.33	0.66
	σ1	1.11	1.07	1.52	0.96	1.61	1.45	1.42	1.26	1.37	1.18
	silt & clay (%)	1.3	1.74	3.32	2.85	3.51	3.46	0.3	5.82	5.90	2.59
	sand%	98.7	98.26	96.68	97.15	96.49	96.54	99.7	94.18	94.10	97.41

[6] License Number: 4178780829412; Order date: Aug 30, 2017; The raw data have been submitted at the MedOBIS Data Repository and can be accessed (available at http://ipt.medobis.eu/resource?r=zoobenthos_in_amvrakikos_wetlands).

Table 3.2: Coverage of the primers used to detect the presence of sulfate-reducing microorganisms (SRMs) in the lagoons of Amvrakikos Gulf (Ionian Sea).[7]

Primer name	Direction	Sequence (5′- 3′)	Length (nt)	Degeneracy	Coverage of core data-set *dsrAB* (%)	Average coverage of core data-set *dsrAB* (%)
1595f	Forward	YCAYGARATCCTBGARCC	18	48	3.17 - 14.86	8.01
1905r	Reverse	CTGGGTRTGRACGATRTTG	19	8	12.93 - 40.94	23.90

The results of the processing of the sequences during the noise removal are shown in Table S3.1. The 148,626 initial raw sequences were clustered into 18,655 high quality sequences, which corresponded to 5912 OTUs at the 90% similarity cutoff; out of those, 2167 were translated to amino acid sequences. The automatic protein model assignment algorithm of RAxML resulted in a log likelihood of −247,061.75. The labelled reference tree including branch labels and query sequences and the classification results show that the majority of the observed OTUs (74%) belong to the *Deltaproteobacteria* supercluster, within which the most abundant is the family *Desulfobacteraceae* (33%), with the Environmental supercluster 1 being second in terms of abundance (25%) (Figure 3.2a). However, when the abundance of the OTUs was taken into consideration, the difference between the groups was augmented (Figure 3.2b); the most abundant OTUs belong, as expected, to the *Deltaproteobacteria* supercluster (83%) followed by the Environmental supercluster 1 (16%).

This pattern was similar when each lagoon was regarded separately (Figure S3.1). Although representatives from the Environmental supercluster 1 were present in all lagoons (21 to 25%), their abundance was relatively lower (11 to 20%), while the *Deltaproteobacteria* supercluster showed higher abundance, which reached 88% in the case of Rodia lagoon. In addition, although the *Desulfatiglans anilini* lineage was present in similar percentages in all lagoons (18 to 27%), its abundance was greater in the a-mesohaline lagoons, i.e. in Tsoukalio and Rodia, and especially in the latter (40%), while levels were the same in the polyhaline and b-mesohaline sampling

[7] Design of the primers was based on genomes of known SRMs, targeting the β-subunit of the dissimilatory sulfite reductase gene (*dsrB*). Coverage of the primers was calculated using the ARB Probe Match Tool. Degeneracy is given as the number of oligonucleotides that comprise the primer. Coverage is calculated as the percentage of positive hits against the 1,292 core dataset.

stations. In the case of the *Desulfobacteraceae* family, the lowest presence (25 to 28%) and abundance (12 to 17%) were found in the a-mesohaline lagoons; by contrast, in the other sampling stations the aforementioned family exhibited a higher presence (33 to 38%) and abundance (22 to 39%). Although the presence of the *Syntrophobacteraceae* family was very low in all lagoons (1 to 2%), it showed higher abundance in Logarou (11%) and Rodia (8%) lagoons. The abundance of the unclassified OTUs of the *Deltaproteobacteria* supercluster was higher in all lagoons (12 to 27%) than their presence (7 to 9%). However, none of the aforementioned variations in the relative abundance percentages of the SRM OTUs were statistically significant (Kruskal-Wallis: $p > 0.05$ for all cases); this could be attributed to the use of data at the phylum level for the Kruskal-Wallis test while differences may be exhibited at a lower taxonomic level.

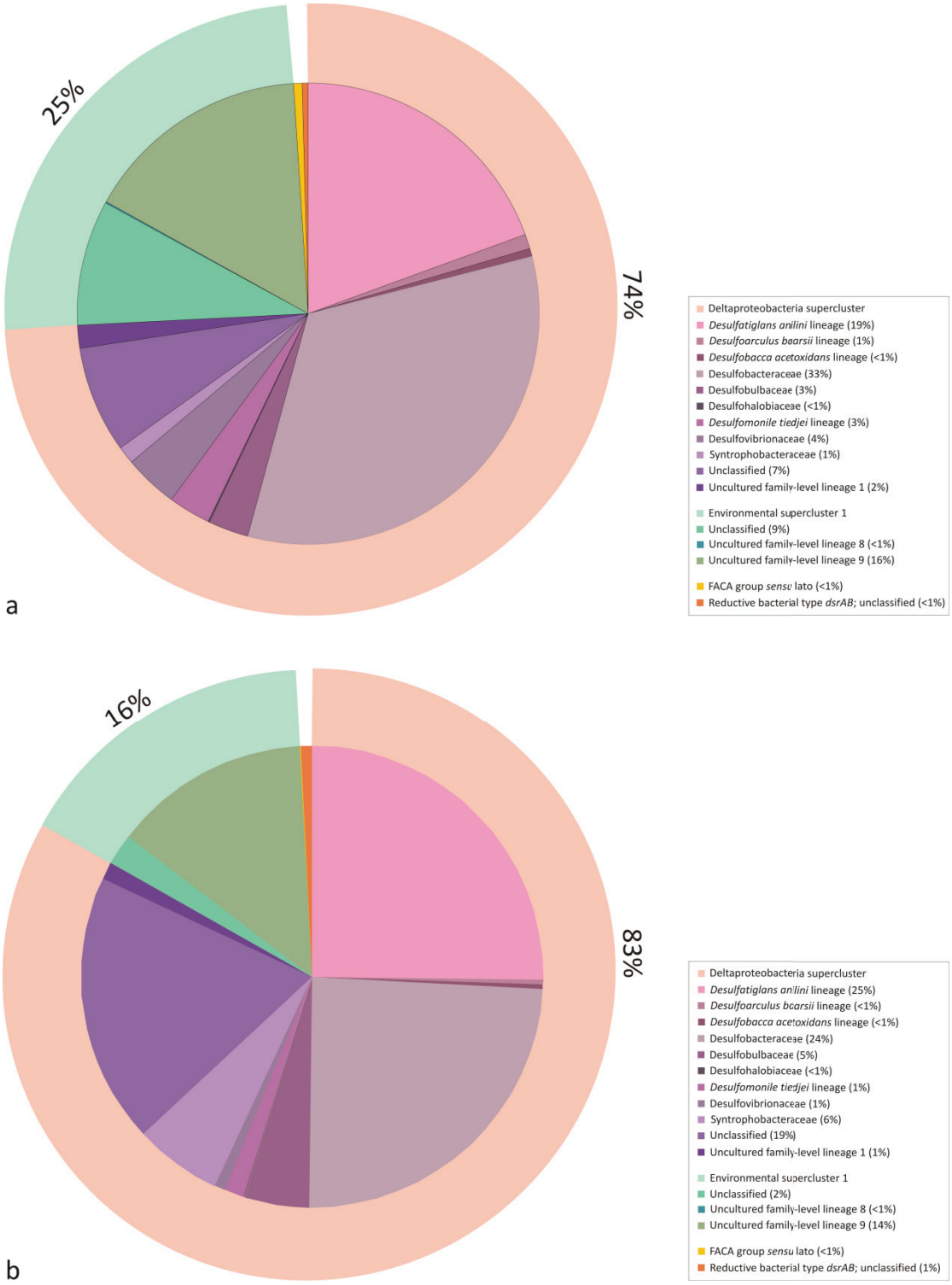

Figure 3.2: Taxonomic classification of the sulfate-reducing microorganism operational taxonomic units (SRM OTUs) retrieved from all the samples based on (a) the presence/absence and (b) the abundance of OTUs.

The number of OTUs that were commonly shared among the lagoons (90% similarity cut-off) is shown in Figure 3.3. A total of 149 OTUs were commonly shared by all 5 lagoons, corresponding to less than 7% of the total number of observed OTUs.

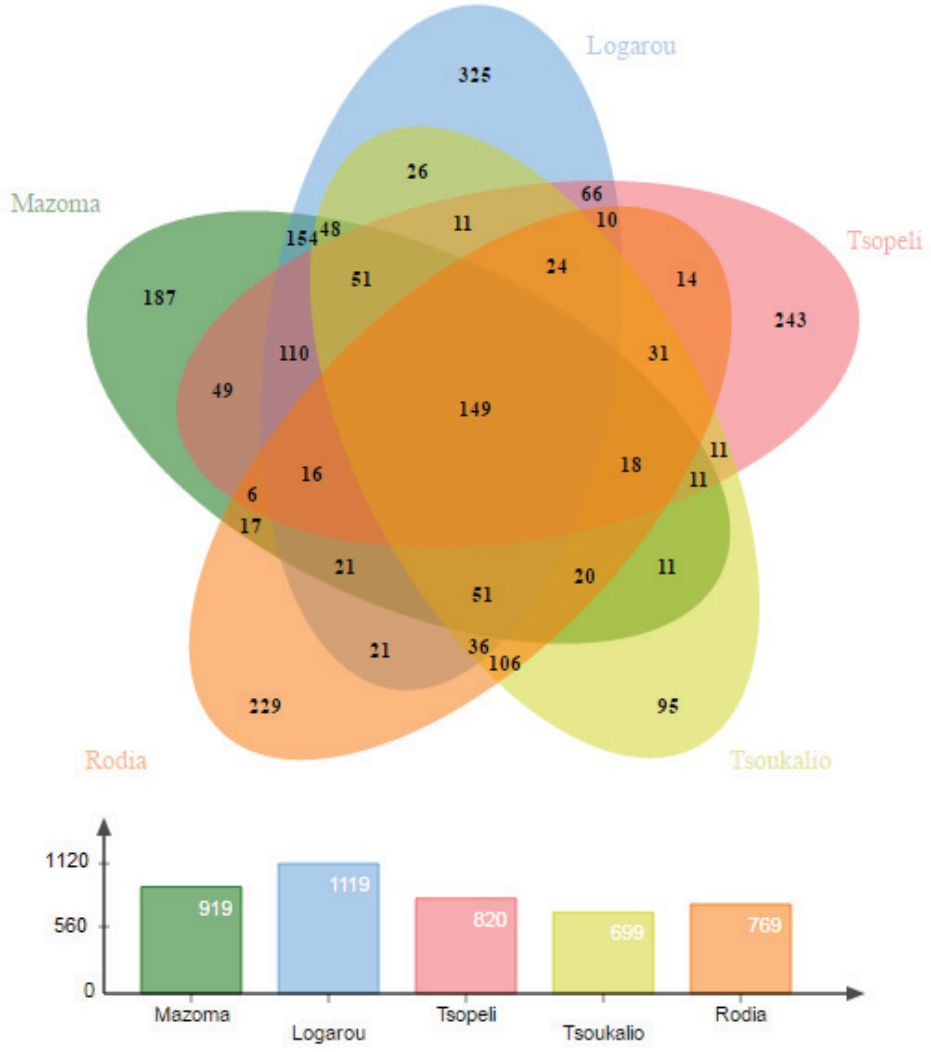

Figure 3.3: Number of OTUs commonly shared among all possible combinations of 2 or more of the 5 studied lagoons (see Figure 3.1) (90% similarity cut-off). The bar chart indicates the total number of OTUs retrieved from each lagoon.

nMDS of the bacterial OTUs spatial pattern (Figure 3.4) showed that the bacterial community pattern differs by lagoon and salinity category. The PERMANOVA test produced significant results for the factors 'lagoon' ($F.Model = 3.5936$, $p < 0.01$), 'lagoon and location' ($F.Model = 3.3443$, $p < 0.01$) and 'salinity category' ($F.Model = 4.0402$, $p < 0.01$). This is also depicted in the Venn diagram for the 3 salinity

categories (Figure 3.5). In addition, salinity and ammonium ions (NH_4^+) (Table 3.3) were the abiotic variables with the highest correlation to the SRM community pattern (ρ_w = 0.575). However, oxygen concentration was not included in the environmental variables that showed significant correlations with the SRM community pattern.

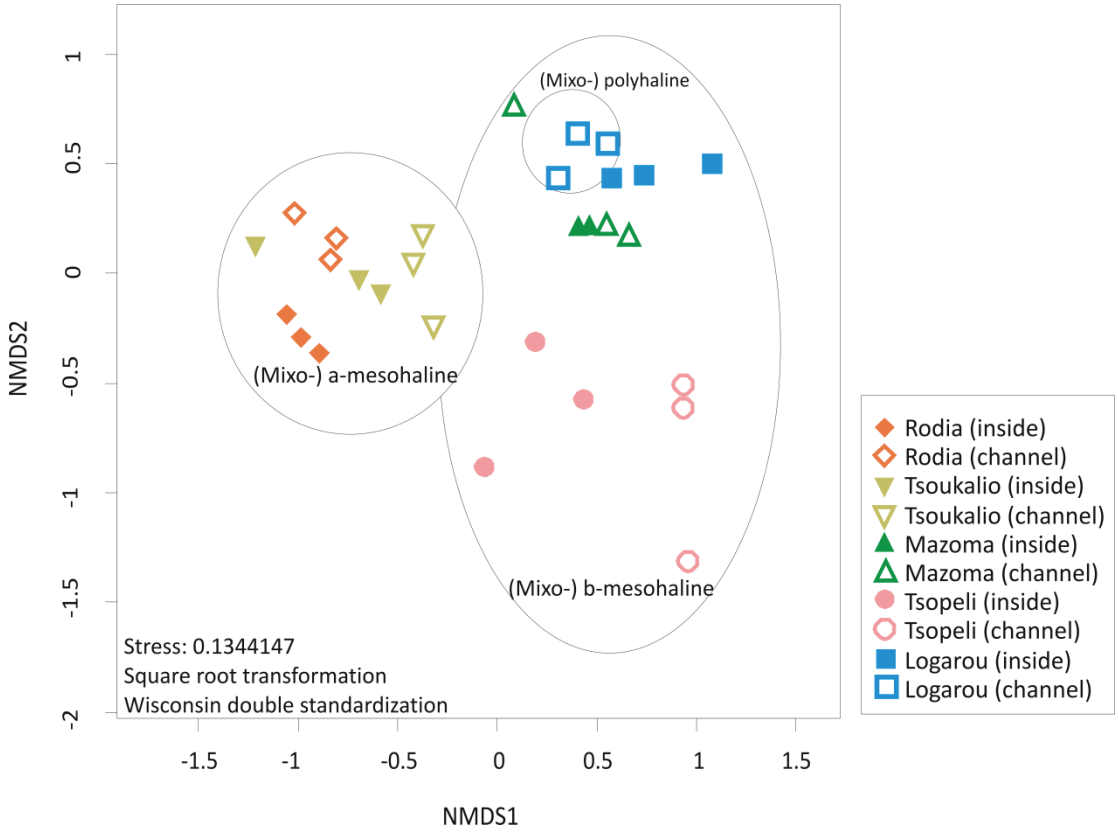

Figure 3.4: Multidimensional scaling of the SRM OTUs (90% similarity cut-off). Symbols indicate sampling stations (2 per lagoon in 5 lagoons; see Figure 3.1). Salinity categories are encircled (see 'Materials and methods: Statistical processing').

Table 3.3: Environmental variables best correlated with the sulfate-reducing microorganism community diversity pattern, as provided by the BIO-ENV analysis ($p < 0.01$). Each row represents the correlation of the microbial community pattern with a different set of environmental variables, indicated by the plus sign. ρ_w: weighted Spearman rank correlation coefficient. NH_4^+: ammonium ion. NO_2^-: nitrite ion. MD: median diameter. σ1: sorting coefficient.

ρ_w	Water			Sediment		
	Salinity	NH_4^+	NO_2^-	MD	Phaeopigments	σ1
0.575	+	+				
0.558	+					
0.537	+	+		+		
0.523	+	+	+			
0.521	+	+	+	+		
0.519	+		+			
0.518	+	+	+	+	+	
0.518	+	+			+	
0.513	+	+				+
0.510	+	+		+	+	

Based on nMDS of the environmental terms that are associated with the SRM OTUs (Figure S3.2), and according to the PERMANOVA test, the samples can be differentiated by all tested factors (lagoon: F.Model = 2.2625, $p < 0.01$; location: F.Model = 2.5827, $p < 0.05$; lagoon and location: F.Model = 4.0728, $p < 0.01$; salinity category: F.Model = 5.0393, $p < 0.01$). This is also evident from Table 3.4, where it is shown that the abundance of the associated environmental terms varied among the sampling stations. Undoubtedly, the most abundant term in all the lagoons was 'sediment' (~44%). The SRM OTUs found in the a-mesohaline lagoons, i.e. in Tsoukalio and Rodia, were associated with the term 'wetland' in higher abundances (~14 to 25%) compared with the other lagoons (~1 to 3%). Similarly, the terms 'hydrothermal vent' and 'acid mine drainage' were found in much lower abundance in the polyhaline and b-mesohaline lagoons (~2 to 5%) than in Tsoukalio and Rodia (~16 to 25%). In addition, Tsopeli lagoon was the only lagoon associated with the terms 'lake' and 'reservoir' (~12%).

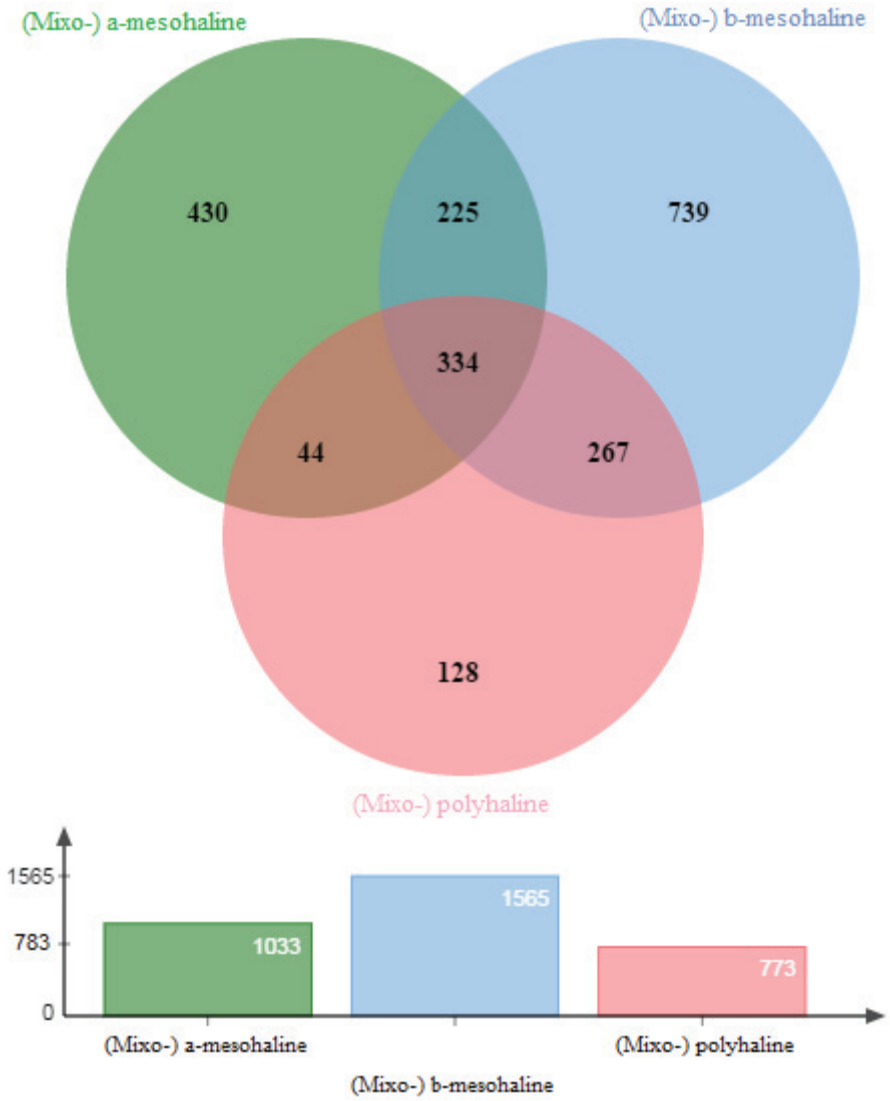

Figure 3.5: Number of OTUs of each salinity category shared with those of the other categories (90% similarity cut-off). The bar chart indicates the total number of OTUs found in each salinity category.

All diversity indices, except for the total SRM relative abundance values (N), were significantly different between the 3 salinity categories and the 5 lagoons (Kruskal-Wallis: $p < 0.05$ for all cases) (Tables S3.2 and S3.3). The post hoc comparisons showed that the diversity indices, in the case of salinity categories, did not differ significantly between (mixo-) bmesohaline and (mixo-) poly haline samples (Table S3.4). However, only Pielou's evenness index (J') was significantly different between the 2 locations (Mann- Whitney: $p < 0.05$) (Table S3.5).

Table 3.4: Abundance percentages of the environmental terms based on Environment Ontology (ENVO) vocabulary associated with the SRM OTUs (90% similarity cut-off) at each lagoon and at all lagoons combined.[8]

ENVO terms	Mazoma	Logarou	Tsopeli	Tsoukalio	Rodia	Lagoons combined
aquifer	7.24	5.75	3.66	9.65	5.19	6.05
biofilter	0.28	0.62		0.08	0.06	0.19
borehole	6.87	5.68	3.24	9.65	5.16	5.95
coast	0.05	0.03	0.12	0.02		0.02
depression	2.57	1.39	0.75	0.46	0.52	0.93
seamount		0.23	0.18		0.02	0.07
ground water		0.08	0.36	0.67	0.24	0.26
harbor	0.78	0.23	1.26	2.15	0.29	0.68
inlet	1.96		0.09	0.24	0.39	0.45
landfill		0.08	0.36	0.67	0.24	0.26
leachate		0.08	0.36	0.67	0.24	0.26
lentic water body	0.78	0.04	12.20		0.36	1.26
lotic water body	3.57	1.72	5.40	16.11	24.92	15.21
Marine biome/sediment/ water body	10.51	4.00	2.25	4.14	1.41	3.40
microbial mat	0.23	0.04		0.40		0.09
gold mine	6.87	5.68	3.24	9.65	5.16	5.95
mud	0.16	0.08				0.03
saline water	0.16				0.03	0.03
sea coast	1.96	0.16	0.09	0.24	0.42	0.49
sediment	52.35	72.47	64.54	31.05	30.41	43.76
sludge		0.04			0.15	0.08
soil	0.08			0.10	0.22	0.13
terrestrial biome/habitat	0.72	0.16	0.84		0.06	0.21
wetland	2.87	1.45	1.08	14.04	24.52	14.24

3.4 Discussion

The majority of the observed OTUs were identified as uncultured; although there are known representatives of sulfate-reducing microorganisms for which sequence data

[8] depression: includes the ENVO terms 'canyon', 'continental shelf', 'drainage basin', 'trough', 'back-arc basin'. lentic water body: includes the ENVO terms 'lake', 'reservoir'. lotic water body: includes the ENVO terms 'hydrothermal vent', 'acid mine drainage'. marine biome/sediment/water body: includes the ENVO terms 'sea', 'marine habitat', 'ocean water'. saline water: includes the ENVO terms 'saline water', 'sea water'. sea coast: includes the ENVO terms 'bay', 'fjord'. wetland: includes the ENVO terms 'saline marsh', 'fen'.

are deposited in public databases, there are still many more that remain to be cultured and described. As expected, the vast majority of SRMs were affiliated to the *Deltaproteobacteria* supercluster, and in particular *Desulfobacteraceae*, which have been shown to exhibit high abundances in marine (Leloup *et al.*, 2009), saline and hypersaline (Foti *et al.*, 2007) and deep sea sediments (Kaneko *et al.*, 2007). However, the abundance of the *Desulfobacteraceae* family fluctuated according to the salinity, i.e. its abundance was lower in the a-mesohaline lagoons and higher in the more saline lagoons, although it has also been reported from freshwater sediments (Wang *et al.*, 2012).

The high abundance of sequences belonging to Environmental supercluster 1, which comprises sequences from uncultured microorganisms (Müller *et al.*, 2015), is indicative of the lack of knowledge of SRM diversity in the sampling sites and in lagoonal habitats in general. In addition, the effect of salinity on the distribution of SRM in the studied habitat is evident from the abundance of family-level Lineage 9 that is composed of many sequences from the marine environment (Müller *et al.*, 2015) and from the absence (Lineages 6 and 10) or very low abundance (Lineage 8) of lineages often detected from freshwater wetlands (Pester *et al.*, 2012). Apart from these findings, when analyzing each lagoon separately, there were certain differences in the abundance of the groups present. Specifically, the *Desulfatiglans anilini* lineage exhibited greater abundance in lower salinities; it has been previously found in both riverine (Suzuki *et al.*, 2014) and marine sediments (Schnell *et al.*, 1989; Ahn *et al.*, 2009), although it has been isolated from marine enrichment cultures, inoculated with mud from the North Sea coast, using brackish water medium (Schnell *et al.*, 1989).

However, these results were undoubtedly affected by the reference dataset used for the classification of the sequences, where most of the included sequences are derived from marine environments, followed by freshwater and other environments (Müller *et al.*, 2015). This succession of environments is also depicted in the retrieved environmental terms that were found to be associated with sequences similar of ones retrieved from the present study; samples were mainly characterized by environmental descriptive terms that could be broadly classified to the marine biome, while terms belonging to the freshwater or brackish biome were found to a

lesser extent. This could suggest that in environments of intermediate salinity concentrations, there is still an unknown component of the SRM diversity that remains to be investigated and incorporated to our knowledge of SRM communities. However, there were certain environmental descriptive terms, such as the term 'wetland', that contributed more to the specificity of the lower salinity lagoons. In addition, the influence of Louros river in the SRMs of Tsopeli lagoon may be deciphered from the association of the OTUs found in this lagoon with the terms 'lake' and 'reservoir'.

The SRM community diversity pattern seems to differ in each lagoon, a finding which is in accordance with previous reports of the total community diversity pattern from the same study sites (Pavloudi *et al.*, 2016). This finding also concurs with those of previous studies, which have shown that SRM communities exhibit biogeographic distribution patterns at small spatial scales and that a homogeneous distribution is not unlikely (Pérez-Jiménez and Kerkhof, 2005). In addition, the SRM community diversity pattern can be clustered according to broad salinity categories; this indicates that salinity is one of the major factors influencing the SRM communities in this habitat which is at the interface of marine and freshwater. This can be also concluded from the significant differentiation of the diversity indices between the salinity categories.

Although sulfate reducers are named after their ability to use sulfate as a terminal electron acceptor, they can in fact use many different electron acceptors, such as nitrate and nitrite (Dalsgaard and Bak, 1994; Moura *et al.*, 1997) or other sulfur compounds (thiosulfate, sulfite and sulfur) (Muyzer and Stams, 2008). Therefore, the detection of *dsrAB* gene sequences in environmental samples should not be regarded per se as actual physiological capability for dissimilatory sulfate/sulfite reduction (Pester *et al.*, 2012; Müller *et al.*, 2015), i.e. the occurrence of high abundance of SRMs does not necessarily reflect the occurrence of sulfate reduction in the respective environment (Muyzer and Stams, 2008). This may be the reason why our results suggest that the concentration of nitrite ions is correlated to the SRM community pattern. In addition, SRMs might be linked to ammonium on a secondary level, by using products of anaerobic protein degradation, which releases acetate, H_2 and ammonium. The lack of correlation with the concentration of total

reduced inorganic sulfur (TRIS), which was also among the tested variables, may be explained by the fact that it is the sum of hydrogen sulfide (H_2S), iron sulfide (FeS), elemental sulfur (S^0) and iron pyrite (FeS_2) (Fossing and Jørgensen, 1989) and thus it cannot reflect only the biogenically produced sulfide (S^{2-}) (Jong and Parry, 2003). However, although there is no conclusive evidence to support this, it can be suggested that the sulfate reducers found at this particular study site and at this particular time point, probably were using sulfate and/or nitrogen compounds as electron acceptors.

3.5 Conclusions

The results of this exploratory study show that the majority of the observed operational taxonomic units (OTUs) belong to the *Deltaproteobacteria* supercluster and more specifically, to the *Desulfobacteraceae* family. Salinity and ammonium ions are the environmental factors that best correlated with the SRM community pattern. Furthermore, the SRM community of the brackish lagoons is differentiated from that of the brackish-marine lagoons and the studied lagoons have distinct SRM communities.

Further investigation is needed to shed light on the functionality of SRMs in lagoonal ecosystems, especially in terms of their viability and competition with each other for the available sulfate, when the latter is insufficient for complete oxidation of organic compounds. In addition, the seasonality of SRMs should be investigated given that the environmental variables that primarily influence the community pattern are subjected to seasonal changes.

Chapter 4: Identification of microbial communities by RNA stable isotope probing and 16S rRNA sequencing from the Benguela coastal upwelling system

Citation: Pavloudi C, Sztejrenszus SY, Kristoffersen JB, Lahajnar, N, De Troch M, Arvanitidis C, Friedrich MW (2017) Identification of microbial communities by RNA stable isotope probing and 16S rRNA sequencing from the Benguela coastal upwelling system. *Environmental Microbiology*, under review

Abstract

The Benguela coastal upwelling system is characterized by the highest primary productivity compared to other upwelling regions, episodic occurrence of free hydrogen sulfide gas and formation of an oxygen minimum zone (OMZ). RNA-based stable isotope probing (SIP) was used to identify nitrate and sulfate reducing microorganisms from three different sediment sampling stations using ^{13}C acetate as labelled substrate.

Labelling patterns of microbial communities, as assessed by high-throughput sequencing of 16S rRNA, varied across SIP incubations and depended on sampling station. When no external electron acceptor was added, an increase in the abundance of *Epsilonproteobacteria* was observed at two stations but of *Gammaproteobacteria* at the third station, which had a much lower water depth. In addition, an increase in *Epsilonproteobacteria* was observed both when nitrate or sulfate were added.

It can be concluded that nitrate stimulated nitrate-reducing, sulfide-oxidizing bacteria, and inhibited the growth of sulfate-reducing bacteria. Furthermore, sulfate addition did not enhance the abundance of known sulfate-reducers, such as *Deltaproteobacteria*. This could be attributed to the competition for electron donors between nitrate-reducers and sulfate-reducers, to the inability of certain sulfate-

reducing bacteria to use acetate as an electron donor or to the short duration of the incubations.

4.1 Introduction

Upwelling is a common feature of the continental shelf, induced by wind, Coriolis effects and Ekman transport (Summerhayes, 2015). The Benguela coastal upwelling system is one of the two main coastal upwelling systems of the southern hemisphere (Moloney *et al.*, 2005) and one of the five coastal upwelling regions globally (Summerhayes, 2015). It is situated along the coast of south western Africa, with its south boundary located east of the Cape of Good Hope and the north reaching Angola waters (Shannon and O'Toole, 2003). In terms of primary productivity, it is the most productive of the subtropical eastern boundary current regions (Carr, 2001), i.e. Humboldt, California and Canary, although in terms of fish catch it is the least productive (Lavik *et al.*, 2009) with Humboldt yielding more than 20 times the fish tonnage of Benguela system (Carr, 2001). This has been attributed to the episodic occurrence of hydrogen sulfide gas (Brüchert *et al.*, 2009) which is a potent respiratory toxin to aquatic organisms (Bagarinao, 1992).

Enhanced primary productivity causes a subsequent increase of sinking organic matter, which in turn promotes increased microbial respiration rates; thus, bottom water in upwelling ecosystems is susceptible to oxygen deficiency (Monteiro *et al.*, 2006; Diaz and Rosenberg, 2008). In cases where oxygen concentration is lower than 20 µM, oxygen minimum zones (OMZs) are formed (Lam and Kuypers, 2011; Wright *et al.*, 2012). Such areas are characterized by unique microbial communities, due to the availability and usage of alternative electron donors and acceptors for respiration (Lam and Kuypers, 2011; Cassman *et al.*, 2012). Heterotrophic denitrification and anaerobic ammonium oxidation (anammox) are prevailing in OMZs and are responsible for massive losses of fixed nitrogen (Gruber and Sarmiento, 1997; Codispoti *et al.*, 2001; Kuypers *et al.*, 2005). The intensity of sulfur cycling is also apparent in OMZs and it has also been argued that there is an active coupling between nitrate reduction and sulfide oxidation (Canfield *et al.*, 2010).

The high concentration of organic carbon in the Namibian shelf (up to 15% in surface sediments; Inthorn *et al.*, 2006b), combined with the low oxygen concentration, renders it an ideal location for the identification of microbial communities involved in anaerobic degradation of organic matter. Organic matter turnover in this area is thought to be mediated primarily by sulfate reduction according to Brüchert *et al.* (2003), with other anaerobic bacterial degradation processes being less significant.

Microorganisms in OMZs have, undoubtedly, been at the forefront of microbial ecology research using different methodologies, such as quantitative PCR (Lam *et al.*, 2009; Pitcher *et al.*, 2011) or cloning and sequencing of functional genes (Kong *et al.*, 2013), metatranscriptomics (Stewart *et al.*, 2012) and metagenomics (Bryant *et al.*, 2012). To directly link microorganisms with their metabolic capabilities, stable isotope probing (SIP) (Friedrich, 2006) was employed in this study.

SIP is based on the incorporation of a substrate enriched in isotopic composition into the cellular compounds of microorganisms that are active during the labelling process. Commonly used molecular markers for SIP are DNA (Radajewski *et al.*, 2000) and RNA (Manefield *et al.*, 2002); since the labelled substrate (e.g. ^{13}C) is incorporated into nucleic acids, it increases their buoyant density and thus, isopycnic centrifugation can be used to separate the labelled from non-labelled molecules (Radajewski *et al.*, 2003; Dumont and Murrell, 2005; Aoyagi *et al.*, 2015). RNA can be synthesized in higher rates than DNA; this, along with the fact that RNA reflects the overall cellular activity while DNA reflects the rate of replication, render the former a more sensitive marker for SIP studies (Manefield *et al.*, 2002).

^{13}C labelled acetate has been widely used as an electron donor in SIP studies in lake (Schwarz *et al.*, 2007) and marine sediments (Vandieken and Thamdrup, 2013), activated sludge (Osaka *et al.*, 2006) and soil (Chauhan and Ogram, 2006; Hori *et al.*, 2010). Acetate is a main product of microbial fermentation that has been shown to be an important substrate for anaerobic terminal electron-accepting processes (Hori *et al.*, 2010; Vandieken and Thamdrup, 2013). One example of these processes is sulfate reduction (Fukui *et al.*, 1997), where acetate and hydrogen are the major electron donors used (Finke *et al.*, 2007), with the former accounting for 40 – 50% of the total sulfate reduced in marine sediments (Sørensen *et al.*, 1981; Parkes *et al.*, 1989). Sulfate reducers and nitrate reducers have been shown to use a variety of

electron donors, such as acetate, propionate, butyrate, methanol, ethanol, glucose and sucrose (Achtnich *et al.*, 1995; Paul *et al.*, 1989; Hallin *et al.*, 1996; Ginige *et al.*, 2005; Osaka *et al.*, 2006; Baytshtok *et al.*, 2009). However, acetate is rather difficult to ferment for thermodynamic reasons; therefore, it can be ideally used as a specific substrate for identifying anaerobic respiratory microbes (Hori *et al.*, 2010).

The aim of the present study was to identify sulfate-reducing and denitrifying microorganisms in the Benguela coastal upwelling system using RNA SIP with ^{13}C labelled acetate. The combination of SIP with subsequent 16S rRNA sequencing was used to reveal the missing link between the structure and function of microorganisms involved in anaerobic degradation of organic matter in the Namibian shelf sediments. The tested hypothesis was that different microbial populations would become labelled, based on the different sampling locations and on the terminal electron acceptors used, which were nitrate and sulfate.

4.2 Materials and Methods

4.2.1. Study area and sampling strategy

The Benguela upwelling area is part of the eastern boundary current system of the South Atlantic subtropical gyre (Lahajnar *et al.*, 2015). The working area included the weak upwelling cell offshore Walvis Bay (23° S), the Terrace Bay (20° S) and Kunene cell (17°15′ S) in the north and the very stable upwelling cell close to Lüderitz (25-27°S) as well as oxygenated water areas at the Namibian - South African border (28°38′ S) in the south.

During R/V METEOR cruise M103/1 in the austral summer season (December 2013 - January 2014), when low to moderate upwelling conditions were prevailing (Lahajnar *et al.*, 2015), sediment samples were collected using an OCTOPUS multicorer that was equipped with eight polyacryl tubes (60 cm length; 10 cm diameter). Samples for nucleic acid extraction were collected from 27 stations along the sampled transects (Table 4.1; Figure 4.1) from the upper 2 cm of sediment cores and stored at -80 ^{o}C until return to the laboratory. Wherever it was possible, replicate samples

were collected from each station. Physicochemical measurements at the sampling stations were conducted by other cruise participants and were submitted at PANGAEA Data Publisher (https://www.pangaea.de/) under the project label GENUS (unpublished datasets #859049 and #854182).

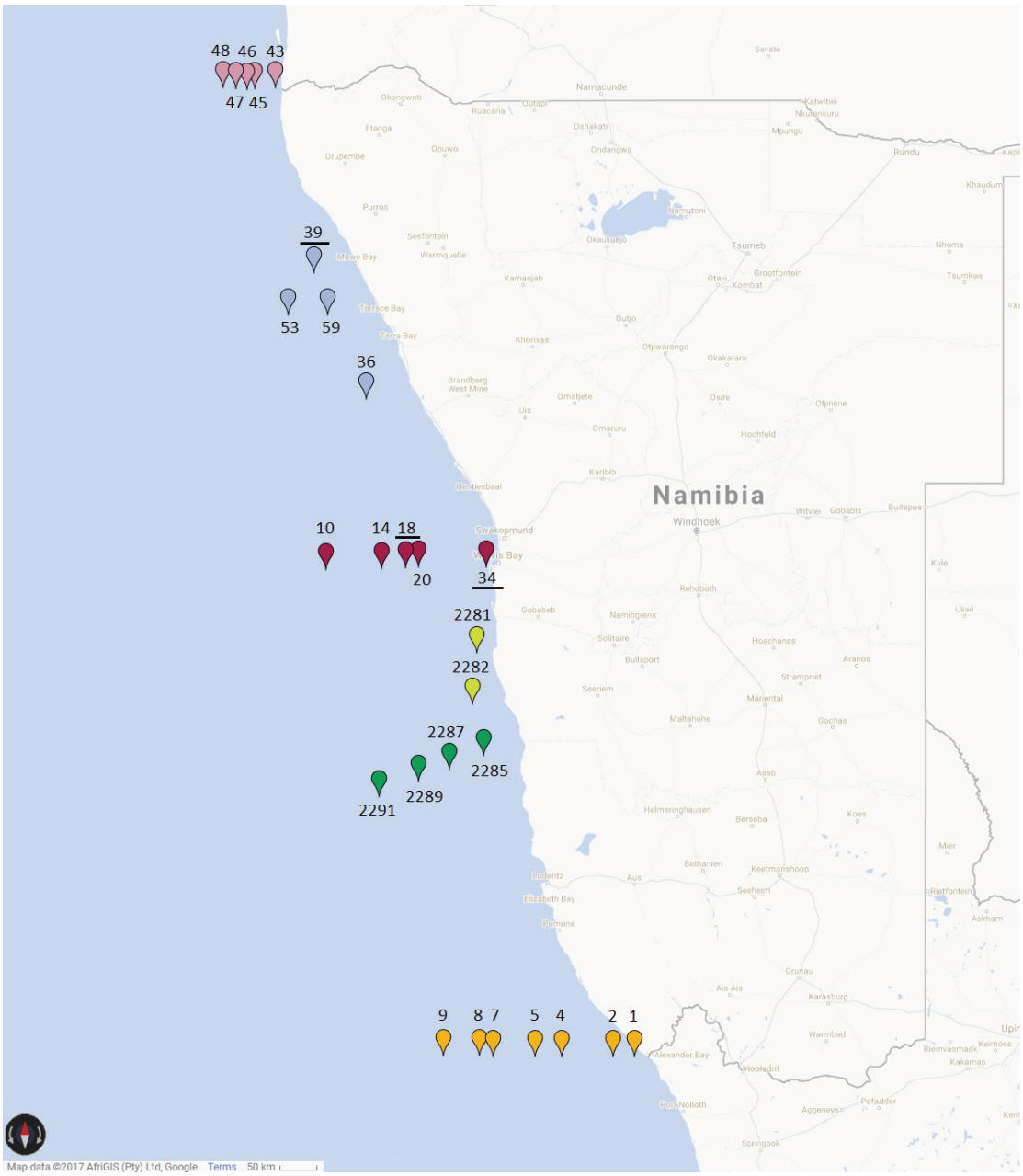

Figure 4.1: Map showing the location of the sampling stations (Map data: Google). Colours are differentiating the sampling stations according to the sampled transect. Sampling stations were SIP was employed are underlined.

Table 4.1: Overview of the stations sampled. SIP: Stable-isotope probing.

Station	Date	Latitude	Longitude	Depth (m)	Number of replicate samples
1	1/1/2014	28° 38.00' S	16° 16.01' E	40.6	1
2	1/1/2014	28° 38.15' S	15° 59.75' E	113.7	1
4	2/1/2014	28° 37.96' S	15° 19.97' E	187.4	1
5	2/1/2014	28° 38.00' S	14° 59.95' E	170.7	3
7	2/1/2014	28° 38.41' S	14° 25.26' E	358.2	1
8	3/1/2014	28° 37.99' S	14° 14.95' E	728	2
9	3/1/2014	28° 38.02' S	13° 47.02' E	2032.8	2
10	5/1/2014	23° 2.28' S	12° 18.59' E	2099.4	3
14	6/1/2014	23° 0.98' S	13° 1.96' E	454.6	2
18***	7/1/2014	23° 0.74' S	13° 20.24' E	350.8	2
20	7/1/2014	23° 0.18' S	13° 30.16' E	233.7	3
34**	9/1/2014	23° 0.01' S	14° 21.98' E	39.3	2
36	9/1/2014	21° 0.72' S	12° 49.96' E	302.3	2
39	10/1/2014	19° 29.63' S	12° 9.99' E	233.2	3
43	11/1/2014	17° 15.04' S	11° 39.84' E	80.9	2
45	12/1/2014	17° 15.47' S	11° 23.94' E	243.9	2
46	12/1/2014	17° 15.82' S	11° 18.06' E	453.8	2
47	12/1/2014	17° 15.49' S	11° 10.04' E	1015.4	3
48	13/1/2014	17° 15.01' S	10° 59.94' E	2115.2	2
53	14/1/2014	19° 59.97' S	11° 49.92' E	408.4	2
59	15/1/2014	19° 59.95' S	12° 19.97' E	212.5	2
2281	28/12/2013	24° 0.02' S	14° 14.90' E	122.3	3
2282	28/12/2013	24° 36.02' S	14° 11.98' E	144.6	3
2285	29/12/2013	25° 11.95' S	14° 19.94' E	153.1	2
2287	29/12/2013	25° 21.03' S	13° 53.79' E	250.1	1
2289*	30/12/2013	25° 29.32' S	13° 30.01' E	700.2	2
2291	31/12/2013	25° 39.99' S	13° 0.00' E	2225.9	1

*: close proximity with station GeoB 12802 (25° 30' S, 13° 27' E) of Evans *et al.* (2017). **: close proximity with stations 1 (25° 51.9' S, 14° 28.9' S E) and 2 (22° 48.9' S, 14° 26.9' E) of Brüchert *et al.* (2003). ***: close proximity with station 8 (23° 11' S, 13° 23.8' S E) of Brüchert *et al.* (2003).

4.2.2. Stable isotope probing (SIP) experiment

Samples from three stations (Table 4.2) from the upper 2 cm of sediment cores were collected for SIP experiments. Sediment slurries were prepared on board by mixing the sediment with autoclaved anoxic sulfate-free artificial seawater (ASW) (containing 26.4 g NaCl, 11.2 g $MgCl_2 \bullet 6H_2O$, 1.5 g $CaCl_2 \bullet 2H_2O$ and 0.7 g KCl per

liter of deionized water) in a ratio of 1:2. The slurries were homogenized anoxically under a stream of N_2 and 15 ml slurry each were transferred into serum bottles.

Table 4.2: General characteristics of the stations were SIP was employed.

Station	Sediment type	Depth (m)	Oxygen concentration (umol/l)	Temperature (^oC)	Salinity (psu)
18	brown grey fine sand (fS)	350.8	59	9.2	34.78
34	dark green mud, H_2S smell, well layered	39.3	32	12.76	35.15
39	dark green mud, foraminifera, shell fragments	233.2	29.5 *	14.17 *	35.39 *

*: values are the average of those measured at the stations 38 (20° 0.64' S, 12° 29.72' E) and 40 (19° 0.54' S, 12° 9.92' E), sampled the same day by the cruise and at about 50 km distance from station 39.

Subsequently, substrates were added to microcosms. Two microcosm experiments were conducted per sampling station (in duplicates) either by adding nitrate or sulfate as electron acceptors, respectively (Table 4.3; Figure S4.1). Sulfate was added to a final concentration of 28 mM. Nitrate was added 4 times at a final concentration of 2.5 mM each time, to avoid the production of high toxic levels of nitrite.
1,2 - $^{13}C_2$ 99% sodium acetate (CAS 56374-56-2, Cambridge Isotope Laboratories, USA) was added to the samples as an electron donor to a final concentration of 5 mM.

Table 4.3: Summary of the stable isotope experiments conducted on board.

	Electron acceptor (substrate)	Electron donor	Sampling	Substrate addition	Sample code
Nitrate microcosms	Nitrate	^{13}C - acetate	Day 7	Day 0, 2, 4, 6	A
		^{12}C - acetate			D
		No addition (control)			H
Sulfate microcosms	Sulfate	^{13}C – acetate		Day 0	B
		^{12}C – acetate			E
		No addition (control)			I
Control microcosms	No addition (control)	^{13}C - acetate		-------	C
		^{12}C – acetate			F
		No addition (control)			G

Along with the ^{13}C labelled samples, a series with unlabelled sodium acetate was used as control (incubation with native substrate, i.e. ^{12}C), as proposed by Dunford and Neufeld (2010); these samples provided a subsequent comparison to ensure that any apparent labelling of nucleic acid was not an artifact of the ultracentrifugation or G+C content density differences in RNA contributing to separation (Neufeld *et al.*, 2007).

In addition, another series of samples was used as a no substrate control, thus without acetate addition, to access background population changes during the SIP incubation (Table 4.3; Figure S4.1).

Serum bottles were sealed with mushroom stoppers and the headspace was flushed with N_2 to achieve incubation at anoxic conditions. Sediment slurries were incubated in the dark at a temperature of ~15 °C for 3 days, when half of the slurry was removed from bottles and stored at -80 °C. The headspace was again flushed with N_2 and the incubation was continued until six days were concluded. At the seventh day, the rest of the slurry was removed from the bottles and stored at -80 °C until further processing.

4.2.2.1 Nucleic acid extraction

RNA and DNA were co-extracted from the last day of all microcosm incubations, and all the other sediment samples, as performed by Henckel *et al.* (1999) and Lueders *et al.* (2004). Approximately 0.7 g of wet sediment were added to a 2-ml screw-cap vial, prefilled with ~0.7 g of baked (3 h, 180 °C) 0.1 mm (diameter) zirconia/silica beads (11079101z, BioSpec, USA). The vials were filled with 750 µl of 120 mM $NaPO_4$ buffer (pH 8) and 250 µl TNS solution (Table S4.1) and placed in a bead beater for 45 s at 6.5 m/s. Immediately after that the vials were centrifuged for 10 min at 20,800 rcf and 4 °C and the supernatants were transferred to new 2-ml vials. For nucleic acid extraction, one volume of phenol/chloroform/isoamylalcohol (P/C/I; 25:24:1; pH 5; Carl-Roth, Karlsruhe, Germany) was added to the aqueous supernatant. Vials were vigorously shaken for 20 s and centrifuged for 5 min at 20,800 rcf and 4 °C. Supernatants were transferred to new 2-ml vials, and one volume of

chloroform/isoamylalcohol (C/I; 24:1; Carl-Roth) was added. Vials were again vigorously shaken for 20 s and then centrifuged for 5 min at 20,800 rcf and 4 °C. Supernatants were transferred to new 2-ml vials and C/I extraction was repeated to successfully remove all phenol remnants. Supernatants were transferred to new 2-ml vials and 1.5 ml of polyethylene glycol (PEG; Table S4.1) was added to precipitate nucleic acids and the vials were centrifuged for 90 min at 20,800 rcf and 4 °C. Supernatants were discarded and the pellets were washed with 1 ml 70% ethanol (4 °C) and centrifuged for 30 min. Supernatants were again discarded, pellets were left for air drying (~5 min) to remove leftover ethanol and resuspended with 50 µl nuclease free diethylpyrocarbonate (DEPC) treated water (Carl-Roth). Each replicate sample was extracted separately.

In order to remove DNA contamination from the microcosm extractions, DNA was digested using RQ1DNase (Promega) using the manufacturer's protocol. The remaining RNA was precipitated with 1 ml PEG followed by centrifugation for 1 h at 20,800 rcf and 4 °C. Supernatants were discarded and the pellets were washed with 500 µl cold 70% ethanol (4 °C) and centrifuged for 30 min. Supernatants were again discarded, pellets were left for air drying (~5 min) to remove leftover ethanol and resuspended with 40 µl nuclease free DEPC treated water.

The absence of DNA, i.e. the success of the DNAse digestions, was verified by failure of 16S rRNA gene targeting PCR amplification. Finally, the RNA extracts were quantified using RiboGreen assay (Invitrogen Quant-iT RiboGreen RNA assay kit, Life technologies).

4.2.2.2 Pure culture standards

Escherichia coli cells were grown on fully ^{13}C-labelled or unlabelled medium (*E. coli* OD2 ^{13}C labelled- 110201102; *E. coli* OD2- 100002, Silantes, Munich, Germany) and RNA was extracted in order to be used as a known ^{13}C-labelled and unlabelled gradient marker during the isopycnic centrifugations.

4.2.2.3 Isopycnic centrifugation

RNA was separated on a density gradient using isopycnic centrifugations. Density gradient centrifugation was done as in Sztejrenszus (2016) who modified the Lueders *et al.* (2004) protocol. Briefly, ~375 ng of each replicate RNA sample, i.e. ~750 ng in total, were mixed with 240 µl of deionized formamide, 6 ml of cesium trifluoroacetate (CsTFA; illustra CsTFA- 17084702, GE Healthcare, UK) and Gradient Buffer (GB; Table S4.1) to a volume of 1.3 ml. For the *E. coli* standard, both labelled and unlabelled RNA were mixed in one sample. Starting density was adjusted to 1.796 – 1.799 g/ml using an AR200 digital refractometer (Reichert Analytical, NY USA). Afterwards, ~6.5 ml of each mixture was transferred to a Beckman polyallomer Quick Seal 16 x 45 mm tube (Catalog number 345830, Beckman Coulter, USA) sealed and spun in a VTI 65.1 vertical rotor in an Optima XE-90 ultracentrifuge (both Beckman Coulter) at 124,000 rcf and 20 °C for 65 h.

After ultracentrifugation, gradient fractions of ~400 µl were collected at a flow rate of 1 ml/min (Aladdin syringe pump, AL-1000, WPI, Berlin, Germany), and density of fractions was measured using a refractometer. Fractionation resulted in 13 – 14 fractions with a density range of 1.77 – 1.84 g/ml. In order to precipitate RNA from fractions, 400 µl isopropanol and 80 µl sodium acetate 3M were added to each one and incubated at -20 ^{o}C overnight, followed by centrifugation at 20,800 rcf and 4 °C for 1 h. Supernatants were removed and 1.2 ml of 70% ethanol (4 °C) was added to each sample and centrifuged for 30 min at 20,800 rcf and 4 °C. Supernatants were again removed, pellets were left for air drying (~5 min), and RNA was resuspended with 20 µl nuclease free DEPC treated water. RNA was quantified fluorometrically using the RiboGreen assay.

From each sample, the isotopically “light” and “heavy” fractions (~1.79 and ~1.82 g/ml, respectively) corresponded to densities where ^{13}C-labelled and unlabelled *E. coli* RNA were found (Figures S4.2-S4.4). RNA from those fractions was transcribed to cDNA using Applied Biosystems High Capacity cDNA Reverse Transcription Kit (Life technologies) following the manufacturer's protocol.

4.2.3 PCR amplification and sequencing of the 16S rRNA gene

cDNA from the isotopically "light" and "heavy" fractions, as well as DNA from the other sediment samples, was amplified using S-DBact-0341-b-S-17 (or 341F) as a forward primer and S-D-Bact-0785-a-A-21-B (or 805RB) as a reverse primer (Herlemann *et al.*, 2011; Klindworth *et al.*, 2013), with a revision in the reverse primer for detection of SAR11 bacterioplankton (Apprill *et al.*, 2015), thus targeting the V3-V4 region of the 16S rRNA gene. In certain cases, in order to enhance nucleotide diversity of the sequencing run and increase the available primer pair combinations, primers having an extra 5 bp barcode were used.

For Illumina library preparation, the two-step PCR approach was used. The first-step PCR was performed with the aforementioned primers containing a universal 5' tail as specified in the Nextera library protocol from Illumina (tail sequence for the forward primer: 5′ - TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG - 3′, tail sequence for the reverse primer: 5′ - GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG - 3′). The amplification reaction mix of the first PCR contained 4 µl 5x KAPA HiFi Fidelity buffer (containing 2 mM Mg^{2+} at 1X), 2 µl trehalose (1 M), 0.5 µl KAPA dNTP Mix (10 mM), 1 µl of each primer (10 µM), 0.5 µl KAPA HiFi HotStart DNA polymerase (1 U/µl) in a final volume of 20 µl per reaction. DNA template concentration was about 10 ng/µl. The first PCR protocol used was the following: 95 ^{o}C for 5 minutes; 30 cycles at 98 ^{o}C for 20 seconds, 57 ^{o}C for 2 minutes, 72 ^{o}C for 1 minute; 72 ^{o}C for 7 minutes. Negative control (PCR blank) samples, i.e. samples with no nucleic acid template, were also included in the first-step PCR in order to assess possible source of contamination in the amplification reactions.

The resulting PCR amplicons (531 bp length), including the negative control samples, were purified using Agencourt AMPure XP (Beckman Coulter, Brea, CA, USA) and were used as templates for the second-step PCR in order to include the indexes (barcodes) as well as the Illumina adaptors.

The amplification reaction mix of the second PCR contained 4 µl 5x KAPA HiFi Fidelity buffer, 0.5 µl KAPA dNTP Mix (10 mM), 2 µl from each primer (10 µM), 0.5 µl KAPA HiFi HotStart DNA polymerase (1 U/µl) in a final volume of 20 µl per reaction. DNA

template concentration was about 20 ng/μl. The second PCR protocol used was the following: 95 ^{o}C for 3 minutes; 8 cycles at 98 ^{o}C for 20 seconds, 55 ^{o}C for 30 seconds, 72 ^{o}C for 30 seconds; 72 ^{o}C for 5 minutes. Amplifications were carried out using T100™ Thermal Cycler (BIORAD). Again, the resulting PCR amplicons (600 bp) were purified and mixed in equimolar amounts to construct the library that was sequenced using a MiSeq Reagent Kit v3 (600-cycles) at the IMBBC (HCMR).

These raw sequence data were submitted to the Sequence Read Archive (SRA) (Leinonen *et al.*, 2011) under the study accession number PRJEB20585 (available at http://www.ebi.ac.uk/ena/data/view/PRJEB20585).

4.2.4 Sequence processing

The raw sequence reads retrieved from all the sediment samples were submitted for processing, using the UPARSE based analysis pipeline (Edgar, 2013), to the Integrated Microbial Next Generation Sequencing (IMNGS) (Lagkouvardos *et al.*, 2016), under the following options: a) number of allowed mismatches in the barcode: 2, b) minimum fastq quality score for trimming of unpaired reads: 3, c) minimum length for single reads or amplicons for paired overlapping sequences: 437, d) maximum length for single reads or amplicons for paired overlapping sequences: 656, e) maximum number of expected errors in paired sequences: 3, f) length of trimming at the forward side of the seqs: 10, g) length of trimming at the reverse side of the seqs: 10 and h) minimum relative abundance of OTU cutoff (0-1): 0.0050.

Pairing, quality filtering and OTU clustering (97% identity) was done by USEARCH 8.0 (Edgar, 2010). Chimera filtering was performed by UCHIME (Edgar *et al.*, 2011) (with RDP set 15 as a reference database). RDP classifier version 2.11 training set 15 (Wang *et al.*, 2007) was used for the taxonomic classification of the OTUs and MUSCLE (Edgar, 2004) was used for the sequence alignment.

4.2.5 Statistical processing

OTUs that were found in the negative controls (Table S4.2) were removed from the samples, in order for a more accurate description of the microbial communities to be made.

A matrix containing the microbial OTUs as variables and sampling stations as samples was constructed. The number of sequences assigned to a given OTU was considered to reflect its relative abundance at each of the 27 sampled stations. The matrix was subsequently standardized (i.e. sample values were divided by the total for each sample), transformed (square root), and used for the calculation of the triangular similarity matrices using the Bray-Curtis similarity coefficient (e.g. Clarke and Warwick, 1994). The similarity matrix was used as an input for the non-metric multidimensional scaling (nMDS) (Clarke, 1993) and permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001), under the null hypothesis that no differences among microbial assemblages exist. The design considered three factors: “transect”, “depth range” and “range of oxygen concentration” (999 permutations). Each factor was tested separately (under a design of unrestricted permutation of raw data) because of data limitations that did not allow for a successful three-factor design to be tested.

A second matrix containing the physicochemical measurements at the sampled stations was also constructed. The values were normalized (i.e. the values for each variable had their mean subtracted and were divided by their standard deviation) and used for the calculation of similarity patterns by the standardized Euclidean distance. In order to test whether the biotic and abiotic matrices were correlated, the RELATE routine (Clarke and Gorley, 2006) was used. In addition, BIO-ENV analysis was used to examine all potential correlations between the biotic and abiotic matrices by employing the weighted Spearman’s rank coefficient (Clarke and Ainsworth, 1993) and after the exclusion of highly correlated ($-0.8 > \rho_w > 0.8$) physicochemical variables.

A suite of diversity indices (Margalef’s species richness, Pielou's evenness, Shannon-Wiener (Pielou, 1969), Simpson) was calculated for each sampling station.

The aforementioned multivariate analyses were performed with the PRIMER (v.6) package developed in Plymouth Marine Laboratories (Clarke and Gorley, 2006). The vegan package (Oksanen *et al.*, 2016) was used for the calculation of diversity indices using R version 3.2.1 (R Core Team, 2015).

4.3 Results

4.3.1 Microbial community composition

In order to assess the composition of the microbial community *in situ*, prior to isotope probing incubations, sediment samples (upper 2 cms) from 27 stations (Figure 4.1) were subjected to 16S rRNA gene amplicon sequencing. An overview of the processing of the obtained sequences is available in Table S4.2. As shown in Figure 4.2, where the relative abundance percentages of each replicate sample have been averaged per sampling station, the majority of the OTUs were classified as *Deltaproteobacteria*, followed by *Gammaproteobacteria* and unidentified *Bacteria*.
When the abundance of the taxonomic groups was compared among the stations (Figure 4.2), some clear patterns were observed. *Epsilonproteobacteria* were highly abundant at station 34 (~7%) and station 48 (~2%); in other stations they were found with much lower abundances. The highest abundances of *Deltaproteobacteria* were found at station 2287 (~33%), station 53 (~27%) and station 45 (~28%), while the lowest were found at station 9 (~4%), station 10 and station 34 (~9% in both cases). Regarding *Gammaproteobacteria*, station 9 and station 43 had the highest abundances (~24 and 25% respectively); the lowest was found at stations 8, 2285, 34 and 39 (~7% on average). *Alphaproteobacteria* were most abundant at station 9 (~30%) and station 10 (~15%). *Actinobacteria* were mainly present at stations 14, 1 and 45 (~17% on average) while almost absent from stations 2282 and 34 (~2% on average). *Flavobacteria* were abundant at station 2281 (~20%) but decreased in abundance in all the other stations. *Ignavibacteria* were more abundant at station 2285, station 2282 (~15% on average) and station 59 (~10%).

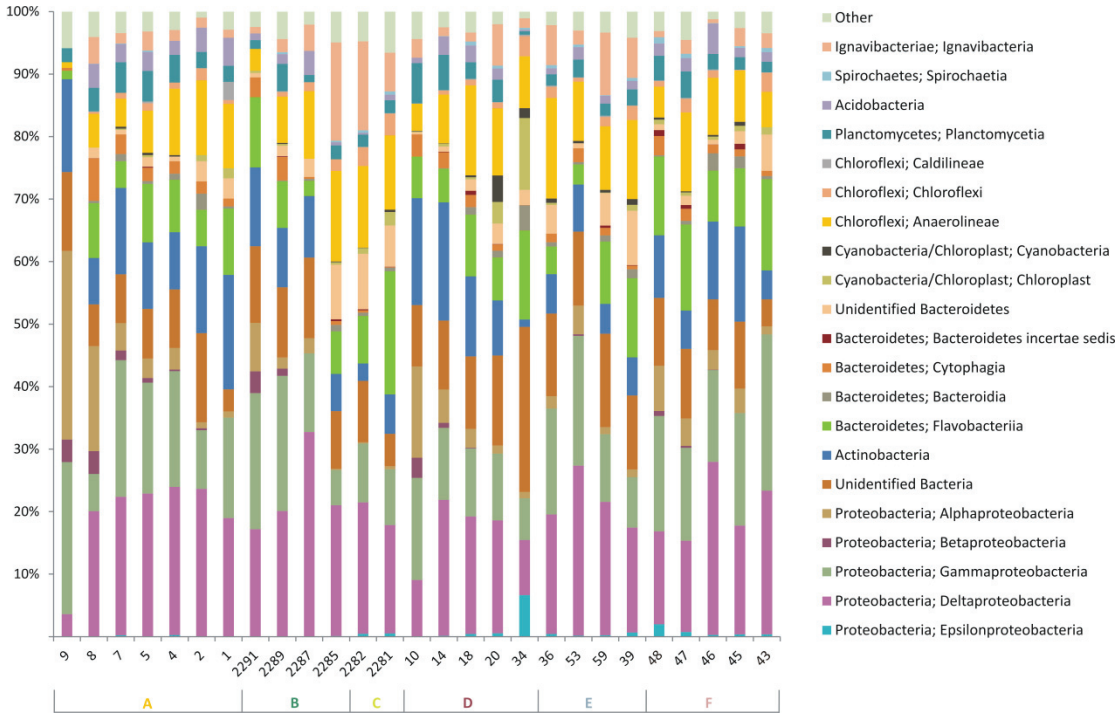

Figure 4.2: The relative abundances (%) of the main microbial taxonomic groups at the sampling stations. A: zonal transect at 28° 38′S. B: cross shelf transect starting at 25°S. C: stations between 23°S and 25°S. D: zonal transect at 23°S. E: stations between 19°S and 21°S. F: zonal transect off Kunene mouth at 17° 15′S. Colours of A, B, C, D, E and F according to sampling transect, based on Figure 4.1.

In terms of diversity (Table S4.3), there was no evidence of a general correlation pattern with the physicochemical measurements of the sampling stations. However, a negative correlation was observed in a few cases (Table S4.4), for instance, the relationship between salinity and total microbial relative abundance values.

The nMDS of the microbial OTUs spatial pattern (Figure 4.3) showed a grouping according to the depth range of the sampling stations. This was also confirmed by the PERMANOVA test which produced significant results for all the tested factors (transect: Pseudo-F = 1.5161, $p < 0.05$, depth range: Pseudo-F = 2.8826, $p < 0.01$, range of oxygen concentration: Pseudo-F = 2.9127, $p < 0.01$). Pairwise tests (Table S4.5) showed that there were significant statistical differences between the lowest ranges of bottom oxygen concentration (up to 30 µmol/l) and the highest ones (from 50 up to 200 µmol/l). In addition, microbial communities of the deepest stations

(1000-2500 m) were significantly different from the ones that were up to 500 m deep.

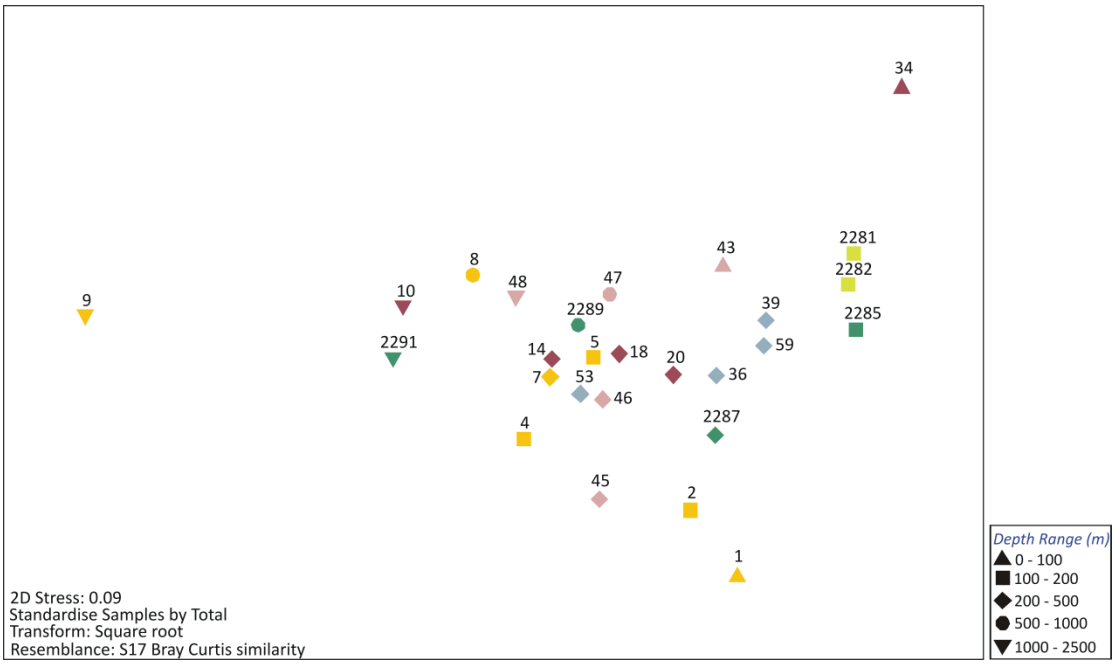

Figure 4.3: nMDS of the microbial OTUs. Symbols according to depth range of the sampling stations. Colour according to sampling transect, based on Figures 4.1 and 4.2.

The average values of the physicochemical variables measured at the sampling stations are presented in Table 4.4. The results of the RELATE routine showed that the biotic matrix is related to the abiotic one ($\rho_w = 0.627$, $p < 0.01$). In more detail, the combination of delta ^{15}N, liquid water content (LWC) and bottom water oxygen concentration are best correlated with the microbial community pattern and are present in all the variable combinations (Table 4.5).

Table 4.4: The average values of the physicochemical variables of the sampling stations.

	Surface Sediment								Bottom Water			
Station	Total Nitrogen (%)	Total Carbon (%)	Total Organic Carbon (%)	Calcium Carbonate (%)	Biogenic Silica (%)	Lithogenic Silica (%)	delta ^{15}N (per mil)	Liquid Water Content (%)	Pressure (dbar)	Oxygen (µmol/l)	Temperature (oC)	Salinity (psu)
1	0.04	0.17	0.14	0.28	4.67	94.75	6.45	23.62	44	63	10.03	34.79
2	0.14	1.42	0.89	4.39	4.44	90.51	6.11	32.51	117.5	78.5	9.61	34.76
4	0.27	11.25	2.24	75.15	2.38	16.49	6.26	55.4	185	104	8.95	34.68
5	0.22	11.95	1.83	84.33	0.92	10.56	6.08	55.91	173	111	9.4	34.73
7	0.3	11.76	2.48	77.38	1.8	16.77	6.59	59.53	369	121	8.64	34.68
8	0.35	11.01	2.79	68.54	3.55	23.3	6.56	55.94	720	150	4.64	34.41
9	0.1	9.68	0.83	73.72	4.18	25.86	8.18	56.37	2043.5	198	3	34.87
10	0.21	10.42	1.74	72.41	4.65	23.79	6.53	61.52	2107	191	3.14	34.9
14	0.26	10.94	1.81	76.08	2.55	18.65	5.86	54.66	454	77	7.65	34.64
18	0.33	11.59	2.53	75.5	3	15.74	5.66	55.75	351	59	9.21	34.78
20	0.68	13.18	5.43	64.62	2	21.9	6.15	61.08	236	16	12.36	35.16
34	1.02	7.8	7.47	2.75	34.21	51.22	7.1	86.72	40	32	12.76	35.15
36	0.87	13.53	7.83	47.5	4.54	35.41	5.81	64.32	296	14	10.55	34.96
39	0.7	9.69	5.94	31.24	10.31	55.33	4.77	65.39	139	29.5	14.17	35.39
43	0.6	5.07	4.49	4.78	20.7	65.92	4.36	81.51	74	22	15.24	35.54
45	0.06	1.86	0.55	10.97	18.22	70.28	5.71	28.92	242.5	23.5	11.63	35.12
46	0.15	1.39	1.14	2.05	13.86	77.55	4.79	35.93	459	48	7.53	34.66
47	0.23	2.41	1.73	5.67	13.18	76.46	5.16	51.89	1016	135	4.04	34.56
48	0.42	3.83	3.07	6.41	17.63	71.54	5.2	69.56	2119	198	3.29	34.92
53	0.59	12.63	4.68	66.29	4.96	23.68	5.11	61.55	410	44	8.07	34.7
59	0.54	11.65	3.61	66.94	5.5	24.46	4.91	59.41	212	10	12.32	35.18
2281	1.46	10.98	9.55	11.88	24.99	51.01	6.08	90.05	114	9	11.81	35.07
2282	1.69	13.93	11.9	16.94	16.11	57.24	4.9	87.86	142.5	16	11.62	35.05

	Surface Sediment								Bottom Water			
Station	Total Nitrogen (%)	Total Carbon (%)	Total Organic Carbon (%)	Calcium Carbonate (%)	Biogenic Silica (%)	Lithogenic Silica (%)	delta ^{15}N (per mil)	Liquid Water Content (%)	Pressure (dbar)	Oxygen (µmol/l)	Temperature (oC)	Salinity (psu)
2285	1.55	13.07	11.57	12.48	16.15	50.97	4.27	84.38	155	7	11.53	35.03
2287	0.57	7.87	7.67	1.71	5.63	78.87	5.32	52.38	251	34	11.21	35.01
2289	1.04	14.38	8.75	46.97	4.22	32.25	5.92	75.9	694	120.5	5.24	34.46
2291	0.96	13.9	7.82	50.67	6.41	32.09	5.98	75.28	2243	197	2.98	34.89

Table 4.5: The environmental variables best correlated with the microbial community diversity pattern, as provided by the BIO-ENV analysis. Each row represents the correlation of the microbial community pattern with a different set of environmental variables, indicated by the plus sign. ρ_w: weighted Spearman rank correlation coefficient.

	Bottom Water		Surface sediment					
ρ_w	Oxygen (µmol/l)	Salinity (psu)	Total Carbon (%)	Total Organic Carbon (%)	Calcium Carbonate (%)	Biogenic Silica (%)	delta ^{15}N (per mil)	Liquid Water Content (%)
0.687	+						+	+
0.684	+					+	+	+
0.669	+	+				+	+	+
0.661	+	+					+	+
0.657	+			+		+	+	+
0.641	+			+			+	+
0.633	+				+		+	+
0.631	+		+			+	+	+
0.630	+	+		+			+	+
0.627	+				+	+	+	+

4.3.2. Identification of microorganisms incorporating ^{13}C acetate

In order to assess the composition of the microbial community using nitrate or sulfate as electron acceptors, two microcosm experiments were conducted (Table 4.3; Figure S4.1) in three sampling stations (Table 4.2). Station 34 was physicochemically different from the other two stations, by having the lowest percentage of calcium carbonate (2.75 %) and pressure (40 dbar) and the highest percentage of liquid water content (86.72 %). The lowest oxygen concentration was found at stations 39 (29.5 µmol/l) and 34 (32 µmol/l). Also, station 18 had the lowest percentage of biogenic (3 %) and lithogenic silica (15.74 %) (Table 4.4).

After six days of incubations, subsequent RNA extractions and isotopic separation by ultracentrifugation, all the isotopically "heavy" gradient fractions (e.g. with density of ~1.82 g/ml cesium trifluoroacetate) of the unlabelled control contained less RNA than the respective fractions of the ^{13}C gradient (Figures S4.2-S4.4), thus confirming the isotopic enrichment and ensuring the successful completion of the SIP experiment (Neufeld *et al.*, 2007). Sequencing of the selected fractions showed differences between the "heavy" and "light" fractions in the ^{13}C acetate amended incubations (Figures 4.4; S4.5-S4.7); since certain taxonomic groups were found in the "heavy" fractions while they were less abundant or absent in the "light" fractions, the incorporation of the label in the ^{13}C acetate amended incubations was verified.

More specifically, in the SIP experiment at station 18, there was a clear increase in the abundance of *Epsilonproteobacteria* in the ^{13}C "heavy" fraction when nitrate was added as an electron acceptor (~74%) compared to their abundance in the "light" fraction of this sample (~34%) (Figures 4.4; S4.5). In addition, *Betaproteobacteria*, *Bacilli* and *Clostridia* were found in the ^{13}C labelled "heavy" fraction while they were absent from the "light" fraction. When sulfate was added as an electron acceptor, there was again an increase in the abundance of *Epsilonproteobacteria* (~57%) compared to the percentage found in the "light" fraction of this sample (~10%). When no external electron acceptors were added in the incubations, the same

increase in the abundance of *Epsilonproteobacteria* was observed (~76%) while in the "light" fraction their relative abundance was only 12%.

The SIP experiment at station 34 (Figures 4.4; S4.6), under nitrate addition, resulted in an increase in the abundance of *Gammaproteobacteria* (~51%) and the appearance of *Betaproteobacteria* (~10%) and *Clostridia* (~23%) in the ^{13}C labelled "heavy" fraction while the abundance of the former was only 10% in the "light" fraction of this sample. The same pattern was observed when no electron acceptors were added and, also, the control incubations were almost identical across the density range of the centrifugation gradient. However, under the addition of sulfate, no change was observed in the taxonomic composition of RNA in the "heavy" fraction when compared to the "light" fraction.

The SIP experiment at station 39 (Figures 4.4; S4.7), with addition of nitrate as electron donor, yielded an increase in the abundance of *Alphaproteobacteria*, *Betaproteobacteria*, *Flavobacteriia* and *Bacilli* in the "heavy" fraction (~16, 20, 10 and 13%, respectively) when compared to their abundance in the "light" fraction (~6% on average). Under the addition of sulfate, the increase in the abundance of *Epsilonproteobacteria* was evident (~89%) compared to the "light" fraction (~9%). When no electron acceptors were added, the "heavy" fraction had higher abundances of *Epsilonproteobacteria*, *Actinobacteria*, *Sphingobacteriia* and *Bacilli* (~32, 21, 10 and 6% respectively) while these groups had an average abundance of 2% in the ^{13}C "light" fraction.

The aforementioned taxonomic groups were actively assimilating the ^{13}C labelled acetate, which resulted in the differentiation of those samples from the ones that were incubated with the native substrate (natural isotopic composition) (Figures S4.5-S4.7).

When compared with the community composition in situ, prior to the SIP incubations, it is evident that at station 18 the abundance of *Epsilonproteobacteria* increased from <1% to 57 – 76% depending on the electron acceptor used. Samples where no electron donor was added, i.e. samples H, I and G (Table 4.3; Figure S4.1), did not differ much from the community composition *in situ*; thus, there were no background population changes derived from the incubation process.

Regarding station 34, there was an apparent increase in the abundance of *Gammaproteobacteria* from ~7% at the in situ community to ~51% under nitrate addition. However, in the case of sulfate addition (samples B, E, I) and where no electron donor was added (H, I, G) the observed pattern was the same; the latter samples differ from the natural community, which might imply that the observed changes are due to the incubation process.

Samples from station 39, under nitrate addition, differentiated from the in situ community regarding the increase in abundances of *Alphaproteobacteria* and *Betaproteobacteria* (from ~1 and <1% to ~16 and 20 %, respectively). Similarly, under sulfate addition, *Epsilonproteobacteria* showed an increase from <1% at the *in situ* community to ~89% and under no addition of electron acceptors there was again an increase of *Epsilonproteobacteria* (~32%) and *Actinobacteria* (~21% from the 6% of the natural community). As for station 18, samples H, I and G suggest that there were only slight background population changes derived from the incubation process.

When the ^{13}C labelled "heavy" and "light" RNA fractions are compared at the three stations (Figure 4.4), it seems that the patterns across the incubations are different, according to sampling station. Although the initial communities prior to the incubation did not differ much between the sampling stations (Figure 4.2), there are certain differences observed in the microbial communities of each experiment (Figure 4.4). For example, when no electron acceptor was added, an increase in the abundance of *Epsilonproteobacteria* was observed at stations 18 and 39 but of *Gammaproteobacteria* at station 34. In addition, the increase in *Epsilonproteobacteria* was observed both when nitrate or sulfate were added.

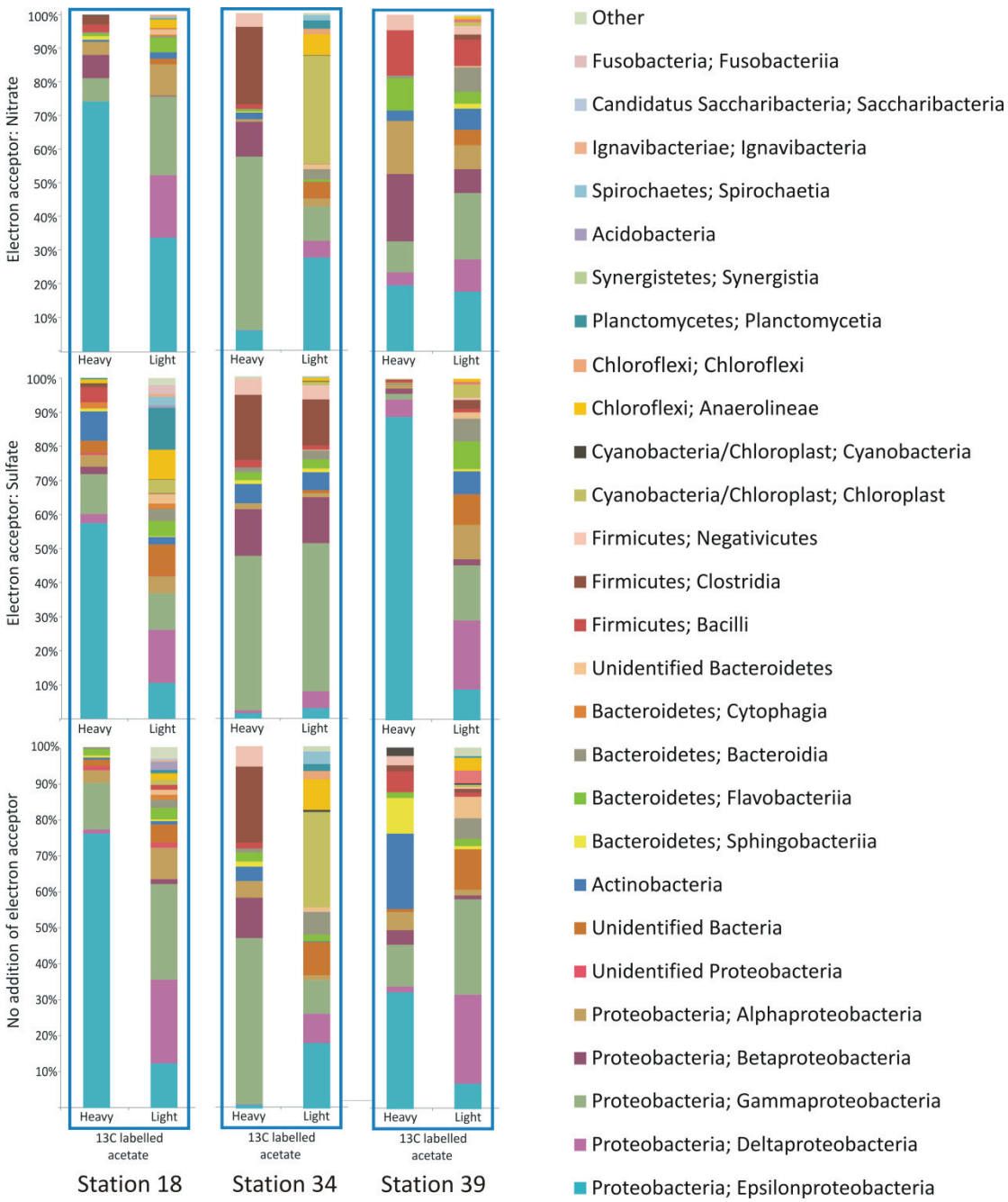

Figure 4.4: The relative abundances (%) of the main microbial taxonomic groups of the "heavy" and "light" fractions of the SIP experiment with addition of ^{13}C labelled acetate at the three stations. Top graph: Incubations with nitrate as an electron acceptor. Middle graph: Incubations with sulfate as an electron acceptor. Bottom graph: Control incubations with no electron acceptor addition.

4.4 Discussion

4.4.1 Microbial community composition *in situ*

Previous studies on the sediment microbial communities of the Benguela upwelling system are very limited (Schippers *et al.*, 2012; Evans *et al.*, 2017). Archaea and Bacteria have been found to exist in equal abundances, while correlating with the sediment organic carbon content (Schippers *et al.*, 2012); however, other studies have shown that Bacteria are more abundant in the surface sediments (Evans *et al.*, 2017). Other studies at the Namibian sediments have focused on the existence and high abundance of large sulfur bacteria, such as *Beggiatoa* spp. and *Thiomargarita namibiensis*, which couple the oxidation of sulfide using nitrate as the terminal electron acceptor (Schulz *et al.*, 1999; Dale *et al.*, 2009), leading to the detoxification of the sulfidic waters (Lavik *et al.*, 2009). However, these bacteria are only present in considerable abundances north of 22°S (Lavik *et al.*, 2009), which excludes four of the six sampled transects of the present study (i.e. transects A to D). *Gamma*- and *Epsilonproteobacteria* have been found to catalyze chemolithotrophic oxidation of sulfide with nitrate (Lavik *et al.*, 2009) although the abundance of the latter was very low in the sampling stations of the present study. *Epsilonproteobacteria* were almost exclusively found at station 34 which was characterized by the highest percentage of biogenic silica and the lowest pressure.

Furthermore, it has been shown that anaerobic ammonium oxidation (anammox), conducted by bacteria belonging to the order of *Planctomycetales* (Wang and Gu, 2013), is a dominant pathway in the Benguela upwelling zone (Kuypers *et al.*, 2005). However, this cannot be suggested from the results of the present study since the abundance of *Planctomycetes*, in general, was very low; thus, anammox does not seem to be an important process due to its scarce representatives in the chosen sampling stations. However, this can be attributed to the choice of primer pair which is not specific for *Planctomycetales*, such as the one used by Chouari *et al.* (2003).

Overall, the majority of the OTUs were classified as *Deltaproteobacteria*, suggesting that sulfate reduction is occurring in the sampling stations, as has been observed by

Brüchert *et al.* (2003); however, this group of microorganisms can use other electron acceptors, such as nitrate and nitrite (Dalsgaard and Bak, 1994; Moura *et al.*, 1997) or other sulfur compounds (thiosulfate, sulfite and sulfur) (Muyzer and Stams, 2008). Microbial communities seem to group according to the range of oxygen concentration and water depth. The relationship of microbial communities with depth is expected since previous studies in the area have shown a decrease in the size of the microbial community with the increase of water depth (Evans *et al.*, 2017). It has also been shown that organic carbon concentration decreases with depth (Inthorn *et al.*, 2006), coinciding with a decrease in sulfate reduction rate (Brüchert *et al.*, 2003). Therefore, this may affect the abundance of different microbial populations in the different sampling stations. However, from the results of the present study, percentages of total organic carbon and total carbon did not correlate with water depth (data not shown).

Our results showed an evident correlation of the microbial community pattern with the physicochemical parameters, and in particularly with delta ^{15}N, sediment water content and oxygen concentration. High values of delta ^{15}N (8 to 15‰) have been found to occur in the diatomaceous mud and to coincide with the OMZ (Nagel *et al.*, 2016); they have been interpreted as a signal of nitrogen consuming anaerobic processes, such as heterotrophic denitrification and anammox, which enrich the residual nitrate in ^{15}N (Pichevin *et al.*, 2005; Emeis *et al.*, 2009). However, there are still patterns of ^{15}N that cannot be explained as such (Nagel *et al.*, 2016).

Different sediment types have shown to be characterized by different delta ^{15}N values and also, delta ^{15}N has shown to be higher in stations with greater depth (Nagel *et al.*, 2016). This would suggest that the observed correlation could reflect an effect of sediment type on microbial composition, apart from dominance of certain nitrogen transformation processes.

4.4.2 Metabolically active microbial populations

SIP was employed to link function and identity of the microorganisms of the sediment samples from the Namibian oxygen minimum zone. SIP involves the

incorporation of "heavy" isotopes into newly synthesized nucleic acids, and can be used to separate newly synthesized from existing DNA or rRNA (Friedrich, 2006; Rettedal and Brözel, 2015). One important issue in SIP studies is the avoidance of cross feeding, i.e. consumption of dead labelled microbial biomass (Gallagher *et al.*, 2005), which results in labelling of microorganisms that did not feed on the provided substrate, i.e. non-acetotrophs in the case of the present study. However, the short duration of the microcosm incubations, and the use of RNA as molecular marker, were considered to have minimized, if not completely avoided, any potential cross feeding. Another important methodological consideration is that SIP inevitably introduces a level of disturbance from the natural conditions, which could favour specific microbial groups (Vandieken and Thamdrup, 2013) In order for safer conclusions to be drawn, control samples were also included in the experiments so that background population changes during the SIP incubation would be detectable.

In all cases of external additions of electron acceptors at station 18, as well as in the case of no addition, an increase of *Epsilonproteobacteria* was observed; the latter is in accordance with previous studies showing that *Epsilonproteobacteria* dominated the anaerobic assimilation of acetate in estuarine sediment SIP incubations (Webster *et al.*, 2010). Since the same pattern was observed when sulfate was added to the incubations, it can be suggested that nitrate, an electron acceptor with higher standard redox potential, was already available at station 18; thus, it could have been used for the oxidation of acetate to a lower threshold concentration than the one that would have been achieved under reduction of sulfate (Achtnich *et al.*, 1995). However, the low concentrations of nitrate in the sediment pore water (Lahajnar *et al.*, 2015) could not have allowed for a complete depletion of the electron donor; hence, acetate would still have been available for sulfate reduction. Nitrate is a favourable electron acceptor in terms of thermodynamics that can stimulate nitrate-reducing, sulfide-oxidizing bacteria (Bentzen *et al.*, 1995), while inhibiting the growth of sulfate-reducing bacteria (García de Lomas *et al.*, 2006). This inhibition can be relieved with the addition of a suitable exogenous electron donor (Achtnich *et al.*, 1995). Nitrate addition has been shown to alter microbial community composition and particularly, to increase the abundance of Beta-, *Gamma*- and *Epsilonproteobacteria* (Chen *et al.*, 2013), which is in accordance with

the results of the present study. Increase of *Beta*- and *Gammaproteobacteria* was also observed under nitrate addition at station 34, while at station 39 an increase in *Alpha*- and *Betaproteobacteria* was observed, which is expected since several genera of denitrifiers are included in these classes (Zumft, 1997; Canfield *et al.*, 2005; Hunter *et al.*, 2006). Also, nitrate surplus has been shown to stimulate large sulfur bacteria, such as the gammaproteobacterium *Thiomargarita namibiensis*, which store nitrate and polyphosphate (poly-P) in their vacuoles (Goldhammer *et al.*, 2010). As observed for station 18, station 39 under sulfate addition showed an increase in *Epsilonproteobacteria*. Sulfate addition did not enhance known sulfate-reducers, such as *Deltaproteobacteria* and *Firmicutes* (Muyzer and Stams, 2008; Müller *et al.*, 2015), although they exist as members of the *in situ* community representing the largest fraction of sequences. Apart from the competition for electron donors among nitrate-reducers and sulfate-reducers (Achtnich *et al.*, 1995), this could be attributed to the fact that certain sulfate-reducing bacteria use lactic, pyruvate, ethanol and some fatty acids as electron donors, instead of acetate (Madigan *et al.*, 2003; Rabus *et al.*, 2006). Hence, increase in the abundance of such genera would not have been favoured by the used electron donor in this experiment. Also, sulfate reduction can be stimulated by alternative unidentified electron donors other than acetate, lactate, propionate and isobutyrate (Finke *et al.*, 2007). In addition, incorporation of labelled acetate into known sulfate-reducers was not evident in similar SIP experiments with comparable duration, suggesting that sulfate reduction in the incubated sediments could be occurring between novel sulfate reducers and sulfur and/or sulfide-oxidizing *Epsilonproteobacteria* (Webster *et al.*, 2010). SIP experiments of longer duration or using deeper sediment layers could have possibly led to enrichment of different microbial communities, and in particular sulfate-reducers, although the highest sulfate reduction rates have been observed in the upper sediment layer (Brüchert *et al.*, 2009). In addition, determining sulfate reduction rates in the presence and absence of acetate could have provided evidence for the existence of sulfate reducers and of their preferred electron donor in these sampling stations.

4.5 Conclusions

It has been suggested that modifications in hydrology and global warming will enhance hypoxic incidents and will expand OMZs (Stramma *et al.*, 2008), thus affecting survival and functioning of macro- and microorganisms (Middelburg and Levin, 2009). Chemolithoautotrophic microbes are expected to be affected by the expansion of hypoxic conditions (Lipsewers *et al.*, 2017), since, in general, anaerobic reactions and pathways will be favoured, at the expense of the aerobic ones. Currently, sulfide-oxidizing bacteria have been shown to control the flux of hydrogen sulfide to the water column (Schulz *et al.*, 1999); however, if sulfate reduction is enhanced, thus leading to increased hydrogen sulfide production, there is a chance that the latter will be largely diffused in the water column (Emeis *et al.*, 2004) and not retained in the sediment in oxidized forms, i.e. as elemental sulfur or sulfate.

Regulation of hypoxic episodes will aid the end-users of coastal upwelling ecosystems to possibly achieve higher productivity, in terms of fish catches, which otherwise is largely compromised by the elevated hydrogen sulfide concentrations. In order to formulate possible predictions on the future state of OMZs, more information is needed in regard to biogeochemical cycles and microbial community functioning, and especially concerning the coupling reactions between nitrogen and sulfur.

Chapter 5: Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies[9]

Citation: Oulas A, Pavloudi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G, Arvanitidis C, Iliopoulos I (2015) Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. *Bioinformatics and Biology Insights,* 9: 75–88, doi: 10.4137/BBI.S12462

Abstract

Advances in next-generation sequencing (NGS) have allowed significant breakthroughs in microbial ecology studies. This has led to the rapid expansion of research in the field and the establishment of "metagenomics", often defined as the analysis of DNA from microbial communities in environmental samples without prior need for culturing. Many metagenomics statistical/computational tools and databases have been developed in order to allow the exploitation of the huge influx of data. In this review article, we provide an overview of the sequencing technologies and how they are uniquely suited to various types of metagenomic studies. We focus on the currently available bioinformatics techniques, tools, and methodologies for performing each individual step of a typical metagenomic dataset analysis. We also provide future trends in the field with respect to tools and technologies currently under development. Moreover, we discuss data management, distribution, and integration tools that are capable of performing comparative metagenomic analyses of multiple datasets using well-established databases, as well as commonly used annotation standards.

[9] For author contributions, please refer to the relevant section. Modified version of the published review.

5.1 Introduction

As has been seen from the previous chapters, there is a variety in the way of sequencing data analysis; different methods and software have been applied in each chapter. In this chapter, the main workflows for the analysis of metagenomic datasets will be presented, along with examples of software and pipelines. In addition, the terminology used in such studies will be discussed in detail in order to avoid confusion and misunderstandings between similar terms or between terms that have been used synonymously without justification.

The advent of next-generation sequencing (NGS) or high-throughput sequencing has revolutionized the field of microbial ecology and brought classical environmental studies to another level. This type of cutting-edge technology has led to the establishment of the field of "metagenomics", defined as the direct genetic analysis of genomes contained within an environmental sample without the prior need for cultivating clonal cultures. Initially, the term was only used for functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample (Riesenfeld *et al.*, 2004), but currently it is also widely applied to studies performing polymerase chain reaction (PCR) amplification of certain genes of interest. The former can be referred to as "full shotgun metagenomics" (Xia *et al.*, 2011), and the latter as "marker gene amplification metagenomics" (i.e., 16S rRNA gene), "targeted metagenomics" or "meta-genetics" (Handelsman, 2009). As the term "metagenomics" has been extensively used for both the aforementioned methods, despite their intrinsic differences, both "shotgun metagenomics" and "marker gene metagenomics" will be analyzed in detail in this chapter.

Such methodologies allow a much faster and elaborative genomic/genetic profile generation of an environmental sample at a very acceptable cost. Full shotgun metagenomics has the capacity to fully sequence the majority of available genomes within an environmental sample (or community). This creates a community biodiversity profile that can be further associated with functional composition analysis of known and unknown organism lineages (i.e., genera or taxa) (Tringe *et al.*, 2005). Shotgun metagenomics has evolved to address the questions of *who* is

present in an environmental community, *what* they are doing (function-wise), and *how* these microorganisms interact to sustain a balanced ecological niche. It further provides unlimited access to functional gene composition information derived from microbial communities inhabiting natural ecosystems.

Marker gene metagenomics is a fast and gritty way to obtain a community/taxonomic distribution profile or fingerprint using PCR amplification and sequencing of evolutionarily conserved marker genes, such as the 16S rRNA gene (Tringe and Hugenholtz, 2008). This taxonomic distribution can subsequently be associated with environmental data (metadata) derived from the sampling site under investigation.

Several types of ecosystems have been studied so far using metagenomics, including extreme environments such as areas of volcanism (Benson *et al.*, 2011; Xie *et al.*, 2011; Kilias *et al.*, 2013; Urich *et al.*, 2014) or other areas of extreme temperature (Bradford *et al.*, 2008; Pearce *et al.*, 2012), alkalinity (Xiong *et al.*, 2012), acidity (García-Moyano *et al.*, 2012; Johnson, 2012), low oxygen (Stevens and Ulloa, 2008; Bryant *et al.*, 2012) and high heavy-metal composition (Chodak *et al.*, 2013; Gołębiewski *et al.*, 2014). This invaluable resource provides an infinite capacity for bioprospecting and allows the discovery of novel enzymes capable of catalyzing reactions of biotechnological commercialization (Segata *et al.*, 2011).

The first metagenomic studies were focused on environments that were believed to be characterized by low-diversity, such as an acid mine drainage (Tyson *et al.*, 2004), human gut microbiome (Breitbart *et al.*, 2003), and water samples from the Sargasso Sea (Venter *et al.*, 2004), mainly due to the unavailability of both high-throughput sequencing technologies at that time and relevant software for the scaffolds' assembly. As more and more researchers entered this new field of study, the need for powerful tools and software became apparent and therefore led to the creation of several such tools.

5.2 Sequencing Technologies

Two commonly used NGS technologies utilized to date are the 454 Life Sciences and the Illumina systems, with the ratio of usage shifting in favor of the latter recently. Both technologies have been widely used in metagenomic studies, and hence it is important to briefly describe their advantages and disadvantages with respect to the sequencing of metagenomics samples.

The 454 pyrosequencer was the first next-generation sequencer to achieve commercial introduction in 2004 (Mardis, 2008). Its chemistry relies on the immobilization of DNA fragments on DNA-capture beads in a water–oil emulsion and then using PCR to amplify the fixed fragments. The beads are placed on a PicoTiterPlate (a fiber-optic chip). DNA polymerase is also packed in the plate, and pyrosequencing is performed (Ronaghi *et al.*, 1998; Ronaghi, 2001). Its main difference from the classic Sanger sequencing is that pyrosequencing relies on the detection of pyrophosphate release on nucleotide incorporation rather than chain termination with dideoxynucleotides. The release of pyrophosphate is conveyed into light using enzyme reactions, which is then converted into actual sequence information (Mardis, 2008).

In the initial years of high-throughput sequencing, scientists embraced the new technology and hence discovered the existence of the "rare biosphere" (Sogin *et al.*, 2006), although the actual existence of the latter has been highly disputed ever since as differentiation between the "rare biosphere" and sequencing errors, i.e. artifacts, cannot be done easily (e.g. Kunin *et al.,* 2010). However, in many cases the apparent assignment of a microbial operational taxonomic unit (OTU) was in fact an attribute of sequencing errors, which caused an overinflation of the diversity estimates (Brown *et al.*, 2015). Noise generated by this 454 pyrosequencing technology affected different aspects of metagenomic data analysis and led to biased results (Rosen *et al.*, 2012).

PCR errors may lead to replicate sequence artifacts, which can cause overestimation of species abundance and functional gene abundance in 16S rRNA and full shotgun metagenomics, respectively. PCR can also generate noise in the form of single base

pair errors (i.e., substitutions, deletions) that can cause frame shifts for protein coding genes in shotgun metagenomics. Moreover, PCR chimeras (sequences generated by undesired end-joining of two or more true sequences) can also affect 16S rRNA metagenomics results with respect to species distribution (Brodin *et al.*, 2013). Sequencing errors can also occur due to the actual chemistry underlining the technology. For example, there is an inherent difficulty in clearly identifying the intensity of 454 pyrosequencing-generated flowgrams. This task becomes even more difficult during the sequencing of homopolymers (Rothberg and Leamon, 2008). The relatively long read length generated by this technology (in comparison to other sequencing technologies) allows a significantly less error-prone assembly in shotgun metagenomics and permits greater annotation accuracy (Wommack *et al.*, 2008; Thomas *et al.*, 2012).

Although 454 has seized being supported by Life Sciences, still one should take into account that there is a large number of existing unpublished datasets that have been generated via this technology. Therefore, it is important to include it in this review and compare it with the other sequencing services that have become more popular over the last years, namely Illumina.

Illumina dye sequencing by synthesis begins with the attachment of DNA molecules to primers on a slide, followed by amplification of that DNA to produce local colonies (Mardis, 2008). This generation of “DNA clusters” is accompanied by the addition of fluorescently labeled, reversible terminator bases (adenine, cytosine, guanine, and thymine) attached with a blocking group (Bentley *et al.*, 2008). The four bases then compete for binding sites on the template DNA to be sequenced, and the nonincorporated molecules are washed away. After each synthesis cycle, a laser is used to excite the dyes, and a high-resolution scan of the incorporated base is made. A chemical deblocking step ensures the removal of the 3’ terminal blocking group and the dye in a single step. The process is repeated until the full DNA molecule is sequenced. Illumina has a variety of sequencing instruments dedicated to different applications. MiSeq, for example, has an output of 15 GB and 25 million sequencing reads of 300 bp in length; clustered fragments can be sequenced from both ends (paired-end sequencing), which can be merged so that 600 bp reads can be obtained. HiSeq2500 has a much greater output (1,000 GB per run) but offers 125 bp reads.

The shorter read length produced by Illumina may increase errors during assembly and, subsequently, the annotation inaccuracies during shotgun metagenomics data analysis (Kircher *et al.*, 2012). In contrast, when analyzing 16S rRNA metagenomics data, this technology obviates the need for time-consuming noise removal algorithms required for pyrosequencing and makes analysis less error-prone (Werner *et al.*, 2011). The greater coverage/yield generally offered by Illumina allows significant decrease of systematic errors. This advantage and the low cost are the delineating factors that have turned Illumina into the preferred high-throughput sequencing technology for metagenomics studies.

Additional sequencing technologies are available and can potentially be used for metagenomic studies. These include the Applied Biosystems SOLiD 5500 W Series sequencer, which offers higher coverage than 454 pyrosequencing but lower than Illumina (~120 GB per run). It allows fragment or mate-paired sequencing; however, it can only guarantee a low error rate for sequencing reads of maximum 50 bp in length (Metzker, 2010). This reduces the possibility of generating a reliable and usable de novo assembly for shotgun metagenomics; but, on the other hand, this technology performs very well when utilizing a reference genome for mapping or assembly of reads. However, using the Exact Call Chemistry (ECC) module, the SOLiD system offers to boost the accuracy of its ligation-based sequencing.

An emerging sequencing technology that may have high impact on the fields of genomics and metagenomics was recently developed by Pacific Biosciences (PacBio) (Metzker, 2010). This technology uses single-molecule real-time (SMRT) sequencing, which is a parallelized single-molecule DNA sequencing by synthesis. SMRT sequencing utilizes the zero-mode waveguide (ZMW), whereby a single DNA polymerase enzyme is fixed to the bottom of a ZMW with a single molecule of DNA as a template. The ZMW is a structure that creates an illuminated observation volume that is small enough to allow the observation of a single nucleotide of DNA (also known as a base) being incorporated by DNA polymerase. Each of the four DNA bases is attached to one of four different fluorescent dyes. When a nucleotide is incorporated by the DNA polymerase, the fluorescent tag is cleaved off, which diffuses out of the observation area of the ZMW where its fluorescence is no longer observable. A detector detects the fluorescent signal of the nucleotide

incorporation, and the base call is made according to the corresponding fluorescence of the dye. PacBio provides much longer read lengths (~10,000 bp) compared to the aforementioned technologies, thus having obvious advantages when addressing issues of annotation and assembly for shotgun metagenomics. PacBio technology uses a process called strobing to perform paired-end read sequencing. Despite the high read length of PacBio, this technology is limited by high error rates and low coverage (albeit at higher throughput than Sanger sequencing).

In addition to the aforementioned technologies, which are based on optics, technologies such as Ion Torrent's semiconductor sequencing benchtop sequencer and Ion Proton are now coming into play. These technologies are based on the use of proton emission during polymerization of DNA in order to detect nucleotide incorporation. This system promises read lengths of >200 bp and relatively high throughput, on the order of magnitude achieved by 454 Life Sciences systems. Additionally, it offers higher quality than 454, especially when sequencing homopolymers, but at a similar cost. Looking into the future, and given that 454 will eventually stop being supported by Life Sciences, it is very likely that former users of the 454 pyrosequencing will switch to Ion Torrent sequencing chemistry, due to the similarities of both (eg, emulsion PCR step) and the significant advantages of the latter.

An even more cutting-edge technology is currently under development by Oxford Nanopore technologies, which is developing "strand sequencing", a method of DNA analysis that could potentially sequence completely intact DNA strands/polymers passed through a protein nanopore. This obviates the need for shotgun sequencing and aims to revolutionize the sequencing industry in the future. Oxford Nanopore intends to commercialize this technology with the Company's GridION™ and MinION™ systems. For metagenomics, this technology can have obvious advantages, as it will eliminate erroneous sequencing caused by shotgun metagenomics and exclude the need for the error-prone assembly step during data analysis (for details, see later). However, nanopore sequencing is at the moment noncommercialized (offered only through the MinION™ Access Program) and is still being optimized on case-by-case basis using specific template and sequencing needs.

Another example of an innovative and very promising technology is the Irys Technology (BioNano Genomics), which uses micro and nanostructures and offers new ways of *de novo* constructing genome maps. The input is DNA labeled at specific sequence motifs that can be used for imaging and identification in IrysChips. These labeling steps result in a uniquely identifiable, sequence-specific pattern of labels to be used for de novo map assembly or for anchoring sequencing contigs.
The main characteristics of the different next generation sequencing technologies are summarized in Table 5.1.

Table 5.1: Comparison of the main different next generation sequencing technologies.

Sequencing Type	Technology	Chemistry	Read Length (bp)	Throughput (Gb)	Runtime (h)
Massively Parallel Sequencing	454	Pyrosequencing	500-1200	1	23
	Illumina	Sequencing By Synthesis	100-600	600-1000	72-192
	SOLiD	Sequencing By Ligation	60	180	336
	Ion Torrent	Proton detection	200-400	2	2-8
Single Molecule Sequencing	PacBio	Real Time Sequencing	50,000	1	2
	Nanopore	Strand sequencing	1,000	10-20	>6

5.3 Shotgun Metagenomics

5.3.1 Assembly of shotgun metagenomics data

Metagenomics studies are commonly applied to investigate the specific genomes (known as well as unknown, both cultured and uncultured) that are present within an environmental community under study. Moreover, when performing full shotgun metagenomics, the complete sequences of protein coding genes (previously characterized or novel) as well as full operons in the sequenced genomes can offer invaluable knowledge on the functional potential of a microbial community. For these reasons, an assembly of shorter reads into genomic contigs and orientation of these into scaffolds is often performed to provide a more compact and concise view

of the sequenced community under investigation. Early attempts at metagenomic data assemblies utilized tools initially implemented for single genome data assemblies. They, therefore, fell short when forced to assemble reads into contigs for metagenomic samples. However, assembly tools have significantly evolved since then, and the current line of tools have been modified and specifically designed to assemble samples containing multiple genomes, thereby rendering them much more affective for the task in hand.

The process of assembling shorter reads into contigs can take two different routes: i) reference-based assembly and ii) *de novo* assembly. The choice of which route to follow depends on the dataset that needs to be analyzed and on the specific needs of each research project. For example, *de novo* assembly could be, in theory, used even if a reference genome exists, if the computational power allows for it.

Reference-based assembly refers to the use of one or more reference genomes as a "map" in order to create contigs, which can represent genomes or parts of genomes belonging to a specific species or genus. Tools such as Newbler (Roche), MIRA 4 (Chevreux *et al.*, 2004) or AMOS, as well as the recent MetaAMOS (Treangen *et al.*, 2013) are commonly used in metagenomics for performing referenced-based assemblies. These tools are not computationally intensive and perform well when metagenomic samples are derived from extensively studied and researched areas. In such cases, sequences from closely related organism would have already been deposited in online data repositories and databases, allowing them to be used as references for the assembly process. Often, assemblies are visually evaluated using genome browser tools such as Artemis (Rutherford *et al.*, 2000). The observation of large gaps in the query genome(s) of the resulting assembly, when comparing to the reference genome(s), can be seen as an indication that perhaps the assembly is incomplete or that the reference genome(s) used are too distantly related to the community under investigation in order to perform optimally.

De novo assembly refers to the generation of assembled contigs using no prior reference to known genome(s) (Paszkiewicz and Studholme, 2010). This task is computationally expensive and relies heavily on sophisticated graph theory algorithms, such as de-Bruijn graphs, which were specifically employed to tackle this job. Tools such as EULER (Pevzner *et al.*, 2001), Velvet (Zerbino and Birney, 2008),

SOAP (Li *et al.*, 2008) and Abyss (Simpson *et al.*, 2009) were amongst the first to perform de novo assembly and are still widely used today. They require computers with large amounts of memory and generally long execution times (depending on the size of the dataset). However, these tools were built with the assumption of assembling a single genome and often underperform when used for metagenome assemblies. Problems arise from 1) variation between similar subspecies, 2) genomic sequence similarity between different species, and 3) difference in abundance for species in a sample also affected by different sequencing depths for individual species. These issues introduce kinks (or branches) in the de Bruijn graph, and have to be addressed in order to improve the assembly.

The next generation of assembly tools, such as MetaVelvet (Namiki *et al.*, 2012) and very recently MetaVelvet-SL (Afiahayati *et al.*, 2014) and Meta-IDBA (Peng *et al.*, 2011) were developed to address these issues. MetaVelvet and Meta-IDBA employ a combined binning and assembly approach to create more accurate assemblies from datasets containing a mixture of multiple genomes. They make use of k-mer frequencies to detect kinks in the de-Bruijn graph and then use these k-mer thresholds to decompose the graph into subgraphs. These tools further assemble contigs and scaffolds based on the decomposed subgraphs, and thus perform a more efficient grouping/assembly of contigs, effectively separating those belonging to different species.

The IDBA-UD algorithm (Peng *et al.*, 2012) was recently developed to additionally address the issue of metagenomic sequencing technologies with uneven sequencing depths. It makes use of multiple depth-relative k-mer thresholds to remove erroneous k-mers in both low-depth and high-depth regions. Comparison of the performances of these tools is often performed using the N50 length score, which is defined as “the length for which the collection of all contigs of that length, or longer, contains at least half of the total of the lengths of the contigs in the assembly” (Miller *et al.*, 2010; Earl *et al.*, 2011). A recent comparison of the latest line of assembly tools shows that IDBA-UD can reconstruct longer contigs with higher accuracy (Peng *et al.*, 2012). However, there is still much room for the improvement of metagenomic assembly algorithms in order for them to conceptually capture the task in hand.

5.3.1.1 Binning tools for metagenomes

Binning is the process of grouping (binning) reads or contigs into individual genomes and assigning the groups to specific species, subspecies, or genus. Binning methods can be characterized in two different ways depending on the information used to group the sequences in hand: 1) Composition-based binning is based on the observation that individual genomes have a unique distribution of k-mer sequences (also denoted as genomic signatures). By making use of this conserved species-specific nucleotide composition, these methods are capable of grouping sequences into their respective genomes. 2) Similarity- or homology-based binning refers to the process of using alignment algorithms such as BLAST or profile hidden Markov Models (pHMMs) to obtain similarity information about specific sequences/genes from publically available databases (eg, NCBI's nonredundant database – nr or PFAM). Thereafter, sequences are binned according to their assigned taxonomic information.

Available composition-based binning algorithms are included in tools such as TETRA (Teeling *et al.*, 2004), S-GSOM (Chan, Hsu, Halgamuge, *et al.*, 2008; Chan, Hsu, Tang, *et al.*, 2008), Phylopythia (McHardy *et al.*, 2007) and its successor PhylopythiaS (Patil *et al.*, 2012), TACOA (Diaz *et al.*, 2009), PCAHIER (Zheng and Wu, 2010), ESOM (Ultsch and Moerchen, 2005; Dick *et al.*, 2009) and ClaMS (Pati *et al.*, 2011), while examples of purely similarity-based binning software include tools such as CARMA (Krause *et al.*, 2008), MetaPhyler (Liu *et al.*, 2011) and SOrt-ITEMS (Monzoorul Haque *et al.*, 2009). Some tools employ similarity-based binning algorithms in their metagenomics analysis pipelines. Examples of such tools are IMG/MER 4 (V M Markowitz *et al.*, 2014), MG-RAST (Meyer *et al.*, 2008; Glass *et al.*, 2010) and MEGAN (Huson *et al.*, 2007; Huson and Mitra, 2012; Huson and Weber, 2013) and will be described in more detail below in section 5.3.1.2.

Certain binning tools employ a hybrid approach using both composition and similarity-based information to group sequences. Some examples of such tools are PhymmBL (Brady and Salzberg, 2009) and MetaCluster (Y Wang *et al.*, 2012; Wang *et al.*, 2014). More innovative binning approaches include co-abundance gene

segregation across a series of metagenomic samples, thus facilitating the assembly of microbial genomes without the need for reference sequences (Nielsen *et al.*, 2014). This new method promises to overcome the usual computational challenges of other binning tools and has been tested for a human gut microbiome.

5.3.1.2 Annotation of metagenomics sequences

Annotation of metagenomes is specifically designed to work with mixtures of genomes and contigs of varying length. Initially, a series of preprocessing steps prepare the reads for annotation. These include: i) Trimming of low-quality reads using platform-specific tools such as the FASTX-Toolkit (Su *et al.*, 2012). Additionally, FastQC (Huson *et al.*, 2007) can provide summary statistics for FASTQ files. Both have been recently integrated into the Galaxy platform (Giardine *et al.*, 2005; Blankenberg *et al.*, 2010; Goecks *et al.*, 2010). SolexaQA (Cox *et al.*, 2010) and Lucy 2 (Li and Chou, 2004) are also used for FASTQ files. Most of these tools make use of Phred or Q quality scores (Ewing and Green, 1998; Ewing *et al.*, 1998), the thresholds of which depend on sequencing technology; ii) Masking of low-complexity reads performed using tools such as DUST (Morgulis *et al.*, 2006); iii) A de-replication step that removes sequences that are more than 95% identical; iv) A screening step performed by some tools (i.e., MG-RAST) in which the pipeline provides the option of removing reads that are near-exact matches to the genomes of a handful of model organisms, including fly, mouse, cow, and human. This is done using mapping tools such as Bowtie 2 (Langmead and Salzberg, 2012).

The next main stage of the annotation pipeline is the identification of genes within the reads/assembled contig, a process often denoted as “gene calling” (V M Markowitz *et al.*, 2014). Genes are labeled as coding DNA sequences (CDSs) and noncoding RNA genes, and certain annotation pipelines (e.g., IMG/MER) also predict for regulatory elements such as clustered regularly interspaced short palindromic repeats (CRISPRs).

CDSs are identified using a number of tools including MetaGeneMark (Zhu *et al.*, 2010), Metagene (Noguchi *et al.*, 2006), Prodigal (Hyatt *et al.*, 2010), Orphelia (Hoff

et al., 2009) and FragGeneScan (Rho *et al.*, 2010), all of which utilize *ab initio* gene prediction algorithms. Often, annotation pipelines use an intersection of these tools to obtain a more informative prediction of the protein coding genes. Gene prediction tools utilize codon information (i.e., start codon – AUG) to identify potential open reading frames and hence label sequences as coding or noncoding. Most tools can be trained by using the desired training sets. For example, FragGeneScan is trained for prokaryotic genomes only, and is used by IMG/MER and MG RAST as well as EBI Metagenomics. It is believed to be one of the most accurate gene-prediction tools currently available. However, like most of these tools, it is expected to have an average prediction accuracy of ∼65%–70%, resulting in multiple genes that are missed altogether (Rho *et al.*, 2010).

Noncoding RNAs such as tRNAs are predicted using programs like tRNAscan (Lowe and Eddy, 1997; Schattner *et al.*, 2005), ribosomal RNA (rRNA) genes (5S, 16S, and 23S) are predicted using internally developed rRNA models for IMG/MER, and MG-RAST uses similarity to compare three known databases (SILVA (Quast *et al.*, 2013), Greengenes (DeSantis *et al.*, 2006) and the Ribosomal Database Project-RDP (Maidak *et al.*, 1996; Cole *et al.*, 2007)) to predict rRNA genes.

The next stage of the annotation pipeline involves functional assignment to the predicted protein coding genes. This is currently achieved by homology-based searches of query sequences against databases containing known functional and/or taxonomic information. Due to the large size of metagenomic datasets, this stage is often very expensive computationally and highly automated. BLAST or other sequence-similarity-based algorithms (Edgar, 2010) often run on high-performance computer clusters. Often, multithreading or other parallel programming approaches are used to divide jobs in multiple central/graphic processing units (CPUs/GPUs). This reduces the running time complexity and significantly speeds up querying execution time.

Some widely used data repositories to obtain annotation for metagenomic datasets include functional annotation databases such as KEGG (Ogata *et al.*, 1999; Du *et al.*, 2014), SEED (Overbeek *et al.*, 2005), eggNOG (Powell *et al.*, 2014), COG/KOG (Tatusov *et al.*, 2000), as well as protein domain databases such as PFAM (Bateman *et al.*, 2000; Finn *et al.*, 2014) and TIGRFAM (Haft *et al.*, 2003). Often, annotation

pipelines make use of multiple databases or composite protein domain databases such as Interpro (Hunter *et al.*, 2009) in order to obtain a more collective, cumulative biological functional annotation.

MG-RAST predicts all genes in the metagenome, and then identifies the best homologs of those genes in the isolate genomes using a tool called BLAT (BLAST-like alignment tool) (Kent, 2002). BLAT misses similarities below 70% identity, so many strong hits to other genes are missed. After the best hits to genes from an isolated genome are identified, all subsequent analysis is done using the genes of the isolate genomes, not the genes of the metagenome at hand. This creates a lot of limitations due to the fact that the analysis is not performed on the original genes of the metagenome but on the "proxy" genes to the isolated genomes instead. The advantage of this method is its speed; the only computationally intensive step is to find the best hits of the metagenomes against the isolates. Once this is done, all other comparisons are already pre-existing. The other major advantage is that the MG-RAST database does not grow in size, as is the case with the IMG/MER database.

IMG/MER also begins with prediction of all genes from the metagenome, but then runs all the computations on those genes rather than on their proxies. This allows the identification of PFAM hits (which is not supported in MG-RAST) and provides much more detailed functional information compared to COGS, which is the only protein families database used in MG-RAST. The major bottleneck for IMG/MER is the exponential growth of the gene number, which is not an issue for MG-RAST since the metagenome genes are not kept for analysis. It is, however, important to use PFAM for functional analysis because by comparing the number of genes from any metagenome that go into COG or PFAM clusters, the second provides significantly higher coverage and therefore allows a much deeper analysis. Another major advantage of IMG/MER is that, since the tool keeps the original metagenome genes, it also keeps the original contigs, which provides synteny information. Therefore, it is far more suitable if one is interested in identifying novel biosynthetic gene clusters (BGCs) in the metagenomes, a type of analysis that may be less viable using MG-RAST. The prediction of BGCs from metagenomics data is recently gaining a great deal of interest due to their potential in biotechnological applications. The possibility to engineer BGCs for the production of secondary metabolites with improved

properties, known for their use in anticancer drugs and antibiotics, offers limitless potential for bioprospecting.

The EBI Metagenomics service (Hunter *et al.*, 2014) is a newly developed web-based portal that uses metadata structures and formats that comply with the Genomic Standards Consortium (GSC) guidelines. Moreover, a novel data scheme currently being hosted by the EBI-EMBL is being adopted by the EBI Metagenomics service. This is known as the European Nucleotide Archive (ENA) (Leinonen *et al.*, 2011) data schema and aims to integrate data derived from sequencing technologies under a consensus, mutually accepted standard. EBI Metagenomics offers a dual shotgun and marker gene analysis service. It allows the extraction of rRNA data from shotgun metagenomic data using tools such as rRNASelector (Lee *et al.*, 2011) for concurrent marker metagenomic analysis. It therefore supports additional 16S rRNA-based analysis tools such as Qiime (Caporaso *et al.*, 2010) (see section on Marker Gene Metagenomics) for the efficient taxonomic assignment of these sequences. For functional analysis and annotation of CDS sequences, EBI Metagenomics uses FragGeneScan to obtain protein coding sequences and thereafter utilizes databases such as Interpro, which is a composite, cumulative system comprised of multiple databases of protein families, and allows for protein domain prediction and functional assignment. EBI Metagenomics provides data archiving via ENA and provides unique accession numbers for submitted datasets. Archiving policies require the data to be made public; however, there is a 2-year period (upon submission) during which the data is kept private pending user publication of analysis results.

CAMERA (Seshadri *et al.*, 2007) is another online cloud computing service that provides hosted software tools and a high-performance computing infrastructure for the analysis of metagenomic data. One advantage of CAMERA is that it allows greater user intervention and flexibility during the analysis process. However, this means that users must have expertise, knowledge, and hands-on experience in metagenomic date analysis per se, in order to ensure correct execution of the pipeline and accuracy of results. Moreover, in order to perform comparative metagenomics using CAMERA, the datasets in hand must be traversed through the CAMERA pipeline, thus making integration of data from different resources more

computationally demanding. MEGAN 5 (Huson *et al.*, 2007) is yet another tool that performs analysis of metagenomic data and offers a wide range of visualization tools for metagenomic annotation results. It supports multiple visualization schemes including functional or taxonomic dendrograms, tag clouds, bar charts, and Krona taxonomic plots (Ondov *et al.*, 2011) that allow hierarchical data to be explored in the form of a zoomable pie chart.

5.4 Marker Gene Metagenomics

It is widely accepted that sequencing of the 16S rRNA gene reflects eubacterial and archaeal evolution (Woese, 1987). Since the introduction of SSU rDNA-based molecular techniques (Amann *et al.*, 1995; Muyzer, 1999; Rusch *et al.*, 2007), the study of microbial diversity in natural environments has advanced significantly. In addition, pyrosequencing of the 16S rRNA gene has been widely applied in the field of microbial ecology (Luna *et al.*, 2007; Jones *et al.*, 2009; Thompson *et al.*, 2011) and has resulted in a great number of sequences deposited in relevant databases, thus enhancing the value of 16S rRNA as the “gold standard” in microbial ecology. While the 16S rRNA gene fragment, containing one or more variable regions, is the preferred target marker gene for bacteria and archaea, this is not the case for fungi and eukaryotes where the preferred marker genes are the internal transcribed spacer (ITS) and 18S rRNA gene, respectively.

Taxonomic analysis for prokaryotes (i.e., bacteria and archaea) is regularly performed using 16S rRNA data derived from varying sequencing technologies (i.e., 454 pyrosequencing as well as Illumina, Solid and Ion Torrent), and, for the purposes of this review, we will list the relevant software to allow analysis for most sequencing technologies. Commonly used tools for 16S rRNA data analysis and denoising include QIIME, Mothur (Schloss *et al.*, 2009), SILVAngs, MEGAN and AmpliconNoise (Quince *et al.*, 2011). Despite the vast availability of algorithms and software for analysis of 16S rRNA metagenomics datasets, QIIME has been established as one of the most popular software (Nilakanta *et al.*, 2014).

It is important to be aware of certain aspects of the terminology required for the efficient analysis of 16S rRNA metagenomics data. These include the following: i) Amplicon: a DNA fragment that is amplified by PCR, e.g., one or more 16S rRNA variable regions, or other marker genes. Most researchers will make use of standard PCR primers; ii) OTU: species distinction in microbiology, typically using rRNA and a percentage of similarity threshold for classifying microbes within the same, or different, OTUs; iii) Barcode: a short DNA sequence that is added to each read during amplification and that is specific for a given sample. This allows samples to be mixed (multiplexed) to reduce sequencing cost. During analysis, sequences need to be demultiplexed, i.e., separated by sample.

Analysis usually requires a reference database that is searched to find the closest match to an OTU from which a taxonomic lineage is inferred. Some widely utilized databases include Greengenes (16S rRNA), Ribosomal Database Project (Cole *et al.*, 2009) (16S rRNA), Silva (Pruesse *et al.*, 2007) (16S + 18S rRNA genes), and Unite (Kõljalg *et al.*, 2013) (ITS), although the usage of Greengenes is not recommended since it is not regularly updated. On other hand, Silva is one of the curated and recommended databases for taxonomic assignment. These databases are less suitable for certain groups of organisms, such as protists and viruses, which are extremely diverse and for which considerably less sequence information is available compared to bacteria. Although the workflow presented in this section was initially developed for the 16S rRNA, it is applicable to functional marker genes and necessary for analysis of functional gene datasets. However, in the latter case, the taxonomic assignment of sequences is performed using functional gene repositories, such as the FunGene (Fish *et al.,* 2013).

5.4.1 Denoising

Denoising is important for 16S rRNA metagenomic data analysis, and it is platform-specific; i.e., certain platforms (e.g., Illumina) require less denoising than others (e.g., pyrosequencing). For example, denoising of 454 pyrosequencing data, despite being computationally expensive, is necessary due to intrinsic errors generated from

pyrosequencing that can give rise to erroneous OTUs. A procedure called “flowgram clustering” removes problematic reads and increases the accuracy of the taxonomic analysis. Several denoising algorithms have been developed so far (Reeder and Knight, 2010; Quince *et al.*, 2011; Bragg *et al.*, 2012; Keegan *et al.*, 2012; Balzer *et al.*, 2013; Brodin *et al.*, 2013; Iyer *et al.*, 2013) but for the purpose of this review three of them will be analyzed in detail.

Denoising is performed very efficiently by AmpliconNoise (Quince *et al.*, 2011) a tool that uses the following basic denoising steps: 1) Filtering of noisy reads: reads are truncated based on the appearance of low signal intensities; 2) Removing pyrosequencing noise: distance between the flowgrams is defined and true sequences and their frequencies are inferred by an expectation-maximization (EM) algorithm; 3) Removing PCR noise: the same ideas are used for removing PCR errors; 4) Chimera identification and removal: for each sequence, exact pairwise alignments are performed to all sequences with equal or greater abundance, which is the set of possible parents. Although a considerable number of sequences is lost during the denoising process, it results in high-quality sequences (Gaspar and Thomas, 2012); however, there has been some debate on the level of stringency required to achieve such high quality (Bakker *et al.*, 2012).

A very popular software for the analysis of microbial communities is QIIME. Initially QIIME was implemented for use of 454 pyrosequencing datasets only, i.e., using sff (Standard Flowgram Format) files, but currently QIIME has been modified to accept the fastq file format, thereby making the analysis of Illumina datasets possible. The QIIME developers provide users with extensive online tutorials for several workflows, and, moreover, QIIME is available as an open-source software package mostly implemented using the programming language PYTHON.

Another widely used software for the analysis of microbial communities is Mothur. It was created from the combination of pre-existing software, such as DOTUR (Schloss and Handelsman, 2005), SONS (Schloss and Handelsman, 2006a) and Treeclimber (Schloss and Handelsman, 2006b), but, due to the community support it has received, currently it incorporates many more algorithms, thus providing the user with a variety of choices.

More recently, a web-based application called SILVAngs was developed, which provides a fully automated analysis pipeline for data derived from rRNA marker gene amplicon sequencing. The analysis workflow is based on 1) Alignment of reads, 2) Quality assessment and filtering of reads, 3) Dereplication, whereby identical sequences are filtered out to avoid overestimation, 4) Clustering and OTU picking using a priori defined thresholds, and 5) Taxonomic assignment of OTUs using the SILVA rDNA database.

The choice of which denoising algorithm to use is largely depends on the user. Once a choice is made, the user should also consider whether to deviate from the default parameters. Parameter adjustment is related to the dataset produced, i.e., which specific 16S rRNA gene region was sequenced and which technology was used to perform the actual sequencing. In addition, it has been suggested that use of different denoising methods can produce significantly different outcomes (Koskinen *et al.*, 2014), which should be taken into careful consideration when comparing studies that have utilized different algorithms for data analysis.

5.4.2 OTU clustering, picking, and taxonomic assignment

After the demultiplexing of the dataset, i.e., the assignment of reads to samples using barcode information, the next step is OTU picking. For bacteria/archaea, it is accepted that OTUs of similarity greater than 97% correspond to the same species, but also other dissimilarity cutoffs can be employed, if needed for the downstream analyses. There are numerous OTU picking strategies: 1) De novo is used if amplicons overlap and if a reference sequence collection is not available. It clusters all reads without using a reference and is quite expensive computationally, hence not very suitable for very large datasets. 2) Closed-reference is used if amplicons do not overlap and if a reference sequence collection is available. This approach discards reads that do not hit a reference sequence. 3) Open-reference is used if amplicons overlap and a reference dataset is available. This method clusters reads against a reference dataset, but if the reads do not match the reference, they are consequently clustered de novo. All the aforementioned are incorporated into

QIIME. There are also other types of OTU clustering and picking strategies being developed (Sun *et al.*, 2009; Hwang *et al.*, 2013; Patin *et al.*, 2013; Preheim *et al.*, 2013); the most appropriate choice for the downstream analysis will depend on the type of data and the user.

Taxonomic assignment of OTUs can be performed using a variety of algorithms. Currently QIIME supports numerous algorithms, such as BLAST, the RDP classifier, RTAX, Mothur classifier, and uclust, to search for the closest match to an OTU from which a taxonomic lineage is inferred. This requires reference databases of marker genes. Some commonly utilized databases include Greengenes (16S rRNA), Ribosomal Database Project (16S rRNA), Silva (16S + 18S rRNA genes), and Unite (ITS).

5.4.3 Statistical analysis and visualization of results

QIIME output includes a representation of a taxonomic tree in Newick format, which can be visualized in several applications, and a file in Biom (Biological Observation Matrix) format (McDonald *et al.*, 2012) representing OTU tables. This file can be imported into MEGAN for visualization or into any other statistical software requiring matrix-type data. In addition, alpha-diversity analysis (diversity within a sample, e.g., Phylogenetic Diversity (PD), Chao (Chao, 1984) etc.) and beta-diversity analysis (diversity across samples, e.g., UniFrac (Lozupone *et al.*, 2006), PCoA), as well as taxonomic composition and phylogenetic analyses, are supported through QIIME. Numerous other tools and software packages exist for performing statistical analysis of metagenomic data. The Primer-E package (Clarke and Gorley, 2006) is commonly utilized by microbial ecologists and allows for multiple multivariate statistical analyses, such as multidimensional scaling (MDS), analysis of similarities (ANOSIM), and hypothesis testing. Recently the R statistical programming language (R Core Team, 2015) has gained immense popularity and is currently widely used for multivariate statistics. Packages such as vegan (Oksanen *et al.*, 2008), phyloseq (McMurdie and Holmes, 2013) and Bioconductor (Gentleman *et al.*, 2004) provide multiple in-built functions and libraries for performing a wide range of statistical

analysis required for metagenomic datasets. While it is out of the scope of this review to thoroughly analyze visualization tools for genomic data, readers are encouraged to visit a recent review article (Pavlopoulos *et al.*, 2013).

5.5 Data Management, Storage, and Sharing

Tools such as IMG/MER, CAMERA, MG-RAST, and EBI metagenomics (which also incorporates QIIME) provide an integrated environment for analysis, management, storage, and sharing of metagenome projects. This requires that a consensus commonly accepted annotation scheme is designed in order to allow for efficient data exchange, integration, sharing, and visualization between different platforms and to further reduce the need for reprocessing of metagenomic datasets, a task which is very expensive computationally.

The GSC is currently investing heavily toward a widely accepted language that shares ontologies and nomenclatures thereby providing a common standard for exchange of data derived from the analysis of metagenomic projects. Toward this goal, MIMS (Minimum Information about a Metagenome Sequence) and MIMARKS (Minimum Information about a MARKer Sequence) (Yilmaz *et al.*, 2011) have been devised, providing a scheme of standard languages for metadata annotation.

5.6 Conclusions

Tools and databases for metagenomic data analysis are currently well on their way to becoming more and more efficient and elaborate (for an overview of the tools most utilized nowadays for metagenomic data analysis, see Table 5.2). Technologies offering increased read length, such as PacBio, or new chemistry, such as Irys Technology and Nanopore Sequencing, are beginning to offer new capabilities to the analysis pipelines and aid in many aspects the assembly as well as the concurrent annotation process. Assembly tools such as IDBA-UD are being developed and increasingly improved to address the specific problem of assembling mixtures of genomes as is eminent for metagenomic samples. Databases like GOLD (Reddy *et al.*,

2014), associated with the IMG/MER portal, can be used as a reference in order to perform validation tests for assembly tools. Moreover, the use of simulated metagenomic datasets has been proposed in order to asses these tools (Mavromatis *et al.*, 2007).

There has been some controversy within the metagenomics community regarding the actual need for performing assembly on metagenomes. One contention is that using clustering algorithms such as cd-hit (Li and Godzik, 2006; Fu *et al.*, 2012) or uclust is sufficient to group similar reads together and thereafter proceed to annotation of these clusters without prior assembly. This clustering approach may allow for more accurate annotation of highly diverse samples containing rare, uncultured genomes that may otherwise be excluded from the assembly process due to their low coverage. One drawback of not performing an assembly may be that complex regulatory elements such as CRISPRs may not be identified successfully (Thomas *et al.*, 2012).

Binning and annotation methods are also constantly being modified and altered to specifically address metagenomic analysis pipelines. A significant improvement of these processes will be achieved upon increase of the genomic repository of cultured as well as uncultured genomes within the public database repertoire. Composition-based as well as similarity-based binning methods, especially those making use of supervised machine learning algorithms (i.e., PhylopithiaS, trained on reference genomes), will become increasingly accurate due to the availability of more reliable information.

Table 5.2: Tools grouped according to their main functionality.

Shotgun metagenomics	**Assembly**	EULER	**Binning**	SOrt-ITEMS	**Annotation**	KEGG	
		Velvet		PhymmBL		MetaCluster TA	
		SOAP		MetaCluster		SEED	
		ABySS	**Annotation**	FASTX-Toolkit		eggNOG	
		MetaVelvet		FastQC		ProViDE (Ghosh *et al.*, 2011)	
		MetaVelvet-SL		SolexaQA		COG/KOG (Tatusov *et al.*, 2003)	
		Meta-IDBA		Lucy 2		PFAM	
		IDBA-UD		DUST		TIGRFAM	
		Newbler (Roche)		Bowtie		MetaPhlAn (Segata *et al.*, 2012)	
		MIRA		MetaGeneMark		HighSSR (Churbanov *et al.*, 2012)	
		Mapsembler (Peterlongo and Chikhi, 2012)		LEfSe (Segata *et al.*, 2011)		Blat	
		ALLPATHS (Butler *et al.*, 2008)		TACOA	**Analysis pipelines**	IMG/MER (Victor M. Markowitz *et al.*, 2014)	
		MetaORFA (Ye and Tang, 2009)		Metagene		MG-RAST	
		MetAMOS		CREST (Lanzén *et al.*, 2012)		MEGAN 5	
	Binning	TETRA		Prodigal		CAMERA	
		S-GSOM		mOTU-LG (Sunagawa *et al.*, 2013)		Parallel-META (Su *et al.*, 2014)	
		PhylopythiaS		Orphelia		EBI Metagenomics	
		TACOA		Kraken (Wood and Salzberg, 2014)		METAREP (Goll *et al.*, 2010)	
		PCAHIER		FragGeneScan		PHACCS (Angly *et al.*, 2005)	
		ESOM		CRT (Bland *et al.*, 2007)			
		ClaMS		NBC (Rosen *et al.*, 2011)			
		CARMA		MyTaxa (Luo *et al.*, 2014)			
		WGSQuikr (Koslicki *et al.*, 2014)		RITA (MacDonald *et al.*, 2012)			
		SPHINX (Mohammed *et al.*, 2011)		PILER-CR (Edgar, 2007)			
		MetaPhyler		tRNAscan (Xinbo Wang *et al.*, 2013)			

Table 5.2: Continued

Marker gene metagenomics	**Standalone software**	QIIME Mothur JAguc (Nebel *et al.*, 2011) M-pick (Xiaoyu Wang *et al.*, 2013) OTUbase (Beck *et al.*, 2011) CopyRighter (Angly *et al.*, 2014) AbundantOTU (Ye, 2011) UniFrac ESPRIT
	Analysis pipelines	SILVA FunFrame (Weisman *et al.*, 2013) PANGEA (Giongo *et al.*, 2010) FastGroupII (Yu *et al.*, 2006) CLOTU (Kumar *et al.*, 2011)
	Denoising	AmpliconNoise DADA (Rosen *et al.*, 2012) JATAC (Balzer *et al.*, 2013) UCHIME (Edgar *et al.*, 2011) Bellerophon (Huber *et al.*, 2004) CANGS (Pandey *et al.*, 2010)
	Databases	SILVA Greengenes Ribosomal Database Project (RDP) Unite

At this stage it is important to mention that, in spite of the best efforts to reconstruct and prepare datasets by 1) quality filtering, 2) performing assemblies, and 3) binning sequences into taxonomically informative groups, annotation pipelines still achieve successful annotation for only ~50% of the sequences under analysis (Gilbert *et al.*, 2010). As mentioned above, the annotation process is highly dependent on the available databases and hence limited by the amount of information that is present within these repositories. Sequences that do not have any similarity with any other sequence existing in a known database are termed "orphan genes" (Lespinet and Labedan, 2005). These genes are believed to be 1) a consequence of sequencing errors and/or reflect the inaccuracy of gene prediction tools, or 2) truly novel genes that have no sequence or function similarity to known genes and may share higher order similarity in the form of protein folds. A lot of work is currently being undertaken in order to shed some light on these unknowns/orphans using various types of information. Some existing tools use pathway information from metagenomic neighbors and also context-depended metabolomic data to assign a functional annotation to unknown genes (Smith *et al.*, 2012; Yamada *et al.*, 2012). Along these lines, the use of metabolomic, metatranscriptomic, and/or metaproteomic data will provide a more elaborate view of the "picture", addressing all aspect of the dogma of life in the metagenomics era. Moreover, single-cell genomics is now becoming increasingly popular by investigating information from sequencing individual cells. The synergy of single-cell genomics with metagenomics can allow a more accurate separation of metagenomics sequences into individual genomes, guided by the single-cell sequencing data.

A wide array of software is currently available to perform each step of the marker gene metagenomics analysis pipeline. What is missing from the literature is a systematic evaluation of software and algorithms that have been used so far and a standardized means of comparing results derived from different workflows. Variation in results can occur due to inconsistencies in a number of factors, such as DNA extraction (Cruaud *et al.*, 2014; Vishnivetskaya *et al.*, 2014), primer pair and amplification region (Kim *et al.*, 2011; Soergel *et al.*, 2012; Klindworth *et al.*, 2013), sequencing platform (Harismendy *et al.*, 2009) and the software used (Sun *et al.*, 2012). All of the aforementioned sources of variation make it very difficult to

compare and obtain trustworthy results. Computational and programming challenges to improve the already available software can be achieved, but only through benchmarks, simulations (Richter *et al.*, 2011) and thorough testing. Initiatives such as the GSC could potentially take over the design of the "Minimum Analysis Requirements of Metagenome Sequences (MARMS)". This will be made up of standardized methodologies and consensus in the choice of software, analysis steps, threshold values, and parameters. Such an initiative would eliminate, or at least minimize, the biases that can be generated by analyzing data using multiple methodologies.

The availability of data software such as EBI Metagenomics, IMG/MER, MG-RAST, and SILVAngs will further allow users with limited computational facilities to perform analysis of metagenomic samples. In comparative metagenomic analyses, one can use tools to compare samples from different ecological niches and extract information that is common and/or unique to a specific environment (Sangwan *et al.*, 2012; D'Argenio *et al.*, 2014). Moreover, the GSC is striving toward the successful integration of analyzed data under a unified and mutually acceptable structure/format that will facilitate the exchange of valuable insights and information in the field of microbial ecology and environmental microbiology.

To sum up, we have created a metagenomics flowchart (Figure 5.1) outlining all the aforementioned basic steps of the analysis pipeline. Analysis can take two different routes depending on the type of sequencing data (marker gene or shotgun metagenomics). Every analysis step shown in the flowchart is complemented by a list of some well-established tools used by the metagenomics community.

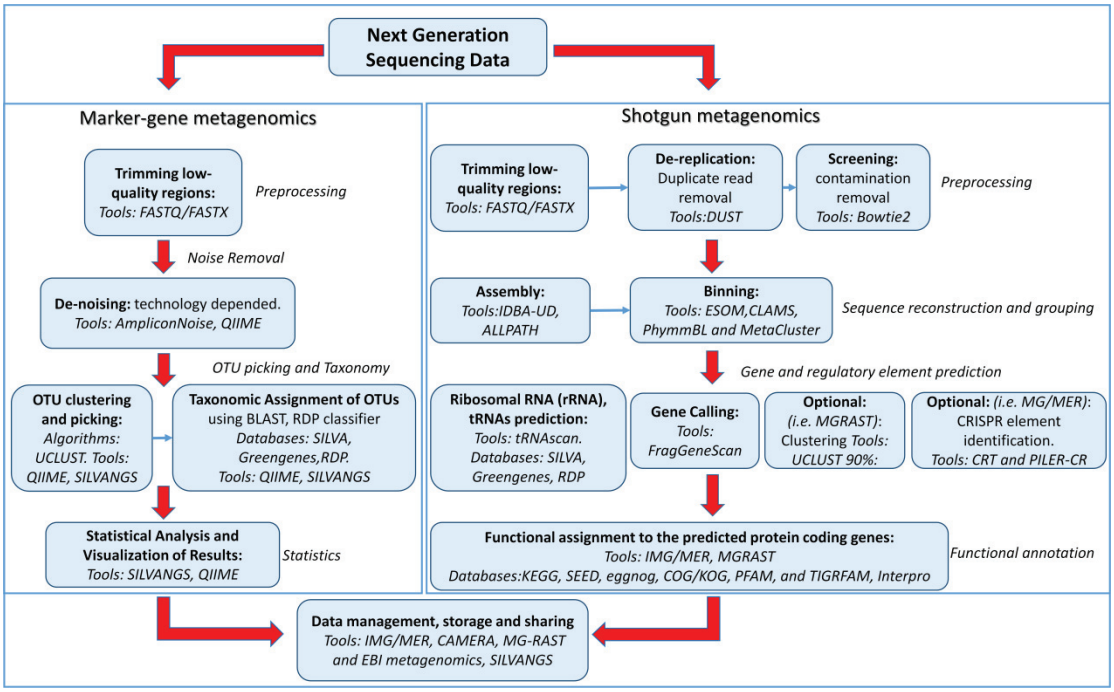

Figure 5.1: Flowchart of basic metagenomics steps and tools currently in practice[10].

[10] The analysis pipeline can take two different routes depending on the type of sequencing data (marker gene or shotgun metagenomics) available. The flowchart outlines the basic steps in the analysis pipeline starting with preprocessing of the data to the final extraction of results and concurrent storage and management of the data. Some popular tools that have been used extensively by the metagenomics community are shown for every step, as a well as the databases and algorithms in common practice.

Chapter 6: General discussion and conclusions

6.1 Molecular tools to characterize marine microbial communities

The aim of this PhD project was to study the microbial community assemblages of hypoxic ecosystems and to identify a potential link between their identity and function, using state-of-the-art approaches such as targeted metagenomics and RNA stable isotope probing.

Although 16S rRNA gene has been considered the golden standard for microbial community studies (Vos *et al.*, 2012), the short reads that are the output of the high-throughput sequencing technologies do not allow for a good taxonomic resolution to be achieved, although this issue has been debated extensively (Liu *et al.*, 2007; Jeraldo *et al.*, 2011). In addition, high-throughput sequencing can only produce relative abundance data as a main output; this, combined with the fact that 16S rRNA is found in multiple copies in each microorganism, may bias the results and conclusions of each study. Also, different conclusions may be derived if the presence or absence of a given OTU is taken into consideration, and not only its relative abundance at a given study site (see Chapter 3). Other genes, which are found in single copies in microorganisms, such as the *rpoB* have been proposed as alternative marker genes for the bacterial diversity assessment (Vos *et al.*, 2012). However, the public databases which are being used for the taxonomic assessment of sequence reads are "biased" towards 16S rRNA, since it has been used extensively as a marker gene for the general "tree of life" creation. Therefore, the usage of another marker gene cannot substitute completely 16S rRNA, i.e. it can only be used as a complement to achieve better taxonomic resolution.

Another obstacle is that the majority of sequence reads in every study can only be classified as uncultured; thus, a high amount of information is of no practical use since no information on the functional aspect of those microorganisms can be derived from such a taxonomic assignment. One solution to this issue would be the choice of protein coding genes, instead of 16S rRNA, that would allow for specific functional groups to be assessed in the studied habitats. Examples of such protein

coding genes are nitrite reductases (e.g. *nirS*) (Bowen *et al.*, 2013), methyl-coenzyme M reductases (e.g. *mcrA*) (Yang *et al.*, 2014) and dissimilatory sulfate reductaces (e.g. *dsrB*), that was chosen for this PhD project (see Chapter 3). However, even in the case of protein coding genes, the produced results are undoubtedly affected by the reference dataset used for the classification of the sequences (Müller *et al.*, 2015).

Apart from the choice of sequencing functional genes, and given the immense number of 16S rRNA sequences deposited in public databases, another useful approach towards the enhancement of high-throughput sequencing studies is the implementation of algorithms and software that can predict functional profiles of microbial communities based on 16S rRNA data (Langille *et al.*, 2013; Aßhauer *et al.*, 2015) (see Chapter 2). Similarly, other types of software, as Seqenv (Sinclair *et al.*, 2016), can add environmental metadata to the studied sequences and therefore, provide information on where the relevant microorganisms have been previously found (see Chapter 3). Combinations of such approaches may increase the amount of knowledge derived from such studies, and therefore the insight in microbial communities.

However, as discussed in Chapter 5, there are more issues that arise from each study focusing on microbial communities and using high-throughput sequencing. The different DNA extraction, PCR and library preparation protocols that exist in the literature, paired with the choice of primer pair and amplification region introduce variations and biases in each study. Apart from the above, the choice of sequencing platform can also influence the outcome of a microbiome study. The first high-throughput sequencing studies were performed using 454 GS FLX (Roche), which was the first commercially available pyrosequencer (see Chapters 3 and 5). This invention gave rise to many more sequencing platforms, with updated and less error-prone chemistry. Currently, the “default” platforms for a microbial community study, as well as for other applications, are manufactured by Illumina (see Chapters 2, 4 and 5). However, there are constant improvements and inventions of other platforms (see Chapter 5).

Moreover, the choice of software for the subsequent data analysis can amplify the variations between the studies; what is being compared between different research papers is in fact incomparable and conclusions drawn from these comparisons

should be cautiously taken in consideration. Different sequencing methodologies and software applications can vary considerably in the produced results (Amore *et al.*, 2016); although certain benchmark studies exist (Sun *et al.*, 2012), the vast availability of software and analysis pipelines has not been compared, evaluated and standardized. The use of online integrated platforms, such as IMNGS (Lagkouvardos *et al.*, 2016), allows users with limited computational facilities to perform analysis of metagenomic samples and facilitates the accurate comparison of metagenomic analyses studies (see Chapter 4). In addition, there is an advantage in using publicly available analysis software, such as R (R Core Team, 2015) (see Chapters 2 and 3), compared to other commercial software (see Chapter 4). In the majority of cases, there is an open source option for the analysis of large-scale data, which is actually more robust and better documented than its commercial twin.

6.2 Microbial community composition in hypoxic sediments

According to this research project, it has been shown that the microbial community diversity pattern can be differentiated based on habitat type, i.e. between riverine, lagoonal and marine environments (see Chapter 2). This is in accordance with studies conducted in similar environments, such as the Baltic Sea (Herlemann *et al.*, 2011).

In addition, a second level of differentiation can be observed when different lagoons are taken into consideration (see Chapter 3); each lagoon hosts a different SRM community, which is consistent with other studies suggesting that SRM communities exhibit biogeographic distribution patterns at small spatial scale (Pérez-Jiménez and Kerkhof, 2005). This finding is in accordance with a previous study in the same sampling sites where it was shown that sediment microbial communities, as revealed by 16S rRNA sequencing, are also differentiated based on the sampled lagoon (Pavloudi *et al.*, 2016).

Furthermore, the abiotic variable that is best correlated with the microbial community pattern is salinity, thus supporting the original hypotheses (see Chapters 2 and 3). However, when the total number of OTUs in different habitats is taken into consideration, a negative linear relationship with salinity is identified concurring with

a model previously described by Attrill (2002), while at the same time questioning the observations of Remane (1934) and Herlemann *et al.* (2011).

As expected, the majority of OTUs can be classified as uncultured (see Chapter 3), *Proteobacteria* and *Bacteroidetes* (see Chapter 2). Nevertheless, taxonomic assignment can reach the level of family, but in very few cases it can reach lower levels. This is indicative of the high diversity exhibited in lagoonal environments, and transitional water ecosystems in general, which have only few representative microorganisms that have been cultured and described.

Regarding the phylum of *Proteobacteria*, the most abundant class is *Deltaproteobacteria*, and in particular with *Desulfobacteraceae*, which have been shown to exhibit high abundances in marine (Leloup *et al.*, 2009), saline and hypersaline (Foti *et al.*, 2007) and deep sea sediments (Kaneko *et al.*, 2007) (see chapter 3). Nevertheless, the abundance of the *Desulfobacteraceae* family was fluctuating according to salinity, i.e. it was higher in more saline stations; thus, despite being the most abundant, *Deltaproteobacteria* are not distributed evenly in transitional water ecosystems (see chapter 3). In addition, the presented results (see Chapter 2) confirm the existence of typical freshwater bacterial taxa (e.g. de Bie *et al.*, 2001), such as the *Betaproteobacteria*, which were almost exclusively present in the inner station and mouth of the Arachthos river. Other microbial taxa, such as the *Gammaproteobacteria*, are adapted to live in higher salt concentrations and osmotic pressure (Wu *et al.*, 2006; Zhang *et al.*, 2014b; Herlemann *et al.*, 2016), although they can dominate brackish habitats, as has been shown in the present study.

In the sediments of the Benguela coastal upwelling system, bacterial species and, in general, taxonomic groups that were expected to be thriving, were absent or present but in very low abundances (see Chapter 4). Examples of such species are the large sulfur bacteria, such as *Beggiatoa* spp. and *Thiomargarita namibiensis*, which couple the oxidation of sulfide using nitrate as the terminal electron acceptor (Schulz *et al.*, 1999; Dale *et al.*, 2009). Also, *Epsilonproteobacteria*, which have been found to catalyze chemolithotrophic oxidation of sulfide with nitrate (Lavik *et al.*, 2009), were found in very low abundances at the *in situ* communities. Similarly, *Planctomycetes* which include the taxonomic order of *Planctomycetales* that catalyze anaerobic ammonium oxidation (Wang and Gu, 2013) were also found in very low abundances.

Since the aforementioned microbial groups have been previously found in high abundances in the study area, it could suggest that there might be a seasonal effect on their appearance and enhancement. Furthermore, different approaches, such as sequencing or quantitative PCR on functional genes, might reveal that representative of these groups actually exist in the samples of the present study. As previously found for lagoonal sediments, the most abundant bacterial taxa were *Deltaproteobacteria*, suggesting that sulfate reduction is occurring in the sampling stations, although other electron acceptors can be used instead (see Chapter 3).

In this environment which is more or less stable in terms of salinity, microbial community patterns were correlated with delta ^{15}N, sediment water content and oxygen concentration. Delta ^{15}N can be interpreted as signal of nitrogen consuming anaerobic processes, whilst reflecting the sediment type (Nagel *et al.*, 2016), which in turn is dependent on the depth and sedimentation of the study area.

Overall, it can be concluded that different examples of hypoxic marine environments have certain similarities regarding their microbial communities. Both of the selected environments are under studied, since a large fraction of the retrieved sequences corresponds to uncultured *Bacteria*. This hightlights the need for more and different type of studies in these habitats since there are still many aspects of their microbial communities that remain undeciphered. In addition, biogeochemical cycling in both the studied ecosystems is in favour of sulfate reducing and nitrate reducing bacteria, although their diversity should be investigated in more detail due to the low resolution of the 16S rRNA gene, especially when the goal is the investigation of the functionality of the microbial communities.

6.3 Functional community composition in hypoxic sediments

Apart from taxonomic diversity of the microbial communities, it has been shown that the studied habitats were functionally distinctive (see Chapter 2). There were more retrieved metabolic pathways in the case of the marine samples and less in the case of the lagoonal and the riverine samples. This is indicative of the number of studies that have been conducted in the marine environment so far, which allowed for more

OTUs to be transferred to KEGG reference organisms. In addition, apart from salinity, in the case of functional community patterns, oxygen concentration was highly correlated with the predicted metabolic pattern of the microbial communities under study, which partially supports the original hypothesis but also augments it.

As mentioned previously, the identification of the environmental descriptive terms that characterize the microbial communities expand the information derived from the sequence reads. In the case of the SRM communities, the majority of environmental terms were classified to the marine biome, but terms belonging to the freshwater or brackish biomes were also found. As expected, the latter were found in stations where a freshwater effect was more evident (see Chapter 3).

Another method that can link directly the function and identity of microorganisms is SIP (Friedrich, 2006) and it was employed in the sediment samples of the Namibian oxygen minimum zone (see Chapter 4). Although SIP is an informative method with vast potential applications, its application is not trivial and broader scale studies are not very feasible. For example, ideally, each application of SIP would require a prior pilot experiment, where different substrates and different durations would be tested in order to conclude safely on the ideal combination of substrate and duration to maximize the end point results and to avoid cross feeding. In the case of Chapter 4, samples were retrieved from the third day of the experiment also in order to test whether this duration was enough for a clear labelling and fraction separation to be observed. However, in this case, the separation of the fractions was not as clear as the one retrieved from the set of samples that were incubated for six days. Therefore, the first set of samples was not processed further.

The isotopic enrichment, i.e. the incorporation of the label in the ^{13}C acetate amended incubations, and the successful completion of the SIP experiment were confirmed. Thus, the produced results on the functionality of the microbial communities at the stations were SIP was employed are trustworthy. Overall, *Epsilonproteobacteria* dominated the anaerobic assimilation of acetate as has been also shown from similar studies (Webster *et al.*, 2010). In addition, enhancement of known sulfate-reducers, such as *Deltaproteobacteria* and *Firmicutes* (Muyzer and Stams, 2008; Müller *et al.*, 2015) was not achieved under sulfate addition. This could be attributed to the competition for electron donors among nitrate-reducers and

sulfate-reducers (Achtnich *et al.*, 1995). Alternatively, it could be due to the fact that acetate cannot be used as an electron donor by all the sulfate-reducing bacteria (Madigan *et al.*, 2003).

Under the expansion of hypoxic conditions, chemolithoautotrophic microbes are expected to be affected (Lipsewers *et al.*, 2017), since, in general, anaerobic reactions and pathways will be favoured. Currently, sulfide-oxidizing bacteria have been shown to control the flux of hydrogen sulfide to the water column in OMZs (Schulz *et al.*, 1999). However, should sulfate reduction be enhanced, concentration of hydrogen sulfide would increase; such an increase would probably be diffused in the water column (Emeis *et al.*, 2004) and not retained in the sediment in oxidized forms.

In addition, the absence or the low abundance of nitrate-reducing sulfide-oxidizing bacteria from transitional water ecosystems, such as lagoons, could be one of the reasons that extensive hypoxic episodes are occurring frequently in such habitats. Hypoxic episodes, and their subsequent effects in aquatic organisms, in the Benguela region can be ameliorated by the presence of nitrate-reducing sulfide-oxidizing bacteria. However, such bacterial groups have not been identified in the studied lagoons which may be enhancing the occurrence of hypoxic events. Thus, enhancement of such microbial groups may decrease the hydrogen sulfide flux from the sediment to the water column and, perhaps, restrict the occurrence and duration of hypoxic episodes.

Moreover, the impact of hypoxic environments in the global biogeochemical cycling of nitrogen and sulfur should be studied more thogoughly; the balance of the cycles in the study sites is leaning towards the anoxic reactions since they are by definition favoured by the prevailing environmental conditions. In the light of climate changes and given that hypoxia is expected to impact more areas (Stramma *et al.,* 2008), the functionality of microbial communities would also be subjected to change. However, this change is still indeterminable despite of each importance.

6.4 Microbial vs Macrobenthic communities in the study sites

It is well known that burrowing macrobenthic organisms may influence significantly the bacterial sediment communities through changes in oxygenation and redox potential of surficial and burrow-lining sediments (Papaspyrou *et al.,* 2006). They increase the available surface area for diffusive exchange between anoxic pore-water and the overlying water, and thus they create new surfaces for microbial colonization at the oxic-anoxic interface (Glud, 2008). In addition, it has been shown that bacterial biofilms can either inhibit or induce larval settlement of macrobenthic organisms (Chan and Walker, 1998; Wieczorek and Todd, 1998; Huggett *et al.,* 2006). Furthermore, research has shown that hypoxia can compromise larval development and settlement of polychaetes (Dubilier, 1988; Shin *et al.,* 2013).

One of the differences between the study sites, i.e. the lagoonal complex of the Amvrakikos Gulf and the Benguela coastal upwelling sediments, is the occurrence and abundance of macrofauna. Although lagoonal benthic communities undergo temporal variations, due to the occurrence of dystrophic crises, there are certain tolerant species that have been found inhabiting the study sites (Nicolaidou *et al.,* 2006). However, changes in environmental conditions, among other biological interactions, seem to have an effect on the abundance of the species (Nicolaidou *et al.,* 2006).

The latest sampling for macrobenthic organisms was conducted during the same sampling campaign that sediment samples were collected for Chapter 3 (http://ipt.medobis.eu/resource?r=zoobenthos_in_amvrakikos_wetlands). It is clear that the macrobenthic diversity pattern differs per lagoon, as there are certain lagoons where there is high abundance of macrobenthos, such as Logarou lagoon, whereas others exhibit a much lower macrobenthic species abundance, such as Tsoukalio lagoon. This is undoubtedly reflected in the sediment abiotic parameters, as in the first case there was a thicker layer of oxygenated sediment, in contrast to the second case. As discussed previously, this could be one of the parameters that have affected the microbial community composition at the studied lagoons.

Research on the presence of macrobenthic organisms in the Benguela coastal upwelling sediments is rather limited and the are has been described as one of the world's regions with "major biodiversity gaps" (Konar *et al.,* 2010). Overall, macrobenthic species richness in the area has been considered to be relatively poor (Sakko, 1998) and it has been shown to be strongly affected by oxygen deficiency (Eisenbarth and Zettler, 2016); macrofauna often exhibit higher diversity abundances at the OMZ edges, rather than in the OMZ itself due to the limited oxygen availability (Levin, 2003). Macrobenthic communities in the Benguela region have been found to differentiate along the water depth, with the deep-sea community exhibiting the highest diversity (Eisenbarth and Zettler, 2016); however, more studies are needed and in a greater extent, i.e. expanding in different gradients, in order to draw safer conclusions on the macrobenthic species diversity of the area. These observations, along with the lack of sediment consistency that was observed in the OMZ, could have affected the existence of the microbial communities that were found at the sampling sites.

6.5 Future perspectives

This PhD project focused on microbial communities in two types of hypoxic ecosystems characterized by high productivity and anthropogenic value: lagoonal ecosystems and OMZs. Although certain aspects of their microbial communities have been identified, there are still many questions that remain unanswered.

Although state-of-the-art techniques were employed, new techniques and approaches are constantly becoming relevant in microbiome studies. Application of protein-SIP (Jehmlich *et al.*, 2008) in the studied habitats would allow for phylogenetic information to be obtained, along with functional information, for the labelled communities in each case. However, there are also limitations in the usage and applicability of protein-SIP since the databases that are available for protein based identification of microorganisms have been developed on cultured representatives. Therefore, identification of unknown proteins from uncultured microbes is still rather troublesome. RNA SIP can be also employed using different

labelled substrates, i.e. electron donors, and different electron acceptors. Usage of ^{15}N or ^{18}O instead of ^{13}C is still an option, although it should be taken into consideration that the resulting labelling pattern would be less clear. Carbon contributes to about 50 % of a cell dry weight, which by far outcompetes oxygen (17 %) and nitrogen (13 %) (Madigan *et al.*, 2003); thus, usage of ^{13}C is recommended for stable isotope studies since incorporation of the isotope in the cell is expected to be greater.

Another technique that could be more informative for future studies is combined usage of metagenomics and metatranscriptomics. This would allow the direct comparison between the available genes in the habitat under study and the actual expressed genes under the prevailing environmental conditions.

In addition, sequencing of different functional genes could allow for the communities involved in each step of the studied biochemical cycles to be identified; thus, coupling between oxidation and reduction reactions could be possibly quantified. Furthermore, the choice of new sequencing technologies, such as Nanopore sequencing (Branton *et al.,* 2008), despite being error-prone to sequencing errors deriving from its chemistry, has the potential to increase the number of reads, as well as their length; thus, overall, taxonomic resolution could be ameliorated and safer conclusions could be drawn.

Apart from the molecular techniques and studies, it is crucial to isolate and culture more microorganisms. Despite being considered as a step backwards, since high throughput sequencing became commercially available and monopolized research papers for many years, culturing is the only method that could clearly fill in the taxonomy gaps and enrich the databases with valuable information. In addition, it would indirectly provide more information since novel approaches, such as the functional prediction of OTUs, would be more precise. The era of metagenomics has provided a plethora of information on the metabolic needs of microorganisms that was unknown before. Hence, in spite of the technical difficulties of this method, new media can be designed and tested in order for more microbes to acquire a scientific name and placement in the phylogenetic tree.

Future research should focus more on the community functioning of such habitats; an increased understanding of the biogeochemical cycles that characterize these

hypoxic ecosystems will perhaps allow for predictions regarding the intensity and direction of the cycling of elements, especially of nitrogen and sulfur given their biological importance. This would facilitate the end-users of these ecosystems in the programming and organization of their activities, such as aquaculture, fishing, deep-sea mining etc., so that they would increase their benefits, e.g. achieve higher productivity in terms of fish catches, and minimize their losses.

Appendix

Supporting Material for Chapter 2

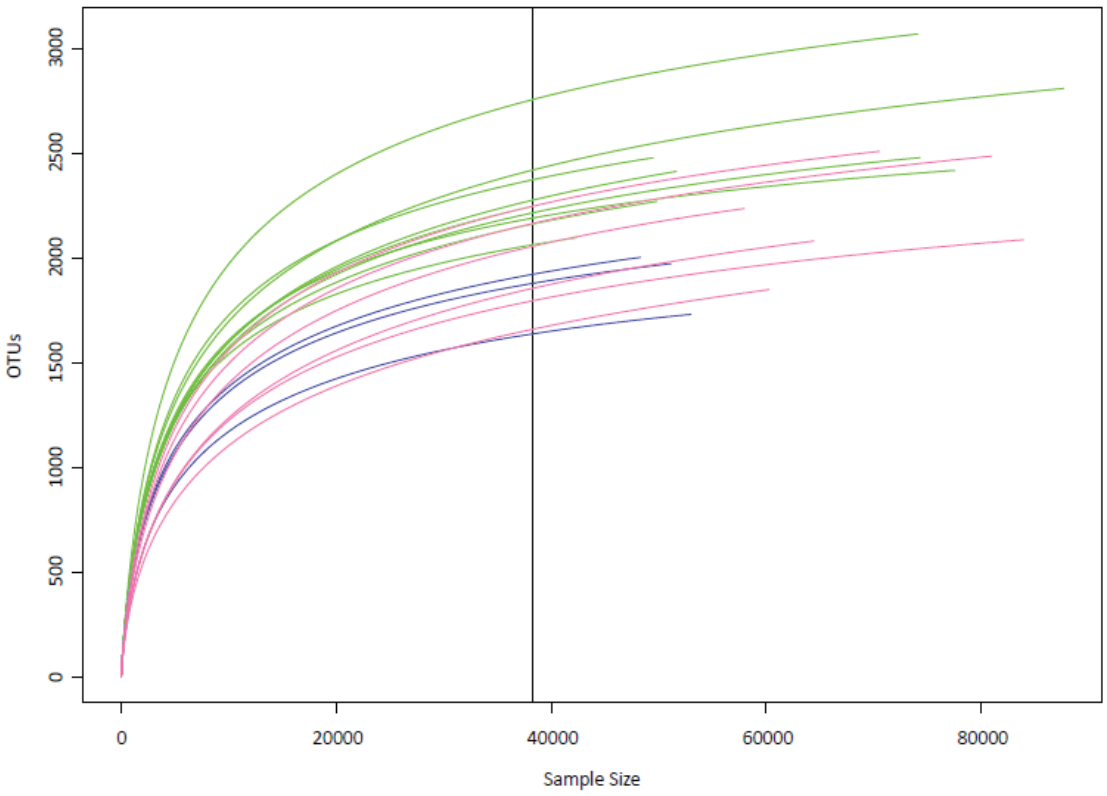

Figure S2.1: Rarefaction curves of the samples. The vertical line (raremax = 38,209) is the smallest number of individuals per sample to extrapolate the expected number of OTUs if all other samples had only that number of individuals. Rarefaction estimated the expected OTU richness in random subsamples of the community, with each subsample having 38,209 individuals. Blue: samples from Kalamitsi station. Pink: samples from Logarou stations. Green: samples from Arachthos stations.

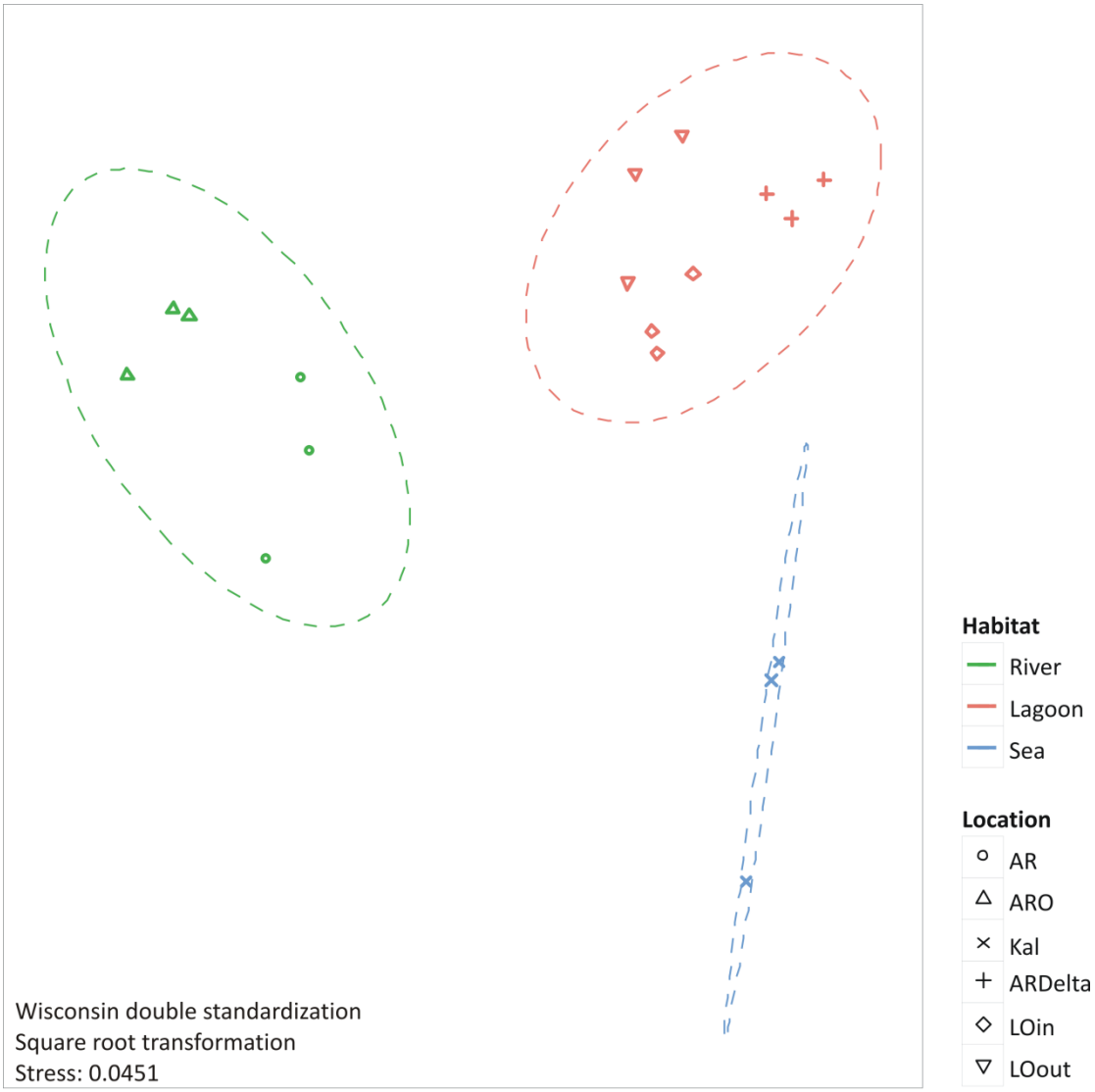

Figure S2.2: nMDS of the similarity matrix of the sampling stations based on the microbial OTUs abundances, with Arachthos Delta grouped with lagoonal samples. Ellipses according to habitat, signs according to location. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

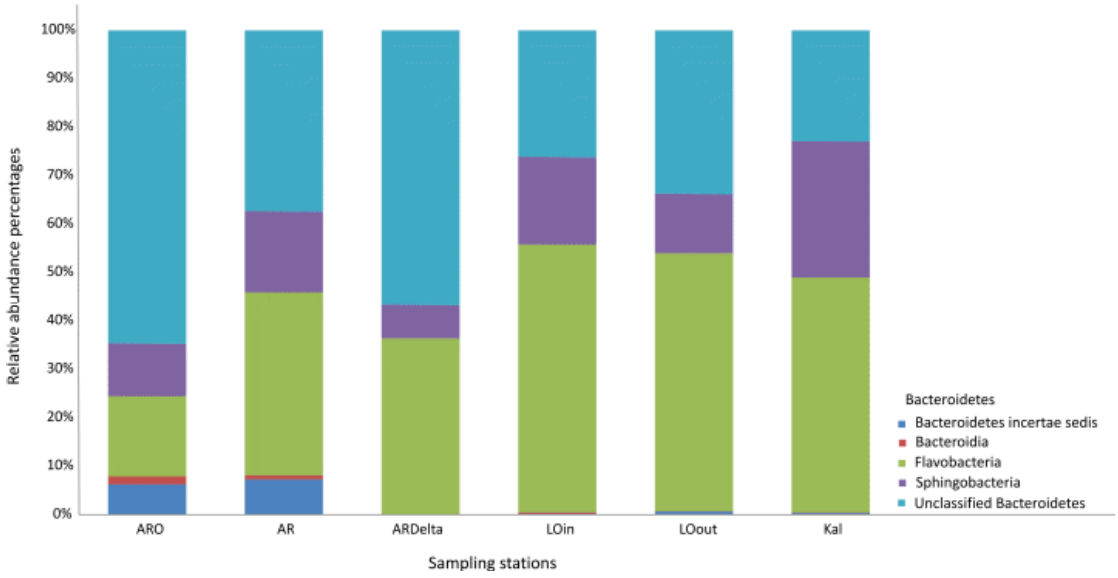

Figure S2.3: Bar chart showing the abundances of the main classes of the Bacteroidetes phylum, at the sampling stations. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

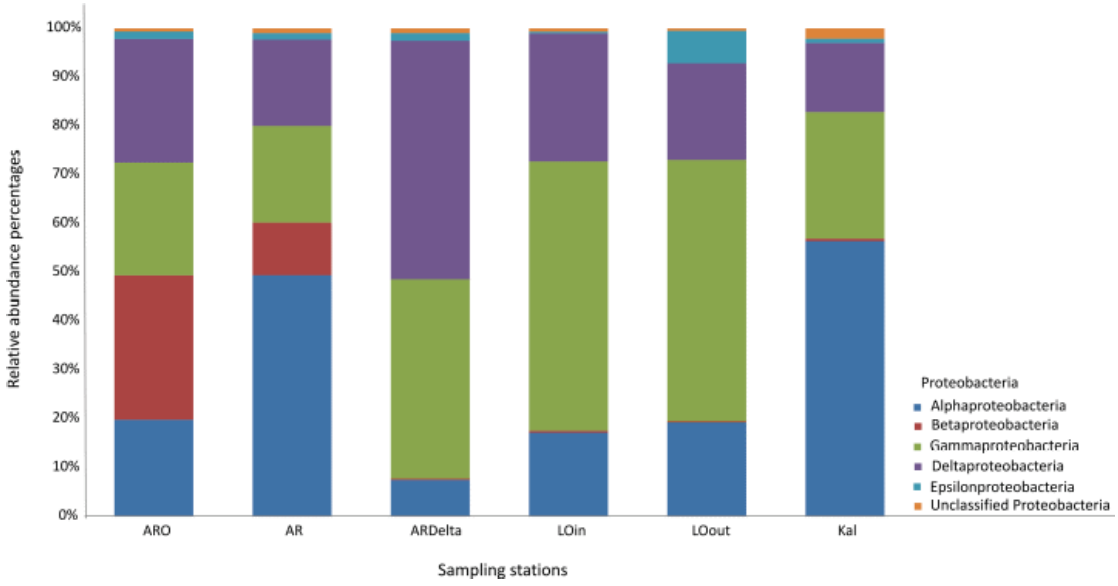

Figure S2.4: Bar chart showing the abundances of the main classes of the Proteobacteria phylum, at the sampling stations. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

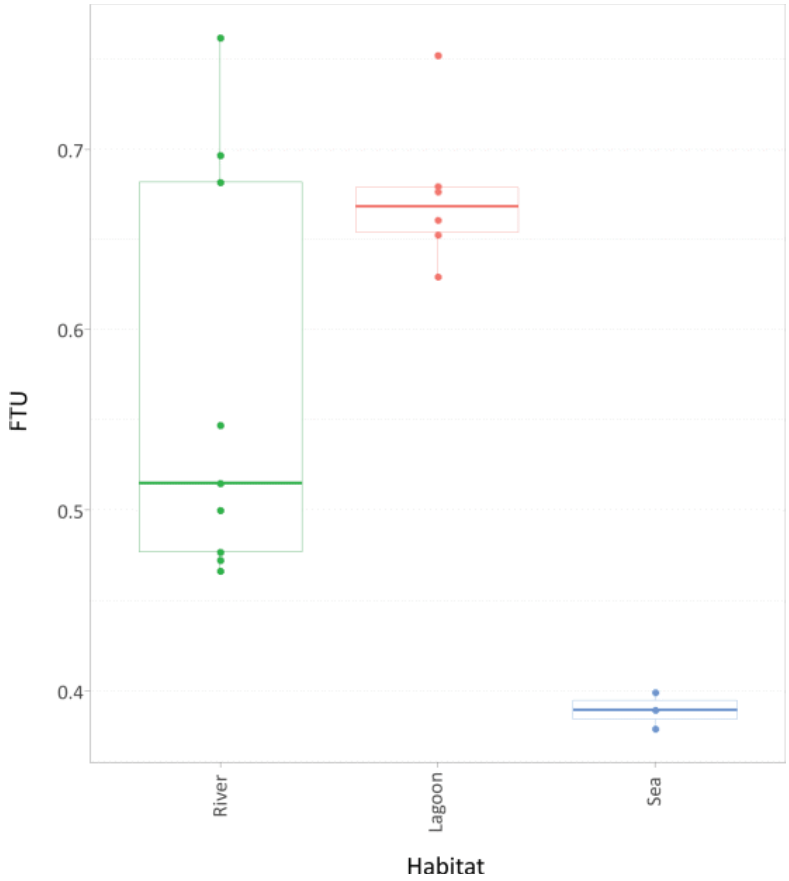

Figure S2.5: The fraction of unexplained taxonomic units per habitat, as it was derived during the functional profiling using UProC KEGG Ortholog reference profiles in long read mode. FTU: fraction of unexplained taxonomic units.

Table S2.1: The results of the processing of the sequences[11]. R: River. L: Lagoon. S: Sea. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi. A, B, C: replicate samples. a: initial number of read pairs. b: number of read pairs after the quality trimming. c: number of read pairs with error-corrected bases. d: number of changed base during error-correction. e: number of bases that failed error-correction. f: total number of bases. g: final number of read pairs. h: number of overlapped reads.

A/A	Sample libraries	a	b	c	d	e	f	g	h
1	R_AR_A	97,520	97,257	103,854	119,220	17,195,293	49,005,323	97,140	85,492
2	R_AR_B	157,246	156,744	150,418	175,130	22,017,363	79,182,431	156,539	138,330
3	R_AR_C	167,487	166,853	157,004	173,927	22,205,133	83,982,212	166,584	147,668
4	R_ARO_A	124,486	124,133	127,234	141,985	19,113,504	62,298,698	123,994	108,362
5	R_ARO_B	148,083	147,533	142,035	158,385	20,558,495	74,288,330	147,315	130,082
6	R_ARO_C	142,040	141,410	138,108	156,909	20,503,832	71,090,930	141,171	124,829
7	R_ARDelta_A	99,760	99,117	127,995	144,962	22,421,254	48,221,228	98,852	83,896
8	R_ARDelta_B	120,241	119,547	150,052	172,265	26,063,598	58,281,725	119,225	100,555
9	R_ARDelta_C	94,144	93,268	119,390	134,307	21,794,657	45,218,135	92,902	78,631
10	L_LOin_A	130,200	129,690	140,728	162,373	22,040,685	65,001,568	129,483	113,625
11	L_LOin_B	108,906	108,537	107,440	122,286	15,054,224	54,614,349	108,375	95,393
12	L_LOin_C	110,955	110,433	114,615	130,986	17,301,282	55,378,030	110,234	97,101
13	L_LOout_A	96,457	96,104	96,991	108,608	14,511,935	48,753,439	95,971	85,503
14	L_LOout_B	144,529	144,022	155,890	183,333	24,598,512	71,687,911	143,792	124,088
15	L_LOout_C	139,149	138,383	131,454	150,722	18,309,051	69,390,494	138,091	121,691
16	S_Kal_A	95,387	94,933	106,728	118,775	16,034,525	47,487,795	94,773	83,828
17	S_Kal_B	100,113	99,642	108,884	123,905	17,366,619	49,583,169	99,461	86,788
18	S_Kal_C	100,822	100,148	112,123	127,559	17,680,003	49,756,703	99,846	87,638
Sum									1,893,500

[11] Details on the sequence processing can be found at the bioinformatics tutorial of Dr. Umer Zeeshan Ijaz, entitled "Illumina Amplicons OTU Construction with Noise Removal" (available at http://www.tinyurl.com/JCBioinformatics).

Table S2.2: The Chi-Square values of the Kruskal-Wallis tests for the diversity indices between the locations and the habitats. OTUs: total number of OTUs. N: total microbial relative abundance values. H′: Shannon-Wiener. J′: Pielou's evenness. d: Margalef's species richness. ACE: Abundance Coverage Estimator. *: $p < 0.05$. **: $p < 0.01$. n.s.: not significant.

	OTUs	N	d	J′	H′(ln)	Chao-1	ACE
Location	n.s.	n.s.	11.129 *	13.28 *	13.767 *	n.s.	n.s.
Habitat	7.596 *	n.s.	8.425 *	13.177 **	12.966 **	7.799 *	7.583 *

Table S2.3: The values of the Mann-Whitney U tests used for the *post-hoc* pairwise significant comparisons, after the Bonferroni correction. OTUs: total number of OTUs. H′: Shannon-Wiener. J′: Pielou's evenness. d: Margalef's species richness. ACE: Abundance Coverage Estimator. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi. *: $p < 0.05$ before the Bonferroni correction. **: $p < 0.017$ in the case of the habitats. n.s.: not significant.

		OTUs	d	J′	H′(ln)	Chao-1	ACE
Location	AR *vs* LOin	n.s.	n.s.	3.971 *	n.s.	n.s.	n.s.
	AR *vs* LOout	n.s.	n.s.	3.971 *	n.s.	n.s.	n.s.
	AR *vs* Kal	n.s.	n.s.	4.091 *	n.s.	n.s.	n.s.
	ARO *vs* LOin	n.s.	n.s.	3.971 *	3.971 *	n.s.	n.s.
	ARO *vs* LOout	n.s.	n.s.	3.971 *	3.971 *	n.s.	n.s.
	ARO *vs* Kal	n.s.	n.s.	n.s.	3.971 *	n.s.	n.s.
	ARDelta *vs* LOin	n.s.	n.s.	3.971 *	n.s.	n.s.	n.s.
	ARDelta *vs* LOout	n.s.	n.s.	3.971 *	n.s.	n.s.	n.s.
	ARDelta *vs* Kal	n.s.	n.s.	4.091 *	n.s.	n.s.	n.s.
Habitat	River *vs* Lagoon	n.s.	n.s.	10.366 **	10.198 **	n.s.	n.s.
	River *vs* Sea	6.231 **	6.231 **	5.658 *	6.319 **	6.231 **	6.231 **
	Lagoon *vs* Sea	n.s.	n.s.	n.s.	n.s.	4.267 *	4.267 *

Table S2.4: The retrieved KEGG metabolic profiles for each sample. R: River. L: Lagoon. S: Sea. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi.

KEGG profile	R_AR	R_ARO	R_ARDelta	L_LOin	L_LOout	S_Kal
K00936	0.009492	0.010892	0.010948	0.009943	0.009910	0.009782
K03296	0.004774	0.004335	0.004524	0.004674	0.004654	0.004741
K06147	0.004574	0.004845	0.003578	0.004387	0.004409	0.004945
K03406	0.004383	0.004567	0.004867	0.003914	0.004070	0.003392
K02014	0.005889	0.004583	0.003118	0.004098	0.004366	0.002918
K00059	0.005071	0.003739	0.002681	0.003471	0.003449	0.004221
K02003	0.003619	0.003430	0.002958	0.003449	0.003521	0.003662
K02004	0.004109	0.003359	0.002483	0.003724	0.003859	0.002732
K02481	0.002720	0.003128	0.003890	0.002903	0.002629	0.002124
K09687	0.003040	0.002796	0.003012	0.002880	0.002884	0.002690
K08884	0.003547	0.002646	0.001737	0.003094	0.002442	0.003366
K00540	0.003274	0.002945	0.002162	0.002733	0.002711	0.002813
K02584	0.002503	0.002949	0.003767	0.002701	0.002469	0.001998
K01897	0.002629	0.002864	0.003237	0.002596	0.002726	0.002214
K07712	0.002546	0.002942	0.003519	0.002683	0.002464	0.002069
K02049	0.002769	0.002554	0.002055	0.002594	0.002618	0.003125
K01533	0.002285	0.002413	0.002485	0.002663	0.002822	0.002299
K01990	0.002368	0.002239	0.002575	0.002510	0.002493	0.002386
K02032	0.002674	0.002288	0.001729	0.002245	0.002406	0.002810
K02488	0.001998	0.002474	0.002583	0.002351	0.002481	0.002227
K13924	0.002027	0.002231	0.002355	0.002464	0.002709	0.002206
K03320	0.001979	0.002255	0.002249	0.002377	0.002539	0.002252
K02667	0.002064	0.002551	0.003093	0.002254	0.002060	0.001621
K10126	0.002092	0.002390	0.002836	0.002223	0.002087	0.001693
K02035	0.002571	0.002066	0.001437	0.001928	0.002126	0.002682
K03701	0.002037	0.001990	0.002144	0.002355	0.002211	0.002005
K07636	0.001876	0.002324	0.002515	0.001986	0.002061	0.001703
K03657	0.001937	0.002061	0.002198	0.002085	0.002132	0.002012
K02013	0.002357	0.002027	0.001926	0.002025	0.001993	0.002073
K03088	0.002905	0.002036	0.001396	0.002191	0.002129	0.001742
K03695	0.001799	0.002039	0.002411	0.002077	0.001929	0.001981
K03046	0.001707	0.001904	0.002330	0.002100	0.002021	0.001997
K01153	0.001566	0.001646	0.002457	0.002176	0.002130	0.001684
K02483	0.002015	0.002174	0.001973	0.001737	0.001877	0.001799
K07714	0.001788	0.002156	0.002814	0.001824	0.001644	0.001330
K06158	0.001975	0.001799	0.001865	0.001965	0.002006	0.001853
K03696	0.001644	0.001973	0.002279	0.001955	0.001779	0.001821
K00257	0.002597	0.001931	0.001886	0.001573	0.001624	0.001735
K03407	0.001934	0.002095	0.002125	0.001807	0.001789	0.001528
K02006	0.001800	0.002006	0.001896	0.001822	0.001903	0.001796
K02337	0.001558	0.001758	0.002256	0.001914	0.001875	0.001715
K13599	0.001758	0.002008	0.002425	0.001764	0.001636	0.001435
K07713	0.001672	0.002028	0.002547	0.001803	0.001635	0.001295
K02519	0.001822	0.001681	0.001893	0.001884	0.001834	0.001779
K01362	0.001758	0.001690	0.001534	0.002021	0.001813	0.002031
K15738	0.001964	0.001676	0.001697	0.001825	0.001844	0.001789
K00666	0.002171	0.001831	0.001846	0.001598	0.001683	0.001494
K02031	0.002058	0.001685	0.001276	0.001677	0.001814	0.002102
K03043	0.001574	0.001616	0.002107	0.001793	0.001801	0.001705
K00128	0.002474	0.001565	0.000926	0.001514	0.001497	0.002269
K15726	0.001974	0.001716	0.001404	0.001854	0.001930	0.001336

KEGG profile	R_AR	R_ARO	R_ARDelta	L_LOin	L_LOout	S_Kal
K01995	0.001783	0.001729	0.001619	0.001568	0.001587	0.001873
K01895	0.001657	0.001647	0.001915	0.001627	0.001679	0.001619
K02028	0.001780	0.001666	0.001460	0.001562	0.001640	0.001994
K07715	0.001590	0.001834	0.002196	0.001660	0.001511	0.001252
K02355	0.001491	0.001678	0.002068	0.001630	0.001577	0.001529
K12266	0.001555	0.001814	0.002234	0.001642	0.001490	0.001221
K02010	0.001835	0.001600	0.001098	0.001602	0.001660	0.002155
K10943	0.001482	0.001750	0.002314	0.001631	0.001479	0.001208
K11085	0.001562	0.001656	0.001620	0.001680	0.001741	0.001574
K01955	0.001439	0.001531	0.002019	0.001634	0.001573	0.001633
K01996	0.001708	0.001555	0.001508	0.001527	0.001564	0.001760
K00382	0.001540	0.001520	0.001589	0.001622	0.001645	0.001595
K00123	0.001473	0.001646	0.002487	0.001305	0.001436	0.001159
K03723	0.001411	0.001481	0.001767	0.001648	0.001591	0.001579
K02482	0.001518	0.001647	0.002398	0.001386	0.001290	0.001171
K00627	0.001883	0.001566	0.001218	0.001557	0.001568	0.001555
K01338	0.001541	0.001534	0.002082	0.001430	0.001532	0.001196
K02469	0.001323	0.001465	0.001612	0.001671	0.001559	0.001655
K01952	0.001339	0.001393	0.001814	0.001596	0.001563	0.001524
K03798	0.001232	0.001543	0.001502	0.001655	0.001407	0.001726
K00265	0.001598	0.001404	0.001204	0.001647	0.001572	0.001604
K00525	0.001407	0.001309	0.001751	0.001447	0.001509	0.001512
K03086	0.001379	0.001423	0.001405	0.001631	0.001372	0.001639
K00548	0.001405	0.001624	0.001381	0.001478	0.001464	0.001444
K01130	0.001456	0.000729	0.000730	0.002189	0.001710	0.001967
K01870	0.001268	0.001347	0.001681	0.001469	0.001461	0.001429
K00626	0.001732	0.001394	0.001488	0.001335	0.001379	0.001309
K01873	0.001229	0.001355	0.001671	0.001467	0.001454	0.001444
K03529	0.001299	0.001467	0.001280	0.001602	0.001455	0.001508
K02056	0.001764	0.001419	0.000835	0.001254	0.001421	0.001871
K05366	0.001494	0.001356	0.001426	0.001365	0.001391	0.001527
K13590	0.001201	0.001425	0.001584	0.001471	0.001595	0.001243
K03070	0.001244	0.001309	0.001629	0.001498	0.001417	0.001389
K00249	0.001959	0.001520	0.001464	0.001082	0.001130	0.001319
K07486	0.000699	0.000535	0.002025	0.001829	0.001805	0.001530
K01999	0.001539	0.001611	0.001214	0.001045	0.001124	0.001651
K03797	0.001221	0.001337	0.001363	0.001484	0.001362	0.001266
K02470	0.001183	0.001169	0.001435	0.001408	0.001422	0.001376
K02052	0.001568	0.001294	0.000831	0.001222	0.001309	0.001740
K03168	0.001287	0.001212	0.001337	0.001377	0.001282	0.001461
K01953	0.001016	0.001200	0.001503	0.001503	0.001465	0.001250
K07814	0.001138	0.001556	0.001738	0.001214	0.001256	0.000996
K01537	0.000946	0.001401	0.001789	0.001253	0.001458	0.001018
K01869	0.001135	0.001246	0.001522	0.001312	0.001307	0.001302
K04043	0.001256	0.001320	0.001261	0.001346	0.001220	0.001416
K08300	0.001475	0.001229	0.001009	0.001398	0.001239	0.001440
K03694	0.001122	0.001225	0.001494	0.001344	0.001310	0.001272
K00615	0.001149	0.001303	0.001606	0.001238	0.001217	0.001237
K03721	0.001198	0.001426	0.001793	0.001261	0.001145	0.000923

Table S2.5: The fraction of unexplained taxonomic units per sample and averaged per habitat, as it was derived during the functional profiling using UProC KEGG Ortholog reference profiles in long read mode. FTU: fraction of unexplained taxonomic units. R: River. L: Lagoon. S: Sea. AR: Arachthos. ARO: Arachthos Neochori. ARDelta: Arachthos Delta. LOin: Logarou station inside the lagoon. LOout: Logarou station in the channel connecting the lagoon to the gulf. Kal: Kalamitsi. A, B, C: replicate samples.

Samples	FTU (%)	Average FTU per location (%)	Average FTU per habitat (%)
R_AR_A	54.68		
R_AR_B	51.49	50.93	
R_AR_C	46.61		
R_ARO_A	49.98		
R_ARO_B	47.68	48.29	56.85
R_ARO_C	47.21		
R_ARDelta_A	76.16		
R_ARDelta_B	69.64	71.32	
R_ARDelta_C	68.17		
L_LOin_A	67.68		
L_LOin_B	67.94	66.95	
L_LOin_C	65.22		67.51
L_LOout_A	75.18		
L_LOout_B	66.04	68.06	
L_LOout_C	62.97		
S_Kal_A	37.9		
S_Kal_B	39.92	38.91	38.91
S_Kal_C	38.92		

Table S2.6: The linear regression equations for the average number of OTUs per taxonomic group, as well as the average number of OTUs irrespective of taxonomic group, at the different salinities. *: $p < 0.05$. **: $p < 0.01$. n.s.: not significant. SW: p-value of Shapiro–Wilk statistic to test for normality in regression residuals (critical value (a = 0.05) for Shapiro–Wilk statistic when n = 6 is 0.7923).

	Taxonomic group	Linear equation	R^2		SW
	All OTUs	y = -13.322x + 2588.4	0.6719	*	0.3366
Archaea	*Crenarchaeota*	y = 0.3862x + 6.7355	0.552	n.s.	
	Euryarchaeota	y = 1.7602x + 42.142	0.3178	n.s.	
	Unclassified *Archaea*	y = 1.3505x + 24.152	0.4197	n.s.	
Bacteria	*Acidobacteria*	y = -1.1499x + 118.77	0.51	n.s.	
	Actinobacteria	y = -2.5228x + 127.28	0.8638	**	0.6748
	Bacteroidetes	y = -5.9149x + 483.59	0.7537	*	0.9983
	Chlorobi	y = -0.2292x + 19.104	0.7611	*	0.8545
	Chloroflexi	y = -1.2697x + 120.66	0.5828	n.s.	
	Cyanobacteria/Chloroplast	y = 0.283x + 23.533	0.1213	n.s.	
	Firmicutes	y = -0.4473x + 36.236	0.7097	*	0.7609
	OD1	y = -0.7155x + 40.82	0.8038	*	0.8457
	Planctomycetes	y = -0.2028x + 91.737	0.0124	n.s.	
	Proteobacteria	y = -6.9127x + 728.14	0.6514	n.s.	
	TM7	y = -0.7108x + 33.437	0.9814	**	0.8424
	Unclassified *Bacteria*	y = 3.824x + 509.41	0.2806	n.s.	
	Verrucomicrobia	y = -1.5924x + 100.18	0.8384	*	0.1759
	Other	y = 0.7423x + 82.471	0.0812	n.s.	
Bacteroidetes	*Bacteroidetes* incertae sedis	y = -0.6164x + 22.668	0.971	**	0.8067
	Bacteroidia	y = -0.3285x + 12.704	0.9297	**	0.6491
	Flavobacteria	y = -0.3244x + 74.333	0.1095	n.s.	
	Sphingobacteria	y = -1.5084x + 121.92	0.5637	n.s.	
	Unclassified *Bacteroidetes*	y = -3.1372x + 251.96	0.7839	*	0.6318
Proteobacteria	*Alphaproteobacteria*	y = -2.8399x + 223.18	0.655	n.s.	
	Betaproteobacteria	y = -2.8083x + 100.92	0.939	**	0.308
	Gammaproteobacteria	y = -0.4691x + 185	0.0823	n.s.	
	Deltaproteobacteria	y = -0.7239x + 190.73	0.2751	n.s.	
	Epsilonproteobacteria	y = -0.1367x + 11.211	0.7642	*	0.5321
	Unclassified *Proteobacteria*	y = 0.0651x + 17.105	0.0245	n.s.	

Supporting Material for Chapter 3

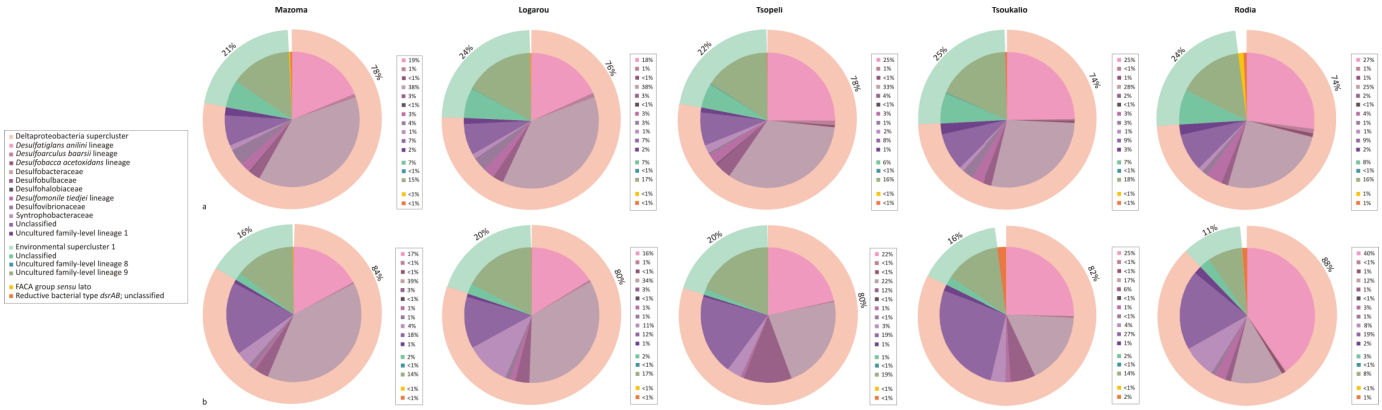

Figure S3.1: Taxonomic classification of the SRM OTUs of the different lagoons: a) based on the presence/absence of OTUs, b) based on the abundance of OTUs.

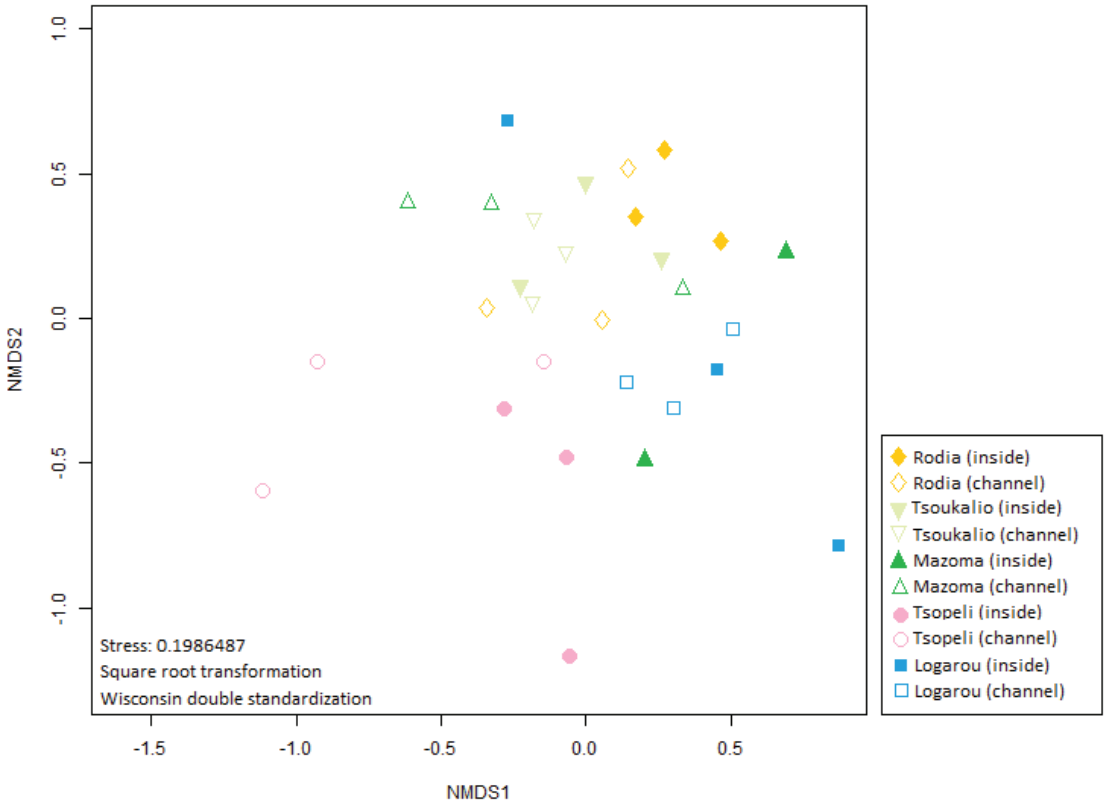

Figure S3.2: Multidimensional scaling of the SRM OTUs (90% similarity cut-off), based on the abundance of the environmental terms that they have been associated with. Data labels according to the location of the sampling station, as in Figure 3.4.

Table S3.1: The results of the processing of the sequences during the noise removal. M: Mazoma lagoon, L: Logarou lagoon, S: Tsopeli lagoon, T: Tsoukalio lagoon, R: Rodia lagoon. 01: Station inside the lagoons. 02: Station in the channel connecting each lagoon to the gulf. A, B, C: replicate samples. a: number of raw sequences. b: number of sequences after the first filtering. c: number of sequences after the removal of sequencing errors. d: number of clustered sequences after the removal of the PCR errors. e: number of chimeras. f: number of clustered sequences after the removal of chimeras (high quality sequences). Sequencing of the sample M_01_A was considered to be faulty, resulting in a high number of errors, thus this sample was not included in the further analyses.

A/A	Sample libraries	a	b	c	d	e	f
1	M_01_B	6,622	4,802	4,363	998	9	989
2	M_01_C	5,798	3,960	3,633	1,005	7	998
3	M_02_A	1,752	1,229	1,020	391	0	391
4	M_02_B	3,617	2,476	2,222	667	4	663
5	M_02_C	8,104	5,264	4,104	914	2	912
6	L_01_A	13,913	6,698	5,689	1,082	217	865
7	L_01_B	5,411	3,120	2,439	850	131	719
8	L_01_C	3,602	2,533	2,242	749	1	748
9	L_02_A	5,191	3,568	3,142	858	1	857
10	L_02_B	4,559	3,083	2,891	967	1	966
11	L_02_C	6,504	4,755	4,179	1,152	3	1,149
12	S_01_A	5,763	4,248	3,012	520	7	513
13	S_01_B	4,057	2,738	2,083	513	5	508
14	S_01_C	1,502	1,067	883	246	1	245
15	S_02_A	6,959	4,546	3,447	995	1	994
16	S_02_B	1,079	691	596	241	1	240
17	S_02_C	3,881	2,868	2,636	846	9	837
18	R_01_A	5,470	3,939	3,442	950	234	716
19	R_01_B	3,311	2,133	1,934	475	1	474
20	R_01_C	2,378	1,794	1,400	339	1	338
21	R_02_A	7,540	5,831	3,276	522	4	518
22	R_02_**B**	5,802	4,488	3,226	560	2	558
23	R_02_C	9,134	6,747	4,350	676	7	669
24	T_01_A	5,064	3,299	2,896	705	2	703
25	T_01_B	1,310	910	761	260	0	260
26	T_01_C	8,200	5,965	3,641	655	7	648
27	T_02_A	3,112	2,121	1,402	388	1	387
28	T_02_B	5,501	4,171	3,089	465	3	462
29	T_02_C	3,490	2,368	1,597	329	1	328
Sum		148,626	101,412	79,595	19,318	663	18,655

Table S3.2: Diversity indices of the samples, based on the 90% similarity cut-off. OTUs: total number of OTUs. N: total SRM relative abundance values. d: Margalef's species richness. J′: Pielou's evenness. H′: Shannon-Wiener. ACE: Abundance Coverage Estimator. M: Mazoma lagoon, L: Logarou lagoon, S: Tsopeli lagoon, T: Tsoukalio lagoon, R: Rodia lagoon. 01: Station inside the lagoons. 02: Station in the channel connecting each lagoon to the gulf. A, B, C: replicate samples.

	OTUs	N	d	J′	H′(log_2)	Chao-1	ACE
M_01_B	469	3,628	57.1	0.8391	7.446	666,1667	656,9447
M_01_C	418	2,585	53.07	0.8249	7.182	608,987	639,5986
M_02_A	257	1,100	36.56	0.8362	6.694	380,0167	436,1893
M_02_B	398	2,054	52.05	0.8535	7.372	617,4925	634,6642
M_02_C	449	4,266	53.6	0.7756	6.834	588,0822	586,8483
L_01_A	383	5,148	44.7	0.8636	7.411	446,8936	441,4868
L_01_B	387	2,434	49.5	0.8561	7.359	576,4068	565,9593
L_01_C	408	2,016	53.49	0.862	7.476	621,1304	612,1865
L_02_A	397	2,595	50.37	0.8149	7.035	611,7619	608,364
L_02_B	422	1,877	55.85	0.8375	7.304	775,9516	731,6545
L_02_C	475	3,461	58.16	0.7638	6.791	698,875	723,8786
S_01_A	311	3,589	37.87	0.7736	6.406	466,1163	463,5426
S_01_B	238	2,153	30.88	0.7323	5.782	325,6923	367,0418
S_01_C	149	926	21.67	0.8404	6.067	195,1613	214,0853
S_02_A	411	3,403	50.42	0.705	6.121	541,679	557,0556
S_02_B	160	619	24.74	0.7355	5.386	277	317,8009
S_02_C	382	1,986	50.17	0.797	6.836	692,082	705,5519
R_01_A	381	3,192	47.1	0.8076	6.924	739,6829	637,4184
R_01_B	239	1,709	31.97	0.8167	6.452	346,8	374,3831
R_01_C	224	1,702	29.97	0.7763	6.061	339,1765	336,6552
R_02_A	312	5,518	36.1	0.676	5.601	430,1458	438,677
R_02_B	300	4,196	35.84	0.7313	6.018	386,8421	410,2137
R_02_C	351	6,194	40.09	0.7025	5.94	488,3077	481,892
T_01_A	311	2,404	39.82	0.7926	6.564	462,25	458,5118
T_01_B	171	812	25.38	0.8046	5.968	317,25	320,797
T_01_C	366	5,426	42.45	0.6938	5.908	491,5091	490,3717
T_02_A	223	1,943	29.32	0.7609	5.935	338,1613	311,9816
T_02_B	292	3,948	35.14	0.7944	6.506	410,825	410,6931
T_02_C	226	2,204	29.23	0.7205	5.634	313,1277	356,1822

Table S3.3: The Chi-Square values of the Kruskal-Wallis tests for the diversity indices between the salinity categories and the lagoons. OTUs: total number of OTUs. N: total SRM relative abundance values. d: Margalef's species richness. J′: Pielou's evenness. H′: Shannon-Wiener. ACE: Abundance Coverage Estimator. *: $p < 0.05$. **: $p < 0.01$. n.s.: not significant.

	OTUs	N	d	J′	H′(log_2)	Chao-1	ACE
Salinity categories	9.099 *	n.s.	9.814 **	6.423 *	8.363 *	7.351 *	8.630 *
Lagoon	14.697 **	n.s.	15.326 **	11.913 *	17.694 **	10.128 *	11.103 *

Table S3.4: The values of the Mann-Whitney U tests used for the *post-hoc* pairwise significant comparisons, after the Bonferroni correction. OTUs: total number of OTUs. N: total SRM relative abundance values. d: Margalef's species richness. J′: Pielou's evenness. H′: Shannon-Wiener. ACE: Abundance Coverage Estimator. *: $p < 0.17$ in the case of the salinity categories. *: $p < 0.01$ in the case of the lagoons. n.s.: not significant.

		OTUs	N	d	J′	H′(log_2)	Chao-1	ACE
Salinity categories	(Mixo-) a-mesohaline *vs* (Mixo-) b-mesohaline	n.s.	n.s.	n.s.	n.s.	37.000 *	n.s.	n.s.
	(Mixo-) a-mesohaline *vs* (Mixo-) polyhaline	0 *	n.s.	0 *	n.s.	1 *	n.s.	1 *
Lagoons	Mazoma *vs* Tsoukalio	n.s.	n.s.	n.s.	n.s.	0 *	n.s.	n.s.
	Tsopeli *vs* Logarou	n.s.	n.s.	n.s.	n.s.	1 *	n.s.	n.s.
	Rodia *vs* Logarou	0 *	n.s.	1 *	n.s.	1 *	n.s.	n.s.
	Tsoukalio *vs* Logarou	0 *	n.s.	0 *	n.s.	0 *	2 *	2 *

Table S3.5: The values of the Mann-Whitney U tests for the diversity indices between the locations. OTUs: total number of OTUs. N: total SRM relative abundance values. d: Margalef's species richness. J′: Pielou's evenness. H′: Shannon-Wiener. ACE: Abundance Coverage Estimator. *: $p < 0.05$. n.s.: not significant.

	OTUs	N	d	J′	H′(log_2)	Chao-1	ACE
Location	n.s.	n.s.	n.s.	59.000 *	n.s.	n.s.	n.s.

Supporting Material for Chapter 4

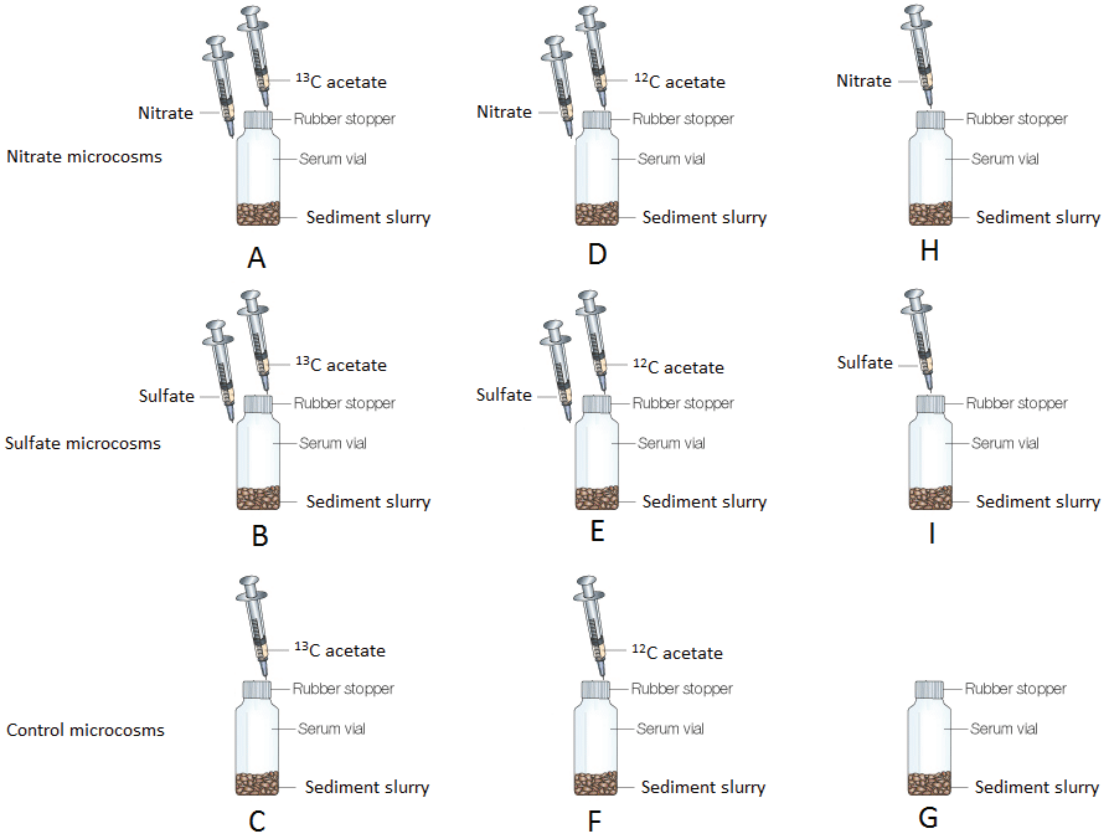

Figure S4.1: Schematic image explaining the SIP experiment (adapted from Dumont and Murrell (2005)). Codes according to Table 4.3.

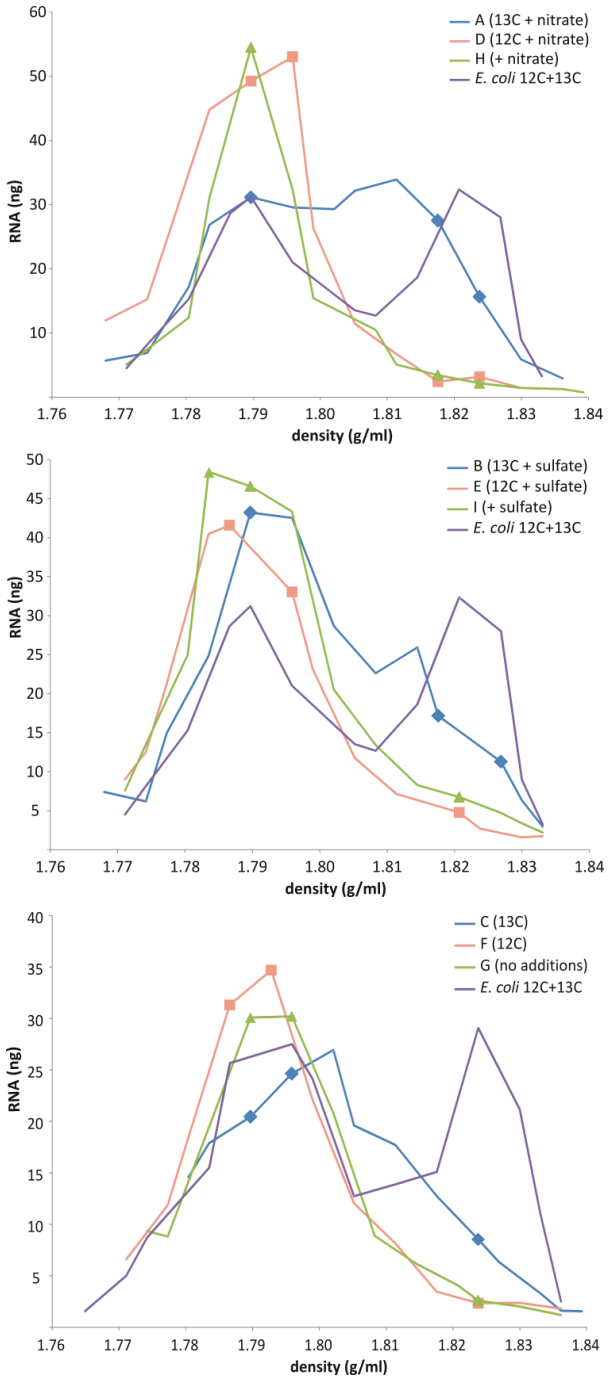

Figure S4.2: Fractionation plots of RNA-SIP experiments for station 18. Plots depict the RNA concentration (ng/µl) *vs* buoyant density (g/ml) for each fraction after ultracentrifugation. RNA from fractions denoted by a symbol was used for 16S rRNA sequencing. For each set, a parallel tube containing *E. coli* RNA was centrifuged and used as marker for identifying "heavy" gradient fraction (~ 1.82 g/ml) and "light" gradient fraction (~ 1.79 g/ml). Top graph: Incubations with nitrate as an electron acceptor. Middle graph: Incubations with sulfate as an electron acceptor. Bottom graph: Control incubations with no electron acceptor addition.

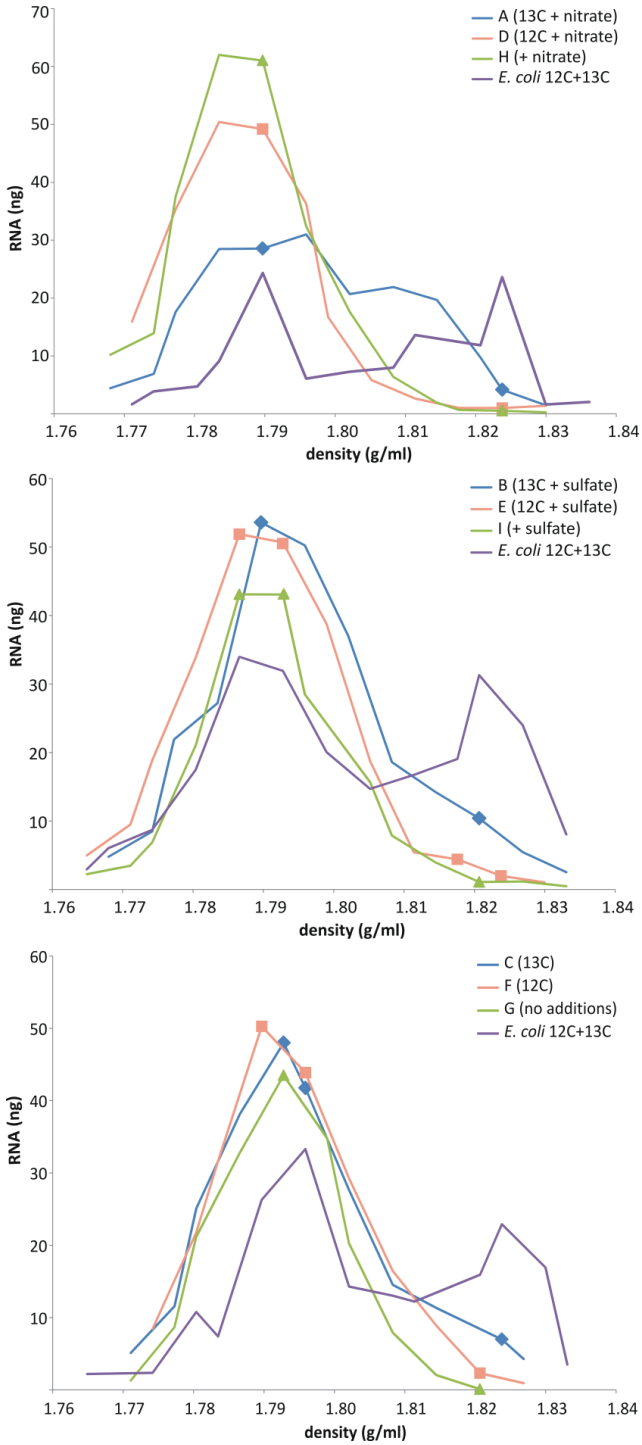

Figure S4.3: Fractionation plots of RNA-SIP experiments for station 34. Plots depict the RNA concentration (ng/µl) *vs* buoyant density (g/ml) for each fraction after ultracentrifugation. RNA from fractions denoted by a symbol was used for 16S rRNA sequencing. For each set, a parallel tube containing *E. coli* RNA was centrifuged and used as marker for identifying "heavy" gradient fraction (~ 1.82 g/ml) and "light" gradient fraction (~ 1.79 g/ml). Top graph: Incubations with nitrate as an electron acceptor. Middle graph: Incubations with sulfate as an electron acceptor. Bottom graph: Control incubations with no electron acceptor addition.

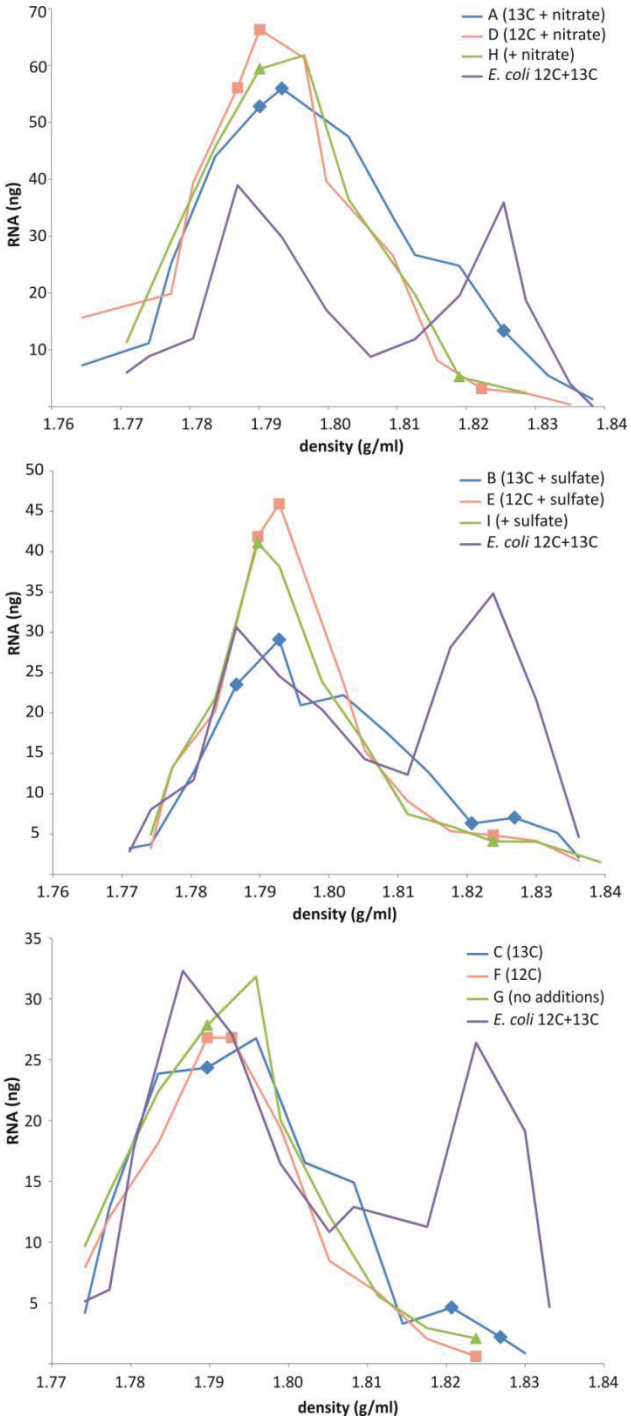

Figure S4.4: Fractionation plots of RNA-SIP experiments for station 39. Plots depict the RNA concentration (ng/µl) *vs* buoyant density (g/ml) for each fraction after ultracentrifugation. RNA from fractions denoted by a symbol was used for 16S rRNA sequencing. For each set, a parallel tube containing *E. coli* RNA was centrifuged and used as marker for identifying "heavy" gradient fraction (~ 1.82 g/ml) and "light" gradient fraction (~ 1.79 g/ml). Top graph: Incubations with nitrate as an electron acceptor. Middle graph: Incubations with sulfate as an electron acceptor. Bottom graph: Control incubations with no electron acceptor addition.

Figure S4.5: The relative abundances (%) of the main microbial taxonomic groups of the "heavy" and "light" fractions of the SIP experiment at station 18. Blue frame: samples with addition of ^{13}C labelled acetate. Pink frame: samples with addition of ^{12}C acetate. Green frame: samples with no electron donor addition. Top graph: Incubations with nitrate as an electron acceptor. Middle graph: Incubations with sulfate as an electron acceptor. Bottom graph: Control incubations with no electron acceptor addition.

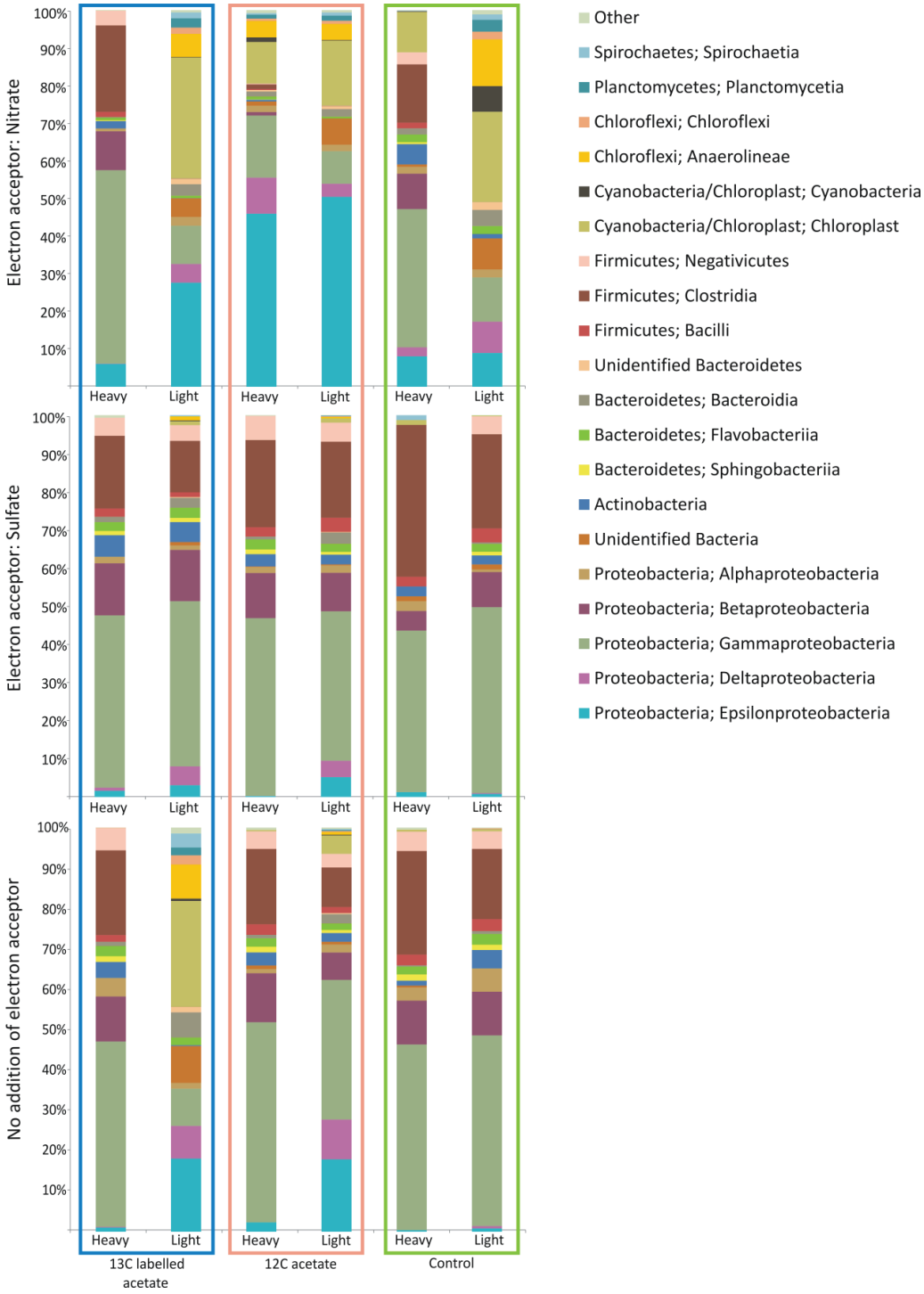

Figure S4.6: The relative abundances (%) of the main microbial taxonomic groups of the "heavy" and "light" fractions of the SIP experiment at station 34. Blue frame: samples with addition of ^{13}C labelled acetate. Pink frame: samples with addition of ^{12}C acetate. Green frame: samples with no electron donor addition. Top graph: Incubations with nitrate as an electron acceptor. Middle graph: Incubations with sulfate as an electron acceptor. Bottom graph: Control incubations with no electron acceptor addition.

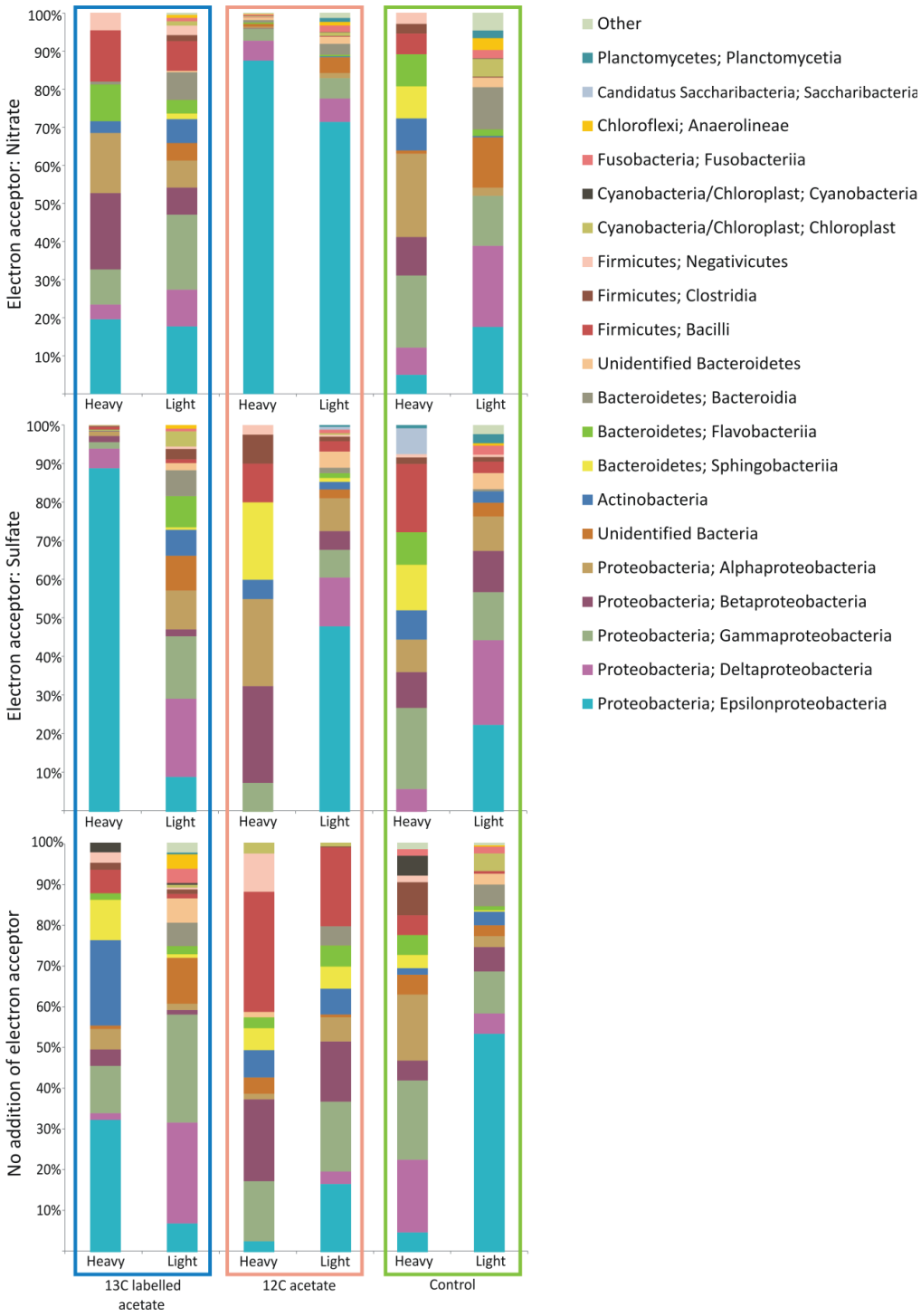

Figure S4.7: The relative abundances (%) of the main microbial taxonomic groups of the "heavy" and "light" fractions of the SIP experiment at station 39. Blue frame: samples with addition of ^{13}C labelled acetate. Pink frame: samples with addition of ^{12}C acetate. Green frame: samples with no electron donor addition. Top graph: Incubations with nitrate as an electron acceptor. Middle graph: Incubations with sulfate as an electron acceptor. Bottom graph: Control incubations with no electron acceptor addition.

Table S4.1: Solutions used for nucleic acid extractions and ultracentrifugation. All solutions were prepared with DEPC treated water, in baked glassware (3 h, 180 °C).

Solutions	Ingredients
120 mM $NaPO_4$ buffer	112.88 mM Na_2HPO_4, 7.12 mM NaH_2PO_4
TNS solution	500 mM Tris-HCl pH 8, 100 mM NaCl, 10 % SDS (w/v)
Polyethylene Glycol (PEG)	30 % (w/v) polyethylene glycol 6000 in 1.6 M NaCl
Gradient buffer (GB)	0.1 M Tris-HCl pH 8, 0.1 M KCl, 1 mM EDTA

Table S4.2: Number of sequences per sample after every processing step, as provided by the IMNGS platform. SIP: stable isotope probing. a, b, c: replicate samples. A, B, C, D, E, F, G, H, I: sample code according to the SIP experiment employed (as in Table 4.3)[12].

Station	Primer pair used for the first-step PCR	Demultiplexing	Merging of paired end reads	Expected error filtering	Chimeras/Artifacts filtering	Abundance of OTUs filtering
1	F1 - R1	5177	1783	765	580	357
2	F1 - R1	7514	2356	872	628	346
4	F1 - R1	9907	3257	1315	963	451
5_a	F1 - R1	159921	62142	23497	16113	5779
5_b	F2 - R2	1576	134	48	38	38
5_c	F2 - R1	983	131	40	28	28
7	F1 - R1	30750	9731	3933	2640	1100
8_a	F1 - R1	328213	106987	42525	20325	7636
8_b	F2 - R2	545	47	19	15	15
9_a	F1 - R1	4175	1529	599	482	245
9_b	F2 - R2	44	8	6	2	2
10_a	F1 - R1	210437	78386	30176	18183	5912
10_b	F2 - R2	448	54	32	20	20
10_c	F2 - R1	1786	240	73	50	50
14_a	F1 - R1	347336	114982	46853	26251	10699
14_b	F2 - R2	758	96	34	18	18
18_a	F1 - R1	60573	22697	8235	4971	2216
18_b	F2 - R2	980	102	26	16	16
20_a	F1 - R1	88865	33058	12466	7180	2792
20_b	F2 - R2	1466	180	64	46	46
20_c	F1 - R1	15494	5031	2184	1481	661

[12] F1: forward primer sequence
(5′ - TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG - 3′).
R1: reverse primer sequence
(5′ - GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACNVGGGTATCTAATCC - 3′).
F2: forward primer sequence including an extra 5 bp barcode
(5′ -TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCTGCCTACGGGNGGCWGCAG - 3′).
R2: reverse primer sequence including an extra 5 bp barcode
(5′ -GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTAGGACTACNVGGGTATCTAATCC - 3′).

Station	Primer pair used for the first-step PCR	Demultiplexing	Merging of paired end reads	Expected error filtering	Chimeras/Artifacts filtering	Abundance of OTUs filtering
34_a	F1 - R1	84710	30105	10650	6603	3140
34_b	F2 - R2	1026	134	50	34	34
36_a	F1 - R1	35457	12271	4663	3068	1287
36_b	F2 - R2	424	34	18	18	18
39_a	F1 - R1	198059	70860	23690	13356	5078
39_b	F2 - R2	494	64	14	10	10
39_c	F2 - R1	2011	206	78	54	54
43_a	F1 - R1	64277	22572	9160	6102	2034
43_b	F2 - R2	240	30	4	4	4
45_a	F1 - R1	22851	8650	3082	2114	866
45_b	F2 - R2	458	52	14	10	10
46_a	F1 - R1	159996	60249	22994	16881	8404
46_b	F2 - R2	752	66	24	18	18
47_a	F1 - R1	182389	67790	27012	14938	5147
47_b	F2 - R2	958	112	40	34	34
47_c	F2 - R1	1714	174	49	23	23
48_a	F1 - R1	56599	21649	8014	5420	2108
48_b	F2 - R2	351	42	16	12	12
53_a	F1 - R1	123069	45561	18398	14412	7712
53_b	F2 - R2	26	0	0	0	0
59_a	F1 - R1	46821	16853	6113	3907	1669
59_b	F2 - R2	828	78	18	14	14
2281_a	F1 - R1	474	56	18	18	18
2281_b	F2 - R2	252239	93277	34313	22605	8485
2281_c	F2 - R1	1683	192	56	28	28
2282_a	F1 - R1	211415	74237	28128	16925	5332
2282_b	F2 - R2	970	82	18	14	14
2282_c	F2 - R1	1321	140	46	32	32
2285_a	F1 - R1	72444	24485	9011	5433	2087
2285_b	F2 - R2	1352	80	20	16	16
2287	F1 - R1	11650	3663	1486	1254	742
2289_a	F1 - R1	294335	106725	45676	28179	11247
2289_b	F2 - R2	308	22	14	4	4
2291	F1 - R1	6332	2201	820	611	303
18_A4	F1 - R1	8921	3575	1716	1648	1584
18_A5	F1 - R1	4370	1603	763	738	719
18_A10	F1 - R1	11139	4250	1725	1551	1337
18_B3	F1 - R1	7555	2803	1145	1083	1017
18_B4	F1 - R1	8553	3552	1629	1565	1522
18_B9	F1 - R1	41824	15410	6541	2776	1123
18_C5	F1 - R1	4231	1569	657	630	596
18_C10	F1 - R1	3768	1382	643	594	535
18_C11	F1 - R1	3228	1250	515	470	396
18_D3	F1 - R1	24233	9629	3994	3781	3682

Station	Primer pair used for the first-step PCR	Demultiplexing	Merging of paired end reads	Expected error filtering	Chimeras/Artifacts filtering	Abundance of OTUs filtering
18_D4	F1 - R1	20205	7895	3232	3049	2949
18_D8	F1 - R1	29991	12064	5655	4978	4082
18_D9	F1 - R1	40741	15797	8085	6546	5527
18_E4	F1 - R1	13516	5326	2343	2272	2205
18_E8	F1 - R1	60874	22089	10370	8018	6121
18_E9	F1 - R1	16341	6236	2655	2411	2263
18_F3	F1 - R1	287	29	12	9	9
18_F8	F1 - R1	281	23	5	2	2
18_F9	F1 - R1	270	25	11	6	6
18_G3	F1 - R1	2790	1042	407	393	383
18_G8	F1 - R1	13479	4844	2173	1809	1379
18_G9	F1 - R1	10805	4192	1739	1601	1265
18_H4	F1 - R1	9722	3776	1664	1564	1510
18_H5	F1 - R1	8364	3362	1512	1447	1403
18_H10	F1 - R1	76397	29007	12524	5862	2208
18_I4	F1 - R1	3503	1346	587	569	552
18_I9	F1 - R1	7076	2693	1025	943	817
18_I10	F1 - R1	6030	2274	853	724	509
34_A3	F1 - R1	53063	20452	8611	8161	7924
34_A9	F1 - R1	111631	41315	17528	8644	6070
34_B4	F1 - R1	30247	12193	4811	4596	4443
34_B9	F1 - R1	33681	12965	5309	4960	4327
34_C4	F1 - R1	16138	6411	2832	2657	2601
34_C8	F1 - R1	46607	16540	7453	3161	1892
34_C9	F1 - R1	73484	26881	12409	5118	3207
34_D3	F1 - R1	46678	18531	8280	6391	4974
34_D9	F1 - R1	104606	40306	19560	11731	9186
34_E3	F1 - R1	107618	41905	14023	13427	12911
34_E4	F1 - R1	30817	11993	5103	4797	4610
34_E8	F1 - R1	38628	15042	6322	5943	5524
34_E9	F1 - R1	44984	17086	7021	6328	5642
34_F4	F1 - R1	26595	10276	4275	4090	3905
34_F8	F1 - R1	29727	11914	5109	4911	4795
34_F9	F1 - R1	33827	13651	6160	4634	3664
34_G4	F1 - R1	21047	8396	3502	3334	3282
34_G9	F1 - R1	38426	15384	6988	6653	6125
34_H3	F1 - R1	34576	13622	5844	5545	5039
34_H9	F1 - R1	113731	38909	15059	6520	3557
34_I4	F1 - R1	608	205	100	78	78
34_I9	F1 - R1	17819	6940	2931	2772	2721
34_I10	F1 - R1	2467	960	405	372	370
39_A3	F1 - R1	17239	6870	2748	2632	2587
39_A8	F1 - R1	6916	2626	1108	1060	997
39_A9	F1 - R1	8257	3229	1297	1241	1172

Station	Primer pair used for the first-step PCR	Demultiplexing	Merging of paired end reads	Expected error filtering	Chimeras/Artifacts filtering	Abundance of OTUs filtering
39_B3	F1 - R1	5167	1969	873	849	827
39_B4	F1 - R1	17214	7018	3220	3028	2911
39_B9	F1 - R1	6136	2380	1005	936	877
39_B10	F1 - R1	2823	1078	423	368	318
39_C3	F1 - R1	2671	997	342	325	321
39_C4	F1 - R1	2426	944	418	406	389
39_C9	F1 - R1	8529	3321	1421	1247	946
39_D4	F1 - R1	14684	6003	2650	2429	2221
39_D9	F1 - R1	28881	11266	5112	4152	3632
39_D10	F1 - R1	48920	19741	9769	6617	4681
39_E3	F1 - R1	3891	1549	705	684	664
39_E8	F1 - R1	5141	1981	825	801	761
39_E9	F1 - R1	4881	1879	889	830	783
39_F3	F1 - R1	7449	2832	1200	1151	1126
39_F8	F1 - R1	5303	1988	826	778	738
39_F9	F1 - R1	4721	1747	700	667	646
39_G3	F1 - R1	3383	1311	523	505	484
39_G9	F1 - R1	4595	1811	809	744	660
39_H3	F1 - R1	15547	6372	2803	2654	2593
39_H7	F1 - R1	45930	17636	6985	4352	2536
39_I3	F1 - R1	6765	2542	1118	1080	1032
39_I9	F1 - R1	4308	1627	718	666	577
Blank_18_SIP	F1 - R1	7044	2628	1104	1025	971
Blank_34_SIP	F1 - R1	14	2	0	0	0
Blank_39_SIP	F1 - R1	3125	1109	494	476	463
Blank_F1-R1	F1 - R1	33367	12182	5370	5004	4748
Blank_F2-R1	F2 - R1	22	4	0	0	0
Blank_F2-R2	F2 - R2	1889	709	219	209	203

Table S4.3: Diversity indices of the samples. OTUs: total number of OTUs. N: total microbial relative abundance values. H′: Shannon-Wiener. J′: Pielou's evenness. d: Margalef's species richness. λ: Simpson.

Samples	OTUs	N	H′(ln)	J′	d	λ
1	89	311	4.154	0.925	15.332	0.977
2	98	309	4.154	0.906	16.919	0.975
4	114	405	4.209	0.889	18.821	0.973
5	242	1755	4.509	0.821	32.261	0.979
7	164	961	4.443	0.871	23.733	0.980
8	198	3545	3.907	0.739	24.103	0.960
9	61	111	3.843	0.935	12.740	0.972
10	179	1915.67	4.207	0.811	23.552	0.974
14	237	4847	4.280	0.783	27.810	0.972
18	198	968	4.378	0.828	28.654	0.976
20	229	1062	4.560	0.839	32.721	0.982
34	171	1473	3.789	0.737	23.303	0.953
36	164	578.5	4.413	0.865	25.627	0.978
39	233	1484	4.546	0.834	31.770	0.980
43	168	848	4.530	0.884	24.767	0.984
45	136	397	4.376	0.891	22.560	0.981
46	220	3225.5	4.359	0.808	27.108	0.976
47	258	1551.33	4.589	0.826	34.981	0.981
48	189	989	4.537	0.866	27.259	0.983
53	182	4880	4.409	0.847	21.312	0.977
59	183	702.5	4.503	0.864	27.767	0.979
2281	236	2497.33	4.296	0.786	30.040	0.972
2282	218	1499.33	4.292	0.797	29.674	0.968
2285	159	859	4.182	0.825	23.387	0.965
2287	113	620	4.135	0.875	17.419	0.972
2289	242	4807	4.501	0.820	28.427	0.978
2291	86	285	3.916	0.879	15.038	0.968

Table S4.4: Statistical significant Pearson correlations between the diversity indices of the samples and the physicochemical variables of the sampling stations. N: total microbial relative abundance values. H′: Shannon-Wiener. J′: Pielou's evenness. d: Margalef's species richness. λ: Simpson. *: $p < 0.05$. **: $p < 0.01$

		Physicochemical variables				
		Total Nitrogen (%)	Liquid Water Content (%)	delta ^{15}N	Salinity (psu)	Total Organic Carbon (%)
Diversity indices	N				- 0.47 *	
	H′			- 0.59 **		
	J′	- 0.39 *	- 0.45 *			
	d			- 0.42 *		
	λ	- 0.40 *				- 0.38 *

Table S4.5: Pair-wise statistical comparisons (PERMANOVA) of microbial communities among the different tested factors. A: zonal transect at 28° 38′S. B: cross shelf transect starting at 25°S. C: stations between 23°S and 25°S. D: zonal transect at 23°S. E: stations between 19°S and 21°S. F: zonal transect off Kunene mouth at 17° 15′S. n. sign.: non significant. *: $p < 0.05$. **: $p < 0.01$

Transect		Depth range		Range of oxygen concentration	
Groups	t	Groups	t	Groups	t
A, B	n. sign.	0-100 m, 100-200 m	n. sign.	0-10 µmol/l, 10-30 µmol/l	n. sign.
A, C	1.7566 *	0-100 m, 200-500 m	1.637 **	0-10 µmol/l, 30-50 µmol/l	n. sign.
A, D	n. sign.	0-100 m, 500-1000 m	n. sign	0-10 µmol/l, 50-100 µmol/l	1.8887 *
A, E	1.4518 *	0-100 m, 1000-2500 m	1.8295 *	0-10 µmol/l, 100-150 µmol/l	2.5629 *
A, F	n. sign.	100-200 m, 200-500 m	n. sign	0-10 µmol/l, 150-200 µmol/l	2.6073 *
B, C	n. sign.	100-200 m, 500-1000 m	n. sign	10-30 µmol/l, 30-50 µmol/l	n. sign.
B, D	n. sign.	100-200 m, 1000-2500 m	2.1683 *	10-30 µmol/l, 50-100 µmol/l	1.4369 *
B, E	n. sign.	200-500 m, 500-1000 m	1.4594 *	10-30 µmol/l, 100-150 µmol/l	1.9505 **
B, F	n. sign.	200-500 m, 1000-2500 m	2.491 **	10-30 µmol/l, 150-200 µmol/l	2.4442 **
C, D	n. sign.	500-1000 m, 1000-2500 m	n. sign	30-50 µmol/l, 50-100 µmol/l	n. sign.
C, E	n. sign.			30-50 µmol/l, 100-150 µmol/l	1.2358 *
C, F	n. sign.			30-50 µmol/l, 150-200 µmol/l	1.6836 *

References

Abell, G.C.J.J., Banks, J., Ross, D.J., Keane, J.P., Robert, S.S., Revill, A.T., and Volkman, J.K. (2011) Effects of estuarine sediment hypoxia on nitrogen fluxes and ammonia oxidizer gene transcription. *FEMS Microbiol. Ecol.* **75**: 111–122.

Achtnich, C., Bak, F., and Conrad, R. (1995) Competition for electron donors among nitrate reducers, ferric iron reducers, sulfate reducers, and methanogens in anoxic paddy soil. *Biol. Fertil. Soils* **19**: 65–72.

Afiahayati, Sato, K., and Sakakibara, Y. (2014) MetaVelvet-SL: an extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. *DNA Res*. **22**: 69–77.

Ahn, Y.-B., Chae, J.-C., Zylstra, G.J., and Häggblom, M.M. (2009) Degradation of phenol via phenylphosphate and carboxylation to 4-hydroxybenzoate by a newly isolated strain of the sulfate-reducing bacterium Desulfobacterium anilini. *Appl. Environ. Microbiol.* **75**: 4248–4253.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Mol. Biol. Rev.* **59**: 143–169.

Amore, R.D., Ijaz, U.Z., Schirmer, M., Kenny, J., Gregory, R., Darby, A.C., et al. (2016) A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics* **17**: 55.

Anastasakis, G., Piper, D.J.W., and Tziavos, C. (2007) Sedimentological response to neotectonics and sea-level change in a delta-fed, complex graben: Gulf of Amvrakikos, western Greece. *Mar. Geol.* **236**: 27–44.

Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**: 32–46.

Angermeyer, A., Crosby, S.C., and Huber, J.A. (2016) Decoupled distance-decay patterns between dsrA and 16S rRNA genes among salt marsh sulfate-reducing bacteria. *Environ. Microbiol.* **18**: 75–86.

Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., et al. (2005) PHACCS, an online tool for estimating the structure and diversity of

uncultured viral communities using metagenomic information. *BMC Bioinformatics* **6**: 41.

Angly, F.E., Dennis, P.G., Skarshewski, A., Vanwonterghem, I., Hugenholtz, P., and Tyson, G.W. (2014) CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* **2**: 11.

Aoyagi, T., Hanada, S., Itoh, H., Sato, Y., Ogata, A., Friedrich, M.W., et al. (2015) Ultra-high-sensitivity stable-isotope probing of rRNA by high-throughput sequencing of isopycnic centrifugation gradients. *Environ. Microbiol. Rep.* **7**: 282–287.

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **75**: 129–137.

Arrigo, K.R. (2005) Marine microorganisms and global nutrient cycles. *Nature* **437**: 349–355.

Arvanitidis, C., Koutsoubas, D., Dounas, C., and Eleftheriou, A. (1999) Annelid fauna of a Mediterranean lagoon (Gialova Lagoon, south-west Greece): community structure in a severely fluctuating environment. *J. Mar. Biol. Assoc. UK* **79**: 849–856.

Aßhauer, K.P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015) Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* **31**: 2882–2884.

Attrill, M.J. (2002) A testable linear model for diversity trends in estuaries. *J. Anim. Ecol.* **71**: 262–269.

Avramidis, P., Iliopoulos, G., Panagiotaras, D., Papoulis, D., Lambropoulou, P., Kontopoulos, N., et al. (2013) Tracking Mid- to Late Holocene depositional environments by applying sedimentological, palaeontological and geochemical proxies, Amvrakikos coastal lagoon sediments, Western Greece, Mediterranean Sea. *Quat. Int.* 1–18.

Bagarinao, T. (1992) Sulfide as an environmental factor and toxicant: tolerance and adaptions in aquatic organisms. *Aquat. Toxicol.* **24**: 21–62.

Bakker, M.G., Tu, Z.J., Bradeen, J.M., and Kinkel, L.L. (2012) Implications of Pyrosequencing Error Correction for Biological Data Interpretation. *PLoS One* **7**:

e44357.

Bakun, A., Field, D.B., Redondo-Rodriguez, A.N.A., and Weeks, S.J. (2010) Greenhouse gas, upwelling-favorable winds, and the future of coastal ocean upwelling ecosystems. *Glob. Chang. Biol.* **16**: 1213–1228.

Balzer, S., Malde, K., Grohme, M.A., and Jonassen, I. (2013) Filtering duplicate reads from 454 pyrosequencing data. *Bioinformatics* **29**: 830–836.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**: 455–477.

Barcina, I., Lebaron, P., and Vives-Rego, J. (2006) Survival of allochthonous bacteria in aquatic systems: a biological approach. *FEMS Microbiol. Ecol.* **23**: 1–9.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014) jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**: 293.

Barnes, R.S.K. (1980) Coastal Lagoons Cambridge University Press, Cambridge, UK.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000) The Pfam protein families database. *Nucleic Acids Res* **28**: 263–266.

Baytshtok, V., Lu, H., Park, H., Kim, S., Yu, R., and Chandran, K. (2009) Impact of varying electron donors on the molecular microbial ecology and biokinetics of methylotrophic denitrifying bacteria. *Biotechnol. Bioeng.* **102**: 1527–1536.

Beck, D., Settles, M., and Foster, J.A. (2011) OTUbase: an R infrastructure package for operational taxonomic unit data. *Bioinformatics* **27**: 1700–1701.

Bellan, G. (1972) Effects of an artificial stream on marine communities. *Mar. Pollut. Bull.* **3**: 74–78.

Ben-Dov, E., Brenner, A., and Kushmaro, A. (2007) Quantification of sulfate-reducing bacteria in industrial wastewater, by real-time polymerase chain reaction (PCR) using dsrA and apsA genes. *Microb. Ecol.* **54**: 439–451.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**: 289–300.

Benlloch, S., Lopez-Lopez, A., Casamayor, E.O., Ovreas, L., Goddard, V., Daae, F.L., et al. (2002) Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ. Microbiol.* **4**: 349–360.

Benlloch, S., Rodríguez-Valera, F., and Martinez-Murcia, A.J. (1995) Bacterial diversity in two coastal lagoons deduced from 16S rDNA PCR amplification and partial sequencing. *FEMS Microbiol. Ecol.* **18**: 267–279.

Benson, C.A., Bizzoco, R.W., Lipson, D.A., and Kelley, S.T. (2011) Microbial diversity in nonsulfur, sulfur and iron geothermal steam vents. *FEMS Microbiol. Ecol.* **76**: 74–88.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–59.

Bentzen, G., Smit, A., Bennett, D., Webster, N., Reinholt, F., Sletholt, E., and Hobsont, J. (1995) Controlled dosing of nitrate for prevention of HS in a sewer network and the effects on the subsequent treatment processes. *Water Sci. Technol.* **31**: 293–302.

Berger, S.A. and Stamatakis, A. (2011) Aligning short reads to reference alignments and trees. *Bioinformatics* **27**: 2068–2075.

Berner, R.A. (1980) Early diagenesis: a theoretical approach. Princeton University Press.

Bernhard, A. (2010) The Nitrogen Cycle: Processes, Players, and Human Impact. *Nat. Educ. Knowl.* **3**: 25.

de Bie, M.J.M., Speksnijder, A.G.C.L., Kowalchuk, G.A., Schuurman, T., Zwart, G., Stephen, J.R., et al. (2001) Shifts in the dominant populations of ammonia-oxidizing ß subclass Proteobacteria along the eutrophic Schelde estuary. *Aquat. Microb. Ecol.* **23**: 225–236.

Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C., and Hugenholtz, P. (2007) CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**: 209.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* **89**: 19.10:19.10.1–19.10.21.

Blümel, M., Süling, J., and Imhoff, J.F. (2007) Depth-specific distribution of Bacteroidetes in the oligotrophic Eastern Mediterranean Sea. *Aquat. Microb.*

Ecol. **46**: 209–224.

Bobrova, O., Kristoffersen, J.B., Oulas, A., and Ivanytsia, V. (2016) Metagenomic 16s rRNA investigation of microbial communities in the Black Sea estuaries in South-West of Ukraine. *Acta Biochim. Pol.* **63**: 315–319.

Bolhuis, H., Fillinger, L., and Stal, L.J. (2013) Coastal microbial mat diversity along a natural salinity gradient. *PLoS One* **8**: e63166.

Bolhuis, H. and Stal, L.J. (2011) Analysis of bacterial and archaeal diversity in coastal microbial mats using massive parallel 16S rRNA gene tag sequencing. *ISME J.* **5**: 1701–1712.

Bouvier, T.C. and del Giorgio, P.A. (2002) Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol. Oceanogr.* **47**: 453–470.

Bowen, J.L., Byrnes, J.E.K., Weisman, D., and Colaneri, C. (2013) Functional gene pyrosequencing and network analysis: an approach to examine the response of denitrifying bacteria to increased nitrogen supply in salt marsh sediments. *Front. Microbiol.* **4**: 1–12.

Bradford, M.A., Davies, C.A., Frey, S.D., Maddox, T.R., Melillo, J.M., Mohan, J.E., et al. (2008) Thermal adaptation of soil microbial respiration to elevated temperature. *Ecol. Lett.* **11**: 1316–1327.

Brady, A. and Salzberg, S.L. (2009) Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat. Methods* **6**: 673–676.

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G.W. (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Methods* **9**: 425–426.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X. and Jovanovich, S.B. (2008) The potential and challenges of nanopore sequencing. *Nature biotech* **26**: 1146–1153.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., and Rohwer, F. (2003) Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J. Bacteriol.* **185**: 6220–6223.

Brodin, J., Mild, M., Hedskog, C., Sherwood, E., Leitner, T., Andersson, B., and Albert,

J. (2013) PCR-Induced Transitions Are the Major Source of Error in Cleaned Ultra-Deep Pyrosequencing Data. *PLoS One* **8**: e70388.

Brongersma-Sanders, M. (1983) Unconsolidated phosphorites, high barium and diatom abundances in some Namibian shelf sediments. In, *Coastal Upwelling Its Sediment Record*. Springer US, pp. 421–437.

Brown, S.P., Veach, A.M., Rigdon-Huss, A.R., Grond, K., Lickteig, S.K., Lothamer, K., et al. (2015) Scraping the bottom of the barrel: Are rare high throughput sequences artifacts? *Fungal Ecol.* **13**: 221–225.

Brüchert, V., Currie, B., and Peard, K.R. (2009) Hydrogen sulphide and methane emissions on the central Namibian shelf. *Prog. Oceanogr.* **83**: 169–179.

Brüchert, V., Jørgensen, B.B., Neumann, K., Riechmann, D., Schlösser, M., and Schulz, H. (2003) Regulation of bacterial sulfate reduction and hydrogen sulfide fluxes in the central Namibian coastal upwelling zone. *Geochim. Cosmochim. Acta* **67**: 4505–4518.

Bryant, J.J.A., Stewart, F.F.J., Eppley, J.J.M., and DeLong, E.F.E. (2012) Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone. *Ecology* **93**: 1659–1673.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., et al. (2008) ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* **18**: 810–820.

Campbell, B.J. and Kirchman, D.L. (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME J.* **7**: 210–220.

Canfield, D.E., Kristensen, E., and Thamdrup, B. (2005) Aquatic geomicrobiology Elsevier Academic Press.

Canfield, D.E., Stewart, F.J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E.F., et al. (2010) A cryptic sulfur cycle in oxygen minimum zone waters off the Chilean coast. *Science* **330**: 1375–1378.

Canfield, D.E. and Thamdrup, B. (2009) Towards a consistent classification scheme for geochemical environments, or, why we wish the term ?suboxic? would go away. *Geobiology* **7**: 385–392.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello,

E.K., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**: 335–336.

Caporaso, J.G., Lauber, C.L., Walters, W. a, Berg-Lyons, D., Lozupone, C. a, Turnbaugh, P.J., et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 1**: 4516–4522.

Carr, M.-E. (2001) Estimation of potential productivity in Eastern Boundary Currents using remote sensing. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **49**: 59–80.

Cassman, N., Prieto-Davó, A., Walsh, K., Silva, G.G.Z., Angly, F., Akhter, S., et al. (2012) Oxygen minimum zones harbour novel viral communities with low diversity. *Environ. Microbiol.* **14**: 3043–3065.

Chambers, L.G., Reddy, K.R., and Osborne, T.Z. (2011) Short-Term Response of Carbon Cycling to Salinity Pulses in a Freshwater Wetland. *Soil Sci. Soc. Am. J.* **75**: 2000–2007.

Chan, A.L. and Walker, G. (1998) The settlement of Pomatoceros lamarckii larvae (Polychaeta: Sabellida: Serpulidae): a laboratory study. *Biofouling* **12**: 71–80.

Chan, C.K., Hsu, A.L., Halgamuge, S.K., and Tang, S.L. (2008) Binning sequences using very sparse labels within a metagenome. *BMC Bioinformatics* **9**: 215.

Chan, C.K., Hsu, A.L., Tang, S.L., and Halgamuge, S.K. (2008) Using growing self-organising maps to improve the binning process in environmental whole-genome shotgun sequencing. *J Biomed Biotechnol* **2008**: 513701.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**: 783–791.

Chao, A. (1984) Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* **11**: 265–270.

Chauhan, A. and Ogram, A. (2006) Phylogeny of acetate-utilizing microorganisms in soils along a nutrient gradient in the Florida Everglades. *Appl. Environ. Microbiol.* **72**: 6837–6840.

Chen, L., Wang, L. yun, Liu, S. juan, Hu, J. yong, He, Y., Zhou, H. wei, and Zhang, X. hui (2013) Profiling of microbial community during in situ remediation of volatile sulfide compounds in river sediment with nitrate by high throughput sequencing. *Int. Biodeterior. Biodegrad.* **85**: 429–437.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A.J., Muller, W.E., Wetter, T., and Suhai, S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* **14**: 1147–1159.

Chiu, C.-H., Wang, Y.-T., Walther, B.A., and Chao, A. (2014) An improved nonparametric lower bound of species richness via a modified good-turing frequency formula. *Biometrics* **70**: 671–682.

Chodak, M., Gołębiewski, M., Morawska-Płoskonka, J., Kuduk, K., Niklińska, M., Gołebiewski, M., et al. (2013) Diversity of microorganisms from forest soils differently polluted with heavy metals. *Appl. Soil Ecol.* **64**: 7–14.

Chouari, R., Le Paslier, D., Daegelen, P., Ginestet, P., Weissenbach, J. and Sghir, A. (2003) Molecular evidence for novel planctomycete diversity in a municipal wastewater treatment plant. *Appl. Environ. Microbiol.* **69**: 7354–7363.

Churbanov, A., Ryan, R., Hasan, N., Bailey, D., Chen, H., Milligan, B., and Houde, P. (2012) HighSSR: high-throughput SSR characterization and locus development from next-gen sequencing data. *Bioinformatics* **28**: 2797–2803.

Clarke, K. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**: 117–143.

Clarke, K. and Ainsworth, M. (1993) A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Ser.* **92**: 205–219.

Clarke, K. and Gorley, R. (2006) PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth.

Clarke, K.R. and Warwick, R. (1994) Similarity-based testing for community pattern: the two-way layout with no replication. *Mar. Biol.* **118**: 167–176.

Cloern, J.E. (2001) Our evolving conceptual model of the coastal eutrophication problem. *Mar Ecol Prog Ser* **210**: 223–253.

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W., et al. (2001) The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Sci. Mar.* **65**: 85–105.

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., et al. (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35**: D169–D172.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**: D141–D145.

Conley, D.J., Humborg, C., Rahm, L., Savchuk, O.P. and Wulff, F. (2002) Hypoxia in the Baltic Sea and basin-scale changes in phosphorus biogeochemistry. *Environ. Sci. Technol.* **36**: 5315–5320.

Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**: 485.

Cruaud, P., Vigneron, A., Lucchetti-Miganeh, C.C., Ciron, P.E., Godfroy, A., and Cambon-Bonavita, M.-A.A. (2014) Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. *Appl. Environ. Microbiol.* **80**: 4626–4639.

Crump, B.C., Peranteau, C., Beckingham, B., and Cornwell, J.C. (2007) Respiratory succession and community succession of bacterioplankton in seasonally anoxic estuarine waters. *Appl. Environ. Microbiol.* **73**: 6802–6810.

D'Argenio, V., Casaburi, G., Precone, V., and Salvatore, F. (2014) Comparative metagenomic analysis of human gut microbiome composition using two different bioinformatic pipelines. *Biomed Res Int* **2014**: 325340.

Dale, A.W., Brüchert, V., Alperin, M., and Regnier, P. (2009) An integrated sulfur isotope model for Namibian shelf sediments. *Geochim. Cosmochim. Acta* **73**: 1924–1944.

Dalsgaard, T. and Bak, F. (1994) Nitrate reduction in a sulfate-reducing bacterium Desulfovibrio desulfuricans, isolated from rice paaddy soil: sulfate inhibition, and regulation. *Appl. Environ. Microbiol.* **60**: 291–297.

Danovaro, R. and Pusceddu, A. (2007) Biodiversity and ecosystem functioning in coastal lagoons: Does microbial diversity play any role? *Estuar. Coast. Shelf Sci.* **75**: 4–12.

Dar, S.A., Yao, L., van Dongen, U., Kuenen, J.G., and Muyzer, G. (2007) Analysis of diversity and activity of sulfate-reducing bacterial communities in sulfidogenic bioreactors using 16S rRNA and dsrB genes as molecular markers. *Appl. Environ.*

Microbiol. **73**: 594–604.

Darr, A., Gogina, M., and Zettler, M.L. (2014) Functional changes in benthic communities along a salinity gradient– a western Baltic case study. *J. Sea Res.* **85**: 315–324.

DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U. S. A.* **89**: 5685–5689.

DeLong, E.F. and Pace, N.R. (2001) Environmental diversity of bacteria and archaea. *Syst. Biol.* **50**: 470–478.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**: 5069–5072.

Devol, A.H. (2003) Nitrogen cycle: Solution to a marine mystery. *Nature* **422**: 575–576.

Dhillon, A., Teske, A., Dillon, J., Stahl, D.A., and Sogin, M.L. (2003) Molecular characterization of sulfate-reducing bacteria in the Guaymas Basin. *Appl. Environ. Microbiol.* **69**: 2765–2772.

Diaz, N.N., Krause, L., Goesmann, A., Niehaus, K., and Nattkemper, T.W. (2009) TACOA: taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *BMC Bioinformatics* **10**: 56.

Diaz, R.J. (2001) Overview of Hypoxia around the World. *J. Environ. Qual.* **30**: 275.

Diaz, R.J. and Rosenberg, R. (2008) Spreading dead zones and consequences for marine ecosystems. *Science* **321**: 926–929.

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.

Dimitriou, P.D., Papageorgiou, N., Arvanitidis, C., Assimakopoulou, G., Pagou, K., Papadopoulou, K.N., et al. (2015) One Step forward: Benthic Pelagic Coupling and Indicators for Environmental Status. *PLoS One* **10**: e0141071.

Draper, N.R. and Smith, H. (1981) Applied regression analysis 2nd ed. John Wiley & Sons, New York. Chicago.

Du, J., Yuan, Z., Ma, Z., Song, J., Xie, X., and Chen, Y. (2014) KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path

analysis model. *Mol Biosyst* **10**: 2441–2447.

Dubilier, N. (1988) H2S—A Settlement Cue or a Toxic Substance For Capitella sp. I Larvae?. *Biol Bull* **174**: 30–38.

Dumont, M.G. and Murrell, J.C. (2005) Stable isotope probing - linking microbial identity to function. *Nat. Rev. Microbiol.* **3**: 499–504.

Dunford, E.A. and Neufeld, J.D. (2010) DNA stable-isotope probing (DNA-SIP). *J. Vis. Exp.* 1–2.

Dupont, C.L., Larsson, J., Yooseph, S., Ininbergs, K., Goll, J., Asplund-Samuelsson, J., et al. (2014) Functional Tradeoffs Underpin Salinity-Driven Divergence in Microbial Community Composition. *PLoS One* **9**: e89549.

Earl, D., Bradnam, K., St John, J., Darling, A., Lin, D., Fass, J., et al. (2011) Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res* **21**: 2224–2241.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.

Edgar, R.C. (2007) PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**: 18.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **647**: 1–5.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Edmonds, J.W., Weston, N.B., Joye, S.B., Mou, X., and Moran, M.A. (2009) Microbial community response to seawater amendment in low-salinity tidal sediments. *Microb. Ecol.* **58**: 558–568.

Eisenbarth, S. and Zettler, M.L. (2016) Diversity of the benthic macrofauna off northern Namibia from the shelf to the deep sea. *J Marine Syst* **155**: 1–10.

Emeis, K.-C., Brüchert, V., Currie, B., Endler, R., Ferdelman, T., Kiessling, A., et al. (2004) Shallow gas in shelf sediments of the Namibian coastal upwelling ecosystem. *Cont. Shelf Res.* **24**: 627–642.

Emeis, K.-C., Struck, U., Leipe, T., and Ferdelman, T.G. (2009) Variability in upwelling intensity and nutrient regime in the coastal upwelling system offshore Namibia: results from sediment archives. *Int. J. Earth Sci.* **98**: 309–326.

Emerson, S. and Hedges, J. (2003) Sediment diagenesis and benthic flux. *In*: Treatise on Geochemistry, 293–319.

Estrada, M. and Marrasé, C. (1987) Phytoplankton biomass and productivity off the Namibian Coast. *South African J. Mar. Sci.* **5**: 347–356.

Evans, T.W., Wörmer, L., Lever, M.A., Lipp, J.S., Lagostina, L., Lin, Y.S., Jørgensen, B.B. and Hinrichs, K.U. (2017) Size and composition of subseafloor microbial community in the Benguela upwelling area examined from intact membrane lipid and DNA analysis. *Orga Geochem* **111**: 86–100.

Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**: 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175–185.

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The Microbial Engines That Drive Earth ’s Biogeochemical Cycles. *Science.* **320**: 1034–1039.

Ferdelman, T.G., Lee, C., Pantoja, S., Harder, J., Bebout, B.M., and Fossing, H. (1997) Sulfate reduction and methanogenesis in a Thioploca-dominated sediment off the coast of Chile. *Geochim. Cosmochim. Acta* **61**: 3065–3079.

Ferentinos, G., Papatheodorou, G., Geraga, M., Iatrou, M., Fakiris, E., Christodoulou, D., et al. (2010) Fjord water circulation patterns and dysoxic/anoxic conditions in a Mediterranean semi-enclosed embayment in the Amvrakikos Gulf, Greece. *Estuar. Coast. Shelf Sci.* **88**: 473–481.

Fierer, N., McCain, C.M.C.M., Meir, P., Zimmermann, M., Rapp, J.M.J.M., Silman, M.R.M.R., and Knight, R. (2011) Microbes do not follow the elevational diversity patterns of plants and animals. *Ecology* **92**: 797–804.

Finke, N., Vandieken, V., and Jørgensen, B.B. (2007) Acetate, lactate, propionate, and isobutyrate as electron donors for iron and sulfate reduction in Arctic marine sediments, Svalbard. *FEMS Microbiol. Ecol.* **59**: 10–22.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., et al. (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**: D222-D230.

Finney, B.P., Emeis, K.-C., Field, D.B., Gutiérrez, D., and Struck, U. (2010) Paleoecological studies on variability in marine fish populations: A long-term perspective on the impacts of climatic change on marine ecosystems. *J. Mar. Syst.* **79**: 316–326.

Fish, J.A., Chai, B., Wang, Q., Sun, Y., Brown, C.T., Tiedje, J.M. and Cole, J.R. (2013) FunGene: the functional gene pipeline and repository. *Front. Microbiol.* **4**: 291.

Fishbain, S., Dillon, J.G., Gough, H.L., and Stahl, D.A. (2003) Linkage of High Rates of Sulfate Reduction in Yellowstone Hot Springs to Unique Sequence Types in the Dissimilatory Sulfate Respiration Pathway. *Appl. Environ. Microbiol.* **69**: 3663–3667.

Fortunato, C.S., Herfort, L., Zuber, P., Baptista, A.M., and Crump, B.C. (2012) Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME J.* **6**: 554–563.

Fossing, H. and Jørgensen, B.B. (1989) Measurement of bacterial sulfate reduction in sediments: evaluation of a single-step chromium reduction method. *Biogeochemistry* **8**: 205–222.

Foti, M., Sorokin, D.Y., Lomans, B., Mussman, M., Zacharova, E.E., Pimenov, N. V, et al. (2007) Diversity, activity, and abundance of sulfate-reducing bacteria in saline and hypersaline soda lakes. *Appl. Environ. Microbiol.* **73**: 2093–2100.

Friedrich, M.W. (2006) Stable-isotope probing of DNA: insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. *Curr. Opin. Biotechnol.* **17**: 59–66.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Fukui, M., Suh, J., Yonezawa, Y., and Urushigawa, Y. (1997) Major substrates for microbial sulfate reduction in the sediments of Ise Bay, Japan. *Ecol. Res.* **12**: 201–209.

Gallagher, E., McGuinness, L., Phelps, C., Young, L.Y., and Kerkhof, L.J. (2005) 13C-carrier DNA shortens the incubation time needed to detect benzoate-utilizing denitrifying bacteria by stable-isotope probing. *Appl. Environ. Microbiol.* **71**: 5192–5196.

García-Moyano, A., González-Toril, E., Aguilera, Á., Amils, R., Aguilera, A., Garc??a-

Moyano, A., et al. (2012) Comparative microbial ecology study of the sediments and the water column of Río Tinto, an extreme acidic environment. *FEMS Microbiol. Ecol.* **81**: 303–314.

García de Lomas, J., Corzo, A., Gonzalez, J.M., Andrades, J.A., Iglesias, E., and Montero, M.J. (2006) Nitrate promotes biological oxidation of sulfide in wastewaters: Experiment at plant-scale. *Biotechnol. Bioeng.* **93**: 801–811.

Gaspar, J.M. and Thomas, W. (2012) The consequences of denoising marker-based metagenomic data. *BMC Proc.* **6**: P11.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**: R80.

Ghai, R., Hernandez, C.M., Picazo, A., Mizuno, C.M., Ininbergs, K., Díez, B., et al. (2012) Metagenomes of mediterranean coastal lagoons. *Sci. Rep.* **2**: 490.

Ghosh, T.S., Mohammed, M.H., Komanduri, D., and Mande, S.S. (2011) ProViDE: A software tool for accurate estimation of viral diversity in metagenomic samples. *Bioinformation* **6**: 91–94.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**: 1451–1455.

Gich, F., Schubert, K., Bruns, A., Hoffelner, H., and Overmann, J. (2005) Specific detection, isolation, and characterization of selected, previously uncultured members of the freshwater bacterioplankton community. *Appl. Environ. Microbiol.* **71**: 5908–5919.

Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., et al. (2010) The taxonomic and functional diversity of microbes at a temperate coastal site: a “multi-omic” study of seasonal and diel temporal variation. *PLoS One* **5**: e15545.

Gilmour, C.C., Riedel, G.S., Ederington, M.C., Bell, J.T., Gill, G.A., and Stordal, M.C. (1998) Methylmercury concentrations and production rates across a trophic gradient in the northern Everglades. *Biogeochemistry* **40**: 327–345.

Ginige, M.P., Keller, J., and Blackall, L.L. (2005) Investigation of an Acetate-Fed Denitrifying Microbial Community by Stable Isotope Probing , Full-Cycle rRNA

Analysis , and Fluorescent In Situ Investigation of an Acetate-Fed Denitrifying Microbial Community by Stable Isotope Probing , Full-Cycle rRNA An. *Appl. Environ. Microbiol.* **71**: 8683–8691.

Giongo, A., Crabb, D.B., Davis-richardson, A.G., Chauliac, D., Mobberley, J.M., Gano, K.A., et al. (2010) PANGEA: pipeline for analysis of next generation amplicons. *ISME J.* **4**: 852–861.

Glass, E., Wilkening, J., Wilke, A., Antonopoulos, D., and Meyer, F. (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **1**: pdb–prot5368.

Glöckner, F.O. and Joint, I. (2010) Marine microbial genomics in Europe: current status and perspectives. *Microb. Biotechnol.* **3**: 523–530.

Glöckner, F.O., Stal, L., Sandaa, R.A., Gasol, J.M., O'Gara, F., Hernandez, F., et al. (2012) Marine Microbial Diversity and its role in Ecosystem Functioning and Environmental Change. Marine Board Position Paper 17. Calewaert, J.B. and McDonough N. (Eds.) Ostend, Belgium.

Glud, R.N. (2008) Oxygen dynamics of marine sediments. *Mar Biol Res* **4**: 243–289.

Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**: R86.

Goldhammer, T., Brüchert, V., Ferdelman, T.G., and Zabel, M. (2010) Microbial sequestration of phosphorus in anoxic upwelling sediments. *Nat. Geosci.* **3**: 557–561.

Gołębiewski, M., Deja-Sikora, E., Cichosz, M., Tretyn, A., and Wróbel, B. (2014) 16S rDNA pyrosequencing analysis of bacterial community in heavy metals polluted soils. *Microb. Ecol.* **67**: 635–647.

Goll, J., Rusch, D.B., Tanenbaum, D.M., Thiagarajan, M., Li, K., Methé, B.A., and Yooseph, S. (2010) METAREP: JCVI metagenomics reports--an open source tool for high-performance comparative metagenomics. *Bioinformatics* **26**: 2631–2632.

Gray, J.S. and Elliott, M. (2009) Ecology of Marine Sediments. From Science to Management Second. Oxford University Press.

Grote, J., Schott, T., Bruckner, C.G., Oliver, F., Jost, G., Teeling, H., et al. (2012)

Genome and physiology of a model Epsilonproteobacterium responsible for sulfide detoxification in marine oxygen depletion zones. *PNAS* **109**: 506–510.

Gruber, N. and Sarmiento, J.L. (1997) Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochem. Cycles* **11**: 235–266.

Guelorget, O. and Perthuisot, J.-P. (1983) Le domaine paralique: expressions géologiques, biologiques et économiques du confinement Presses de l'École Normale Supérieure.

Guelorget, O. and Perthuisot, J.-P. (1992) Paralic ecosystems. Biological organization and functioning. *Vie milieu* **42**: 215–251.

Gutierrez-Zamora, M.-L. and Manefield, M. (2010) An appraisal of methods for linking environmental processes to specific microbial taxa. *Rev. Environ. Sci. Bio/Technology* **9**: 153–185.

Haft, D.H., Selengut, J.D., and White, O. (2003) The TIGRFAMs database of protein families. *Nucleic Acids Res* **31**: 371–373.

Hallin, S., Rothman, M., and Pell, M. (1996) Adaptation of denitrifying bacteria to acetate and methanol in activated sludge. *Water Res.* **30**: 1445–1450.

Handelsman, J. (2009) Metagenetics: spending our inheritance on the future. *Microb. Biotechnol.* **2**: 138–139.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., et al. (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10**: R32.

Hawkes, C.V. and Keitt, T.H. (2015) Resilience vs. historical contingency in microbial responses to environmental change. *Ecol. Lett.* **18**: 612–625.

Hedges, J. and Stern, J. (1984) Carbon and nitrogen determinations of carbonate-containing solids. *Limnol. Oceanogr.* **29**: 657–663.

Henckel, T., Friedrich, M., and Conrad, R. (1999) Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA, particulate methane monooxygenase, and methanol dehydrogenase. *Appl. Environ. Microbiol.* **65**: 1980–1990.

Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**: 1571–1579.

Herlemann, D.P.R., Lundin, D., Andersson, A.F., Labrenz, M., and Jürgens, K. (2016) Phylogenetic Signals of Salinity and Season in Bacterial Community Composition Across the Salinity Gradient of the Baltic Sea. *Front. Microbiol.* **7**: 1–13.

Highton, M.P., Roosa, S., Crawshaw, J., Schallenberg, M., and Morales, S.E. (2016) Physical Factors Correlate to Microbial Community Structure and Nitrogen Cycling Gene Abundance in a Nitrate Fed Eutrophic Lagoon. *Front. Microbiol.* **7**: 1691.

Hoff, K.J., Lingner, T., Meinicke, P., and Tech, M. (2009) Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Res* **37**: W101-W105.

Hori, T., Müller, A., Igarashi, Y., Conrad, R., and Friedrich, M.W. (2010) Identification of iron-reducing microorganisms in anoxic rice paddy soil by 13C-acetate probing. *ISME J.* **4**: 267–278.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.

Huggett, M.J., Williamson, J.E., de Nys, R., Kjelleberg, S. and Steinberg, P.D. (2006) Larval settlement of the common Australian sea urchin Heliocidaris erythrogramma in response to bacteria from the surface of coralline algae. *Oecologia* **149**: 604–619.

Hunter, E.M., Mills, H.J., and Kostka, J.E. (2006) Microbial community diversity associated with carbon and nitrogen cycling in permeable shelf sediments. *Appl. Environ. Microbiol.* **72**: 5689–5701.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**: D211-D215.

Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., et al. (2014) EBI metagenomics--a new resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res* **42**: D600-D606.

Huse, S.M., Mark Welch, D.B., Voorhis, A., Shipunova, A., Morrison, H.G., Eren, A.M., and Sogin, M.L. (2014) VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinformatics* **15**: 41.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of

metagenomic data. *Genome Res.* **17**: 377–386.

Huson, D.H. and Mitra, S. (2012) Introduction to the analysis of environmental sequences: metagenomics with MEGAN. *Methods Mol Biol* **856**: 415–429.

Huson, D.H. and Weber, N. (2013) Microbial community analysis using MEGAN. *Methods Enzym.* **531**: 465–485.

Hwang, K., Oh, J., Kim, T.-K., Kim, B.K., Yu, D.S., Hou, B.K., et al. (2013) CLUSTOM: A Novel Method for Clustering 16S rRNA Next Generation Sequences by Overlap Minimization. *PLoS One* **8**: e62623.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

International Symposium on the Classification of Brackish Waters (1958) The Venice system for the classification of marine waters according to salinity. *Limnol. Oceanogr.* **3**: 346–347.

Inthorn, M., Mohrholz, V., and Zabel, M. (2006) Nepheloid layer distribution in the Benguela upwelling area offshore Namibia. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **53**: 1423–1438.

Inthorn, M., Wagner, T., Scheeder, G. and Zabel, M., (2006b) Lateral transport controls distribution, quality, and burial of organic matter along continental slopes in high-productivity areas. *Geology* **34**: 205–208.

Iyer, S., Bouzek, H., Deng, W., Larsen, B., Casey, E., and Mullins, J.I. (2013) Quality score based identification and correction of pyrosequencing errors. *PLoS One* **8**: e73015.

Jehmlich, N., Schmidt, F., von Bergen, M., Richnow, H.-H., and Vogt, C. (2008) Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures. *ISME J.* **2**: 1122–1133.

Jeraldo, P., Chia, N., and Goldenfeld, N. (2011) On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys. *Environ. Microbiol.* **13**: 3000–3009.

Jiang, L., Zheng, Y., Peng, X., Zhou, H., Zhang, C., Xiao, X., and Wang, F. (2009) Vertical distribution and diversity of sulfate-reducing prokaryotes in the Pearl River estuarine sediments, Southern China. *FEMS Microbiol. Ecol.* **70**: 249–262.

Jiang, X.-T., Peng, X., Deng, G.-H., Sheng, H.-F., Wang, Y., Zhou, H.-W., and Tam, N.F.-Y. (2013) Illumina Sequencing of 16S rRNA Tag Revealed Spatial Variations of Bacterial Communities in a Mangrove Wetland. *Microb. Ecol.* **66**: 96–104.

Johnson, D.B. (2012) Geomicrobiology of extremely acidic subsurface environments. *FEMS Microbiol. Ecol.* **81**: 2–12.

Joint, I., Muhling, M., and Querellou, J. (2010) Culturing marine bacteria – an essential prerequisite for biodiscovery. *Microb. Biotechnol.* **3**: 564–575.

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* **3**: 442–453.

Jong, T. and Parry, D.L. (2003) Removal of sulfate and heavy metals by sulfate reducing bacteria in short-term bench scale upflow anaerobic packed bed reactor runs. *Water Res.* **37**: 3379–3389.

Jørgensen, B.B. (1982) Mineralization of organic matter in the sea bed—the role of sulphate reduction. *Nature* **296**: 643–645.

Joshi, N. and Fass, J. (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). [Software]. Available at https://github.com/najosh.

Kan, J., Hanson, T.E., Ginter, J.M., Wang, K. and Chen, F. (2005) Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline systems* **1**: 7.

Kaneko, R., Hayashi, T., Tanahashi, M., and Naganuma, T. (2007) Phylogenetic Diversity and Distribution of Dissimilatory Sulfite Reductase Genes from Deep-Sea Sediment Cores. *Mar. Biotechnol.* **9**: 429–436.

Kapsimalis, V., Pavlakis, P., Poulos, S.E., Alexandri, S., Tziavos, C., Sioulas, A., et al. (2005) Internal structure and evolution of the Late Quaternary sequence in a shallow embayment: The Amvrakikos Gulf, NW Greece. *Mar. Geol.* **222–223**: 399–418.

Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**: 511–518.

Keegan, K.P., Trimble, W.L., Wilkening, J., Wilke, A., Harrison, T., D'Souza, M., and Meyer, F. (2012) A Platform-Independent Method for Detecting Errors in Metagenomic Sequencing Data: DRISEE. *PLoS Comput. Biol.* **8**: e1002541.

Kent, W.J. (2002) BLAT--the BLAST-like alignment tool. *Genome Res* **12**: 656–664.

Kilias, S.P., Nomikou, P., Papanikolaou, D., Polymenakou, P.N., Godelitsas, A., Argyraki, A., et al. (2013) New insights into hydrothermal vent processes in the unique shallow-submarine arc-volcano, Kolumbo (Santorini), Greece. *Sci. Rep.* **3**: 1–13.

Kim, M., Morrison, M., and Yu, Z. (2011) Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microbiol. Methods* **84**: 81–87.

Kircher, M., Sawyer, S., and Meyer, M. (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**: e3.

Kirchman, D.L., Dittel, A.I., Malmstrom, R.R., and Cottrell, M.T. (2005) Biogeography of major bacterial groups in the Delaware estuary. *Limnol. Oceanogr.* **50**: 1697–1706.

Kjeldsen, K.U., Loy, A., Jakobsen, T.F., Thomsen, T.R., Wagner, M., and Ingvorsen, K. (2007) Diversity of sulfate-reducing bacteria from an extreme hypersaline sediment, Great Salt Lake (Utah). *FEMS Microbiol. Ecol.* **60**: 287–298.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**: e1.

Knoppers, B. (1994) Aquatic Primary Production. In, Kjerfve,B. (ed), *Coastal Lagoon Processes*. Elsevier, pp. 243–286.

Kõljalg, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., et al. (2013) Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **22**: 5271–5277.

Konar, B., Iken, K., Pohle, G., Miloslavich, P., Cruz-Motta, J.J., Benedetti-Cecchi, L., Kimani, E., Knowlton, A., Trott, T., Iseto, T. and Shirayama, Y. (2010) Surveying nearshore biodiversity. Wiley-Blackwell, Chichester. *In:* Life in the world's oceans: diversity, distribution, and abundance, 27–41.

Kondo, R. and Butani, J. (2007) Comparison of the diversity of sulfate-reducing bacterial communities in the water column and the surface sediments of a

Japanese meromictic lake. *Limnology* **8**: 131–141.

Kong, L., Jing, H., Kataoka, T., Buchwald, C., and Liu, H. (2013) Diversity and spatial distribution of hydrazine oxidoreductase (hzo) gene in the oxygen minimum zone off Costa Rica. *PLoS One* **8**: e78275.

Koskinen, K., Auvinen, P., Björkroth, K.J., Hultman, J., Bjoerkroth, K.J., and Hultman, J. (2014) Inconsistent denoising and clustering algorithms for amplicon sequence data. *J. Comput. Biol.* **22**: 1–9.

Koslicki, D., Foucart, S., and Rosen, G. (2014) WGSQuikr: fast whole-genome shotgun metagenomic classification. *PLoS One* **9**: e91784.

Kountoura, K. and Zacharias, I. (2011) Temporal and spatial distribution of hypoxic/seasonal anoxic zone in Amvrakikos Gulf, Western Greece. *Estuar. Coast. Shelf Sci.* **94**: 123–128.

Krause, L., Diaz, N.N., Goesmann, A., Kelley, S., Nattkemper, T.W., Rohwer, F., et al. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res.* **36**: 2230–2239.

Kreuzer-Martin, H.W. (2007) Stable Isotope Probing: Linking Functional Activity to Specific Members of Microbial Communities. *Soil Sci. Soc. Am. J.* **71**: 611–619.

Kristiansen, K.D., Kristensen, E., and Jensen, E.M.H. (2002) The Influence of Water Column Hypoxia on the Behaviour of Manganese and Iron in Sandy Coastal Marine Sediment. *Estuar. Coast. Shelf Sci.* **55**: 645–654.

Kruskal, W.H. and Wallis, W.A. (1952) Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**: 583–621.

Kumar, S., Carlsen, T., Mevik, B.-H., Enger, P., Blaalid, R., Shalchian-Tabrizi, K., and Kauserud, H. (2011) CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinformatics* **12**: 1–9.

Kunin, V., Engelbrektson, A., Ochman, H. and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118–123.

Kuypers, M.M., Sliekers, A.O., Lavik, G. and Schmid, M. (2003) Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**: 608.

Kuypers, M.M.M., Lavik, G., Woebken, D., Schmid, M., Fuchs, B.M., Amann, R., et al.

(2005) Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 6478–6483.

Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., and Clavel, T. (2016) IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci. Rep.* **6**: 33721.

Lahajnar, N. (2011) RV MARIA S. MERIAN, Cruise Report MSM17/L3, Geochemistry and Ecology of the Namibian Upwelling System.

Lahajnar, N., Mohrholz, V., Angenendt, S., Ankele, M., Annighöfer, M., Beier, S., et al. (2015) Geochemistry and Ecology of the Namibian Upwelling System NAMUFIL: Namibian Upwelling Filament Study - Cruise No. M103 - December 27, 2013 – February 11, 2014 – Walvis Bay (Namibia) - Walvis Bay (Namibia). METEOR-Berichte, M103.

Lam, P. and Kuypers, M.M.M. (2011) Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. *Ann. Rev. Mar. Sci.* **3**: 317–345.

Lam, P., Lavik, G., Jensen, M.M., van de Vossenberg, J., Schmid, M., Woebken, D., et al. (2009) Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 4752–4757.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., et al. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**: 814–821.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lanzén, A., Jørgensen, S.L., Huson, D.H., Gorfer, M., Grindhaug, S.H., Jonassen, I., et al. (2012) CREST – Classification Resources for Environmental Sequence Tags. *PLoS One* **7**: e49334.

Lavik, G., Stührmann, T., Brüchert, V., Van der Plas, A., Mohrholz, V., Lam, P., et al. (2009) Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457**: 581–584.

Leduc, G., Herbert, C.T., Blanz, T., Martinez, P., and Schneider, R. (2010) Contrasting evolution of sea surface temperature in the Benguela upwelling system under natural and anthropogenic climate forcings. *Geophys. Res. Lett.* **37**: L20705.

Lee, J.H., Yi, H., and Chun, J. (2011) rRNASelector: a computer program for selecting

ribosomal RNA encoding sequences from metagenomic and metatranscriptomic shotgun libraries. *J Microbiol* **49**: 689–691.

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., et al. (2011) The European Nucleotide Archive. *Nucleic Acids Res.* **39**: 1–4.

Leloup, J., Fossing, H., Kohls, K., Holmkvist, L., Borowski, C., and Jørgensen, B.B. (2009) Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity related to geochemical zonation. *Environ. Microbiol.* **11**: 1278–1291.

Lespinet, O. and Labedan, B. (2005) Orphan enzymes? *Science* **307**: 42.

Levin, L. (2003) Oxygen minimum zone benthos: adaptation and community response to hypoxia. *In:* Oceanography and marine biology: an annual review **41**: 1–45.

Li, J., Purdy, K.J., Takii, S., and Hayashi, H. (1999) Seasonal changes in ribosomal RNA of sulfate-reducing bacteria and sulfate reducing activity in a freshwater lake sediment. *FEMS Microbiol. Ecol.* **28**: 31–39.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**: 713–714.

Li, S. and Chou, H.-H. (2004) LUCY2: an interactive DNA sequence quality trimming and vector removal tool. *Bioinformatics* **20**: 2865–2866.

Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Lipsewers, Y.A., Vasquez-Cardenas, D., Seitaj, D., Schauer, R., Hidalgo Martinez, S., Sinninghe Damsté, J.S., et al. (2017) Impact of Seasonal Hypoxia on Activity and Community Structure of Chemolithoautotrophic Bacteria in a Coastal Sediment. *Appl. Environ. Microbiol.* **83**: e03517–e03616.

Liu, B., Gibbons, T., Ghodsi, M., Treangen, T., and Pop, M. (2011) Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics* **12 Suppl 2**: S4.

Liu, W.T.T., Marsh, T.L.L., Cheng, H., and Forney, L.J.J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **63**: 4516–4522.

Liu, X.-Z., Zhang, L.-M., Prosser, J.I., and He, J.-Z. (2009) Abundance and community structure of sulfate reducing prokaryotes in a paddy soil of southern China under different fertilization regimes. *Soil Biol. Biochem.* **41**: 687–694.

Liu, Z., Lozupone, C., Hamady, M., Bushman, F.D., and Knight, R. (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* **35**: e120.

Logares, R., Bråte, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K., and Rengefors, K. (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol.* **17**: 414–422.

Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.

Lozupone, C.A. and Knight, R. (2007) Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 11436–11440.

Lozupone, C., Hamady, M., and Knight, R. (2006) UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 371.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**: 1363–1371.

Lueders, T., Dumont, M.G., Bradford, L., and Manefield, M. (2016) RNA-stable isotope probing: From carbon flow within key microbiota to targeted transcriptomes. *Curr. Opin. Biotechnol.* **41**: 83–89.

Lueders, T., Manefield, M., and Friedrich, M.W. (2004) Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ. Microbiol.* **6**: 73–78.

Lüke, C., Speth, D.R., Kox, M.A.R., Villanueva, L., and Jetten, M.S.M. (2016) Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. *PeerJ* **4**: e1924.

Luna, R.A., Fasciano, L.R., Jones, S.C., Boyanton, B.L.J., Ton, T.T., and Versalovic, J. (2007) DNA pyrosequencing-based bacterial pathogen identification in a pediatric hospital setting. *J. Clin. Microbiol.* **45**: 2985–2992.

Luo, C., Rodriguez-R, L.M., and Konstantinidis, K.T. (2014) MyTaxa: an advanced

taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res.* **42**: e73.

MacDonald, N.J., Parks, D.H., and Beiko, R.G. (2012) Rapid identification of high-confidence taxonomic assignments for metagenomic data. *Nucleic Acids Res.* **40**: e111.

Madigan, M.T., Martinko, J.M., and Parker, J. (2003) Brock biology of microorganisms 10th editi. Pearson Education, Inc.

Madigan, M.T., Martinko, J.M., Stahl, D.A., and Clark, D.P. (2012) Brock: Biology of Microorganisms 13th Editi. Benjamin Cummings, San Francisco, CA.

Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J., and Woese, C.R. (1996) The Ribosomal Database Project (RDP). *Nucleic Acids Res* **24**: 82–85.

Manefield, M., Whiteley, A.S., Griffiths, R.I., and Bailey, M.J. (2002) RNA Stable Isotope Probing , a Novel Means of Linking Microbial Community Function to Phylogeny. *Appl. Environ. Microbiol.* **68**: 5367–5373.

Mann, H.B. and Whitney, D.R. (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Stat.* **18**: 50–60.

Mardis, E.R. (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **9**: 387–402.

Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., et al. (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42**: D568–D573.

Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., et al. (2014) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* **42**: D560-D567.

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., and Neufeld, J.D. (2012) PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**: 31.

Mavromatis, K., Ivanova, N., Barry, K., Shapiro, H., Goltsman, E., McHardy, A.C., et al. (2007) Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nat. Methods* **4**: 495–500.

McDonald, D., Clemente, J.C., Kuczynski, J., Rideout, J.R., Stombaugh, J., Wendel, D., et al. (2012) The Biological Observation Matrix (BIOM) format or: how I learned

to stop worrying and love the ome-ome. *Gigascience* **1**: 7.

McHardy, A.C., Martin, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nat Methods* **4**: 63–72.

McLusky, D. and Elliott, M. (2004) The estuarine ecosystem: ecology, threats and management Oxford University Press on Demand.

McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**: e61217.

Meinicke, P. (2015) UProC: tools for ultra-fast protein domain classification. *Bioinformatics* **31**: 1382–1388.

Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**: 31–46.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Middelburg, J.J., Soetaert, K., Herman, P.M. and Heip, C.H. (1996) Denitrification in marine sediments: A model study. *Global Biogeochem Cycles* **10**: 661–673.

Middelburg, J. and Levin, L. (2009) Coastal hypoxia and sediment biogeochemistry. *Biogeosciences* **6**: 1273–1293.

Miller, J.R., Koren, S., and Sutton, G. (2010) Assembly algorithms for next-generation sequencing data. *Genomics* **95**: 315–327.

Mohammed, M.H., Ghosh, T.S., Singh, N.K., and Mande, S.S. (2011) SPHINX--an algorithm for taxonomic binning of metagenomic sequences. *Bioinformatics* **27**: 22–30.

Moloney, C.L., Jarre, A., Arancibia, H., Bozec, Y.-M., Neira, S., Roux, J.-P., and Shannon, L.J. (2005) Comparing the Benguela and Humboldt marine upwelling ecosystems with indicators derived from inter-calibrated models. *ICES J. Mar. Sci.* **62**: 493–502.

Monteiro, P.M.S., van der Plas, A., Mohrholz, V., Mabille, E., Pascall, A., and Joubert, W. (2006) Variability of natural hypoxia and methane in a coastal upwelling

system: Oceanic physics or shelf biology? *Geophys. Res. Lett.* **33**: L16614.

Monteiro, P.M.S., van der Plas, A.K., Mélice, J.-L., and Florenchie, P. (2008) Interannual hypoxia variability in a coastal upwelling system: Ocean–shelf exchange, climate and ecosystem-state implications. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **55**: 435–450.

Monzoorul Haque, M., Ghosh, T.S., Komanduri, D., and Mande, S.S. (2009) SOrt-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics* **25**: 1722–1730.

Morgulis, A., Gertz, E.M., Schaffer, A.A., and Agarwala, R. (2006) A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol* **13**: 1028–1040.

Mosier, A.C. and Francis, C.A. (2008) Relative abundance and diversity of ammonia-oxidizing archaea and bacteria in the San Francisco Bay estuary. *Environ. Microbiol.* **10**: 3002–3016.

Mosier, A.C. and Francis, C.A. (2010) Denitrifier abundance and activity across the San Francisco Bay estuary. *Environ. Microbiol. Rep.* **2**: 667–676.

Moura, I., Bursakov, S., Costa, C., and Moura, J.J. (1997) Nitrate and nitrite utilization in sulfate-reducing bacteria. *Anaerobe* **3**: 279–290.

Müller, A.L., Kjeldsen, K.U., Rattei, T., Pester, M., and Loy, A. (2015) Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* **9**: 1152–1165.

Murase, J., Hordijk, K., Tayasu, I., Bodelier, P.L.E., H, S., JHP, H., et al. (2011) Strain-specific incorporation of methanotrophic biomass into eukaryotic grazers in a rice field soil revealed by PLFA-SIP. *FEMS Microbiol. Ecol.* **75**: 284–290.

Muyzer, G. (1999) DGGE/TGGE a method for identifying genes from natural ecosystems. *Curr. Opin. Microbiol.* **2**: 317–322.

Muyzer, G. and Stams, A.J.M. (2008) The ecology and biotechnology of sulphate-reducing bacteria. *Nat. Rev. Microbiol.* **6**: 441–454.

Nagel, B., Gaye, B., Lahajnar, N., Struck, U., and Emeis, K.-C. (2016) Effects of current regimes and oxygenation on particulate matter preservation on the Namibian shelf: Insights from amino acid biogeochemistry. *Mar. Chem.* **186**: 121–132.

Nakagawa, T., Hanada, S., Maruyama, A., and Marumo, K. (2002) Distribution and

diversity of thermophilic sulfate-reducing bacteria within a Cu-Pb-Zn mine (Toyoha, Japan). **41**: 199–209.

Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* **40**: e155.

Naqvi, S.W.A., Voss, M., and Montoya, J.P. (2008) Recent advances in the biogeochemistry of nitrogen in the ocean. *Biogeosciences* **5**: 1033–1041.

Nebel, M.E., Wild, S., Holzhauser, M., Hüttenberger, L., Reitzig, R., Sperber, M., and Stoeck, T. (2011) Jaguc - a software package for environmental diversity analyses. *J. Bioinform. Comput. Biol.* **9**: 749–773.

Nelson, T.M., Streten, C., Gibb, K.S. and Chariton, A.A. (2015) Saltwater intrusion history shapes the response of bacterial communities upon rehydration. *Sci Total Environ* **502**: 143–148.

Neubauer, S.C., Franklin, R.B., and Berrier, D.J. (2013) Saltwater intrusion into tidal freshwater marshes alters the biogeochemical processing of organic carbon. *Biogeosciences* **10**: 8171–8183.

Neufeld, J.D., Dumont, M.G., Vohra, J., and Murrell, J.C. (2007) Methodological considerations for the use of stable isotope probing in microbial ecology. *Microb. Ecol.* **53**: 435–442.

Nicolaidou, A., Petrou, K., Kormas, K. and Reizopoulou, S. (2006) Inter-annual variability of soft bottom macrofaunal communities in two Ionian Sea lagoons. *Marine Biodiv* 89–98.

Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., et al. (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**: 822–828.

Nikolenko, S.I., Korobeynikov, A.I., and Alekseyev, M.A. (2013) BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14**: S7.

Nilakanta, H., Drews, K.L., Firrell, S., Foulkes, M.A., and Jablonski, K.A. (2014) A review of software for analyzing molecular sequences. *BMC Res. Notes* **7**: 830.

Noguchi, H., Park, J., and Takagi, T. (2006) MetaGene: prokaryotic gene finding from

environmental genome shotgun sequences. *Nucleic Acids Res* **34**: 5623–5630.

Nordberg, K., Filipsson, H.L., Gustafsson, M., Harland, R., and Roos, P. (2001) Climate, hydrographic variations and marine benthic hypoxia in Koljö Fjord, Sweden. *J. Sea Res.* **46**: 187–200.

Nübel, U., Garcia-Pichel, F., Clavero, E., and Muyzer, G. (2000) Matching molecular diversity and ecophysiology of benthic cyanobacteria and diatoms in communities along a salinity gradient. *Environ. Microbiol.* **2**: 217–226.

O'Hara, R.B. (2005) Species richness estimators: how many species can dance on the head of a pin? *J. Anim. Ecol.* **74**: 375–386.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **27**: 29–34.

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2016) vegan: Community Ecology Package. R package version 2.4-1. https://CRAN.R-project.org/package=vegan.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., et al. (2008) The vegan package. **10**: 2008.

Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**: 385.

Osaka, T., Yoshie, S., Tsuneda, S., Hirata, A., Iwami, N., and Inamori, Y. (2006) Identification of Acetate- or Methanol-Assimilating Bacteria under Nitrate-Reducing Conditions by Stable-Isotope Probing. *Microb. Ecol.* **52**: 253–266.

Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N., Kotoulas, G., et al. (2015) Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform. Biol. Insights* **9**: 75.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J. V, Chuang, H.Y., Cohoon, M., et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691–5702.

Painchaud, J., Therriault, J., and Legendre, L. (1995) Assessment of salinity-related mortality of freshwater bacteria in the saint lawrence estuary. *Appl. Environ. Microbiol.* **61**: 205–208.

Palmer, T.A., Montagna, P.A., Pollack, J.B., Kalke, R.D., and DeYoe, H.R. (2011) The

role of freshwater inflow in lagoons, rivers, and bays. *Hydrobiologia* **667**: 49–67.

Pandey, R.V., Nolte, V., and Schlötterer, C. (2010) CANGS: a user-friendly utility for processing and analyzing 454 GS-FLX data in biodiversity studies. *BMC Res. Notes* **3**: 1–7.

Papaspyrou, S., Gregersen, T., Kristensen, E., Christensen, B. and Cox, R.P. (2006) Microbial reaction rates and bacterial communities in sediment surrounding burrows of two nereidid polychaetes (Nereis diversicolor and N. virens). *Mar Biol* **148**: 541–550.

Papatheodorou, G., Hasiotis, T., and Ferentinos, G. (1993) Gas-charged sediments in the Aegean and Ionian Seas, Greece. *Mar. Geol.* **112**: 171–184.

Parkes, R.J., Gibson, G.R., Mueller-Harvey, I., Buckingham, W.J., and Herbert, R. a (1989) Determination of the Substrates for Sulphate-reducing Bacteria within Marine and Esturaine Sediments with Different Rates of Sulphate Reduction. *J. Gen. Microbiol.* **135**: 175–187.

Paszkiewicz, K. and Studholme, D.J. (2010) De novo assembly of short sequence reads. *Brief. Bioinform.* **11**: 457–472.

Pati, A., Heath, L.S., Kyrpides, N.C., and Ivanova, N. (2011) ClaMS: A Classifier for Metagenomic Sequences. *Stand Genomic Sci* **5**: 248–253.

Patil, K.R., Roune, L., and McHardy, A.C. (2012) The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS One* **7**: e38581.

Patin, N. V., Kunin, V., Lidström, U., and Ashby, M.N. (2013) Effects of OTU Clustering and PCR Artifacts on Microbial Diversity Estimates. *Microb. Ecol.* **65**: 709–719.

Paul, J.W., Beauchamp, E.G., and Trevors, J.T. (1989) Acetate, propionate, butyrate, glucose, and sucrose as carbon sources for denitrifying bacteria in soil. *Can. J. Microbiol.* **35**: 754–759.

Pavlopoulos, G.A., Oulas, A., Iacucci, E., Sifrim, A., Moreau, Y., Schneider, R., et al. (2013) Unraveling genomic variation from next generation sequencing data. *BioData Min.* **6**: 13.

Pavloudi, C., Oulas, A., Vasileiadou, K., Sarropoulou, E., Kotoulas, G., and Arvanitidis, C. (2016) Salinity is the major factor influencing the sediment bacterial communities in a Mediterranean lagoonal complex (Amvrakikos Gulf, Ionian Sea). *Mar. Genomics* **28**: 71–81.

Pearce, D.A., Newsham, K.K., Thorne, M.A.S., Calvo-Bado, L., Krsek, M., Laskaris, P., et al. (2012) Metagenomic analysis of a southern maritime antarctic soil. *Front. Microbiol.* **3**: 403.

Pelikan, C., Herbold, C.W., Hausmann, B., Müller, A.L., Pester, M., Loy, A., et al. (2016) Diversity analysis of sulfite- and sulfate-reducing microorganisms by multiplex dsrA and dsrB amplicon sequencing using new primers and mock community-optimized bioinformatics. *Environ. Microbiol.* **18**: 2994–3009.

Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.

Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2011) Meta-IDBA: a de Novo assembler for metagenomic data. *Bioinformatics* **27**: i94-i101.

Pérez-Jiménez, J. and Kerkhof, L. (2005) Phylogeography of Sulfate-Reducing Bacteria among Disturbed Sediments, Disclosed by Analysis of the Dissimilatory Sulfite Reductase Genes (dsrAB). *Appl. Environ. Microbiol.* **71**: 1004–1011.

Pérez-Ruzafa, A. and Marcos, C. (2012) Fisheries in coastal lagoons: An assumed but poorly researched aspect of the ecology and functioning of coastal lagoons. *Estuar. Coast. Shelf Sci.* **110**: 15–31.

Pester, M., Knorr, K.-H.H., Friedrich, M.W., Wagner, M., and Loy, A. (2012) Sulfate-reducing microorganisms in wetlands - fameless actors in carbon cycling and climate change. *Front. Microbiol.* **3**: 72.

Peterlongo, P. and Chikhi, R. (2012) Mapsembler, targeted and micro assembly of large NGS datasets on a desktop computer. *BMC Bioinformatics* **13**: 48.

Pevzner, P.A., Tang, H., and Waterman, M.S. (2001) An Eulerian path approach to DNA fragment assembly. *Proc. Natl. Acad. Sci. U. S. A.* **98**: 9748.

Pichevin, L., Martinez, P., Bertrand, P., Schneider, R., Giraudeau, J., and Emeis, K. (2005) Nitrogen cycling on the Namibian shelf and slope over the last two climatic cycles: Local and global forcings. *Paleoceanography* **20**: PA2006.

Pielou, E.C. (1969) An introduction to mathematical ecology. Wiley-Interscience, New York.

Pinhassi, J. and Berman, T. (2003) Differential growth response of colony-forming alpha- and gamma-proteobacteria in dilution culture and nutrient addition

experiments from Lake Kinneret (Israel), the eastern Mediterranean Sea, and the Gulf of Eilat. *Appl. Environ. Microbiol.* **69**: 199–211.

Pitcher, A., Villanueva, L., Hopmans, E.C., Schouten, S., Reichart, G.-J., and Sinninghe Damsté, J.S. (2011) Niche segregation of ammonia-oxidizing archaea and anammox bacteria in the Arabian Sea oxygen minimum zone. *ISME J.* **5**: 1896–1904.

Plante, C.J., Coe, K.M., and Plante, R.G. (2008) Isolation of surfactant-resistant bacteria from natural, surfactant-rich marine habitats. *Appl. Environ. Microbiol.* **74**: 5093–5099.

Polymenakou, P.N., Lampadariou, N., Mandalakis, M., and Tselepides, A. (2009) Phylogenetic diversity of sediment bacteria from the southern Cretan margin, Eastern Mediterranean Sea. *Syst. Appl. Microbiol.* **32**: 17–26.

Poulos, S., Collins, M.B., and Ke, X. (1993) Fluvial/wave interaction controls on delta formation for ephemeral rivers discharging into microtidal waters. *Geo-Marine Lett.* **13**: 24–31.

Poulos, S.E., Collins, M.B., and Lykousis, V. (1995) Late Quaternary Evolution of Amvrakikos Gulf, Western Greece. *Geo-Marine Lett.* **15**: 9–16.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., et al. (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res* **42**: D231-D239.

Preheim, S.P., Perrotta, A.R., Martin-Platero, A.M., Gupta, A., Alm, E.J., Perrott, A.R., et al. (2013) Distribution-Based Clustering: Using Ecology to Refine the Operational Taxonomic Unit. *Appl. Environ. Microbiol.* **79**: 6593–6603.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**: 7188–7196.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-D596.

Quince, C., Lanzen, A., Davenport, R.J., and Turnbaugh, P.J. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**: 38.

Rabus, R., Hansen, T.A., and Widdel, F. (2006) Dissimilatory sulfate-and sulfur-reducing prokaryotes. In, *The Prokaryotes*. Springer New York, pp. 659–768.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* **403**: 646–649.

Radajewski, S., McDonald, I.R., and Murrell, J.C. (2003) Stable-isotope probing of nucleic acids: A window to the function of uncultured microorganisms. *Curr. Opin. Biotechnol.* **14**: 296–302.

Ranjard, L., Poly, F., Lata, J.C., Mougel, C., Thioulouse, J., and Nazaret, S. (2001) Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability. *Appl. Environ. Microbiol.* **67**: 4479–4487.

Reddy, T.B., Thomas, A.D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., et al. (2014) The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res* **43**: D1099–D1106.

Reeder, J. and Knight, R. (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat. Methods* **7**: 668–669.

Reizopoulou, S., Thessalou-Legaki, M., Nicolaidou, A., and Biology, M. (1996) Assessment of disturbance in Mediterranean lagoons: An evaluation of methods. *Mar. Biol.* **125**: 189–197.

Remane, A. (1934) Die brackwasserfauna. In, *Verhandlungen der Deutschen Zoologischen Gesellschaft*. Deutschen Zoologischen Gesellschaft im Frankfurt am Main, Leipzig, pp. 34–74.

Rettedal, E.A. and Brözel, V.S. (2015) Characterizing the diversity of active bacteria in soil by comprehensive stable isotope probing of DNA and RNA with H218O. *Microbiologyopen* **4**: 208–219.

Rho, M., Tang, H., and Ye, Y. (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38**: e191.

Richter, D.C., Ott, F., Auch, A.F., Schmid, R., and Huson, D.H. (2011) MetaSim: A Sequencing Simulator for Genomics and Metagenomics. *Handb. Mol. Microb. Ecol. I Metagenomics Complement. Approaches* **3**: 417–421.

Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**: 525–552.

Rodriguez-Valera, F., Ventosa, A., Juez, G., and Imhoff, J.F. (1985) Variation of environmental features and microbial populations with salt concentrations in a multi-pond saltern. *Microb. Ecol.* **11**: 107–115.

Roegner, G.C., Needoba, J.A., and Baptista, A.M. (2011) Coastal Upwelling Supplies Oxygen-Depleted Water to the Columbia River Estuary. *PLoS One* **6**: e18672.

Rolston, A. and Dittmann, S. (2009) The Distribution and Abundance of Macrobenthic Invertebrates in the Murray Mouth and Coorong Lagoons 2006 to 2008 CSIRO: Water for a Healthy Country National Research Flagship, 89., Canberra, Australia.

Ronaghi, M. (2001) Pyrosequencing Sheds Light on DNA Sequencing. *Genome Res.* **11**: 3–11.

Ronaghi, M., Uhlén, M., and Nyrén, P. (1998) A sequencing method based on realtime pyrophosphate. *Science* **281**: 363–365.

Rosen, G.L., Reichenberger, E.R., and Rosenfeld, A.M. (2011) NBC: the Naive Bayes Classification tool webserver for taxonomic classification of metagenomic reads. *Bioinformatics* **27**: 127–129.

Rosen, M.J., Callahan, B.J., Fisher, D.S., and Holmes, S.P. (2012) Denoising PCR-amplified metagenome data. *BMC Bioinformatics* **13**: 283.

Rothberg, J.M. and Leamon, J.H. (2008) The development and impact of 454 sequencing. *Nat. Biotechnol.* **26**: 1117–1124.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., et al. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**: e77.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.

Sakko, A.L. (1998) The influence of the Benguela upwelling system on Namibia's

marine biodiversity. *Biodivers. Conserv.* **7**: 419–433.

Sangwan, N., Lata, P., Dwivedi, V., Singh, A., Niharika, N., Kaur, J., et al. (2012) Comparative metagenomic analysis of soil microbial communities across three hexachlorocyclohexane contamination levels. *PLoS One* **7**: e46219.

Schattner, P., Brooks, A.N., and Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* **33**: W686–W689.

Schidlowski, M. (1989) Evolution of the sulphur cycle in the Precambrian. *In:* Evolution of the global biogeochemical sulphur cycle, 3–19.

Schippers, A., Kock, D., Höft, C., Köweker, G. and Siegert, M. (2012) Quantification of microbial communities in subsurface marine sediments of the Black Sea and off Namibia. *Front. Microbiol.* **3**.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., and Quince, C. (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**: e37–e37.

Schloss, P.D. and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**: 1501–1506.

Schloss, P.D. and Handelsman, J. (2006a) Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl. Environ. Microbiol.* **72**: 6773–6779.

Schloss, P.D. and Handelsman, J. (2006b) Introducing TreeClimber, a test to compare microbial community structures. *Appl. Environ. Microbiol.* **72**: 2379–2384.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**: 7537–7541.

Schnell, S., Bak, F., and Pfennig, N. (1989) Anaerobic degradation of aniline and dihydroxybenzenes by newly isolated sulfate-reducing bacteria and description of Desulfobacterium anilini. *Arch. Microbiol.* **152**: 556–563.

Scholz, M.B., Lo, C.-C., and Chain, P.S. (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr.*

Opin. Biotechnol. **23**: 9–15.

Schulz, H.N., Brinkhoff, T., Ferdelman, T.G., Mariné, M.H., Teske, A., and Jørgensen, B.B. (1999) Dense Populations of a Giant Sulfur Bacterium in Namibian Shelf Sediments. *Science* **284**: 493–495.

Schwarz, J.I.K., Lueders, T., Eckert, W., and Conrad, R. (2007) Identification of acetate-utilizing Bacteria and Archaea in methanogenic profundal sediments of Lake Kinneret (Israel) by stable isotope probing of rRNA. *Environ. Microbiol.* **9**: 223–237.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**: R60.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**: 811–814.

Segnini de B, M.I., Gómez, I., Brito, L., Acosta, V., and Troccoli, L. (2015) Microbial activity in surface sediments of Chacopata-Bocaripo lagoon axis, Sucre State, Venezuela. *Mar. Pollut. Bull.* **91**: 483–490.

Seshadri, R., Kravitz, S.A., Smarr, L., Gilna, P., and Frazier, M. (2007) CAMERA: a community resource for metagenomics. *PLoS Biol* **5**: e75.

Severin, I., Confurius-Guns, V., and Stal, L.J. (2012) Effect of salinity on nitrogenase activity and composition of the active diazotrophic community in intertidal microbial mats. *Arch. Microbiol.* **194**: 483–491.

Shannon, L.V. and O'Toole, M.J. (2003) Sustainability of the Benguela: ex Africa semper aliquid novi. In, Hempel,G. and Sherman,K. (eds), *Large Marine Ecosystems of the World: Trends in Exploitation, Protection and Research*., pp. 227–253.

Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika* **52**: 591–611.

Shin, P.K.S., Leung, J.Y.S., Qiu, J.W., Ang, P.O., Chiu, J.M.Y., Thiyagarajan, V. and Cheung, S.G. (2013) Hypoxia induces abnormal larval development and affects biofilm–larval interaction in the serpulid polychaete Hydroides elegans. *Marine Poll Bull* **76**: 291–297.

Sievert, S.M., Kiene, R.P., and Schulz-Vogt, H.N. (2007) The Sulfur Cycle. In, *Ocenography*., pp. 117–123.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., and Birol, I. (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**: 1117–1123.

Sinclair, L., Ijaz, U.Z., Jensen, L.J., Coolen, M.J.L., Gubry-Rangin, C., Chroňáková, A., et al. (2016) Seqenv: linking sequences to environments through text mining. *PeerJ* **4**: e2690.

Singh, B.K., Bardgett, R.D., Smith, P. and Reay, D.S. (2010) Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nat. Rev. Microbiol.* **8**: 779.

Skarlato, S.O. and Telesh, I.V. (2017) The development of the protistan species-maximum concept for the critical salinity zone. *Russ J Mar Biol* **43**: 1–11.

Smith, A.A., Belda, E., Viari, A., Medigue, C., and Vallenet, D. (2012) The CanOE strategy: integrating genomic and metabolic contexts across multiple prokaryote genomes to find candidate genes for orphan enzymes. *PLoS Comput Biol* **8**: e1002540.

Soergel, D.A.W., Dey, N., Knight, R., and Brenner, S.E. (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.* **6**: 1440–1444.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* **103**: 12115–12120.

Sørensen, J., Christensen, D., and Jørgensen, B.B. (1981) Volatile Fatty acids and hydrogen as substrates for sulfate-reducing bacteria in anaerobic marine sediment. *Appl. Environ. Microbiol.* **42**: 5–11.

Staley, C., Unno, T., Gould, T.J., Jarvis, B., Phillips, J., Cotner, J.B., and Sadowsky, M.J. (2013) Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J. Appl. Microbiol.* **115**: 1147–1158.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stander, G.H. (1964) The Benguela current off South West Africa. *Investig. Rep. Mar. Res. Lab. Southwest Africa* **12**: 1–43.

Stevens, H. and Ulloa, O. (2008) Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environ. Microbiol.* **10**: 1244–1259.

Stewart, F.J., Ulloa, O., and Delong, E.F. (2012) Microbial metatranscriptomics in a permanent marine. **14**: 23–40.

Stramma, L., Johnson, G.C., Sprintall, J., and Mohrholz, V. (2008) Expanding Oxygen-Minimum Zones in the Tropical Oceans. *Science* **320**: 655–658.

Su, X., Pan, W., Song, B., Xu, J., and Ning, K. (2014) Parallel-META 2.0: Enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS One* **9**: 1–13.

Su, X., Xu, J., and Ning, K. (2012) Parallel-META: efficient metagenomic data analysis based on high-performance computation. *BMC Syst. Biol.* **6 Suppl 1**: S16.

Summerhayes, C. (2015) Upwelling. In, *Encyclopedia of Marine Geosciences*. Springer Netherlands, Dordrecht, pp. 1–20.

Sun, Y., Cai, Y., Huse, S.M., Knight, R., Farmerie, W.G., Wang, X., and Mai, V. (2012) A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Brief. Bioinform.* **13**: 107–121.

Sun, Y., Cai, Y., Liu, L., Farrell, M.L., McKendree, W., and Farmerie, W. (2009) ESPRIT: Estimating Species Richness Using Large Collections of 16 rRNA Pyrosequences. *Nucleic Acids Res.* **37**: 1–13.

Sunagawa, S., Mende, D.R., Zeller, G., Izquierdo-Carrasco, F., Berger, S.A., Kultima, J.R., et al. (2013) Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* **10**: 1196–1199.

Suzuki, D., Li, Z., Cui, X., Zhang, C., and Katayama, A. (2014) Reclassification of Desulfobacterium anilini as Desulfatiglans anilini comb. nov. within Desulfatiglans gen. nov., and description of a 4-chlorophenol-degrading sulfate-reducing bacterium, Desulfatiglans parachlorophenolica sp. nov. *Int. J. Syst. Evol. Microbiol.* **64**: 3081–3086.

Sztejrenszus, S.Y. (2016) Effect of humic substances on microbial community composition and iron reduction in marine sediments (Doctoral dissertation, Universität Bremen).

Tagliapietra, D., Sigovini, M., and Magni, P. (2012) Saprobity: a unified view of benthic succession models for coastal lagoons. *Hydrobiologia* **686**: 15–28.

Tagliapietra, D., Sigovini, M., Volpi Ghirardini, A., and Ghirardini, A.V. (2009) A review of terms and definitions to categorise estuaries, lagoons and associated environments. *Mar. Freshw. Res.* **60**: 497–509.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725–2729.

Tang, X., Gao, G., Zhu, L., Chao, J., and Qin, B. (2009) DNA extraction procedure affects organic-aggregate-attached bacterial community profiles from a shallow eutrophic lake. *Can. J. Microbiol.* **55**: 776–782.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E. V, et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E. V (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33–36.

Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., Glockner, F.O., and Glöckner, F.O. (2004) TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics* **5**: 163.

Telesh, I., Schubert, H., and Skarlato, S. (2013) Life in the salinity gradient: Discovering mechanisms behind a new biodiversity pattern. *Estuar. Coast. Shelf Sci.* **135**: 317–327.

Telesh, I., Schubert, H., and Skarlato, S. (2011) Revisiting Remane's concept: evidence for high plankton diversity and a protistan species maximum in the horohalinicum of the Baltic Sea. *Mar. Ecol. Prog. Ser.* **421**: 1–11.

Telesh, I., Schubert, H. and Skarlato, S. (2013) Life in the salinity gradient: discovering mechanisms behind a new biodiversity pattern. *Estuar. Coast. Shelf Sci.* **135**: 317–327.

Telesh, I.V., Schubert, H. and Skarlato, S.O. (2015) Size, seasonality, or salinity: What drives the protistan species maximum in the horohalinicum?. *Estuar. Coast.*

Shelf Sci. **161**:102–111.

Thomas, T., Gilbert, J., and Meyer, F. (2012) Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* **2**: 3.

Thompson, F.L., Bruce, T., Gonzalez, A., Cardoso, A., Clementino, M., Costagliola, M., et al. (2011) Coastal bacterioplankton community diversity along a latitudinal gradient in Latin America by means of V6 tag pyrosequencing. *Arch. Microbiol.* **193**: 105–114.

Treangen, T.J., Koren, S., Sommer, D.D., Liu, B., Astrovskaya, I., Ondov, B., et al. (2013) MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biol* **14**: R2.

Tringe, S.G.S.G., von Mering, C., Kobayashi, A., Salamov, A.A.A.A., Chen, K., Chang, H.W.H.W.H.W., et al. (2005) Comparative metagenomics of microbial communities. *Science* **308**: 554–557.

Tringe, S.S.G. and Hugenholtz, P. (2008) A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. Microbiol.* **11**: 442–446.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.

Ultsch, A. and Moerchen, F. (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. *Tech. Rep. Dep. Math. Comput. Sci. Univ. Marbg.*

Urich, T., Lanzén, A., Stokke, R., Pedersen, R.B., Bayer, C., Thorseth, I.H., et al. (2014) Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics. *Environ. Microbiol.* **16**: 2699–2710.

Vairavamurthy, M.A., Orr, W.L. and Manowitz, B. (1995) Geochemical transformations of sedimentary sulfur: An introduction. *In:* Geochemical Transformations of Sedimentary Sulfur, 1–14.

Vandieken, V. and Thamdrup, B. (2013) Identification of acetate-oxidizing bacteria in a coastal marine surface sediment by RNA-stable isotope probing in anoxic slurries and intact cores. *FEMS Microbiol. Ecol.* **84**: 373–386.

Varsos, C., Patkos, T., Oulas, A., Pavloudi, C., Gougousis, A., Ijaz, U., et al. (2016)

Optimized R functions for analysis of ecological community data using the R virtual laboratory (RvLab). *Biodivers. Data J.* **4**: e8357.

Vasileiadou, K., Pavloudi, C., Kalantzi, I., Apostolaki, E.T., Chatzigeorgiou, G., Chatzinikolaou, E., et al. (2016) Environmental variability and heavy metal concentrations from five lagoons in the Ionian Sea (Amvrakikos Gulf, W Greece). *Biodivers. Data J.* **4**: e8233.

Vasileiadou, K., Pavloudi, C., Sarropoulou, E., Fragopoulou, N., Kotoulas, G., and Arvanitidis, C. (2016) Unique COI haplotypes in hediste diversicolor populations in lagoons adjoining the Ionian Sea. *Aquat. Biol.* **25**: 7–15.

Vasileiadou, K., Sarropoulou, E., Tsigenopoulos, C., Reizopoulou, S., Nicolaidou, A., Orfanidis, S., et al. (2012) Genetic vs community diversity patterns of macrobenthic species: preliminary results from the lagoonal ecosystem. *Transitional Waters Bull.* **6**: 20–33.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J. a, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

Viaroli, P., Bartoli, M., Giordani, G., Naldi, M., Orfanidis, S., and Zaldívar, J.-M. (2008) Community shifts, alternative stable states, biogeochemical controls and feedbacks in eutrophic coastal lagoons: a brief overview. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **18**: S105–S117.

Vishnivetskaya, T.A., Layton, A.C., Lau, M.C.Y., Chauhan, A., Cheng, K.R., Meyers, A.J., et al. (2014) Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. *FEMS Microbiol. Ecol.* **87**: 217–230.

Vitousek, P.M., Cassman, K., Cleveland, C., Crews, T., Field, C.B., Grimm, N.B., et al. (2002) Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry* **57**: 1–45.

Vos, M., Quince, C., Pijl, A.S., de Hollander, M., and Kowalchuk, G. a (2012) A Comparison of rpoB and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity. *PLoS One* **7**: e30600.

Wagner, M., Loy, A., Klein, M., Lee, N., Ramsing, N.B., Stahl, D.A., and Friedrich, M.W. (2005) Functional marker genes for identification of sulfate-reducing

prokaryotes. *Methods Enzymol.* **397**: 469–489.

Wagner, M., Roger, A.J., Flax, J.L., Brusseau, G.A., and Stahl, D.A. (1998) Phylogeny of Dissimilatory Sulfite Reductases Supports an Early Origin of Sulfate Respiration. *J. Bacteriol.* **180**: 2975–2982.

Wang, J. and Gu, J.-D. (2013) Dominance of Candidatus Scalindua species in anammox community revealed in soils with different duration of rice paddy cultivation in Northeast China. *Appl. Microbiol. Biotechnol.* **97**: 1785–1798.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**: 5261–5267.

Wang, Q., Quensen, J.F., Fish, J.A., Lee, T.K., Sun, Y., Tiedje, J.M., and Cole, J.R. (2013) Ecological patterns of nifH genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *MBio* **4**: e00592-e00613.

Wang, X., Wang, Y., Yue, B., Zhang, X., and Liu, S. (2013) The complete mitochondrial genome of the *Bufo tibetanus* (Anura: Bufonidae). *Mitochondrial DNA* **24**: 186–188.

Wang, X., Yao, J., Sun, Y., and Mai, V. (2013) M-pick, a modularity-based method for OTU picking of 16S rRNA sequences. *BMC Bioinformatics* **14**: 43.

Wang, Y., Leung, H., Yiu, S., and Chin, F. (2014) MetaCluster-TA: taxonomic annotation for metagenomic data based on assembly-assisted binning. *BMC Genomics* **15 Suppl 1**: S12.

Wang, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012) MetaCluster 5.0: a two-round binning approach for metagenomic data for low-abundance species in a noisy sample. *Bioinformatics* **28**: i356–i362.

Wang, Y., Sheng, H.-F., He, Y., Wu, J.-Y., Jiang, Y.-X., Tam, N.F.-Y., and Zhou, H.-W. (2012) Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol.* **78**: 8264–8271.

Ward, B.B. and Jensen, M.M. (2014) The microbial nitrogen cycle. *Front. Microbiol.* **5**: 2–3.

Webster, G., Rinna, J., Roussel, E.G., Fry, J.C., Weightman, A.J., Parkes, R.J., et al.

(2010) Prokaryotic functional diversity in different biogeochemical depth zones in tidal sediments of the Severn Estuary, UK, revealed by stable-isotope probing. *FEMS Microbiol. Ecol.* **72**: 179–197.

Weeks, S.J., Currie, B., Bakun, A., and Peard, K.R. (2004) Hydrogen sulphide eruptions in the Atlantic Ocean off southern Africa: implications of a new view based on SeaWiFS satellite imagery. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **51**: 153–172.

Weisman, D., Yasuda, M., and Bowen, J.L. (2013) FunFrame: Functional Gene Ecological Analysis Pipeline. *Bioinformatics* 1–2.

Werner, J.J., Zhou, D., Caporaso, J.G., Knight, R., and Angenent, L.T. (2011) Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *ISME J.* **6**: 1273–1276.

Whiteley, A.S., Manefield, M., and Lueders, T. (2006) Unlocking the "microbial black box" using RNA-based stable isotope probing technologies. *Curr. Opin. Biotechnol.* **17**: 67–71.

Wickham, H. (2009) ggplot2: elegant graphics for data analysis Springer New York.

Wieczorek, S.K. and Todd, C.D. (1998) Inhibition and facilitation of settlement of epifaunal marine invertebrate larvae by microbial biofilm cues. *Biofouling* **12**: 81–118.

de Wit, R. (2008) Microbial diversity in the Bassin d'Arcachon coastal lagoon (SW France). *Hydrobiologia* **611**: 5–15.

Woese, C.R. (1987) Bacterial Evolution. *Microbiol. Rev.* **51**: 221–271.

Wommack, K.E., Bhavsar, J., and Ravel, J. (2008) Metagenomics: read length matters. *Appl. Environ. Microbiol.* **74**: 1453–63.

Wood, D.E. and Salzberg, S.L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**: R46.

Wright, J.J., Konwar, K.M., and Hallam, S.J. (2012) Microbial ecology of expanding oxygen minimum zones. *Nat. Rev. Microbiol.* **10**: 381–394.

Wu, Q.L., Zwart, G., Schauer, M., Kamst-van Agterveld, M.P., and Hahn, M.W. (2006) Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. *Appl. Environ. Microbiol.* **72**: 5478–5485.

Xia, L.C., Cram, J.A., Chen, T., Fuhrman, J.A., and Sun, F. (2011) Accurate Genome Relative Abundance Estimation Based on Shotgun Metagenomic Reads. *PLoS One* **6**: e27992.

Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S.M., Meng, J., et al. (2011) Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J.* **5**: 414–426.

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J., et al. (2012) Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ. Microbiol.* **14**: 2457–2466.

Xu, J. (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol. Ecol.* **15**: 1713–1731.

Yakushev, E. V. and Newton, A. (2012) Introduction: Redox Interfaces in Marine Waters. Springer Berlin Heidelberg, pp. 1–12.

Yamada, T., Waller, A.S., Raes, J., Zelezniak, A., Perchat, N., Perret, A., et al. (2012) Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours. *Mol Syst Biol* **8**: 581.

Yang, S., Liebner, S., Alawi, M., Ebenhöh, O., and Wagner, D. (2014) Taxonomic database and cutoff value for processing mcrA gene 454 pyrosequencing data by MOTHUR. *J. Microbiol. Methods* **103**: 3–5.

Ye, Y. (2011) Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment. *Proceedings. (IEEE. Int. Conf. Bioinformatics Biomed).* **2010**: 153–157.

Ye, Y. and Tang, H. (2009) An ORFome assembly approach to metagenomics sequences analysis. *J. Bioinform. Comput. Biol.* **7**: 455–471.

Yentsch, C.S. and Menzel, D.W. (1963) A method for the determination of phytoplankton chlorophyll and phaeophytin by fluorescence. *Deep Sea Res. Oceanogr. Abstr.* **10**: 221–231.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., et al. (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat. Biotechnol.* **29**: 415–420.

Yu, Y., Breitbart, M., McNairnie, P., and Rohwer, F. (2006) FastGroupII: a web-based

bioinformatics platform for analyses of large 16S rDNA libraries. *BMC Bioinformatics* **7**: 57.

Zehr, J.P. and Kudela, R.M. (2011) Nitrogen cycle of the open ocean: from genes to ecosystems. *Ann. Rev. Mar. Sci.* **3**: 197–225.

Zehr, J.P.J.P.J.P. and Ward, B.B.B. (2002) Nitrogen Cycling in the Ocean: New Perspectives on Processes and Paradigms. *Appl. Environ. Microbiol.* **68**: 1015–1024.

Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–829.

Zhang, J., Yang, Y., Zhao, L., Li, Y., Xie, S., and Liu, Y. (2015) Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. *Appl. Microbiol. Biotechnol.* **99**: 3291–3302.

Zhang, L., Gao, G., Tang, X., and Shao, K. (2014) Can the freshwater bacterial communities shift to the "marine-like" taxa? *J. Basic Microbiol.* **54**: 1264–1272.

Zhang, L., Gao, G., Tang, X., and Shao, K. (2014b) Impacts of different salinities on bacterial biofilm communities in fresh water. *Can. J. Microbiol.* **60**: 319–326.

Zheng, H. and Wu, H. (2010) Short prokaryotic DNA fragment binning using a hierarchical classifier based on linear discriminant analysis and principal component analysis. *J Bioinform Comput Biol* **8**: 995–1011.

Zhu, W., Lomsadze, A., and Borodovsky, M. (2010) Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**: e132.

Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., Welch, D.B.M., et al. (2011) Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. *PLoS One* **6**: e24570.

Zumft, W.G. (1997) Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**: 533–616.

Biosketch

Education and Training

- BSc in Biology (Specialization in Environmental Biology), 01/09/2003 - 27/07/2009

  Aristotle University of Thessaloniki (Greece)

  Overall qualification: 6.75 (very good)
- MSc in Environmental Biology-Management of Terrestrial and Marine Resources, 01/10/2009 - 16/11/2012

  University of Crete, Hellenic Centre for Marine Research, Natural History Museum of Crete (Greece)

  Overall qualification: 9.52 (excellent)
- PhD student at the MARES Joint Doctoral Programme on Marine Ecosystem Health & Conservation, 01/10/2013 - 29/09/2017

  University of Ghent (Belgium), University of Bremen (Germany), Hellenic Centre for Marine Research (Greece)

Scholarships / Awards

- EuroMarine Individual Fellowship for participation to the EMBO Practical Course on Bioinformatics and genome analyses, 5-17 June 2017, Thessalonica (Greece)
- EMBO Short Term Fellowship for a visit to the Microbial Ecophysiology group of the University of Bremen
- FEMS Meeting Grant for participation in the 4th EMBO Workshop on Microbial Sulfur Metabolism, 12-15 April 2015, Helsingør (Denmark)

Selected Seminars / Workshops

- Participation in the EMBO Practical Course on “Breathless microbes: Techniques and theory in anaerobic microbiology”, Wageningen (Netherlands), 24/07/2017 - 28/07/2017
- Participation in the EMBO Practical Course on “Bioinformatics and genome analyses”, Thessalonica (Greece), 05/06/2017 - 17/06/2017

- Participation in the "Dredging & the environment" workshop, Ghent (Belgium), 28/03/2017 - 30/03/2017
- Participation in the Ecosystem Network Analysis (ENA) Tutorial Workshop, Bremen (Germany), 05/09/2016
- Participation in the "MaCuMBA Summer School 2015: Sampling, Isolation & Cultivation of Marine Microorganisms", Texel (The Netherlands), 12/07/2015 - 24/07/2015
- Course on "Likelihood and Bayesian approaches in Biology", instructor: Dr. Mark Holder, Hellenic Centre for Marine Research, Heraklion (Greece), 15/06/2015 - 18/06/2015
- EMBOS Workshop on Pilot Studies, Hellenic Centre for Marine Research, Heraklion (Greece), 15/12/2014 - 19/12/2014
- Participation in the WG 2&3 Hackathon 2 Workshop "From Signals to Environmentally Tagged Sequences III – SeqEnv III", for the COST Action ES1103 – Microbial ecology & the earth system: collaborating for insight and success with the new generation of sequencing tools, Hellenic Centre for Marine Research, Heraklion (Greece), 22/09/2014 - 25/09/2014
- Participation in the "International Summer School in Ecological Data Analysis (ECODAR)", Hellenic Centre for Marine Research, Heraklion (Greece), 07/07/2014 - 12/07/2014
- Participation in the "Micro B3 summer school: from sampling to analyzing microbial diversity & function", Hellenic Centre for Marine Research, Heraklion (Greece), 26/05/2014 - 06/06/2014

Publications

Selected Journal articles

- Faulwetter S, Markantonatou V, **Pavloudi C**, Papageorgiou N, Keklikoglou K, Chatzinikolaou E, Pafilis E, Chatzigeorgiou G, Vasileiadou K, Dailianis T, Fanini L, Koulouri P, Arvanitidis C (2014). Polytraits: A database on biological traits of marine polychaetes. Biodiversity Data Journal 2: e1024. doi: 10.3897/BDJ.2.e1024

- Faulwetter S, Papageorgiou N, Koulouri P, Fanini L, Chatzinikolaou E, Markantonatou V, **Pavloudi C**, Chatzigeorgiou G, Keklikoglou K, Vasileiadou K, Basset A, Pinna M, Rosati I, Reizopoulou S, Nicolaidou A, Arvanitidis C (2015). Resistance of polychaete species and trait patterns to simulated species loss in coastal lagoons. Journal of Sea Research (special issue on Protecting Marine Biodiversity to Preserve Ecosystem Functioning: a Tribute to Carlo Heip), 98: 73-82, doi: 10.1016/j.seares.2014.09.003
- Pafilis E, Frankild SP, Schnetzer J, Fanini L, Faulwetter S, **Pavloudi C**, Vasileiadou A, Leary P, Hammock J, Schulz K, Parr CS, Arvanitidis C, Jensen LJ. (2015) ENVIRONMENTS and EOL: identification of Environment Ontology terms in text and the annotation of the Encyclopedia of Life. Bioinformatics, 31(11):1872-4. doi: 10.1093/bioinformatics/btv045
- Hummel H, Frost M, Juanes JA, Ruci S, Cakmak T, Louizidou P, Andersen SF, Eriksson A, Bordeyne F, Vandenbosch F, Kochmann J, **Pavloudi C**, et al. (2015) A comparison on the degree of implementation of marine biodiversity indicators by European countries in relation to the Marine Strategy Framework Directive (MSFD). Journal of the Marine Biological Association of the United Kingdom (special issue on the 49th EMBS), 95 (8): 1519-1531, doi: 10.1017/S0025315415000235
- **Pavloudi C**, Oulas A, Vasileiadou K, Sarropoulou E, Kotoulas G, Arvanitidis C (2016) Salinity is the major factor influencing the sediment bacterial communities in a Mediterranean lagoonal complex (Amvrakikos Gulf, Ionian Sea). Marine Genomics, 28: 71-81, doi: 10.1016/j.margen.2016.01.005
- Hummel H, van Avesaath P, Wijnhoven S, Kleine Schaars L, Degraer S, Kerckhof F, Bojanic N, Skejic S, Vidjak O, Rousou M, Orav-Kotta H, Kotta J, Jourde J, Pedrotti ML, Leclerc JC, Simon N, Rigaut-Jalabert F, Bachelet G, Lavesque N, Arvanitidis C, **Pavloudi C**, et al (2016) Geographic patterns of biodiversity in European coastal marine benthos. Journal of the Marine Biological Association of the United Kingdom (special issue on the 50th EMBS), doi: 10.1017/S0025315416001119

- Puente A, Guinda X, Juanes J, Echavarri-Erasun B, Ramos E, Degraer S, Kerckhof F, Bojanic N, Rousou M, Orav-Kotta H, Kotta J, Jourde J, Pedrotti ML, Leclerc JC, Simon N, Bachelet G, Lavesque N, Arvanitidis C, **Pavloudi C**, et al (2016) The role of physical variables in biodiversity patterns of intertidal macroalgae along European coasts. Journal of the Marine Biological Association of the United Kingdom (special issue on the 50th EMBS). doi: 10.1017/S0025315416001673
- Dal Bello M, Leclerc JC, Benedetti Cecchi L, Arvanitidis C, van Avesaath P, Bachelet G, Bojanic N, Como S, Coppa S, Coughlan J, Crowe T, Degraer S, Espinosa F, Faulwetter S, Frost M, Guinda X, Jankowska E, Jourde J, Kerckhof F, Kotta J, Lavesque N, De Lucia G, Magni P, de Matos V, Orav-Kotta H, **Pavloudi C**, et al (2016) Consistent patterns of spatial variability between NE Atlantic and Mediterranean rocky shores. Journal of the Marine Biological Association of the United Kingdom (special issue on the 50th EMBS). doi: 10.1017/S0025315416001491
- Kotta J, Orav-Kotta H, Jänes H, Hummel H, Arvanitidis C, van Avesaath P, Bachelet G, Benedetti Cecchi L, Bojanic N, Como S, Coppa S, Coughlan J, Crowe T, Dal Bello M, Degraer S, Juanes J, de Matos V, Espinosa F, Faulwetter S, Frost M, Guinda X, Jankowska E, Jourde J, Kerckhof F, Lavesque N, Leclerc JC, Magni P, **Pavloudi C**, et al (2016) Essence of the patterns of cover and richness of intertidal hard bottom communities: a pan-European study. Journal of the Marine Biological Association of the United Kingdom (special issue on the 50th EMBS). doi: 10.1017/S0025315416001351
- **Pavloudi C**, Christodoulou M, Mavidis M (2016) Macrofaunal assemblages associated with the sponge *Sarcotragus foetidus* Schmidt, 1862 (Porifera: Demospongiae) at the coasts of Cyprus and Greece. Biodiversity Data Journal (LifeWatchGreece collection), 4: e8210, doi: 10.3897/BDJ.4.e8210
- Vasileiadou K, **Pavloudi C**, Kalantzi I, Apostolaki ET, Chatzigeorgiou G, Chatzinikolaou E, Pafilis E, Papageorgiou N, Fanini L, Konstas S, Fragopoulou N, Arvanitidis C (2016) Environmental variability and heavy metal concentrations from five lagoons in the Ionian Sea (Amvrakikos Gulf, W

Greece). Biodiversity Data Journal (LifeWatchGreece collection), 4: e8233, doi: 10.3897/BDJ.4.e8233

- Vasileiadou K, **Pavloudi C**, Sarropoulou E, Fragopoulou N, Kotoulas G, Arvanitidis C (2016) Unique COI haplotypes in *Hediste diversicolor* populations in lagoons adjoining the Ionian Sea. Aquatic Biology, 25: 7-15, doi:10.3354/ab00654
- Varsos C, Patkos T, Oulas A, **Pavloudi C**, Gougousis A, Ijaz UZ, Filiopoulou I, Pattakos N, Vanden Berghe E, Fernández-Guerra A, Faulwetter S, Chatzinikolaou E, Pafilis E, Arvanitidis C (2016) Optimized R functions for analysis of ecological community data using the R virtual laboratory (RvLab). Biodiversity Data Journal (LifeWatchGreece collection), 4: e8357, doi: 10.3897/BDJ.4.e8357
- Sinclair L, Ijaz UZ, Jensen LJ, Coolen MJL, Gubry-Rangin C, Chroňáková A, Oulas A, **Pavloudi C**, Schnetzer J, Weimann A, Ijaz A, Eiler A, Quince C, Pafilis E (2016) Seqenv: linking sequences to environments through text mining. PeerJ 4:e2690, doi: 10.7717/peerj.2690
- Vasileiadou K, **Pavloudi C**, Camisa F, Tsikopoulou I, Fragopoulou N, Kotoulas G, Arvanitidis C (2017) Genetic diversity of Nephtys hombergii (Phyllodocida, Polychaeta) associated with environmental factors in a highly fluctuating ecosystem. Journal of the Marine Biological Association of the United Kingdom. doi: 10.1017/S0025315416001910
- **Pavloudi C**, Faulwetter S, Keklikoglou K, Vasileiadou K, Chatzinikolaou E, Mavraki D, Nikolopoulou M, Bailly N, Rousou M, Kotta J, Bachelet Guy, et al (2016) Taxonomic *vs* functional patterns across European coastal soft-sediment benthic habitats. Journal of the Marine Biological Association of the United Kingdom. *Submitted*

Selected Poster presentations

- Microbial functional diversity and genetic diversity patterns of benthic polychaetes. Christina Pavloudi, Katerina Vasileiadou, Anastasis Oulas, Georgios Kotoulas, Christos Arvanitidis. 50th European Marine Biology Symposium (EMBS), 21-25 September 2015, Helgoland (Germany)

- Correlation of the intraspecific diversity patterns of benthic invertebrates with the microbial community functioning. Katerina Vasileiadou, Christina Pavloudi, Anastasis Oulas, Georgios Kotoulas, Christos Arvanitidis. XV Congress of the European Society for Evolutionary Biology, 10-14 August 2015, Lausanne (Switzerland)
- Sulfate-reducing bacteria in Mediterranean lagoons: similarities and disparities between the different biogeographic areas. Christina Pavloudi, Anastasis Oulas, Katerina Vasileiadou, Christos Arvanitidis. 4th EMBO Workshop on Microbial Sulfur Metabolism, 12-15 April 2015, Helsingør (Denmark)
- Combination of classic biodiversity analyses with novel algorithms which extract environmental information from microbial sequences. Christina Pavloudi, Anastasis Oulas, Katerina Vasileiadou, Christos Arvanitidis. 6th Conference of the Mikrobiokosmos Society, 03-05 April 2015, Athens (Greece)

Selected Oral presentations

- Microbial community interacting with macrobenthic populations in transitional water ecosystems. Christina Pavloudi, Katerina Vasileiadou, Evangelia Chatzinikolaou, Georgios Kotoulas, Mercedes González-Wangüemert, Christos Arvanitidis. ECSA 56 Coastal systems in transition: From a 'natural' to an 'anthropogenically-modified' state, Bremen (Germany)
- Integrated management of coastal ecosystems: a new approach. Katerina Vasileiadou, Christina Pavloudi, Christos Arvanitidis. Euromarine Foresight Symposium “Future Coast – Europe”, Berlin (Germany)
- Prokaryotic abundance, diversity and phoX gene distribution across the Mediterranean Sea. Anastasia Tsiola, Paraskevi Pitta, Tatiana M. Tsagaraki, Christina Pavloudi, Georgios Kotoulas. 2015 Aquatic Sciences Meeting (ASLO), Granada (Spain)

Professional Experience

- 01/01/2016 - 31/10/2016

JERICO-NEXT - Joint European Research Infrastructure network for Coastal Observatory – Novel European eXpertise for coastal observaTories

Taxonomic identification of benthic samples, Application of ecological and molecular techniques in microbial communities

- 26/12/2013 - 18/01/2014

  GENUS - Geochemistry and Ecology of the Namibian Upwelling System

  Participation in the benthic sampling with the Research Vessel Meteor (Cruise M103/1)

- 15/11/2013 - 31/12/2015

  LifeWatchGreece

  Collection and management of biodiversity data

- 01/03/2013 - 30/11/2013

  Encyclopedia of Life (EOL) Rubenstein Fellow 2013 – ENVIRONMENTS: Discovering habitat terms in EOL Contents.

  PhD student – Collaborator

- 01/12/2012 - 14/11/2013

  EUBON - Building the European Biodiversity Observation Network

  Collection and management of biodiversity data, Application of ecological and molecular techniques in microbial communities

Organization of Workshops and Symposia

- EMBOS Workshop on Pilot Studies, 15/12/2014 - 19/12/2014. Member of the organizing committee
- Micro B3 summer school: from sampling to analyzing microbial diversity & function, 26/05/2014 - 06/06/2014. Member of the organizing committee

Participation in proposal writing for research projects

- *ADRIONPort project*

  Submitted at the 1st call of the Interreg Adrion 2014-2020 programme (25/03/2016)

  Approved at the 1st stage of evaluation (Admissibility and Eligibility check)

- *RECONNECT project*

Submitted at the 1st call of the Interreg Balkan-Med 2014-2020 programme (26/04/2016)

Approved for funding

- *PureLag project*

  Submitted at the 2nd call of the Interred Med programme

  Approved at the 1st stage of evaluation (Admissibility and Eligibility check)

  Under evaluation