Reviews and synthesis: Carbon capture and storage monitoring – an integrated biological, biophysical and chemical approach

N. Hicks¹,*, U. Vik²,*, P. Taylor³, E. Ladoukakis⁴, J. Park⁵, F. Kolisis⁴, H. Stahl⁶, and K. S. Jakobsen²

¹Scottish Association for Marine Science, Oban, Argyll, PA37 1QA, UK
²Centre for Ecological and Evolutionary synthesis (CEES), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway
³SAMS Research Services Limited, Lismore Suite, Malin House, The European Marine Science Park, Oban, Argyll, PA37 1S, UK
⁴National Technical University of Athens, School of Chemical Engineering, Laboratory of Biotechnology, 9 Iroon Polytechniou Street, Zografou 157 80, Athens, Greece
⁵Norwegian Geotechnical Institute (NGI), Sognsveien 72, 0855 Oslo, Norway
⁶College of Sustainability Sciences and Humanities, Zayed University, P.O. Box 19282, Dubai, UAE

*These authors contributed equally to this work.
Abstract

Carbon capture and storage (CCS) is a developing technology that seeks to mitigate against the impact of increasing anthropogenic carbon dioxide (CO$_2$) production by capturing CO$_2$ from large point source emitters. After capture the CO$_2$ is compressed and transported to a reservoir where it is stored for geological time scales. Potential leakages from CCS projects, where stored CO$_2$ migrates through the overlaying sediments, are likely to have severe implications on benthic and marine ecosystems. Nonetheless, prokaryotic response to elevated CO$_2$ concentrations has been suggested as one of the first detectable warnings if a CO$_2$ leakage should occur. Applying properties of prokaryotic communities (i.e. community composition and metabolic status) as a novel CO$_2$ monitoring application is highly reliable within a multidisciplinary framework, where deviations from the baseline can easily be identified.

In this paper we review current knowledge about the impact of CO$_2$ leakages on marine sediments from a multidisciplinary-based monitoring perspective. We focus on aspects from the fields of biology, geophysics, and chemistry, and discuss a case study example. We argue the importance of an integrative multidisciplinary approach, incorporating biogeochemistry, geophysics, microbial ecology and modelling, with a particular emphasis on metagenomic techniques and novel bioinformatics, for future CCS monitoring. Within this framework, we consider that an effective CCS monitoring programme will ensure that large-scale leakages with potentially devastating effects for the overlaying ecosystem are avoided. Furthermore, the multidisciplinary approach suggested here for CCS monitoring is generic, and can be adapted to other systems of interest.
1 Sediments and the role of microbial communities

Marine sediments play a vital role in global biogeochemical cycles, particularly in the carbon cycle (Harley et al., 2006; Crain et al., 2008). The biogeochemical processes that occur in these sediments are driven by physical parameters and the presence and metabolic activity of organisms that dwell in and on the sediment surface. The oceans are a huge sink for carbon, and as the carbon reaches the seabed, a large proportion is sequestered in the sediment, particularly in the deep sea sediments where light does not penetrate to the benthos.

Sinking phytodetritus form a significant portion of carbon supply to the benthos, although recent studies suggested that benthic microbes may be a carbon sink in the benthic food web (van Oevelen et al., 2006; Pozzato et al., 2013). Changes in environmental variables such as light, temperature, pH flow and organic matter can modify the contribution of species to ecosystem processes (Biles et al., 2003; Ouellette et al., 2004; Widdicombe and Needham, 2007; Teal et al., 2013), and are likely to have huge implications for benthic systems.

The role of macrofauna in benthic biogeochemical processes (e.g. nutrient flux, oxygen cycling, redox reactions) is well documented, and the presence and activity of macrofauna enhances benthic-pelagic coupling (Bulling et al., 2010). However, many of the processes stimulated by macrofaunal activity, such as nutrient cycling, are mediated by microbial activity (Prosser, 2007; Gilbertson et al., 2012). Whilst macrofaunal impacts are often quantified in terms of biomass or species diversity, estimating biomass for microbial communities are not as informative for determining contribution to key biogeochemical processes. Molecular techniques are increasingly used to infer microbial diversity and abundance, and to identify key genes and metabolic pathways involved in environmental processes from a continuously range of habitats. High throughput sequencing (HTS) methods of amplicons and metagenomics, and subsequent bioinformatics analyses (Fig. 2), have revolutionized the field of microbial ecology in the last decade, enabling assessment of communities of even the most complex...
samples, which harbour millions of species, through nucleic acid sequence identification. Whereas amplicon sequencing is commonly used for identifying members of the communities, metagenomics is increasingly applied to reveal important processes of prokaryotic communities. The technique provides information on important genes, functions and enzymes of the environmental niche. HTS methods have been applied in studies of prokaryotic communities of pockmarks (Haverkamp et al., 2012, 2014); methane seeps (Rike et al., 2013); coal oil point seeps (Håvelsrud et al., 2011); continental margins (Biddle et al., 2008); ocean basins (Biddle et al., 2011); ridges (Lee et al., 2014); hydrothermal vents (Urich et al., 2014) and from natural deep-sea CO$_2$ seeps (Yanagawa et al., 2013). In addition, microbial community response to environmental catastrophes, such as the Deep Water Horizon oil spill (Kimes et al., 2013; Mason et al., 2014), as well as an in situ sub-surface CO$_2$ leak (Tait et al., 2014) have been investigated. The majority of the above mentioned studies represent baseline and observational investigations, describing prokaryotic communities and metabolic pathways of specific environments, and providing a “snapshot” of the communities in time and space. Even though they only give a glimpse of the community at a site, such baseline studies are extremely valuable, as potential follow-up studies can track changes in prokaryotic communities related to possible environmental changes (Håvelsrud et al., 2013). Furthermore, the studies of microbial communities response to environmental disasters clearly demonstrates that HTS based methods can have many applications regarding environmental monitoring. It is, however, extremely important to link sequence-based data to gathered meta-data, so that the distribution of community members and their metabolic potential can be related to environmental conditions. In 2014, Parkes et al. performed a meta-analysis of 32 independent nucleic acid-based studies from marine sediments. The meta analysis demonstrated that the analysed prokaryotic communities were linked to sediment type or geographic province, likely reflecting site specific geochemical and physical conditions, such as oxygen, sulphate, methane hydrate, organic and inorganic carbon content, mineralogy water and sediment depth. Furthermore, the few HTS studies that have investigated prokaryotic
communities and metabolic pathways along environmental gradients in marine sediments have all detected changes corresponding to these gradients (Table 1). Mason et al. (2014) investigated the impact of oil deposition of the Deep Water Horizon spill on sediment surface microbial communities at distances of 0.3–256 km from the wellhead. They found that the composition of prokaryotic communities largely correlated with the content of total petroleum hydrocarbons and inorganic nitrogen concentrations. Furthermore, the deposition of hydrocarbons from the oil spill increased the metabolic repertoire of the microorganisms, particular those responsible for degradation of alkanes and aromatic hydrocarbons. However, the function or metabolic potential of microbial communities does not necessarily always increase with added concentrations of compounds or gases. Yanagawa et al. (2013) studied the metabolically active microbial communities of sediments in a gradient from naturally occurring high CO₂-seeps to hydrothermally unaffected sediments. They found that the density, diversity and metabolic activity of the sediments’ microbial communities decreased toward the CO₂ rich deeper zones. Recently, Tait et al. (2014) studied the response of surface microbes to an in situ CO₂ release experiment using HTS of 16S rRNA amplicons, and found that there was an increase in abundance of 16S rRNA per gram sediments, accompanied by changes in the activity of bacterial taxa, that could be detected after 14 days of gas release both at the epicentre and as far as 25 m away from the epicentre. Such studies clearly illustrate that microbial communities’ rapid response, in terms of compositional and functional turnover, to environmental shifts is measurable by use of HTS methods.

2 Consequences of increased CO₂ in the ocean

Rising atmospheric CO₂ levels have been directly linked to an increase in acidity in seawater (ocean acidification, OA) due to the ocean’s ability to absorb ~ 30 % of this CO₂ (Doney et al., 2009; Drinkwater et al., 2010) consequently reducing the ocean pH (Feely et al., 2008). A growing body of research on the effects of OA has demonstrated the predominantly negative effects on marine organisms (Kroeker et al., 2010,
2013), and ecosystem processes such as primary production and nutrient cycling (Brown et al., 2010; Dossena et al., 2012). To date, most research has focused on “open ocean” species and ecosystems (Duarte et al., 2013). The effects of elevated CO$_2$ on benthic systems and their contribution to biogeochemical cycling remain less understood, with the exception of a few studies which have focused on macrofaunal impacts (Bulling et al., 2010; Hicks et al., 2011; Godbold and Solan, 2013) but lacked an in depth benthic microbial investigation. A recent study simulated CCS conditions (discussed in Sect. 6 of this paper), and included an integrated microbial analysis of the benthos, but this was limited to biomass and abundance indices using PCR (Tait et al., 2014). The few studies dedicated to microbial effects of OA have been limited to pelagic or symbiotic bacterial communities (Webster et al., 2013; Endres et al., 2014) and based on changes in bacterial abundance. However, a complete understanding of microbial response to changes in CO$_2$ levels would require an integration of microbial ecology and biogeochemistry. Whilst initial CO$_2$ focused research examined effects of OA from an initial atmospheric source of CO$_2$, another growing area of CO$_2$ research has stemmed from the Carbon Capture and Storage (CCS) field.

3 Prokaryotic response to elevated CO$_2$ levels

Prokaryotes are known for their versatile metabolism, including many species that are chemolithoautotrophs, which are capable of assimilating energy from a chemical reaction based on inorganic substrates as electron donors. In anaerobic marine sediments, a large fraction of the prokaryotes are chemolithoautotrophs, and several of these are able to assimilate CO$_2$ into organic carbon. To date, six metabolic pathways that assimilate CO$_2$ into cellular material have been identified in prokaryotes (Table 2). Three of these are anaerobic, and only one of these metabolic pathways is shared between bacteria and archaea.

Prokaryotic response to elevated CO$_2$ levels have been studied using HTS methods in terrestrial (He et al., 2010) and marine environments (Yanagawa et al., 2013; Tait et al., 2014).
et al., 2014; Taylor J. D. et al., 2014). In all of these studies, the prokaryotes respond to elevated CO\textsubscript{2} levels by altered community structure and changes in their functional repertoire. Tait et al. (2014) found that there were detectable shifts in prokaryote abundances 14 days into an in situ CO\textsubscript{2} release experiment. The properties of prokaryotic response to elevated CO\textsubscript{2} concentrations have values that extend beyond basic research. Håvelsrud (et al., 2013) suggested that changes in prokaryotic sediment communities might be among the first detectable warnings if a CO\textsubscript{2} leakage should occur. In a baseline study overlaying a potential CO\textsubscript{2} storage site in the North Sea, Håvelsrud and colleagues identified prokaryotic taxa and genes known to be involved in CO\textsubscript{2} fixation using a shot-gun metagenomic approach (Håvelsrud et al., 2013). Six of the 27 most abundant taxa identified in this study were known to fixate CO\textsubscript{2}, and furthermore enzymes belonging to the three aerobic carbon fixation pathways (the Wood Ljungdahl pathway, the reductive tricarboxylic acid cycle and the 3-hydroxypropionate/4-hydroxybutyrate cycle) were also identified. This information can be applied in CCS monitoring approaches, where specific taxa and metabolic pathways associated with CO\textsubscript{2} assimilation can be targeted and sequenced using HTS methods, and deviations from a baseline can be detected.

4 Carbon Capture Storage (CCS)

CCS is a rapidly developing technology that seeks to mitigate against the impact of anthropogenic CO\textsubscript{2} production by capturing CO\textsubscript{2} from large point source emitters, such as power stations or cement works. After capture, the CO\textsubscript{2} is compressed and transported to a reservoir where it is stored for geological time scales (IPCC, 2005).

Such reservoirs typically consist of deeply buried porous and permeable rock, which is blanketed by at least one layer of impermeable rock, creating a seal commonly known as the “cap rock”. On the most basic level, the reservoir, the cap rock and the “overburden” – the rock between the cap rock and the Earth’s surface – is shaped in one of a range of possible geometries, each of which means that any fluid injected into the
reservoir and which tends to migrate upwards due to buoyancy will be trapped inside the structure (Fig. 1). The most obvious of these is the syncline, typically discussed in introductory texts on oil and gas exploration (Hunt, 1996).

It is tautology that if compressed CO₂ (either as a compressed gas or as a supercritical fluid) is to be injected into a reservoir for long-term storage, then certain criteria must be met by the reservoir. Firstly, the reservoir must conform to injectivity requirements – for any given project, the maximum fluid injection rate must be higher than the maximum CO₂ capture rate. Secondly, the reservoir must be sealed well enough that the risk of gas escape, for example through a fractured cap rock, is acceptably low, or at least within tolerable probability over defined timescales. Thirdly, the injection process itself must not operate out-with the acceptable parameters of the reservoir and the cap rock seal – any injection will cause a pressure wave or disturbance which will propagate through the reservoir in time. This may come close to, but must not exceed, the fracture pressure of either the reservoir or the cap rock seal. Over time, it is probable that injected CO₂ will firstly dissolve in pre-extant pore fluids within the reservoir (whether this is saline briny water or hydrocarbons) and may eventually, over tens of thousands of years, precipitate onto sediment grains within the reservoir as carbonate (Halland et al., 2014).

Aside from the environmental benefits for the development of CCS there is also at least one strong financial motivation; enhanced oil recovery (EOR). When producing oil from a reservoir, it is common practice to inject a fluid into the reservoir (IPCC, 2005). This fluid (commonly sea water or brine from a saline aquifer) serves two purposes: firstly it replaces a volume of oil that has been extracted and thus serves to maintain reservoir pressure and aid production, but the second reason is that it can also be used to “sweep” oil from distant areas of the reservoir towards production wells. Injecting CO₂ as a substitute for the fluid will fulfil both of these purposes, but as stated above, it will also dissolve in the oil. This reduces the viscosity of the oil and its surface tension, allowing greater production of the reservoir, with estimates as high as approximately 25% more oil (IPCC, 2005).
All aspects of CCS, from anticipated capture volume to storage capacity, are linked not only to the immediate site where the CO₂ is injected and stored, but also to the whole subsurface (i.e. reservoir formation, caprock, under-/overburden, seabed sediment and seawater-column – see Fig. 1). This results in various mechanical and biogeochemical responses at each section of the subsurface and at different time scales. CCS projects may have a positive or negative effect on these processes, which in turn will have a knock on effect on the services they provide to society. It is imperative that set protocols are clearly implemented and rigorously followed to ensure maximum advantage (e.g. reduction of CO₂ emissions and utilisation of captured CO₂ for EOR) and minimum disruption to the environment and subsurface.

5 Consequences of a CCS leakage or accident

Despite the many precautions that will undoubtedly be taken prior to implementation of CCS, there remains a possibility of a leak from an injection facility, whether this is caused by an unexpectedly weak cap rock or through anthropogenic conduits through the cap rock, such as pre-existing oil production wells, exploration wells, water production facilities and so on (Bachu and Watson, 2009). Such leakages have implications for the wider environment, whether the facility is land based or off-shore. To minimise the risk of large-scale environmental change being initiated by a long-term leak, it is imperative that operators of such facilities set up and maintain extensive monitoring programmes to quickly identify such leaks. The first step will be that of geophysical investigation and routine monitoring. The sub-surface geology should be detailed by seismic, electromagnetic and gravity surveys to a high level of detail prior to initiation of CCS injection (Arts et al., 2008; Park et al., 2014). These surveys should be regularly repeated to detect changes in the geology of the reservoir, the cap rock and the overburden indicative of CO₂. This will be supplemented by injectivity tests and pressure checks of existing conduits.
However, this is only one part of what should be a comprehensive and multi-disciplinary monitoring programme, particularly for off-shore facilities, due to the unique challenges of access and detection presented by the marine environment. Current monitoring programmes rely heavily on modelling predictions (Blackford et al., 2014a) and largely lack thorough in-situ measurements. A multifaceted approach should include physical and chemical data on the sea floor sediments, the ecology and biology of the benthos as well as metagenome data that describes and characterizes changes in the composition and the activity of the benthic microbial community assemblage. Such a monitoring programme may be more sensitive to small scale, or incipient leaks due to the responses of certain parameters, allowing high-resolution and early detection aspects to any such monitoring project.

6 Environmental impacts of CCS – a case study

The quantification of the environmental impact of leakages from CCS is not trivial (Blackford et al., 2014b). There have been many attempts to use environmental modelling techniques to examine the probable impact of a leak from a sub seabed CCS reservoir (Blackford et al., 2008, 2013). However, the modelling approach has several well-known weaknesses regarding the examination of ecological response of, and biological impacts on, key species in response to excess CO$_2$, particularly where detailed species level information is lacking or contradictory (Widdicombe and Spicer, 2008). More precise and observational data can be gathered from mesocosm or experimental based studies, where CO$_2$ treatments are initiated and the impact of a strictly controlled environmental parameter is observed and quantified with a high degree of accuracy (Payán et al., 2012a, b; Queirós et al., 2015). However, these experimental methods have several shortcomings in as much as they are tightly constrained and cannot replicate the complexity of the natural environment, so care must be taken when interpreting the results (Benton et al., 2007).
Environmental parameters and benthic community composition have been studied around natural CO₂ seeps (Caramanna et al., 2011, 2013), but they too are not ideal for the accurate quantification of environmental impact of a new CO₂ seep. They are volcanic in origin and the gas is often contaminated with other compounds such as methane or hydrogen sulphide (Pearce, 2006; Voltattorni et al., 2009). They are also pre-existing natural phenomena and therefore a baseline study cannot be conducted to accurately quantify their impact. Additionally, since these systems are volcanic in nature, it is difficult to stop the gas flow, and therefore it is impossible to study the rate of recovery after an impact, unless a site stops releasing CO₂ by chance while under observation.

A field scale experiment was designed and conducted that simulated the impact of CO₂ leaking from a sub-seabed reservoir (Taylor P. et al., 2014). This experiment validated monitoring programmes and investigated the sensitivity of various monitoring designs to change caused by the CO₂ release (Blackford et al., 2014a, b).

The experiment took place in the summer of 2012 on the west coast of Scotland (Blackford and Kita, 2013; Blackford et al., 2014b; Taylor P. et al., 2014), and began with the drilling of a well from a land location into seabed sediments. The well terminated in 11 m of sediment with the seafloor 12 m below mean sea level. Over the course of 37 days, 4200 kg of CO₂ gas was injected into the seabed sediments (Taylor P. et al., 2014). Changes in benthic processes and characteristics (Lichtschlag et al., 2014; Taylor P. et al., 2014b); behaviour of macrofauna species (Kita et al., 2014); and microbial response (Tait et al., 2014) were examined. These parameters were compared between the release site and a reference site some 450 m distant from the CO₂ release point, and unaffected by the released CO₂ (Lichtschlag et al., 2014) and studied for up to a year after the initial release of CO₂ (Taylor P. et al., 2014b). Furthermore, the following monitoring programmes were used (Blackford et al., 2014b); geophysical monitoring of gas propagation (Cevatoglu et al., 2014); and modelling of bubble dynamics was carried out (Dewar et al., 2014, 2013).
The role of metagenomics in CCS monitoring systems

Although traditional biomass measurements can provide a large quantity of information, as seen in the CCS case study (Tait et al., 2014), these methods of ascertaining microbial community response alone are not sufficient for CCS monitoring. Investigating the prokaryote community composition and metabolic pathways using DNA provides a cutting edge technique that build on established and accepted molecular methods, including HTS, and combined with metadata provides further information that is key to identifying small scale changes in microbial response e.g. to a CO₂ leak. The main advantage of HTS methods is to provide far more detailed data on microbial assemblages and their subsequent response to environmental change.

The wealth of genomic information acquired from a metagenomic analysis is extracted by means of DNA sequencing and subsequent bioinformatics analyses (i.e. full documentation of the nucleotide sequences that constitute the metagenome). Sequencing techniques have evolved rapidly in the last decades (Metzker, 2010), and to date it is possible to exploit high-throughput protocols to achieve exceptionally high yields for only a fraction of the cost of traditional processes (see Sanger et al., 1977). This results in generation of huge amounts of data. However, due to the corresponding increase in the amount of data produced as techniques increase yield, it has become necessary to implement protocols of bioinformatic tasks for the complicated effort of handling such big datasets. This challenge is, for simplicity, illustrated by a six-step metagenomic pipeline, including quality control, assembly, gene detection and gene annotation, followed by taxonomic analysis (see Fig. 2).

To reduce errors through user handling, the results produced from following this protocol should be electronically stored (on a backed up computer database) to allow advanced data handling and meta-processing. As the datasets get larger, the complexity and time entailed in completing each of the bioinformatic tasks also increases. To counter this, bioinformatics pipelines have been developed that can automate these
workflows through user-friendly interfaces, making it easier to handle and analyze these datasets (as detailed in Fig. 2).

Current bioinformatic approaches rely on platforms introducing automated workflows that incorporate and run analytical tools in a consecutive order. This need for an automated solution became apparent at the beginning of large scale sequencing projects, resulting in the development of specific pipelines (Almeida et al., 2004; Harrington et al., 2010; Angiuoli et al., 2011) dedicated to the analysis of genomic data from a single organism. However, since the advent of HTS technologies and the subsequent prevalence of metagenomic projects, many of these solutions quickly became inadequate in handling the magnitude and complexity of the generated data. The current bioinformatic arsenal of pipelines able to take up the challenge of analyzing a metagenomic sequencing dataset include numerous tools that are already available (Meyer et al., 2008; Arumugam et al., 2010; Angiuoli et al., 2011; Pilalis et al., 2012; Markowitz et al., 2014) and many others currently in development. Since each pipeline has specific benefits and drawbacks, it remains up to the researcher to decide upon the most appropriate one based on the type of data and subsequent required analysis.

Such knowledge of bioinformatics and bioinformatics pipelines can be applied to an integrated CCS monitoring system, allowing collection of environmental samples (e.g. sediments from a CCS site) and through use of a bioinformatics pipeline, can identify whether certain bacterial assemblages, such as those that favor elevated CO₂ conditions, are present. Automated bioinformatics pipelines make analytical tools available for novice users, providing researchers with an advantage over other sequencing techniques, and thus can be modified for use within a CCS monitoring programme. This means a simple sediment sample from a CCS site can be analysed using these HTS methods and metagenomics knowledge, and could indicate the presence of a CO₂ leak.
8 Future developments

Monitoring CCS projects through an integrated metagenomic multidisciplinary approach has enormous potential, and can be implemented in marine and terrestrial subsurface CCS projects world-wide. Furthermore, metagenomic approaches have a vast potential in a wide array of other environmental monitoring applications, such as hydrocarbon detection and effects of oil spills, pollutive agents and environmental effects, both in terrestrial and marine environments. For CCS monitoring in particular, there are, however, still several issues that need to be addressed. These include a cautious optimization and standardization of molecular methods, excluding as many as possible of the known biases associated with nucleic acid extraction, PCR amplification, sequencing, and an automatisation of these molecular methods into a user-friendly setting or even better into a single instrument. Such an instrument may be designed to include all steps from sampling to analysis of the samples. Furthermore, CCS specific automated bioinformatic pipelines need to be developed for assessment of the prokaryotic communities. Lastly, accurate migration models of the stored CO$_2$ should also be established and applied in conjunction with direct observational techniques.

A major challenge in order to develop monitoring methods that can be applied to a wide range of CCS projects is the large heterogeneity of sediments. To overcome this challenge, targeted baseline investigations are of pivotal importance. As development of sampling procedures are likely to improve and become more effective and cheaper, and CCS specific automated and low-threshold bioinformatic workflows for molecular analyses are developed, we argue that such a multidisciplinary approach will be of significant value in future monitoring of CCS projects.

Author contributions. N. Hicks, H. Stahl and P. Taylor reviewed all benthic CCS literature and reviewed the case study. U. Vik and K. S. Jakobsen reviewed the microbial literature and developed the start of the bioinformatics pipeline, E. Ladoukakis and F. Kolisis developed the bioinformatics pipeline and incorporated the metagenomics experience. J. Park developed the monitoring models and reviewed the current monitoring literature. N. Hicks and U. Vik prepared the manuscript with contributions from all co-authors.
Acknowledgements. The authors would like to thank A. G. Rike for initiating and writing the project proposal for COVERALL and the Research Council of Norway for funding (grant number 230433).

References


Caramanna, G., Voltattorni, N., and Maroto Valer, M. M.: Is Panarea Island (Italy) a valid and cost-effective natural laboratory for the development of detection and monitoring tech-

8926


Park, J., Vanneste, M., Bohloli, B., Viken, I., and Bjørnarå, T. I.: In Situ Resistivity of CO$_2$ Plume at Sleipner from CSEM and Gravity Data, EAGE Publications BV, Stavanger, 2014.


Table 1. Overview of high throughput sequencing studies where marine prokaryotic communities have been investigated along gradients.

<table>
<thead>
<tr>
<th>Gradient</th>
<th>Location</th>
<th># samples</th>
<th>Sample type</th>
<th>Extraction type</th>
<th>Method</th>
<th>Type of study</th>
<th>Sequenced region</th>
<th>Platform</th>
<th>Bioinformatics</th>
<th>Key finding</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical gradient</td>
<td>Peru Margin</td>
<td>1 Core, 4 depths</td>
<td>Sediment samples</td>
<td>MoBio UltraClean Microbial DNA kit</td>
<td>qPCR</td>
<td>Whole genome amplification (WGA)</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td>Significant changes with depth in community composition</td>
<td>Biddle et al. (2008)</td>
</tr>
<tr>
<td>Temporal and spatial CO₂ gradient</td>
<td>Ardmucknish Bay, Scotland</td>
<td>4 sites, 6 timepoints, 5 replicates</td>
<td>Sediment samples</td>
<td>RNA PowerSoil Total RNA Isolation kit</td>
<td>qPCR</td>
<td>16S rRNA</td>
<td>V1-V3 (27Fmod and 519Rmodbio)</td>
<td>454</td>
<td>Qlime</td>
<td>Increases in abundance of bacterial, archaeal and cyanobacterial 16S rRNA during CO₂ release in zone 1 and 2</td>
<td>Tait et al. (2014)</td>
</tr>
<tr>
<td>Seawater pH/CO₂ gradient</td>
<td>Levante Bay, Italy</td>
<td>3 sites (6 samples per site) 18 samples</td>
<td>Biofilms on pyroclastic rock</td>
<td>PowerSoil DNA isolation kit</td>
<td>PCR</td>
<td>16S rRNA</td>
<td>V4 (515F-806R)</td>
<td>Ion Torrent PGM</td>
<td>Higher number of OTUs at high levels of CO₂</td>
<td>Taylor-P. et al. (2014a)</td>
<td></td>
</tr>
<tr>
<td>Seawater pCO₂</td>
<td>Levante Bay</td>
<td>3 sites (7 samples at each site)</td>
<td>Sediment bacterial communities</td>
<td>PowerSoil DNA isolation kit</td>
<td>PCR</td>
<td>16S rRNA</td>
<td>V1-V3</td>
<td>454</td>
<td>Qlime</td>
<td>Mothur</td>
<td>Kerfashi et al. (2014)</td>
</tr>
<tr>
<td>High CO₂ seepage sites to unaffected marine sediments</td>
<td>Yanaguni Knoll IV hydrothermal field</td>
<td>5 sites</td>
<td>Metabolically active microbial communities in CO₂-seep sediment samples</td>
<td>RNA PowerSoil Total RNA Isolation kit</td>
<td>PCR</td>
<td>16s rRNA</td>
<td>EUB27F, EUB338Rmix and UNIV530F-mix and ARCO912Rmix</td>
<td>454</td>
<td>RDP</td>
<td>Less bacterial diversity at high CO₂ seepage sites compared to those of low CO₂ seepage sites</td>
<td>Yanagawa et al. (2013)</td>
</tr>
</tbody>
</table>
Table 2. Overview of the six CO₂ assimilation metabolic pathways.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Alternative name</th>
<th>CO₂-fixing enzymes</th>
<th>Aerobe/ anaerobe</th>
<th>Domain</th>
<th>Found in members of (ref)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calvin-Benson-Bassham cycle</td>
<td>Reductive pentose phospahet cycle</td>
<td>RubisCO</td>
<td>Aerobe</td>
<td>Bacteria</td>
<td>Cyanobacteria, alpha-proteobacteria, beta-proteobacteria, γ-proteobacteria, Firmicutes, Chloroflexi</td>
</tr>
<tr>
<td>Arnon-Buchanan cycle</td>
<td>Reductive citric acid cycle/rTCA</td>
<td>2-Oxoglutarate synthase</td>
<td>Anaerobe</td>
<td>Nitrospirea, Aquificales, Chlorobiales, ε-proteobacteria, δ-proteobacteria, γ-proteobacteria, α-proteobacteria</td>
<td></td>
</tr>
<tr>
<td>Wood-Ljungdahl pathway</td>
<td>Reductive acetyl-CoA pathway</td>
<td>Acetyl-CoA synthase-CO dehydrogenase</td>
<td>Anaerobe</td>
<td>Bacteria</td>
<td>Spirochaetes, Firmicutes, Plactomycetes, δ-Proteobacteria, Euryarchaeota</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Formylmethanofuran dehydrogenase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pyruvate synthase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-Hydroxy-propionate bicycle</td>
<td>Acetyl-CoA and propionyl-CoA carboxylase</td>
<td>Acetyl-CoA and propionyl-CoA carboxylase</td>
<td>Aerobe</td>
<td>Bacteria</td>
<td>Chloroflexi</td>
</tr>
<tr>
<td>3-Hydroxypropionate-4-hydroxybutyrate cycle</td>
<td>Acetyl-CoA and propionyl-CoA carboxylase</td>
<td>Acetyl-CoA and propionyl-CoA carboxylase</td>
<td>Aerobe</td>
<td>Archaea</td>
<td>Crenarchaeota</td>
</tr>
<tr>
<td>Dicarboxylate-4-hydroxybutyrate cycle</td>
<td>Pyruvate synthase PEP carboxylase</td>
<td>Pyruvate synthase PEP carboxylase</td>
<td>Anaerobe</td>
<td>Archaea</td>
<td>Crenarchaeota</td>
</tr>
</tbody>
</table>
Figure 1. A typical offshore CCS schematic – note the “overburden” includes all rock layers over the cap rock, from the top of the storage formation up to the surface of the seabed. Image courtesy of SCCS®.
The flow chart to the left illustrates important steps of HTS analyses of microbial communities using an amplicon approach (left flow) and a metagenomic approach (right flow), and highlights the complexity introduced in each step. First of all, the heterogeneity of each sediment sample (1) may facilitate for millions prokaryotes in each gram of sediment. Crushing and homogenization of the sample makes the DNA of the prokaryotic communities more accessible for DNA extraction (2). After the DNA is extracted, the amplicon workflow proceeds with amplification of targeted sequences (barcodes) of the DNA in a PCR (3). These barcodes are optimally invariable within taxa but variable between taxa. The PCR-amplification step (3) introduces biases, including primer bias and amplification bias due to e.g. length differences and differences in AT/GC content. Moreover, during PCR the polymerase enzyme makes errors and introduces 'artificial' mutations and combines different templates into chimeric sequences. During the high throughput sequencing (HTS; 4) other errors are introduced into the sequence (depending on sequencing method). The resulting raw sequences (5) therefore needs to be quality-checked and errors that are of non-biological origin needs to be filtered out during the quality control (6). Sequence clustering (7) is a crucial step that currently imposes challenges, as taxa delimitation in nature does not follow the exact strict criteria set in the analyses (such as % sequence similarity). In other words, one species may be form several clusters (i.e. splitting), or several species may appear in the one cluster (i.e. lumping). During the taxonomic annotation process (8) the sequence clusters are annotated with taxonomic information through e.g. BLAST. This annotation is also hampered with several difficulties, including poor representation of many taxa in the reference sequence libraries. The metagenomic workflow following the DNA extraction (2) progresses with a random fragmentation of the extracted DNA (9) before the fragments are sequenced on a HTS platform (4). After the raw sequences (5) have been quality controlled (6), fragments originating from the same taxa may be assembled into larger segments (contigs; 10). At this step, the sequencing depth is crucial, as deeper sequencing will result in more overlaps between different fragments, which again will increase the reliability of the resulting contigs. At step 11, the content of the contigs are searched to identify regions that encode genes. The resulting genes are then assigned functions (12) through BLAST searching against existing databases as well as through pattern based and novel machine-learning methodologies that do not rely on homology identification. Lastly, particular regions of the metagenome may provide an additional means of assigning the sequences to particular taxa (13).

**Figure 2.** Overview of the workflow of high throughput sequencing of amplicons and metagenomes.
Archaea – Domain of life constituting prokaryotic microorganisms

Automated workflow — see bioinformatic pipeline

Benthic-pelagic coupling – processes that occur over the sediment-water interface e.g. biogeochemical cycling

Bioinformatics – A sub-discipline of biology and computer science concerned with the acquisition, storage, analysis, and dissemination of biological data, most often DNA and amino acid sequences.

Bioinformatic pipeline — A set of bioinformatic tasks that are configured to run consecutively in an automated way.

Cap rock – an impermeable formation located above a storage formation that prevents injected CO$_2$ from escaping or leaking.

Cap rock/well integrity – Used to estimate and analyze the geomechanical properties and behaviors of cap rock and well in the context of risk analysis (e.g. leakage, drilling hazards).

CCS – Carbone Dioxide Capture and Storage or Sequestration.

Chemolithoauotrophes – organisms that utilize chemicals (chemo) from the bedrock (litho) as an energy source for making their own (auto) food (troph).

Enzyme – A biological catalyst and is almost always a protein. It speeds up the rate of a specific chemical reaction in the cell. A cell contains thousands of different types of enzyme molecules, each specific to a particular chemical reaction.

High throughput sequencing (HTS) – Sequencing where more than one sample can be processed at the same time, often applied or sequencing methods such as Illumina, 454, PacBio.

Injectivity – A measure of how much fluid is injected for a given time and pressure, defined as the ratio of CO$_2$ injection rate to net injection pressure in wellbore.

Macrofauna – invertebrates that live within or on the sediment or hard substrate; often classified by size (often defined as organisms greater than 500mm)

Mesocosm – container/tank used as an experimental tool to manipulate and control the natural environment

Metabolic pathway – series of biochemical reactions occurring within a cell

Metagenomics – The study of genetic material from mixed templates, such as from environmental samples.

Meta-processing — Data processing that involves handling and filtering of large datasets, advanced search queries and statistical analysis.

Microbe – A common term for single celled microscopic organisms, from all domains of life.

Nucleic acid – DNA and RNA

Overburden – Denotes all formations above a storage formation up to the top surface or seabed/seafloor.

PCR amplification — Polymerase chain reaction (PCR) - a laboratory technique used to amplify DNA sequences by using short DNA sequences (primers) to select the portion of the genome to be amplified.

Phytodetritus – organic matter consisting of dead/dying photosynthetic organisms (e.g. phytoplankton) that falls from the surface to the seabed (benthos)

Prokaryote – Single-celled organisms from the two domains of life: Bacteria and Archaea.

Storage formation – a reservoir that is used to store any kind of fluids or waste (e.g. cutting injection, captured CO$_2$, etc.)

Supercritical CO$_2$ – A fluid state of carbon dioxide where it is held at or above its critical temperature (304.25 K) and critical pressure (72.9 atm or 7.39 MPa).

Underburden – Denotes all formations below a reservoir or storage formation.

Figure 3. Glossary.