Metabolic roles of uncultivated bacterioplankton lineages in the Northern Gulf of Mexico ”dead zone”

**General information from article**

*Dead zone* = marine regions with low dissolved oxygen (DO) concentrations. One of the largest of these dead zones occur in the northern Gulf of Mexico (nGOM). This results from the eutrophication (övergödning) of bacterioplanktion respiration and stratification. Microbial respiration reduces oxygen to levels unfit for many animals (=hypoxia). Hypoxia results from oxygen consumption by aerobic microbes combined with stratification that prevents reoxygenation of bottom waters, where the microbes are fueled by organic matter from phytoplankton absorbing nitrogen input. The hypoxic zones have become more widespread due to spread of nitrogen-based fertilizers. nGOM is the largest eutrophication-driven hypoxic zones in the world, making it fit as a model system. The nitrogen runoff from Mississippi and Atchafalaya Rivers lead to this bottom water hypoxia. But there has not been much study of the microbial metabolisms underlying this phenomenon.

Some bacterial groups live in these regions, but little is known about their metabolic roles in the ecosystem. The authors made a metagenomic analysis to assemble genomes of bacteria found in the different regions of the dead zone. They use metatranscriptiomic data to assess their metabolic activity. They thereby use a coupled shotgun metagenomic and metatranscriptomic approach to assess the metabolic potential of: Marine Group II Euryarchaeota (MGII), SAR406 and SAR202. These are poorly understood “microbial dark matter” lineages. Previous studies showed that Thaumarchaea dominated in nGOM, and also these lineages. The basic functions of these groups have become more clear, but they might contain sublineages with specific metabolic roles. None of these have been studied in detail in shallow coastal waters and in context of seasonal hypoxia. Authors want to investigate:

* Specific contribution of the lineages to biogeochemical cycling in nGOM during hypoxia
* Evaluate the relative similarity of these organisms to their counterparts elsewhere

**Result**

They selected 6 samples with varying DO: from suboxic to oxic (more oxygen). Initial assembly and binning recovered 76 genomes. 20 genomes were assigned to uncultivated microbial dark matter groups. They defined MGII, SAR406 and SAR202 into sublineages based on average aa identity, GC content, clade structure in ribosomal protein tree and 16S rRNA genes. In their metagenomic data MGII were more abundant in lower oxygen samples than fully oxic samples. The majority of genomes encoded for in the metagenomic data was aerobic, chemoheterotrophic metabolism and with no genes for nitrogen/sulfur respiration expect for nitrite reductase in one genome. They especially found in the MGII cytochrome c oxidase expression.

After assembling and binning data they receieved 20 bins and they associated different genome characteristics of those. Such as how complete it is, % contaminants, number of scaffolds, longest scaffold, size, no of genes, GC-content and so on. Carbohydrate active enzyme (CAZy) genes provide information about relationship between microbes and carbon sources. They found a few, such as glycosyltransferases GT, with activities related to cellular synthesis. This could thus indicate actively growing cells.

SAR406 represented more than 5% of population in some locations during hypoxia. They found greater metagenomic reads in suboxic samples relative to dysoxic or oxic samples. SAR406 genomes encoded a capacity for aerobic respiration. They found expression of cytC oxidases even in lowest oxygen samples. SAR406 can be divided into group A and B where B encodes both high- and low-affinity cytC oxidases. High-affinity oxidases were not found in group A – indication of sublineage-specific optimization for different oxygen regimes.

They have reconstructed the metabolism of the 3 groups, based on 3-4 complete genomes. They have also made a figure showing the expression of predicted respiratory genes and expression of CAZy genes.

They recovered almost 3 complete SAR202 Chloroflexi genomes (still present in lower abundance than MGII and SAR406). The SAR202 genome had more major facilitator superfamily transporters than others. The MFS genes transport numerous substrates such as sugars, amino acids. They can through coupling with ion gradient be associated with uptake or export of compounds. The genomes had number of duplicated genes, many genes were annotated as galactonate dehydratase which catalyze a step in pathway to utilize D-galactonate in central carbon metabolism. The large number of genes in these categories likely indicate some divergence for alterantive roles. The SAR202 had low amount of CAZy genes relative to other genomes.

The most abundant organisms were MGII, SAR406 and SAR202 clades – but they also recovered genomes from several groups that were previously undetected in nGOM or very rare. These taxa might not contribute to the biomass for populous (folkrik) clades, their genomes provide important insight into their functional potential during hypoxia. The bin 13 genome had highet relative activity compared to other genomes – low abundance does not automatically equate to low metabolic impact. Bin 13 had aerobic respiration with both high- and low-affinity cytC oxidases. The low-affinity oxidases contributed more reads in the samples where we could detect expression. They reconstructed the candidate phylum members in bin 13. Many bin 13 genes were among the most highly expressed in all samples.

Bin 50 and 48 were lower in abundance than SAR202 genomes, with no observable trend associated with oxygen levels. They investigated expression of genes in samples containing more oxygen and less. They encoded motility, aerobic respiration, glycolysis. The bin 50 genome was among the most active in the analysis, it had the highest expression of cytC oxidase. Bin 48 and 50 had abundant CAZy genes, suggesting highly flexible metabolic repertoire for carbon acquisition. They contain capacity for breakdown of carbon-based substrates.

**Discussion**

They define roles for MGII, SAR406, SAR202, Bin 13 and Bins 50/48 as aerobic heterotrophs. They also observed concurrent expression of genes associated with anaerobic metabolism in SAR406, SAR202, MGII and Bins 50/48 in suboxic samples with lowest DO concentrations. An organism’s set of CAZy genes often gives insights into its nutrition sensing and acquisition. All taxa examined in the study had predicted chemoorganoheterotrophic metabolism. The CAZy genes foundin the genomes suggest that SAR406, SAR202, Bin 13 and 50/48 participate in degradation of organic matter. This matches the general model of hypoxic zone oxygen consumption resulting from sinking organic matter provided by algal blooms in surface waters.

MGII is a broadly distributed archael clade with members found in marine and sedimentary environments. Work during 2012 and 2013 hypoxia indicated a spread of archaeal taxa both in Thaumarchaea and MGII phyla. They found many MGII in lower oxygen sample in hypoxic zone – surprising since they are often found in aerobic environments. However, oxygen was still present even in the lowest DO samples. Previously, MGII have been shown to be dominant in water environments associated with blooms in productivity. The nGOM MGII appear to metabolically similar to those. The availability of organic matter, thought to be substrates for MGII, probably explains their abundance.

SAR406 have previously shown to be present in marine and sedimentary. They are commonly found in deeper ocean waters and prefer lower oxygen concentrations. Their data define roles for SAR406 in the eutrophication-driven hypoxia of the nGOM: their data confirm the predicted aerobic metabolism and sulfur reduction. SAR406 are likely to degrade cellulose, starch, agar, peptidoglycan. During nGOM hypoxia, SAR406 members degrade complex carbohydrates fueled by aerobic respiration and are supplied with anaerobic respiration of nitrate, nitrite and sulfur.

Members of the SAR202 clade of chloroflexi inhabit a variety of marine environments. The nGOM hypoxia SAR202 genomes had CAZy genes involved in degradation of chitin and pectin.

Bins 50/48 provide novel genome data for bacterioplankton in nGOM hypoxia. The ribosomal protein tree provides evidence that these taxa belongs to Latescibacteria. The bins represented the largest genome, and contained numerous of genes suggestion degradation of complex organic matter. Since cytC oxidase genes were found in the bins, these organisms do then likely have aerobic, potentially facultatively anaerobic, chemoorganoheterotrophic metabolism with roles in complex carbon degradation. Bin 50/48 predict aerobic metabolism as well, although only with low-affinity cytC oxidases.

**Information important to analysis**

The study reconstructs multiple genomes from uncultivated bacterioplankton during nGOM hypoxia. They had 6 samples representing hypoxic (4) and oxic (2) dissolved oxygen concentrations. They had different stations: D1, D2, D3, E2, E2A and E4. They collected nucleic acids (DNA, RNA) from these stations. DNA and RNA were sequenced separately with Illumina HiSeq. DNA sequencing resulted in 400 million reads that were reduced after adaptors were removed using Scythe. Low-quality reads were trimmed with Sickle. Metagenomic reads from all samples were pooled, assembled and binned. Binning produced 76 genomes, where 20 were assigned to lineages with uncultivated representatives using checkM, ribosomal protein trees and 16S gene sequences.

The Metagenomic and metatranscriptomic sequencing reads from samples were mapped to binned contigs using BWA. Bins were examined for contamination and completeness with CheckM,

Taxonomy for each bin was assigned using ribosomal protein tree. For bins not having enough ribosomal proteins, assignments were made based on marker gene tree as part of the checkM analysis. 16S rRNA genes were identified via CheckM and werew aligned using BLASTN.

The dataset worked on by the authors is too big so I will work on a subset of the data.

**Steps in analysis**

1. Metagenomic assembly
2. Binning of assembly
3. Phylogenetic placement of bins
4. RNA expression analysis of assembled genomes

**Summary**

**Aim**

The goal was to define the metabolic roles of uncultivated bacterioplankton lineages in nGOM. The dead zones are increasing in number and severity around the globe with deleterious effects on ecology and economics. It’s therefore of interest to get better insight into these areas.

**Type of biological data**

They had 6 samples with hypoxic and oxic dissolved oxygen concentrations from different stations. They collected seawater where DNA and RNA were extracted from by placing a filter in a lysing matrix. They collected all nucleic acid they could find – metagenomes and metatranscriptomes. They had different levels, come from another paper.

Took sequences from earlier study. For this study they measured nutrients. They used DO levels.

Metadata – all additional information expect for the sequences. Conditions during the sampling process. Data describing data.

**Type of sequencing data**

They collected both DNA and RNA data. Sequencing with Illumina to generate 100 bp reads. The reads were then trimmed and the low-quality reads were removed - which reduced the number of reads. Metagenomic reads were after that pooled, assembled and binned. Abundance of taxa in samples was quantified by evaluating mapped reads using RPKM – reads per kilobase per million.

Reconstruction of genome – metagenomic pool of sequences with separate genomes that is a critical step. That is the assembly and binning.

Sort reads depending on property.

The first step are the most ciritcal, everything builds on that.

Metabolic analysis is probably the most relevant analysis. This also includes the annotation of genes and metabolic construction. Without this you can’t do any analysis.

Assembly gives contigs and binning gives genomes.

The assembly will use all reads and produce some contigs. The binning produced genomes.

**Most relevant analyses**

**Conclusions**

**A lot was aerobic prokaryotic – surprising. The level was still too high for anaerobic organisms to survive. A lot of them had the ability to switch between aerobic and anaerobic**

**Most used nitrogen and sulfate as energy source.**

**They didn’t have a goal to find a single conclusions.**

Subset of data: We’re only using some % of the total amount of reads.

**Terms**

Anthropogenic activity = antropogena processer är sådana som kan härledas ur mänskliga aktiviteter.

nGOM = northern Gulf of Mexico

Eutrophication = övergödning

Hypoxia = refers to low oxygen conditions - oxygen levels unfit for many animals. It’s when DO < 2 mg l-1 / ~62.5 micromole kg-1. Oxygen depletion is a phenomenon that occurs in acquatic environments as dissolved oxygen (DO; molecular oxygen dissolved in water) becomes reduced to a point where it becomes detrimental (skadlig) to aquatic organisms living in the system. Hypoxia leads to reduction of reproduction of fish.

Anoxia = total depletion (reducering) of level of oxygen. Extreme form of hypoxia. An aquatic system lacking DO is termed anaerobic, reducing or anoxic. A system with low concentration, between 1-30% saturation, is called hypoxic or dysoxic.

Oxygen depletion can result from number of factors, but mos toften is due to pollution and eutrophication in which plant nutrients and phytoplankton blooms are encouraged. Phytoplankton through photosyntehsis will raise DO saturation, the dense population of a bloom reduces the DO saturation during the night by respiration. When phytoplankton cells die they sink to the bottom and are decomposed by bacteria – something that further reduces DO. Another is stratification.

Stratification = the arrangements of vegetation in layers. The vertical layering of a habitat. Stratification could be that lakes develop two discrete layers of water of different temperatures. Seasonal stratification?

Plankton = freely circulating organisms in water that lack active mobility or whose movement capacity is so insignificant that they depend on the currents to move. Microscopic animals, algae and bacteria are often referred to as plankton. Bacterioplankton referes to the bacterial component of the plankton that drifts in the water.

autochthonous = autokton, term för de som härstammar från ett specifikt område.

Prevalent = allmänt förekommande

Benthic zone = ecological region at the lowest level of water such as ocean, lake, stream.

Benthic contribution = bottenbidrag

Euphotic zone = layer closer to surface that receives enough light for photosynthesis to occur. Beneath lies the disphotic zone, which is illuminated so poorly that rates of respiration exceed those of photosynthesis.

Uncultivated = icke-odlade

Cytochrome C oxidase = transmembrane protein complex found in bacteria, archaea and eukaryotes (mitochondria). It’s the last enzyme in the respiratory electron transport chain of cells. It receives an electron from each of four cytochrome C molecules and converting the molecular oxygen to two molecules of water.

CAZy = carbohydrate-active enzymes. Enzymes involved in the synthesis, metabolism and recognition of complex carbohydrates.

concatenated = sammanbunden, ihoplänkad