

## Final report VISTA 2012

### Project title

### **Subsurface metagenomics, functional microbial diversity analysis and gene discovery in deep and hot petroleum reservoirs**

Project director:	Birkeland, Nils-Kåre, University of Bergen
Post-doc/ scholar:	Keen, Jeffrey Thomas
Project duration:	15.04.2009 – 14.04-2012
Technical contact person in Statoil:	Kotlar, Hans Kristian
Division head:	Kotlar, Hans Kristian
Project number:	6504

#### **1. En ½ A4 side med populærfremstilling av prosjektet**

Deep petroleum reservoirs represent an extreme microbial environment shown to harbor a diversity of anaerobic and thermophilic bacteria and archaea from a variety of physiological groups of which the vast majority have not been cultivated and characterized. The main objectives of this project was to obtain a better understanding of the diversity, genetics and physiology of these extremophilic microorganisms as well as to explore their possible biotechnological potential mainly by using a metagenomic approach. Extraction of suitable amounts of DNA of sufficient quality for metagenomic analyses from oil-water mixtures produced by the Valhall and Troll C petroleum reservoirs turned out to be much more difficult than anticipated although this has been done in the past<sup>1</sup>. After more than two years of work, we finally managed to purify small amounts of DNA from Troll C production waters, but prior to 454 sequence analysis we had to amplify it using a commercial Whole Genome Amplification kit, with a possibility for PCR bias. The sequencing resulted in >465 000 reads with a total of >229 Megabases (Mbp). Following assembly, 1920 large contigs were obtained with an N50 value of 12590, the largest contig being 137987 bases, and with a total of 7 Mbp of assembled sequences. Use of MG-RAST for annotation and analysis identified 77% of the genes as annotated protein-coding genes of known function, 19.5% as hypothetical genes and 3.1% (87) as previously unknown genes. 95.5% of the genes were scored as belonging to Bacteria, while 3.6% to the Archaeal domain (mostly methanogens and Thermococci). The bacterial annotations belonged phylogenetically to a large range of taxa, but most genes were affiliated with Gram-positive groups (>42%) (mostly Thermoanaerobacterales and Clostridiales) and a range of Proteobacteria (>35%) and Thermotogae (6.7%). A more thorough analysis of the sequences is underway.

*Anaerophaga thermohalophila* is a moderate thermophilic gram-negative fermentative bacterium able to grow at salt concentrations up to 12% and believed to be a promising organism for microbially enhanced oil recovery. Through molecular analyses, we have previously identified this organism in oil well production waters. We succeeded in isolating a strain of this species from a separation tank on the Valhall platform. Genome sequencing was, however, difficult, yielding very short reads, and thus low genome coverage. For comparison with the *A. thermohalophila* type strain, we also did genome sequencing of this isolate, as deposited in DSMZ. The sequencing of the type strain was much more successful, yielding 4.3 Mbp in 212 contigs. Genome analysis and comparison between the strains is underway.

---

<sup>1</sup> Dahle et al. (2008): ANTON LEEUW INT J G, 93; 47-49

**2. Har det skjedd endringer i målsetning underveis?/ Oppnådde en det som ble lovet?**

Due to very low yield of DNA from oil-well production waters the metagenomic sequence analyses were significantly delayed. We therefore did not fulfill our aims of the project, i.e. to identify genes and products that can be useful for the petroleum industry. This was also caused by an extensively long queue for commercial sequencing at the Norwegian DNA sequencing centre in Oslo.

**3. Publikasjoner**

Results have not yet been published, but we are planning to publish two papers:

1. A genome announcement in *J. Bact.* on the *Anaerophaga thermohalophilum* genome sequence, and
2. A paper on the Troll C oil-reservoir metagenome.

**4. Refleksjoner om videreføring (fra prosjektleder)**

The genome and metagenome data sets are currently being further analysed using bioinformatics, and the project is thus being continued by the project leader using his research group resources. The 7 Mb of metagenomic sequence data that we obtained at the end of the project period represents only a small fraction of the complete Troll C metagenome, and further sequencing should therefore be performed in order to maximize the outcome of this project and to better justify the resources used. For this purpose we would need at least one man year and running costs for additional sequencing.