FINAL CRUISE REPORT

May 6, 2013

1. Cruise Particulars

U.S. Dept. of State CRUISE No.:	DOS File #: 2009-025
Ship Name:	S/V *Sorcerer II*
Operating Institute or Agency:	J. Craig Venter Institute
Project Title:	Sorcerer II Global Expedition http://www.jcvi.org/cms/research/projects/gos/overview/
Cruise Dates (Inclusive):	Aug 1 - Sept 15, 2009
Chief Scientist:	Dr. J. Craig Venter
Affiliation	J. Craig Venter Institute 9704 Medical Center Drive Rockville, MD 20850
Phone/Fax	(Ph) +1-858-200-1810
E-mail	rfriedman@jcvi.org
Research program sponsors	Beyster Family Foundation Fund of the San Diego Foundation; Life Technologies Foundation.

2. Participating Personnel Dr. J. Craig Venter (Chief Scientist), Jeffrey Hoffman (Scientist), Charles Howard (Captain), John Henke (First Mate), Jeremiah Niles (Deck hand), Sarah Dyste (Permitting), Karen McNish (Chef), Karolina Innbergs (Scientist).

3. Program Overview and Objectives:

Since 2003, the J. Craig Venter Institute has been engaged in a Global Ocean Sampling Expedition to explore the taxonomic and functional biodiversity of the microorganisms that inhabit the world's oceans. The research results included in this Final Cruise Report are part of *The J. Robert Beyster and Life Technologies Foundation 2009-2010 Research Voyage,* a two year Expedition to the Baltic, Mediterranean, and Black Seas.

Over the course of the two year Expedition, microbial samples were collected by filtration from over 300 aquatic environments in 12 nations and from international waters. The associated environmental data (e.g., salinity and dissolved oxygen) indicate that many of these samples are dramatically different from those collected during the GOS Circumnavigation in 2003-2005. Using improved computational biology platforms, we observed dramatic changes in the types of organisms found across these gradients. For example, a massive shift from alphaproteobacteria to betaproteobacteria was observed with the transition from marine to freshwater systems, but with little functional change. Oxygen gradients were associated with

both phylogenetic and functional shifts. Assembly-based analyses have also functionally characterized upwards of 100 previously unobserved and still uncultivated microbes.

At the conclusion of this project, we will have extensively cataloged the microbial diversity and function of nearly all global oceans and marginal seas. We have also begun to develop a mechanistic understanding for the distribution of microbes and function that could be applied to all aquatic environments.

Sampling: Microbial samples were collected by serial filtration from over 300 environmental sites (black cruise track, below). Data on numerous environmental characteristics (such as temperature, salinity, oxygen, photosynthetic biomass, photosynthetic activity, sulfide, nitrogen, phosphorus, and silica) were also collected at each site.

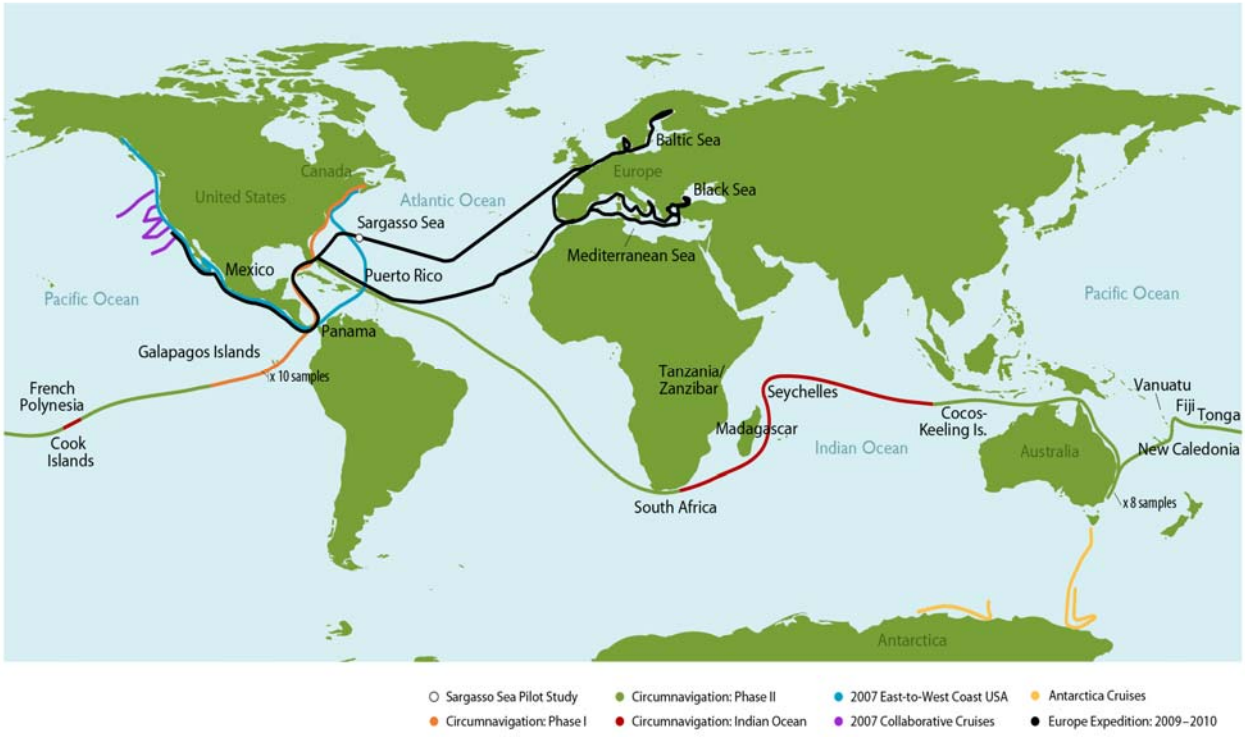

Figure 1: The J. Robert Beyster and Life Technologies Foundation 2009-2010 Research Voyage of the Sorcerer II Expedition (black cruise track) displayed with previous voyages of JCVI's Global Ocean Sampling (GOS) Expedition.

The European GOS expedition substantially expanded the diversity of aquatic environments that have been sampled. From a simple perspective of temperature and salinity, select samples from the Baltic and Black Sea are clearly environmentally distinct from samples from the Pacific and Atlantic oceans. (Figure 2, below.)

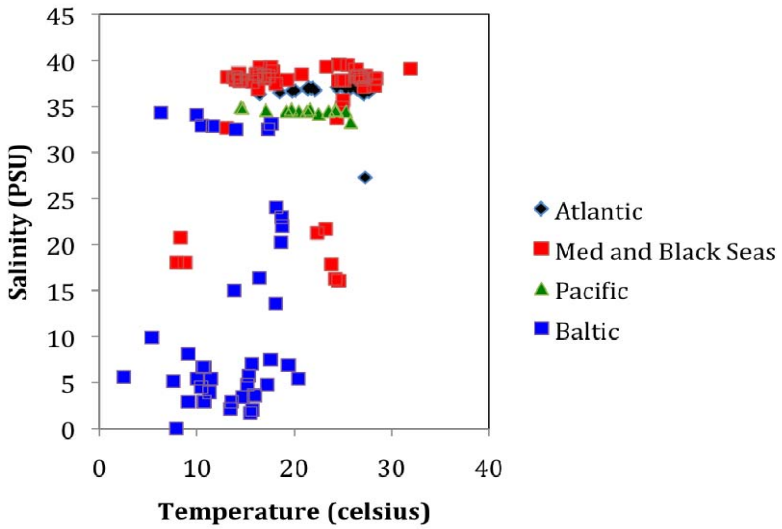

Figure 2: Temperature and salinity of a subset of GOS samples from the Atlantic and Pacific oceans, along with the Baltic, Mediterranean, and Black Seas.

Sequencing: Over fifty full plates of 454 sequencing have generated over 200 metagenomes from 63 GOS sites, with the Solid platform being tested on 3 sites. The GOS sites were chosen both to foster collaborations and maximize the environmental diversity of the samples. These samples include collaborative sampling efforts with the University of Stockholm, Plymouth Marine Laboratory, the Max Plank Institute, the Alfred Wengener Polar Institute, the University of Konstanz, the Center for Advanced Studies-Blanes, Universitat de Girona, Universidad Miguel Hernandez , Institute de Ciencies del Mar en Barcelona, University of Athens, and Stazione Zoologica Anton Dohrn of Naples, in addition to samples from international waters in the Atlantic Ocean.

Sequencing efforts were weighted towards metagenomic sequencing, though a substantial proportion has been dedicated to metatranscriptomic sequencing, which can simultaneously provide phylogenetic (who is there) and functional (what are they doing) information.

In collaboration with Life Technologies, an early test of the Ion Torrent PGM platform was conducted on a sample where a 454 metagenome was available for comparison. The PGM produced far more data at a lower cost and also generated a large set of assembled DNA. Encouraged by these results, the PGM is currently being used for deep sequencing of metagenomes from Spanish coastal waters.

Bioinformatic tools and analyses: Next-generation sequencing technologies produce large volumes of data at very low sequencing cost; however, the sequencing reads produced by these technologies are shorter and contain larger number of errors compared to earlier generation Sanger sequencing. Coupled with the

complexity of many microbial communities, metagenomic sequence data presents unique computational and analytical challenges.

To analyze these data, we have developed an extensive, high-throughput metagenomic data analysis infrastructure. Our infrastructure includes tools for: (i) assembling metagenomic data, (ii) mapping sequence data to genomes or metagenomes (iii) identification and taxonomic classification of 16S genes and over 30 other phylogenetic markers, (iv) identification of protein coding genes and novel gene families, (v) functional annotation of predicted genes, (vi) comparison of metagenomic samples via their taxonomic or functional profiles, (vii) exploring genome structure variation and genomic diversity of populations represented in a metagenomic sample, and (viii) linking sequence based observations to the environmental variables measured during sample collection.

We have developed JCVI Metagenomics Reports (METAREP), an open source application that provides a graphical web browser interface for analyzing and comparing massive sets of annotated metagenomes. The ease of use is such that a computational biology background is not required to visualize and analyze "omic" datasets once they have been annotated. The web interface also allows for efficient sharing of data with collaborators who often do not have the required computational background or resources to deal with raw sequence data.

For all of the sequencing to date, automated functional and phylogenetic annotation of all sequences has been completed and extensive manual interpretations have been conducted.

Publications reporting results from the two year European Expedition are currently being written. Results from the first leg of the Global Ocean Sampling Expedition were published in a March 2007 in a special edition of PLoS Biology (http://www.ploscollections.org/article/browseIssue.action?issue=info%3Adoi%2F10.1371%2Fissue.pcol.v06.i02). The "PLoS Biology: Ocean Metagenomics Collection" contains three scientific articles analyzing 6.3 billion base pairs of sequence data generated from the first 41 sites. The samples were collected as Sorcerer II sailed from Canada down the east coast of the USA, through the Panama Canal to the Pacific Ocean and into French Polynesia.

4. Research Activities in Waters of Norway

Barcode	Latitude (DM)	Longitude (DM)	Time Date	Depth (m)	Temp. (C)	Salinity (PSU)	Oxygen (umol/Kg)	pH	Chlorophyll (ug chl a/L)
692	59d27.10396'N	10d32.09702'E	7/29/2009 2:00	0.3	18.74	22	5.7	8.5	2.2
693	59d27.10396'N	10d32.09702'E	7/29/2009 2:00	55	10.48	32.94	6.3	8.2	0.36

Table 1: Sample locations and metadata.

Figure 3: Chart showing samples collected in Norway maritime boundaries.

At each station a Seabird CTD that can measure temperature, depth, salinity, pH, chlorophyll, and oxygen concentration down to 120 meters in depth was lowered into the water. Typically, at each station, a 400L sample was collected at the surface and a 200L water sample was collected from below the surface based on the CTD water column profile. The collected microbes were size fractionated by serial filtration through 200 µm nytex net, 3 µm, 0.8 µm, and 0.1 µm membrane filters, and finally a 50 Kda cut-off tangential flow filter. The filters, with the captured organisms, were placed in a -80 ^{o}C freezer on the research vessel until transport back

to the laboratory. On return to the United States, the filters were subjected to enzymatic lysis to collect the DNA. Extracted DNA was sequenced using a combination of sequencing technologies.

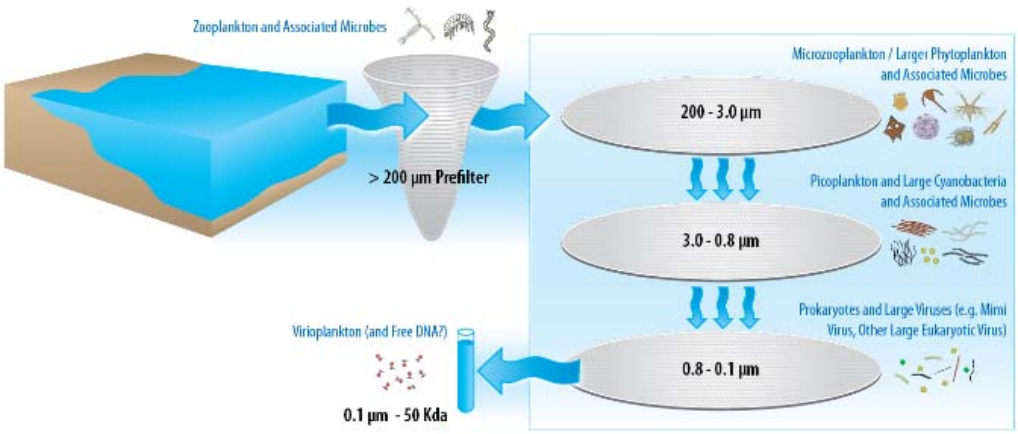

Figure 4: Representation of serial filtration process of water samples.

5. Results

To date no samples collected in the waters of the Norway have been sequenced and analyzed. We have provided analysis from 71 Baltic Sea samples that have been sequenced and analyzed as part of this study that demonstrate the microbial diversity in that region.

Sample Site Description	Sample Id	Sample Depth (m)	Sample Habitat	Filter Size (µm)
Sweden Bottenhavets-Station C-3	GS659	1	coastal	0.8
Sweden Bottenhavets-Station C-3	GS659	1	coastal	3
Sweden Bottenhavets-Station C-3	GS659	1	coastal	0.1
Sweden Bottenhavets, Station C-3	GS660	12	coastal	3
Sweden Bottenhavets, Station C-3	GS660	12	coastal	0.1
Sweden Bottenhavets, Station C-3	GS660	12	coastal	0.8
Sweden Bottenhavets, Station C-3	GS660	12	coastal	<0.1
Sweden Bottenviken, Station A5	GS665	1	coastal	0.8
Sweden Bottenviken, Station A5	GS665	1	coastal	3
Sweden Bottenviken, Station A5	GS665	1	coastal	0.1
Sweden Bottenviken, Station A5	GS666	10	coastal	0.1
Sweden Bottenviken, Station A5	GS666	10	coastal	0.8
Sweden Bottenviken, Station A5	GS666	10	coastal	3
Sweden Bottenviken, Station A5	GS666	10	coastal	<0.1
Sweden Lake TorneTrask, Abisko	GS667	0.3	lake sample	0.8
Sweden Lake TorneTrask, Abisko	GS667	0.3	lake sample	3
Sweden Lake TorneTrask, Abisko	GS667	0.3	lake sample	0.1
Sweden Lake TorneTrask, Abisko	GS667	0.3	lake sample	<0.1
Sweden N.Kvarken, Station B-3	GS673	0.3	coastal	0.8
Sweden N.Kvarken, Station B-3	GS673	0.3	coastal	3
Sweden N.Kvarken, Station B-3	GS673	0.3	coastal	0.1
Sweden N. Kvarken, Station B-3	GS674	16	coastal	0.1
Sweden N. Kvarken, Station B-3	GS674	16	coastal	0.8
Sweden N. Kvarken, Station B-3	GS674	16	coastal	3
Sweden Helcom site 4	GS677	9	coastal	0.1
Sweden Helcom site 4	GS677	9	coastal	0.8
Sweden Helcom site 4	GS677	9	coastal	3
Sweden Helcom site 4	GS678	74	coastal	0.1
Sweden Helcom site 4	GS678	74	coastal	<0.1
Sweden Helcom site 4	GS678	74	coastal	0.8
Sweden Helcom site 4	GS678	74	coastal	3
Sweden Off south tip of Oland, bloom occuring	GS679	0.3	coastal	3
Sweden Off south tip of Oland, bloom occuring	GS679	0.3	coastal	<0.1

Sample Site Description	Sample Id	Sample Depth (m)	sample habitat	Filter Size (µm)
Sweden Off south tip of Oland, bloom occuring	GS679	0.3	coastal	0.8
Sweden Off south tip of Oland, bloom occuring	GS679	0.3	coastal	0.1
Sweden Off south tip of Oland, bloom occuring	GS680	4	coastal	3
Germany German Baltic Sample-IOW Station	GS681	0.3	coastal	0.1
Germany German Baltic Sample-IOW Station	GS681	0.3	coastal	0.8
Germany German Baltic Sample-IOW Station	GS681	0.3	coastal	3
Germany German Baltic Sample. IOW Station	GS682	24	coastal	0.1
Germany German Baltic Sample. IOW Station	GS682	24	coastal	0.8
Germany German Baltic Sample. IOW Station	GS682	24	coastal	3
Denmark Danish sample site DMU 939	GS683	0.3	coastal	0.1
Denmark Danish sample site DMU 939	GS683	0.3	coastal	3
Denmark Danish sample site DMU 939	GS683	0.3	coastal	0.8
Denmark Danish Sample Site DMU 939	GS684	15	coastal	3
Denmark Danish Sample Site DMU 939	GS684	15	coastal	0.1
Denmark Danish Sample Site DMU 939	GS684	15	coastal	0.8
Sweden Kattegat-Helcom site	GS685	0.3	coastal	0.1
Sweden Kattegat-Helcom site	GS685	0.3	coastal	0.8
Sweden Kattegat-Helcom site	GS685	0.3	coastal	3
Sweden Kattegat-Helcom site	GS686	20	coastal	0.1
Sweden Kattegat-Helcom site	GS686	20	coastal	0.8
Sweden Kattegat-Helcom site	GS686	20	coastal	3
Sweden Lysekil Fjord-Alsback Station	GS687	0.3	coastal	0.1
Sweden Lysekil Fjord-Alsback Station	GS687	0.3	coastal	0.8
Sweden Lysekil Fjord-Alsback Station	GS687	0.3	coastal	3

Sample Site Description	Sample Id	Sample Depth (m)	sample habitat	Filter Size (µm)
Sweden Lysekil Fjord-Alsback Station	GS688	30	coastal	0.1
Sweden Lysekil Fjord-Alsback Station	GS688	30	coastal	0.8
Sweden Lysekil Fjord-Alsback Station	GS688	30	coastal	3
Sweden Lysekil Fjord- Alsback Station	GS689	72	coastal	0.1
Sweden Lysekil Fjord- Alsback Station	GS689	72	coastal	<0.1
Sweden Lysekil Fjord- Alsback Station	GS689	72	coastal	0.8
Sweden Lysekil Fjord- Alsback Station	GS689	72	coastal	3
Denmark DMU 1005, Denmark	GS694	0.3	coastal	0.1
Denmark DMU 1005, Denmark	GS694	0.3	coastal	0.8
Denmark DMU 1005, Denmark	GS694	0.3	coastal	3
Denmark DMU 1005, Denmark	GS695	19	coastal	0.1
Denmark DMU 1005, Denmark	GS695	19	coastal	<0.1
Denmark DMU 1005, Denmark	GS695	19	coastal	0.8
Denmark DMU 1005, Denmark	GS695	19	coastal	3

Table 2: The 71 samples from the Baltic Sea that were sequenced and analyzed as part of this study.

Species	Number of Sequence Reads	Percentage of Total Sample Sequence Reads
Candidatus Pelagibacter ubique	341984	4%
Alpha proteobacterium HIMB114	204431	2%
Unresolved	129039	1%
Micromonas sp. RCC299	107136	1%
Candidatus Pelagibacter sp. HTCC7211	101965	1%
Chthoniobacter flavus	93989	1%
Flavobacteria bacterium MS024-2A	86282	1%
Candidatus Puniceispirillum marinum	84669	1%
Psychroflexus torquis	81564	1%
Marine gamma proteobacterium HTCC2080	79096	1%

Table 3: Top 10 microbial species found in the 71 samples from the Baltic Sea sequenced and analyzed as part of this study.

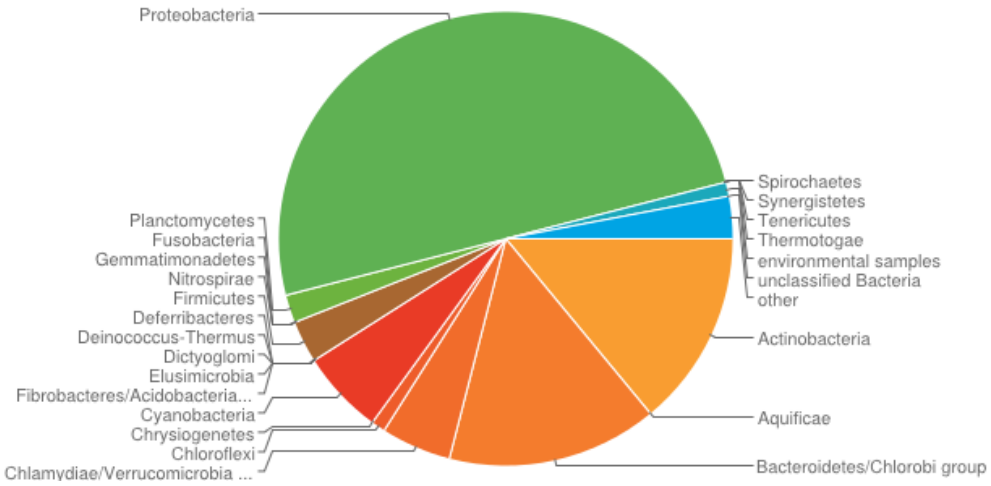

Figure 5a: Taxonomic distribution of Bacteria at the phylum level from 71 Baltic Sea samples sequenced and analyzed as part of this study.

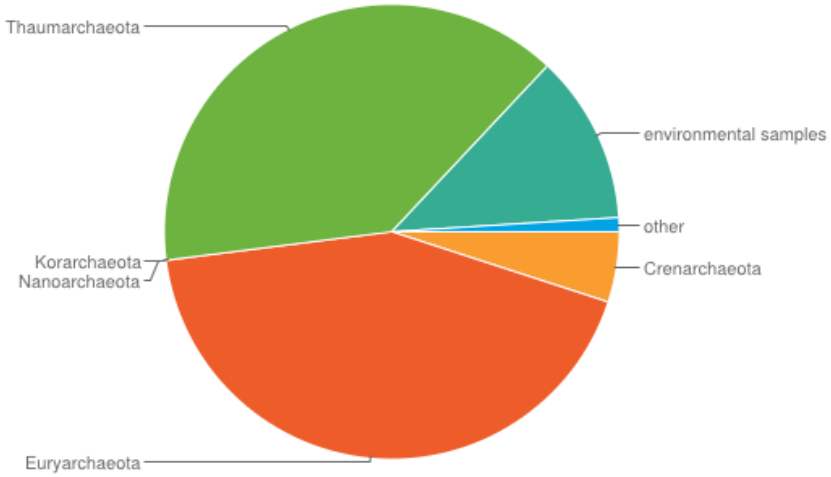

Figure 5b: Taxonomic distribution of Archaea at the phylum level from 71 Baltic Sea samples sequenced and analyzed as part of this study.

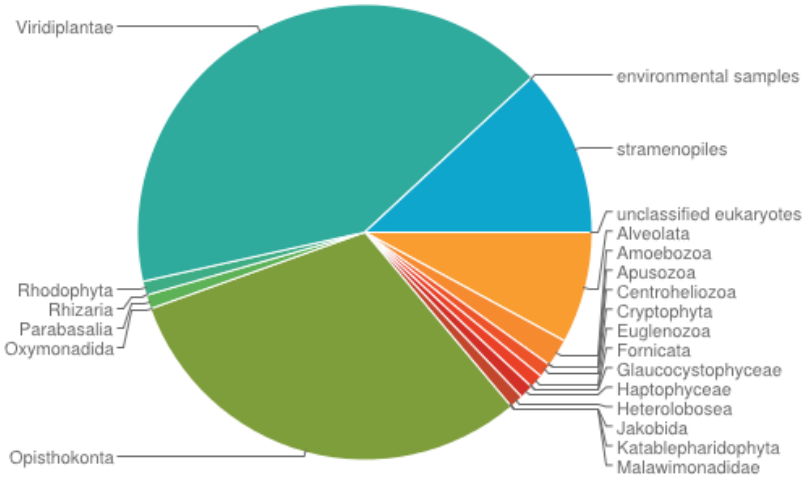

Figure 5c: Taxonomic distribution of Eukaryotes at the phylum level from 71 Baltic Sea samples sequenced and analyzed as part of this study.

6. Key Global Ocean Sampling Expedition Publications

Allen LZ, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LD, Thiagarajan M, Rusch DB, Nealson KH, Williamson SJ, Venter JC, Allen AE. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. ISME J. 2012 Jan 26.

Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Alexander Richter R, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS, Nealson K, Friedman R, Craig Venter J. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J. 2011 Dec 15.

Wu D, Wu M, Halpern A, Rusch DB, Yooseph S, Frazier M, Venter JC, Eisen JA. Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. PLoS One. 2011 Mar 18;6(3).

Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M, Johnson J, Montgomery R, Ferriera S, Beeson K, Williamson SJ, Tovchigrechko A, Allen AE, Zeigler LA, Sutton G, Eisenstadt E, Rogers YH, Friedman R, Frazier M, Venter JC. Genomic and functional adaptation in surface ocean planktonic prokaryotes. Nature. 2010 Nov 4;468(7320):60-6.

Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. Characterization of Prochlorococcus clades from iron-depleted oceanic regions. Proc Natl Acad Sci U S A. 2010 Sep 14;107(37):16184-9. Epub 2010 Aug 23.

Shaw, A. K., Halpern, A. L., Beeson, K., Tran, B., Venter, J. C., Martiny, J. B. It's all relative: ranking the diversity of aquatic bacterial communities. Environ Microbiol. 2008 Sep 10.

Yooseph, S., Li, W., Sutton, G. Gene identification and protein classification in microbial metagenomic sequence data via incremental clustering. BMC Bioinformatics. 2008 Apr 10; 9(1): 182.

Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., Andrews-Pfannkoch, C., Fadrosh, D., Miller, C. S., Sutton, G., Frazier, M., Venter, J. C. The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples. PLoS ONE. 2008 Jan 23; 3(1): e1456.

Sharon, I., Tzahor, S., Williamson, S., Shmoish, M., Man-Aharonovich, D., Rusch, D. B., Yooseph, S., Zeidner, G., Golden, S. S., Mackey, S. R., Adir, N., Weingart, U., Horn, D., Venter, J. C., Mandel-Gutfreund, Y., Beja, O.

Viral photosynthetic reaction center genes and transcripts in the marine environment. Isme J. 2007 Oct 01; 1(6): 492-501.

Yutin, N., Suzuki, M. T., Teeling, H., Weber, M., Venter, J. C., Rusch, D. B., Beja, O. Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. Environ Microbiol. 2007 Jun 01; 9(6): 1464-75.

Douglas, B.R., Halpern, A.L., Heidelberg, K.B., Sutton, G., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Howard, C.H., Foote, C., Dill, B.A., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freemen, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y., Zhang, S., Bafna, V., Falcon, L.I., Souza, V., Bonilla, G., Eguiarte, L.E., Karl, D.M., Nealson, K., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo, G., Ferrari, M.R., Friedman, R., Strausberg, R.L., Frazier, M., and Venter, J.C. The Sorcerer II Global Ocean Sampling Expedition: The Northwest Atlantic through the Eastern Tropical Pacific. PLoS Biology. March 2007; 5(3): 0398-0431.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C.S., Li H., Mashiyama, S.T., Joachimiak, M.P., van Belle, C., Chandonia, J., Soergel, D.A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B.J., Bafna, V., Friedman, R., Brenner, S.E., Godzik, A., Eisenberg, D., Dixon, J.E., Taylor, S.S., Strausberg, R.L., Frazier, M., and Venter, J.C. Expanding the Universe of Protein Families. PLoS Biology. March 2007; 5(3): 0432-0466.

Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C., Manning, G. Structural and functional diversity of the microbial kinome. PLoS Biology. March 2007; 5(3): e17.

Venter, JC, Remington, K, Heidelberg, JF, Halpern, AL, Rusch, D, Eisen, JA, Wu, D, Paulsen, I, Nelson, KE, Nelson, W, Fouts, DE, Levy, S, Knap, AH, Lomas, M W, Nealson, K, White, O, Peterson, J, Hoffman, J, Parsons, R, Baden-Tillson, H, Pfannkoch, C, Rogers, YH, Smith, HO. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 304 (5667): 66-74.