

27621

Prokaryotic Gene Discovery, Metagenomics and Pangenomics

Thomas Sicheritz-Pontén
Carsten Friis
Dave Ussery

*Center for Biological Sequence Analysis
Department of Systems Biology
Technical University of Denmark*



NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

Outline

- People
- Course structure
- Learning objectives
- Examination
- Short Intro to the Curriculum

Main Teachers

Thomas Sicheritz-Ponten
Associate Professor
Head of Metagenomics Group



Main Teachers

Carsten Friis Rundsten

Assistant Professor

Comparative Microbial Genomics Group



Main Teachers

Dave Ussery
Associate Professor
Head of Comparative Microbial Genomics Group



Other Teachers

Kristoffer Kiil

PhD student

Comparative Microbial Genomics Group



Marcelo Bertalan

PhD student

Metagenomics Group

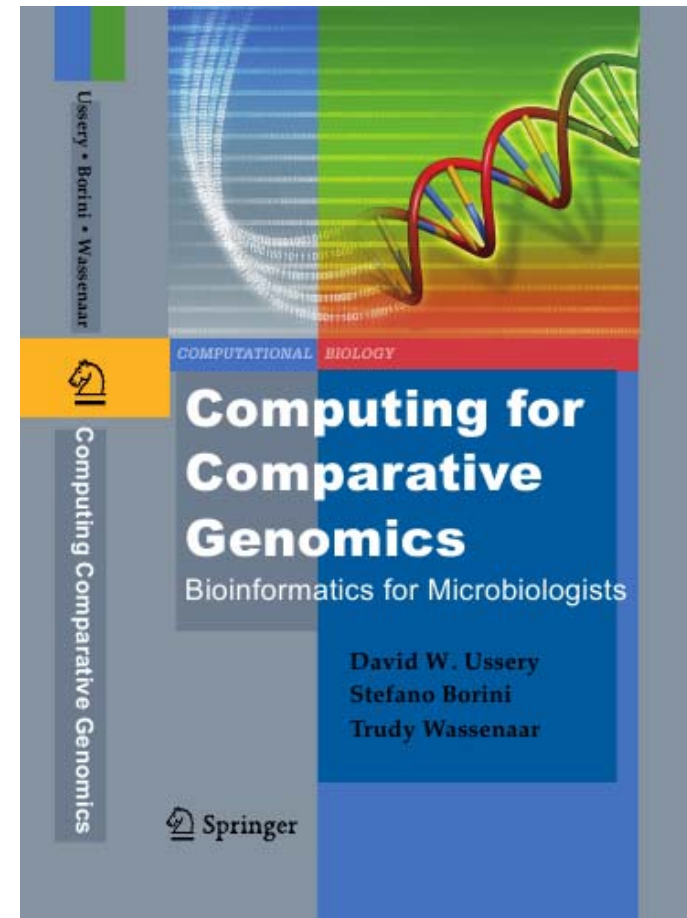
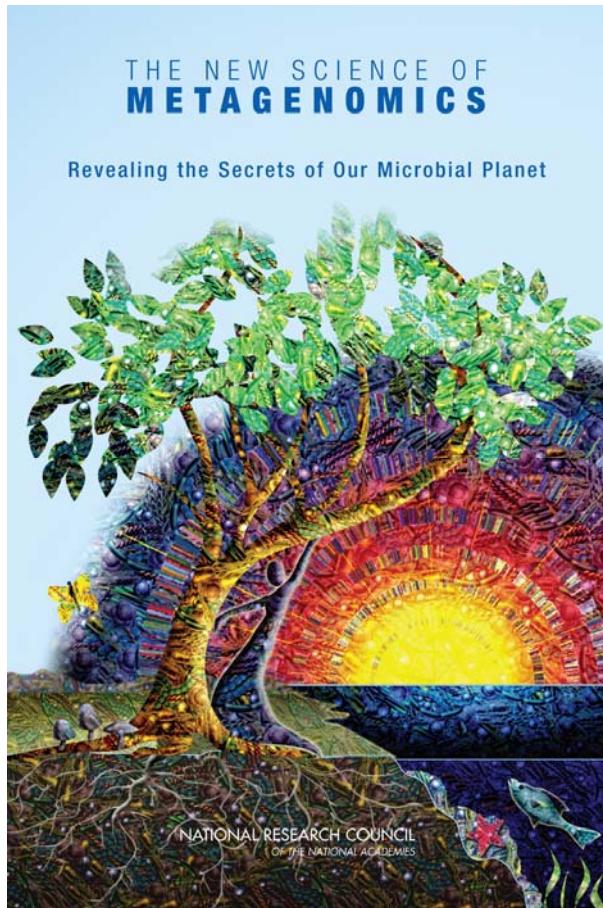


Ulrik Plate

PhD student

Metagenomics Group

Text books



Content of the Course

Course Philosophy

- Cutting Edge Science
- Specialist Knowledge (Guest Lecturers)
- Dialog, not monolog

Student Projects

- Groups of 2-3 Students
- Problem Formulation
- Project Presentation

Course Structure

- Lectures
- Labs/Computer exercises
- Lab Journals **must** be handed in

Exam

- 4 Hrs. Written Exam
- No Aid!

Course Web Page

<http://www.cbs.dtu.dk/dtucourse/metagenomics/27621.php>

The screenshot shows a Mozilla Firefox browser window with the following details:

- Title Bar:** Prokaryotic Gene Discovery, Metagenomics and Pangenomics - Mozilla Firefox
- Address Bar:** <http://www.cbs.dtu.dk/dtucourse/metagenomics/27621.php>
- Navigation:** Back, Forward, Home, Reload buttons.
- Bookmarks:** Most Visited, Customize Links, Free Hotmail, Windows Marketplace, Windows Media, Windows, DTV Proxy.
- Open Tabs:** Environmental genome shotgun..., The microbial pan-genome. [Cur..., CBS Prokaryotic Gene Discover..., ScienceDirect - Current Opinion..., Environmental Genome Shotgun...
- Navigation Menu:**
 - EVENTS
 - NEWS
 - RESEARCH GROUPS
 - CBS PREDICTION SERVERS
 - CBS DATA SETS
 - CBS BIOINFORMATICS TOOLS
 - PUBLICATIONS
 - CBS COURSES
 - BIOINFORMATICS EDUCATION PROGRAM
 - OTHER BIOINFORMATICS LINKS
 - STAFF
 - CONTACT
 - ABOUT CBS
 - INTERNAL
- Page Content:**

CBS >> [CBS Courses](#) >> Prokaryotic Gene Discovery, Metagenomics and Pangenomics

27621 - Prokaryotic Gene Discovery, Metagenomics and Pangenomics

Course Programme - Autumn 2009

Practical Information

The aim of this course is to give the students a solid knowledge of methods for determining microbial meta-, core- and pan-genomes, based on the genome sequences of a set of related organisms or microbial communities. They should be able to understand and apply this information to practical uses, such as screening large segments of sequenced DNA extracted from wide-ranging environmental samples, classification of bacterial species as well as the design of high-density microarray chips. Furthermore, it is a goal that the students learn to formulate interesting problems/questions that can be addressed by pangenome-chip and advanced metagenomics analysis.

Lectures and exercises will take place at the Technical University of Denmark, building 208, room 062, in Lyngby.

Exam

Written Exam, 4 hours
 Date: To be announced

Reading material:

Exact details as to which chapters and which papers are considered curriculum will be given as the course progresses.

 - [The new science of metagenomics](#)
 - [Computing for Comparative Microbial Genomics](#)
 - Any articles handed out during lectures
 - Everything talked about during the lectures and labs

Learning Objectives

- Understand and explain the differences between conventional, meta- and pan-genomics.
 - Understand and apply bioinformatic tools on meta- and pan-genomic data.
 - Estimate 16S rRNA biodiversity and the size of the total core- and pan-genomes
 - Produce a list of genes representing the core- and pan-genomes.
 - Independently design a metagenomic project and a pan-genomic microarray chip.
 - Explain the process of microbial gene discovery
 - Explain the basic principles of the different sequencing methods
-
- Plan and carry out a written project, including the stating of and concluding of hypotheses
 - Present a scientific project as a poster to the teacher and other students
 - Present a scientific project orally and in English using a PowerPoint presentation

To Pass the Course

- Participation in the weekly lectures/exercises
 - All Computer Lab Reports must be Approved
- Exam
 - 50 % of the grade
- Project work
 - Oral Presentation
 - Group Work, 2-3 Students
 - 50 % of the grade

Project Details

You Write your own Problem Formulation

- And present it in front of class

You present project as an Oral Presentation (Powerpoint style)

- Cross-examined by Teachers and Other Students


Hand in Project Summary Prior to Presentation

- ~0.5-1 A4 page
- Opponent Group

Coffee and Computers

Coffee break ~13:45 – 14:00

Portable Computers to borrow from CBS
(fetch between 12-13 and return latest 17:00)



CENTERFO
R BIOLOGI
CAL SEQU
ENCE ANA
LYSIS CBS

27621

Prokaryotic Gene Discovery, Metagenomics and Pangenomics

Thomas Sicheritz-Pontén
Carsten Friis
Dave Ussery

*Center for Biological Sequence Analysis
Department of Systems Biology
Technical University of Denmark*



The First Whole Genome

Display Show Sort By Send to

All: 1 Review: 0

1: [Science](#). 1995 Jul 28;269(5223):496-512.

Comment in:

- [Science. 1995 Jul 28;269\(5223\):468-70.](#)
- [Science. 1995 Sep 29;269\(5232\):1805.](#)
- [Science. 1996 Mar 1;271\(5253\):1302-3; author reply 1303-4.](#)
- [Science. 1996 Mar 1;271\(5253\):1302; author reply 1303-4.](#)

Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.

[Fleischmann RD](#), [Adams MD](#), [White O](#), [Clayton RA](#), [Kirkness EF](#), [Kerlavage AR](#), [Bult CJ](#), [Tomb JE](#), [Dougherty BA](#), [Merrick JM](#), et al.

Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium *Haemophilus influenzae* Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The *H. influenzae* Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.

PMID: 7542800 [PubMed - indexed for MEDLINE]

Intraspecies variation in bacterial genomes: the need for a species genome concept

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC

Search PubMed for [Advanced](#)

Display AbstractPlus Show 20 Sort By Send to

All: 1 Review: 1

1: [Trends Microbiol.](#) 2000 Sep;8(9):396-401.

Intraspecies variation in bacterial genomes: the need for a species genome concept.

[Lan R, Reeves PR.](#)

Dept of Microbiology, Bldg G08, University of Sydney, NSW 2006, Sydney, Australia.

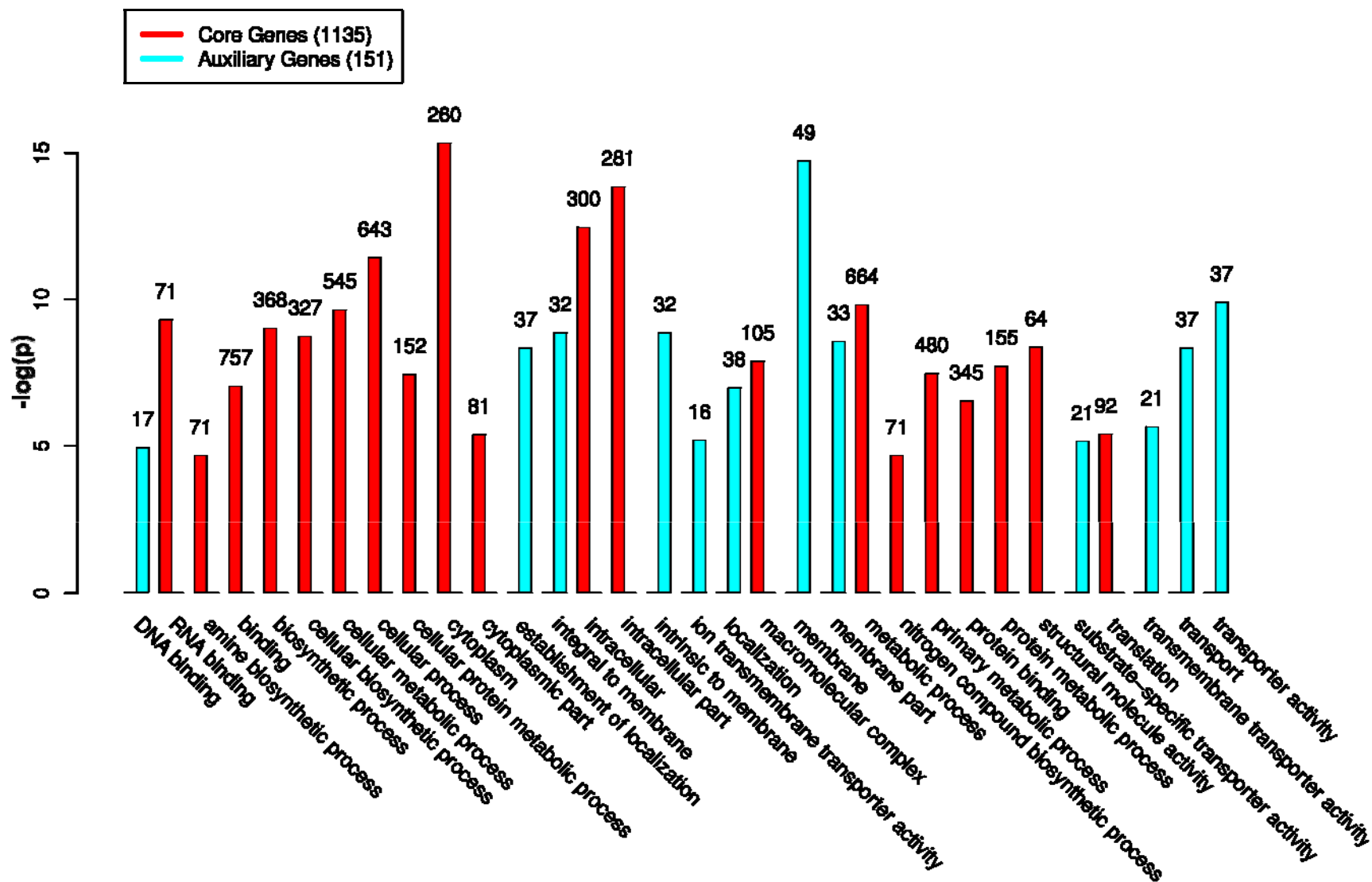
Bacterial populations are clonal. Their evolution involves not only divergence between orthologous genes but also gain of genes from other clones or species, which has only recently been widely appreciated through macrorestriction mapping, genomic subtraction and complete genome sequencing. Genes can also be lost in response to selection or by random mutation after becoming redundant. The bacterial genome is a dynamic structure and intraspecies variation needs to be included in genome analysis if we are to gain insight into the full species genome.

PMID: 10989306 [PubMed - indexed for MEDLINE]

The species genome concept

- Even the limited data available on pair-wise comparisons show that up to 20% of the DNA in one strain can be absent in another. The total amount of such DNA is, at present, unknown, but it is part of the genome of that particular species. Clearly, comparison with different strains will add to the DNA in this class.
- If we accept the concept of the species genome, comprising all genes found in the species, then the genes of any individual will include two components: the core set of genes and the auxiliary genes. Genes found in most individuals, which we can call the core set of genes for that species, are the genes that determine those properties characteristic of all members of the species.
- Additionally, each strain will have some auxiliary genes, which determine properties found in some but not all members of the species. The distinction will not be absolute but we believe it provides a useful framework.
- Suitable boundaries might become obvious as our knowledge of intraspecies variation grows, but we suggest as a starting point that genes found in 95% or more of isolates form the core set and genes found in 1–95% of isolates form part of the auxiliary set of genes; those present in 1% are provisionally treated as foreign genes or genes being lost from the species. The cutoffs are arbitrary, and with better knowledge one would define the lower cutoff in terms of genes that persisted in the species for long enough to show that their presence was maintained by selection.

Overrepresented gene ontology terms for core and auxiliary genes



The microbial pan-genome

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC

Search PubMed for [Advanced](#)

Display AbstractPlus Show 20 Sort By Send to

All: 1 Review: 1

1: [Curr Opin Genet Dev](#), 2005 Dec;15(6):589-94. Epub 2005 Sep 26.

The microbial pan-genome.

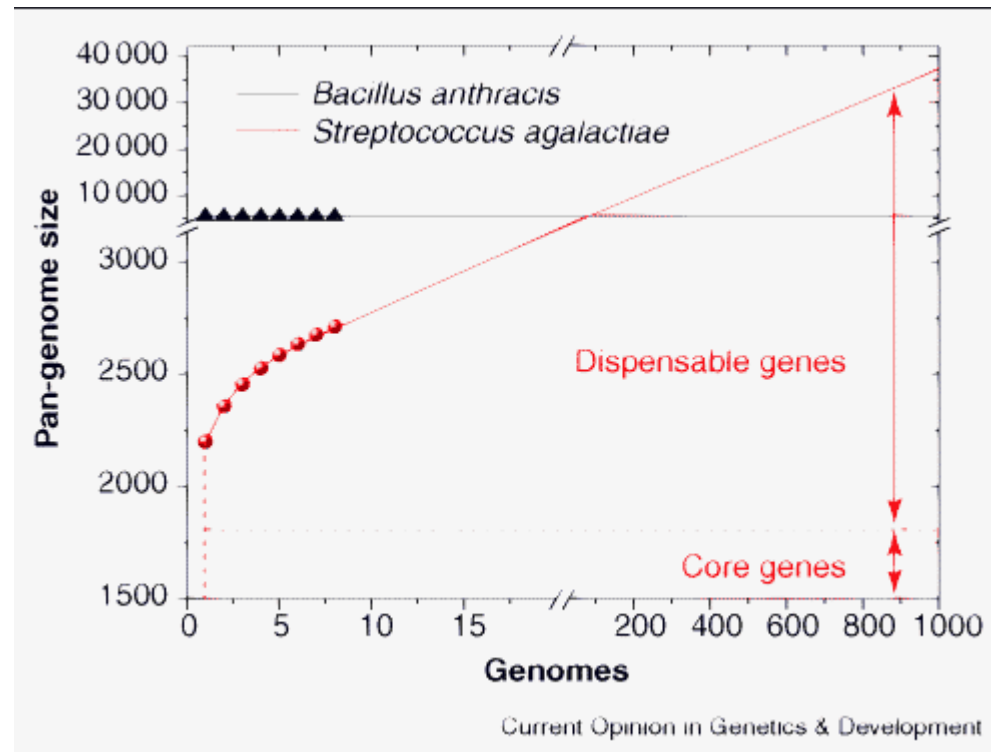
[Medini D](#), [Donati C](#), [Tettelin H](#), [Masignani V](#), [Rappuoli R](#).

Immunobiological Research Institute of Siena (IRIS), Chiron Vaccines, via Fiorentina 1, 53100 Siena, Italy.

A decade after the beginning of the genomic era, the question of how genomics can describe a bacterial species has not been fully addressed. Experimental data have shown that in some species new genes are discovered even after sequencing the genomes of several strains. Mathematical modeling predicts that new genes will be discovered even after sequencing hundreds of genomes per species. Therefore, a bacterial species can be described by its pan-genome, which is composed of a "core genome" containing genes present in all strains, and a "dispensable genome" containing genes present in two or more strains and genes unique to single strains. Given that the number of unique genes is vast, the pan-genome of a bacterial species might be orders of magnitude larger than any single genome.

PMID: 16185861 [PubMed - indexed for MEDLINE]

The genes of a species pan-genome



J. Craig Venter



Environmental genome shotgun sequencing of the Sargasso Sea

Display Show Sort By Send to

All: 1 Review: 0

1: [Science](#). 2004 Apr 2;304(5667):66-74. Epub 2004 Mar 4.

Comment in:

[Science](#). 2004 Apr 2;304(5667):58-60.

Environmental genome shotgun sequencing of the Sargasso Sea.

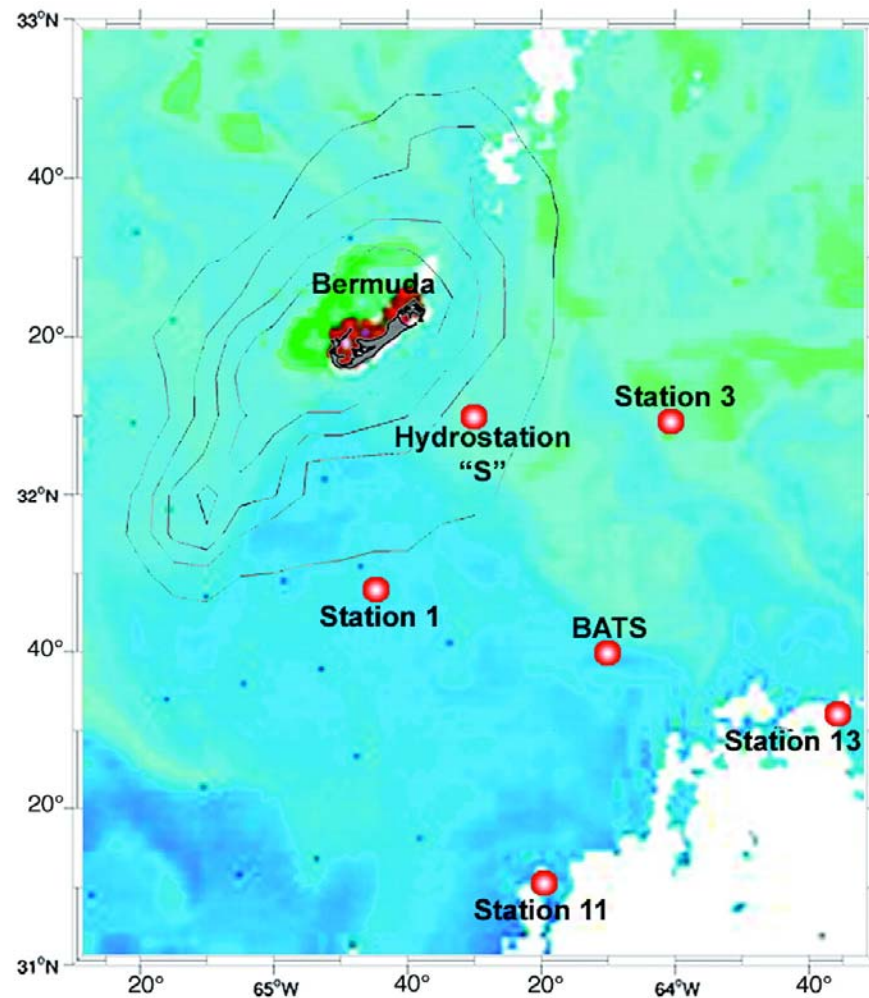
[Venter JC](#), [Remington K](#), [Heidelberg JF](#), [Halpern AL](#), [Rusch D](#), [Eisen JA](#), [Wu D](#), [Paulsen I](#), [Nelson KE](#), [Nelson W](#), [Fouts DE](#), [Levy S](#), [Knap AH](#), [Lomas MW](#), [Nealson K](#), [White O](#), [Peterson J](#), [Hoffman J](#), [Parsons R](#), [Baden-Tillson H](#), [Pfannkoch C](#), [Rogers YH](#), [Smith HO](#).

Institute for Biological Energy Alternatives, 1901 Research Boulevard, Rockville, MD 20850, USA.
 jcventer@tcag.org

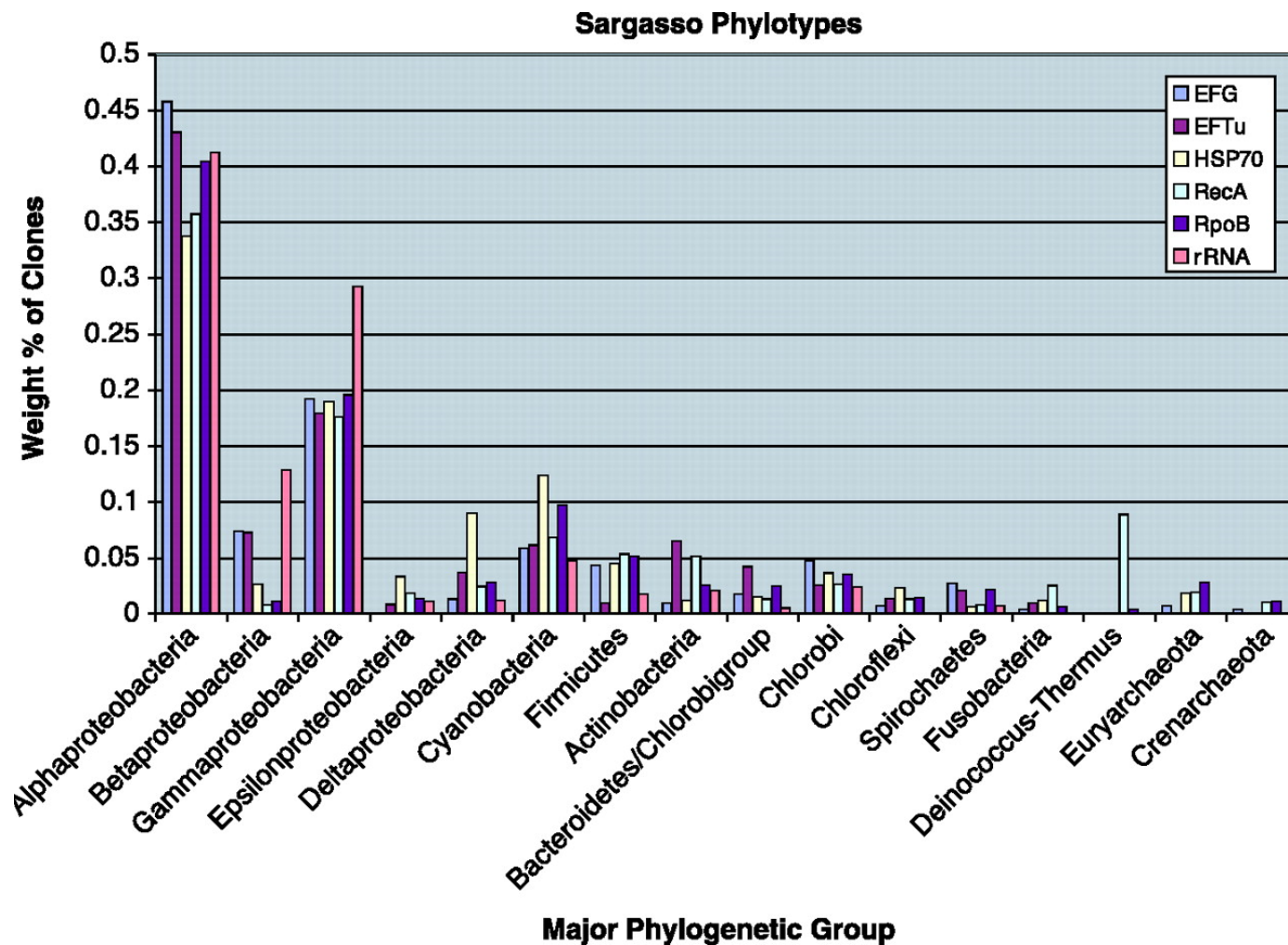
We have applied "whole-genome shotgun sequencing" to microbial populations collected en masse on tangential flow and impact filters from seawater samples collected from the Sargasso Sea near Bermuda. A total of 1.045 billion base pairs of nonredundant sequence was generated, annotated, and analyzed to elucidate the gene content, diversity, and relative abundance of the organisms within these environmental samples. These data are estimated to derive from at least 1800 genomic species based on sequence relatedness, including 148 previously unknown bacterial phylotypes. We have identified over 1.2 million previously unknown genes represented in these samples, including more than 782 new rhodopsin-like photoreceptors. Variation in species present and stoichiometry suggests substantial oceanic microbial diversity.

PMID: 15001713 [PubMed - indexed for MEDLINE]

Sargasso Sea Samplings



Phylogenetic diversity of Sargasso Sea



Sargasso Sea Samplings

