

EXPLORING THE OCEAN'S MICROBES: SEQUENCING THE SEVEN SEAS

Marvin E. Frazier¹, Douglas B. Rusch¹, Aaron L. Halpern¹, Karla B. Heidelberg¹, Granger Sutton¹, Shannon Williamson¹, Shibu Yooseph¹, Dongying Wu², Jonathan A. Eisen², Jeff Hoffman¹, Charles H. Howard¹, Cyrus Foote¹, Brooke A. Dill¹, Karin Remington¹, Karen Beeson¹, Bao Tran¹, Hamilton Smith¹, Holly Baden-Tillson¹, Clare Stewart¹, Joyce Thorpe¹, Jason Freeman¹, Cindy Pfannkoch¹, Joseph E. Venter¹, John Heidelberg², Terry Utterback¹, Yu-Hui Rogers¹, Shaojie Zhang³, Vineet Bafna³, Luisa Falcon⁴, Valeria Souza⁴, German Bonilla⁴, Luis E. Eguiarte⁴, David M. Karl⁵, Ken Neelson⁶, Shubha Sathyendranath⁷, Trevor Platt⁷, Eldredge Bermingham⁸, Victor Gallardo⁹, Giselle Tamayo¹⁰, Robert Friedman¹, Robert Strausberg¹, J. Craig Venter¹

¹ *J. Craig Venter Institute, Rockville, Maryland, United States Of America*

² *The Institute For Genomic Research, Rockville, Maryland, United States Of America*

³ *Department of Computer Science, University of California San Diego*

⁴ *Instituto de Ecologia Dept. Ecología Evolutiva, National Autonomous University of Mexico Mexico City, 04510 Distrito Federal, Mexico*

⁵ *University of Hawaii, Honolulu, United States of America*

⁶ *Dept. of Earth Sciences, University of Southern California, Los Angeles, California, United States of America*

⁷ *Dalhousie University, Halifax, Nova Scotia, Canada*

⁸ *Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama*

⁹ *University of Concepción, Concepción, Chile*

¹⁰ *University of Costa Rica, San Pedro, San José, Republic of Costa Rica*

The J. Craig Venter Institute's (JCVI) environmental genomics group has collected ocean and soil samples from around the world. We have begun shotgun sequencing of microbial samples from more than 100 open-ocean and coastal sites across the Pacific, Indian and Atlantic Oceans. These data are being augmented with deep sequencing of 16S and 18S rRNA and the draft sequencing of ~150 cultured marine microbial species. The JCVI is also developing and refining bioinformatics tools to assemble, annotate, and analyze large-scale metagenomic data, along with the appropriate database infrastructure to enable directed analyses. The goals of this Global Ocean Survey are to better understand microbial biodiversity; to discover new genes of ecological importance, including those involved in carbon cycling; to discover new genes that may be useful for biological energy production; and to establish a freely shared, global environmental genomics database that can be used by scientists around the world.

Using newly developed metagenomic methods, we are able to examine not only the community of microorganisms, but the community of genes that enable them to capture energy from the sun, remove

carbon dioxide from the air, take up organic carbon, and cycle nitrogen in its various forms through the ecosystem. To date, we have discovered many thousands of new microbial species and millions of new genes, with no apparent slowing of the rate of discovery. This data will be of great value for the study of protein function and protein evolution. The goal of this new science, however, is not to merely catalog sequences, genes and gene families, and species for their own sake. We are attempting to use these new data to better understand the functioning of natural ecosystems. Environmental metagenomics examines the interplay of perhaps thousands of species present and functioning at a point in space and time. Each individual sequence is no longer just a piece of a genome. It is a piece of an entire biological community. This is a resource that can be mined by microbial ecologists worldwide to better understand biogeochemical cycling. Moreover, within this data set is a huge diversity of previously unknown, energy-related genes that may be useful for developing new methods of biological energy production.

We acknowledge the DOE, Office of Science (DE-FG02-02ER63453), the Gordon and Betty Moore Foundation, the Discovery Channel and the J. Craig Venter Science Foundation for funding to undertake this study. We are also indebted to a large group of individuals and groups for facilitating our sampling and analysis. We thank the Governments of Canada, Mexico, Honduras, Costa Rica, Panama, and Ecuador and French Polynesia/France for facilitating sampling activities. All sequencing data collected from waters of the above named countries remain part of the genetic patrimony of the country from which they were obtained.

Canada's Bedford Institute of Oceanography provided a vessel and logistical support for sampling in Bedford basin. The Universidad Nacional Autónoma de México (UNAM) facilitated permitting and logistical arrangements and identified a team of scientists for collaboration. The scientists and staff of the Smithsonian Tropical Research Institute (STRI) hosted our visit in Panama. Representatives from Costa Rica's Organization for Tropical Studies (Jorge Arturo Jimenez and Francisco Campos Rivera), the University of Costa Rica (Jorge Cortés) and the National Biodiversity Institute (INBio) provided assistance with planning, logistical arrangements and scientific

analysis. Our visit to the Galapagos Islands was facilitated by assistance from the Galapagos National Park Service Director, Washington Tapia, the Charles Darwin Research Institute, especially Howard Snell and Eva Danulat. We especially thank Greg Estes (guide), Héctor Cháuz Campo (Institute of Oceanography of the Ecuador Navy) and a National Park Representative, Simon Ricardo Villemar Tigrero for field assistance while in the Galapagos Islands. Martin Wilkalski (Princeton) and Rod Mackie (University of Illinois) provided advice for target regions in the Galapagos to sample. We thank Matthew Charette (Woods Hole Oceanographic Institute) and Dave Karl (University of Hawaii) for nutrient analysis work and advice. We also acknowledge the help of Michael Ferrari and Jennifer Clark for assistance in acquiring the satellite images. The U.S. Department of State facilitated Governmental communications on multiple occasions. John Glass (JCVI) provided valuable assistance in methods development. Tyler Osgood (JCVI) facilitated many of the vessel related technical needs. We gratefully acknowledge Dr. Michael Sauri, who oversaw medical related issues for the crew of the Sorcerer II. Finally, special thanks also to the captain and crew of the S/V Sorcerer II.