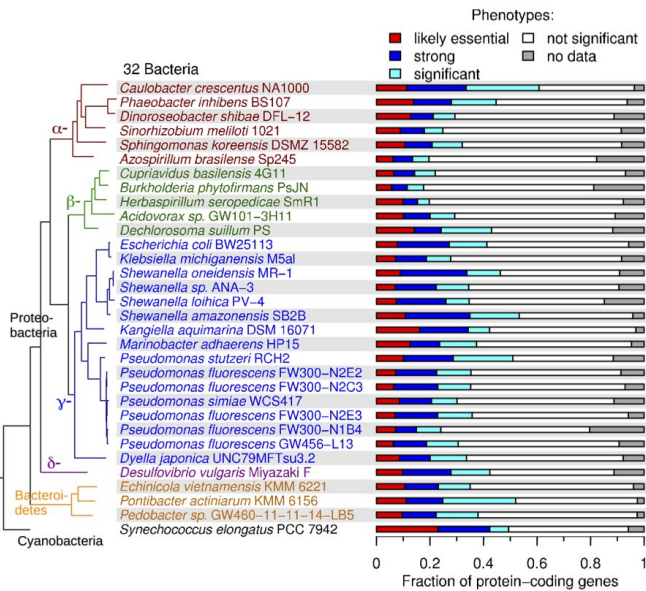


# Exploration of diverse bacteria signals big advance for gene function prediction

16 May 2018



scale, genome-wide assays of gene importance across many conditions. The study, "Mutant Phenotypes for Thousands of Bacterial Genes of Unknown Function," has been published in the journal *Nature* and is by far the largest functional genomics study of bacteria ever published.

"This is the first really large, systematic experimental effort to try to assign functions to [bacterial genes](#) of unknown function," said study senior author and biologist Adam Deutschbauer of Berkeley Lab's Biosciences Area. "We are tackling the problem that biology is up against and recognizes: It is super easy to sequence, but we cannot currently assign confident functions for the majority of [genes](#) identified by sequencing. Our experimental data provides an anchor that other researchers could use to make a more informed inference about protein function."

Berkeley Lab bioscientists applied high-throughput genetics to 32 different bacteria (left) and identified phenotypes for thousands of genes of previously unknown function (right). Credit: Berkeley Lab

Tested on nearly three dozen bacteria from various genera, the workflow combined high-throughput genetics and comparative genomics to identify mutant phenotypes for thousands of genes with previously unknown functions.

In the air, beneath the ocean's surface, and on land, microbes are the minute but mighty forces regulating much of the planet's biogeochemical cycles. To better understand their roles, scientists work to identify these microbes and to determine their individual contributions. While advances in sequencing technologies have enabled researchers to access the genomes of thousands of microbes and make them publicly available, no similar shift has occurred with the task of assigning functions to the genes uncovered.

## Technology to understand Earth's genetic potential

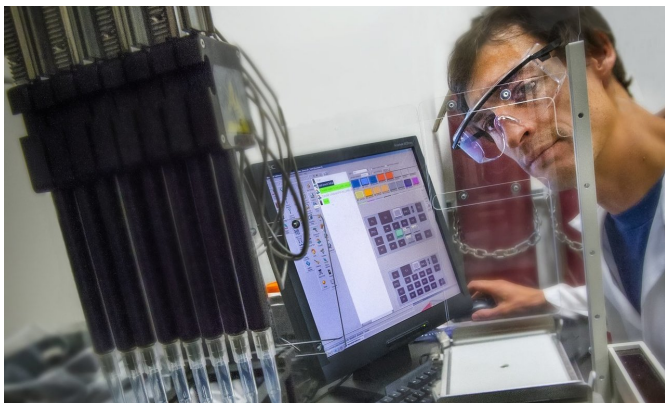
The team worked with 32 bacteria, including plant-growth promoting bacteria and a cyanobacterium relevant for biofuels production, as well as bacteria involved in bioremediation. "Typically, researchers work on functional analysis of individual genomes, from a limited number of 'workhorse' bacteria," said JGI scientist Matt Blow, the study's co-corresponding author. "This is because of the limited capacity of [functional analysis](#) approaches compared with high-throughput sequencing. Here, you have data from 32 different bacteria at once, capturing more microbial diversity."

To help overcome this bottleneck, scientists at Lawrence Berkeley National Laboratory (Berkeley Lab), including researchers at the U.S. Department of Energy (DOE) Joint Genome Institute (JGI), have developed a workflow that enables large-

To more efficiently generate mutant libraries for each bacterium, the team refined a DNA bar-code

sequencing approach known as RB-TnSeq (randomly bar-coded transposon sequencing). "The implications of this work are that it could be scaled with proper investment and coordination—in combination with other methods—to have substantial benefit for understanding the genetic potential of the Earth," said Adam Arkin, senior faculty scientist and co-corresponding author.

suggested they might be part of the same pathway. For example, they found that genes with the uncharacterized protein domain UPF0126 were important for growth on glycine in 11 different bacteria, suggesting that this protein domain is involved in transporting glycine across the cell membrane. Studying such conserved associations, he added, demonstrates the value in identifying phenotypes for homologous genes across multiple bacterial species.



Scientists including Adam Deutschbauer, pictured, used automation to achieve the results of this large-scale project. Credit: Roy Kaltschmidt/Berkeley Lab

"The technology behind this project was developed to elucidate the genetic functions of all the organisms we are collecting in the field and to understand importance for organism fitness in diverse environments," he added, speaking as co-director of Berkeley Lab's ENIGMA Scientific Focus Area, DOE Office of Science's largest and longest-running environmental biology program. "We believe that to understand means—given appropriate data—you should be able to predict, control, and design behavior in the system of interest."

### Conserved phenotypes suggest functional associations

Deutschbauer pointed out that the resulting large data set allowed the team to glean insights from conserved phenotypes across organisms, and also look for co-fitness patterns among the genes, cases where two genes had similar patterns of phenotypes across all conditions, a correlation that

"A comparative functional genomics study of bacteria was not really possible before because large genetic data sets were available for only a few bacteria, and the ones that did exist were not typically generated with the same technology, same methodology, or the same metadata, so it's hard to do comparisons," he said. "Although we experimentally studied a relatively small number of bacteria compared to the diversity present in nature, our data is of relevance across all bacteria. For example, about 12 percent of all uncharacterized proteins across [bacteria](#) have a homologous protein with a functional phenotypic association in our data set."

The data set is publicly accessible for comparative analyses at [fit.genomics.lbl.gov](http://fit.genomics.lbl.gov), a web workbench developed by Morgan Price, the study's lead author, who has also developed powerful tools such as PaperBlast to help interpret results.

Arkin also sees future benefits toward integrating this data set into systems like the JGI's [IMG/M system](#) and the [DOE Systems Biology Knowledgebase \(KBase\)](#), the first large-scale bioinformatics system that allows users to upload, analyze, and share information within a single integrated environment.

"These data sets provide a fantastic opportunity for innovations in data science to predict biological function," said Arkin, who is KBase's CEO and lead primary investigator. "At KBase, we are already working with JGI to integrate data like this together with phylogenetic, homology, and chemical similarity relationships to propagate this information across the tree of life, and to project, for example, improved metabolic models for organisms and communities so we can predict the conditions that

most impact growth."

**More information:** Morgan N. Price et al, Mutant phenotypes for thousands of bacterial genes of unknown function, *Nature* (2018). [DOI: 10.1038/s41586-018-0124-0](https://doi.org/10.1038/s41586-018-0124-0)

Provided by Lawrence Berkeley National Laboratory

APA citation: Exploration of diverse bacteria signals big advance for gene function prediction (2018, May 16) retrieved 18 June 2019 from <https://phys.org/news/2018-05-exploration-diverse-bacteria-big-advance.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*