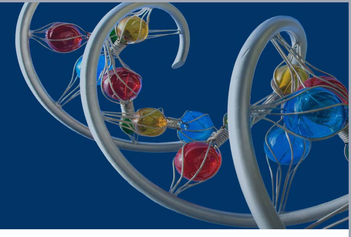
Functional Encyclopedia of Bacteria and Archaea

M. J. Blow[1], A. M. Deutschbauer[2], C. A. Hoover[1], J. Lamson[2], L.A.Pennacchio[1], M. N. Price[2], J. Waters1, K. M. Wetmore[2], J. Bristow[1], A. P. Arkin[2]
[1.] D.O.E. Joint Genome Institute, Lawrence Berkeley National Laboratory
[2.] Physical Biosciences Division, Lawrence Berkeley National Laboratory

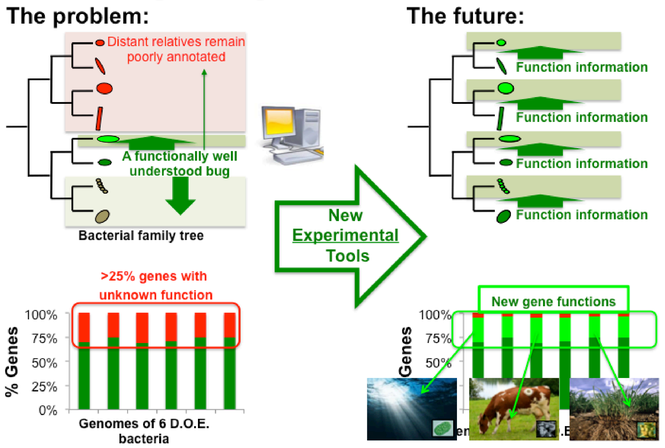
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE



Summary

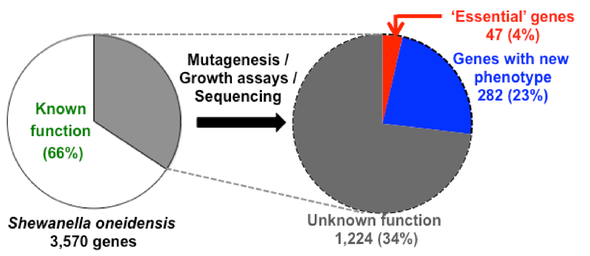
Bacteria and Archaea exhibit a huge diversity of metabolic capabilities with fundamental importance in the environment, and potential applications in biotechnology. However, the genetic bases of these capabilities remain unclear due largely to an absence of technologies that link DNA sequence to molecular function. To address this challenge, we are developing a pipeline for high throughput annotation of gene function using mutagenesis, growth assays and DNA sequencing. By applying this pipeline to annotate gene function in 50 diverse microbes we hope to discover thousands of new gene functions and produce a proof of principle 'Functional Encyclopedia of Bacteria and Archaea'.

Challenges in gene function annotation



Computational annotation of gene function relies on transfer of information between closely related sequences. As these are often absent, a 'typical' genome has >25% genes with no annotated function. New experimental tools capable of high-throughput annotation of novel gene functions are therefore required. Novel annotations would provide direct biological insights and serve as a resource for improving traditional automated annotation approaches.

Gene function annotation using Tn-Seq



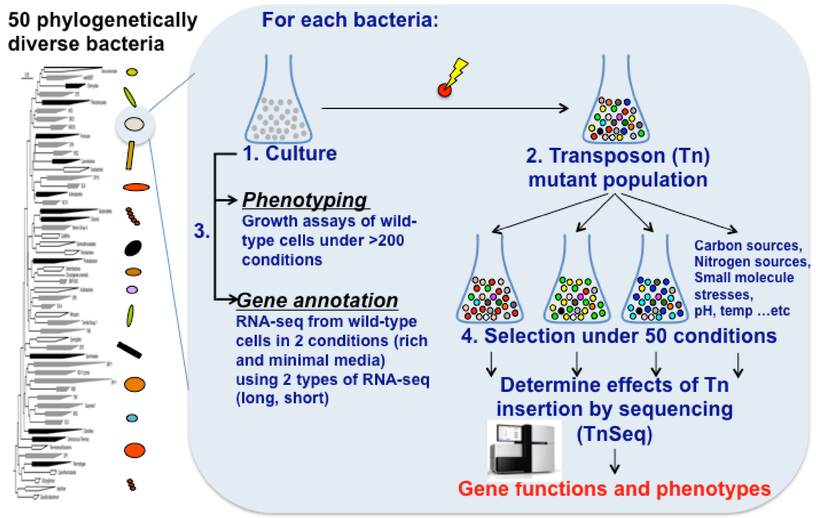
New annotations for >25% genes with unknown functions

In a proof of principle study, Deutschbauer et al. (ref 1) used transposon mutagenesis, growth assays and sequencing to annotate the function of >25% previously 'unknown' genes in *S. Oneidensis*. Here we propose to scale this approach to 50 diverse organisms.

Reference

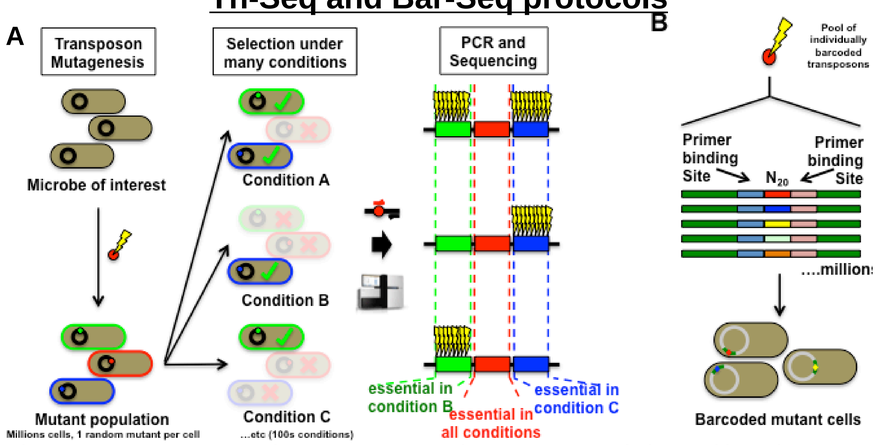
1. A. M. Deutschbauer et al., *PLoS Genetics* **7** (11), 238 (2011). e1002385

FEBA Gene function annotation workflow



Candidate bacteria were selected primarily based on phylogenetic diversity, but include several organisms with relevance to D.O.E. missions. Successfully cultured bacteria (1) will be subject to transposon mutagenesis through which we aim to obtain at least one mutant strain for every gene in the genome (2). Successfully mutagenized organisms will be subject to high throughput growth assays to determine wild-type phenotypic capabilities, and RNA-Seq to annotate gene structures (3). Mutant populations will then be subject to growth under relevant conditions followed by high-throughput sequencing using the Tn-Seq and Bar-Seq approaches (4 and below). Downstream computational analyses will determine fitness effects of each mutation and infer gene function.

Tn-Seq and Bar-Seq protocols

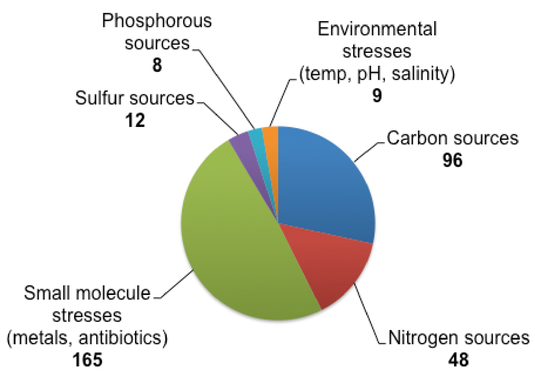


A) Transposon mutagenesis generates complex mutant populations with at least one inactivating mutation for every gene in the genome. Mutant populations are grown under a diverse set of growth conditions. Cells with inactivating mutations in genes important for survival in a condition will fall to low frequency in the final population. Changes in the abundance of individual strains are determined by high throughput sequencing and used to infer gene function. B) Bar-Seq. For more rapid characterization of mutant populations, transposons have been engineered to contains 20 random nucleotides flanked by common priming sites. These can be monitored by a simple PCR and sequencing assay.

FEBA Bacteria

Phylum	Organism	Cultured?	Transposon mutants ?	Full TnSeq library
Acidobacteria	Terriglobus roseus KBS 63, DSM 18391	YES		
Actinobacteria	Actinospica robiniae DSM 44927			
	Beutenbergia cavernae HKI 0122, DSM 12333			
	Corynebacterium glutamicum Kalinowski ATCC 13032	YES		
	Cryptosporangium arvum YU 629-21, DSM 44712			
	Promicromonospora kroppenstedtii RS16, DSM 19349	YES		
	Jiangella gansuensis YIM 002, DSM 44835			
	Saccharomonospora viridis P101, DSM 43017			
	Segniliparus rotundus CDC 1076, DSM 44985			
	Rubrobacter radiotolerans DSM 5868			
	Patulibacter minatonensis KV-614, DSM 18081	YES		
Aquificae	Hydrogenobacter thermophilus TK-6, DSM 6534			
Bacteroidetes	Aquiflexum balticum BA160, DSM 16537	YES		
	Belliella baltica BA134, DSM 15883	YES		
	Cyclobacterium marinum DSM 745			
	Echinicola vietnamensis KMM 6221, DSM 17526	YES		
	Adhaeribacter aquaticus MBRG1.5, DSM 16391			
	Emticicia oligotrophica GPTSA100-15, DSM 17448			
	Flectobacillus major VKMB-859, DSM 103			
	Flexibacter litoralis Fx I1, DSM 6794	YES		
	Hymenobacter roseosalivarius AA-718, DSM 11622	YES		
	Pontibacter actiniarum KMM 6156, DSM 19842	YES		
	Runella slithyformis LSU4, DSM 19594			
	Spirosoma linguale DSM 74			
	Fluviicola taffensis RW262, DSM 16823	YES		
	Owenweeksia hongkongensis DSM 17368			
	Gillisia limnaea R-8282, DSM 15749	YES		
	Weeksella virosa 9751, DSM 16922			
	Aequorivita sublithincola QSSC9-3, DSM 14238	YES		
	Joostella marina En5, DSM 19592	YES		
	Niabella soli JS13-8, DSM 19437	YES		
	Niastella koreensis GR20-10, DSM 17620			
	Haliscomenobacter hydrossis O, DSM 1100	YES		
	Pedobacter heparinus HIM 762-3, DSM 2366			
	Saprospira grandis HR1, DSM 2844			
Euryarchaeota	Halorhabdus utahensis AX-2, DSM 12940	YES		
Firmicutes	Alicyclobacillus acidocaldarius acidocaldarius 104-IA, DSM 446			
Planctomycetes	Planctomyces brasiliensis IFAM 1448, DSM 5305			
	Planctomyces limnophilus Mu 290, DSM 3776			
Proteobacteria	Dinoroseobacter shibae DFL-12, DSM 16493	YES	YES	
	Phaeobacter gallaeciensis BS107	YES	YES	
	Desulfovibrio vulgaris Miyazaki F	YES	YES	
	Sulfurospirillum deleyianum 5175, DSM 6946			
	Shewanella amazonensis SB2B	YES	YES	
	Shewanella oneidensis MR-1	YES	YES	
	Escherichia coli BW25113	YES	YES	
	Alcanovorax jadensis	YES	YES	
	Kangiella aquimarina	YES	YES	
	Pseudomonas stutzeri RCH2	YES	YES	YES
	Thiothrix nivea JP2, DSM 5205			
Spirochaetes	Turneriella parva H, DSM 21527			
Thermi	Deinococcus peraridilitoris KR-200, DSM 19664			
	Deinococcus pimensis KR-235, DSM 21231	YES		
	Deinococcus hopiensis KR-140, DSM 18049			
	Meiothermus chliarophilus ALT-8, DSM 9957			
10 Phyla	**55 organisms 'in hand'**	**27 cultured organisms**	**9 mutagenized organisms**	**1 high complexity mutant library**

FEBA Growth conditions



~300 total conditions

Progress so far

- Synthesis of a library of barcoded transposons for mutagenesis
- Development of a sequencing pipeline for high throughput Tn-Seq and Bar-Seq.
- Proof of principle mutagenesis and sequencing of *P.Stutzeri* to reveal new gene functions in this bacteria.
- Generation of transposon mutants for 8 new bacterial species

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.