

16th International Microbial Genomes Conference



September 14-18, 2008

UCLA Conference Center
Lake Arrowhead
California

keys change everywhere

SCIENTIFIC PROGRAM

SUNDAY, SEPTEMBER 14

- 4:00-6:00 pm Arrival and Check-in at Lake Arrowhead Conference Center
- 6:15-7:45 pm Dinner (Dining Room)
- Opening of Meeting** (Pineview Room)
- 7:45-8:05 pm Jeffrey H. Miller
University of California, Los Angeles
“Welcome”
- 8:05-9:00 pm **Keynote Address**
James C. Liao
University of California, Los Angeles, CA
“Non-fermentative Pathways for Synthesis of Branched-Chain
Higher Alcohols as Biofuels”
- 9:00 pm Reception (Iris Room)

MONDAY, SEPTEMBER 15

- 7:45-8:30 am Breakfast (Dining Room)
- Opening Session** (Pineview Room)
- 8:45-9:00 am Introduction/Announcements
- Session I** **Microbial communities I: Biodiversity / Interaction / Evolution**
- 9:00-9:30 am Jonathan Eisen
University of California, Davis, CA
“A Genomic Encyclopedia of Bacteria and Archaea”
- 9:30-10:00 am Bernhard Palsson
University of California, San Diego, CA
“The Genetic Basis for Adaptive Evolution in *E. coli*”
- 10:00-10:30 am George Weinstock
The Genome Center at Washington University, St. Louis, MO
“The Human Microbiome: Progress and Challenges”

SCIENTIFIC PROGRAM ORGANIZERS

Dr. Jeffrey H. Miller, Chair
University of California, Los Angeles

Dr. George Weinstock
The Genome Center at Washington University, St. Louis, MO

Dr. Cheryl Kerfeld
DOE Joint Genome Institute, Walnut Creek, CA

Dr. Elisabeth Raleigh
New England BioLabs, Beverly, Massachusetts

Dr. Jizhong Zhou
University of Oklahoma, Norman, Oklahoma

Dr. Fredrick Blattner
University of Wisconsin, Madison, Wisconsin

Dr. Ashlee Earl
Harvard Medical School, Boston, Massachusetts

ACKNOWLEDGEMENTS

The organizers of the conference gratefully acknowledge the contributions of the following for their support:

Department of Energy

National Institute of Allergy and Infectious Diseases
National Institute of Health

New England Biolabs

Scarab Genomics, LLC

DNASTAR, INC.

CONTACT NUMBER

The Arrowhead Conference Center Phone number is (909) 337-2478

10:30-10:50 am	Break
10:50-11:20 am	Gary Siuzdak The Scripps Research Institute, La Jolla, CA "Metabolomics Reveals Large Effect of Gut Microflora on Biochemistry and Activation of a Host Response"
11:20-11:50 am	Jim Bristow DOE Joint Genome Institute, Walnut Creek, CA "Microbial Sequencing for Biofuels Applications"
11:50 am-12:20 pm	K. Eric Wommack University of Delaware, Newark, DE "Making sense of the chaff: What will metagenomic approaches tell us about viral ecology?"
12:30 pm	Lunch (Dining Room)
4:00-6:00 pm	Poster Session (Lakeview Room) Social/Mixer (Lakeview Room)
6:15-7:45 pm	Dinner (Dining Room)
Session II	New Genomes and Strategies for Genome Sequence Annotation
7:45-8:15 pm	Valérie de Crécy-Lagard University of Florida, Gainesville, FL "Making Sense of Genomes: Linking Gene and Function by Comparative Genomics"
8:15-8:45 pm	Trent Northen The Scripps Research Institute, La Jolla, CA "High Throughput Mass Spectrometry Based Metabolomic and Enzymatic Assays for Functional Genomics"
8:45-9:00 pm	Break
9:00-9:30 pm	Matteo Pellegrini University of California, Los Angeles, CA "New Methods for Processing High-throughput Sequencing Data: Improving the Solexa/Illumina Data Analysis Pipeline"

9:30-10:00 pm Sabeeha S. Merchant
University of California, Los Angeles, CA
**"Transcriptomics of Nutritional Copper Homeostasis in
Chlamydomonas"**

TUESDAY, SEPTEMBER 16

7:45-8:30 am Breakfast (Dining Room)

Session III Pathogens / Resistance

8:45-9:00 Lynn L. Silver
LL Silver Consulting, LLC, Springfield, NJ
"The State of Antibacterial Discovery in 15 minutes"

9:00-9:30 am Julian Parkhill
Wellcome Trust Sanger Institute, Cambridge, UK
**"Discovering Variation in Genetically Monomorphic Bacteria:
SNPs and the Evolution of Salmonella Typhi"**

9:30-10:00 am Tiffany Williams
Baylor College of Medicine, Houston, TX
**"Global Health and Next Generation Sequencing
Technologies: *Streptococcus pneumoniae* serotype 1 in Africa"**

10:00-10:30 am Kim Lewis
Northeastern University, Boston, MA
"Persister Cells and Biofilm Resistance"

10:30-10:50 am Break

10:50-11:20 am Jeffery F. Miller
University of California, Los Angeles, CA
"Diversity-Generating Retroelements"

11:20-11:50 am Steven A. Benner
Foundation for Applied Molecular Evolution, Gainesville, FL
"Molecular Paleoscience"

11:50 am-12:20 pm Joe Zhou
University of Oklahoma, Norman, OK
**"Metagenomics Insights of the Feedback Responses of a
Grassland Ecosystem to Elevated Atmospheric CO₂"**

12:30 pm Lunch (Dining Room)

4:00-6:00 pm	Poster Session (Lakeview Room) Social/Mixer (Lakeview Room)
6:15- 7:45 pm	Dinner (Dinning Room)
Session IV	Microbial Communities II: Metagenomics/Biodiversity
7:45- 8:15 pm	Heather Allen University of Wisconsin-Madison, WI “Using Functional Metagenomics to Discover Antibiotic Resistance Genes in Natural Environments”
8:15- 8:45 pm	David A. Relman Stanford University School of Medicine, Stanford, CA “Response of the Human Distal Gut Mircobiota to Disturbance: The Effect of Antibiotics”
8:45- 9:00 pm	Break
9:00- 9:30 pm	Ashlee Earl Harvard Medical School, Boston, MA “<i>Bacillus subtilis</i> Biofilm Diversity”
9:30-10:00 pm	Mary E. Lidstrom University of Washington, Seattle, WA “Coupling Function to Phylogeny via Single-Cell Phenotyping”

WEDNESDAY, SEPTEMBER 17

7:45-8:30 am	Breakfast (Dining Room)
Session V	Bioenergetics / Regulatory and Metabolic Pathways / Modeling
8:45-9:15 am	John Dueber University of California, Berkeley, CA “Use of Synthetic Protein Scaffolds to Balance Pathway Flux of Engineered Metabolic Pathways”
9:15-9:45 am	Maria L. Ghirardi National Renewable Energy Laboratory, Golden, CO “Hydrogen Fuel Production by Microalgae: Issues and Future Directions“

9:45-10:15 am	Caroline S. Harwood University of Washington, Seattle, WA "Redirection of Metabolism of Hydrogen Production"
10:15-10:35 am	Break
10:35-11:05 am	Byung-Kwan Cho University of California, San Diego, CA "The Reconstruction of the Transcriptional Regulatory Network in <i>E. coli</i>"
11:05-11:35 am	E. Virginia Armbrust University of Washington, Seattle, WA "Molecular Insights into Silicon Bioprocesses in Marine Diatoms"
11:35 am-12:05 pm	Athanasios Typas University of California, San Francisco "High-throughput Quantitative Analysis of Genetic and Chemical-Genetic Interactions in <i>E. coli</i>"
12:05 – 12:20 pm	Devaki Bhaya Carnegie Institution for Science, Stanford, CA "Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes."
12:30 pm	Lunch (Dining Room)
Session VI A	Undergraduate Education Programs - Insights
4:30-5:00 pm	Erin Sanders-Lorenz University of California, Los Angeles, CA "Integrating Discovery-based Undergraduate Research Experiences into UCLA Courses Using a Collaborative Curriculum Model"
5:00-5:30 pm	Cheryl Kerfeld DOE Joint Genome Institute, Walnut Creek, CA "The JGI Microbial Genome Annotation Program"
5:30-6:00 pm	Fredrick Blattner University of Wisconsin, Madison, Wisconsin "Insights from the Genomes of Commonly Used Lab Strains"
6:15-7:45 pm	Dinner

Session VI B	Genomics / New Methods
7:45-8:15 pm	Colin J. Ingham Wageningen University, Wageningen, The Netherlands “Reinventing the Petri Dish: Miniaturized Tools for High Throughput Microbial Culture”
8:15-8:45 pm	Maureen Hillenmeyer Stanford University, Stanford, CA “The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes”
8:45-9:15 pm	Simon Prochnik DOE – Joint Genome Institute, Walnut Creek, CA “The Genome Sequence of the Deep-Branching Amoeboflagellate <i>Naegleria gruberi</i> Reveals Ancestral Eukarotic Functions”
9:15 pm	Reception/Party (Iris Room)

THURSDAY, SEPTEMBER 18

7:30-8:20 am	Breakfast (Dining Room)
Session VII	Regulatory Mechanisms / New Methods
8:30-9:00 am	Elizabeth Fozo National Institute of Allergy and Infectious Diseases, Bethesda, MD “Regulating Bacterial Expression of Short Hydrophobic Toxic Proteins with Small RNAs”
9:00-9:30 am	Mariusz Nowacki Princeton University, Princeton, NJ “RNA-mediated Epigenetic Programming of a Genome-Rearrangement Pathway”
9:30-9:55 am	Barry L. Wanner Purdue University, West Lafayette, IN “Development of the www.EcoliHub.org Information Resource”

9:55-10:20	Hirotsada Mori Nara Institute of Science and Technology, Nara, Japan "Systematic Analysis of Genetic Interaction of <i>Escherichia coli</i>"
10:45 am	Check-out
11:00 am	Departure of 1 st Conference Bus for LAX
12:00 pm	Lunch (Dining Room)
1:15 pm	Departure of Vans to LAX

SPEAKER ABSTRACTS

Non-Fermentative Pathways for Synthesis of Branched-Chain Higher Alcohols as Biofuels

James C. Liao

Department of Chemical and Biomolecular Engineering

University of California, Los Angeles, CA 90095, USA

Global energy and environmental problems have stimulated increased efforts in synthesizing biofuels from renewable resources. Compared to the traditional biofuel, ethanol, higher alcohols offer advantages as gasoline substitutes because of their higher energy density and lower hygroscopicity. In addition, branched-chain alcohols have higher octane numbers compared to their straight-chain counterparts. However, these alcohols cannot be synthesized economically using native organisms. Here we present a metabolic engineering approach using *Escherichia coli* to produce higher alcohols including isobutanol, 1-butanol, 2-methyl-1-butanol, 3-methyl-1-butanol and 2-phenylethanol from a renewable carbon source, glucose. This strategy leverages the host's highly active amino acid biosynthetic pathway and diverts its 2-keto acid intermediates for alcohol synthesis. In particular, we have achieved high yield, high specificity production of isobutanol from glucose. The strategy enables the exploration of biofuels beyond those naturally accumulated to high quantities in microbial fermentation.

A genomic encyclopedia of bacteria and archaea

Jonathan A. Eisen

U. C. Davis and the Joint Genome Institute

There is a glaring gap in microbial genome sequence availability – the currently available genome sequences show a highly biased phylogenetic distribution compared to the extent of microbial diversity known today. This bias has resulted in a major limitations in our knowledge of microbial genome complexity and our understanding of the evolution, physiology and metabolic capacity of microbes. Although there have been small efforts in sequencing genomes from across the tree of life for microbes, there are no systematic efforts.

There are many reasons why phylogenetic based sequencing in theory should be of great benefit including: (a) improved identification of protein families and orthology groups across species, which will improve annotation of other microbial genomes (b) improved phylogenetic anchoring of metagenomic data, (c) gene discovery (which tends to be maximized by selecting phylogenetically novel organisms, (d) a better understanding of the processes underlying the evolutionary diversification of microbes (e.g., lateral gene transfer and gene duplication) (e) a better understanding of the classification and evolutionary history of microbial species and (f) improved correlations of phenotype and genotype in microbes.

Based on the potential benefits, we (JGI) has commenced a pilot project to create a Genomic Encyclopedia of Bacteria and Archaea (GEBA). In this pilot, we plan to sequence ~100 genomes selected based on their phylogenetic novelty. This is being done at two phylogenetic scales. About 60 of the genomes are from across the breadth of bacteria and archaea. The remaining 40 genomes are from within the Actinobacteria. By doing this two tiered selection we can test both the value of breadth from across the bacteria and archaea as well as the value of filling in the phylogenetic gaps within a single phyla.

In my talk I will summarize the project and report on the sequencing and analysis of the first 56 genomes. I will discuss how we are using this pilot to test protocols that could be used for a scale up of the GEBA project or for any other large scale microbial sequencing project. In addition I will discuss how collaborations with culture collections can be valuable in such a project. Finally, I will report on the results of tests of the value of phylogenetic based sequencing.

The Genetic Basis for Adaptive Evolution in *E. coli*

Bernhard Palsson

University of California, San Diego, CA

This talk will focus on recent results obtained from genomic re-sequencing of *E. coli* after adaptive evolution to identify all of the acquired sequence changes. These experiments were conceived, designed and the outcomes predicted based on a genome-scale reconstruction of *E. coli*'s metabolic network. The discovered mutations were characterized in vivo by allelic replacement in the starting strain [1]. The mutations were then used as genetic markers in competition experiments to measure the relative fitness gains between adaptive endpoints and between same-gene mutations, by measuring allelic frequency as a function of evolutionary time [2]. These experiments enable multi-scale investigation of adaptation; including molecular, systems and population biology.

Key refs:

1. Herring, C.D., et al., Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet*, 2006. 38(12): p. 1406-1412.
2. Applebee, M.K., M.J. Herrgard, and B.O. Palsson, Impact of Individual Mutations on Increased Fitness in Adaptively-evolved strains of *Escherichia coli*. *J Bacteriol*, 2008.

The Human Microbiome: Progress and Challenges

George Weinstock for the NIH Human Microbiome Project Consortium
The Genome Center, Washington University School of Medicine
St. Louis, Missouri, USA

The human body contains ten times or more microbes than human cells, and this microbiome is expected to have profound consequences for human health and disease. The microbes live in communities in the many different environments the body offers, at internal sites such as the gastrointestinal tract, sites exposed to the external environment such as the skin, and intermediate sites such as the oral cavity. Within each of these are many sub-environments that house communities that differ in structure. Thus the oral cavity has both subgingival and supragingival communities, that can vary from tooth to tooth in an individual. Often these regions contain hundreds of different species, varying in prevalence over orders of magnitude. It is the goal of the Human Microbiome Project to define these communities, correlate how changes in their structure contribute to health or cause disease, and understand the mechanisms for these effects. To begin to approach this, the genome centers are Washington University, Baylor College of Medicine, the Broad Institute, the J. Craig Venter Institute, and the NIH Sequencing Center are working collaboratively on a series of projects that both address technical and biological issues.

Metabolomics reveals large effect of gut microflora on biochemistry and activation of a host response

William R. Wikoff¹, Eric C. Peters², Gary Siuzdak¹

¹Department of Molecular Biology and The Center for Mass Spectrometry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

²Genomics Institute of the Novartis Research Foundation, San Diego, California 92121

Gut microflora (microbiome) have a significant impact on human health. Untargeted metabolomics was used to investigate the effect of gut microbes on plasma biochemistry in a mouse model. Plasma from germ free mice was compared with wild type using a novel combination of metabolomics methods to obtain global coverage. More than 100 plasma molecules were unique to wild type, arising from bacterial metabolism. The concentration of at least ten percent of all endogenous compounds changed by 50% or more simply from the effect of the gut microbiome. Three effects of the microbiome were observed: endogenous mammalian metabolism is altered, unique bacterial metabolites are produced by the microbiome which appear directly in the blood, and bacterial metabolites produce a host response via conjugation in phase II drug metabolism pathways. There is thus a complex effect on blood biochemistry, in which the metabolic pathways of bacteria and host partially intersect. Two pathways were particularly affected: tryptophan/indole metabolism, and metabolism of phenyl compounds. Plasma serotonin, although not produced by bacteria, was almost 3-fold higher in wild type mice, while tryptophan and acetyl-tryptophan levels were reduced. Indoxyl sulfate and indole-3-propionic acid were present only in wild type mice. Two types of phase II metabolism were observed: sulfation and glycine conjugation. Phenyl compounds generated in the gut by bacterial action are conjugated to sulfated forms such as phenyl sulfate and p-cresol sulfate in the liver and appear in plasma. Exogenous aromatic bacterial compounds such as benzoic, phenylacetic and cinnamic acids are conjugated with glycine to produce hippuric acid, phenylacetyl glycine and cinnamoyl glycine, respectively. We demonstrate the surprisingly large effect of the gut microbiome on plasma biochemistry, the overlap between bacterial and mammalian metabolism, and the fuzzy line between endogenous and exogenous.

Microbial Sequencing for Biofuels Applications

James D. Bristow, Deputy Director, Joint Genome Institute, Lawrence Berkeley Ntl. Lab.

The Dept of Energy has set a goal of replacing 30% of US transportation fuel needs with biofuels by 2030. The cornerstone of this goal is the production of 60 billion gallons of ethanol from 1 billion tons of biomass. This aggressive goal will require significant enhancements to current biomass degradation and fermentation strategies. JGI's microbial genome program will contribute to this effort through sequencing and analysis of cellulose-degrading bacterial isolates, microbial communities obtained from a variety of environments (including insect and animal gut), and biomass degrading fungi. In addition, JGI is sequencing evolved species that display improved ethanol production from complex sugar mixtures and ethanol tolerance. Examples of these projects will be discussed.

Making sense of the chaff: What will metagenomic approaches tell us about viral ecology?

K. Eric Wommack, Assoc. Professor
University of Delaware

High-throughput sequencing of viral genomic nucleic acids from environmental samples, is adding exquisite genetic detail to our view of natural viral assemblages. Indeed, as compared to microbial consortia, viral assemblages are ideally suited to metagenomic approaches owing to the small coding-dense nature of viral genomes. For example, we now appreciate that in aquatic environments the phenomena of 'viral photosynthesis' is common and widespread; and that the extraordinary genotypic diversity seen in local viral assemblages may rival that of the entire global marine virome. It is quite possible that dsDNA viruses alone are among the greatest reservoirs of genetic diversity in the biosphere. Despite these and other notable advances, one consistent message we've obtained from the analysis of viral sequence data (metagenome or genome) is that the majority of viral genes are completely novel or show homology to only genes of unknown function. Ironically, it is the divergent nature of viral genes that has restricted the utility of short-read sequencing for functional characterization of natural viral assemblages. Considering the rapid production and turnover rates of viral assemblages, it is quite probable that viral genes are among the most actively expressed and replicated genes on earth. Thus, a significant challenge for the coming third decade of the 'third age of phage' is to leverage metagenomic resources for an increasingly sophisticated view of the role of viral processes within microbial communities. This talk will review the current status of viral metagenomics and propose paths forward to meeting this challenge.

Viral metagenomics

454 GS20
data not
good

AEM 74:1145

Increased
coverage
does not
help

3/20/08
Prokaryotic
Half full of
education
paper

4/20/08
don't know
how few
bacterial
genes

(5) compact
multiple

Hy metagenomics
The "Everest effect"
because it's there

"Who cares for the
complex"

"The Noah effect" - let's collect
all possible genes

Eisen? - how many phage in E are detected

Making sense of genomes: linking gene and function by comparative genomics

Valérie de Crécy-Lagard, University of Florida, Gainesville, FL

Even if linking gene and function has been one of the major goals of biology in the last 100 years, the genome sequencing efforts of the last decade have revealed how little is known about the relationship between DNA sequences and biological functions. In the best genetically characterized organisms, a third of the genes have no assigned function and this number can climb to two-thirds in many less studied species. Traditionally the link between gene and function was inferred from genetically or biochemically derived association but the availability of nearly 1000 genomes has led to the emergence of accessible comparative genomics methods. These methods integrate several types of data and use filters, sieves and associations to make predictions that can then be tested experimentally. An unknown gene's function may thus be predicted from those of its associates: the 'guilt by association' principle. In practice it is often 'guilt by multiple association' as genes can be associated in several ways and analyzing more than one of these improves predictions. Associations that can be derived from whole genome data sets include: gene clustering, gene fusion events, phylogenetic occurrence profiles or signatures, and shared regulatory sites. Post-genomic experimental sources such as protein interaction networks, gene expression profiles and phenomics data can also be used to find associations. Application of these methods both to conserved families of unknown function and to functions that lacked any associated genes will be discussed.

High Throughput Mass Spectrometry Based Metabolomic and Enzymatic Assays for Functional Genomics

Trent R. Northen*^{†§}, Oscar Yanes[†], Linh Hoang[†], Steven M. Yannone[§], Jason Raymond[‡], Jinq-Chyi Lee[‡], Chi-Huey Wong[‡], John Tainer^{†§}, Gary Siuzdak^{†§}

[§]Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA. [†]Department of Molecular Biology and Scripps Center for Mass Spectrometry, [‡]Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA; [¶]School of Natural Sciences, University of California, Merced, 5200 North Lake Road, Merced, CA 95343 USA

Mass spectrometry's ability to efficiently generate intact biomolecular ions in the gas phase has led to a wide range of biological applications and is recently being applied for global metabolite profiling ('metabolomics'). This makes it particularly well suited for functional genomic studies since the metabolome is perhaps the 'omic' most closely related to phenotype. We recently introduced Nanostructure-Initiator Mass Spectrometry (NIMS, Nature 2007), a new high throughput and high sensitivity tool for spatially defined mass analysis which complements existing methods by enabling the analysis and imaging of metabolites from tissues, single cells, and microbial communities. NIMS utilizes 'initiator' molecules trapped within nanostructured surfaces or 'clathrates' to release/ionize intact molecules adsorbed from the surface upon laser or ion beam irradiation. This technology has now been extended as a Nanostructure-Initiator Mass Spectrometry enzymatic (Nimzyme, PNAS 2008) assay where enzyme substrates are immobilized on the mass spectrometry surface using fluorophilic phase interactions. This 'soft' immobilization allows efficient desorption/ionization and allows surface washing to reduce signal suppression from complex biological samples as a result of the preferential retention of the tagged products and reactants. In addition, it is shown that Nimzyme can differentiate multiple and competing enzymatic activities. These attributes make it complementary to metagenomic studies by allowing sequencing efforts to be focused on the samples with the most interesting or desirable activities. This approach was used to detect and screen optimal pH, temperature, and enzyme inhibition from a hot springs microbial community. The combination of this and metagenomic analysis was used to putatively ID the hydrolase of interest. NIMS and Nimzyme are well suited for both metabolite and activity profiling complex biological systems with extremely high sensitivity.

New methods for processing high-throughput sequencing data: improving the Solexa/Illumina data analysis pipeline

Shawn Cokus, Matteo Pellegrini
MCD Biology, UCLA

High-throughput sequencing has undergone remarkable increases in efficiency over the past few years. Some of the most efficient machines are those produced by Illumina, which sequence approximately 1-3 billion bases every three days as millions of ~36 base reads. We have developed novel techniques to process and interpret such data that extend the standard data analysis pipeline. There are two main areas we address with our software: a probabilistic base calling module and a mapping tool that aligns short reads to reference sequences while accounting for base call quality.

The first aspect of our software involves the estimation of the sequence of each read. Our base calling tool starts with Solexa base calls and attempts to correct systematic errors in these using multidimensional Gaussian mixture models. Starting with Solexa base calls we estimate the distribution of each base using Gaussian hyper-ellipsoids. These are then used to infer the probability that each base in a read is an A, C, G or T.

Our second tool aligns these probabilistic reads to sequenced genomes using suffix trees. Full use is made of information available at every base of every read as to the probability of A, C, G, and T, and scoring of reads is statistically grounded via algorithmic, non-heuristic consideration of whole-genome likelihoods. Our tool also supports sodium bisulfite-converted reads suitable for study of DNA cytosine methylation at the single base level.

We apply these tools to a variety of data sets. The first involves the estimation of DNA methylation in *Arabidopsis* using bisulfite sequencing. The second involves the measurement of the transcriptional landscape in *Chlamydomonas*. We use this data to estimate transcript levels of each gene, differential expression in differing conditions and to correct existing gene models.

Transcriptomics of nutritional copper homeostasis in *Chlamydomonas*

Sabeeha S. Merchant^{*1,3}, David Casero², Madeli Castruita¹, Shawn Cokus², Christian Haudenschild⁴, Steven Karpowicz¹, Janette Kropat¹ and Matteo Pellegrini^{2,3}. Departments of ¹Chemistry and Biochemistry and ²Molecular, Cell and Developmental Biology, ³Molecular Biology Institute and DOE Institute of Genomics and Proteomics, UCLA, Los Angeles, CA 90095, ⁴Illumina, Inc., Hayward, CA.

Chlamydomonas, a chlorophyte alga in the green plant lineage, is a choice model organism for the study of chloroplast-based photosynthesis and cilia-based motility. The 121 Mb draft genome sequence, determined at 13X coverage is estimated to encode approximately 15,000 protein coding genes. Besides the pathways for oxygen evolving photosynthesis, dark respiration of acetate and hydrogen production, the gene repertoire reveals less-studied pathways for fermentative metabolism, suggestive of extraordinary metabolic flexibility. The operation of these bioenergetic pathways is dependent on metal cofactors like copper, iron, manganese and zinc, and accordingly these elements are essential nutrients for *Chlamydomonas*. In a copper-deficient environment, *Chlamydomonas* will modify the photosynthetic apparatus by substituting a heme protein - Cyt *c*₆ - for an abundant copper protein - plastocyanin - that accounts for about half of the intracellular copper. This modification is viewed as a copper sparing mechanism and is dependent on a plant specific transcription factor CRR1. We have used digital gene expression (DGE) and RNA-Seq methodology to characterize the *Chlamydomonas* transcriptome under steady conditions of various degrees of copper-deficiency and in a bloom situation where cells deplete the copper as they divide. Both methods are quantitative and show excellent correlation with real time PCR indicative of a large dynamic range relative to microarrays. Direct vs. indirect responses to copper-deficiency are distinguished by comparison of the *crr1* transcriptome to that of wild-type cells. The analyses indicate previously unknown modifications of the photosynthetic apparatus and the potential for modification of bioenergetic pathways.

The State of Antibacterial Discovery in 15 Minutes

Lynn Silver, LL Silver Consulting

While new antibacterial drugs have been developed in the past ten years, most have been improvements upon existing classes. The pursuit of novel chemical classes inhibiting novel targets has been a mainstay of the discovery effort, but has yielded few developable candidates. In my view, the obstacles to success in the area include poor choice of targets, poor quality of chemicals used for screening, and the lack of a sustained effort to address the problems of bacterial entry and avoidance of efflux.

Discovering Variation in Genetically Monomorphic Bacteria: SNPs and the Evolution of *Salmonella* Typhi

Julian Parkhill, The Wellcome Trust Sanger institute, Cambridge, UK

Salmonella enterica serovar Typhi (*S. Typhi*), is the causative agent of typhoid fever, and is responsible for around 600,000 annual deaths worldwide. Recent studies have shown *S. Typhi* to be highly monomorphic, with only 88 SNPs found in nearly 200 gene fragments in 100 strains [Roumagnac I (2006) *Science*, **314**, 1301]. This extreme lack of variation is due to a recent clonal expansion of the organism, which is estimated to be around 10-40,000 years old. This lack of variation makes it extremely difficult to do fine-scale epidemiology and genotype-phenotype association in this important pathogen.

To discover more variation, we have generated whole genome sequence data from a further ~20 *S. Typhi* strains, using both Roche/454 and Illumina 1G machines. Using this approach, we have identified over 2,000 novel SNPs and other variants. The analysis of these variants shows that *S. Typhi* is evolving almost entirely neutrally, with little evidence for recombination or genes under positive selection. The consequences of this for our understanding of the biology and epidemiology of *S. Typhi* will be discussed.

Global Health and Next Generation Sequencing Technologies: *Streptococcus pneumoniae* serotype 1 in Africa

Tiffany Williams, Nick Loman, Lori Synder, Chinelo Ebruke, Wendy Keitel, Richard Adegbola, Mark Pallen, Martin Antonio and George Weinstock

Next generation sequencing technologies provide opportunity to improve global health through increased understanding of the molecular epidemiology, disease pathogenesis, and host-microbe interactions of infectious diseases. *Streptococcus pneumoniae* (*Spn*), the etiologic agent of invasive pneumococcal disease (IPD), is responsible for a significant amount of global morbidity and mortality, particularly within Africa and Southeast Asia. The 91 serotypes of *Spn* vary in their ability to colonize the nasopharynx and/or cause IPD. *Spn* serotype 1 (*Spn*-1) is rarely found in carriage and strongly associated with IPD. Within Africa, *Spn*-1 is the most common IPD-causing serotype. In addition to serotyping, multi-locus sequence typing (MLST) allows genetic grouping of pneumococci into sequence types (ST). Epidemiologic studies have shown the predominant African *Spn*-1 sequence types, ST217 and ST618, segregate into a geographically restricted clonal complex and cause severe IPD with poor clinical outcomes. These same isolates cause meningitis outbreaks within the “African meningitis belt” with clinical features similar to meningococcal epidemics known for their clinical severity and high mortality rates. Furthermore, a Gambian clinical trial evaluating a 9-valent pneumococcal glycoconjugate vaccine failed to show a reduction in *Spn*-1 IPD in contrast to significant decreases observed for other vaccine serotypes. It is clear these geographically restricted *Spn*-1 are genetically distinct and linked to clinically severe IPD not seen within the Americas or Europe. Better understanding of the genomic composition of *Spn*-1 within this clonal complex and its influence on clinical phenotype will help decrease pneumococcal mortality and ensure vaccine efficacy within Africa. Evidence supports the hypothesis that the clinical phenotype of IPD-causing *Spn*-1 B clonal complex isolates is determined by unique genetic elements identifiable via genomic sequencing and comparative genomic analysis. We aim to establish a pipeline which integrates the microbial sequencing infrastructure and expertise of a US-based genome sequencing center with the epidemiologic and microbiologic expertise of an African-based research institution in order to fully realize the potential next generation sequencing technologies bring to global health. Previous pneumococcal sequencing projects have focused on isolates from developed countries while the majority of IPD burden is located within developing countries. Thus work on *Spn*-1 strains from The Gambia may provide a more appropriate perspective into bacterial virulence and disease pathogenesis within this region.

Persister Cells and Biofilm Resistance

Kim Lewis, Northeastern University, Boston, MA

Bacterial populations produce a small number of dormant persister cells that exhibit multidrug tolerance. Persisters are not mutants and are produced stochastically from regular cells. All resistance mechanisms do essentially the same thing – prevent the antibiotic from hitting a target. By contrast, tolerance apparently works by shutting down antibiotic targets which protects cells from killing. The number of persisters in a growing population of bacteria rises at mid-log and reaches a maximum of ~1% at stationary state. Similarly, slow-growing biofilms produce substantial numbers of persisters. The ability of a biofilm to limit the access of the immune system components, and the ability of persisters to sustain an antibiotic attack could then account for the recalcitrance of such infections in vivo and for their relapsing nature. *P. aeruginosa* biofilm infections of cystic fibrosis patients is incurable with existing antibiotics. We find that late isolates of *P. aeruginosa* CF patients are high-persistence (*hip*) mutants. The genome sequences of early vs. late isolates provide insights into the mechanism of persister formation. Similarly, in vitro selection of *hip* mutants in *E. coli* combined with whole genome sequencing is providing information on the mechanism of persister formation. A transcription profile of isolated persisters indicates downregulated biosynthetic pathways, consistent with their dormant nature, and overexpression of toxin/antitoxin (TA) modules. Stochastic overexpression of toxins that inhibit essential functions such as translation may contribute to persister formation.

Diversity-Generating Retroelements

Jeff F. Miller

Department of Microbiology, Immunology and Molecular Genetics
UCLA School of Medicine
Los Angeles, CA, USA

Host-parasite interactions are often driven by mechanisms that promote genetic variability. In the course of our studies on bacterial pathogenesis, we discovered a group of temperate bacteriophages that generate diversity in a gene that specifies tropism for receptor molecules on host *Bordetella* species, which cause respiratory infections in humans and other mammals. This microevolutionary adaptation is produced by a novel “diversity-generating retroelement” (DGR) that combines the basic retroelement life cycle of transcription, reverse transcription and integration with site-directed, adenine-specific mutagenesis. Central to this process is a reverse transcriptase-mediated exchange between two repeats, one serving as an donor template (TR) and the other as a recipient of variable sequence information (VR). Recent work has focused on the genetic basis of diversity-generation. The directionality of information transfer is determined by the initiation of mutagenic homing (IMH) sequence present at the 3’ end of VR. We have demonstrated that DGR function occurs through a TR-containing RNA intermediate by a unique target-primed reverse transcription mechanism that precisely regenerates target sequences. This non-proliferative, “copy and replace” mechanism enables repeated rounds of protein diversification and optimization of ligand-receptor interactions. The potential utility of DGRs is illustrated by the identification of over 40 related elements in bacterial, phage, and plasmid genomes. DGRs are present in human pathogens (*Treponema*, *Legionella* spp.), human commensals (*Bacteroides*, *Bifidobacterium* spp.), green sulfur bacteria (*Chlorobium*, *Prosthecochloris* spp.), cyanobacteria (*Trichodesmium*, *Nostoc* spp.), magnetotactic bacteria (*Magnetospirillum* spp.), and many other diverse species. DGRs comprise a new family of retroelements with the potential to confer powerful selective advantages to their host genomes.

Molecular Paleoscience

Steven A. Benner

Foundation for Applied Molecular Evolution, Gainesville FL

Since the development of the MasterCatalog, which pre-computed evolutionary histories for all protein families in the genome sequence database, scientists at the FfAME have developed experimental paleogenetics and paleogenomics for small genomes. By resurrecting ancient proteins from extinct organisms for study in the laboratory, these fields allow the experimental method to be brought to bear on historical hypotheses. Especially interesting are hypotheses that connect the biology of single celled organisms with their immediate and planetary ecosystems. This allows genomics to address "function" in a Darwinian sense. This talk will develop the methods, and then illustrate these by understanding microbial function late in the Age of Dinosaurs, and then approximately two billion years ago deep in the history of eubacteria.

Metagenomics Insights of the Feedback Responses of a Grassland Ecosystem to Elevated Atmospheric CO₂

Meiying Xu^{1,2}, Zhili He¹, Laurie E. Kellogg¹, Sanghoon Kang¹, Liyou Wu¹, Peter B. Reich³, Sarah Hobbie³, and **Jizhong Zhou^{1,*}**

¹Department of Botany and Microbiology, Institute of Environmental Genomics, University of Oklahoma, Norman, OK 73072, ²Guangdong Institute of Microbiology, Guangzhou, China, 510070, ³Department of Forest Resources, the University of Minnesota, St. Paul, MN 55108
(*jzhou@ou.edu; Phone: 405-325-6073; Fax: 405-325-7552)

Understanding the responses and mechanisms of biological communities to global environmental changes, especially anthropogenic change such as elevated atmosphere carbon dioxide, is a central issue in ecology and for society. Although the fertilization effects of elevated CO₂ on plant growth and productivity are well documented, its influences on belowground ecosystem functioning processes and associated microbial communities are poorly understood. Here, we used various integrated metagenomics approaches to characterize the responses of the belowground grassland microbial communities to elevated CO₂ in a multifactor FACE (Free Air Carbon Dioxide Enrichment) experimental facility, BioCON (Biodiversity, CO₂, and Nitrogen deposition), at the Cedar Creek Natural History area in Minnesota. A total of 24 soil samples grown with 16 plant species from both elevated and ambient CO₂ were analyzed with GeoChip 3.0, PhyloChip, 454 tag sequencing, and OminiLog. The GeoChip used contains ~25,000 probes and covers ~47,000 sequences for 292 gene families, which is the most comprehensive functional gene arrays for biogeochemical and ecological studies. Our results showed that elevated CO₂ significantly altered the belowground microbial community functional structure after ten-year field exposure of grass species to CO₂ although the overall microbial community diversity appears to be not affected. Also, Canonical Correspondence Analysis (CCA) indicated that CO₂, plant biomass and soil chemistry played significant roles in shaping microbial community structure. In addition, further analysis indicated that the abundance of the functional genes involved in nitrogen fixation, denitrification and carbon fixation are significantly increased. These results have important implications for the feedback responses of grassland ecosystems to elevated CO₂ as well as for the global climate modeling.

Using functional metagenomics to discover antibiotic resistance genes in natural environments

Heather K. Allen. University of Wisconsin, Madison.

Despite the threat posed by antibiotic resistance in infectious bacteria, little is known about the diversity, distribution, and origins of resistance genes. Soil and insects may be important environmental reservoirs of antibiotic resistant bacteria, but their significance has not been explored. We undertook functional metagenomic analyses of a remote Alaskan soil and of culturable bacterial isolates from gypsy moth midguts to investigate antibiotic resistance genes in two environments with different potential to affect human health. The Alaska site is on an island in the Tanana River in the Bonanza Creek Long-Term Ecological Research Station near Fairbanks. We selected 12 Gigabases of DNA in eight metagenomic libraries for resistance to β -lactam antibiotics. We report that this soil is a reservoir for β -lactamases that function in *Escherichia coli*, including ancestral β -lactamases and the first bifunctional β -lactamase. In contrast to soil, which is sedentary, gypsy moths are an invasive species in the northeastern United States that feed on a wide variety of plants as they move westward. To identify antibiotic resistance genes associated with gypsy moth midgut bacteria, we built metagenomic libraries from a pool of 44 cultured isolates from wild and lab-reared gypsy moth larvae. We identified three types of genes responsible for resistance to diverse antibiotics: an efflux pump, a transcriptional regulator, and an extended-spectrum β -lactamase. Our findings suggest that even in the absence of selective pressure imposed by anthropogenic activity, a soil microbial community in an unpolluted site harbors unique and ancient β -lactam resistance determinants, and an insect midgut carries genes encoding diverse mechanisms of antibiotic resistance. All of the genes reported here confer resistance on *E. coli* without tampering with the native gene expression machinery on the clones or in the host, demonstrating the potential for resistance genes from the environment to compromise human health if transferred to pathogens.

- what if
have a diff
flx in
E. coli then
in natural
host
- flx/metagenomics
 - β -lactam resistance
 - Get wide diversity using flx/screen
 - Mistake w/ ancestrally related statement
- "Lynne earlier - ask was that
today or yesterday"

Response of the human distal gut microbiota to disturbance: the effect of antibiotics

David A. Relman

Stanford University, Medicine and Microbiology & Immunology, Stanford, CA

The human intestinal microbiota is essential to the health of the host, and plays a role in nutrition, development, metabolism, pathogen resistance, and regulation of immune responses. Antibiotics may disrupt these coevolved interactions, leading to acute or chronic disease in some individuals. Our understanding of antibiotic-associated disturbance of the microbiota has been limited by the poor sensitivity, inadequate resolution, and significant cost of current methods. The use of pyrosequencing technology to generate large numbers of 16S rDNA sequence tags circumvents these limitations, and has been shown to reveal previously unexplored aspects of the 'rare biosphere'.

As part of a larger study, we have investigated the distal gut bacterial communities of three healthy humans before and after treatment with ciprofloxacin, obtaining more than 7,000 full-length rRNA sequences and over 900,000 pyrosequencing reads from two hypervariable regions of the rRNA gene. We have found that the taxonomic information obtained with these methods is concordant. Pyrosequencing of the V6 and V3 variable regions identified 3300-5700 taxa that collectively accounted for over 99% of the variable region sequence tags that could be obtained from these samples. Ciprofloxacin treatment influenced the abundance of about a third of the bacterial taxa in the gut, decreasing the taxonomic richness, diversity and evenness of the community. However, the magnitude of this effect varied among individuals, and some taxa showed interindividual variation in the response to ciprofloxacin. While differences of community composition between individuals were the largest source of variability between samples, we found that two unrelated individuals share a surprising degree of community similarity. In all 3 individuals, the taxonomic composition of the community closely resembled its pre-treatment state by 4 weeks after the end of treatment, but several taxa failed to recover within 6 months.

These pervasive effects of ciprofloxacin on community composition contrast with the reports by participants of normal intestinal function, and with prior assumptions of only modest effects of ciprofloxacin on the intestinal microbiota. These observations support the hypothesis of functional redundancy in the human gut microbiota. The rapid return to the pretreatment community composition is indicative of factors promoting community resilience, the nature of which deserves future investigation.

(cause or effect)
in some environments the function
is very robust yet in others
the diversity is at the top

Other features
- unevenness
- few archaea
- individuality
- spatial pattern

Bacillus subtilis Biofilm Diversity

Ashlee M. Earl^{1*}, ²Richard Losick and ¹Roberto Kolter. ¹Harvard Medical School, Boston, MA, ²Harvard University, Cambridge, MA.

The bacterial species *Bacillus subtilis* has been a workhorse for the study of molecular genetics and bacterial development for decades. As with most model organisms, there has been a trade-off, sacrificing breadth for depth, in our analysis of this species by focusing almost exclusively on one strain, *B. subtilis* 168. Although we have an unparalleled understanding of the inner workings of this bacterium, we now recognize that the genotypic and phenotypic potential of the species cannot be accounted for by the characterization of one strain alone. Our recent comparative analysis of globally diverse *B. subtilis* strains revealed not only marked genome diversity among species members, but also remarkable phenotypic diversity. This is particularly true when isolates are grown under conditions that encourage the development of biofilms; structured communities of microbes adhered to one another and to surfaces by a self-produced extracellular matrix. Considering the ecological significance of life within biofilms, we were struck by the impressive range in biofilm morphologies. While a great deal is currently known regarding both the structural and regulatory players involved in forming these complex multicellular communities, this insight has largely come from the study of one strain. We are currently working to unravel the genetic basis for the observed phenotypic variation among members of this species using both comparative genome analysis and more classic genetic techniques. Our analyses thus far reveal that the major structural and regulatory elements are highly conserved among strains. We now speculate that differences in the expression patterns of major biofilm regulators account for the observed morphological variation.

"a wonderful scientist
as of course are all these
people"

→ inside-outs. & edge
samples dolphins in 50 Bay
sea water is months very different

→ compare what you
get w/ sanger vs 454
at same cost

The redal sweeps
are here in far color

If try to let
the picture
The talking
an 11
stay out
of the way
Our model
system de jon
winkles and
worms and
spaghetti
and then in
Jeffrey Dahmer

Structure
of biofilm

That
And this
is my
cheesy
analogy
here.

MCA
regulators
the nature

Coupling Function to Phylogeny via Single-Cell Phenotyping

M. Lidstrom*, L. Chistoserdova, S. Levine, S. McQuaide, M. Kalyuzhnaya, University of Washington, Seattle, United States of America

One of the major challenges in understanding microbial community function in natural environments is linking genome-level sequences to function and role in the ecosystem. This area is often referred to as "ecogenomics", and requires not only the extraction of functional information from metagenomic databases, but also the understanding of functional metabolic units, often called metabolic modules, and their role in natural systems. It is the intersection of these two approaches that provides the possibility to couple genomics and function in natural microbial communities. We are using an approach that couples a function-based live cell presorting step to single-cell analysis at both the physiological and genetic levels. This approach allows culture-independent enrichment of live cells involved in specific functions, analysis of a variety of phenotypic capabilities at the single cell level, then targeting of those cells that test positive for specific functions for further culture-dependent and sequencing analysis. Our test case involves C1 cycling in Lake Washington, a site for which we have carried out a variety of culture-dependent and culture-independent analyses of the communities involved in C1 cycling. More recently, we have generated a metagenomic database enriched by SIP for strains involved in C1 cycling. We have developed a prototype single cell analysis system for carrying out the upstream screening, technology created within the Single Cell Observatory at the University of Washington. This system is effective in identifying cells capable of respiring C1 compounds.

Use of Synthetic Protein Scaffolds to Balance Pathway Flux of Engineered Metabolic Pathways.

John E. Dueber*, Gabriel C. Wu, G. Reza Malmirchegini, Adeeti V. Ullal, Jay D. Keasling. University of California, Berkeley.

Synthetic biology strives to introduce control over biological systems. Engineering metabolic pathways in *E. coli* holds particular promise in harnessing the impressive abilities of living cells to catalyze difficult chemical reactions for the purpose of synthesizing chemicals in an inexpensive and ecologically friendly manner. A hallmark of natural metabolism is tight flux regulation (via mechanisms such as negative and positive feedback). However, engineered metabolic pathways are usually devoid of these flux-regulating mechanisms and yields can suffer as a result. We improved the product yield of an engineered metabolic pathway that synthesizes mevalonate by introducing modular scaffolds that co-localize pathway enzymes. Flux through this pathway was balanced by optimizing the relative ratios of the pathway enzymes co-localized to the scaffold. This work shows that modular control over pathway flux can be introduced post-translationally in a systematic manner that should prove to be both generalizable to other metabolic pathways and scalable for increased numbers of enzymes.

Hydrogen Fuel Production by Microalgae: Issues and Future Directions

Maria L. Ghirardi*¹, Michael Seibert¹, Matthew Posewitz², Paul King¹, Arthur Grossman³, Matthew Wecker⁴, Alexandra Dubini¹, Florence Mus³, Christine English¹, Sergey Kosourov¹, Eric Johnson⁵. ¹National Renewable Energy Laboratory (NREL), Golden CO; ²Colorado School of Mines, Golden, CO; ³Stanford University, Menlo Park, CA; ⁴Genebiologics, Boulder, CO; ⁵Johns Hopkins University, Baltimore, MD.

Some microalgae are able to photoproduce H₂ gas from water under anaerobic conditions, thus representing a renewable source of fuel. Two [FeFe]-hydrogenase genes have been reported in the green alga, *Chlamydomonas reinhardtii*. Their expression and activity are regulated by O₂, and they may both be involved in fermentative pathways as well. I will discuss recent gene expression microarray studies done in collaboration with Stanford University and the Colorado School of Mines and will present research directions at NREL aimed at identifying and solving major metabolic and rate-limiting steps in H₂ photoproduction, including the enzyme's high sensitivity to O₂. Recent work will be presented on (a) structural and mechanistic aspects of [FeFe]-hydrogenases, (b) transcriptional regulation of algal [FeFe]-hydrogenases, and (c) issues related to application of green algae for alternative fuel production at high light conversion efficiencies.

Redirection of metabolism for hydrogen production

Yasuhiro Oda¹, Frank W. Larimer², Patrick S. G. Chain³, Federico Rey¹, Loren Hauser², Stephan Braatsch⁴, J. Thomas Beatty⁴, Dale A. Pelletier², Amy L. Schaefer¹ and Caroline S. Harwood^{1*}, University of Washington, Seattle¹; Oak Ridge National Laboratory, Oak Ridge²; Lawrence Livermore National Laboratory, Livermore³; University of British Columbia⁴

Rhodopseudomonas and other purple nonsulfur photosynthetic bacteria have received attention as potential biocatalysts for nitrogenase-mediated biohydrogen production because they can derive ATP to drive this thermodynamically unfavorable reaction from the abundant resource of solar energy. The first *Rhodopseudomonas* genome sequence, that of strain CGA009¹, revealed a surprising richness of metabolic versatility that we exploited to generate mutants in which metabolism is redirected so that cells produce large amounts of hydrogen constitutively under all growth conditions². CGA009 encodes three nitrogenase isozymes³, and this will present additional opportunities for improving hydrogen production once we better understand how the expression and activities of each of these enzymes is regulated. Recently we have obtained and analyzed the complete genome sequences of four additional *Rhodopseudomonas* strains. The sequenced strains form a distinct taxonomic unit, but there is also considerable strain-to-strain variation. There are 3,319 core genes, 70% of the genes in each genome, shared by four or more strains. Between 10 and 18 percent of the genes in each genome are strain-specific. The diversity of strain-specific attributes of the *Rhodopseudomonas* strains provides new opportunities to analyze the integrated functioning of the hundreds of proteins involved in photophosphorylation, carbon metabolism, reductant transfer and nitrogenase synthesis that are needed for hydrogen production. For example, studies of how one of the strains, BisA53, synthesizes and controls the activities of its record number of seven light harvesting complexes in response to light intensity and quality should expand our current limited understanding of how genetic engineering might be applied to maximize the efficiency of light absorption for photosynthesis. The expanded abilities of another strain (BisB5) to degrade aromatic compounds of the type derived from lost-cost agricultural feed stocks provides opportunities to study and improve the ability of cells to simultaneously utilize diverse electron-donating substrates for hydrogen production. Also, our discovery from its genome sequence that a third strain (BisB18) can produce substantial amounts of hydrogen by pyruvate fermentation in the dark as well as via nitrogenase in light, has obvious implications for developing a hydrogen production process that operates continuously over a 24 h day.

¹Larimer, F. W., P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M. Land, D. A. Pelletier, J. T. Beatty, A. S. Lang, F. R. Tabita, J. L. Gibson, T. E. Hanson, C. Bobst, J. Torres y Torres, C. Peres, F. Harrison, J. Gibson and C. S. Harwood. 2004. The genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nature Biotechnology* 22:55-61.

²Rey, F. E., E. K. Heiniger and C. S. Harwood. 2007 Redirection of metabolism for biological hydrogen production. *Appl. Environ. Microbiol.* 73:1665-1671.

³Oda Y, Samanta SK, Rey FE, Wu L, Liu X, Yan T, Zhou J, Harwood CS. 2005. Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*. *J Bacteriol.*;187:7784-94.

Reconstruction of the transcriptional regulatory network in *E. coli*

Byung-Kwan Cho and Bernhard O. Palsson

Department of Bioengineering, University of California, San Diego

Over the past decade or so, dramatic developments in our ability to experimentally determine the contents and functions of genomes have taken place. In particular, high-throughput technologies are now allowing the genome-scale reconstruction of transcriptional regulatory networks. ChIP-chip and mRNA expression data have become two primary high-throughput methods most useful for determining the regulatory networks of entire genomes. In the work presented here, high-density tiling arrays have been used to perform location analysis for the DNA-binding proteins in *E. coli* and for high resolution expression profiling. The proteins examined include RNA polymerase, sigma factors, and broad acting transcription factors. In addition, we have further developed a method for the genome wide determination of transcriptional start site (TSS) using Solexa sequencing. The TSS data was then integrated with RNA polymerase ChIP-chip and tiled array expression data in order to fully elucidate the *E. coli* transcriptome.

Cool introduction

RNA P. binding

- static map - w/ rifampicin
- dynamic map - across genes

Lots of cool stuff but
no mention of prior work

Took all data



Network

Significant binding w/o expression

High-throughput quantitative analysis of genetic and chemical-genetic interactions in *E. coli*

Athanasios Typas*, Robert J. Nichols, Nevan J. Krogan and Carol A. Gross. University of California, San Francisco.

Large-scale genetic interaction studies provide the basis for defining gene function and pathway architecture. Recent advances in the ability to generate double mutants on a genome-wide basis in the budding yeast have dramatically accelerated the acquisition of such genetic interaction information and the biological inferences that follow. Here, we describe a method based on F-driven conjugation, which allows for high-throughput generation of double mutants in *E. coli*. This method, termed Genetic Interaction ANalysis Technology for *E. coli* (GIANT-coli), permits us to systematically generate and array double mutant cells on solid media in high-density arrays. We show that colony size provides a robust and quantitative output of cellular fitness and that GIANT-coli can recapitulate known synthetic interactions and identify numerous new negative (e.g. synthetic lethality) and positive (e.g. suppressive) relationships. Furthermore, chemical genetic profiling, a sister approach to GIANT-coli that involves screening its single-gene knockout library in the presence of different drugs or stresses, complements our genetic interaction data and provides further insights on gene function and organization.

Took 2 comprehensive single-gen KO libs
Hfr conjugator
can get SOS

1000000
SOS
100 papers
out next
year!!

no nondescript
genetic

sensors
w/ internet
access via
a cable

Günther

If I have time I will tell you
about a dream

(-every 8th breath thank a doctor)
(-while we still have poles)

-Silicon walks up species spectra
-silica precipitation in different species
is non-homologous
-connection between iron + Si

were the
non-10d genes
mixed in
other
species
what about
down
regulation

Integrating Discovery-based Undergraduate Research Experiences into UCLA Courses Using a Collaborative Curriculum Model.

Erin R. Sanders-Lorenz*, Ann M. Hirsch, and Jeffrey H. Miller. University of California, Los Angeles.

Undergraduate science education should mirror the collaborative nature of research and reflect how scientific hypotheses are evaluated and results are communicated in the 21st century. Such an undertaking requires the development of teaching methods that actively engage students in the creative process of scientific inquiry, provide skills necessary for success in the modern research laboratory, as well as foster excitement about the discovery process central to scientific research. To this end, UCLA's Department of Microbiology, Immunology, and Molecular Genetics launched a project-based laboratory course entitled 'I, Microbiologist: A Discovery-based Undergraduate Research Course in Microbial Ecology and Molecular Evolution' in which students explore microbial diversity within environmental samples. Students generate 16S rDNA data sets, which they use to build phylogenetic trees, characterizing the composition of bacterial communities inhabiting soil samples collected from sites in the Los Angeles area. While some students investigate oil-degrading thermophiles that reside far below the Earth's surface, others target the rhizosphere of plants hypothesized to respond to the activities of nitrogen-fixing bacteria. Cultivation of microbes on media lacking a nitrogen source produces isolates that are screened for the presence of nitrogenase genes or proteins. Students in a plant biology course offered by the Molecular, Cell and Developmental Biology department investigate the role microbes play in plant development wherein they cultivate plants inoculated with putative nitrogen-fixers isolated by 'I, Microbiologist' students and assess the effect on plant growth. Here, students learn not only about biological nitrogen fixation, plant-microbe interactions, and root biology, but also become active participants in microbial energy research, exploring ways by which these organisms could be commercially exploited as biofertilizers. Taken together, this collaborative research experience represents a model for undergraduates in the life sciences to learn the process of scientific inquiry, and for educators to engender student interest and enthusiasm about microbiological research.

The JGI Microbial Genome Annotation Program

Cheryl Kerfeld,* DOE Joint Genome Institute and University of California, Berkeley/

The JGI has developed a set of tools and faculty training modules to integrate microbial genome annotation across the undergraduate life sciences curriculum, from introductory molecular biology to capstone microbiology and biochemistry courses. The annotation platform includes IMG-EDU, a new version of IMG with special features such as six frame translation viewer. IMG-EDU and links to other databases are embedded in a web/wiki portal that guides students through the annotation process and records their findings. The annotation platform also provides tools for instructors to assign genes and pathways and to evaluate students' annotations. This will enable colleges and universities nationwide to "adopt a genome" for manual curation and student research. Students at 12 colleges and universities nationwide are currently part of the pilot phase of the program.

Insights from the Genomes of Commonly Used Lab Strains.

T. Durfee^{1,2*}, S. Baldwin², R. Nelson², D. Frisch^{1,3}, B. Mau¹, G. Plunkett III¹, V. Burland¹, N. Perna¹ & F.R. Blattner^{1,2,3}, ¹University of Wisconsin-Madison, ²DNAStar, Madison, WI, ³Scarab Genomics, Madison, WI

Much of basic biology has been gleaned from the study of *E. coli* using strains closely related to wild type isolates as well as those that have undergone extensive genetic manipulation. The whole genome sequences of MG1655 and W3110 provide the basis for understanding much of *E. coli* biology, but undocumented variation in other strains and the impact those changes have on physiology remains largely unknown. Next-gen sequencing technologies and improved assembly software are making the determination of bacterial genome sequences far more routine. Using Next-gen and Sanger data sets in combination with a new assembly engine, SeqMan N-Gen, we have assembled ten new *E. coli* genomes including those from both K-12 and B lineages. The genomes represent a survey of both strains closely related to natural isolates and those derived from multiple steps of classic genetics. Together, these genomes provide a resource for understanding and, in some cases, re-evaluating the genetic basis for various phenotypes.

Reinventing the Petri Dish: Miniaturized Tools for High Throughput Microbial Culture

C.J. Ingham^{*}, W.M. deVos & J.E.T. van Hylckama Vlieg. Laboratory of Microbiology, Wageningen, Top Institute Food and Nutrition and MicroDish B.V.

A microbial growth format has been created that allows up to a million individual samples to be cultured. The base of this culture chip is a thin strip of porous aluminium oxide (PAO) which allows microbial cells to grow on the upper surface. PAO is an excellent material for culturing microorganisms; it is flat, inert and uniquely porous. The walls of the chip which segregate samples are created by laminating the PAO with a acrylic film. The acrylic film is selectively removed by reactive ion etching. The result is to create culture areas as small as 7 x 7 μm on the base, with walls 10 μm high. The PAO chips can be scanned by microscopy. Individual strains or mutants can be isolated by a micromanipulator. The PAO chips are versatile and have been used for viable counting and in various examples of high-throughput screening. The design is flexible and the chips permit rapid changes in the environment of cultures grown on them and do not desiccate. Additional tools for microbial culture are in development, including a "velvet pad" designed to replicate patterns of microcolonies grown on PAO.

Ingham, C.J., A. Sprenkels, J. Bomer, D. Molenaar, A. van den Berg, J.E.T. van Hylckama Vlieg, and W.M. de Vos. 2007. *Proc. Natl. Acad. Sci. USA* 46:18217-18222.

The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes

Maureen E. Hillenmeyer*, Eula Fung, Jan Wildenhain, Sarah E. Pierce, Shawn Hoon, William Lee, Michael Proctor, Robert P. St.Onge, Mike Tyers, Daphne Koller, Russ B. Altman, Ronald W. Davis, Corey Nislow, Guri Giaever. Stanford University.

Genetics aims to understand the relation between genotype and phenotype. However, because complete deletion of most yeast genes (~80%) has no obvious phenotypic consequence in rich medium, it is difficult to study their functions. To uncover phenotypes for this nonessential fraction of the genome, we performed 1144 chemical genomic assays on the yeast whole-genome heterozygous and homozygous deletion collections and quantified the growth fitness of each deletion strain in the presence of chemical or environmental stress conditions. We found that 97% of gene deletions exhibited a measurable growth phenotype, suggesting that nearly all genes are essential for optimal growth in at least one condition. When two genes display a high level of co-fitness (i.e., they show share a similar fitness profile across all conditions), we found that they tend to share Gene Ontology function. Pairs of compounds that co-inhibit (i.e., induce similar genome-wide fitness signatures) tend to be structurally similar and belong to the same therapeutic class. Using machine learning methods, fitness profiles for yeast deletion strains can be used to predict interactions between gene products and compounds.

The genome sequence of the deep-branching amoeboflagellate *Naegleria gruberi* reveals ancestral eukaryotic functions

Simon Prochnik (1), Lillian Fritz-Laylin (2), Chandler Fulton (3), Zacheus Cande (2), Alan Kuo (1), Igor Grigoriev (1), Scott Dawson (4), Daniel Rokhsar (1,5).

1 DOE-Joint Genome Institute, Walnut Creek, CA; 2 MCB, UC Berkeley, Berkeley, CA; 3 Dept. Biology, Brandeis University, Waltham, MA; 4 Microbiology, UC Davis, Davis, CA; 5 CIG, UC Berkeley, Berkeley, CA

Naegleria gruberi is a free-living, heterotrophic, unicellular eukaryote that, like human cells, can use both flagellar and amoeboid forms of motility. Here we present an analysis of the ~45 million nucleotide *N. gruberi* genome and its ~15,000 protein-coding genes. The genome encodes a rich repertoire of cytoskeletal, signaling, and metabolic modules. With minimal assumptions about the poorly-resolvable deep eukaryotic phylogeny, the presence of these genes in the *Naegleria* genome implies an ancestral eukaryote with similar, high levels of complexity. In particular, this ancestor appears to have been capable of both amoeboid and flagellar motility; and anaerobic metabolism. Many of these modules were independently lost in diverse eukaryotic lineages, including the relatively closely related parasitic trypanosomes.

Regulating Bacterial Expression of Short Hydrophobic Toxic Proteins with Small RNAs.

Elizabeth Fozo, Mitsuoki Kawano, Fanette Fontaine, Sofia Mendieta, Kenneth Rudd and Gisela Storz

More and more regulatory small RNAs (sRNAs) are being discovered in both prokaryotes and eukaryotes. In *Escherichia coli* MG1655, computational as well as experimental evidence suggest that at least 80 sRNAs are encoded by the genome.

The Sib sRNAs are unusual group of RNAs found in *E. coli* and closely related bacteria. Most sRNAs genes are unique sequences; however, the Sib RNAs are encoded by five highly homologous genes in *E. coli* MG1655. The number of Sib genes can vary in different species and even in different strains of *E. coli*. We confirmed that all five Sib RNAs are expressed in *E. coli* MG1655, and no phenotype was observed for a five *sib* deletion strain. However, the plasmid-encoded copy of a specific Sib RNA had to be expressed at high levels in a strain carrying a deletion of the same *sib* gene or cell growth ceased. Further examination of the Sib sequences revealed the presence of conserved open reading frames encoded opposite each *sib* gene. These antisense genes encode highly hydrophobic 18-19 amino acid proteins (Ibs). Overproduction of the Ibs proteins was highly toxic to *E. coli*, but this toxicity can be overridden by induction of the antisense Sib RNA. Two other RNAs encoded divergently from each other were similarly found to encode a small, toxic, hydrophobic protein (ShoB) and an antisense RNA regulator (OhsC). The properties of the Ibs and ShoB proteins are very similar to those of the Ldr and TisB proteins of *E. coli* in that they are small, toxic, hydrophobic proteins whose expression is regulated by sRNAs. Whole genome expression analysis showed that overexpression of IbsC, ShoB, LdrD and TisB has both overlapping and unique consequences for the cell. Ongoing studies to further investigate the regulation, the causes of toxicity upon overexpression, phenotypes associated with endogenous levels of the genes, as well as the genome-wide distribution of these toxin-antitoxin modules should begin to provide insight into cellular roles these small, potentially toxic, hydrophobic proteins.

RNA-mediated Epigenetic Programming of a Genome-Rearrangement Pathway.

M. Nowacki*, V. Vijayan, Y. Zhou, T.G. Doak, K. Shetty, K. Schotanus, L.F. Landweber. Princeton University

Genome-wide DNA rearrangements occur in many eukaryotes but are most exaggerated in ciliates, making them ideal model-systems for epigenetic phenomena. During development of the somatic macronucleus, *Oxytricha trifallax* destroys 95% of its germ-line, severely fragmenting its chromosomes, and then unscrambles hundreds of thousands of remaining fragments by permutation or inversion. Our recent findings, (Nowacki et al. 2008 Nature 451:153-8) demonstrate that DNA or RNA templates can orchestrate these genome rearrangements in *Oxytricha*, supporting an epigenetic model for sequence-dependent comparison between germ-line and somatic genomes. A complete RNA cache of the maternal somatic genome may be available at a specific stage during development to provide a template for correct and precise DNA rearrangement. We show the existence of maternal RNA templates that could guide DNA assembly, and that disruption of specific RNA molecules disables rearrangement of the corresponding gene. Injection of artificial templates reprograms the DNA rearrangement pathway, suggesting that RNA molecules guide genome rearrangement. Furthermore, the occasional transfer of point mutations in these RNA templates to the rearranged molecules provides a mechanism for stable inheritance of acquired, spontaneous somatic mutations (in either DNA sequence or alternative splicing pattern), without altering the germline genome. This mechanism for inheritance beyond the conventional DNA genome can epigenetically transfer information across multiple generations, hinting at the power of RNA molecules to shape genome information.

Development of the www.EcoliHub.org Information Resource

Barry L. Wanner¹, Walid Aref¹, Kirill A. Datsenko¹, Sara Ess¹, Michael R. Gribskov¹, Daisuke Kihara¹, Sangtae Kim¹, Hirotada Mori², Ali Roumani¹, Dawn R. Whitaker¹

¹Purdue University, West Lafayette, IN, ²Nara Institute of Science and Technology, Ikoma, JAPAN.

We envision a future where a dynamic community of online resources (both information and computational resources) act cooperatively, with each resource focusing on its strength and expertise, and linking to the strength and expertise of other resources. This vision of collaboration and sharing is part of a suite of strategies that are collectively now referred to as Web 2.0. EcoliHub will not replace existing information resources; the goal is to add value to these resources by: 1) improving the ability to share information and computational services among resources, 2) allowing resources to be combined (piped together) in new ways, without requiring additional development effort by the provider, 3) improving the community's ability to find information and resources, 4) providing new information and resources that 'fill in the gaps' between existing resources and improve the quality of information provided by all participating *E. coli* resources. Current status of the EcoliHub project will be described. The EcoliHub project is funded by NIH GMS U24 GM077905.

Systematic Analysis of Genetic Interaction of *Escherichia coli*.

Hirotsada Mori^{1,2}, Rikiya Takeuchi¹, Fusamitsu Yanagihara¹, Natsuko Yamamoto¹, Kenji Nakahigashi², Kirill A. Datsenko³, Athanasios Typas⁴, Andrew Emilli⁵, Jack Greenblatt⁵, Carol A. Gross⁴, Barry L. Wanner³

¹Nara Institute of Science and Technology, Nara 630-0101, Japan, ²Institute of Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0017 Japan, ³Purdue University, West Lafayette, IN 47907, USA., ⁴University of California at San Francisco, San Francisco, CA 94158-2517, USA., ⁵University of Toronto, Toronto, Ontario M5G 1L6, Canada

Robustness is an important fundamental property of biological systems. Common mechanisms that give rise to robustness depend on alternative or bypass pathways of metabolism and other redundant biological processes. Also, elucidation of epistatic relationships among genes can give insight into the understanding of physiological networks. Synthetic lethality or sickness by double gene knockout mutations is one of the most powerful methods for analyzing robustness. To make this possible in *Escherichia coli*, which is one of the best organisms to understand cellular systems comprehensively based on the vast amount of accumulation of biological knowledge, we setup an easy and reliable system for construction double knockout strains by conjugation and for analysis of their growth effects.

The comprehensive single gene knockout library (Keio collection) was already established. To make double knockout deletion mutants, we started to develop the second single gene deletion library and efficient conjugation system. The second single gene deletion library, named ASKA deletion collection, carrying a different (chloramphenicol) resistance cassette with additional features, including (1) turbo GFP fusion with the initiation codon of the target gene, (2) a modified FLP recombination site (FRT1) site, which does not recombine with FRT, and (3) a 20-nt molecular bar code. For conjugation system, a fragment carrying *tra* genes and *oriT* of wild-type F plasmid were combined with *oriRg* replicon. This plasmid can replicate only in cells producing *pir* gene product. About 300bp chromosomal fragment of the target site of integration is cloned into this plasmid. The resultant plasmid is integrated into the chromosome, which does not code *pir* gene, and change host cell to Hfr. Using these systems, systematic analysis of genetic interaction of *E. coli* is now underway.

POSTER ABSTRACTS

Genetic Dissection of an Exogenously-Induced Biofilm in Laboratory and Clinical Isolates of *E. coli*.

S. Amini* & S. Tavazoie. Princeton University, Princeton, NJ 08540.

Biofilms are associated with the majority of microbial infections. However, developing effective strategies for controlling them requires a better understanding of the underlying biology beyond what currently exists. Using a novel approach, we have induced formation of a robust biofilm in *Escherichia coli* by utilizing an exogenous source of poly-*N*-acetylglucosamine (PNAG) polymer, a major virulence factor of many pathogens. Through a comparative microarray profiling of competitive selections carried out in both transposon insertion and over-expression libraries, we have revealed the genetic basis of PNAG-based biofilm formation. Our observations suggest that electrostatic interactions between PNAG and surface structures of *E. coli* are dominant determinants of biofilm formation. However, various regulatory systems and surface structures can modulate the phenotype in both lab strain and clinical isolates of *E. coli*. The fact that the majority of clinical isolates which produced PNAG also responded to the exogenously produced version of the polymer suggests that PNAG-induced biofilm formation is a biologically and possibly clinically important phenomenon.

***Burkholderia pseudomallei* Type III secretion effectors**

C.T. French, M. Liu*, D. Arambula*, J.F. Miller. University of California Los Angeles

B. pseudomallei employs three type III secretion systems (T3SSs) to deliver protein effectors into host cells, thereby promoting intracellular invasion, survival, and cell to cell spread. We devised bioinformatic screens to allow the identification of *B. pseudomallei* T3SS effector genes by the presence of conserved eukaryotic functional domains, similarity to other T3SS effectors, and genomic co-localization with class IA chaperones. To date, we have identified 24 putative T3SS effector loci in the genome of *B. pseudomallei*. Expression phenotypes for four *Burkholderia* effector protein (Bep) candidates - BepB, BepD, BepE, and BepF, have been examined by confocal microscopy following transfection into mammalian cells. BepB, BepE and BepF result in altered cellular morphology, the reorganization of cellular actin and changes in substrate adhesion, respectively. BepD displays remarkable colocalization with cytoplasmic microtubules which was eliminated by treatment with nocodazole, and BepD also localized to the spindle apparatus of mitotic cells. The N-terminus of BepD was found to mediate T3SS-dependent secretion of a β -lactamase reporter protein into HeLa cells using enteropathogenic *E. coli* and *Bordetella bronchiseptica* as surrogate T3SS secretion-containing bacterial hosts. Experiments are in progress to determine the roles of these factors in the intracellular life cycle of *B. pseudomallei*. As part of the work to screen and identify new T3SS effectors, we have also developed a k-mer-based method for rapid phylotyping of related bacterial strains and identification of novel sequences using unassembled sequencing reads.

Antibiotic-Resistance Genes in the Metagenome of Urban Wetland Sediments.

K.F. Archer, D.J. Arriola*, J.B. Laroya, C.R. Ryan, N.G. Singh and K.R. Unruh, & D.E. Cummings (PI). Point Loma Nazarene University, Dept. of Biology, San Diego, CA.

Runoff from urban watersheds contains many biological pollutants including antibiotic-resistant bacteria and the genes coding for resistance (AR genes) whose fates are largely unknown. The purpose of this study is to determine the extent to which AR genes are introduced and accumulate in southern California's coastal wetlands. DNA was extracted from surface sediments of coastal wetlands in southern California and background sites and examined for AR genes using the polymerase chain reaction (PCR). Specific AR genes examined were *tetA/C* and *qnrA* encoding resistance to tetracyclines and (fluoro)quinolones, respectively. Sediments from the sewage-contaminated Tijuana River Estuary (TRE) tested positive for both *tetA/C* and *qnrA* 24 hours after a winter storm; however, AR genes were virtually undetectable after 10 dry weeks. Sediments collected from the TRE every 2-3 days during a three-week winter storm series suggested that the abundance of AR genes is dependent on river flow and precipitation. qPCR protocols are currently being developed to quantify AR genes in these samples. Cloned AR genes from the TRE during the storm series shared 99-100% sequence similarity with plasmid-borne *qnrA* or *tetA/C* genes from clinical isolates. Notably, AR genes were undetectable by PCR at all sites, regardless of sewage control, during the dry season. However, dry-season sediments enriched in the presence of tetracycline were positive for both *tetA/C* and *qnrA*, suggesting that AR genes may be linked on the same genetic elements. Winter sediments from the TRE enriched in the presence of five separate antibiotics, including tetracycline, were positive for both *qnrA* and *tetA/C* in 9 out of 10 samples, further suggesting linkage of multiple AR genes. Taken as a whole, these data suggest that AR genes, possibly including multidrug-resistance plasmids, are being introduced to coastal wetlands through urban runoff, potentially creating a long-term reservoir of AR genes in the natural environment.

Presence of prophages in *Streptococcus equi* and *Streptococcus zooepidemicus*

S. Artiushin* & J. F. Timoney. University of Kentucky, Lexington.

S. equi and *S. zooepidemicus* of Lancefield group C share DNA homology greater than 90%. However, the antigenically homogeneous and clonal *S. equi* is an obligate parasite of *Equidae* for which it is a contagious and virulent pathogen, whereas *S. zooepidemicus* is antigenically heterogeneous and found as mucosal commensal in a variety of host animals causing occasional opportunistic invasions. Recently the presence of phage associated pyrogenic mitogens SePE-H and SePE-I has been demonstrated for *S. equi*. In addition, subtractive hybridization was instrumental in detection of 3 additional prophages in the *S. equi* genome. The aims of this work were to study the distribution of these prophages in isolates of *S. equi* and *S. zooepidemicus* and investigate their potential contributions to virulence. All prophage associated sequences except Se-ph1 were detected in *S. equi* but not in *S. zooepidemicus*. Se-ph1 sequences were found only in some *S. equi*. Pyrogenic exotoxins SeeL, SeeM and SePE-I and SePE-H are associated with prophages Se-ph3 and Se-ph4 respectively. Se-ph2 encodes Phospholipase A, another potential virulence factor. Gene of streptodornase has been also recognized in the genome of Se-p9, a temperate phage recently sequenced in our laboratory. Interestingly, almost identical proteins are encoded by phages of human pathogen *S. pyogenes*.

In conclusion, evolution of *S. equi* from an ancestral *S. zooepidemicus* may have been driven by acquisition of prophages. Considering the ubiquity of *S. equi* in horse populations and the fact that strangles has been known at least since the Roman era together with evidence of similarity between phages of *S. equi* and *S. pyogenes*, formation of *S. equi* as a virulent host adapted pathogen probably took place during domestication of the horse.

Ecological insights into marine microbial metagenomics data

Barberan A.*, A. Fernandez-Guerra, R.S. Etienne, J.L. Green, E.O. Casamayor, and B.J.M. Bohannan. Centre d'Estudis Avançats de Blanes, Spanish Council of Scientific Research, SPAIN

Community genomics are revealing an unprecedented level of microbial taxonomic diversity and metabolic novelty in the world's oceans. However, the rates of information collection and information interpretation are entirely uncoupled. A synergy between ecological theory and metagenomics may help bridge this mismatch, by providing theoretical and analytical tools that can reveal microbial community patterns and the processes that underlie them. Here, we applied 3 different theoretical approaches to marine microbial environmental genomics data. First, we examined community taxonomic and functional abundance distributions along a depth gradient, using Hubbell's neutral theory of biodiversity to make inferences about the processes of community assembly. Second, we explored microbial trait-based biogeography (e.g., the distribution of G+C content, rRNA copy number, functional diversity, etc) along a surface oceanic transect. Finally, we designed a bioinformatics pipeline to compare community phylogenetic features reconstructed from different genes.

Study of *Escherichia coli* Single Gene Knockout Strains Susceptibility to Genotoxic Agents

E. Becket*, C. Tamae, F. Chen, and J. H. Miller. Department of Microbiology, Immunology, and Molecular Genetics. University of California, Los Angeles.

While chemotherapeutic regimens are widely used in the treatment of cancer, negative side effects, secondary mutagenesis, and incidences of drug resistance are often seen. One way to counteract these adverse effects is to discover improved chemotherapeutic regimens that would potentiate cancer-treatment drugs. Potential directions of study would be to determine the therapeutic mechanisms of these drugs and their respective cellular targets. We developed a high-throughput screening method using a 4,000 strain single gene knockout KEIO collection of *Escherichia coli* to identify mutants which are hypersensitive to a range of genotoxic agents: Bleomycin (BLEO), Cisplatin (CPT), ICR-191 (ICR), 5-azacytidine (5AZ), Zebularine (ZEB), and 5-Bromouracil (5BU). From this screen, 171 strongly susceptible mutants were discovered, from which a sensitivity profile with these six agents was created. This profile enabled the knockout collection of 4,000 strains to be reduced to a set of 171 strains that allow for rapid screening against a wide range of agents at varying concentrations. From these data, the mechanisms of action for each drug can be discovered and, ultimately, protein inhibitors will be designed or discovered so as to devise combinatorial regimens that allow for lower doses of chemotherapeutic agents.

Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes.

John F. Heidelberg¹, William C. Nelson¹, Thomas Schoenfeld², and Devaki Bhaya^{*3}

¹ Dept. of Biological Sciences, Wrigley Institute for Environmental Studies, University of Southern California, Wrigley Marine Science Center, Avalon, CA 90704, ² Lucigen Corporation, 2120 W. Greenview Drive, Middleton, WI 53562, ³ Carnegie Institution for Science, Department of Plant Biology, 260 Panama Street, Stanford, CA 94305

CRISPR arrays and associated *cas* genes are widespread in bacteria and archaea and confer acquired resistance to viruses. To examine viral immunity in the context of naturally evolving microbial populations we analyzed genomic data from two thermophilic *Synechococcus* isolates (*Syn* OS-A and *Syn* OS-B') as well as a prokaryotic metagenome and viral metagenome derived from microbial mats in hot springs at Yellowstone National Park. Two distinct CRISPR types, distinguished by the repeat sequence, are found in both the *Syn* OS-A and *Syn* OS-B' genomes. The genome of *Syn* OS-A contains a third CRISPR locus with a distinct repeat sequence, which is not found in *Syn* OS-B', but appears to be shared with other microorganisms that inhabit the mat. The CRISPR repeats identified in the microbial metagenome are highly conserved, while the spacer sequences (hereafter referred to as "viriotopes" to emphasize their critical role in viral immunity) were mostly unique and had no high identity matches when searched against GenBank. Searching the viriotopes against the viral metagenome, however, yielded several matches with high similarity some of which were within a gene identified as a likely viral lysozyme/lysin protein. Analysis of viral metagenome sequences corresponding to this lysozyme/lysin protein revealed several mutations all of which translate into silent or conservative mutations which are unlikely to affect protein function, but may help the virus evade the host CRISPR resistance mechanism. These results demonstrate the varied challenges presented by a natural virus population, and support the notion that the CRISPR/viriope system must be able to adapt quickly to provide host immunity. The ability of metagenomics to track population-level variation in viriope sequences allows for a culture-independent method for evaluating the fast co-evolution of host and viral genomes and its consequence on the structuring of complex microbial communities.

Alex to RNAi

- found a nearly 100%
one in Roseflexus

it does live check by jowl
with the cyanobacteria

western
"changes"
when she
meant
↓ fitness

11 spacers
is a vague
insulting

Experimental Mapping of Ligands with Binding Proteins in a group of ABC transporters from *Rhodopseudomonas palustris*

¹Sarah E. Giuliani, ¹Ashley M. Frank ²Loren Hauser & ^{1*}Frank R. Collart

¹Biosciences Division, Argonne National Laboratory, Lemont, IL, and ²Oak Ridge National Laboratory, Oak Ridge, TN

We evaluated a fluorescent thermal shift (FTS) assay as an approach to improve gene/protein functional assignments. Targets were selected from the bacterial ABC transporter family, focusing on transporter components specifically annotated as periplasmic binding proteins. ABC transporters are widely distributed in bacterial organisms and transport a diverse spectrum of ligands. Target proteins (48) were selected from *Rhodopseudomonas palustris*, a metabolically versatile bacterium commonly found in soils and water. The targets primarily represented a single COG (COG0683) and were generally annotated as proteins binding branched-chain amino acids, urea, nitrate or aliphatic sulfonates. Targets were cloned into vectors for heterologous expression and the most soluble proteins were scaled for purification with 67% of the total set purified. FTS assay results from this group of binding proteins indicate a capability for specific functional assignment of ~50% of the screened targets. The assays reveal that these proteins bind a diverse set of ligands representative of the functional and metabolic diversity of this organism. Specific binding was detected for several functional chemical categories including aromatic compounds, polyamines, dicarboxylate metabolites, and amino acids. We are currently expanding our ligand library to extend the functional assignments for this specific COG category. The specific experimental assignments illustrate the limitations of current sequence based approaches for functional assignment. We suggest that the FTS assay can improve sequence based approaches by generation of functional assignments for a set of homologs that can be used to define sequence motifs and improve annotation assignments based on sequence alignments. The results also demonstrate the ability of the FTS assay to provide biologically significant and specific annotation for this important class of proteins and suggest that this approach can be generally applied to other families and ligand binding proteins from *R. palustris* and other genomes.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

A Functional Metagenomic Analysis of Antibiotic Resistance Genes from Soil in a Wisconsin Apple Orchard

Justin Donato*, Luke Moe, Brandon Converse, Flora Berklein, Keith Smart, Patricia McManus, and Jo Handelsman. University of Wisconsin - Madison

Human health depends on a delicate balance between a person and the microbes that reside within them. Despite their role in maintaining human health, many human diseases are caused by bacteria that are natural residents of the human body. In the fight against bacterial pathogens, one of the most important weapons available to physicians is a broad range of chemotherapeutic agents. However, a significant limitation to the successful use of these drugs is the development of antibiotic resistance. This generates a serious need to identify novel antibiotics to overcome resistance as well as a need to develop new strategies to manage resistance.

Surprisingly, given the risk of selecting for resistant bacteria, antibiotics are widely used in modern agricultural practices. The clinically relevant antibiotic, streptomycin, is used to control bacterial disease on apple trees. To study the effects of streptomycin treatment on antibiotic resistance genes in the soil, we constructed metagenomic libraries directly from the soil samples without amplification or enrichment of the bacteria present. The resulting libraries contained >400,000 clones averaging approximately 30 kb each, or more than 12 Gb of environmental DNA. The libraries were introduced into *E. coli* and selected for resistance to ten antibiotics representing three different classes of antibiotic compounds. We identified genes conferring resistance to all three classes of antibiotics tested. For the clones resistant to β -lactam antibiotics, we found genes homologous to known β -lactamases in addition to genes that were not previously associated with resistance. Further study will characterize the impact of streptomycin on the diversity of these resistance genes.

The YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA: bioinformatic identification and experimental validation.

Basma El Yacoubi^{*1}, Fabio Agnelli², James. R. Williamson², Paul Schimmel², Manal A. Swairjo² and Valerie de Crécy-Lagard¹. ¹University of Florida, Gainesville Florida and ²The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA

Using a combination of bioinformatic, genetic, structural, and biochemical approaches, we showed that the universal protein family YrdC/Sua5 is involved in the biosynthesis of threonylcarbamoyladenosine (t⁶A). t⁶A is a universal modification of the anticodon loop of tRNAs coding for ANN codons that imparts a unique structure to the tRNA anticodon loop enhancing its binding to the ribosome *in vitro*. Contradictory reports on the essentiality of both the *yrdC* gene of *Escherichia coli* and the *sua5* gene of *Saccharomyces cerevisiae* lead us to definitely prove (1), that *yrdC* is essential in *E. coli* and (2), that *sua5* is dispensable in yeast. However deleting the *sua5* gene leads to severe growth phenotypes. We show that YrdC binds ATP and preferentially binds tRNA^{Thr} only lacking the t⁶A modification. This work lays the foundation for elucidating the function of a protein family found in every sequence genome to date and, for the first time, understanding the role of t⁶A *in vivo*.

Module Level Analysis of Genome-wide Datasets

H. Goodarzi*, O. Elemento & S. Tavazoie. Princeton University, Princeton, NJ 08540.

Whole-genome approaches have revolutionized the study of biological systems across many disciplines. These methods facilitate analyses through parallel data collection for thousands of genes simultaneously. The real challenge, however, is to extract biologically meaningful insights from such datasets. Mapping gene-level data to pathway and process level modules is the first step in relating large-scale observations to previously existing knowledge. We propose a novel computational framework, based on mutual-information, to reveal pathway modules that show non-random patterns across the entire distribution of genome-wide observations (e.g. gene expression microarrays). This framework is embodied in a new algorithm, called PAGE (Pathway Analysis of Gene Expression). The notion of mutual-information enables PAGE to analyze both continuous (e.g. transcript levels) and discrete (e.g. co-expression cluster indices) datasets with no *a priori* statistical assumptions. Another advantage of this framework is that under-representation (i.e. when a genetic module is less frequent than predicted by chance) is treated on par with over-representation of pathway modules across the data, leading to substantial increases of both sensitivity and specificity. We have compiled a predefined set of modules in *E. coli* based on available annotation (genes sharing a given GO term), regulons (genes downstream of a given transcription factor) and stress responses (genes participating in the initiation and maintenance of different stress responses). Application of our approach to *E. coli* gene expression datasets demonstrates the power and versatility of our framework. We anticipate PAGE to facilitate pathway level exploration of large-datasets across all microbial species of basic, medical, and industrial importance.

Diversity Generating Retroelements

Huatao Guo*, Longping V. Tse, Roman Barbalat, Mari Gingery, Minghsun Liu, Sameer Sivaamnuaiphorn, Min Xu, Sergei Doulatov, and Jeff F. Miller. University of California, Los Angeles, CA.

Host-parasite interactions are often driven by mechanisms that promote genetic variability. In the course of our studies on bacterial pathogenesis, we discovered a group of temperate bacteriophages that generate diversity in a gene that specifies tropism for receptor molecules on host *Bordetella* species, which cause respiratory infections in human and mammals. This microevolutionary adaptation is produced by a novel “diversity-generating retroelement” (DGR) that combines the basic retroelement life cycle of transcription, reverse transcription and integration with site-directed, adenine-specific mutagenesis. Central to this process is a reverse transcriptase-mediated transposition process from a donor template repeat (TR) to a recipient variable repeat (VR). Recent work has focused on the genetic basis of diversity-generation. The directionality of information transfer is determined by the initiation of mutagenic homing (IMH) sequence present at the 3’ end of VR. We have demonstrated that DGR function occurs through a TR-containing RNA intermediate by a unique target-primed reverse transcription mechanism that precisely regenerates target sequences. This non-proliferative, “copy and replace” mechanism enables repeated rounds of protein diversification and optimization of ligand-receptor interactions. The potential utility of DGRs is illustrated by the identification of about 100 related elements in bacterial, phage, and plasmid genomes. DGRs are present in human pathogens (*Treponema*, *Legionella* spp.), human commensals (*Bacteroides*, *Bifidobacterium* spp.), green sulfur bacteria (*Chlorobium*, *Prosthecochloris* spp.), cyanobacteria (*Trichodesmium*, *Nostoc* spp.), magnetotactic bacteria (*Magnetospirillum* spp.), and many other diverse species. DGRs comprise a new family of retroelements with the potential to confer powerful selective advantages to their host genomes. In addition to shedding light on DGR function, our results suggest novel approaches for DGR-based genetic engineering.

A Simple Method for Preparing DNA Sequencing Templates using SeqPrep96 DNA Binding Microplate.

Y. Yang, H.R. Hebron & J. Hang*. EdgeBio, Gaithersburg, MD.

DNA binding matrix was immobilized on the surface of a 96-well microplate for easy, rapid and cost-effective sample preparation for DNA sequencing. DNA was captured and purified with the SeqPrep plates, without intense use of centrifugation, vacuum, or magnet. Two procedures were developed based on enzymatic and alkaline lysis of bacterial cells respectively. For enzymatic cell lysis format, one SeqPrep plate was used for bacterial growth, DNA purification and DNA storage, which eliminate plate transfer steps. Additionally, both cell lysis and DNA capture were accomplished in one step using a single buffer. For the procedure adapted to alkaline lysis method, cells were grown and lysed by a modified alkaline lysis process in a deep well culture block. Without conventional clarification, crude lysate is directly used for DNA isolation in SeqPrep plate. After DNA capture in SeqPrep plate and the simple washing steps, DNA was eluted in 40 μ l. Purified DNA samples were successfully used in DNA sequencing with high consistence and reproducibility. When using 10 μ l sequencing reaction with 3 μ l sample and 0.25 μ l BigDye v3.1 Terminator (1/16th dilution, i.e. 1/32nd reduction), the sequencing results gave average success rate of 90-95% and read length of 700 bases or above. The method is fully automatable and convenient for manual operation as well. It enables easy and quick high throughput production of DNA with purity and yields both sufficient for high quality DNA sequencing at a significantly reduced cost.

Characterization of *Salmonella enterica* Serotypes Using a Novel High Density 85 Genome Microarray and Antimicrobial Susceptibility Testing

S. A. Jackson*, H. C. Harbottle*, M. K. Mammel, I. R. Patel, T. J. Barnaba, S. L. Ayers, S. Zhao, T. A. Cebula, J. E. LeClerc
Food and Drug Administration, Laurel, MD.

Salmonellae are recognized as the leading foodborne bacterial pathogen in the United States, causing over 1.4 million cases of illness annually. The ability of this pathogen to rapidly adapt to novel environmental niches has recently been made apparent by their contamination of foods not traditionally associated with enteric illness, such as fresh produce. More problematic is the fact that multidrug resistance (MDR) is being increasingly identified among numerous *Salmonella* serotypes, potentially compromising antimicrobial therapy. To investigate their genomic diversity and antimicrobial resistance, an Affymetrix® DNA microarray containing over 2.5 million features was developed. It represents 177,000 genes from 35 whole genomes and 58 plasmids/integrations of *Salmonella*; 178,000 genes from 28 whole genomes and 46 plasmids of *E. coli*; and 10 whole genomes and 16 plasmids of *Shigella*; 13,000 intragenic regions; and 2400 antibiotic resistance genes and 2200 virulence genes from various bacterial sources. Sequence analysis identified over 85,000 unique genes, each of which is represented on the array using 11 independent 25-mer probe pairs. Additionally, a tiling strategy was utilized to identify genomic changes at the single nucleotide level in approximately 2400 core genes highly conserved in *Salmonella*. Using this array, over 100 temporally and geographically diverse *Salmonella* isolates having varying serotypes and antimicrobial resistance profiles were analyzed to investigate genomic content and the presence of antimicrobial resistance elements. All strains were phenotypically tested for susceptibility to 17 antimicrobials. The results demonstrate the power of a comprehensive DNA microarray representing pangenomic elements for identification and characterization of foodborne pathogens and their associated resistomes. The array serves as a diagnostic tool for identification of strains important to public health, provides a means to delimit the core genes of a pathogenic group, identifies genes that specify the particular 'personality' of strains, and provides a detailed correlation between antimicrobial resistance, gene content, and expressed phenotype.

Keywords: Microarray ; *Salmonella* ; Antimicrobial resistance

* These two authors contributed equally to this work.

Presenting Author:

S.A.Jackson and H. C. Harbottle
U.S. Food and Drug Administration
8301 Muirkirk Rd
Laurel, MD 20708
Phone 301-210-7807
FAX: 301-210-7807
scott.jackson1@fda.hhs.gov

Reticulate Evolution of the SSU-rRNA gene in prokaryotes

C.W. Herbold*, J.A. Servin, R.G. Skophammer, J.A. Lake

Evolutionary relationships among prokaryotic taxa are generally examined using bifurcating phylogenetic trees. Although there is a bulk of genomic data that appears to violate a tree-like model of evolution in the prokaryotes, it has been generally asserted that a “vertical-inheritance” tree exists, onto which horizontal/lateral transfer events may be mapped. Often times in these analyses, it is assumed that genes are a fundamental unit of inheritance and are transferred as a single unit from donor to recipient lineages. Using a gene that has been used extensively for determining “vertical-inheritance” trees, the SSU-rRNA gene, we challenge this assumption by showing that the SSU-rRNA gene itself undergoes reticulate evolution, with sections within a single gene sequence having been inherited from different lineages. Rare examples of this phenomenon have been previously reported within the literature, however our analysis indicates that transfer of rRNA gene segments may play a more central role in the evolution of this gene. These findings suggest that phylogenetic trees are probably insufficient to describe the evolution of the SSU-rRNA gene and provides an explanation for the lack of resolution power between eubacterial phyla. Despite these issues, a wealth of phylogenetic information exists within the SSU-rRNA gene that can be used to build robust evolutionary graphs/networks rather than trees. Here we provide examples of this approach using a set of archaeal eocytes (Crenarchaeota). This approach promises to provide a more complete map of prokaryotic evolution and to help resolve questions regarding the deep relationships among taxonomic groups.

Analysis of the bovine rumen microbiota, its cellulolytic genes and proteins associated with biofuel crops.

M. Hess*^{1,2}, T. Zhang^{1,2}, S. Tringe^{1,2}, R. Mackie³ and E. Rubin^{1,2}

¹Joint Genome Institute, Walnut Creek, CA, ²Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, ³University of Illinois, Urbana-Champaign, Ill

Current bioethanol production from lignocellulosic biomass is a relatively inefficient process requiring energy-intensive chemical and physical pretreatments. The identification of robust cellulolytic enzymes (e.g. cellulases and hemicellulases) is a necessity to achieve the ambitious goal to replace 30% of the national petroleum based gasoline with bioethanol by 2030.

Switchgrass (*Panicum virgatum*), a perennial, warm-season grass is native to most states of the U.S. and has drawn a lot of attention as a promising biofuel crop. Foregut fermenters such as cattle and sheep possess an area within the gastrointestinal tract that is well separated from the acid-secreting portion of the stomach. This fermentation chamber harbors a microbial community that is able to degrade the plant celluloses, hemicelluloses, pectins, fructosans, starches and other polysaccharides to monomeric and dimeric sugars. The bovine rumen exceeds a volume of 100 L and represents an enormous, easily accessible and manipulable system to study the microbial community and its biocatalysts adapted to the degradation of selected biofuel crops such as *P. virgatum*.

Metagenomics and Metatranscriptomics of the microbial community associated with biofuel crops will allow us to identify the microbes responsible for lignocellulose degradation and the genes that are overexpressed during fiber hydrolysis. The combination of Metagenomics, Metatranscriptomics, and functional screening assays will help to identify the genes and proteins required for the efficient lignocellulose breakdown within the bovine rumen.

In summary, the findings of our project will provide significant insight into the process of lignocellulose degradation and facilitate the construction of genetically modified organisms (i.e. *Escherichia coli*, *Saccharomyces cerevisiae*, and *Sulfolobus solfataricus*) and ultimately the development of an efficient process to convert lignocellulosic biomass into bioethanol at industrial scale.

The role of a novel *Pseudomonas aeruginosa* stress response in innate aminoglycoside resistance.

Aaron Hinz*, Samuel Lee, and Colin Manoil. University of Washington, Seattle.

Pseudomonas aeruginosa is an opportunistic pathogen which chronically infects the airways of cystic fibrosis patients. The aminoglycoside antibiotic tobramycin is used to treat *P. aeruginosa* infections, but high intrinsic resistance can limit the effectiveness of the drug. In the course of investigating this resistance, we identified a two-component regulator (encoded by *amgRS*; aminoglycoside resistance), whose inactivation results in hypersensitivity to tobramycin and other aminoglycosides.

The role of AmgRS in tobramycin resistance was investigated by using transcriptional arrays to identify genes regulated by AmgRS. The expression of three membrane-associated proteases and six hypothetical membrane proteins was induced by tobramycin in an *amgRS*-dependent manner, suggesting a role for AmgRS in proteolysis of membrane proteins. In addition, the aminoglycoside sensitivity of *amgRS* mutants was accentuated by both overexpression of membrane proteins and inactivation of AmgRS-independent proteases. We propose that tobramycin-induced mistranslation activates the AmgRS response by increasing the number of misfolded proteins in the cell membrane, and that proteases upregulated by AmgRS function to preserve membrane integrity.

Production of hydrogen gas from light and an inorganic electron donor: use of thiosulfate by *Rhodopseudomonas palustris*

Jean J. Huang*, Erin Heiniger, and Caroline S. Harwood
Department of Microbiology, University of Washington, Seattle

Thiosulfate occurs naturally as an intermediate of microbial sulfur metabolism, is formed abiotically in thermal springs, and is also generated as a product of industrial processes including the refinement of natural gas and kraft paper production. We determined that the purple bacterium *Rhodopseudomonas palustris* can grow and fix nitrogen photolithoautotrophically using thiosulfate as an electron donor. These observations are interesting for several reasons. Electrons transfer from an inorganic compound to nitrogenase is fundamentally interesting. In addition, *R. palustris* produces hydrogen as an obligate aspect of nitrogenase activity and the use of an inorganic donor expands the range of potential electron donating compounds for hydrogen production beyond biomass-based electron donors. We characterized *R. palustris* culture growth rates and yields on thiosulfate under nitrogen-fixing conditions, determined the *in vivo* nitrogenase activity with this electron donor using gas chromatography and acetylene reduction assays, and determined hydrogen production rates and yields under non-growing culture conditions in comparison to a variety of organic donors. *R. palustris* is able to efficiently use electrons from thiosulfate in the presence of bicarbonate to produce hydrogen gas under non-growing conditions. Additionally, our work with transposon mutants enabled the identification of a gene cluster, *fixABCX* (RPA4062-4605) that is necessary for photolithoautotrophic growth using thiosulfate under nitrogen-fixing conditions, but not under fixed-nitrogen replete conditions. These genes are likely involved in transport of electrons from thiosulfate to nitrogenase.

Functional Indications for Unknown Proteins in the GreenCut

Steven Karpowicz^{1*}, Arthur Grossman², Matteo Pellegrini¹, and Sabeeha Merchant¹.

¹University of California, Los Angeles and ²Carnegie Institution.

Many chloroplast proteins remain unstudied and have unknown function. The GreenCut is a group of 349 nucleus-encoded proteins conserved in the land plants *Arabidopsis thaliana* and *Physcomitrella patens* and the algae *Chlamydomonas reinhardtii* and *Ostreococcus spp.* but not present in non-photosynthetic organisms. Only 144 of these proteins have known (K) or inferred (KI) function, while 205 have predicted (UP) or wholly unknown (U) function. To better understand the purpose of the unknown proteins and to identify fruitful targets for functional investigation by experiments, a bioinformatic analysis of the GreenCut proteins was undertaken. Predicted unknowns were classified into defined functional groups using MapMan and Pfam domain descriptors, revealing 13 unknown transcription factors and 11 substrate transporters of unknown specificity. Results from Illumina whole transcriptome analysis of both *Chlamydomonas* cells and *Arabidopsis* roots indicate known genes are more highly expressed than unknowns in photosynthetic *Chlamydomonas* cells compared to non-photosynthetic roots. *Arabidopsis* tissue-development microarray data suggest particular unknowns are involved in tissue-specific metabolic activities, such as 5 root-expressed substrate translocation proteins. A search for GreenCut homologs in 33 fully-sequenced cyanobacterial species revealed 61 GreenCut genes are conserved in at least 30 of 33 cyanobacteria. Our computational analysis of the GreenCut has already advised us on choice of genes for experimental investigation, and we expect continuing work will reveal other interesting targets for functional studies.

NCBI Microbial Genome Validation Tool.

W. Klimke*, A. Badretdinov, L. Riley, I. Mizrahi, and T. Tatusova. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland.

Increasing genomic data has led to a vast increase in understanding prokaryotic biology. Although the data is increasing constantly, the quality of the data needs to be maintained and curated. The National Center for Biotechnology Information, in collaboration with a number of organizations, is working to establish a set of annotation standards for prokaryotic genomes. Quantitative validation of the annotation is an important first step and NCBI has created a prokaryotic genome validity check tool that is available online and as a standalone command line package to submitters of genomic records to both GenBank and the Prokaryotic Genome Automatic Annotation Pipeline. The validation tool does a number of checks including: overlaps for coding regions and structural RNAs, missing structural RNAs or RNAs on the wrong strand, truncated proteins, and potential frame-shifts. This validation tool has also been used to improve the annotation on a number of Reference Sequence (RefSeq) records, thereby improving the annotation quality for anyone utilizing the RefSeq resources in their scientific research.

IslandPick and IslandViewer: Tools for the computational identification and visualization of genomic islands.

M.G.I. Langille* & F.S.L. Brinkman

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada

Genomic islands (GIs; clusters of genes of probable horizontal origin) play an important role in medically important adaptations of bacteria. Therefore, the identification of GIs has become a particular focus when examining a genome for its notable new properties and several computational tools have been developed to predict islands in sequenced genomes. The majority of these tools utilize the naturally occurring genome sequence biases that exist between bacterial strains to identify regions that appear to have a foreign sequence composition. In contrast to searching for anomalous regions using sequence signatures, GIs can also be predicted using multiple genomes and identifying regions that have a limited phylogenetic distribution suggesting that they most likely have horizontal origins. We have constructed stringent datasets of GIs and non-GIs (phylogenetically conserved regions) using a comparative genomics approach, called IslandPick, and we have used these to test the accuracy of several sequence composition based GI prediction tools. In addition, we present IslandViewer, the first web accessible interface that facilitates viewing and downloading of multiple pre-computed GI datasets predicted by the three currently most accurate methods.

Antibiotic Sensitivity Profiles Determined with an *Escherichia coli* Gene Knockout Collection

A. Liu*, L. Tran, K. Lee, L. Chinn, E. Park, & J. H. Miller. Department of Microbiology, Immunology, and Molecular Genetics. University of California, Los Angeles.

Antibiotic resistant bacteria are on the rise and pose a major health problem with more people dying of antimicrobial resistance than of AIDS in the US since 2005. One approach to combat this issue is to improve the efficacy of existing antibiotics with combination therapy. Potential targets for such drugs would be bacterial proteins that provide intrinsic antibiotic resistance. We have extended previous studies and tested the entire KEIO collection of close to 4,000 single gene knockouts in the *Escherichia coli* knockout collection to identify genes whose loss increases sensitivity to one of twelve different antibiotics (ciprofloxacin, CPR; rifampicin, RIF, vancomycin, VAN, ampicillin, AMP; sulfamethoxazole, SFX, gentamicin, GEN; metronidazole, MET; streptomycin, STR; fusidic Acid, FUS; tetracycline, TET; chloramphenicol, CAM; and erythromycin, ERY). We used high through-put screening of one or more subinhibitory concentrations of each antibiotic, and reduced more than 75,000 data points to a set of 161 strains that display significantly increased sensitivity to at least one of the antibiotics. These strains allow one to rapidly define a “sensitivity profile” fingerprint for each additional antibiotic to be tested and characterize these antibiotics with respect to their mechanisms of action. We used this redacted set of strains to define such a profile for nine additional antibiotics (spectinomycin, SPC; cephradine, CEP; aztreonam, AZT; colistin, COL; nitrofurantoin, NIT; neomycine, NEO; enoxacin, ENO; tobramycin, TOB; and cefoxitin, CEF). Comparison of the profiles for the 21 antibiotics reveals several patterns. Although some of the gene knockout mutants are relatively specific to the type of antibiotic tested, others are hypersensitive to more than one antibiotic. Collectively, these data provide targets for the design of “co-drugs” that can potentiate existing antibiotics and increase bacterial killing efficiencies at levels less toxic to humans.

The *Acidiphilium cryptum* genome reveals capacities for metal transformation and mineral colonization

David Cummings and David Kerk. Point Loma Nazarene University.

David Sims and Paul Richardson. DOE-Joint Genome Institute.

Brandon Briggs, Michael Swenson, and Timothy S. Magnuson*. Idaho State University.

Some metal-contaminated environments are under acid conditions, nonetheless, acidophilic iron respiring bacteria have been detected in such environments, and a potential means for microbial remediation of metals under acid conditions is feasible. The genome of the strict acidophile *Acidiphilium cryptum* JF-5 was sequenced by JGI, and a variety of interesting and significant findings were made. The main chromosome of *A. cryptum* was about 3.9 Mb in size, containing 4005 ORFs. Surprisingly, *A. cryptum* also possessed an 'extragenome' of 8 plasmids ranging in size from 1 kb to 200 kb. Plasmid-encoded mobility and replication related sequences were discovered, as well as partial non-expressed sequences corresponding to mercury resistance and cation transport. On the main chromosome, genes were found encoding redox proteins thought to be important in Fe respiration. These genes included 12 genes encoding both membrane-associated and periplasmic cytochromes c, several of which can reduce Cr(VI) in vitro. Significant differences in cytochrome c gene numbers and predicted protein structures from *Shewanella* and *Geobacter* were noted, suggesting different mechanisms for respiratory Fe-reduction. Genes for metal transformation were discovered, including chromate reductase and arsenate reductase. Chromate reductase was shown to be expressed only under Cr(VI)-grown conditions. Biofilm-related genes, encoding capsular exopolysaccharide synthesis and pili/flagella assembly, were present, and biofilm formation was observed. *A. cryptum* is thus substantially different genetically and biochemically than its neutrophilic counterparts, but is well-suited for both aerobic and anaerobic lifestyles in environments where toxic metals are prevalent.

ClpX Protease Contributes to Antimicrobial Peptide Resistance and Virulence Phenotypes of *Bacillus anthracis*.

S. M. McGillivray^{1*}, C. M. Ebrahimi¹, N. Fisher², M. Sabet¹, D.X. Zhang¹, Y. Chen³, N. M. Haste¹, R. Aroian¹, R. L. Gallo¹, D. G. Guiney¹, A.M. Friedlander², T.M. Koehler³ and V. Nizet¹. University of California, San Diego, La Jolla CA¹; United States Army Medical Research Institute of Infectious Diseases, Frederick MD²; University of Texas-Houston Health Science Center³, Houston TX

Bacillus anthracis, the causative agent of anthrax, resists many host defenses during infection. Although anthrax toxin and capsule, located on two large plasmids, play important roles in the pathogenesis of this disease, evidence indicates that chromosomal genes also contribute. We developed a novel transposon system using the *HimarI* transposase gene to generate a random chromosomal library of *B. anthracis* Sterne. Mutations affecting candidate toxins were identified in screens for hemolytic and proteolytic activities. A transposon mapping to the ClpX protease yielded a nonproteolytic and nonhemolytic phenotype and targeted allelic replacement mutagenesis confirmed a role for ClpX. *B. anthracis* lacking ClpX showed increased sensitivity to cathelicidin and antimicrobial peptides and increased susceptibility to whole blood killing. The *B. anthracis* ClpX mutant showed decreased virulence in skin infection and inhalational models of infection. Conversely, cathelicidin-deficient mice showed increased susceptibility to *B. anthracis* infection. Deletion of the ClpX gene in the fully virulent *B. anthracis* Ames strain also significantly attenuated virulence. We conclude that ClpX is essential for *B. anthracis* virulence, and functions at least in part by thwarting host cathelicidin mediated defenses. Inhibiting ClpX may provide a novel target for *B. anthracis* therapy.

Microbial Genomics with Pyrosequencing and Celera Assembler

Arthur L. Delcher¹, Sergey Koren², Jason R. Miller^{2*}, Eli Venter², Brian P. Walenz², Granger Sutton².

¹University of Maryland, USA. ²The J. Craig Venter Institute, Maryland, USA.

DNA sequencing platforms differ in cost, accuracy, read length, typical coverage, and the variety of available paired-end protocols. By all of these measures, the pyrosequencing platform is intermediate to Sanger platforms and short-read platforms. Compared to short-read technologies, pyrosequencing can resolve longer genomic repeats by virtue of its intermediate-length reads and mate distances. Even longer repeats could be resolved, at higher cost, by combinations of pyrosequencing data with long-read, large-insert data provided by Sanger sequencing. The Celera Assembler has been modified to support hybrid sequencing strategies that combine pyrosequencing with long-range mates. The new pipeline is named CABOG, for Celera Assembler with the Best Overlap Graph. It supports the latest mate pair protocol for the 454 FLX pyrosequencing platform. The software was tested on various combinations of three data types: Sanger mate pairs; FLX unpaired reads; FLX paired-end reads. In tests on four microbial genomes, CABOG generated longer contigs than any other software tested. CABOG was better able to exploit the mate constraints provided by paired-end reads from either platform. It was able to resolve more repeats and to build larger contigs and scaffolds. CABOG's low rate of contig mis-assembly was reduced in the presence of sufficient mate pair data. Assemblies of different ratios of paired and unpaired reads indicated a successful strategy: FLX reads complemented by 1X genome coverage in Sanger or FLX long-range mates. The new software is available as open-source from <http://wgs-assembler.sf.net> under the GNU Public License.

The JGI Contribution to Whole Microbial Genome Sequencing.

A. Christine Munk^{1*}, David Sims¹, Lynne Goodwin¹, Cliff Han¹, David Bruce¹, Tom Brettin¹, Alla Lapidus², Susan Lucas², Patrick Chain², Alex Copeland², and Chris Detter¹.
¹JGI-Los Alamos National Laboratory, ²JGI-Production Genomics Facility

The DOE Joint Genome Institute (JGI) was created in 1997 to unite the expertise and resources in genome mapping, DNA sequencing, technology development, and information sciences pioneered at the DOE genome centers at Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), and Los Alamos National Laboratory (LANL).

Since that time, there has been an exponential increase in the number of whole microbial genome sequences that have been made publicly available through the National Center for Biotechnology Information (NCBI) as a result of the work of the JGI. There has been a similar increase in the proportion of the whole genomes that have been submitted by the JGI as compared to all other centers. To date, approximately 250 whole microbial genomes are available at www.ncbi.nlm.nih.gov/genomes/lproks.cgi?view=1 that were completed by the JGI. This accounts for over 1/3 of all of the genomes, and 2.5 times the input of the next finishing center.

The JGI continues to strive to be the world leader in its contribution to this resource of basic microbiological knowledge, and to pave the way to next generation genome sequencing techniques and technologies.

LA-UR#081494

Modeling promoters of alternative sigma factors in *E. coli*

Vivek Mutalik^{*1}, Carol Gross^{1,2}, and Virgil Rhodius¹

(1) Department of Microbiology and Immunology, and (2) Department of Cell and Tissue Biology, University of California at San Francisco, San Francisco, CA 94158

Identification of potential promoters from genomic sequences is a key element in uncovering transcriptional regulatory circuits and understanding how genes are differentially regulated within these networks. We are developing models that correctly predict the strength of promoters transcribed by the alternative sigma factors sigma-E and sigma-32. We chose to examine this set because these promoters have slightly higher information content than those transcribed by the housekeeping sigma. We used a library of 60 “short” natural promoters (sequences -35 to +20) that contained the core -10 and -35 motifs, and 60 “long” natural promoters (sequences -65 to +20) that also includes the upstream UP element, representing all of the promoters known to be transcribed by sigma-E. A similar library of 50 short and long promoters representing all known sigma-32 promoters was also used. We measured their *in vivo* promoter strengths using promoter::gfp fusion reporter assays. Our experimental results reveal the contribution of core and UP elements towards promoter strength. Differences in hierarchy of promoter strengths at different growth phases indicate a role of additional regulatory factors influencing the promoter activity. We then modeled these results using a series of iterations, including position-specific weight matrices without or with a contribution for the activity of the promoter, and partial least square regression analysis. Thus far, our analysis has concentrated on the sigma-E short promoters. We find that more sophisticated models can better capture promoter strength. We are currently expanding these models to the other data sets we have and will then use them to see if they also allow better promoter prediction in *E. coli* and related genomes.

Chemical Genetic Profiling of the *E. coli* Deletion Library

Robert J. Nichols*¹, Athanasios Typas¹, Angela Wong¹, Sueyoung Lee¹, Nevan J. Krogan² and Carol A. Gross¹

¹Department of Microbiology and Immunology, University of California at San Francisco, USA

²Department of Cellular and Molecular Pharmacology & The California Institute for Quantitative Biomedical Research, University of California at San Francisco, USA

The quantitative assessment of growth in response to a variety of conditions is known as chemical genetic profiling¹. Profiling entire single-gene knockout libraries, such as the *E. coli* Keio collection², has the potential to yield valuable insights into the functional organization of cellular responses to stress. Briefly, single gene deletion strains of *E. coli* that respond similarly to a given panel of conditions suggest a functional link between those genes. Likewise, conditions that affect a common subset of the deletion library may have a common mode of action. Here we present a preliminary analysis of chemical genetic profiles of the Keio collection in response to approximately 30 conditions.

1. Parsons, A. B. et al. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* 126, 611-25 (2006).
2. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2, 2006 0008 (2006).

Metagenomic Biodiversity and Biogeography: Theory and Practice

James P. O'Dwyer*, Elizabeth B. Perry* and Jessica L. Green, University of Oregon.

“The advent of the metagenomic era presents two interrelated challenges to ecologists. The first is the need to characterize biodiversity and biogeography using metagenomic data, which requires the development of new ecological metrics. In turn, the development of these metrics must also inform and be informed by the development of new ecological theory. The synthesis of these two aims will provide us with a way both to analyze and to understand the ecological implications of metagenomic datasets, and in this project we focus on the Sorcerer II Global Ocean Sampling expedition. Our preliminary analyses use the metagenomic similarity metric introduced by Rusch et al (*PLoS Biology* 2007), and we have found that genomic similarity shows significant distance-decay patterns along several important environmental gradients. The overall metagenomic similarity between samples is significantly correlated with geographic distance, temperature, chlorophyll *a* concentration, and ocean depth, while other environmental ‘metadata’ such as sample depth and salinity do not correlate with metagenomic similarity. Although these results are indicative of important ecological processes, we have subsequently identified certain weaknesses in this similarity metric. The second part of our project focuses on developing new metrics to describe both within-sample metagenomic diversity and between-sample metagenomic similarity, and how these metrics may be compared with predictions from ecological theory.”

Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*

B.K. Cho, C.L. Barrett, E.M. Knight, Y.S. Park^{*} & B. Ø. Palsson. University of California, San Diego

Broad acting transcription factors (TFs) in bacteria form regulons. Here we present a four-step method to fully reconstruct the leucine-responsive protein (Lrp) regulon in *Escherichia coli* K-12 MG 1655 that regulates nitrogen metabolism. Step one is comprised of obtaining high-resolution ChIP-chip data for Lrp, the RNA polymerase and expression profiles under multiple environmental conditions. We identified 138 unique and reproducible Lrp-binding regions and classified their binding state under different conditions. In the second step the analysis of these data revealed six distinct regulatory modes for individual open reading frames (ORFs). In the third step, we used the functional assignment of the regulated ORFs to reconstruct four types of regulatory network motifs around the metabolites that are affected by the corresponding gene products. In the fourth step, we determine how leucine, as a signaling molecule, shifts the regulatory motifs for particular metabolites. The physiological structure that emerges shows the regulatory motifs for different amino acid fall into the traditional classification of amino acid families thus elucidating the structure and physiological functions of the Lrp-regulon. The same procedure can be applied to other broad acting TFs opening the way to full bottom-up reconstruction of the transcriptional regulatory network in bacterial cells.

Identification of Genes Involved in *Variovorax paradoxus* EPS Surface Phenotypes

Michael Pehl*¹, W. David Jamieson², Karen Kong², Paul Orwin¹.

¹Department of Biology, California State University San Bernardino

²Department of Biology and Biochemistry, University of Bath

Variovorax paradoxus is a ubiquitous soil microorganism that plays an important role within the soil microbial community. *Variovorax paradoxus* strain EPS was isolated from the nature preserve area behind CSU San Bernardino, and displays a mucoid phenotype on laboratory media. Approximately 25000 mutants from a previously generated transposon library were screened based on colony morphology and then tested for altered biofilm physiology. Out of approximately 400 mutants initially identified in the biofilm screen, about 120 were determined to have a stable alteration in biofilm growth. These insertions were examined directly by rescue cloning and sequencing from the Tn5 insertion. A number of genes associated with exopolysaccharide production, including at least two unique glycosyl transferases, two NAD-epimerases, and an O-acetyltransferase were identified. One of the glycosyl transferases identified was interrupted in numerous unique mutations that were all captured in this screen. Gene products involved in the regulation and transport of exopolysaccharides were also found in this group. Another gene that was disrupted at several unique sites in separately identified mutants was *pi/Y1*, a putative surface adhesion involved in abiotic surface attachment. A putative sensor histidine kinase, potential c-di-GMP homeostasis modifiers including a diguanylate cyclase and phosphodiesterase, along with numerous other unique insertions were identified in this manner. A high percentage of our mutants with altered colony morphology and altered biofilm formation also displayed different swarming phenotypes. Our results suggest that *V. paradoxus* colony morphology, biofilm physiology, and swarming motility are coordinated surface activities, and that the switch between sessile and motile surface behaviors is an important facet of *V. paradoxus* life in the soil environment.

Aerobic Fermentation of *D*-Glucose by an Evolved Cytochrome Oxidase Deficient *Escherichia coli* Strain

Vasiliy A. Portnoy*, Markus J. Herrgård, Bernhard Ø. Palsson. University of California, San Diego.

Fermentation of glucose to *D*-lactic acid by an evolved *E. coli* mutant deficient in three terminal oxidases under aerobic growth condition is reported in this work. Cytochrome oxidases (*cydAB*, *cyoABCD*, *cbdAB*) were removed from the *E. coli* K12 MG1655 genome resulting in the ECOM3 (*E. coli* Cytochrome Oxidase Mutant) strain. Removal of cytochrome oxidases reduced the oxygen uptake rate of the knock-out strain by nearly 85%. Moreover, the knock-out strain was initially incapable of growing on M9 minimal media. After subjecting the ECOM3 strain to adaptive evolution on glucose M9 medium for 60 days, the growth rate equivalent to anaerobic wild type *E. coli* was achieved. Our findings demonstrate that three independently adaptively evolved ECOM3 populations acquired different phenotypes: one produced lactate as a sole fermentation product while the other two strains exhibited a mixed acid fermentation under oxic growth conditions with lactate remaining as the major product. The homofermenting strain showed the *D*-lactate yield of 0.8 g/g from glucose. Gene expression and *in silico* model-based analysis was employed to identify perturbed pathways and explain phenotypic behavior. Significant upregulation of *ygiN* (quinol monooxygenase) and *sodAB* (superoxide dismutase) explain the remaining oxygen uptake that was observed in evolved ECOM3 strains. *E. coli* strains produced in this study showed the ability to produce lactate as a fermentation product from glucose as well as undergo mixed-acid fermentation under oxic conditions.

Role of RpoN in the Transcription Regulation Network in *Geobacter sulfurreducens*

Yu Qiu^{*1}, Karsten Zengler¹, Byung-Kwan Cho¹, Larisa Kagan¹, James Elkins¹, Derek R. Lovley², and Bernhard Ø Palsson¹, ¹*Dept. of Bioengineering, University of California San Diego, CA;* ²*University of Massachusetts, Amherst, MA*

Geobacter sulfurreducens is capable of transferring electrons to a variety of electron acceptors, including Fe(III), U(IV), and the surface of electrodes, making it the candidate of choice for bioremediation of contaminated environments and harvesting electricity from waste organic matter. In depth understanding of how *G. sulfurreducens* functions will have great impact on optimizing bioremediation and energy harvesting applications. A major step toward this goal is the establishment of a genome-scale transcriptional regulatory network (TRN) which allows in depth understanding of fundamental aspects of chromosome structure, DNA replication and repair, response to stress and regulation of metabolism.

Towards a comprehensive understanding of the TRN in *G. sulfurreducens*, genome-wide binding patterns of RNA polymerase and sigma factors were determined by chromatin immunoprecipitation (ChIP) coupled with high-density tiling arrays (ChIP-chip). Patterns of RNA polymerase (RNAP), σ^{70} and σ^{54} binding were compared from cells grown with acetate under fumarate and Fe(III)-reducing conditions. All ChIP-chip results were complemented by gene expression profiles using high-density tiling arrays. We found 953 and 945 binding sites of σ^{70} under fumarate and Fe(III)-reducing conditions, respectively. Almost all the genes that were expressed under fumarate conditions but not in Fe(III)-reducing conditions have a σ^{54} binding sites, suggesting a major role of RpoN in the differential gene expression between these two conditions.

Statistical Properties of the Codon Adaptation Index

A.C. Retchless* & J.G. Lawrence. University of Pittsburgh, Pittsburgh.

The Codon Adaptation Index is a widely used measure of translational selection, which enables predictions of gene expression levels and the rate of gene evolution. Here, we provide an analysis of the expected distributions of CAI values in the absence of translational selection. This analysis reveals biases in the CAI that could confound its interpretation. Consideration of the expected distribution allows us to standardize observed CAI values, which increases the usefulness of the CAI as a predictive measurement while clarifying the relationship of CAI values across genomes. Finally, these insights enable us to propose a new algorithm to identify relative adaptiveness of synonymous codons based solely on the protein-coding sequences present in a genome.

Seasonal Dynamics of Antarctic Bacterioplankton Revealed by Metagenomics.

C.S. Riesenfeld*, J.J. Grzymski, & A.E. Murray. Earth and Ecosystem Sciences, Desert Research Institute, Reno, N.V.

Background: The Antarctic Peninsula marine waters (~64°S latitude) experience strong temporal variability in physical parameters which drive strong gradients in biological productivity in the upper ocean. Little is known about the bacterial and archaeal metabolic and genomic differences between the period of maximal productivity in the summer and the Antarctic winter. **Methods:** Two large-insert metagenomic libraries were created from marine bacterioplankton (<2.5 mm fraction) from February (summer) and August (winter) samples collected in Antarctic Peninsula waters. Bi-directional end-sequencing (~37,000 reads) generated 7 and 9.6 Mb of coding DNA for the summer and winter libraries, respectively. **Results:** A sizeable fraction of the data represents genetic diversity not currently represented in large public databases, while other sequences resemble portions of completely sequenced marine archaeal and bacterial genomes (e.g., *Nitrosopumilus maritimus*, *Polaribacter irgensii*, *Roseobacter denitrificans* and *Pelagibacter ubique*). Comparison of the two metagenomic libraries revealed differences in overall nucleotide composition, phylogenetic distribution of BLAST hits and predicted functions (e.g., frequency of protein families). One conspicuous difference between the libraries was the GC% distribution of end-sequence reads. In the case of the summer sample, the distribution was skewed towards GC-rich sequences, which can be explained (in part) by the fact that *Roseobacter*-affiliated sequences were abundant in the summer sample. Sequence clusters were identified that affiliated exclusively with the summer (360 COGs) or winter (584 COGs) data sets. The Antarctic winter library encodes a more diversified metabolic capacity than the summer library, including pathways for carbon fixation, and nitrogen and sulfur utilization that were not detected in the summer library whereas the summer metagenome encodes more signal transduction and transport capabilities. **Conclusion:** The Antarctic summer and winter metagenomes revealed major differences in diversity of organisms and functions in temporally distinct marine bacterioplankton communities.

DdrD: A Component of UV and mitomycin resistance in *Deinococcus radiodurans* R1

Jana R. Robins*, Kathiresan Selvam, and John R. Battista
Department of Biological Sciences, Louisiana State University, Baton Rouge, LA

Five novel gene products have been linked to the ionizing radiation (IR) resistance of *D. radiodurans* R1. The transcripts of these genes, designated *ddrA* (DR0423), *ddrB* (DR0070), *ddrC* (DR0003), *ddrD* (DR0326), and *pprA* (DRA0346) are the five most highly induced in response to IR and desiccation (a stress that like IR induces DNA double strand breaks). Mutants created by deleting these five loci alone and in all possible combinations revealed that each gene product contributed, at least partially, to IR resistance, but the functions of these proteins has remained obscure. In the present study, the same collection of mutants is analyzed to determine if these gene products also affect *D. radiodurans*' capacity to survive exposure to UV light or mitomycin C. We report that DdrD contributes to overlapping processes that contribute to UV and mitomycin resistance. When both DdrD and PprA are inactivated, cells are as much as 1000-fold more sensitive to the lethal effects of UV light relative to their wild type parent. We also report that DdrD and DdrC must be inactivated to observe sensitivity to mitomycin C. The role of this protein does not appear related to either excision repair or homologous recombination, suggesting it mediates a previously undefined mechanism that facilitates tolerance of at least three distinct DNA damaging agents.

Genomic studies for the discrimination and assay of gene content of commercial products of *Bacillus thuringiensis*

P. Shwed*, J. Crosthwait, V. Seligy, Safe Environments Programme, Health Canada

Bacillus thuringiensis (Bt) is a spore-forming gram positive bacterium and member of the *Bacillus cereus* (Bc) group, which comprises four species, and includes strains known to have pathogenic and/or beneficial properties. At least 15 fully sequenced Bc group genomes are publicly available, but none include genomes of commercial Bt strains such as of the subspecies *israelensis* (Bti) and *kurstaki* (Btk). Comparative analyses of the available Bt and Bc genomes reveal a high level of gene synteny, which includes several human/animal virulence factors. We are interested in delineating precise molecular-level differences between commercial Bt and Bc group genomes to determine novelty and allow the ability to track strains in the environment.

This presentation provides an overview of two genomic studies we have carried out towards discrimination and assay of the genetic content of biotech products containing strains of Bti and Btk. Microarrays were designed for Bc ATCC 14579 (type strain) chromosomal, *Bacillus spp.* plasmid and unique oligonucleotide typing probes, predicted from Bc group annotated genomes (National Centre for Biotechnology information database). Genomic DNA from Bc4579 and biotech strains were compared using fluorescent and electrochemical probes. The scoring for presence of virulence genes and factors was most consistent with the latter probe. In addition, examination of DNA sequence of Bt biotech strains in regions that correspond to the Bc 14579 genomic islands, resulting from horizontal gene transfer, indicated absence of some Bc 14579 genes and gain of transposable elements and novel open reading frames. Collectively, the results of these studies validate the relatedness of Bt strains to Bc 14579, demonstrate the presence of potential virulence genes and indicate DNA sequences that may be used to discriminate commercial products of Bt.

New Insights into Methylotrophic Metabolism and Regulation in *Methylobacterium extorquens* AM1

Elizabeth Skovran, Lauren D. Palmer and M.E. Lidstrom
Department of Chemical Engineering, University of Washington

Methylotrophic bacteria are of interest due to their environmental significance as well as their potential to be used in biocatalysis applications for the production of chemicals from renewable biofeedstocks such as methanol. Two approaches were used to find new genes required for methylotrophic growth and links to other metabolic pathways in *Methylobacterium extorquens*. First, global gene expression from multi-carbon chemostat grown cells (succinate) and single carbon chemostat grown cells (methanol and methylamine) was compared. Second, microarrays were used to follow gene expression after switching cells from succinate to methanol growth. These data indicated a putative operon of importance for methylotrophic growth containing genes encoding the following putative functions: oxalate/formate antiporter, hydrolase, reductase and acylCoA synthetase. Loss of the reductase caused an inability to grow on C1 compounds and resulted in methanol sensitivity likely due to the accumulation of a toxic intermediate. Secondary mutations located in this operon prevented the formation of this intermediate and restored the ability to grow on C1 substrates. Second site null mutations in the Serine Cycle activator, QscR, also allowed for partial growth of the reductase mutant on C1 substrates and relieved methanol sensitivity. Gel shift assays and gene expression studies suggested that QscR may directly activate the reductase operon, partially explaining how mutations in *qscR* relieve the methanol sensitivity phenotype of the reductase mutant. Further gene expression studies indicated that lesions in *qscR* and the reductase result in opposing regulation of C1 metabolism genes which is partially remedied in the *qscR* reductase double mutant. Further characterization of these mutants will reveal much about methylotrophic metabolism and its regulation.

Digging a Little Deeper: Characterizing Bacterial Diversity in Soil Samples Excavated From One-Mile Under Terminal Island in Los Angeles, California.

K .Soni^{1*}, B. Kirkpatrick¹, J. Young², and E. Sanders-Lorenz¹. ¹ University of California, Los Angeles, ² Terralog Technologies USA, Inc.

The soil is a complex natural system, harboring an extensive community of microorganisms, which coexist together to create a unique terrestrial environment of astounding ecological and phylogenetic diversity. As part of a renewable energy project that involves a partnership between the City of Los Angeles and Terralog Technologies USA, Inc., two deep wells, a vertical well and one deviated well, were drilled on Terminal Island located in the L.A. basin. The project strategy includes a field experiment, in which biosolids will be injected into the vertical well where they will undergo thermal, anaerobic biodegradation, producing methane that can be captured by the deviated well and utilized as a biofuel. During drilling, sandstone and mudstone drill cutting samples were collected from between 1500 to 5500 feet below the soil surface. A drill cutting sample (4080 - 4100 ft depth) and a core sample (5004.2 - 5005.2 ft depth) were shared with the students in the "I, Microbiologist" laboratory course at UCLA. An assortment of bacteria was cultivated using enrichment conditions that mirrored the deep subsurface environment, which students predicted to have little to no oxygen, high temperature, high pressure, high salinity, and to be rich in minerals and pervaded with hydrocarbons such as petroleum. Although hundreds of colonies were obtained on the original enrichment plates, many were difficult to propagate beyond this first step. Genomic DNA was harvested from a few microbial isolates. Subsequently, a mixture of colonies representing the cultivatable community was pooled for genomic DNA extraction. When possible, total metagenomic DNA was purified directly from the samples. The genomic DNA sources were subjected to PCR and sequencing of 16S rRNA genes. DNA sequences were compared to those of classified organisms using NCBI-BLAST and RDP-II. A comprehensive picture of the soil bacterial community was obtained by building phylogenetic trees.

The Two Roles of a Signal Peptidase, SipW, in Biofilm Formation by *Bacillus subtilis*.

Rebecca Terra* & Beth Lazazzera. University of California, Los Angeles.

A biofilm is a community of microbial cells encased in a self-produced, exopolymeric matrix that forms at air-liquid (i.e. pellicles) or surface-liquid interfaces. In *Bacillus subtilis*, the type I signal peptidase, SipW, is required for the formation of both surface-adhered and pellicle biofilms. It has previously been reported that SipW has two functions in biofilm formation: processing of its target proteins TasA, a major component of the biofilm matrix, and YqxM, and a second, independent function. In this study, this second TasA-independent function was expounded on through the characterization of a *sipW* suppressor mutant strain that forms surface-adhered biofilms. This strain had acquired a loss-of-function mutation in *sinR*. Similar *sinR* mutations had been shown to suppress the pellicle biofilm defect of other mutant strains due to de-repression of the *eps* operon, which encodes for exopolysaccharide production. We similarly showed that de-repression of the *eps* operon was sufficient to suppress the biofilm defect of a *sipW* mutant, as a strain expressing *eps* from a heterologous promoter reproduced the suppressor phenotype. We further found that the phenotypes of the *sipW* and *tasA* mutants were identical when *eps* was de-repressed, suggesting that de-repression of the *eps* operon bypassed the TasA-independent function of SipW. Interestingly, we found that this TasA-independent function of SipW was associated with biofilms adhered to a surface and not with pellicle biofilms. In addition, we found that an *eps* mutation was epistatic to a *sipW* mutation, suggesting that *eps* is downstream of *sipW* in the same genetic pathway. Taken together, our data suggests a model in which the TasA-independent function of SipW is involved in regulating expression of *eps*. This work provides insight into how bacteria coordinate production of different components of the biofilm matrix.

Genome-scale reconstruction of *E. coli*'s transcriptional and translational machinery: A knowledge-base and its mathematical formulation

Ines Thiele *, Ronan M. T. Fleming, Neema Jamshidi, and Bernhard Ø. Palsson
University of California, San Diego

Metabolic network reconstructions are available for many organisms. They represent a scaffold for high-throughput data integration and can be used to computationally interrogate network properties; however, these reconstructions do not explicitly account for macromolecules. Here, we present the first genome-scale network reconstruction of *E. coli*'s transcriptional and translational machinery that is sequence-specific and accounts for all chemical transformations necessary to produce functional gene products. Legacy data from more than 500 publications and 3 databases were used to include details such as stable RNA maturation and modification, protein complex formation and iron-sulfur cluster assembly. The resulting reconstruction represents the first comprehensive knowledge base about these important cellular functions in *E. coli* and is unique in its scope. We demonstrate the use of this reconstruction, which can be converted to a mathematical format, 1) to quantitatively integrate various high-throughput data sets, such as uptake rates, transcription rates, and gene expression data, as constraints on the network, and 2) to compute functional network states of the data-constrained network and compare them to independent observations. Computed ribosome production rates compared very well with observed rates. Functional modules in the network were determined and many were found to contain gene products from multiple subsystems highlighting the functional interaction of proteins and stoichiometric coupling of their production rates. This first comprehensive reconstruction of *E. coli*'s transcriptional and translational machinery has many latent applications, and has the potential to have a similar impact on systems biology as its metabolic predecessors.

Single Virus Genomics: New Technology for Viral Ecology

L. Zeigler, R. Lasken, and S. J. Williamson*. J. Craig Venter Institute, San Diego, CA.

Whole genome amplification and sequencing of single microbial cells has revolutionized microbial ecology by allowing researchers to directly examine the genomic contents of individual cells in the absence of prior cultivation efforts. The field of environmental viral metagenomics has gained momentum over the past five years, but sequencing of individual environmental viral genomes is still dependent on the establishment of cultivable virus-host systems. In order to circumvent the potential selection bias that is inherent to the cultivation process, we have developed methods to amplify the entire genomes of single virus particles in preparation of sequencing. As a proof of concept, high titers of a bacteriophage T4 and lambda mixture were reduced to single virus particles using a combination of flow cytometry and serial dilution. Discreet virus particles were immobilized in a thin layer of agarose and epifluorescence and confocal microscopy were used for single virus particle verification. Multiple displacement amplification (MDA) was performed *in situ* to amplify the whole genomes of embedded virus particles. The immobilized amplified viral genomes were visualized using epifluorescence microscopy and validated through multiplex PCR amplification of the T4-specific gp23 major head protein and the lambda-specific integrase gene. Imminent future directions include full genome sequencing of one amplified T4 phage genome and one amplified lambda phage genome for complete substantiation of this technique. We intend to extend this technique to naturally occurring populations of marine viroplankton. Molecular fingerprinting (via RAPD-PCR) and neural network training on the extracted image variables will be used to identify unique viruses that are poised for sequence analysis. The ability to generate an extensive environmental viral reference genome library provides a first step towards constraining the amazing level of viral genotypic diversity witnessed within environmental samples and will be highly complementary to existing viral metagenomic data that currently suffers from a lack of robust assembly.

A Simple, Fast, and Accurate Method of Phylogenomic Inference

Martin Wu * & Jonathan A. Eisen. University of California, Davis

The explosive growth of genomic data provides an opportunity for increased usage of protein markers for phylogenetic inference. In response, we have developed an automated pipeline for phylogenomic analysis (AMPHORA) that overcomes the existing bottlenecks limiting large-scale protein phylogenetic inference. We demonstrated its high throughput capabilities and high quality results by constructing a genome tree of 578 bacterial species and by assigning phylotypes to 18,607 protein markers identified in metagenomic data collected from the Sargasso Sea.

Sampling and Preservation of Thermophilic Bacteria from a mile deep

Ben Attai¹, Omar Moghaddam¹, Shahrouzeh Saneie^{1*}, Mike Bruno², and Jean Young^{2*}. ¹ City of Los Angeles, Bureau of Sanitation, ²Terralog Technologies USA Inc

Most biosolids in the US are currently managed by land application. As the ratio of urban to farm areas increase, this practice becomes more difficult and restricted. The City of Los Angeles and Terralog are testing an innovative technology for converting biosolids into clean energy by deep well injection and thermal biodegradation. A slurry mixture of treated, non hazardous municipal sludge and water is being injected into a high permeability, unconsolidated sandstone formation at about a mile deep under the Terminal Island Treatment Plant, operated by the City of Los Angeles, Bureau of Sanitation. The biosolids injected will continue to biodegrade to CH₄ and CO₂ under in situ condition of 159°F. One injection and one monitoring wells were drilled in June to July 2007, and the injection well completed in January 2008. Drill cuttings from both wells, 12.4ft conventional core sample at the injection zone from the monitoring well, and 8 side-wall cores samples from the injection well were collected. Additionally, pressurized formation fluid at injection depth were taken and analyzed for geochemical composition. Periodic fluid samples will be obtained from the monitoring well to quantify for CH₄ generation and CO₂ sequestration.

A third monitoring well is planned for drilling in about a year. These wells are all permitted by US EPA under UIC (Underground Injection Control) program.

DIRECTORY

Heather Allen
University of Wisconsin-Madison
1550 Linden Drive, 1645 Microbial
Sciences Building
Madison, WI 53706
Tel: (608) 262-8735
hkallen@wisc.edu

Sasan Amini
Princeton University
Department of Molecular Biology
L206 Carl C. Icahn Laboratory,
Washington Road, Princeton University,
Princeton, NJ 08544
Tel: (609) 258-5738
Fax: (609) 258-8004
sasan@princeton.edu

Diego Arambula
University of California, Los Angeles
615 Charles E. Young Dr., BSRB 254
Los Angeles, CA 90095
Tel: (310) 825-0986
Fax: (310) 267-2774
shokenryu@hotmail.com

E. Virginia Armbrust
School of Oceanography
Box 357940
University of Washington
Seattle, WA 98195
Tel: (206) 616-1783
Fax: (206) 685-6651
armbrust@ocean.washington.edu

David J. Arriola
Point Loma Nazarene University
3900 Lomaland Drive
San Diego, CA 92106
Tel: (619) 849-2205
darriola100@pointloma.edu

Sergey Artiushin
University of Kentucky
Gluck Equine Research Center
Lexington, KY 400546-0099
Tel: (859) 257-4757 ext. 81111
Fax: (859) 257- 5169
scartil@email.uky.edu

Albert Barberan
Center for Ecology & Evolutionary
Biology (CEEB)
Acces Cala St Francesc, 14
17300 Blanes, SPAIN
Tel +34 972 33 6101
Fax +34 972 33 7806
abarberan@ceab.csic.es

Wojciech Bartkowski
University of California, Los Angeles
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 825-4009
Fax: (310) 206-5231
wojciech@lifesci.ucla.edu

John R. Battista
Louisiana State University
202 Life Sciences Bldg
Baton Rouge, LA 70803
Tel: (225) 578-2810
Fax: (225) 578-2597
jbattis@lsu.edu

Kate Bayliss
170 E. 6th St. Su. #132; Box #766
Claremont, Ca 91711
Tel: (509) 670-4646
Ksr02006@mymail.pomona.edu

Elinne Becket
405 Hilgard Ave.
Molecular Sciences Bldg. #1602G
Los Angeles, CA 90095
Tel: (310) 825-8924
elinneb@gmail.com

Steven Benner
The Foundation for Applied Molecular
Evolution
Tel: (352) 271-7005
sbenner@ffame.org

Devaki Bhaya
Carnegie Institution
260 Panama St., Stanford, CA 94305
Tel: (650) 325-1521
Fax: (650) 325-6857
dbhaya@stanford.edu

Fredrick Blattner
University of Wisconsin-Madison
425G Henry Mall, Rm 5430
Madison, WI 53706
Tel: (608) 890-0191
Fax: (608) 262-2976
fred@genome.wisc.edu

Jim Bristow
DOE Joint Genome Institute
2800 Mitchell Drive, B400
Walnut Creek, CA 94598-1632 USA
Ph: 925-296-5693
Fax: 925-296-5620
JBristow@lbl.gov

Jean Challacombe
Joint Genome Institute, Los Alamos
Bioscience Division
Mail Stop M888
Los Alamos, NM 87545
Tel: (505) 665-1485
Fax: (505) 665-3024
jchalla@lanl.gov

Yun-juan (Janet) Chang
Oak Ridge National Laboratory
P.O Box, MS 6164
Oak Ridge, TN 37831-6164
Tel: (865) 574-8912
Fax: (865) 576-5332
yjs@ornl.gov

Casey Chen
University of Southern California
School of Dentistry
925 W. 34th St.
Los Angeles, CA 90089-0641
Tel: (213) 740-1075
ccchen@usc.edu

Byung-Kwan Cho
University of California San Diego
9500 Gilman Dr. Dept. 0412
La Jolla, CA 92093-0412
Tel: (858) 534-5668
Fax: (858) 534-5722
bcho@ucsd.edu

Sandra Clifton
Washington University
4444 Forest Park Ave
Campus Box 8501
St. Louis, MO 63108
Tel: (314) 286-1800
Fax: (314) 286- 1810
sclifton@watson.wustl.edu

Frank Collart
Argonne National Lab
Bldg 202
Bioscience Division
9700 S. Cass Ave
Lemont, IL 60439
Tel: (630) 252-4859
Fax: (630) 252-3387
fcollart@anl.gov

Balint Csorgo
Hungary Academy of Science
Institute of Biochemistry
6726 Szeged, Temesvari krt. 62.,
Hungary
Tel: 36-62-599 600
Fax: 36-62-433 506
csorgob@brc.hu

Carla Cueva
California State University, Los Angeles
2015 Easy Ave
Long Beach, CA 90810
Tel: (562) 506-6277
miluska13@gmail.com

Valérie de Crécy-Lagard
University of Florida
Microbiology and Cell Science
Building 981, Museum Road
Gainesville, FL 32611-0700
vcrecy@ufl.edu

Justin Donato
University of Wisconsin
Microbial Sciences Building
1550 Linden Drive
Madison, WI 53706
Tel: (608) 262-8735
Fax: (608) 265-5289
jdonato@wisc.edu

John E. Dueber
University of California, Berkley
327 Stanley Hall
Berkley, CA 94720
Tel: (510) 664-4200
jdueber@berkeley.edu

Ashlee Earl
Harvard Medical School
200 Longwood Ave, D-1 #219
Boston, MA 02115
Tel: (617) 432-4359
Fax: (617) 738-7664
Ashlee_earl@hms.harvard.edu

Jonathan Eisen
UC Davis Genome Center
Genome and Biomedical Sciences
Facility Room 5311
451 East Health Sciences Dr.
Davis, CA 95616-8816
Tel: (530) 752-3498
Fax: (530) 754-9658
jaeisen@ucdavis.edu

Basma El Yacoubi
University of Florida
Microbiology and Cell Science
Building 981, Museum Road
Gainesville, FL 32611-0700
basma@ufl.edu

Aria Eshraghi
University of California, Los Angeles
609 Charles E. Young Dr. East MSB
1602G
Los Angeles, CA 90095-0001
Tel: (310) 206-7517
Fax: (310) 206-5231
Aria.e@ucla.edu

Daisy Evans
University of California, Los Angeles
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 825-4009
Fax: (310) 206-5231
elvenwarriorqueen@yahoo.com

Tamas Feher
Institute of Biochemistry, Biological
Research Center Hungarian Academy of
Sciences
6726 Szeged, Temesvari krt. 62.,
Hungary
Tel: 36-62-599 600
Fax: 36-62-433 506
fehert@brc.hu

Sorel Fitz-Gibbon
University of California, Los Angeles
Molecular & Medical Pharmacology
Crump Institute for Molecular Imaging
700 Westwood Plaza
Box 951770
Los Angeles, CA 90095-1770
Tel: (310) 206-5585
sorel@ucla.edu

Elizabeth Foza
NIH
Building 18T, Room 101
18 Library drive, MSC 5430
Bethesda, MD 20892-5430
Tel: (301) 496-7082
fozoeliz@mail.nih.gov

Richard Fredendall
California State University San
Bernardino
5500 University Pkwy.
San Bernardino, CA 92407
Tel: (909) 537-5405
fredendr@csusb.edu

Maria Ghirardi
National Renewable Energy Laboratory
Basic Sciences Center
1617 Cole Blvd.
Golden, CO 80401
Tel: (303) 384-6312
Fax: (303) 384-6150
Maria_Ghirardi@nrel.gov

Hani Goodarzi
The Lewis-Sigler Institute for Integrative
Genomics, Princeton University
Carl C. Icahn Laboratory, Washington
Road, Princeton University
Princeton, NJ 08544
Tel: (609) 258-5738
Fax: (609) 258-8004
goodarzi@princeton.edu

Huatao Guo
University of California, Los Angeles
615 Charles E. Young Dr., BSRB 254
Los Angeles, CA 90095
Tel: (310) 825-0986
Fax: (310) 267-2774
htguo@ucla.edu

Cliff Han
Los Alamos National Lab (JGI) M888
Los Alamos, NM 87544
Tel: (505) 665-1724
Fax: (505) 665-3024
Han_cliff@lanl.gov

Jun Hang
EdgeBio
201 Perry Parkway, Suite 5
Gaithersburg, MD 20877
Tel: (301) 990-2685
Fax: (301) 990-0881
jhang@edgebio.com

Heather Harbottle
U.S. Food and Drug Administration
8401 Muirkirk Rd
Laurel, MD 20708
Tel: (301) 210-4246
Fax: (301) 210-4685
heather.harbottle@fda.hhs.gov

Caroline S. Harwood
University of Washington
1705 NE Pacific Street
Department of Microbiology-Box
357242
Health Sciences Building, Room K-
335B
Seattle, WA 98195-7242
Tel: (206) 221-2848
Fax: (206) 221-5041
csh5@u.washington.edu

Loren Hauser
Oak Ridge National Laboratory
One Bethel Valley Road, P.O. Box 2008
Oak Ridge, TN 37831-6164
Tel: (865) 241-1960
Fax: (865) 576-5332
hauerlj@ornl.gov

Craig Herbold
University of California Los Angeles
BOX 951570, 242 MBI Bldg
Los Angeles, CA 90095-1570
cherbold@ucla.edu

Chris Herring
Mascoma Corporation
16 Cavendish Court
Lebanon, NH 03766
Tel: (603) 676-3320
Fax: (603) 676-3321
tlivingstone@mascoma.com

Matthias Hess
Joint Genome Institute
2800 Mitchel Dr.
Walnut Creek, CA 94598
Tel: (925) 296-5746
Fax: (925) 296- 5720
mhess@lbl.gov

Maureen Hillenmyer
Stanford University
Stanford Genome Technology Center
Address: 855 California Avenue, Palo
Alto, CA 94304
Tel: (650) 812-1973
maureenh@stanford.edu

Aaron Hinz
University of Washington
MCB Program
Box 357275
Seattle, WA 98195-7275
Tel: (206) 543-7795
Fax: (206) 685-8174
ahinz@u.washington.edu

Jean Huang
University of Washington
1705 NE Pacific Street
Department of Microbiology-Box
357242
Health Sciences Building, Room K-
335B
Seattle, WA 98195-7242
Tel: (206) 221-2848
Fax: (206) 221-5041
Jeanh4@u.washington.edu

Colin Ingham
Wageningen University and Research
Centre
Laboratory of Microbiology
P.O. Box 8033,
6700 EJ Wageningen, The Netherlands
colin.ingham@wur.nl

Scott A. Jackson
U.S. FDA
8301 Muirkirk Rd.
MOD-1 HFS-025
Laurel, MD 20708
Tel: (301) 210-7807
scott.jackson@fda.gov

Sergey Kalachikov
Columbia University
1150 St Nicholas Ave., rm 406,
New York, NY 10032
Tel: (212) 851-5193
Fax: (212) 851-5215
Sk363@columbia.edu

Steven Karpowicz
University of California, Los Angeles
Chemistry & Biochemistry Dept.
BOX 951569, 5070A Young
Los Angeles, CA 90095-1569
Tel: (310) 825-8300
Fax: (310) 206-1035
skarp@chem.ucla.edu

Cheryl A. Kerfeld
DOE Joint Genome Institute
Education/ Structural Genomics
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: (925) 296-5691
Fax: (925) 296-5752
ckerfeld@lbl.gov

William Klimke
National Center for Biotechnology
Information
6As.43D-49, Building 45,
Bethesda, Maryland, 90894
Tel: (301) 496-4859
Fax: (301) 402-9651
klimke@ncbi.nlm.nih.gov

Christopher Lam
California State University, Los Angeles
5151 State University Drive
Los Angeles, CA 90032
Tel: (323) 343-2040
ibillgates@yahoo.com

Miriam Land
Oak Ridge National Laboratory
One Bethel Valley Road, P.O. Box 2008
Oak Ridge, TN 37831-6164
Tel: (865) 241-6231
Fax: (865) 576-5332
landml@ornl.gov

Morgan Langille
Department of Molecular Biology and
Biochemistry
Simon Fraser University
8888 University Dr. Burnaby, BC,
Canada V5A 1S6
Tel: (778) 782-5414
Fax: (778) 782-5583
mlangill@sfu.ca

Kim Lewis
Northeastern University
360 Huntington Avenue
Mugar 405
Boston, MA 02115
Tel: (617) 373-8238
Fax: (617) 373-3724
k.lewis@neu.edu

Huiying Li
University of California, Los Angeles
Molecular & Medical Pharmacology
Crump Institute for Molecular Imaging
700 Westwood Plaza
Box 951770
Los Angeles, CA 90095-1770
Tel: (310) 206-5585
huiying@mednet.ucla.edu

James C. Liao
University of California, Los Angeles
Chemical & Biomolecular Engineering
BOX 951592, 5531 BH
Los Angeles, CA 90095-1592
Tel: (310) 825-1656
Fax: (310) 206-4107
liaoj@ucla.edu

Mary E. Lidstrom
University of Washington
Office of Research
Box 351202
Seattle, WA 98195-1202
Tel: (206) 685-7641
Fax: (206) 685-9210
lidstrom@u.washington.edu

Anne Liu
University of California, Los Angeles
405 Hilgard Ave.
Molecular Sciences Bldg. #1602G
Los Angeles, CA 90095
Tel: (310) 825-8924
Anneliu9@ucla.edu

Cindy M. Liu
Translational Genomics Research
Institute
3051 W Shamrell Blvd., Ste 106
Flagstaff, AZ 86001
Tel: (928) 226-6364
Fax: (928) 226-6360
cliu@tgen.org

Minghsun Liu
University of California, Los Angeles
615 Charles E. Young Dr.
254 BSRB c/o J. F. Miller
Los Angeles, CA 90095
lmh@ucla.edu

Tim Magnuson
Idaho State University
Room 326A Gale Life Science
921 South 8th Avenue, Stop 8007
Pocatello, ID 83209-8007
Tel: (208) 282-5014
magntimo@isu.edu

Eric Mathur
Synthetic Genomics
111149 N. Torrey Pines Rd. Ste. 100
La Jolla, CA 92037
Tel: (858) 754- 2903
Fax: (858) 754-2988
emathur@syntheticgenomics.com

Shauna McGillivray
University of California San Diego
Cellular & Molecular Medicine East,
Rm 1066
9500 Gilman Dr. Mail code 0687
La Jolla, CA 92093-0687
Tel: (858)534-7408
Fax: (858) 534-5611
smcgilli@ucsd.edu

Sabeeha Merchant
University of California, Los Angeles
Chemistry & Biochemistry Dept.
BOX 951569, 5070A Young
Los Angeles, CA 90095-1569
Tel: (310) 825-8300
Fax: (310) 206-1035
sabeeha@chem.ucla.edu

Tony Michael
Institute of Food Research
Norwich Research Park, Colney
Norwich, NR4 7UA
United Kingdom
Tel: 44 1603 255356
Tony.Michael@bbsrc.ac.uk

Jason Rafe Miller
J. Craig Venter Institute
9704 Medical Center Drive
Rockville MD 20850-3343 USA
Tel: (301) 795-7903
Fax: (301) 795-7060
jmiller@jcvl.org

Jeffery F. Miller
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 206-7926
jfmiller@ucla.edu

Jeffrey H. Miller
University of California Los Angeles
Dept. of Microbiology
405 Hilgard Ave.
Los Angeles, CA 90095
Tel: (310) 825-8460
Fax: (310) 206-3088
jhmiller@microbio.ucla.edu

Jenna Morgan
Joint Genome Institute 445H
2800 Mitchell Drive
Walnut Creek, CA 95698
Tel: (510) 499-4649
Fax: (925) 296-5850
jlmorgan@lbl.gov

Hirotsada Mori
Nara Institute of Science and
Technology
8916-5 Takayama, Ikoma,
Nara 630-0101, Japan
Tel: 81-743-72-5660
Fax: 81-743-72-5669
hmori@gtc.naist.jp

Irina Morozova
Columbia University
1150 St Nicholas Ave., rm 406
New York, NY 10032
Tel: (212) 851-5162
Fax: (212) 851-5215
Im198@columbia.edu

A. Christine Munk
Joint Genome Institute-Los Alamos
National Laboratory
MS M888 Los Alamos National Lab
Los Alamos, NM 87545
Tel: (505) 667-0779
cmunk@lanl.gov

Vivek Mutalik
University of California, San Francisco
Genentech Hall, Room S376
600 16th Street
University of California at San Francisco
San Francisco, CA 94158-2517
Tel: (415) 476-1493
Vivek.Mutalik@ucsf.edu

Nan Li
DNA STAR, Inc.
3801 Regent Street
Madison, WI 53705
Tel: (608) 258-7420
Fax: (608) 258-7439
toms@dnastar.com

Robert Nichols
University of California at San Francisco
Gross Laboratory
Genentech Hall Room S376
600 16th St., MC 2200
San Francisco, CA 94154-2517
Tel: (415) 476-1493
Robert.Nichols@ucsf.edu

Trent Northen
The Scripps Research Institute
Center for Mass Spectrometry BCC007
10550 North Torrey Pines Road
La Jolla, CA 92037
Tel: (858) 784-9415
Fax: (858) 784-9496
trentnorth@gmail.com

Mariusz Nowacki
Princeton University
Guyot Hall room 323
Princeton, NJ 08544-1003
Tel : (609) 258-1933
Fax : (609) 258-7892
mnowacki@princeton.edu

James O'Dwyer
University of Oregon
Center for Ecology & Evolutionary
Biology
1025 University St, Pacific 335
5289 University of Oregon
Eugene, OR 97403-5289
jodwyer@uoregon.edu

Paul Orwin
California State University San
Bernardino
5500 University Pkwy.
San Bernardino, CA 92407
Tel: (909) 537-5405
porwin@csusb.edu

Bernhard Palsson
University of California, San Diego
9500 Gilman Drive Dept. 0412
La Jolla, CA 92093-0412
Tel: (858) 534-5668
Fax: (858) 534-5722
bpalsson@be-research.ucsd.edu

Young Seob Park
University of California, San Diego
9500 Gilman Drive Dept. 0412
La Jolla, CA 92093-0412
Tel: (858) 534-5668
Fax: (858) 534-5722
yspark@bioeng.ucsd.edu

Julian Parkhill
The Wellcome Trust Sanger Institute
Pathogen Sequencing Unit
Genome Campus Hinxton
Cambridge, CB 10 ISA
United Kingdom
Tel: 44-1223-834-244
Fax: 44-1223-494-919
Parkhill@sanger.ac.uk

Michael Pehl
California State University San
Bernardino
5500 University Pkwy.
San Bernardino, CA 92407
Tel: (909) 537-5405
mpehl@csusb.edu

Matteo Pellegrini
University of California, Los Angeles
Molecular, Cell & Developmental
Biology
BOX 951606
621 Charles E. Young
Los Angeles, CA 90095-1606
Tel: (310) 825-0012
matteop@mcdb.ucla.edu

Liz Perry
University of Oregon
Center for Ecology & Evolutionary
Biology (CEEB)
1025 University St, Pacific 335
5289 University of Oregon
Eugene, OR 97403-5289
eperry1@uoregon.edu

Brian Pflieger
University of Wisconsin Madison
3629 Engineering Hall
1415 Engineering Dr
Madison, WI 53706
Tel: (608) 890-1940
Fax: (608) 262-5434
pflieger@eagr.wisc.edu

Irving J. Phillips
California State University, Los Angeles
Department of Biological Sciences
5151 State University Drive
Los Angeles, CA 90032
Tel: (323) 343-2040
rasguear@hotmail.com

Vasiliy A. Portnoy
University of California, San Diego
500 Gilman Dr. Dept 0412
La Jolla, CA 92093-0412
Tel: (858) 534-5668
Fax: (858) 534-5722
vasiliy.portnoy@gmail.com

Gyorgy Posfai
Institute of Biochemistry, Biological
Research Center Hungarian Academy of
Sciences
67226 Szeged
Temesvari krt. 62., Hungary
Tel: 36-62-599 600
Fax: 36-62-433 506
posfai@brc.hu

Lance B. Price, PhD
Translational Genomics Research
Institute
3051 W. Shamrell Blvd., Suite 106
Flagstaff, AZ 86001
Tel: (928) 226-6371
Fax: (982) 226-6360
lprice@tgen.org

Simon Prochnik, PhD.
University of California Berkeley
Room 545 L. S. A. Bldg. #3200
Berkeley, CA 94720-3200
Tel: (510) 643-9944
Fax: (510) 643-9947
seprochnik@lbl.gov

Yu Qiu
University of California, San Diego
9500 Gilman Drive Dept. 0412
La Jolla, CA 92093-0412
Tel: (858) 534-5668
Fax: (858) 534-5722
yuqiu@ucsd.edu

Elisabeth Raleigh
New England BioLabs, Inc.
240 County Road
Ipswich, MA 01938
Tel: (978) 380-7238
Fax: (978) 921-1350
Raleigh@neb.com

David A. Relman
Stanford University
VA Palo Alto Health Care System 154T
Building 101, Room B4-185
3801 Miranda Avenue
Palo Alto, CA 94304
Tel: (650) 852-3308
Fax: (605) 852-3291
relman@stanford.edu

Adam Retchless
University of Pittsburgh
352 Crawford Hall
Pittsburgh, PA 15260
Tel: (412) 624-4205
acr7@pitt.edu

Toby Richardson
Synthetic Genomics
111149 N. Torrey Pines Rd. Ste. 100
La Jolla, CA 92037
Tel: (858) 754-2909
Fax: (858) 754-2988
trichardson@syntheticgenomics.com

Christian Riesenfeld
Desert Research Institute
Division of Earth and Ecosystem
Sciences
2215 Raggio Parkway
Reno, NV 89512
Tel: (775) 674-7583
Fax: (775) 673-7635
csr@dri.edu

Jana Robins
Louisiana State University
107 Life Sciences Building
Baton Rouge, LA 70803
Tel: (225) 578-5226
Fax: (225) 578-2597
Jrobi44@lsu.edu

Erin Sanders-Lorenz
Dept. of Microbiology
University of California Los Angeles
1602 Molecular Science Bldg.
Los Angeles, CA 90095
Tel: (310) 825-1783
Fax: (310) 206-5231
erinsl@microbio.ucla.edu

Shahrouzeh Saneie
City of Los Angeles
Department of Public Works
Bureau of Sanitation
Regulatory Affairs Division
12000 Vista Del Mar
Playa Del Rey, CA 90293
Tel: (310) 648-5239
Fax: (310) 648-5115
Shahrouzeh.Saneie@lacity.org

Philip Shwed
Health Canada
Safe & Environments Program
Environmental Health Center, PL0803A
50 Columbine Driveway, Rm 207
Tunney's Pasture, Ottawa
Ontario, Canada KIA OL2
Tel: (613) 941-7374
Fax: (613) 941-8530
Phil_shwed@hc-sc.gc.ca

Lynn L. Silver
LL Silver Consulting, LLC
3404 Park Place
Springfield, NJ 07081
Tel: (973) 218-1466
silverly@comcast.net

Gary Siuzdak
The Scripps Research Institute
Center for Mass Spectrometry BCC007
10550 North Torrey Pines Road
La Jolla, CA 92037
Tel: (858) 784-9415
Fax: (858) 784-9496
siuzdak@scripps.edu

Betsy Skovran
University of Washington
616 NE North Lake Place
Benjamin Hall Room 440
Seattle, WA 98105
Tel: (206) 616-6954
Fax: (206) 616-5721
eskovran@u.washington.edu

Karina Soni
University of California Los Angeles
1602 Molecular Science Bldg.
Los Angeles, CA 90095
Tel: (310) 825-1783
Fax: (310) 206-5231
ksoni@ucla.edu

Cindy Tamae
University of California, Los Angeles
405 Hilgard Ave.
Molecular Sciences Bldg. #1602G
Los Angeles, CA 90095
Tel: (310) 825-8924
ctamae@gmail.com

Stephanie Tan
California State University, Los Angeles
Department of Biological Sciences
5151 State University Drive
Los Angeles, CA 90032
Tel: (626) 382-9039
stephietan@gmail.com

Becky Terra
University of California, Los Angeles
Molecular Sciences 4610B
405 Hilgard Ave.
Los Angeles, CA 90095
Tel: (310) 794-4986
bterra@ucla.edu

Jill Terra
University of California, Los Angeles
609 Charles E. Young Drive East
1602 G MSB
Los Angeles, CA 90095
Tel: (310) 206-7517
Fax: (310) 206-5231
jillkatpot@yahoo.com

Ines Thiele
University of California, San Diego
9500 Gilman Drive Dept. 0412
La Jolla, CA 92093-0412
Tel: (858) 534-5668
Fax: (858) 534-5722
ines.thiele@gmail.com

Katherine Tran
University of California, Los Angeles
405 Hilgard Ave.
Molecular Sciences Bldg. #1602G
Los Angeles, CA 90095
Tel: (310) 825-8924
Ktran101@ucla.edu

Lilian Tran
University of California, Los Angeles
405 Hilgard Ave.
Molecular Sciences Bldg. #1602G
Los Angeles, CA 90095
Tel: (310) 825-8924
Lmtran08@gmail.com

Athanasios Typas
University of California San Francisco
Genentech Hall, Room S376
600 16th Street
University of California at San Francisco
San Francisco, VA 94158-2517
Tel: (415) 476-1493
athanasios.typas@ucsf.edu

James Wang
Cepheid
904 Caribbean Dr.
Sunnyvale, CA 94089
Tel: (408) 400-8261
jwang@cepheid.com

Barry L. Wanner
Purdue University
915 West State St.
Lilly Hall Rm 1-228
West Lafayette, IN 47907
Tel: (765) 494-8034
Fax: (765) 496-1189
blwanner@purdue.edu

George Weinstock, Ph.D.
The Genome Center at Washington
University
4444 Forest Park Ave.
Campus Box 8501
St. Louis, MO 63108
Tel: (314) 286-1879
gweinsto@watson.wustl.edu

Tiffany Williams
Baylor College of Medicine
Human Genome Sequencing Center,
Alkek N1519,
One Baylor Plaza
Houston, TX 77030
Tel: (713) 798-6539
Fax: (713) 798-5741
tmwillia@bcm.edu

Shannon Williamson
J. Craig Venter Institute
Director of Environmental Virology
VI West
Tel: (858) 200-1804
swilliamson@jcv.org

Sydnor Withers
Great Lakes Bioenergy Research Center
University of Wisconsin-Madison
3529 Microbial Sciences Building
Madison, WI 53706

K. Eric Wommack
University of Delaware
Delaware Biotechnology Institute
15 Innovation Way
Newark, DE 19711
Tel: (302) 831-4362
Fax: (302) 831-3447
wommack@dbi.udel.edu

Martin Wu
University of California, Davis
UC Davis Genome Center
451 Health Sciences Drive
5315 GBSF
Davis, CA 95616
Tel: (530) 752-8409
Fax: (530) 754-9658
martinwu@ucdavis.edu

Jean Young
Terralog
332 E. Foothill Blvd., suite B
Arcadia, CA 91006
Tel: (626) 305-8460
jtyoung@terralog.com

Alex Hon-Tsen Yu
University of California Davis
Genome center
Tel: (530) 752-8409
ayu@ntu.edu.tw

Yang Yu
Mascoma Corporation
16 Cavendish Court,
Lebanon, NH 03766
Tel: (603) 676-3320 x 1185
Fax: (603) 676-3321
tlivingstone@mascoma.com

Joe Zhou
Institute for Environmental Genomics
University of Oklahoma
101 David L. Boren Blvd.
Norman, OK 73072
Tel: (405) 325-6073
Fax: (405) 325-7552
jzhou@ou.edu

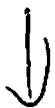
Anthony J. Zuccarelli
Loma Linda University
Office of Research Affairs
24888 Prospect Ave
Loma Linda, CA 92350
Tel: (909) 558-8544
Fax: (909) 558-0244
azuccarelli@llu.edu

Persistent Typology of Families

all vs all



diff cutoff



clusters



compare w/ TDA families

multiple samples

↓
X'gem fans

OTus
E1 E2 E3
1 2 1
1 0 1
2 0 1
3 10 10

Quote of the Day

"We are the environment
We live the phenotype"

David Relman
contemplating the
human microbiome