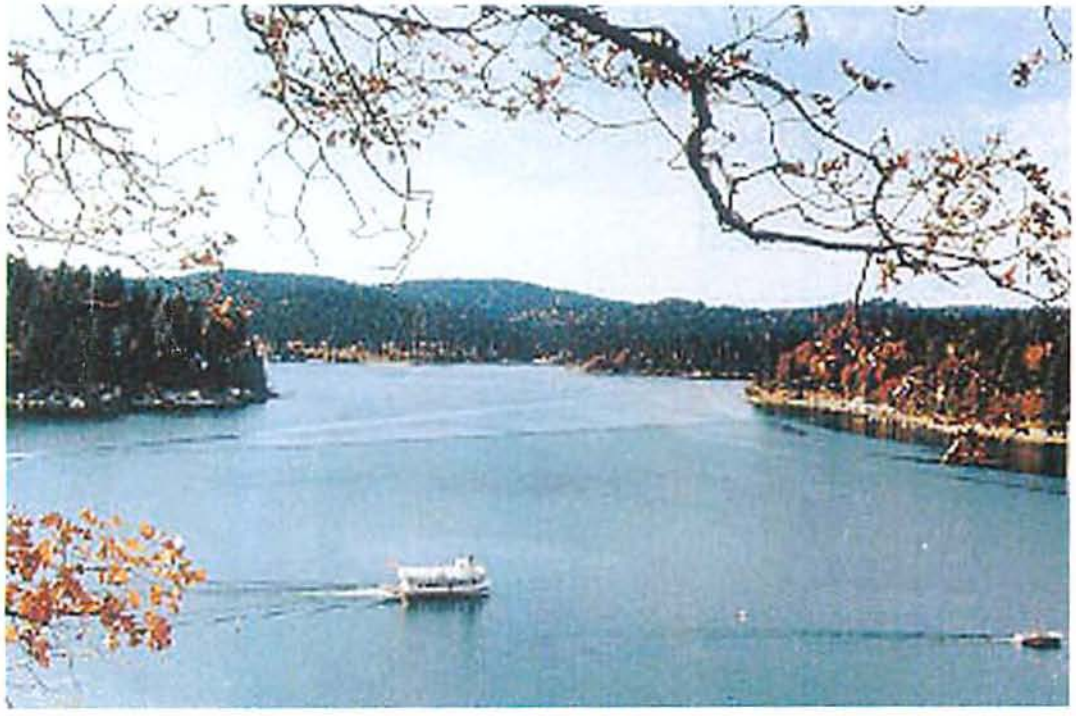
Jonathan Eisen

14th International

Microbial Genomics Conference



September 24-28, 2006

UCLA Conference Center

Lake Arrowhead

California

10:30-10:50 am	Break
10:50-11:20 am	Bernhard Palsson – University of California, San Diego, CA **"The Reconstruction of the TRN in *E.coli* and its Plasticity with Adaptation"**
11:20-11:50 am	Fredrick Blattner – University of Wisconsin, Madison, WI **"From genomes to Designed Genomes: The *E. coli* Reduction Experience"**
11:50 am-12:20 pm	**Special Topics Lecture:** Jeff F. Miller University of California, Los Angeles, CA **"Tropism Switching in Pathogens Defines a Family of Diversity-Generating Retroelements"**
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 Sequencing and Genomics Methods**
7:45-8:00 pm	Jeffrey H. Miller University of California, Los Angeles, CA **"KOFS: Knockout Fusion Screens for the *E. coli* KO Collection"**
8:00-8:30 pm	Aled Edwards University of Toronto, Toronto, Ontario **"Large Scale Biochemical Approaches to Annotated Genomes"**
8:30-9:00 pm	John H. Leamon 454 Life Sciences Corp, Branford, CT **"Routine, One Day Whole Genome Bacterial Sequencing"**
9:00-9:15 pm	Break
9:15-9:45 pm	Christiane Honisch Sequenom, Inc., San Diego, CA **"Development of New Methods for Comparative Sequence Analysis"**

9:45-10:15 pm	Zaida (Zan) Luthey-Schulten Univeristy of Illinois at Urbana,Champagne, Urbana, IL **"Evolutionary Profiles from the QR Factorization of Multiple Sequence Alignments"**

TUESDAY, SEPTEMBER 26

7:45-8:30 am	Breakfast (Dining Room)
Session III	**Resources/ Pathogens/ Biothreats**
8:45-9:15 am	Barry Wanner Purdue University, West Lfayette, IN **"EcoliHub, K-12 Information Resources – Under Construction"**
9:15-9:45 am	Eric Eisenstadt The Institute for Genomic Research, Rockville, MD **"On Using Genomics to Combat Agents of Infectious Diseases"**
9:45-10:15 am	Julian Parkhill Welcome Trust Sanger Institute, Cambridge, UK **"Comparative Genomics and Genome Organization of Microbial Pathogens"**
10:15-10:35 am	Break
10:35-11:05 am	David Raskin Harvard Medical School, Boston, MA **"Functional Genomic Studies of *Vibrio cholerae* Essential Genes"**
11:05-11:35 am	Jacques Ravel The Institute for Genomic Research, Rockville, MD **"Genomic Projects at TIGR Involving Pathogens and Biothreats"**
11:35 am-12:05 pm	Victor Nizet UCSD School of Medicine, La Jolla, CA **"Genetic Analysis of Virulence Factors Allowing Streptococcus and Staphylococcus to Escape Phagocyte Killing"**
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 I: Metagenomics/Biodiversity**
7:45- 8:15 pm	Gregory L. Challis University of Warwick, Coventry, UK **"Genome Mining in *Streptomyces coelicolor*"**
8:15- 8:45 pm	Nathan C. VerBerkmoes University of California, Berkeley, CA **"Strain resolved Proteomics Reveals that Recombination Shapes the Genomes of Acidophilic Bacteria in Natural Consortia"**
8:45- 9:00 pm	Break
9:00- 9:30 pm	Jessica Green University of California, Merced **"Probing Patterns of Microbial Diversity"**
9:30-10:00 pm	Curtis A. Suttle University of British Columbia, Vancouver, BC **"Metagenomic Analysis of Viral Diversity in the Sea"**

WEDNESDAY, SEPTEMBER 27

7:45-8:30 am	Breakfast (Dining Room)
Session V	**Microbial Communities II: Biodiversity / Interaction/ Evolution**
8:40-9:10 am	Jonathan Eisen University of California, Davis, CA **"Genomics of Mutualistic Symbioses and the Origin of Novelty"**
9:10-9:40 am	Richard E. Lenski Michigan State University, East Lansing, MI **"Experimental Evolution"**

9:40-10:10 am	Julie Huber Marine Biological Laboratory. Woods Hole, MA **"Microbial Diversity in the Deep Sea and the Underexplored Rare Biosphere"**
10:10-10:30 am	Break
10:30-11:00 am	Margaret A. Riley University of Massachusetts, Amherst, MA **"Microbial Phylogenies, Genomics, and the Bacterial Species Concept"**
11:00-11:30 am	Siv Andersson University of Uppsala, Uppsala, Sweden **"Alpha-Proteobacterial Genome Complexity"**
11:30 am-12:00 pm	Derek Lovley University of Massachusetts, Amherst, MA **"Systems Biology Approach to Groundwater Bioremediaton and Microbial Electricity Production"**
12:00-12:30 pm	Zakee Sabree University of Wisconsin-Madison, WI **"Getting Down and Dirty with Antibiotic Resistance: Cloning the Soil Resistome"**
12:30 pm	Lunch (Dining Room)
Session VI A Discovery	**Genomic Undergraduate Education Programs/ Gene Function**
4:30-5:00 pm	Cheryl Kerfeld University of California, Los Angeles, CA **"Combined Laboratory Course Effort to Sequence Complete Genome"**
5:00-5:30 pm	Erin Sanders-Lorenz University of California, Los Angeles, CA **"Making Individual Phylogenomic Trees in a Laboratory Course"**
5:30-6:00 pm	Tina K. Van Dyk DuPont, Wilmington, DE **"From Genome-Wide Reporter Gene Arrays to Gene Function Discovery and Biosensor Development"**

6:15-7:45 pm	Dinner
Session VI B	**Cell/Cell Communication**
7:45-8:15 pm	Chris Waters Princeton University, Princeton, NJ **"The *Vibrio harveyi* Quorum-Sensing System Uses Shared Regulatory Components to Discriminate Between Multiple Autoinducers"**
8:15-8:45 pm	Marvin Whiteley The University of Texas, Austin, TX **"Membrane Vesicles Traffic Signals and Facilitate Group Activity in *Pseudomonas aeruginosa*"**
8:45-9:15 pm	Tracy Teal California Institute of Technology, Pasadena, CA **"Spatiometabolic Stratification of *Shewanella oneidensis* Biofilms"**
9:15 pm	Reception/Party (Iris Room)

THURSDAY, SEPTEMBER 28

7:30-8:15 am	Breakfast (Dining Room)
Session VII	**New Model Organisms/Late Breaking Advances**
8:30-9:00 am	Robert Gunsalus Univeristy of California, Los Angeles **"First Genomic View of 'The Missing Link' in Anaerobic Food Chains: *Syntrophus aciditrophicus*"**
9:00-9:30 am	Jizhong Zhou University of Oklahoma, Norman, OK **"Genomics of Ethanol Producing Bacteria"**
9:30-10:30 am	Poster Talks
10:45 am	Check-out
11:00 am	Departure of 1st Conference Bus for LAX
12:00 pm	Lunch (Dining Room)
1:15 pm	Departure of Vans to LAX

The JGI: Metagenomics and Beyond

Eddy Rubin, MD, PhD
Joint Genome Institute

The Joint Genome Institute (JGI) has more than 25 metagenomic projects underway. Two projects where sequence data has revealed insights into biological systems that have otherwise proven difficult to study include: the symbiotic relationship between a microbial community and a gutless worm and a 4000 year old Neanderthal. For the worm we used a metagenomic approach to describe the symbiotic microbial community in this organism lacking a mouth, gut, and nephridia. Through shotgun sequencing we were able to assemble the four predominant symbionts. Metabolic pathway reconstruction revealed the mechanisms by which the worm was able to outsource waste and energy management to the associated microbes. In the second study we describe the deep sequencing and analysis of genomic DNA extracted from bones of 40,000 year old Neanderthal. These studies, demonstrating that analytically significant quantities of ancient genomic sequence can be obtained, revealed the evolutionary relationship between Neanderthals and Homo sapiens.

Diversity-Generating Retroelements

Jeff F. Miller
University of California Los Angeles

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 genetic element that combines the basic retroelement life cycle of transcription, reverse transcription and integration with site-directed, nucleotide-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. Based on patterns of marker transfer in response to variant selective pressures, we propose that a TR reverse transcript is integrated into VR as a single non-coding strand and then partially converted to the parental VR sequence. This allows the diversity-generating system to minimize variability to the subset of bases under selection, and provides an opportunity to maximize receptor affinity through iterative rounds of optimization. Using the *Bordetella* phage cassette as a signature, we have identified numerous related elements in diverse bacterial genomes, including *Vibrio*, *Bacteroides*, *Treponema*, *Bifidobacterium*, *Trichodesmium* and *Nostoc* species. These comprise a new family of retroelements, designated "Diversity-Generating Retroelements," with the potential to confer powerful selective advantages to their host genomes.

Approaches for Developing Hydrogen Photoproducing Organisms and Processes

Michael Seibert
National Renewable Energy Laboratory, Golden, CO 80401

Several years ago we discovered that the application of a nutrient stress (removal of sulfate) with the green alga, *Chlamydomonas reinhardtii*, resulted in the direct coupling of photosynthetic water oxidation to hydrogenase-linked H_2 production for a period of 4 days. The mechanism involves the co-occurrence of oxygenic photosynthesis, anaerobic fermentation, and respiration when photosystem II O_2-evolution activity is significantly down-regulated. Hydrogen photoproduction was first demonstrated as an anaerobic batch process, but process improvements have included (a) optimization of the culture and light conditions, (b) conversion to a continuous process with H_2 production maintained for up to 6 months, (c) improvement in continuity as well as efficiency by immobilizing the algae on glass fibers, and (d) development of a new immobilization technique using low cost materials, suitable for scale-up. While helpful in learning how to produce H_2 in different types of photobioreactor systems, only a fraction of the photosynthetic potential of the alga can be used when sulfate is missing. In order to utilize the maximum photosynthetic potential of the algae, we must eliminate the sulfur stress, address the O_2-sensitivity of [FeFe]-hydrogenases, and operate the organism under aerobic conditions.

To this end we have developed Molecular Dynamics computational methods for simulating H_2 and O_2 gas diffusion in [FeFe]-hydrogenases. These studies have identified two well-defined pathways by which O_2 can access and inactivate the catalytic site, and many more pathways for H_2 to escape the protein. A strategy to protect the catalytic site from O_2 is to employ site-directed mutagenesis to restrict access of O_2 along the two pathways. To identify candidate amino acid residues for mutagenesis, we generated Potential Mean Force maps, which plot the free energy for O_2 placed at positions along the two pathways. These show regions of low (cavities) and high (barriers) energy, and provide us with the identity of potential specific residues to mutate. Site-directed mutants have been generated, the mutated proteins expressed in *E. coli* using an [FeFe]-hydrogenase expression system developed at NREL, and the initial results are encouraging.

Futhermore, we have started to develop a better understanding of the factors that promote H_2 production in algae by global transcript profiling and qPCR methods. These demonstrate significant changes in the transcript levels of many genes associated with photosynthesis, electron transport, proton transport, metabolism, fermentation, translational regulation, posttranslational modification, transcriptional regulation, stress response, signal transduction, as well as a variety of other physiological processes when H_2 is produced. The work is elucidating pathways used by *C. reinhardtii* during anaerobiosis and will provide insights into how mutants, altered in normal H_2 metabolism, acclimate to H_2-production conditions. More detailed knowledge of the metabolic and regulatory context that facilitates H_2 production will be necessary to understand and also ultimately eliminate current limitations in H_2-production yields.

PW King, K Kim, S Kosourov, S. Smolinski, P-C Maness, and ML Ghirardi (NREL); A. Dubini, L Nagy, J Meuser, and MC Posewitz (CSM); J Cohen and K Schulten (U. Ill); AS Federov and AA Tsygankov (RAS, Russia); J Gosse and M. Flickinger (U. Minn.); and F Mus and A Grossman (Carnegie Inst.) have all been intimately involved in this research. Sponsorship by the U.S. DOE Office of Science; the HFC&IT Program; and the AFOSR is greatly appreciated.

Engineering microbes for production of affordable anti-malarial drugs

Jay D. Keasling
Departments of Chemical Engineering & Bioengineering
University of California

Physical Biosciences Division
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

Malaria infects 300–500 million people and causes 1-2 million deaths each year, primarily children in Africa and Asia. More than half of the deaths occur among the poorest 20% of the world's population. One of the principal obstacles to addressing this global health threat is a lack of effective, affordable drugs. The chloroquine-based drugs that were used widely in the past have lost effectiveness because the *Plasmodium* parasite which causes malaria has become resistant to them. The faster-acting, more effective artemisinin-based drugs — as currently produced from plant sources — are too expensive for large-scale use in the countries where they are needed most.

We have metabolically engineered *E. coli* to produce high levels of mono-, sesqui-, and diterpenes, most notably the sesquiterpene precursor to artemisinin, amorphadiene. The result of these studies is an *E. coli* host capable of producing 1,000,000-fold higher levels of amorphadiene than the strains and expression systems that had been available previously. The engineered strain contains a heterologous mevalonate-based terpene biosynthetic pathway and an amorphadiene cyclase gene resynthesized with the *E. coli* codon usage. Recently, we cloned the final steps in the artemisinin biosynthetic pathway and engineered yeast to produce artemisinic acid at high levels. The development of this technology will eventually reduce the cost of artemisinin-based combination therapies significantly below their current price.

The reconstruction of the TRN in E. coli and its plasticity with adaptation

BO Palsson, BK Cho, E. Knight, A. Joyce, C. Herring, K. Applebee, T. Patel, A. Ragunathan

Department of Bioengineering
University of California--San Diego

This talk will describe the systematic reconstruction of the transcriptional regulatory network (TRN) through the use of literature data, and new data from dual perturbation experiments that generate ChIP-chip and expression profiling. The data suggests the existence of new ORFs and active regulation of sRNA molecules. Re-sequencing following laboratory adaptation to maximal growth rates on glycerol and lactate show systematic changes in cAMP and ppGpp regulation through mutations in TFs, RNAP and metabolic enzymes. The use of this data for genome-scale reconstruction and in silico model building will be discussed

From genomes to designed genomes: The *E. coli* reduction experience.

György Pósfai[1,2], Guy Plunkett III[2,3,4], Tamás Fehér[1], David Frisch[2,4], Günther M. Keil[5], Kinga Umenhoffer[1], Vitaliy Kolisnychenko[1], Buffy Stahl[2], Monika de Arruda[2], Valerie Burland[2,4], Shamik S. Sharma[6], Sarah W. Harcum[7], and Frederick R. Blattner[1,2]

[1]Institute of Biochemistry, Biological Research Center, Szeged, Hungary
[2]Scarab Genomics LLC, Madison, WI, USA
[3]Department of Genetics, University of Wisconsin, Madison, WI, USA
[4]Genome Center of Wisconsin, University of Wisconsin, Madison, WI, USA
[5]Federal Research Centre for Virus Diseases of Animals, Institute of Molecular Biology, Friedrich-Loeffler Institutes, Greifswald-Insel Riems, Germany
[6]Department of Chemical & Biomolecular Engineering, Clemson University, Clemson, SC, USA
[7]Department of Bioengineering, Clemson University, Clemson, SC, USA

As one of the best understood and thoroughly analyzed organisms, *Escherichia coli* K-12 is the platform of choice for genetic, biochemical, and metabolic simulation research. It is also used commercially, for the production of metabolites such as amino acids and proteins of therapeutic or commercial interest, and for the production of DNA for gene therapy, DNA vaccines, and iRNA. But because *E. coli* evolved in both animal intestines and the environment, parts of its genome are unnecessary — possibly even counterproductive — for some applications. In addition, mobile DNA elements, disseminated throughout the genome, mediate recombination events such as transposition and horizontal gene transfer.

Proposing that a reduced genome might improve metabolic efficiency and decrease the redundancy among *E. coli* genes and regulatory circuits, we have used synthetic biology to trim the *E. coli* K-12 genome by making a series of planned, precise deletions. The multiple-deletion series (MDS) strains, with net genome reductions of 15% or more, were designed by identifying non-essential genes and sequences for elimination, including recombinogenic or mobile DNA and cryptic virulence genes, while preserving good growth profiles and protein production. By means of a rational design strategy we avoided a loss of robustness that would result from more extensive deletions or an attempt to construct a minimal genome. We have constructed genetically stable "tabula rasa" strains with robust metabolic performance, to which genes for practical applications may be added.

Genome reduction also led to unanticipated beneficial properties: high electroporation efficiency and accurate propagation of recombinant genes and plasmids that were unstable in other strains. Eradication of stress-induced transposition has stabilized the MDS genomes and provides some of the new properties.

Use of Group II Introns for Large Scale Engineering of Chromosomes

Greg Davis*, Melissa Spears, Erik Eastlund, Kevin Kayser.

Sigma-Aldrich Corporation, Biotechnology R&D, Genomics, 2909 Laclede Ave., St. Louis, MO, 63103
gdavis@sial.com.

The TargeTron™ Gene Knockout System is a novel prokaryotic functional genomics tool based on the LI.LtrB group II intron from *Lactococcus lactis*. Like DNA transposons, mobile group II introns can inactivate genes by insertion, however, recent advances in group II intron research have enabled insertion to be site-specific. A key step in the mobility of group II introns is the formation of an RNA-protein complex (RNP). A somewhat accurate analogy is that the RNP acts like a programmable restriction enzyme, with the added activity of inserting the RNA component into the top strand of a cleavage site. Since site-specificity is primarily dictated by base pairing using the RNA component of the RNP, this allows re-targeting of the RNP by a rapid and routine PCR mutagenesis step. The insertion of group II introns is minimally dependent on host factors, making them applicable to a broad range of bacteria. To date, the TargeTron system has successfully disrupted genes in *Escherichia coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Clostridium perfringens*, *Shigella flexneri*, *Salmonella typhimurium*, *Serratia marcescens*, and *Lactococcus lactis*.

Many re-targeted introns are so efficient that selection is not required, allowing screening for insertional mutants by colony PCR. This eliminates the need to remove selection markers and expedites the creation of multiple knockouts. In addition to knockouts, knock-ins are also possible since heterologous DNA can be cloned into the intron and taken to specific genomic sites by user designed introns. This feature was recently exploited to deliver Cre-*loxP* recombinase sites to specific locations within the *E. coli* chromosome allowing rapid deletion of the *lac* operon as well as an extended 34 kb deletion encompassing 28 genes. With the help of new host-specific intron expression vectors (such as those recently validated for *S. aureus* and *C. perfringens*), this intron-recombinase method will provide a novel and efficient tool for large scale chromosomal engineering in bacterial hosts not easily manipulated by homologous recombination.

SCIENTIFIC PROGRAM ORGANIZERS

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

Dr. Frederick Blattner
University of Wisconsin-Madison

Dr. Jonathan Eisen
University of California, Davis

Dr. Jizhong Zhou
University of Oklahoma

Dr. Monica Riley
Woods Hole

Dr. Elisabeth Raleigh
New England Biolabs

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

Sigma-Aldrich Corporation

New England Biolabs

Scarab Genomics, LLC

DNASTAR, INC.

OpGen Technologies

CONTACT NUMBER

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

SCIENTIFIC PROGRAM

SUNDAY, SEPTEMBER 24

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** Edward M. Rubin DOE Joint Genome Institue, Walnut Creek, CA **"The JGI: Metagenomics and Beyond"**
9:00 pm	Reception (Iris Room)

MONDAY, SEPTEMBER 25

7:45-8:30 am	Breakfast (Dining Room)
Opening Session	(Pineview Room)
8:45-9:00 am	Introduction/Announcements
Session I	**Syngenomics: Pathways/Modeling/Synthetic Biology I**
9:00-9:30 am	Jay Keasling – University of California, Berkeley, CA **" Engineering *E. coli* to Produce New Pathways"**
9:30-10:00 am	Michael Seibert National Renewable Energy Laboratory, Golden, CO **" Approaches to Developing Biological H(2)-Photoproducing Organisms and Processes"**
10:00-10:30 am	Greg Davis Sigma-Aldrich, St. Lous, MO **"Use of Group II Introns for Large Scale Engineering of Chromosomes"**

KOFS: Knockout Fusion Screens for the *E. coli* KO Collection

Jeffrey H. Miller
Department of Microbiology, Immunology, and Molecular Genetics, and
The Molecular Biology Institute, University of California, Los Angeles, CA

In collaboration with Dr. Barry Wanner's group at Purdue University, we have constructed and employed transcriptional fusions to screen the E. coli complete knockout collection that consists of close to 4,000 deletion strains. We have constructed 20 fusions that place the lacZ gene under the control of different promoters involved in DNA replication, repair, and recombination that are carried on conjugatable plasmids. We have developed high throughput screening methods that allow us to easily screen the entire collection with each fusion, to look for deletion/KO strains that affect the control of these genes. The data from this type of approach complements that from microarrays, in that here we are looking at a selected set of genes (the recombination and repair genes) and monitoring their response to 4,000 different conditions (the strains with each gene, in series, deleted).

Large scale biochemical approaches to annotate genomes

Aled Edwards, University of Toronto

The biochemical and structural biology communities have recently turned their attention to the study of proteins on a larger scale than previously imagined possible. Many international projects, e.g. NIH-funded Protein Structure Initiative, are devoted to characterize proteins whose structures could not have been predicted using sequence-based methods. In the pilot phase of the PSI project between 2000 and 2005, the structures of more than 1000 such proteins were determined. Already 336 new structures have been determined in the first year of the second phase of the PSI, at a cost/structure far lower than could have been achieved in academia. These structures were selected to have broad leverage; on average more than a hundred new sequences can be modeled based on each PSI-determined structure. Other efforts, such as one in Toronto, screen the purified proteins for catalytic activity. Using a battery of relatively simple assays and the stocks of purified proteins, more than 250 new enzymes have been characterized to date with a team of 4 people. Larger scale implementation of these biochemical screens will doubtless be a cost-effective way to provide experimental annotation to new sequences. Going forward, it would make sense not only for the genome community to become more familiar with these various protein-based efforts, but also to play a proactive role in advising the experimentalists.

Routine, One Day Whole Genome Bacterial Sequencing

John H. Leamon
454 Life Sciences, Branford, CT

454 Life Sciences has developed a sequencing platform that allows the simultaneous sequencing of 200,000-400,000 individual target sequences in a single five hour run (Nature, 2005 Sep 15; 437, p376-80). The technology uses emulsion-based instant clonal amplification of target DNA onto 28-micron beads, followed by deposition into the wells of a chemically etched 50-micron pitch fiber-optic faceplate. Using sequencing-by-synthesis and detection of the released pyrophosphate, in excess of 100 bases of sequence information can be acquired per bead.

With this throughput, projects such as whole genome shotgun sequencing and assembly of microbial genomes can be completed by a single individual within a week. To accomplish assembly of the contigs we have developed a facile paired end library that works on the 454 system. By combining 100 bp reads and paired ends we de novo assembled E.coli into 140 contigs collapsed into 20 multi-contig-scaffolds (and 8 contigs) with 99.998% accuracy.

Additionally, the 454 sequencing system's throughput and lack of bias permits sequencing of complex DNA populations in unprecedented depth and detail. This has enabled 16S rDNA-based microbial community profiling (BMC Genomics, 2006,7:57), identification of pathogenic features in viral strains (Infection and Immunity, 2006, 74:8, p4694-4707), detect the single mutation responsible for evolution of a novel social behavior in the bacterium *Myxococcus xanthus* (Nature, 2006, May 18; 441,p310-314).

The 454 system is highly scalable and with the potential to increase throughput by employing higher bead densities and smaller beads per run. Furthermore, the read length can be extended beyond the current 100 bp reads and data from longer reads will be discussed.

Development of New Methods for Comparative Sequence Analysis

Christiane Honisch
Sequenom, Inc., San Diego, CA

Large-scale genome DNA-sequencing projects have reshaped microbiology by providing a rapidly increasing number of genome reference sequences. Based on this information a majority of sequencing applications in genomics will be focused on comparative sequencing with important applications for the characterization of microbes in the clinical, industrial and biodefense environment.

Studies typically involve hundreds to thousands of organisms and require standardized protocols and automated, high-throughput technologies. These requirements are met the MassARRAY System, which has become one of the leading technologies for high-throughput analysis and high fidelity measurements of nucleic acid variations. New developments of MALDI-TOF MS based comparative sequence analysis provide a tool, which is ideally suited for large-scale microbial testing on multiple genomic regions.
The application combines the sensitivity of PCR, *in vitro* transcription and base specific-cleavage of a genomic region with the accuracy and reproducibility of MALDI-TOF MS for detection and resolution of resulting nucleotide compomer mixtures. Mass signal pattern of compomer masses are acquired representing reproducible specific mass spectrometric fingerprints of the sample and region. Algorithms compare the detected list of experimental compomer masses with a calculated list of compomer masses from a set of reference sequences in a database. Peak pattern changes reflect microheterogeneities and are utilized to detect, identify and localize single base pair sequence changes.

Validation of MALDI-TOF MS comparative sequence analysis has shown concordance of accuracy and throughput with dideoxy sequencing in 16S rDNA based typing and multi-locus sequence typing. It has successfully been applied in the tracking of genomic alternations associated with adaptive evolution and environmental change and in the development of new informative marker sets for microbial identification.

Evolutionary Profiles from the QR Factorization of Multiple Sequence Alignments

Anurag Sethi, Patrick O'Donoghue, Jonathan Montoya, Elijah Roberts, Evan Rosenfeld, and Zaida Luthey-Schulten
Department of Chemistry, University of Illinois at Urbana-Champaign, Illinois.

We present a new algorithm to generate complete evolutionary profiles that represent the topology of the molecular phylogenetic tree of homologous group of proteins or nucleic acids. The multidimensional Structure QR and Sequence QR algorithms remove redundancy from a multiple alignment and order the structures or sequences by increasing linear dependence, resulting in a minimal set of sequences and/or structures that spans the evolutionary space of the homologous group. We observe a general trend that these smaller, more evolutionarily balanced profiles have comparable, and in many cases better performance in database searches than conventional profiles containing hundreds of sequences, constructed in an iterative and computationally intensive procedure. We used this approach to identify a putative class II cysteinyl-tRNA synthetase (CysRS) in methanogenic archaea that eluded previous annotation studies [1]. This enzyme was subsequently shown to be involved in an indirect mechanism for aminoacylation of cysteine that requires two enzymes [2]. The enzyme, now known as SepRS, first charges $tRNA^{Cys}$ with O-phosphoserine (Sep), a precursor of the cognate amino acid, and Sep-tRNA:Cys-tRNA synthase (SepCysS) converts Sep to Cys in a tRNA dependent reaction. Using both sequence and structure-based phylogenetic analysis of proteins involved in the direct and the indirect pathways of cysteine aminoacylation, we showed that both these pathways existed at the last universal common ancestral state (LUCAS) [3]. We are currently applying Sequence and Structure QR algorithms to study the coevolution of ribosomal RNA and proteins. We also suggest how the algorithms can guide genome sequencing projects in their selection of the next organism with the "newest" information.

[1] Sethi, A., O'Donoghue, P, & Luthey-Schulten, Z. (2005) *Proc. Natl. Acad. Sci. USA,* **102**, 4045–4050.
[2] Sauerwald, A., Zhu, W., Major, T. A., Roy, H., Palioura, S., Jahn, D., Whitman, W., III, J. R. Y., Ibba, M, & S^ll, D. (2005) *Science,* **307**, 1969–1972.
[3] O'Donoghue, P., Sethi, A., Woese, C. R, & Luthey-Schulten, Z. (2005) *Proc. Natl. Acad. Sci. USA,* **102**, 19003–19008.

Ecoli Genome wiki

EcoliHub, K-12 Information Resources – under construction

Barry L. Wanner[1], Walid Aref[1], Michael Gribskov[1], Daisuke Kihara[1], Tyrrell Conway[2], James C.Hu[3], Peter D. Karp[4], and Hirotada Mori[5]

[1]Purdue University, West Lafayette, IN; [2]University of Oklahoma, Norman, OK; [3]TAMU, College Station, TX; SRI Inc., Menlo Park, CA; [5]Nara Institute of Science and Technology 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

Hosting open source Not clear how used

More basic biochemistry, molecular biology, and genetics are arguably known of *Escherichia coli* K-12 than of any other organism. The sheer wealth of this information rangingfrom physiology responses in the environment down to the atomic level makes it exceedingly difficult, even for experts, to grasp and utilize it fully. While many databases, websites, and computational tools have been developed to enable access to and interrogation of these different kinds of biological data, none is comprehensive and gaps exist. The quality and quantity of these resources also vary widely both within and between them. Many have redundant, conflicting, and out-of-date information, impacting their reliability, especially when the data cannot be traced back to sources for resolution. In addition, the information systems are disorganized, overwhelmed by existing knowledge, and unprepared for handling new information from highthroughput experimentation (genome-wide transcription, translation, and metabolite profiling, large-scale structural biology, enhanced imaging of living cells, and the like).

We are developing an integrated "one-stop-shopping" *E. coli* community information resource, called "EcoliHub," to permit full use of our existing knowledge and to enable new discoveries leading to deeper understanding of life processes. Moreover, since most cellular processes are universal, these tools and the advances they allow will have important impact on human health, through their application to pathogenic bacteria, especially enteropathogenic *E. coli*, *Shigella*, and *Salmonella* species.

EcoliHub is being designed for seamless and transparent bi-directional connections with cooperating and interoperable *E. coli* informatics databases (EcoCyc and Genobase), as well as with major biological database like ERIC, NCBI's Entrez Gene, UniProt, and KEGG. It will provide an interactive chat room to facilitate community-driven developments that meet the needs of the users. EcoliHub will offer on-the-fly computational tools for molecular andstructural analysis, like 3D-PSSM, BLAST/PSI-BLAST, BLOCKS, Pfam, and PredictProtein. It will maintain a depository of related *E. coli* informatics from numerous specialized databases, like Brenda, GeneMark, GenProtEC, GO, LipoP, MEROPS, PORES, RegulonDB, Superfamily, and TransportDB. Schemas will be designed to annotate and track changes to biological entities within them. A module for biosamples (strains, plasmids, etc.) will be developed to facilitate their acquisition from national repositories (like the Coli Genetic Stock Center) and other sources. A web-based architecture will be implemented for access to intuitive software for data mining and high throughput functional studies.

Supported by NIH U24 GM077905 activation date June 1, 2006

On using genomics to combat agents of infectious diseases

Eric Eisenstadt
The Institute for Genomic Research

The Institute for Genomic Research is a not-for-profit research institute established in 1992 with activities in structural, functional, and comparative analysis of genomes and gene products in viruses, bacteria, archaea, and eukaryotes (both plant and animal). I will provide a brief overview of infectious disease research and development at TIGR, the services we provide the infectious disease research community via NIAID supported resource centers, and examples of current efforts involving the use of new genome sequencing technologies. To illustrate the practical implications and applications of genome sequence information, I will also present some examples of recent work on the sequencing of multiple isolates of Group B Streptococcus and the full-length genome sequence analysis of hundreds of influenza A genomes, both human and avian. Finally, I will present an outline of an idea for using TIGR's high-throughput functional genomics platforms coupled to a specific modern tissue engineering platform to create an automated in vitro pipeline to identify potential human vaccine candidates.

Genomic variation in host-pathogen interactions

Julian Parkhill

The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Individual clones within a genomic shotgun are samples from a population, and any variation in the population will therefore be represented in the shotgun sequences. Bacterial pathogens often encode directed variation mechanisms that enable diversity generation, and even clonal growths of such bacteria can exhibit significant variation produced from such systems. Looking at variation within shotgun sequences allows the elucidation of the extent of known variation mechanisms, and the inference of novel mechanisms. These points will be illustrated with examples from several organisms including *Campylobacter*, *Bacteroides* and *Tropheryma*.

Recombination

Phase variation

All sequences is metagenomics

4Mb genome

700,000 reads

10^6 g DNA

Functional genomic studies of Vibrio cholerae essential genes

David M. Raskin* and John J. Mekalanos. Harvard Medical School.

Almost half of the genes of *Vibrio cholerae* have no known function. We are interested in identifying the essential genes of unknown function, and characterizing their function. We used a non-saturating transposon mutagenesis to identify essential genes, those required for growth on rich media, and identified 100 genes as essential including 20 genes of unknown function. We are currently determining the complete set of essential genes by integrating an inducible promoter upstream of each gene. This method has given us a set of conditional mutants for each of these genes. Lack of growth in the absence of the inducer indicates essentiality. We then use these conditional mutants for functional genomic studies of the genes of unknown function. We are using microarrays to determine transcriptional profiles of strains depleted for the essential gene product. These transcriptional profiles have been used to determine how cells respond to loss of the essential gene product while still viable. These responses often included stress responses or changes in expression of sets of genes involved in a specific function. This data, when coupled with other data such as protein-protein interactions, localization, phenotypes of mutants, can be used to effectively characterize the function of essential genes of unknown function.

Genome sequencing projects at TIGR involving pathogens and biothreat agents

Jacques Ravel
The Institute for Genomic Research

Under the umbrella of the Microbial Sequencing Center, a NIAID funded project, TIGR has undertaken the sequencing of the genomes a large number of microbial pathogens and biothreat agents. Of interest are the genome sequences of eight additional *Yersinia pestis* strains, which are being sequenced to complement the several *Y. pestis* genomes known to be finished or at a draft stage. In addition, the genome sequence of an isolate of *Yersinia pseudotuberculosis* responsible for Far East Scarlet-like Fever is shedding light into the role of genomic plasticity and pathogenicity. While *Yersinia pestis* is thought to be niche specific, an interesting link has been discovered between a *Salmonella* multi-drug resistance isolates and a *Yersinia pestis* isolate from Madagascar, whereas both isolates carry an almost identical plasmid encoding for several antibiotic resistance genes. These projects and those of *Salmonella enterica*, *Bacillus anthracis* and *Bacillus cereus* among others will be discussed.

Genetic Analysis of Virulence Factors Allowing Streptococcus and Staphylococcus to Escape Phagocyte Killing

Victor Nizet, M.D.
Division of Pharmacology & Drug Discovery
University of California, San Diego School of Medicine
Skaggs School of Pharmacy & Pharmaceutical Sciences

Staphylococcus aureus (SA) and Group A Streptococcus (GAS) are Gram-positive bacteria each associated with a wide spectrum of invasive infections in humans. To better understand disease pathogenesis, we have coupled precise, in-frame allelic xchange mutagenesis with heterologous gene expression to establish "molecular Koch's postulates" for individual SA or GAS virulence factors. This approach provides living reagents for subsequent analysis in tissue culture and small animal models of infection. For example, a key element of innate host defense against infection is the bactericidal capacity of specialized leukocytes known as neutrophils. Recently, we have applied molecular genetic approaches to demonstrate (a) how SA uses its eponymous golden pigment to avoid oxidant-based neutrophil intracellular killing, and (b) how invasive strains of GAS express a secreted DNAse to avoid escape neutrophil extracellular traps (NETs). Definition of these virulence factors identify novel therapeutic strategies, which aim not to kill bacteria directly, but instead to render the pathogens susceptible to our normal innate defense mechanisms.

Genome mining in Streptomyces coelicolor

Gregory L. Challis, Department of Chemistry, University of Warwick,
Coventry CV4 7AL, UK

Analyses of numerous microbial genome sequences have identified gene clusters encoding enzymes typically involved in the biosynthesis of complex natural products that are not associated with production of known metabolites. One example is Streptomyces coelicolor, where numerous "cryptic" or "orphan" gene clusters encoding nonribosomal peptide synthetase, polyketide synthase, sesquiterpene synthase and other biosynthetic enzymes are present in the genome. In this lecture I will discuss the genome mining approaches investigated by us and othersfor identifying the products of such gene clusters using Streptomyces coelicolor as an example.

Strain Resolved Proteomics Reveals that Recombination Shapes the Genomes of Acidophilic Bacteria in Natural Consortia

*Nathan C. VerBerkmoes[1], I. Lo[2], V. Denef[2], M. Shah[1], G.W. Tyson[2], E.E. Allen[2], R.J. Ram[2], M.P. Thelen[3], R.L. Hettich[1], and J.F. Banfield[2]; 1) Oak Ridge National Laboratory 2) University of California, Berkeley, 3) Lawrence Livermore National Laboratory

Microbial communities play key roles in the Earth's biogeochemical cycles. Our knowledge of the structure and function of these communities is limited because analyses of microbial physiology and genetics have been largely confined to isolates grown in laboratories. Recent acquisition of genomic data directly from natural samples has begun to reveal the genetic potential of communities (Tyson, Nature 2004) and environments (Venter, Science 2004). The ability to obtain whole or partial genome sequences from microbial community samples has opened up the door for microbial community proteomics. We have developed and applied a combined proteogenomic approach using genomics and mass spectrometry-based proteomic methods (Ram, Science 2005). We evaluated gene expression, identified key activities, and examined partitioning of metabolic functions in a natural acid mine drainage (AMD) microbial biofilm community through the identification of over 2,000 proteins from a natural environment. Recently, we extended this approach to peptide resolved proteomics to distinguish, genome-wide, strain-specific expressed protein variants so as to resolve genetic evidence for recombination in the dominant member of a natural multi-species biofilm. MS-based proteomic discrimination relied upon cultivation-independent reconstruction of near-complete genomic datasets for two related *Leptospirillum* group II bacteria from the same system. Variants were detected by peptide resolved proteomics suggesting recombination of chromosomal regions tens to hundreds of kilobases in length in a third unique environmental sample.

* Recombination

Spatial scaling of microbial biodiversity

Jessica Green
School of Natural Sciences, University of California, Merced, USA

A central goal in ecology is to understand the spatial scaling of biodiversity. Patterns in the spatial distribution of organisms provide important clues about the underlying mechanisms that structure ecological communities and are central to setting conservation priorities. Although microorganisms comprise much of Earth's biodiversity, little is known about their biodiversity scaling relationships relative to what is known regarding plants and animals. Here, I discuss current knowledge of microbial diversity at local and global scales. I focus on three spatial patterns: the distance-decay relationship (how community composition changes with geographical distance), the taxa-area relationship, and the local:global taxa richness ratio. Recent empirical analyses of these patterns for microorganisms suggest that there are biodiversity scaling rules common to all forms of life.

Metagenomic Analysis of Viral Diversity in the Sea

Curtis A. Suttle
University of British Columbia, Vancuver, BC

There are an estimated 4×10^{30} viruses in the world's oceans, which if stretched end-to-end would easily transverse our galaxy. They are the least known and most genetically diverse life forms in the sea. Pyrosequencing (454) on 184 virus communities from the Arctic Ocean (n=57), British Columbia Coast (n=85), Gulf of Mexico (n=41), and Bermuda (n=1) showed that the most geographically distant samples were also the most genetically distant. Most sequences (97.5%) were not similar to those in the NCBI database (E<0.001), and the most frequent hits were to relatively few phage. In the Arctic, ~1% of the hits had significant similarity and of these, >30% were to prophages. The top hits from other locations were to cyanophages. Bermuda had the highest number of hits to known sequences (>6%) with ~50% of these to cyanophages, and ~25% to ssDNA viruses (Microviridae). Metagenomic analysis was also done on two marine RNA virus communities. Most sequence fragments had no significant similarity (tblastx e value < 0.001) to sequences in the NCBI database. Assembly and phylogenetic analyses suggest that one community was dominated by a diverse group of picorna-like viruses. In contrast, picorna-like viruses were undetected in the other community, but viruses that are distant relatives of those that infect higher plants were present. Marine RNA virus communities appear to be significantly less diverse than dsDNA marine phage communities although they have a similar population structure. The lack of genome richness, and the small average genome size allowed the complete genome sequences of three viruses to be reassembled. These results emphasize that we are only beginning to describe the genetic diversity in the virosphere.

Nothing new.
All say 454 is as good as Sanger. BS.

Genomics of Mutualistic Symbioses and the Origin of Novelty

Jonathan A. Eisen
University of California Davis

One of the most straightforward, and fundamentally important, mechanism by which organisms can acquire new functions is by engaging in mutualistic symbioses with other species. The best studied such symbioses are those involving the mitochondria and plastids of eukaryotes. However, it is difficult to learn about how such symbioses originated by studies of organelles since these symbioses started billions of years ago. I will discuss studies from my group in which we have used genome sequencing to characterize more recently evolved mutualistic symbioses involving a diverse array of host and symbiont species.

Experimental evolution: phenotypic and genetic changes during a long-term experiment with *Escherichia coli*

Richard E. Lenski

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824-4320, USA

Twelve initially identical populations of *E. coli* B have been propagated in a simple environment for more than 15 years. The goals of the study are to examine the reproducibility of evolutionary changes and better understand the coupling between phenotypic and genomic evolution. We have examined the dynamics of adaptation as measured by competitive fitness [1], documented increasing ecological specialization of the evolving organisms over time [2], observed the emergence of mutator phenotypes [3], and so on. We have employed various approaches, from tracking IS elements [4] to examining changes in global expression profiles [5], to identify many of the mutations responsible for these changes, including some that affect global regulatory networks [5-7]. Collaborative efforts have begun to obtain whole-genome sequences in order to identify all of the mutations substituted in some of the evolving populations.

1. Lenski, R.E., and Travisano, M. (1994) Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *PNAS* 91:6808-6814.

2. Cooper, V.S., and Lenski, R.E. (2000) The population genetics of ecological specialization in evolving *E. coli* populations. *Nature* 407:736-739.

3. Sniegowski, P.D., Gerrish P.J., and Lenski, R.E. (1997) Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* 387:703-705.

4. Schneider, D., Duperchy, E., Coursange, E., Lenski, R.E., and Blot, M. (2000) Long-term experimental evolution in *Escherichia coli*. IX. Characterization of IS-mediated mutations and rearrangements. *Genetics* 156:477-488.

5. Cooper, T.F., Rozen D.E., Lenski, R.E. (2003) Parallel changes in gene expression after 20,000 generations of evolution in *E. coli*. *PNAS* 100:1072-1077.

6. Crozat, E., Philippe, N., Lenski, R.E., Geiselmann, J., and Schneider, D. (2005) Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169:523-532.

7. Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A., Lenski, R.E. (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *PNAS* 103:9107-9112.

*Now doing sequencing
Estimated 3 mutations / genome

Microbial Diversity in the Deep Sea and the Underexplored Rare Biosphere

Julie A. Huber, Hilary G. Morrison, David Mark Welch, Susan M. Huse, Phillip R. Neal, and Mitchell L. Sogin

Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA, USA 02543
E-mail: *jhuber@mbl.edu*; Telephone: 508-548-3705, x6616; Fax: 508-457-4727

The world's oceans are teeming with microscopic life forms. The staining of cells with DNA-binding dyes (DAPI and acridine orange) coupled with epifluorescence microscopy demonstrated that nominal cell densities exceed 10^5/ml of sea water. Extrapolations of these numbers predict that the oceans harbor 3.6 x 10^{29} microbial cells with cellular carbon of ~3 x 10^{17} grams. Given the enormous populations of microbes with seemingly unlimited metabolic diversity, the accumulation of mutations during the past 3.5 billion years should have led to very high levels of genetic diversity and phenotypic variation. By adopting a massively parallel 454 tag sequencing strategy, we show that bacterial communities of deep water masses of the North Atlantic and diffuse flow hydrothermal vents are one to two orders of magnitude more complex than previously reported for any microbial environment. A relatively small number of different populations dominate all samples, but thousands of low-abundance populations account for most of the observed phylogenetic diversity. This "rare biosphere" is very ancient and may represent a nearly inexhaustible source of genomic innovation. Members of the rare biosphere are highly divergent from each other and at different times in earth's history may have had a profound impact on shaping planetary processes.

* Some issues with this

small genome
odd organism
variation in the population
all occured in two regions of the chromosome

R. Green

Spatial patterns of biodiversity

"Publications, students, etc."

	Microbes	Plants and Animals
Pop Size	High	Low
Dispoersal	High	Low
Speciation and extinction	High	Low

Have you done DOUTR on Sargasso data for the V6 region?

Is there variaiton within taxa?

3% cutoff for whole molecule would give a very different cutoff for V6

Peg Riley

All that is interesting to study is in essence invisible to the eye

Is transfer cooon enough t bring species back together
Or do lineages diverge without limit

Microbes not only have a lot of sex, they have a lot of weird sex

Successful gene transfer most likely to be retained when selection is very strong

Bacterial phenotypes are CLEARLY discrete not continuous

I'm a TIGR wannabe

Ecoli database

Trying to get it up and running in October

"Not trying to say anything bad about anyone"

Esienstadt --- not a very interesting talk - mostly a TIGR ad

Julian Parkhill

Genetic variation in host pathogen interactions

Says he has been doing metagenomics all along - single organism metagenomics

Typical 4 Mb genome
Shotgun 70,000 reads
10-6 grams of DNA
approximate number of genomes to make library 2.2 x 10^8
Average number of clone sused - 3.5 x 10^4
Therefore all reads come from different cells

If there is variation in the popualtion you will capture it

For example, C. jejuni
- runs of Gs vary in numebr of copies
- phase variation
- slippage occurs incredibly frequently
- likely due to lack of MMR
- where are the phase variable genes? - most in surface

Another example
- shotgun assembly anomalies
- some due to phage indel polymoprhisms

Invertable promoters in B. fragilis

"the huamn guts are a real mileu of stuff"

<<<<Eisen comments - ths could lead to variation within a single cell
in its lifetime IF (a) repair or (b) polyploid

Examples - restruction modification system in B. fragilis

<rate differrs - can they estimate from sequencing>

T. whipplei

To her, what jmped out from comparative genomics was how similar the cores of genomes were

What explains the core vs. the flux

"Core genome hypothesis"

Rare transfer between close relatives
May include housekeeping
95% or more of the isolates should have it

Auxillary genes serve to adapt isolates to local niches

Makes predictions about mechanisms
- core genes will diverge over time and this will limit recombination
- should have lower diversity within vs. between taxa
- auxillary genes should maintain long term homogeneity

How test?

Need genomes of closely realted taxa that can coexist in time and space
Gordon E. coli islates
Picked seven housekpeeing genes
- multiple isolates from each taxa ALWAYS cluster

Compared multiple Ec and Se genomes

Used genome flux analysis for find auxillary genes
- core genes match Ec and Se classification
- auxillary do not

What is core in one species mayb eauxilary in another

COMPARATIVE EVOLUTIONARY GENOMICS CONSORTIUM

Proposes MLST can distinguish bacterial species
Says this can work for uncultured?

Sacee

Functional metagenomic screens

"They are good at eating a lot of snacks"

Universal donors versus recipients is important to udnerstand for heterologous expressions

SYntrophs - Rob Gunsalas

Lab Meeting 10/2/01

Tom Turner + Maggie

Bacillus subtilis Genome Diversity

Ashlee M. Earl[1*], M.J. Rosovitz[2], David A. Rasko[3], Richard Losick[4], Roberto Kolter[1] and Jacques Ravel[2]

[1]Harvard Medical School, Boston, MA, [2]The Institute for Genomic Research, Rockville, MD, [3]University of Texas Southwestern Medical Center at Dallas, Dallas, TX, [4]Harvard University, Biological Laboratories, Cambridge, MA.

Bacillus subtilis is a non-pathogenic spore-former that is also arguably one of the best known and most extensively studied Gram-positive bacteria. And while we know a great deal about *B. subtilis* at the molecular level (especially as it relates to strain Bsu168, the only fully sequenced representative), relatively little is known about this species' ecology and evolution. Using microarray-based comparative genomic hybridization (M-CGH) to examine 19 geographically distinct *B. subtilis* strains we have determined that there is considerable genetic heterogeneity among members of this species; nearly one-third of Bsu168-specific genes exhibited variability as measured by the microarray hybridization intensities. Variable loci included those encoding proteins involved in antibiotic production, cell wall synthesis, sporulation and germination suggesting a role for these genes in environmental adaptation. While M-CGH is a powerful technique for exploring whole genome diversity among bacterial species, there are clear limitations. For example, gene absence versus gene divergence is not always obvious and M-CGH is not able to forecast what genes, if any, a strain possesses that are not already found within the reference genome. In an effort to gain a more complete understanding of *B. subtilis* genome diversity we are sequencing the genomes of three additional *B. subtilis* and one phylogenetically close relative, *B. mojavensis*. Progress from this effort will be discussed. It is our hope that comparative analyses using these newly sequenced genomes will shed further light on the physiology, evolution and ecology of this important group of organisms as well as expand our understanding of gene flow and speciation among sympatric soil Bacilli.

Tu1C

A Genome-scale Metabolic Reconstruction of *E. coli* K-12 MG1655: A Curated Framework for Systems Analysis

A.M. Feist*, J.L. Reed & B.Ø. Palsson. Department of Bioengineering, University of California – San Diego, La Jolla, CA, U.S.A.

Network reconstruction has been a common denominator in systems biology. Taking advantage of the extensive characterization of *E. coli*, we reconstructed the metabolic network of *E. coli* K-12 MG1655 and examined its application for systems analysis. Specifically, we characterized the reconstruction, used the reconstruction for computations of phenotypes and determined its use in interpreting high-throughput experimental data. This reconstruction is based on previous work [1,2] and additional information from published studies. The reconstruction is fully compartmentalized into three distinct cellular compartments and includes 1254 of the predicted 4578 ORFs in *E. coli* (27%), over 2000 reactions, 1400 metabolites and thermodynamic estimates for metabolites and reactions. The final reconstruction was generated through a collaborative effort with EcoCyc (SRI International) and the Computational and Mathematical Biotechnology Laboratory (P.I. Vassily Hatzimanikatis). After converting the reconstruction to a computational model, we identified the important parameters needed for using the model as a predictive tool (using a constraint-based approach) and further performed a sensitivity analysis on these variables. Using the model, we analyzed the essential genes predicted under minimal media growth conditions [3]. This work represents a significant enhancement from the previous work because of its expanded coverage of metabolism, inclusion of reaction and metabolite thermodynamic data and characterization of the variables that significantly affect modeling simulations.

1. Edwards, J.S., and Palsson, B.Ø., (2000) The *Escherichia coli* MG1655 *in silico* Metabolic Genotype: Its Definition, Characteristics, and Capabilities. Proc. Nat'l Acad. Sci. USA, 97: 5528-5533.
2. Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.Ø., (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biology. 4(9): R54.1-R54.12.
3. Joyce, A.R., Reed, J.L., White, A., Edwards, R., Osterman, A., Baba, T., Lesley, S.A., Palsson, B.Ø., Agarwalla, S., (2006) Experimental and computational assessment of conditionally essential genes in *E. coli*. Journal of Bacteriology, To Appear.

Evolution of spectral diversity in nymphalid butterflies through gene duplication and positive selection

Francesca D. Frentiu, Marilou P. Sison-Mangus and Adriana D. Briscoe
Dept. of Ecology and Evolutionary Biology, UC Irvine, CA 92697.

Butterflies form an important group in terrestrial ecosystems because of their role in pollination. Many aspects of butterfly behaviour are mediated by visual cues but visual systems in this group remain poorly understood. Our aim was to characterize the spectral diversity of visual pigments in the long wavelength range and investigate their evolution in a butterfly family, the Nymphalidae. Physiological measurements indicated significant spectral diversity, with peak sensitivities in the range of 515 – 565 nm. We then cloned the long-wavelength (L) opsin genes encoding for the proteins that form part of the visual pigments. We found two instances of gene duplication in a set of 32 species, suggesting this process may be partly responsible for some of the spectral diversity detected. Phylogenetic relationships among 33 opsin gene sequences were reconstructed using maximum likelihood (ML) methods. Comparisons of the ratio of synonymous to non-synonymous substitutions using the opsin gene tree indicated several amino acid sites are under positive selection, which may be driving the spectral diversification observed. Homology models of nymphalid opsins constructed using the bovine rhodopsin crystal structure indicated that some of these sites are located in functionally important regions of the protein.

Selection for Efficiency of Transcription and Translation Are Correlated in Bacteria.

Jeff L. Froula and M. Pilar Francino* Evolutionary Genomics Department, DOE Joint Genome Institute

Because binding of RNAP to misplaced sites could compromise the efficiency of transcription, natural selection should regulate the distribution of DNA motifs capable of RNAP-binding across the genome. Here we analyze the distribution of the –10 promoter motifs that bind the σ^{70} subunit of RNAP in 42 bacterial genomes. We show that selection on these motifs operates in different directions in regulatory vs. nonfunctional sequences. Furthermore, the avoidance of –10 sites in nonfunctional DNA correlates with the number of tRNA genes in the genome, a parameter that tracks translational speed and accuracy. This indicates that genome-level properties affecting the efficiency of transcription and translation respond in an integrated manner to optimize gene expression. Finally, we show that selection is generating –10 sites in the coding sequences of many genomes, rather than eliminating them as in the nonfunctional DNA. Pauses induced by these sites could effect regulatory roles throughout the length of a transcriptional unit.

Comparative Sequence Analysis of Shewanella Methyl-Accepting Chemotaxis Proteins

Michael C Gilson* & Nicole T Perna. University of Wisconsin-Madison.

Chemotactic bacteria respond to a changing environment via a multimodular protein network that will sense and adapt to variable substrate concentrations. Chemosensory receptors on the cell surface and cytoplasm regulate the transduction of environmental signals to intracellular response regulators which, in turn, induce changes in motility—biasing motion toward carbon sources and electron acceptors and away from toxins or microenvironments which yield decreased productivity of the electron transport chain. In this study we address the metabolically diverse, environmental microbe, *Shewanella oneidensis* MR-1, whose chemotactic capabilities have recently come under increased scrutiny. Our study utilizes a bioinformatic approach that complements recent *in vitro* investigations into the substrate specificity of MR-1's chemosensory repertoire. We conducted a hidden markov model (HMM) search of the recently published MR-1 genome for putative methyl-accepting chemotaxis proteins (MCP). These putative MCP were identified by their highly-conserved, C-terminal signaling domain as represented in the HMM query profile. This search identified 25 putative MCP, in sharp contrast to the related gammaproteobacterium, *E. coli*, which possesses only five. A two part, knowledge base (KB) approach was used to further characterize these 25 sequences. Functional sites and domains were predicted using InterPro, an integrated profile database and search tool. For the second part of our KB approach the amino acid sequences were pruned of their C-terminal signaling domain, which promiscuously aligned to distantly related but functionally diverse proteins in the database. The truncated, N-terminal portions were searched against the NCBI non-redundant protein database and yielded high scoring matches predominantly to other *Shewanellae* but also to more distant genera. Our data support the hypothesis that this paralogous family of proteins expand and contract by duplication and deletion of genes, undergo horizontal transfer across diverse lineages, and likely recombine within individual genomes. Thus, the evolutionary history is complex, but unraveling it can provide insights into how this organism came to have this particular taxis repertoire, potentially revealing selection for this behavioral response that may be characteristic of this environment in general, or at least of this particular lineage.

ıental genome analysis of the *Alvinella pompejana* epibiont consortia.

ırzymski[1]*, Alison E. Murray[1], Mihailo Kaplarevic[2], Robert Feldman[3], ıbara J. Campbell[4], Guang R. Gao[2] and S. Craig Cary[4]. [1]Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, Nevada. [2]University of Delaware, Delaware Biotechnology Institute, Newark, DE. [3]SymBio Corp. Menlo Park, CA. [4]University of Delaware, College of Marine Sciences, Lewes, De.

A shotgun library created with DNA extracted from the predominantly ε-proteobacteria epibionts that live on the dorsal surface of *Alvinella pompejana,* a hydrothermal vent worm, was sequenced and produced 301,000 quality-filtered sequences. Using strict assembly parameters, these sequences assembled into 26,000 contigs and ~100,000 singletons. The sequences were annotated using FgenesB (Softberry) and a custom annotation pipeline. The dataset consisted of 103,000 predicted proteins and 484 SSU-rRNA genes and 918 tRNAs. There were at least 5 ε-proteobacteria SSU rRNA phytotypes (at 99% sequence identity). These results correspond to the diversity of recA and other functional genes in the dataset. To assess the diversity of predicted proteins in the epibiont community protein clustering was performed at 40% identity (CD-HIT). The 103,000 predicted proteins clustered into 36,221 clusters of which 27,583 were singletons. The high percentage of singletons agreed well with an analysis of COG functional categories; there were 1560 different COGs represented in the environmental genome but 85% of the predicted proteins did not fall into a functional category. This is high in comparison to the combined whale fall community genome (3 datasets; Tringe et al. 2005) of which 45% of the predicted proteins did not fall into a functional category. When eight whole ε-proteobacterial genomes were considered the same way, they had a combined total of 32% of the predicted proteins that were not assigned to a functional cateogory. An epibiont "core" genome that consisted of clusters of predicted proteins that were found at least 6 times (at 40% identity) in the dataset was defined. There were 2814 clusters that met this criterion. This core genome was rich in cell wall, membrane and envelope biosynthesis proteins - not surprising given the complex sheath that protects the bacteria on the back of the worm. In general, the COG distribution of this "core" genome corresponds well to the distribution of functional classes in the pooled ε-proteobacteria including very reduced representation of transcription COGs: 1.5% in "core" epibiont, 2.6% in ε-proteobacteria and 4.9% in all bacterial genomes sequenced thus far. Community genomes require a combination of data reduction (assembly, protein clustering) and comparative analysis of functional distributions (i.e. COGs) to facilitate analysis of these complex datasets. We utilized a unique combination of tools and a data pipeline to help discern unique characteristics of the unique microbial community.

Identifying Novel Autoinducer-Dependent Target Genes of the *Vibrio cholerae* Quorum Sensing Regulon

Brian K. Hammer & Bonnie L. Bassler, Department of Molecular Biology,
Princeton University, Princeton NJ 08544

Quorum sensing (QS) is a process by which bacteria alter gene expression in response to the accumulation of extracellular signaling molecules called autoinducers (AIs). In the human pathogen *Vibrio cholerae*, parallel QS systems converge to control expression of the master transcriptional activator HapR. HapR positively and negatively regulates the expression of many genes including those involved in biofilm formation and virulence. , To define the *V. cholerae* QS regulon, a plasmid library containing random DNA fragments fused to a luciferase transcriptional reporter was introduced into an AI-deficient *V. cholerae* strain, and the transformants were screened for promoters that displayed differential regulation in the presence or absence of AIs. 25 quorum sensing target genes were identified; 11 of these genes have been previously identified, whereas 14 genes were novel. One newly identified QS-repressed target gene, *rho*, is predicted to encode Rhomboid, a protein involved in production of an extracellular signal molecule required for development in both prokaryotes and metazoans. Two additional QS-repressed target genes (*vca0681* and *vca0895*) encode HD-GYP proteins with predicted phosphodiestase activity capable of cleaving the intracellular second messenger molecule 3',5'-cyclic diguanylic acid (c-di-GMP). Consistent with studies of other *V. cholerae* phosphodiesterases, overexpression of VCA0681 reduces polysaccharide expression and biofilm formation. One QS-activated target identified here (*vca0939*), encodes a GGDEF protein that synthesizes c-di-GMP. Surprisingly, *vca0939* expression is repressed by AIs, however, this occurs via a HapR-independent route suggesting that additional, unidentified signal transduction pathways exist in *V. cholerae* to couple QS to downstream gene expression.

The genome sequence of *Haloferax volcanii*, a model archaea

Hartman, A.[1], J. Badger[2], J. Maupin-Furlow[3], M. Pohlschroder[4], T. Lowe[5], C. Daniels[6], H.M. Khouri[2], J.Robinson[2], J. Eisen[2,7]

The 4.1 Mb genome of the extreme halophilic archaea, *Haloferax volcanii,* contains a 4 Mb chromosome and 4 plasmids, pHv1-4 (86, 6.4, 442, 690 kb, respectively.) *H. volcanii* is readily transformable and genetically tractable, thus making it an ideal model organism for the study of archaeal biology. Additionally, the comparative use of 4 other halophilic archaeal genomes enables broad phylogenomic contrasts and correlations. We have incorporated functional genomic and proteomic information to improve our auto-annotation and better define the post-translational proteome and the twin-arginine transport pathway uniquely utilized by the halophilic archaea. *H. volcanii* and other halophiles have exceptionally acidic proteomes to counteract the potentially denaturing effects of a high salt intracellular environment. The signature of this adaptation might be related to the high GC content of these organisms and perhaps partially explains how this biological adaptation occurred. *H. volcanii* is highly related to the recently sequenced *Haloquadratum walsbyi*. In fact, 66% of annotated ORFs are common to both organisms. While *H. walsbyi*'s genome is considered to be more streamlined and adapted to a specific ecological niche, *H. volcanii* appears to be a halophilic generalist with a wider degree of ecological flexibility.

1 Johns Hopkins University, 3400 N. Charles St, Baltimore, MD 21218
2 The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD, 20850
3 Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611-0700
4 Department of Biology, University of Pennsylvania, Philadelphia, PA 19104 USA
5 Biomolecular Engineering, University of California, 1156 High Street, Santa Cruz, CA 95064
6 Department of Microbiology, Ohio State University, 484 West 12th Ave., Columbus, Ohio USA 43210-1292
7 Genome Center, Dept. of Evolution and Ecology, Dept. of Medical Microbiology and Immunology, University of California, Davis, CA, 95616

Comparative Genomics with Optical Mapping: Identifying Genomic Differences in Clinical Isolates of Methicillin-Resistant *Staphylococcus aureus*

John Henkhaus*[1], Adam Briska[1], Emily Zentz[1], Sanjay Shukla[2]
[1]OpGen Technologies; [2]Molecular Microbiology Laboratory, Marshfield Clinic Research Foundation

Staphylococcus aureus is a major nosocomial and community-acquired pathogen whose rapidly evolving genome presents an important clinical challenge. The complete genome sequence from several staphylococcal isolates indicates that many of the genes involved in virulence are primarily carried on mobile genetic islands (GIs). Identification and eventual characterization of GIs and the genes they harbor is a necessary step towards understanding and effectively combating staphylococcal disease. Optical Mapping rapidly generates high-resolution, ordered restriction maps of whole genomes. "Optical Maps" provide a comprehensive view of genome architecture; therefore, map-based comparisons identify a variety of genomic rearrangements including insertions, deletions, duplications, inversions, etc. In this study we present the whole genome maps of five unsequenced Methicillin-Resistant Staphylococcus aureus (MRSA) isolates (Wisconsin strains WI-23, WI-33, WI-99, WI-591 and USA300-114) and their comparison to five sequenced genomes (N315, MW2, COL, Mu50, USA300-FPR3757). A variety of genomic differences were identified in each of the unsequenced isolates including the presence of unique genomic islands. The map data showed that most of the Wisconsin strains are closely related to the sequenced strain MW2 and USA300-0114 clusters most closely with the sequenced strains COL and USA300-FPR3757.

Phylogenetic analysis as a tool for testing selective advantage of single nucleotide polymorphisms.

Manzour Hernando Hazbón*, Magali Cavatore, Alifiya S. Motiwala, Michael Brimacombe and David Alland. University of Medicine and Dentistry of New Jersey, Newark, New Jersey.

The genome of *Mycobacterium tuberculosis* continuously adapts to the environment by selecting for single nucleotide polymorphism (SNPs) that confer selective advantage, such as increased fitness, virulence and drug resistance. There are numerous drug resistance-associated SNPs in genes involved in the mechanisms of drug resistance. However, the role of most of these SNPs remains unclear. Lengthy and cumbersome genetic studies are required to establish the biological role of each SNP. Based on the results of a previous work, we selected a set of nine SNPs that allowed classifying *M. tuberculosis* into seven SNP cluster groups (SCGs) and five subgroups. The genotype at positions *katG*315, *kasA*G269S, *kasA*G312S, *ahpC*D73H, *ndh*V18A, *inhA*S94A, *inhA*I194T, the promoter regions of *ahpC* and *inhA* were obtained as well as the SCGs for over 1000 clinical isolates of *M. tuberculosis*. We hypothesize that if a SNP confers a selective advantage it will be independently selected several times in the population, being detected in different SCGs. In the case of neutral SNPs, it should be the product of a rare event detected only in isolates belonging to one particular SCG or branch. Mutations at position *katG*315 (known to confer resistance to the drug isoniazid (INH)) were found in resistant isolates belonging to all groups, corroborating its role in INH resistance. The mutations *kasA*G269S, *kasA*G312S, *ahpC*-46, *ahpC*D73H and *ndh*V18A have been detected in both INH resistant and susceptible isolates. Each of these mutants was detected in single SCGs, suggesting that they do not confer any selective advantage. Mutants at positions *inhA*S94A, *inhA*I194T and all other mutants in the promoter regions of *ahpC* and *inhA* were detected only in resistant isolates belonging to several SCGs. This result suggests that these SNPs confer a selective advantage. In conclusion, our results suggest that phylogenetic analysis could be used to determine if a particular SNP confers a selective advantage.

Development and Evaluation of a Comprehensive Functional Gene array for Environmental Studies

Z. He[1,2], C. W. Schadt[2], T. Gentry[2], J. Liebich[3], S.C. Song[2], X. Li[4], and J. Zhou[1,2]

[1]The University of Oklahoma, Norman, OK, [2]Oak Ridge National Laboratory, Oak Ridge, TN, [3]Forschungszentrum Julich GmbH, Julich, Germany, [4]Perkin Elmer Life and Analytical Sciences, Boston, MA

To detect and monitor functions of microbial organisms in their environments, functional gene arrays (FGAs) have been used as a promising and powerful tool. In this study, we have constructed the second generation of FGA, called FGA2.0 that contains 23,843 oligonucleotide (50mer) probes and covers more than 10,000 sequences of targeted genes, which are involved in nitrogen, carbon, sulfur cycling and metabolism, metal reduction and resistance, and organic contaminant degradation. Several new strategies have been implemented in probe design, array construction and data analysis. Gene sequences were automatically retrieved by key words. A newly developed oligonucleotide design program CommOligo was used to select gene-specific and group-specific probes, and multiple probes were designed for each gene sequence or each group of highly homologous sequences. All designed oligonucleotides were verified and output in a 96-well format for direct order placement of oligonucleotide synthesis. To ensure the array specificity, the array has been systematically evaluated using different targets and environmental samples. The results demonstrate that such an array can provide specific analysis of microbial communities in a rapid, high-through-put and cost-effective fashion.

A Software Package for the Design and Analysis of Microbial Functional Gene Arrays

Zhili He[1,2], Ye Deng[1], Christopher W. Schadt[2], Xingyuan Li[3], Jizhong Zhou[1,2]

[1]University of Oklahoma, Norman, OK, [2]Oak Ridge National Laboratory, Oak Ridge, TN, [3]Perkin Elmer Life and Analytical Sciences, Boston, MA

A package has been developed to comprehensively analyze microbial functional genes for the design of functional gene arrays (FGAs). The package contains multiple modules to download sequences, remove low-homology sequences, design oligonucleotide probes, verify probe specificity, output selected oligonucleotide probes, and finally store data in local databases. The gene download module retrieves functional gene sequences from publicly available databases (e.g. GenBank) using BLAST algorithms and key words, or a combination of key words with AND, OR, and NOT operations as a query for a particular functional gene. Since some downloaded sequences may be mis-annotated or not related to the functional gene of interest, the homology search module removes those low-homology sequences, and also discards duplicate sequences. All processed sequences are used to design oligonucleotide probes using CommOligo, which designs single or multiple oligonucleotides for each sequence or each group of homologous sequences. Prior to oligonucleotide design, a multiple sequence alignment (MSA) may be conducted to select conserved regions to be used for probe selection. The probe check module confirms the specificity of all designed oligonucleotides using the same criteria as probe design against larger databases (e.g. GenBank). The plate production module outputs qualified probes in a 96-well plate format, which can be directly input into oligonucleotide synthesis systems of commercial companies. Finally, all information for sequences, probes and arrays are stored in databases, which can be easily retrieved by Search, Link and other functions. Since sequences are continuously deposited into public databases. This package is able to automatically update all information periodically so that it is kept consistent and up-to-date. Experimental details and analysis tools will be integrated into this package as well as development continues. This package will greatly facilitate the analysis of microbial functional gene sequences and arrays.

Molecular Characterization of a Basidiomycete Yeast Environmental Isolate: How Little We Know

Les M. Hoffman
EPICENTRE Biotechnologies, Madison WI

Ten to thirty percent of environmental fungi are culturable, as opposed to the bacteria, which are largely not culturable. Environmental yeasts should, by extension, be an attractive set of organisms for genomics studies. They also should also possess relatively small genomes and short introns. A salmon pink contaminant of bacterial agar plates was isolated and characterized at the molecular level. The organism was exceptionally impervious to chemical or enzymatic lysis. The sequence of the 18S ribosomal gene region was identical to that of Rhodotorula slooffiae, a pigmented yeast in the Uridiniomycetes family of the class Basidiomycetes. Carotenoids, including astaxanthin, are antioxidant and anticancer compounds found abundantly in Rhodotorula sp. Astaxanthin is the major carotenoid used for pigmentation of commercially raised salmon.

Rhodotorula genomic DNA was used to construct a library in the fosmid vector pCC2, without sizing the DNA fragments obtained with the MasterPure DNA Isolation Kit. The resultant library, containing an approximately five-fold representation of the genome, was analyzed by random fosmid end sequencing. In an attempt to clarify the phylogenetic relation of the yeast to other fungi, I examined random fosmid end sequences. BLASTX homologies to conceptually translated proteins, however, showed no consistent relationship to any family, or even to the class Ascomycetes or Basidiomycetes. Translations of
R. slooffiae proteins matched most often those of Cryptococcus neoformans, a Hymenomycete which is not closely related to R. slooffiae according to rDNA sequences. This initial study of a yeast genome demonstrated the limitations imposed by the lack of curated genomic data for common environmental fungi.

Genome Sequencing of Dechlorinating Bacterium *Desulfitobacterium hafniense* Y51 and Development of its DNA Microarray

K. Inatomi[1*], S. Yamamoto[1], X. Peng[1], H. Nonaka[1], M. Inui[1], K. Furukawa[2] & H. Yukawa[1]
[1]Research Institute of Innovative Technology for the Earth, Kyoto, JAPAN, [2]Kyushu University, Fukuoka, JAPAN.

Halogenated organic compounds such as tetrachloroethene (PCE) and trichloroethene (TCE) are toxic and persistent ground water contaminants that pose risks to human health. Anaerobic dechlorinating microorganisms that can dechlorinate these contaminants to nontoxic products of ethene have been identified in contaminated sites. The two most prominent dechlorinating genera are *Dehalococcoides* and *Desulfitobacterium. Dehalococcoides ethenogenes* 195 dechlorinates PCE to ethene using only hydrogen as an electron donor and chlorinated ethenes as electron acceptors (halorespiration_. In contrast, *Desulfitobacterium* strains can not only utilize electron acceptors other than chlorinated compounds but also utilize several organic acids as electron donors. In this poster, we report the complete genomic sequence of the *Desulfitobacterium hafniense* Y51and compare it to that of *D. ethenogenes 195*. The development of a DNA microarray composed of 5,060 gene sets is also presented.

The complete genome of *D. hafniense* Y51 is a 5,727,534-bp circular chromosome harboring 5,060 predicted protein coding sequences (G+C 47.4%). The large genome reveals a more versatile microorganism than *D. ethenogenes* 195 (1,469,720-bp, 1,580-CDSs) with a narrow metabolic repertoire. The *D. hafniense* Y51genome has only two reductive dehalogenase genes, compared to 17 genes of *D. ethenogenes* 195._Although the dehalogenase of *D. hafniense* Y51 (PceA) show 24% similarity with that of *D. ethenogenes* 195 (TceA), they exhibit Tat signal peptides and binding motifs of two iron-sulfur clusters. To identify genes involved in dehalorespiration, we performed DNA microarray analysis composed of 5,060 genes. RNA isolated from *D. hafniense* Y51cells grown in the presence of 0.3 mM TCE or PCE for 16 hr was used to analyze gene expression. This research was supported by New Energy and Industrial Technology Development Organization (NEDO), Japan.

Profiling of the Gut Microbiota of Identical with Crohn's Disease Using Proteomics and Microbiomics

Janet Jansson[1]*, Nathan VerBerkmoes[2], Manesh Shah[2], Robert Hettich[2], Johan Dicksved[1], Jonas Halvarsson[3], Curt Tysk[3], Lars Engstrand[4] and Magnus Rosenquist[1].[1]Swedish University of Agricultural Sciences, Uppsala, Sweden, [2]Oak Ridge National Laboratories, Oak Ridge, TN, [3]Örebro University Hospital, Örebro, Sweden, [4]Swedish Institute for Infectious Disease Control, Solna, Sweden.

Crohn's disease is a prevalent chronic inflammatory disorder of the gastrointestinal tract. Although great advances have been made and animal models are available, the mechanisms behind development of Crohn's disease are still poorly understood. Specific bacterial inhabitants of the gastrointestinal tract may cause development of Crohn's disease in susceptible individuals. The response observed also depends on the patient genotype. Animal models have demonstrated that different species of bacteria cause variations of inflammatory bowel disease (IBD) in hosts with the same genetic background. Our material, derived from monozygotic twins, provided us the unique possibility to examine the correlation between gastrointestinal bacteria and IBD. Fecal samples were collected from 6 discordant pairs (only one of the twins has Crohns), 4 concordant pairs (both have the disease) and 8 healthy pairs. The fecal microbiota of each individual was fingerprinted using terminal restriction fragment length polymorphism (T-RFLP) and multivariate statistics was used to determine whether there were any similarities or differences between the individuals. In addition, GC fractionation was used to dissect the community DNA of 6 individuals prior to T-RFLP analysis to identify rare members of the communities. The same 6 fecal samples were analyzed by shotgun proteomics to obtain representative proteomes of the fecal microbiota. Currently over 1000 proteins were identified by mass spectrometry with high representation from *Bifodobacterium* and *Bacteroides* species. As the subjects in the proposed study are both discordant and concordant identical twins, the results highlight how the composition of the microbiota impacted the disorder independently of the genetics of the subjects.

Probing the genotype-phenotype relationship in *E. coli* strains adapted to growth on glycerol

A.R. Joyce[1,2]*, A.M. Maggay[2], C. Herring[2], J. Zhang[3], R. Landick[4], & B.Ø. Palsson[2]. [1]Bioinformatics Program, [2]Department of Bioengineering, University of California – San Diego, La Jolla, CA, U.S.A. [3]Department of Biomolecular Chemistry, [4]Department of Bacteriology, University of Wisconsin, Madison, WI, U.S.A.

Laboratory evolution has long served as a powerful research tool for studying genotype-phenotype relationships in microbes. In this study, we build on prior work that comprehensively identified the sequence changes responsible for the acquisition of a selective growth advantage of five different *E. coli* strains adapted to glycerol-supplemented minimal growth medium [1]. In order to assess further the compendium of mutations that can confer selective growth advantage on glycerol-supplemented medium, 45 additional adaptively evolved strains were generated through serial passage of log-phase cultures for 300-400 generations and previously identified target loci were re-sequenced. As previously observed [2], all strains exhibited a dramatic change in phenotype as evidenced by a doubling in growth rate (_~ 0.25 to _~ 0.5). Furthermore, the vast majority of strains harbor mutations in glycerol kinase (*glpK*) and RNA polymerase (RNAP) (*rpoB* or *rpoC*). Mutations in *glpK* and *rpoB* are typically single nucleotide polymorphisms whereas indels (9-36 bp in-frame deletions) are typically observed in *rpoC*. Many of the *glpK* mutations were previously shown to enhance glycerol uptake by relieving allosteric inhibition [1]. In addition, recent *in vitro* work shows that two of the most common deletions observed in *rpoC* suppress RNAP pausing by an order of magnitude. In sum, this work more firmly establishes that the acquisition of only a few mutations, specifically related to substrate uptake in *glpK* coupled with a global transcriptional regulatory modification in the form of RNAP mutations, drive the large phenotype change. Further, this work also provides insight into the manner in which *E. coli* navigates the adaptive landscape for growth on glycerol.

1. Herring CD, Raghunathan A, Honisch C, et. al. (2006) Comparative genome sequencing of Escherichea coli allows bacterial evolution to be observed on a laboratory timescale. (Under Review).
2. Fong SS, Joyce AR, & Palsso BØ. (2005) Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. Genome Res **15**(10):1365-72.

Increasing mevalonate production by engineering the metabolism of *Escherichia coli*

Connie (Yisheng) Kang, Jay Keasling. QB3, University of California- Berkerley

Synthesize the anti-malarial drug, artemisinin, precursor mevalonate in *Escherichia coli* branches from acetyl Co-A which is the entry point to TCA cycle. We have been using genetic and environmental manipulations to redirect carbon flux from the endogenous central metabolic pathways (CMPs) to the heterologous pathway precursors. Deletion the production pathways of acetate, which is an undesirable product of excessive glycolytic flux, result in excretion of pyruvate rather than help increase mevalonate production. Heterologous mevalonate pathway from *Enterococcus faecalis* is more effective to draw carbon flow than that of *Saccharomyces cerevisiae.* Providing limited amount of nitrogen source also efficiently cut carbon flux to biomass and redirect it to mevalonate production. We are also performing metabolic flux analysis using ^{13}C-labeled glucose. This information will help us determine how carbon flux through native metabolic pathways is affected by the presence of heterologous pathways, allowing us to identify and correct bottlenecks in the artemisinin production pathway. Successful completion of this project will provide insights into the metabolic status of living cells under different conditions and help us build a robust bacterium capable of producing high levels of artemisinin from cheap carbon sources.

THE UCLA Undergraduate Genomics Research Initiative: Sequencing and preliminary analysis of the genome of *Ammonifex degensii*

Cord D. Kirshner*, Kimberly M. Houck*, Liliane Idylle, Christopher H. Shin, Dominic L. Taylor, Thuy Nguyen*, Ahmed Qureshi,Gaston Pfluegl, Robert W. Simons, Shelley N-M Thai, & Cheryl A. Kerfeld.
University of California, Los Angeles

The thermophilic bacterium *Ammonifex degensii* is a strictly anaerobic, chemolithoautotroph that grows by nitrate ammonification or by sulfate/sulfur reduction. This microbe hails from a terrestrial solfataric hotspring in Indonesia. Other bacterial and archaeal nitrate ammonifiers occupy similar biotopes such as deep-sea hydrothermal vents and marine sedimentary layers. *A. degensii* empirically stains Gram-negative yet is phylogenetically related to the low G+C subgroup of Gram-positive bacteria based on its 16S rDNA sequence. Interestingly, *A degensii* contains ~ 85% glycerol diethers in its lipid bilayer.

The UCLA Undergraduate Genomics Research Initiative (UGRI) is sequencing *A. degensii's* genome. The UGRI is precedent-setting undergraduate education in which students conduct cutting-edge research in genomic biology and biotechnology. The UGRI is a collaborative, cross-discipline effort among undergraduates in general education and lower- and upper division life science courses.

The UGRI is currently in the sequencing phase of the project. Preliminary sequence data and analyses suggest that *A. degensii* may be a spore forming acetogen. In addition, roughly 9% of *A. degensii's* putative proteins most closely resemble archaeal proteins. The UGRI's collaborative sequencing effort is aimed at further elucidating the genetic basis of *A. degensii's* metabolic diversity.

Modeling the Metal Reduction Pathway of *Shewanella Oneidensis* MR-1 from Time Series Data using the S-system

R. Lall* & J. Mitchell.BACTER Institute, University of Wisconsin at Madison

Shewanella Oneidensis strain MR-1 had been used previously in laboratory studies of metal reduction on isolation from anaerobic sediments. Concentrations of several metals including FE(II)NTA, FE(II)citrate, CrO4, Co(III)EDTA, UO2 and TcO4 were measured at selected time points. In addition, concentrations were measured for uranyl acetate in the presence of varying ratios of geothite and ferrihydrite. The goal of the present study is to determine suitable kinetic parameters for a mathematical model of the metal reduction pathway in *Shewanella Oneidensis* that is consistent with the observed data. We chose for this purpose an S-system model, as it possesses a mathematical form that is flexible enough to accommodate typically encountered kinetic and regulatory behaviors, yet explicit enough for the topology of the underlying network to be inferred from the values of its parameters. The results of this work have yielded a model that compares the reduction rates of metals and mixed metals. For metals, parameters were estimated using a finite difference approach in combination with multiple linear regression. For mixed metals, parameter estimation and curve fitting are accomplished with a generalized least squares formulation that handles systems of ordinary differential equations and is implemented in Matlab. It consists of an optimization algorithm (Levenberg-Marquardt, LSQCURVEFIT) and a numerical ODE solver (Runge-Kutta, ODE45). Simulations with the estimated parameters indicate that the model captures the experimental data quite well.

The literature contains models of metal reduction in *S. Oneidensis* that are based on Monod kinetics which become statistically overparametrized resulting in large uncertainties in their parameters. In the current work, statistical analysis has indicated that confidence limits of the parameters obtained are in the range of BST models.

Insights into adaptation to the deep sea from the genomes of piezophilic bacteria.

Federico M. Lauro* & Douglas H. Bartlett

Center for Marine Biotechnology and Biomedicine, Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0202, USA.

Gradients of physicochemical factors determine to a great extent the growth rate and survival of bacteria in the deep sea. One could speculate that the genetic modifications required to adapt to specific depth zones should be substantial. However, the lack of available genome information on deep-sea isolates has prevented the testing of this hypothesis.
Here we present the preliminary results of the analysis of the first 5 genomes of psychropiezophilic (cold- and pressure-loving) bacterial isolates and the comparison to the genomes of the closest non-piezophilic relatives.

In general the genomes of deep-sea bacteria show a high ratio of rRNA operon/genome size and remarkably large intergenic regions compatible with the hypothesis of an opportunistic (r-strategy) lifestyle with a high degree of gene regulation. Other hallmarks of all piezophilic genomes are the presence of a large number of genes involved in membrane unsaturation, such as Δ-9 desaturases and PUFA gene clusters, and the absence of photolyase genes.

On the contrary, the comparison of the orthologous protein-coding portion of the genome shows relatively few differences between dramatically diverse depth-specific ecotypes (bathytypes). Therefore, the colonization of new depth niches might rely primarily on changes in gene regulation in addition to the acquisition of a limited number of specific genomic regions. We are testing this hypothesis by transferring depth-specific genes between different bathytypes of *Photobacterium profundum*.

For example, 16S rRNA sequence comparisons indicate the presence of piezophile-specific insertions to the stems of helices 10 and 11. The introduction on a multi-copy plasmid of the long-stem ribotype from the deep bathytype *P. profundum* SS9 enhances growth of the shallow-water bathytype 3TCK at high hydrostatic pressure.

Conversely the introduction of the deoxyribodipyrimidine photolyase (*phr*) gene from the shallow-water bathytype 3TCK into the deep-sea strain SS9 confers enhanced UV tolerance that is dependent on blue-light irradiation (photoreactivation).

Construction of the L-threonine overproducing strain using reduced-genome *Escherichia coli.*

Jun Hyoung Lee*, B. H. Sung, S. H. Lee, J. Y. Lee, K. H. Kang, and S. C. Kim. Korea Advanced Institute of Science and Technology, Daejeon, Korea.

Genome engineering, including the restructuring and minimization of a microbial genome, has been recently used for the improvement of industrial strains. Especially, the minimization of a genome provides many advantages over conventional approaches for the strain improvement, such as simplified metabolism without physiological compromise, reduced production of unwanted by-product, and increased genome stability. Recently, a reduced-genome *Escherichia coli* MDS42 (*Science*, 312, 1044-1046) has been constructed with normal growth pattern, increased transformation efficiency, and no IS-mediated mutagenesis or genomic rearrangement. In this study, this reduced-genome *E. coli* has been reengineered to increase the productivity of an essential amino acid L-threonine, by overexpressing the feedback resistant threonine operon, deleting the threonine dehydrogenase gene and threonine transporter genes, and introducing the mutant threonine exporter gene. The resulting strain, MDS42-thr, shows 40% increase in threonine productivity compared to a wild-type *E. coli*-derived strain. This result clearly indicates that elimination of unnecessary genes for cell growth from a genome can increase the productivity of the strain significantly by reducing the metabolic burden on cells.

Phenotypic Switch Using Artificial Transcription Factors in *Escherichia coli*

Ju Young Lee*, Bong Hyun Sung, Jun Hyoung Lee, Sang Hee Lee, Kui Hyeon Kang, and Sun Chang Kim. Korea Advanced Institute of Science and Technology, Daejeon, Korea

Now that the genomes of many organisms have been sequenced, the next goal is to annotate the products of newly identified genes and to engineer the desired phenotypes of organisms at will. To facilitate this goal, we have developed novel artificial transcription factors capable of up- or down-regulating the expression of genes regardless of endogenous transcription factors in *Escherichia coli,* using zinc finger DNA binding proteins with distinct DNA-binding specificities and the *E. coli* cyclic AMP receptor protein (CRP). By random assembly of 53 different types of zinc fingers, we have constructed more than 1.4×10^6 DNA binding domains composed of three zinc fingers, and have fused each of them to CRP as an effector domain. With the resulting artificial transcription factors, various phenotypic changes were induced in *E. coli,* and new strains with a phenotypic change have screened for industrially important traits, such as growth improvement and the resistance to heat or cold shock. These results and the general applicability of this platform demonstrate that these artificial transcription factors can be used a powerful tool for functional genomics and phenotypic engineering.

Deletion of non-essential genomic segments in *Escherichia coli* using a transposon mariner- and gd-mediated bidirectional deletion system

Sang Hee Lee*, B. H. Sung, J. H. Lee, J. Y. Lee, K. H. Kang, and S. C. Kim. Korea Advanced Institute of Science and Technology, Daejeon, Korea

Numerous genome information obtained by genomic studies after complete genome sequencing of diverse species has been used to define and understand the cellular life at the molecular level. To facilitate the understanding of the life, we have developed a powerful bidirectional deletion system for identifying essential genes and minimizing bacterial genomes. The technique, which produces nested sets of deletions efficiently, involves a hybrid transposable element that includes the components of two transposable elements, mariner and gd, as well as two counterselectable genes, *tetR* and *sacB*, for phenotypic selection. We have made a large pool of independent transposon insertion mutants in *Escherichia coli* using the transposable element mariner, and have mapped over 600 transposon insertion sites on *E. coli* chromosome. By the induction of gd transposase, a bidirectional deletion was performed from the gd ends inserted to the adjacent chromosomal DNA with various distances, and deletion mutants were selected on the medium containing either kanamycin or sucrose. We obtained *E. coli* strains in which various genomic segments were deleted by a serial or simultaneous bidirectional deletion. This provides a robust technology for eliminating dispensable genes and constructing a minimal genome.

Phylogenetic Diversity of a Soil Microbial Community

T. H. Lee*, K. Flummerfelt, and E.R. Sanders-Lorenz. University of California, Los Angeles.

The soil harbors an extensive community of microorganisms comprised of bacteria, fungi, and other lower eukaryotes, which coexist together to create a unique terrestrial environment of astounding ecological and phylogenetic diversity. To explore microbial diversity within this environment, we reconstructed the phylogeny of a bacterial community using small subunit ribosomal RNA (SSU rRNA) genes. Soil samples were collected from the base of a cabbage tree (*Cussonia natalensis*) in the UCLA Botanical Garden. An assortment of bacteria were cultivated using enrichment media, and genomic DNA was harvested from 24 microbial isolates. Total DNA representing the entire metagenome was also purified directly from the same soil sample. Both the cultivation-dependent and environmental genomic DNA samples were subjected to PCR and sequencing of SSU rRNA genes. DNA sequences were compared to those of classified organisms in the Genbank database using BLAST. A more detailed phylogenetic picture of the soil community was obtained by building evolutionary distance-based trees, which were compared to determine how well the isolates represented of the diversity of microorganisms in the environmental sample. The analysis revealed that many of the isolates belong to well-characterized taxonomic groups, including the class *Gammaproteobacteria* and the phyla *Firmicutes* and *Actinobacteria*. Other isolates belong to classes having few cultivated representatives such as the *Flavobacteria* and *Sphingobacteria*. Because soil microorganisms are prolific manufacturers of antibiotic substances, isolates were screened for antibiotic production. Two produced antibiotics inhibiting growth of an indicator bacterial strain. Isolates were also surveyed for resistance to antibiotics. Seven displayed resistance to a single drug, while one showed multiple-resistance. These phenotypic observations demonstrate a range of ecological strategies employed by microbes in this habitat.

IMG/M: An Experimental Metagenome Management and Analysis System

Natalia Ivanova[1], Victor M. Markowitz[2], Frank Korzeniewski[2], Krishna Palaniappan[2], Ernest Szeto[2], Inna Dubchak[3], **Athanasios Lykidis**[1*], Iain Anderson[1], Konstantinos Mavromatis[1], Victor Kunin[4], Hector Garcia Martin[4], Philip Hugenholtz[4], and Nikos Kyrpides[1].
[1]Genome Biology Program, Joint Genome Institute, USA; [2]Biological Data Management and technology Center, Lawrence Berkeley National Laboratory, USA; [3]Genomics Division, Lawrence Berkeley National Laboratory, USA; [4]Microbial Ecology Program, Joint Genome Institute, USA

We have developed an experimental metagenome data management and analysis system, IMG/M, based on the Integrated Microbial Genomes (IMG) system. In addition to ING's isolate genomes, IMG/M includes metagenome sequence data generated from several environmental microbiome samples which comprise a representative set in terms of species diversity, abundance of dominant organism(s) and sequencing depth. The IMG/M data analysis tools are the result of extending the IMG data analysis tools to handle metagenome data. Additional tools allow examining profiles of functional annotations across microbial communities and isolate organisms of interest, and analyzing strain-level heterogeneity within a species population in metagenome data. IMG/M shows that although data processing of metagenome sequences needs to overcome numerous challenges due to the complex nature and inherent incompleteness of the data, successful analysis can be conducted in the context of a comprehensive data management and analysis system that provides support for data review and revision. IMG/M has been used for completing the analysis of biological phosphorus removing (EBPR) sludge communities and for studying the metagenomes of several microbial communities recently sequenced by JGI, including the hydrogen-producing consortium colonizing the termite hindgut. IMG/M is available at http://img.jgi.doe.gov/m

This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research; the University of California; and Lawrence Berkeley, Lawrence Livermore and Los Alamos National Laboratories.

The National Microbial Pathogen Data Resource (NMPDR): A Bioinformatics Platform for Research Support

L. McNeil* & C. Reich. University of Illinois, Urbana, IL.

The National Microbial Pathogen Data Resource (NMPDR) is one of eight Bioinformatics Resource Centers funded by the National Institute of Allergy and Infectious Disease (NIAID) to provide the comprehensive bioinformatics environment needed to support research in biodefense, and emerging and re-emerging pathogens. The NMPDR focus organisms, all Category B priority pathogens, include the diarrheagenic bacteria *Campylobacter jejuni*, *Vibrio cholerae*, *V. parahaemolyticus*, *V. vulnificus*, and *Listeria monocytogenes*. Also included are *Staphylococcus aureus*, *Streptococcus pneumoniae* and *S. pyogenes* (Group A Strep).

NMPDR contains the complete genomes of nearly 50 strains of these pathogens that are the focus of our curators, as well as more than 400 other genomes that provide a broad context for comparative analysis. NMPDR integrates genomes with expertly curated biological subsystems to provide consistent genome annotations. Subsystems are sets of functional roles related by a biologically meaningful organizing principle, built over large collections of genomes. Investigators can browse subsystems and reactions to develop accurate reconstructions of metabolic networks, pathogenicity markers, virulence factors, etc.

Organism summary pages contain information about the focus pathogens, from textbook descriptions to recent findings. The organism summary pages are designed to provide user services and a collaborative environment for investigators. A literature aggregator provides access to the most recent scientific developments. Epidemiological data will be added presently. NMPDR provides a comprehensive bioinformatics platform, with tools and viewers for genome analysis. Widely used bioinformatics tools, as well as especially designed functionalities, are integrated into the system. Results of precomputed gene clustering analyses can be retrieved in tabular or graphic format with one-click tools. NMPDR unique tools include Signature Genes, which finds the set of genes in common or that differentiates two groups of organisms. Essentiality data collated from genome-wide studies have been curated. Drug target identification and high-throughput, *in silico*, compound screening are in development.

Transformation of *Vibrio cholerae* with Large Chromosomal Regions and Superintegron Cassettes via Natural Competence

Michael C. Miller* & Gary K. Schoolnik. Stanford University School of Medicine.

The ability of *Vibrio cholerae* to thrive under diverse conditions is due in part to its plastic genome. Chitin, a prominent marine carbon source, induces natural competence, allowing *V. cholerae* to incorporate DNA into its genome via homologous recombination. Here we characterize genetic exchange between strains and demonstrate that the genome is regional in susceptibility to recombination.

V. cholerae strains induced to competence by growth on chitin were transformed with gDNA from various *V. cholerae* strains. Transformants were selected by complementation of a metabolic pathway or gain of antibiotic resistance and recombination events detected by comparative genome hybridization, transposon site hybridization, or PCR.

Isolates of environmental *V. cholerae* transformed with gDNA from a clinical strain were able to acquire genes encoding a transporter for $(GlcN)_2$ or mannose-6-phosphate isomerase. Recombination events as large as 42kb were detected. The relative recombination frequency across the genome was assayed by transforming with gDNA from a *V. cholerae* mariner library. The pathogenicity islands were depleted in recombination events, while the superintegron was enriched. Transformation at the superintegron was due in part to integrase-mediated mobilization of integron cassettes from within the donor DNA superintegron to the primary attachment site of the recipient superintegron.

Natural competence allows the mobilization of complex functions encoded in large segments of DNA between *V. cholerae* lineages. Exchanges mediated by homologous recombination occur preferentially in regions of the chromosome with *Vibrio*-typical sequence structure. DNA taken up by natural competence can also serve as a substrate for integration at the superintegron. Genetic exchange via natural competence explains in part the extraordinary level of genomic diversity observed in *Vibrio* populations.

Features and applications of a new set of *E. coli* K12 deletion mutants under construction

H. Mori1, N. Yamamoto1, K. A. Datsenko2, T. Nakamichi1, T. Baba1, B. L. Wanner2.

1. Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara 630-0101, Japan.
2. Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.

We recently described the construction of a complete set of single gene deletion mutants of all non-essential genes in E. coli K-12, called the Keio collection (Baba et al.2006) Features of this collection are: 1) mutants are deleted of the entire open reading frame from codon two to six codons from the C-terminus, 2) deleted genes are replaced with kanamycin resistance gene to facilitate transfer with the target ORF, and 3) eviction of the kanamycin resistance gene by FLP-FRT site specific recombination leaves behind an in-frame deletion from the Met start codon to sixth residue from C-terminus.

To perform genetic network analysis with the Keio collection, we carefully designed the new deletion mutant series that will enable us to extend system-level studies of *E.coli* with more flexibility. Features of the new set1) mutants are deleted similar to the Keio collection, 2) deleted genes are replaced with chloramphenicol resistance gene instead of kanamycin resistance to permit construction of double mutants, 3) each mutant has an in-frame fusion to the Met start codon of an enhanced variant of the green fluorescent protein called venus (Nagai et al. 2002),4) mutants contain mutant FRT sites called FRT1 that cannot recombine with FRT of Keio collection, and 5) mutants containa20-nt molecular bar code near the C-terminus of the deleted gene.

We have now developed candidate template plasmids for this construction and have completed pilot tests for feasibility. Our new mutant set will be transferable by conjugation and be useful for single cell monitoring of the target gene expression.

We will report on our recent progress and schedule and how preliminary result for future large scale approaches.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol, 2, 2006 0008.http://www.nature.com/msb/journal/v2/n1/full/msb4100050.html

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K. and Miyawaki, A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat Biotechnol, 20, 87-90.

RpoS Regulon of *Legionella pneumophila.*

[1,2]Pampou, S.*, [1]Morozova, I., [1]Morozov, P., [1]Russo, J.J., [2]Shuman, H.A., [1]Kalachikov, S.
[1]Columbia Genome Center and [2]Microbiology Department, Columbia University, New York, NY 10032

The *rpoS* gene encodes the alternative sigma factor RpoS or s^s, the main stationary phase transcriptional regulator that confers adaptation to nutrient limitation in Legionella and other bacteria. To elucidate the overall effect of RpoS-mediated regulation on Legionella physiology during the transition to stationary phase, we performed genome-wide gene expression analysis of wild-type Legionella strains and an isogenic *rpoS* mutant strain. The transition from a free-living to an intracellular lifestyle is accompanied by acute changes in nutrient availability as well as changes in pH, metal concentrations, and toxic substances. Based on previous analyses, RpoS plays an important role during both exponential and stationary phases of growth and is necessary for *Legionella*'s survival inside protozoan hosts.

We found that during growth in rich media, the expression levels of about 20% of the Legionella genes are altered upon entry into stationary phase. In contrast, many fewer changes in gene expression accompany the cessation of growth of the *rpoS* mutant strain. Comparison of expression profiles between the wild-type and the mutant strains shows that RpoS exhibits both positive and negative regulatory effects. While 159 genes are repressed by RpoS during exponential growth, this repression is largely lifted during transition to the stationary phase. Transcription of another 220 genes is induced by RpoS in stationary phase. We were able to identify a putative RpoS-dependent promoter sequence for these genes. The two sets of RpoS-dependent genes encode proteins involved in energy metabolism, amino acid degradation, detoxification, ammonia assimilation and cell wall lipids and peptidoglycan synthesis. In addition, the expression levels of 75 Legionella transcription factors are altered upon cessation of exponential growth, and of these, 52 were dependent on RpoS.

These results highlight the importance of understanding the regulatory networks that respond to nutrient availability and environmental signals for intracellular pathogens.

Connecting Quantitative Regulatory-Network Models to the Genome

Y. Pan*, T. Durfee & M. Craven. University of Wisconsin, Madison.

Genetic regulatory networks are one of the key interacting networks in the living cell. In recent years many methods have been developed to reconstruct such networks from a variety of high-throughput data. A significant limitation of current approaches, however, is that the inferred models do not represent many of the things that biologists can directly manipulate, such as the growth medium and the genomic sequence of the organism being studied. Our work is aimed at developing methods for inferring regulatory-network models that are more directly connected to the genome and environment of the cells being modeled, so biologists can directly manipulate them to make predictions on gene expression. Currently, we are considering *E. coli* genes that are differentially expressed in carbon-shift experiments as a case study. Our approach applies non-linear Michaelis-Menten equations to describe how the transcription rates of genes are quantitatively dependent on their regulators' states and to explain how kinetic binding between each gene's promoter and its regulators affects transcription rates. We have extended the approach of Nachman *et al.* such that we learn models of binding parameters as a function of genome-sequence features. Our machine-learning methods for inferring such models show the improved expression level predictions over baseline models and previous approaches.

Functional Genomic analysis of *Variovorax paradoxus* strain EPS using Tn5 mutagenesis

Michael Pehl, Andrew Cajigal, Karen Kong, and Paul M. Orwin
California State University, San Bernardino

Abstract

Variovorax paradoxus strain EPS was cultivated from the soil in the natural preserve at CSUSB. The strain was selected for its mucoid phenotype and oligotrophic growth. Further experiments demonstrated that it forms biofilms readily in culture. Our laboratory has developed a high throughput system for transposon mutagenesis using pOT182:Tn5 to generate a library of transconjugants containing marked insertions in the *Variovorax* genome. Validations using recovered flanking DNA from insertions as well as analysis of promoterless *lacZ* expression suggest that our library is representative of the total genome. The Tn5 construct we use includes an *E. coli* plasmid *ori* to allow for rescue cloning and rapid identification of the disrupted gene. Because very little genomic information is available for *V. paradoxus*, this method represents a rapid and efficient way to identify genes important in functions of interest, such as the formation of biofilms, and growth under low nutrient conditions. Our phenotypic screens of approximately 25000 mutants have yielded putative insertion mutants altering pigmentation, growth rate on low nutrient agar, hydrolysis of Tween 80, and colony morphology. Experiments have identified mutants within these subsets that have alterations in their ability to form biofilms in a 96-well plate biofilm screen. Further experiments using the promoterless *lacZ* gene in this Tn5 construct have demonstrated differential gene expression on rich (YE) and poor (10% YE) agars. These findings will allow for development of *V. paradoxus* as a tool for bioremediation and biocatalysis.

From genomes to designed genomes: The *E. coli* reduction experience.

György Pósfai[1,2], Guy Plunkett III[2,3,4], Tamás Fehér[1], David Frisch[2,4], Günther M. Keil[5], Kinga Umenhoffer[1], Vitaliy Kolisnychenko[1], Buffy Stahl[2], Monika de Arruda[2], Valerie Burland[2,4], Shamik S. Sharma[6], Sarah W. Harcum[7], and Frederick R. Blattner[1,2]

[1]Institute of Biochemistry, Biological Research Center, Szeged, Hungary
[2]Scarab Genomics LLC, Madison, WI, USA
[3]Department of Genetics, University of Wisconsin, Madison, WI, USA
[4]Genome Center of Wisconsin, University of Wisconsin, Madison, WI, USA
[5]Federal Research Centre for Virus Diseases of Animals, Institute of Molecular Biology, Friedrich-Loeffler Institutes, Greifswald-Insel Riems, Germany
[6]Department of Chemical & Biomolecular Engineering, Clemson University, Clemson, SC, USA
[7]Department of Bioengineering, Clemson University, Clemson, SC, USA

As one of the best understood and thoroughly analyzed organisms, *Escherichia coli* K-12 is the platform of choice for genetic, biochemical, and metabolic simulation research. It is also used commercially, for the production of metabolites such as amino acids and proteins of therapeutic or commercial interest, and for the production of DNA for gene therapy, DNA vaccines, and iRNA. But because *E. coli* evolved in both animal intestines and the environment, parts of its genome are unnecessary — possibly even counterproductive — for some applications. In addition, mobile DNA elements, disseminated throughout the genome, mediate recombination events such as transposition and horizontal gene transfer.

Proposing that a reduced genome might improve metabolic efficiency and decrease the redundancy among *E. coli* genes and regulatory circuits, we have used synthetic biology to trim the *E. coli* K-12 genome by making a series of planned, precise deletions. The multiple-deletion series (MDS) strains, with net genome reductions of 15% or more, were designed by identifying non-essential genes and sequences for elimination, including recombinogenic or mobile DNA and cryptic virulence genes, while preserving good growth profiles and protein production. By means of a rational design strategy we avoided a loss of robustness that would result from more extensive deletions or an attempt to construct a minimal genome. We have constructed genetically stable "tabula rasa" strains with robust metabolic performance, to which genes for practical applications may be added.

Genome reduction also led to unanticipated beneficial properties: high electroporation efficiency and accurate propagation of recombinant genes and plasmids that were unstable in other strains. Eradication of stress-induced transposition has stabilized the MDS genomes and provides some of the new properties.

Computational Approach to Metabolic Engineering in Microbes

V.A. Portnoy[1]*, M.J. Herrgard[1], S.Van Dien[2], B.Ø. Palsson[1] [1]University of California – San Diego, La Jolla, CA, U.S.A. [2]Genomatica Inc.,San Diego, CA, U.S.A

Engineering microbes to produce desired products has a potential to bring a lot of advancement to biotechnological and pharmaceutical areas. There have been a number of successful studies aimed on production of simple metabolites such as ethanol, succinate, glycerol etc. from prokaryotic and eukaryotic organisms [1,2]. The goal of this study is to enhance production of ethanol from *S. cerevisiae* by enchasing anaerobic fermentation through deletion of genes required for assembly of ATP synthase. To this end, two genes, ATP2 and AEP1, were selected by the OptKnock [3] computational framework as possible deletion candidates for ethanol surplus production. *In silico* metabolic model *i*ND750 [5] was used as a basis for the OptKnock design of ethanol overproducing strains. The product of ATP2 is directly involved in the formation of F1-ATP synthase, while AEP1 is a translational regulator required for expression of the mitochondrial ATPase subunit 9 in yeast [4]. Deletion of either of these two genes does not affect viability of the organism, but leads to the phenotypical change resulting in the increase of ethanol production. Computational approaches such as an *in silico* modeling and optimization based algorithms were used to predict growth rates and ethanol production of selected mutants. Experimental and computational results demonstrated increased production of ethanol with respect to glucose uptake by close to 50 % for mutants; however, experimental growth rates of selected strains were significantly lower expected. Due to coupling of ethanol secretion to growth rate we will use adaptive evolution as a way of increasing growth rate and therefore ethanol production. Direct identification of target deletions demonstrated in this study allows obtaining predicted phenotypes and production of desired products. This methodology could be further applied to other well known organisms such as *Escherichia coli* and *Bacillus cereus*.

1. Kalscheuer, R, Stolting, T, & Steinbuchel, A. (2006). Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology*, 152(Pt), 2529-2536.
2. Wang, Q, Chen, X, Yang, Y, et al. (2006). Genome-scale *in silico* aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production. *Applied microbiology and biotechnology*,
3. Burgard, AP, Pharkya, P, & Maranas, CD. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, 84(6), 647-57.
4. Takeda, M, Vassarotti, A, & Douglas, MG. (1985). Nuclear genes coding the yeast mitochondrial adenosine triphosphatase complex. Primary sequence analysis of ATP2 encoding the F1-ATPase beta-subunit precursor. *The Journal of biological chemistry*, 260(29), 15458-65.
5. Duarte, N.C., *et al.* Reconstruction and Validation of *Saccharomyces cerevisiae* iND750, a Fully Compartmentalized Genome-scale Metabolic Model, *Genome Research*, 14: 1298-1309 (2004)

Evolution of Structure and Function of the Aminoacyl-tRNA synthetases

Sethi, Anurag, O'Donoghue, Patrick, and Luthey-Schulten, Zaida

Department of Chemistry, University of Illinois, A550 Chemical & Life Sciences Lab, Box 45-6, 600 S. Mathews Av., Urbana, IL 61801

Keywords: aminoacyl-tRNA synthetase, indirect pathway,

The evolution of structure in the aminoacyl-tRNA synthetases (AARS) has been studied using a measure of structural similarity, Q_H, and shown to be congruent with the topology of sequence-based phylogenetic analysis. We have also developed the Sequence QR and Structure QR algorithms to remove redundancy from a multiple alignment and orders the protein sequences or structures by increasing linear dependence. The minimal set of sequences and/or structures spans the evolutionary space of the homologous group of proteins and allows us to generate complete evolutionary profiles that best represents the topology of the full molecular phylogenetic tree. These structure-based profiles outperformed other sequence-based methods for finding distant homologs and were used to identify a putative class II cysteinyl-tRNA synthetase (CysRS) in some methanogenic archaea that eluded previous annotation studies. This enzyme was subsequently shown to be involved in an indirect mechanism for aminoacylation of cysteine which requires two enzymes. The enzyme, now known as SepRS, first charges $tRNA^{Cys}$ with O-phoshposerine (Sep), a precursor of the cognate amino acid, and Sep-tRNA:Cys-tRNA synthase (SepCysS) converts Sep to Cys in a tRNA dependent reaction. The phylogenetic analysis of proteins involved in the indirect and direct pathways (class I CysRS) revealed that both these pathways were as ancient as the root of the universal phylogenetic tree. In addition, we also make suggestions, based on bioinformatic analysis, and homology modeling about the active site configurations of SepRS and SepCysS, a possible reaction mechanism for SepCysS, and it's interaction with tRNA.

EcoCyc: A Model Organism Database Reaches a Curation Milestone

A. Shearer, C. Bonavides-Martinez, J. Collado-Vides, B. Contreras, S. Gama-Castro, J. Ingraham, A. Johnson, I. Keseler, M. Krummenacker, I. Martinez-Flores, S. Paley, I. Paulsen, M. Peralta-Gil, A. Santos-Zavaleta, Q. Ren, P. Karp. SRI International; Center of Genomics, National Autonomous University of Mexico; University of California, Davis; The Institute for Genomic Research.

EcoCyc, a model organism database for *Escherichia coli K-12* (www.EcoCyc.org), is a full-fledged molecular encyclopedia describing the biology of *E. coli*. As of version 10.5 (August, 2006), we have reached the significant milestone of having manually curated all *E. coli* genes. Following extensive literature searches for each gene, minireview comments were written covering all of the 3,257 *E. coli* genes for which published experimental information could be identified; these comments can be found in the EcoCyc protein and RNA pages linked to those genes. This version of EcoCyc cites 14,269 publications that form the basis of curation describing 197 metabolic and signaling pathways, 4,854 reactions including metabolic, transport and binding reactions, 1,276 enzymes, 211 transporters, 1,662 transcription units and 4,128 proteins. Our extensive coverage is the product of continuous, literature-based curation backed by evaluations and recommendations from experts on specific research areas. Since reaching the complete gene coverage milestone, we are now engaged in the continuing process of updating older entries and adding new types of information as well as new ways to use that information. EcoCyc is a comprehensive, searchable and computable pathway-genome database that is freely available in a number of formats and can be redistributed in whole or in part.

Evaluating the metabolic pathways predicted by PathoLogic: a Bayesian approach for confidence score calculation

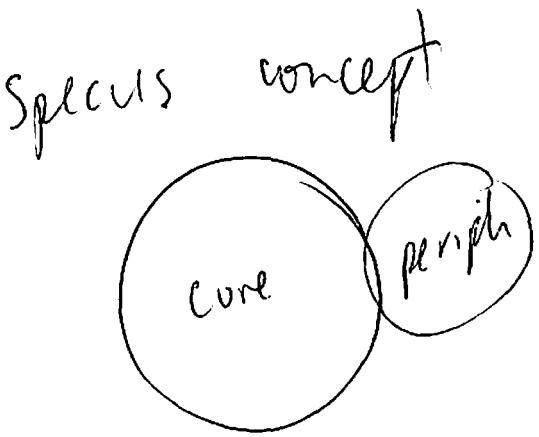
J. Shi*, M. G. Walker, P.D. Karp, S. N. Cohen. Biomedical Informatics Training Program, Stanford University.

There are approximately 200 computationally-derived pathway-genome databases in the BioCyc collection (http://www.biocyc.org) that were created using the PathoLogic software. How can we evaluate the validity of these predicted metabolic pathways? Previous research has used manual curation based on published literature and clinical tests data to validate the pathways. However, many genomes have very limited literature to start with, and the manual validation is labor-intensive. Here, we describe a Bayesian algorithm to calculate a confidence score for each predicted metabolic pathway by incorporating genome sequence, functional annotation, and gene expression data. We developed the algorithm using the literature-based pathways for *E. coli* in the EcoCyc database. The positive training data are the known pathways; the negative training data are randomly generated combinations of genes. For each pathway, we calculate a confidence score. We then evaluate the algorithm by determining its ability to distinguish the positive and negative training data (true pathways vs. randomly generated pathways). The algorithm separates the true pathways from the randomly generated (p = 5e-22, t-test). The area under the ROC curve for the algorithm is 0.95 (where the area under the ROC for a perfect classifier is 1.0 and 0.5 for a classifier that fails to separate the two groups. We then applied this algorithm to StreptoCyc, a pathway-genome database for *Streptomyces coelicolor*. We evaluated the algorithm's ability to distinguish predicted pathways in StreptoCyc from randomly generated combinations of genes. The area under the ROC curve is 0.94 and the p-value is 8e-16 (t-test) using the confidence scores to separate the predicted pathways from the randomly generated pathways in StreptoCyc.

Phylogenetics, Genomics and a Bacterial Species Concept

M.A. Riley
University of Massachusetts, Amherst.

The goal of this work is to evaluate the core genome hypothesis, which posits that there is a core set of shared genes that define a bacterial species. Although it is clear that mechanisms exist for abundant and widespread genetic transfer between microbial lineages, the observation of phenotypic clustering argues for genomic stability and cohesion. To evaluate the importance of genomic and evolutionary stability versus genomic flux, we employ population and comparative genomic methods. Such analyses suggest that, for at least *E. coli* and *S. enterica*, there is a core genome that is shared within, but not between, these two related species. If the core genome hypothesis holds for many bacterial lineages, then it may be possible to revise the existing Biological Species Concept originally proposed by Ernst Mayr such that is can be usefully applied to bacteria.



SPEAKER ABSTRACTS

On the Complexity of alpha-Proteobacterial Genomes

Siv G. E. Andersson*, Eva Berglund[1], Björn Brindefalk[1], Alexandra Calteau[1], Nam-Huyk Cho[2], Myung-Sik Choi[2], Alistair Darby, Carolin Frank[1], Hans-Henrik Fuxelius[1], Ik-Sang Kim[2], Lisa Klasson[1], Hillevi Lindroos[1], Björn Sällström[1], Johan Viklund[1], Olga Vinnere-Pettersson[1]. Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Norbyvägen 18C, 772 36 Uppsala, Swedem
[2]Department of Microbiology and Immunology, Seoul National University College of Medicine, 28 Yongon-Dong, Chongno-Gu, Seoul 110-799, Korea
E-mail: Siv.Andersson@ebc.uu.se

Metagenomics has shown that the a-proteobacteria is one of the most highly abundant bacterial subdivisions in the environment. With a ten-fold genome size range, this subdivision provides examples of the simplest as well as the most complex bacterial genomes. A hypothesis is that the large population sizes of bacteria pose barriers to the evolution of complex genomes. The availability of multiple genomes from closely related strains and species offers an excellent opportunity to re-examine hypotheses about how genome complexity relates to factors such as host range size, mode of transmission, bottleneck effects and population size. Our main model systems are intracellular bacteria of the orders Rickettsiales and Rhizobiales, all of which belong to the alpha-proteobacteria and have undergone reductive genome evolution. Members of *Rickettsia* are transmitted among hosts by blood-sucking arthropods whereas *Wolbachia* are maternally transmitted with no mammalian host stage. *Orientia* is an intermediate that can alternative between maternal inheritance and horizontal transmission. We discuss our genomic comparisons of these bacteria at the micro- and macro scales and place them in the context of the different lifestyle characteristics. Although gene loss is the dominating theme, extra-ordinary examples of horizontal gene transfer and intra-genomic expansion of conjugative transfer systems and host-interaction genes have been discovered. A genomic comparison of *Bartonella* species adapted to humans, cats and small rodents has revealed a gradual reduction in genome complexity with the successive crossing of host species barriers. We discuss the diversity of the natural population of *Bartonella henselae*, the agent of cat-scratch disease, at the whole-genome level. We suggest that evasion of the host immune response is a selective force that drives rearrangements and recombination events across and within the genomic islands. The dynamics of intracellular bacterial genomes is summarized in a model that highlights general trends in response to population size changes during the adaptation of alpha-Proteobacteria to eukaryotic host cell environments.

Systems Biology Approach to Groundwater Bioremediation and Microbial Electricity Production

Derek R. Lovley
Department of Microbiology, University of Massachusetts, Amherst, MA

Geobacter species are specialists in coupling the complete oxidation of organic compounds with extracellular electron transfer onto insoluble, extracellular electron acceptors. Molecular ecology studies have demonstrated that *Geobacter* species are the predominant organisms in many sediments and subsurface environments in which the oxidation of organic matter coupled to the reduction of Fe(III) oxides or electrodes is important. They have proven to be useful agents for the bioremediation of subsurface environments contaminated with a variety of organic or metal contaminants. Furthermore, they are by far the most effective microorganisms available in pure culture for converting organic compounds into electricity in microbial fuel cells. The genomes of multiple pure cultures of *Geobacter* species, as well as the genome sequences of uncultured *Geobacter* species living in subsurface environments, are now available as are genome-based *in silico* models of these microorganisms. Iterative experimental and *in silico* modeling studies have identified key genes for diagnosing the metabolic state of *Geobacter* species. Monitoring the *in situ* expression of these genes during the bioremediation of uranium-contaminated groundwater has revealed important environmental stresses and nutrient limitations that if alleviated with the proper amendments to groundwater may enhance the uranium bioremediation process. A systems biology approach has also been taken to understanding and improving electricity production. Whole-genome analysis of gene expression of *Geobacter* species harvesting electricity coupled with genetic studies have revealed the likely electrical contacts between *Geobacter* and the electrode surface and have guided recent studies that have substantially increased the power output of microbial fuel cells. The *in silico Geobacter* model has guided reengineering of the central metabolism of *Geobacter* to promote increased respiration and current production. Directed evolution studies have improved both power production and rates of metal reduction and one key mutation associated with this has already been identified.

Title: Getting Down and Dirty with Antibiotic Resistance: Cloning the Soil Resistome

Zakee Sabree[1], Heather Allen[1], Lakita Hopkins[2] and Jo Handelsman[1]
Department of Plant Pathology, UW-Madison, Madison WI 53706; Department of Biology, Dillard University, 70112

Antibiotic resistance has been studied in bacteria isolated from clinical and agricultural sites, yet little is known about the prevalence, diversity, or spread of antibiotic resistance determinants in unperturbed sites. To explore antibiotic resistance in such a site, we built metagenomic libraries from soil samples taken from a pristine site -- floodplain islands in the Tanana River near Fairbanks, Alaska. The libraries were screened in an *E. coli* host that contained a cloned copy of the gene encoding the major sigma factor from *Acidobacterium capsulatum,* a member of the phylum Acidobacterium, which is highly abundant in soil but largely unknown in the laboratory due to its members' recalcitrance to culturing. Over 350,000 clones from large-insert fosmid libraries were subjected to selection with eight beta-lactams, and based on restriction digest analyses, >200 unique clones were found. In some of these clones, the resistance phenotype appears to be *A. capsulatum* RpoD-dependent. These results are intriguing because of their implications about the soil as a reservoir of antibiotic resistance genes and for the approach of using a transcription factor to expand the range of genes from anonymous organisms that are expressed in *E. coli.*

The UCLA Undergraduate Genomics Research Initiative:
A Combined Laboratory Course Effort to Sequence a Microbial Genome

Cheryl A. Kerfeld* & Robert W. Simons
University of California, Los Angeles

The UCLA Undergraduate Genomics Research Initiative (UGRI, http://www.lsic.ucla.edu/ugri/) is a collaboration among life science students (at UCLA and at partner institutions) to sequence a microbial genome. The program involves nearly 2000 students annually. The participating courses are united by a novel research methods course, "Principles and Practices of Genomic Research." This course is interdisciplinary; exploring diverse concepts unified by the genome project, including topics in physics, bioinformatics biochemistry, evolution, microbiology. Furthermore, students in the course have the opportunity to master techniques and concepts while learning to work as a team. Results over the past three years of the UGRI show that doing real research motivates students; they report gains in their learning and their interest in and enthusiasm for life sciences research. The UGRI model for collaborative undergraduate research can be adopted at a range of institution types to provide research experience for a large number of undergraduates.

Amnonifex degensii

LS167
-principles + practice of genome research

Undergraduate Research in Microbial Molecular Biology, Ecology, Diversity, and Evolution – Making Individual Phylogenomic Trees in a Laboratory Course.

E.R. Sanders-Lorenz*. University of California, Los Angeles.

Changes in biotechnology are driving a nationwide research initiative to transform undergraduate science education, in which the goal is to devise teaching methods that generate excitement about the discovery process central to scientific research. To incorporate this initiative into its curriculum, the Department of Microbiology, Immunology, and Molecular Genetics (MIMG) launched a project-based laboratory course entitled 'Undergraduate Research in Microbial Molecular Biology, Ecology, Diversity, and Evolution' in which students explored microbial diversity within environmental samples. This course, which has the potential to be precedent-setting in terms of its educational impact, was devised using the protocols from the research program termed "I, Microbiologist" developed by Professor Jeffery H. Miller and his associate Krystle Ziebell. In launching the course, we explored the feasibility of converting the "I, Microbiologist" research experience into a formal classroom laboratory setting. Like the participants of "I, Microbiologist", students in the course contributed to cutting edge research and experienced the thrill of discovery within only a few weeks of starting their projects. Students each implemented a discovery-based research project, reconstructing the phylogeny of a unique soil bacterial community based on the analysis of SSU rRNA genes. The lab was complemented by a seminar-style didactic module, which covered exciting topics in current microbiological literature. A subset of students enrolled in an honors section, where they participated as collaborators in a bilateral interface with the genomic biology course LS 187, in which honor students learned DNA sequencing techniques. Since we were interested in whether the overall course format affected student impressions about research, their knowledge of lab techniques, and their interest in scientific research careers, an evaluation was conducted throughout the term by way of anonymous student questionnaires. Altogether, our efforts resulted in the successful adaptation of the "I, Microbiologist" research program into the MIMG laboratory curriculum.

8/10/2010

From Genome-Wide Reporter Gene Arrays to Gene Function Discovery to Biosensor Development

Tina Van Dyk, DuPont Company, Wilmington, DE.

A genome-wide bioluminescent reporter gene array, LuxArray, was developed for *Escherichia coli* using a collection of 8000 random fusions of *Escherichia coli* genomic DNA to a *Photorhabdus luminescens luxCDABE* reporter. The DNA sequence of each of junction was used to precisely map these fusions, from which a non-redundant group of 849 *luxCDABE* gene fusions was selected. This highly parallel cell-based platform for gene expression analysis can be used on agar plates or in liquid medium and provides advantages and synergies with DNA array analyses. The LuxArray in combination with other genome-wide tools in *E. coli* enabled facile gene function discovery for an efflux transport system and regulatory protein. Furthermore, as bioluminescence is an ideal signal for whole cell biosensors, rapid development of sensors for specific responses is possible. One example is a biosensor for oxygen depletion in an integrated microbioreactor.

"The *Vibrio harveyi* Quorum-Sensing System Uses Shared Regulatory Components to Discriminate Between Multiple Autoinducers"

Christopher M. Waters and Bonnie L. Bassler
Princeton University, Department of Molecular Biology
Princeton, NJ

Bacteria communicate using chemical signals known as autoinducers (AIs) in a process called quorum sensing (QS). *Vibrio harveyi* possesses a well defined QS system that integrates information from three distinct AIs (HAI-1, CAI-1, and AI-2) into a common phosphorelay pathway. Interestingly, these three AIs show marked differences in relatedness to *V. harveyi* of the species which produce them: HAI-1 is only made by *V. harveyi*, CAI-1 is made by many vibrios, and AI-2 is made by over half of all sequenced bacteria. To determine how *V. harveyi* integrates the information from these three AIs, 50 AI regulated promoters from *V. harveyi* were identified using a FACS based screen. Analysis of these promoters in different AI input conditions revealed distinct classes of responses. Whereas some genes required all AIs for significant regulation, others showed regulation to single AI input states. Additionally, the Qrr sRNAs and LuxR, key regulatory intermediates of the *V. harveyi* QS circuit, responded in a graded fashion to different AI input states. These data suggest that *V. harveyi* utilizes the information contained by these three AIs in the extracellular environment to initiate distinct patterns of gene expression in response to differences in the relatedness of the surrounding community.

Membrane Vesicles Traffic Signals and Facilitate Group Activity in *Pseudomonas aeruginosa*

Marvin Whiteley
University of Texas, Austin

Many bacteria use extracellular signals to communicate and coordinate social activities, a process referred to as quorum sensing. Many quorum signals have significant hydrophobic character, and how these signals are trafficked between bacteria within a population is not understood. Here we show that the opportunistic human pathogen *Pseudomonas aeruginosa* packages the signaling molecule 2-heptyl-3-hydroxy-4-quinolone (pseudomonas quinolone signal; PQS) into membrane vesicles that serve to traffic this molecule within a population. Removal of these vesicles from the bacterial population halts cell-cell communication and inhibits PQS-controlled group behavior. We also show that PQS actively mediates its own packaging and the packaging of other antimicrobial quinolines produced by *P. aeruginosa* into vesicles. These findings illustrate that a prokaryote possesses a signal trafficking system with features common to those used by higher organisms and outlines a novel mechanism for delivery of a signal critical for coordinating group behavior in *P. aeruginosa*.

Spatiometabolic stratification of *Shewanella oneidensis* biofilms

T.K. Teal*, D.P. Lies, B.J. Wold & D.K. Newman
California Institute of Technology.

Biofilms, or surface-attached microbial communities, are both ubiquitous and resilient in the environment. Although genetic mechanisms of biofilm growth, development and detachment have been studied, little is understood about metabolism and the spatiotemporal stratification of gene expression within biofilms. Studies utilizing the LIVE/DEAD stain suggest that large sub-populations of cells within biofilms are dead. We hypothesized that these cells may instead persist and retain the capacity to dynamically regulate their metabolism. To test this, we measured the expression of genes involved in growth activity and anaerobic metabolism, imaging fluorescent protein expression from gene reporter constructs over the course of biofilm development. Quantitative analysis of these images revealed remarkably consistent patterns of growth activity and metabolism during development, dependent on the size of the biofilm colony. Within domains of growth inactive cells, genes typically upregulated under anaerobic conditions are expressed well after growth has ceased. These findings reveal that, far from being dead, the majority of cells in mature *S. oneidensis* biofilms have actively turned on metabolic programs appropriate to their local microenvironment and developmental stage.

First Genomic View of "The Missing Link" in Anaerobic Food Chains: *Syntrophus aciditrophicus*

R.P. Gunsalus‡ †, L. Rohlin‡, H. Moutakki†, U. Kim‡, R. Krupp‡, L. Rio-Hernandez†, J. Sieber†, and C. Struchtemyer†, A. Bhattacharyya§, J.W. Campbell§, and M. McInerney †

‡Dept. Microbiology, Immunology, and Molecular Genetics, UCLA, §Integrated Genomics, Inc, Chicago, †Dept. Botany and Microbiology, U. Oklahoma, Norman

Under anaerobic conditions in natural environments, a wide range of organic compounds including alcohols, fatty acids, aromatic acids, organic acids such as lactate and glycolate, many amino acids, sugars, and hydrocarbons are degraded syntrophically. The ubiquity of syntrophic metabolism emphasizes that metabolic cooperation among microbial species is required for complete destruction of organic matter and that the catalytic unit of anaerobic metabolism is a consortium. Many syntrophic associations are highly organized, multicellular structures where the partners are in close physical proximity to each other, little is known about the molecular mechanisms involved in the formation and maintenance of these catalytic units.

The completed genome sequence of *Syntrophus aciditrophicus* SB, a fatty and aromatic acid-degrading syntrophic bacterium, provides the first glimpse of the composition and architecture of this class of anaerobic microbes. It appears to posses unique electron transfer and energy-transducing systems needed to exist on marginal energy economies of a syntrophic lifestyle. The 3.18 MB genome contains 3,169 genes of which 1,618 genes (51%) were assigned functions. They were organized into pathways where most all pathways considered essential for viability of a typical Gram-negative organism were detected. Genomic analysis also suggests novel approaches to degrade fatty and aromatic acids. A distinctive feature of syntrophic metabolism is the need for reverse electron transfer, and the presence of a Na^+-translocating electron transfer complex, menaquinone, and membrane-bound Fe-S proteins suggest a mechanism to accomplish this task. The genome predicts the potential to generate ion gradients by multiple H^+ or Na^+ - translocating ATP synthetases and pyrophosphates, glutaconyl-CoA decarboxylase, and a formate cycle involving periplasmic and cytosolic formate dehydrogenases. The genome sequence of *S. aciditrophicus* SB reveals the essential features required for growth with fatty and aromatic acids by a syntrophic lifestyle.

The genomic data also provide insight towards defining the orgin and roles of many genes being revealed by environmental DNA sequencing projects.

Genomics of Ethanol Producing Bacteria

Jizhong Zhou[1], C. L. Hemme[1,2], Q. He[3], M. W. Fields[4], E. H. Saunders[5], M. L. Land[2], L. J. Hauser[2], A. L. Lapidus[6], C. S. Han[5], Zhiguo Fang[1], Yoolbong Sung[1], Zhili He[1], D. C. Bruce[5], and P. Richardson[6]

[1]The Institute for Environmental Genomics, University of Oklahoma, Norman, OK, [2]Oak Ridge National Laboratory, Oak Ridge, TN, [3]Temple University, Philadelphia, PA, [4]Miami University, Oxford, OH, [5]Los Alamos National Laboratory, Los Alamos, NM, [6]DOE Joint Genome Institute, Walnut Creek, CA.

Recent volatility in the global energy markets has sparked renewed interest in the development of alternative fuel technologies to reduce American reliance on fossil fuels. One promising technology that has generated much interest is the generation of bio-produced ethanol from cellulosic materials such as switch grass. To this end, the genomes of three strains of ethanol-producing bacteria *Thermoanaerobacter ethanolicus* 39E, *Th. ethanolicus* X514 and *Clostridium cellulolyticum* H10 have been sequenced. Strain 39E was originally isolated from thermal springs in Yellowstone National Park while the metal-reducing strain X514 was obtained from a deep subsurface location geographically isolated for ~200 mya. *C. cellulolyticum* was isolated from decaying grass. The preliminary annotations for these strains were compared to the annotations for the completed genomes of *T. tengcongensis* (*Thermoanaerobacter* strains) and *C. thermocellum* (*C. cellulolyticum*). All species studied encode systems for the breakdown of glycogen and xylan, but only the *Clostridium* strains encode genes for cellulase, which is confirmed by the inability of all *Thermoanaerobacter* strains to utilize insoluble cellulosic materials. A _-xylosidase gene implicated in degradation of xylan and hemicellulose was previously cloned from *T. ethanolicus* JW200 and this gene was found to be highly conserved in strain 39E but significantly diverged in strain X514. The *T. ethanolicus* strains also encode components of the Entner-Doudroff pathway which are lacking in *T. tengcongensis*. Interestingly, strain X514 lacks a methylglyoxal shunt that is present in the other two *Thermoanaerobacter* species, which may explain the observed higher ethanol fermentation efficiency in cultures of strain X514 supplemented with excess carbohydrate substrates, as compared to cultures of other strains tested in our laboratory. In many fermenting bacteria, the methylglyoxal shunt is employed to limit energy production when carbohydrates are plentiful. Thus, the loss of the shunt in X514 may represent an adaptation of the strain to the presumably energy-poor conditions of the deep subsurface. On the other hand, the ability of strain X514 to sustain high fermentation rates at high carbohydrate concentrations is advantageous for efficient ethanol fermentation processes, where high substrate concentrations are highly desirable to facilitate process control. The preliminary results suggest that all three organisms utilize similar carbon metabolism pathways but may have evolved differences in the carbon metabolism and energy flux pathways in response to particular environmental pressures. Finally, the *Thermoanaerobacter* strains utilize a novel bifunctional 2°-alcohol dehydrogenase to catalyze the terminal step in ethanol production.

POSTER ABSTRACTS

Analysis of the Fitness Landscape of adaptively-evolved *Escherichia coli* strains using Competition Experiments

M.K. Applebee* & B.Ø. Palsson
Department of Bioengineering, University of California San Diego, La Jolla, CA, U.S.A.

The relative fitness among a set of five *Escherichia coli* strains adaptively-evolved for growth on M9-glycerol was assessed by direct competition. Previous work [1] demonstrated that these strains all represent endpoint adaptations to the same environment. However, each strain harbors a unique set of mutations, and they have different growth-rates, indicating they are not of equal fitness. Competition experiments were employed to explore this observation. These competitions combined cultures of two strains in a 1:1 population ratio, and continued the mixed culture for 3-5 days using passaging and aerobic growth at 30 C [2]. The population frequency of daily samples was measured using Allele Frequency Estimation [1,2], which allows the measured mutation to serve as the strain identifier. Results quantified the relative fitness among the endpoint-strains in terms of the rate of fixation, and the resulting hierarchy mirrored increasing growth-rates among the strains. Two genes (*GlpK* & *RpoB/C*) acquired mutations in multiple endpoint strains. Which mutation to a gene gave the greatest fitness-advantage was determined by competitions between constructed strains carrying the mutations individually. Interestingly, the most beneficial individual mutations were not from the fittest endpoint strains. Additionally, when compared to biochemical-activity measurements of the mutant GlpK enzymes [2], the relative fitness of neither the endpoint-strains nor the individual GlpK-mutant strains correlated to increasing V_{max}. These results all suggest that the combined effect of accumulated mutations is more crucial than their individual impact. Further, competitions between strains harboring multiple mutations from an endpoint illustrated a fitness advantage in two cases (*DapF* mutation in strain GE; *MurE* mutation in strain GD) where individually the mutation produced no significant growth-rate increase. Current work involves combining mutations from different endpoint strains, to further explore how the genomic background determines a mutation's affect on fitness. Further work may extend this process to new strains, including adaptations to other environments.

1. Herring CD, Raghunathan A, Honisch C, et. al. "Comparative genome sequencing of *Escherichea coli* allows bacterial evolution to be observed on a laboratory timescale." *Submitted* (Under Review)
2. Honisch C, Raghunathan A, Cantor CR, Palsson B Ø, van den Boom D. "High-throughput mutation detection underlying adaptive evolution of *Escherichea coli*-K12." *Genome Research* 14:2495-2502, 2004

Nitrogen Cycle Functional Metagenomics and Diversity in a Marine Microbial Community of the Yucatan's Channel

German Bonilla, Luisa Falcon, Luis Eguiarte and Valeria Souza
Instituto de Ecología UNAM, Dept. Ecologia Evolutiva, CU AP 70-275, Coyoacan 04510Mexico DF.

Biochemical cycles are performed mainly by bacteria, most of which remain unknown to science. We analyzed the functional potential of a marine bacterial community to perform a theoretical nitrogen cycle. We adopted a metagenomic approach that included shotgun sequencing and PCR amplification of certain key genes. 16SrRNA revealed the abundance of Gammaproteobacteria, while Cyanobacteria resulted underrepresented un shotgun database. Comparison of shotgun, PCR and TRFLP's data points towards a meaningful bias in each method. Main inorganic phases of the cycle were absent, including nitrification, denitrification and ammonification. Surprisingly, nitrogen fixing *nifH* genes were also absent from shotgun database and appeared only with PCR. Nitrogenases belonged to Cyanobacteria, particularly to genuses *Trichodesmium* and *Cyanothece*. NH4 incorporation and flux are well represented, while exit phases, depicted by the degradation of arginine into urea and anaerobic ammonia oxidation, are absent. Hence, the system does not seems to be limited by nitrogen though it posess a well developed nitrogen recycling and reusing phase. The community prefers NAAP's over chitin as high-molecular weight nitrogen sources, which suggests the compartmentalization of the cycle along the water column. When compared against Sargasso's database, we did not find any significant difference between the relative frequencies of the cycle even though there is a great difference in the depth coverage between the two databases. This suggests the homogeneity of the functional capacities amongst marine bacterial communities in oligotrophic, subtropical oceanic surface waters.

The Utilization of Optical Maps in Whole Genome Finishing

Adam Briska*[1], John Henkhaus[1], Emily Zentz[1], Nancy Miller[2], Brad Barbazuk[3], Craig Darby[4], Heidi Goodrich-Blair[5], Zijin Du[2], Steve Forst[6], Barry Goldman[2], Brad Goodner[7], Stacie Norton[2], Steve Slater[8]
[1]OpGen Technologies; [2]Monsanto Company; [3]Donald Danforth Plant Science Center; [4]University of Alabama; [5]University of Wisconsin-Madison; [6]University of Wisconsin-Milwaukee; [7]Hiram College, [8]Arizona State University

Whole-genome sequencing strategies typically involve bidirectional shotgun sequencing of variable-sized insert libraries followed by contig ordering and gap closure. The latter stages remain labor intensive and costly and are complicated by repetitive sequences as well as "unclonable" genomic segments. Optical Mapping – a technique for rapidly generating whole-genome, ordered restriction maps from single DNA molecules – addresses these issues by providing high-resolution chromosome scaffolds to which contigs are anchored and sorted for rapid genome finishing. By ordering contigs, sequence gaps are readily identified. Importantly, comparison of sequence contigs to Optical Maps finds misassemblies and other structural anomalies which obfuscate sequence finishing. In this study we present the Optical Maps of two *Xenorhabdus* species, *X. boviennii* and *X. nematophila,* which were used to validate the sequence completed sequence of *X. boviennii* and expedite gap closure and finishing in *X. nematophila*. *X. boviennii* was finished by traditional methods and then compared to the Optical Map. Two major misjoins were identified in the sequence assembly at the location of large sequence repeats. The Optical Map for *X. nematophila* was applied at early stage finishing. A massively missambled contig was identified and resolved into 4 minor contigs. In total, twenty five sequence contigs were anchored to the Optical Map and then rapidly closed with traditional approaches. The application of the Optical Map for *X. nematophila* reduced the time, reagent costs and FTE requirements when compared to *X. boviennii*. The Optical Map also provided a sequence-independent means to validate the completed genomes for each species.

Mapping microbial biodiversity using phylogenetic approaches.

J. Bryant*, B. Enquist & J. Green. University of California, Merced.

A fundamental characteristic of mountain ecosystems is a change in climate and biota with elevation. Observed changes in species distributions with elevation provide insight into the mechanisms that create and maintain biodiversity along with providing baseline information to gauge the potential effects of climate change on biological diversity. Although microorganisms comprise much of the Earth's biodiversity and play critical roles in ecosystem functioning, little is currently known about microbial biodiversity along elevation gradients. We use ecological and phylogenetic approaches to explore how soil microbes are distributed across an elevation gradient in the Colorado Rockies, using Acidobacteria as a target group. Three soil cores were taken at five sites across an elevation gradient yielding a total of 15 samples. 16s clone libraries were constructed and sequenced with DNA extracted from each soil sample. We examined how biodiversity is phylogenetically structured in an environment by doing a pair wise comparison of: 1) species turnover at multiple taxonomic resolutions and 2) P-test results implemented with TreeClimber. Species turnover is a measure of the dissimilarity in species composition between two soil samples. The P-test uses parsimony based methods to measure ancestor community relatedness between samples. Our results show that Acidobacteria turnover increased with increased geographic distance. The slope of the turnover versus geographic distance did not change significantly between different taxonomic resolutions. This suggests that species turnover patterns are maintained at multiple levels of the phylogenetic tree. P-test results revealed that ancestral community relatedness significantly decreased with increasing distance between samples. These results are consistent with the turnover results and show that samples geographically closer to one another are more phylogenetically related than samples separated by large distances.

The complete genome sequence of Bartonella grahamii.

Alexandra Calteau*, Eva Berglund, Carolin Frank, Olga Vinnere Pettersson, Ann-Sofie Eriksson, Kristina Näslund and Siv Andersson. Department of Molecular Evolution, Evolutionary Biology Centre, Uppsala University, Sweden

Bartonella bacteria live in the bloodstream of mammals and are transmitted among their hosts by blood-sucking arthropods. The two human pathogens species, *B. henselae* and *B. quintana,* have already been sequenced. We are currently working on the finishing and the annotation of the complete genome sequence of *B. grahamii.* This species, naturally found within rodent species, has rarely been associated with human disease. Its genome is larger than those of the two human pathogens. We have compared the genomes of the three *Bartonella* species. It appears that no genes are unique to *B. henselae* and *B. quintana.* Moreover genes in these species considered to be of importance for virulence are present in even higher copy numbers in *B. grahamii.* Since *B. grahamii* harbors all putative virulence determinants present in the clinical isolates it probably has the potential to evolve into a human pathogen.

RegTransBase (RTB) - a database of regulatory sequences and interactions in prokaryotic genomes.

Michael Cipriano
Lawrence Berkeley National Laboratory

In the studies on bacterial regulation the final decision of whether to include each putative site in a particular regulon is made after detailed inspection and consultation with relevant scientific literature by a human expert. RegTransBase (RTB), a manually curated database of regulatory interactions, captures the knowledge in published scientific literature using a controlled vocabulary. Although a number of databases describing interactions between regulatory proteins and their binding sites are currently being maintained, they focus mostly on the model organisms E.coli and B.subtilis, or are entirely computationally derived. RTB describes a large number of regulatory interactions reported in many organisms and contains the following types of experimental data: investigating the activation or repression of a gene's (or operon's) transcription by an identified direct regulator; regulation of the gene's (or operon's) expression on the post-transcriptional level; mapping of a promoter or terminator; characterization of an operons' structure (co-transcription, complementation etc.); determining the transcriptional regulatory function of a protein (or RNA) directly binding to DNA (RNA); mapping of the binding site of a regulatory protein; characterization of a regulatory mutation; prediction of the binding sites of a regulatory protein, and others. Currently, RTB content is derived from ~3000 relevant articles describing over 7000 experiments in relation to 128 microbes. It contains data on the regulation of ~7500 genes and evidence for ~6500 interactions with ~650 regulators. RTB contains over 160 manually created position weight matrices (PWM) that span over 60 species and represent 27 families of organisms. After extensive testing, these PWMs will be used to search for new regulatory sites in related organisms.
An intuitive user interface will make this knowledge freely accessible to the larger microbiological research community. We will present examples of using the database to answer questions on regulation in prokaryotes.
RTB is a collaborative project with the group of Mikhail Gelfand at IITP, Moscow

Aligning genomes with lineage-specific rates of rearrangement and gene flux

Aaron E. Darling*, Bob Mau, and Nicole T. Perna. University of Wisconsin-Madison.

Genomes evolve via large-scale and local mutational processes that include nucleotide substitution and indels in addition to rearrangement and gene flux: lateral acquisition, gene duplication, and loss. The relative contribution made by each mutation type to evolution of microbial genomes has yet to be described in a unified framework. Through application of existing genome comparison methods [1] we elucidate the relative rates of various mutational processes in a group of sequenced Enterobacteriacae. Pairwise comparisons indicate that genomic rearrangements may be observed much more frequently in some lineages than in other lineages [2]. Given strong evidence for lineage-specific mutation rates, we develop a new method for multiple genome alignment that accounts for variable mutation rates when scoring candidate alignments. At a high level, the algorithm consists of four steps: (1) generation of high-scoring local-multiple-alignments, (2) estimation of pairwise breakpoint and genome content distances, (3) progressive multiple genome alignment according to a gene content guide tree, and (4) iterative refinement. We call the new method "Progressive Mauve."

We evaluate the quality of the resulting genome alignments computed on simulated data sets and draw comparison to previous methods for genome comparison [3,4]. In general, Progressive Mauve scales more gracefully than the original Mauve algorithm and accommodates large data sets with many more taxa. In particular, the new method identifies and aligns regions conserved among subsets of the genomes under study—an important improvement over the original Mauve alignment algorithm. On data sets with substantial amounts of genomic rearrangement, Progressive Mauve significantly outperforms the TBA alignment algorithm [5], reducing alignment error by 10x-400x.

Finally, we apply the new alignment algorithm to finished genomes of E. coli, Shigella, Salmonella, Yersinia, and Erwinia. We find substantial amounts of both inter- and intra-specific genomic variability. Genes annotated with some functional classes appear to be significantly over-conserved, while others are under-conserved.

[1] Darling ACE, Mau B, Blattner FR, Perna NT.(2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Research 14(7):1394-1403.
[2] Deng W, Burland V, Plunkett G 3rd, Boutin A, Mayhew GF, Liss P, Perna NT, Rose DJ, Mau B, Zhou S, Schwartz DC, Fetherston JD, Lindler LE, Brubaker RR, Plano GV, Straley SC, McDonough KA, Nilles ML, Matson JS, Blattner FR, Perry RD. (2002) Genome sequence of Yersinia pestis KIM. J Bacteriol. 184(16):4601-11.
[3] Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. (2003) Glocal alignment: finding rearrangements during alignment. Bioinformatics. 19 Suppl 1:i54-62.
[4] Bray N, Pachter L. (2004) MAVID: constrained ancestral alignment of multiple sequences. Genome Research 14(4):693-9.
[5] Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED,

Haussler D, Miller W (2004) Aligning multiple genomic sequences with the threaded blockset aligner. Genome Research. 14(4):708-15

The Construction and Use of KOFS, Knockout Fusion Screens for Use with the E. coli KO collection.

Jie Shao, Barry Wanner, Purdue University, West Lafayette, IN
Cindy Tamae, Katherine Kim, Tamar Sardarian, Anne Liu, Martha Pastuszka, Hanjing Yang, Bindu Patel, and Jeffrey H. Miller, University of California, Los Angeles, CA

We have constructed and employed transcriptional fusions to screen the E. coli complete knockout collection that consists of close to 4,000 deletion strains. We have constructed 20 fusions that place the lacZ gene under the control of different promoters involved in DNA replication, repair, and recombination that are carried on conjugatable plasmids. We have developed high throughput screening methods that allow us to easily screen the entire collection with each fusion, to look for deletion/KO strains that affect the control of these genes. The data from this type of approach complements that from microarrays, in that here we are looking at a selected set of genes (the recombination and repair genes) and monitoring their response to 4,000 different conditions (the strains with each gene, in series, deleted).

Accelerating the Reconstruction of Genome-scale Metabolic Networks

Richard A Notebaart[1], Frank HJ van Enckevort[1,2], Christof Francke[1,3], Roland J Siezen*[1,2,3] and Bas Teusink[1,2,3]
[1]Center for Molecular and Biomolecular Informatics, Radboud University Nijmegen, P.O.Box 9010, 6500GL Nijmegen, The Netherlands; [2]NIZO food research, P.O.Box 20, 6710BA, Ede, The Netherlands; [3]Wageningen Center for Food Sciences, P.O.Box 557, 6700AN Wageningen, The Netherlands

The genomic information of a species allows for the genome-scale reconstruction of its metabolic capacity. Such a metabolic reconstruction gives support to metabolic engineering, but also to integrative bioinformatics and visualization. Sequence-based automatic reconstructions require extensive manual curation, which can be very time-consuming. Therefore, we present a method to accelerate the time-consuming process of network reconstruction for a query species. The method exploits the availability of well-curated metabolic networks and uses high-resolution predictions of gene equivalency between species, allowing the transfer of gene-reaction associations from curated networks.

We have evaluated the method using *Lactococcus lactis* IL1403, for which a genome-scale metabolic network was published recently. We recovered most of the gene-reaction associations (i.e. 74 – 85%) which are incorporated in the published network. Moreover, we predicted over 200 additional genes to be associated to reactions, including genes with unknown function, genes for transporters and genes with specific metabolic reactions, which are good candidates for an extension to the previously published network. In a comparison of our developed method with the well-established approach Pathologic, we predicted 186 additional genes to be associated to reactions. We also predicted a relatively high number of complete conserved protein complexes, which are derived from curated metabolic networks, illustrating the potential predictive power of our method for protein complexes.
We show that our methodology can be applied to accelerate the reconstruction of genome-scale metabolic networks by taking optimal advantage of existing, manually curated networks. As orthology detection is the first step in the method, only the translated open reading frames (ORFs) of a newly sequenced genome are necessary to reconstruct a metabolic network. When more manually curated metabolic networks will become available in the near future, the usefulness of our method in network prediction is likely to increase.

Microbial Genome Finishing Techniques Employed at the Joint Genome Institute - Los Alamos National Laboratory (JGI-LANL).

D. R. Sims*, O. Chertkov, A.C. Munk, H. Kiss, E. Saunders, L. S. Thompson, J. Brainard, and C.S. Han. Los Alamos National Laboratory, Los Alamos, NM.

The mission of the JGI is to provide integrated high-throughput sequencing and computational analysis to enable genomic-scale/systems-based scientific approaches to DOE-relevant challenges in energy and the environment. In order to contribute to the accomplishment of this mission, JGI-LANL continues to refine methods to produce finished microbial genome sequences from drafted sequences. Upon receipt of a drafted genome from JGI-Walnut Creek, a LANL-developed program, dupFinisher [available from han_cliff@lanl.gov], uses paired reads to resolve repetitive regions. When resolved, dupFinisher creates a fake read spanning the repeat. Additionally, dupFinisher selects primers on appropriate clones designed to bridge the unresolved repeats. Two automated cycles of lanlAutoFinish are then run to choose primer walks using a variety of chemistries designed to close captured gaps, extend contigs and improve read quality. When a project has been resolved to no more than six scaffolds, PCR primers near the end of each scaffold are paired in all possible combinations to close uncaptured gaps. In some cases, 454 reads are used to align scaffold gaps. Hard stops, duplications and misassemblies are the most common challenges to the finishing process, each requiring a different approach. For hard stops, finishers take a graduated approach from least to most expensive: DMSO, dGTP, commercial kit [Amersham Biosciences, PN# 25-6401-01], an outside entity, and shatter libraries. Transposon bombs are used to solve complex duplications by drafting a clone that spans the duplication of interest. Consed is the main computer aids in identifying and resolving misassemblies. Quality control of finished genomes is conducted at JGI-Stanford University, and the genome is annotated at JGI-Oak Ridge. JGI-LANL has finished over 50 microbial genomes using these methods, representing approximately 250 Mb of microbial DNA sequence with an average error rate nearing 0.0 per 10,000 bases. These methods continue to evolve, providing an efficient method for finishing bacterial genomes.

Interpreting Metagenomic Data Using Oligonucleotide Signatures

David Soergel (1), David Tulga (2), and Steven Brenner (1,2,3)
(1) Biophysics Graduate Group,
(2) Department of Bioengineering, and
(3) Department of Plant and Microbial Biology, University of California, Berkeley

Environmental shotgun sequences are accruing at an ever-increasing rate and promise to revolutionize our understanding of microbial community ecology. However, we presently lack many of the computational methods needed to interpret these data.

We are developing algorithms for understanding metagenomic data using modern methods in computational statistics, with emphasis on use of "genomic signatures" to identify the species origin of sequence reads. Each microbial species has a surprisingly specific signature consisting of the relative frequencies of short oligonucleotides in its genome. Here, we develop and validate methods for extracting biological knowledge from distributions of oligonucleotides in metagenomic data sets. From a given sample or set of samples, these methods may provide estimates of species numbers and abundances, strain heterogeneity, phylogenetic relationships, lateral gene transfer, and changes of all of these over time and space.

In addition to helping to answer basic questions about microbial evolution, these methods will have widespread practical applications, particularly in human health, biosensing, bioremediation, and improving metagenomic sampling efficiency.

Rapid Accumulation of Genomic Heterogeneity Within Clonal Isolates of *C. jejuni*.

Eduardo N. Taboada[1*], Joanne MacKinnon[1], Christian Luebbert[2], John H. E. Nash[2], and Kris Rahn[1]. [1] Laboratory for Foodborne Zoonoses, Public Health Agency of Canada;
[2] Institute for Biological Sciences, National Research Council of Canada

Genome sequencing has revealed the remarkable amount of genetic diversity that can be encountered in bacterial genomes. In particular, the comparison of genome sequences from multiple isolates of the same species has uncovered an astonishing level of intraspecies genomic variability that is characterized by significant differences in gene content, with many hypervariable genes found clustered in genomic islands.

Our group has been involved in a large-scale study of the genome dynamics of the human enteric pathogen *Campylobacter jejuni*. We have examined the gene content of several hundred isolates by means of microarray-based comparative genomic hybridization. We have previously shown that that roughly 8-10% of the genes in the *C. jejuni* NCTC 11168 genome strain show considerable intraspecies variability and are often absent in a significant proportion of strains tested. While examining the potential for exploiting the conservation status of these genes in a molecular typing context, we have uncovered significant levels of genomic heterogeneity within groups of strains assumed to be clonal based on epidemiological and conventional molecular typing evidence.

Our results suggest that variability in the gene content within genomic islands can accumulate within relatively short time frames. As many hypervariable genes encode potential virulence determinants, these findings have significant implications in the area of risk assessment since the exact gene content of strains cannot be predicted by examining a small number of loci using conventional molecular typing approaches.

Converting protein sequences to auditory signals by assigning amino acids to musical notes

Rie Takahashi, Jeffrey H. Miller

Department of Microbiology, Immunology & Molecular Genetics and The Molecular Biology Institute, University of California, Los Angeles, California 90095

The goal of this work is to convert genome-encoded protein sequences into musical notes. Although there have been previous efforts to do this, one of the main problems has involved the awkwardness of the 20 note scale that results from a one-to-one ratio of amino acid-to-music note arrangement. This 20 note scale generates music with large jumps between two consecutive notes, sometimes referred to as "Alien music". Some other concerns include assigning rhythm, dynamics, and accompaniment according to the characteristics of the sequence. We derived a reduced 13 note scale based on hydrophobicity and pairing of like amino acids, and using three-note chords to differentiate between members of amino acid pairs. A rhythm has been encoded into the music note sequence according to the codon distribution used in the genome-encoded protein sequence, allowing each amino acid to be represented by different note durations. The result is a set of rules that produces musical compositions that can be applied to any protein sequence. The conversion will help make protein sequences more approachable and tangible for the general public, young children, and more easily accessible by the blind research population. It is also possible that auditory patterns can be used in the future to search for patterns in protein sequences that may not have been noticed otherwise. We show and allow one to listen to examples of several proteins translated into music by these methods including Thymidylate Synthase A, Lac Y Permease, and the Huntingtin protein.

Strain Variation in Mitochondrial RNA Editing of *Trypanosoma cruzi*

L.L. Isadora Trejo Martinez*, Sean Thomas, Scott J. Westenberger, David A. Campbell, and Nancy R. Sturm. University of California, Los Angeles.

Chagas disease is caused by *Trypanosoma cruzi*, a single-celled, eukaryotic parasite of the Order Kinetoplastida. This intracellular parasite gives rise to digestive tract and heart malfunction, however only 30% of infected individuals develop the symptoms of chronic disease several decades after exposure. Chagas disease affects nearly 18 million people in South and Central America and there is no vaccine or therapy to combat it. Using information from the whole-genome shotgun sequencing of *T. cruzi* strains CL Brener and Esmeraldo, we have compiled two mitochondrial maxicircle DNA sequences, the equivalent of our mitochondrial DNA. The mitochondrial genome is composed of a few dozen maxicircles (~25 kb), interwoven in a complex network among tens of thousands of catenated minicircles (1.42 kb). Maxicircle mRNA transcripts undergo a unique form of RNA editing that creates start codons, resolves frameshifts, and, in the most extreme cases, reveals entire open reading frames *via* directed uridine insertions and deletions. The CL Brener and Esmeraldo sequence analysis indicated that 12 edited genes are present within each isolate, similar to *Trypanosoma brucei*, the causative agent of African Sleeping Sickness. We have characterized RNA editing events of the cytochrome oxidase subunit III mRNA in the CL Brener strain. Based on this sequence, we have identified minicircle- and maxicircle-encoded guide RNAs that provide the specific directions for editing and can examine guide RNA strain variability. This information can be compared with RNA editing events in other trypanosomes, allowing an evaluation of the origins and potential benefits (or disadvantages) of the process.

Characterization of Large-Insert DNA Libraries for Comparative Community Genomics of the Human Gut Microbiota.

Parag A.Vaishampayan*[1], J. Froula[1], Jenna Morgan[1], Howard Ochman[2], M. Pilar Francino[1]. [1] Evolutionary Genomics Department, DOE Joint Genome Institute, Walnut Creek, CA. [2] Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ

The human intestinal microbiota is composed of 10^{13} to 10^{14} microorganisms whose collective genome contains at least 100 times as many genes as our own genome. The diverse community of gut microbes plays an essential role in human health, contributing to the digestive process, promoting gut maturation, modulating the immune system, and interacting with pathogens in several complex ways. Diet, mode of birth and other environmental conditions can influence colonization and development of the microbiota of an infant gut. We have initiated efforts to study the complex community structure and coding capabilities of the intestinal microbiota of the human infant and adult.

Human fecal samples from two healthy anonymous volunteers (mother and infant) were collected at the University of Arizona in 2004 and 2005, in collaboration with the laboratory of Howard Ochman. We have constructed four large-insert fosmid libraries in total covering 4 Gbp of community DNA. In a fosmid end sequencing approach including 5337 sequence tags, we have shown a shift in the microbiota in infant fecal samples. One month old infant fecal sample shows dominance of *Bacteroides* sp. and *Escherichia coli*, while eleven month old infant sample exhibits dominating *Bifidobacterium* sp. population. The maternal intestinal flora is a source of bacteria for the neonatal gut. We have analyzed the microbiota of infant's mother showing increase in *Bifidobacterium* sp. count, eleven months after delivery. The shift in maternal microbiota reflects similar changes in the infant gut flora. The comparison of random metagenome reads with available complete genomes of bacterial species will suggest diversity among the dominating genus in each library. The diversity of putative protein-encoding genes, as reflected by their distribution into different COG clusters was also studied. We have also prepared 16s rDNA libraries from the DNA used for constructing the large-insert libraries. Overall, the information on phylogenetic diversity and coding capabilities will provide an ample biological background describing the GI microbial community and will pinpoint fosmids for targeted sequencing.

Reductive Evolution as a Controlling Event in *Orthopoxvirus* Speciation and Biology.

C. Wang, R. C. Hendrickson, & E. J. Lefkowitz*, University of Alabama at Birmingham

Poxviruses are highly successful pathogens known to infect a variety of hosts. Previous studies of orthologous gene sets have established the evolutionary relationships between members of the *Poxviridae* family. But it is still not entirely clear how variations between family members arose and gave rise to current viruses that possess distinct biological properties. Using a newly developed computational analysis pipeline, the Poxvirus Genome Annotation System, we re-predicted the gene sets for all sequenced genomes in the genus *Orthopoxvirus*. Employing sequence similarity comparisons together with comparisons of syntenic gene maps, we established the relationships between the gene sets of these viruses. These techniques allowed us to unambiguously identify the gene loss/gain events that have occurred over the course of orthopoxvirus evolution. We also examined the similarities of both coding regions and regulatory elements (promoter sequences) to further delineate the subtle differences between these closely related species.

Our results demonstrate that the last common ancestor of current orthopoxvirus species probably contained a gene set resembling that of current isolates of Cowpox virus, and that this gene set represents a super-set of genes present in all other species. These species, including Variola virus and Monkeypox virus, contain gene complements that are subsets of that coded for by Cowpox virus. There are no genes coded for by any of these other species that are not also present in Cowpox virus. All other differences between Orthopoxvirus isolates can be described in terms of variations in protein sequences and variations in promoter sequences. Our results support a theory of reductive evolution in which the reduction in size of the core gene set of a putative ancestral virus, similar to modern-day Cowpox virus, played a critical role in speciation, host-range, and pathogenesis.

Metagenomic analysis of microbial symbionts in a gutless worm

Tanja Woyke[1,2], Hanno Teeling[3], Natalia N. Ivanova[1], Marcel Hunteman[3], Michael Richter[3], Frank Oliver Gloeckner[3,4], Dario Boffelli[1,2], Iain J. Anderson[1], Kerrie W. Barry[1], Harris J. Shapiro[1], Ernest Szeto[1], Nikos C. Kyrpides[1], Marc Mussmann[3], Rudolf Amann[3], Claudia Bergin[3], Caroline Ruehland[3], Edward M. Rubin[1,2,†] & Nicole Dubilier[3,†]

[1]DOE Joint Genome Institute, Walnut Creek, California 94598, USA.
[2]Lawrence Berkeley National Laboratory, Genomics Division, Berkeley, California 94720, USA.
[3]Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany.
[4]International University Bremen, 28759 Bremen, Germany.

† Correspondence

Most eukaryotes are inhabited by commensal, mutualistic, or parasitic bacteria, yet our understanding of the interactions that drive these associations is hampered by our inability to cultivate most host-associated microbes. Here, we used a metagenomic approach to describe the symbiotic microbial community in the marine oligochaete worm *Olavius algarvensis*, a host lacking a mouth, gut, and nephridia. Through shotgun sequencing and nucleotide signature binning we were able to assemble two nearly complete and two partial genomes of the four predominant symbionts. Metabolic pathway reconstruction revealed that these are sulfur-oxidizing and sulfate-reducing bacteria, all of which are capable of carbon fixation, providing the host with multiple sources of nutrition. Molecular evidence for the uptake and recycling of worm waste products by the symbionts suggests how the worm could afford to eliminate its excretory system, an adaptation unique among known free-living animals. We propose a model which describes how the versatile metabolism within this bacterial consortium provides an optimal energy supply to the host as it shuttles between the upper oxic and lower anoxic coastal sediments which it inhabits.

A Complex Bacterial Community Living in Pitcher Plant Fluid

Yogiara1*, Faculty of Biotechnology, Atma Jaya Catholic University, Indonesia
Antonius Suwantoand Maggy T. Suhartono, Bogor Agricultural University, Indonesia

Nepenthes spp. is a tropical pitcher plant that has a pitcher containing fluid as a unique habitat for microorganisms. Until recently, only few reports on the diversity of microbes living in the pitcher fluid. The purpose of this study was to compare bacterial community of several pitcher plants species. We collected fluid samples from five species of pitcher plants, i.e. *Nepenthes ampullaria, N. gymnaphora, N. mirabilis, N. rafflesiana*, and *N. reinwardtiana*. The levels of bacterial community were analyzed employing amplified ribosomal DNA analysis (ARDRA). The phylotype richness and frequency distribution or relative abundance (evenness) of the sample were investigated using Shannon-Weiner diversity indices. ARDRA analysis showed that pitcher plant fluid was found to be inhabited by a complex bacterial community. Fluid sample of *N. ampullaria* TJ1-2 from Tanjung Selor, East Kalimantan, has the highest diversity index. ARDRA results showed that bacteria occupied the fluid were very diverse and scattered among the pitcher plants species. There was no correlation of bacterial community between neither the plant species itself nor the location were the plant is grown. Our preliminary result also suggested that bacterial community of pitcher plant fluid might not derived from bacterial community in phyllosphere of pitcher plant leaf surfaces but they seem to develop the community itself simultaneously during pitcher's lid opening and might be a minor affection of phyllosphere bacteria.

Key word: *Nepenthes*, bacterial community, ARDRA

DIRECTORY

Sharon Aguila
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
Sjag25@msn.com

Timothy Allen
University of Virginia
Department of Biomedical Engineering
Box 800759, Health System
Charlottesville, VA 22908
Tel: 434/982-6751
Fax: 434/982-3870
teallen@virginia.edu

Siv Andersson
Department of Molecular Evolution
University of Uppsala
Evolutionary Biology Center
Uppsala, 752 36 Sweden
Tel: 46-18-471-4379
Siv.andersson@ebc.uu.se

Kenyon Applebee
University of California, San Diego
9500 Gilman Drive Dept. 0412
La Jolla, CA 92093-0412
Tel: 858/534-5668
Fax: 858/534-5722

Kayo Arima
PRAGMA-AIST Joint Program
University of California, San Diego
California Institute for
Telecommunication and Information
Technology
9500 Gilman Dr.
La Jolla, CA 92093
Tel: 858/822-4649
Fax: 858/822-3610
karima@ucsd.edu

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

Adam Briska
OpGen Technologies, Inc.
510 Charmany Dr., Ste 151
Madison, WI 53719
Tel: 608/441-8106
Fax: 608/441-8101
abriska@opgen.com

C. Titus Brown
Dept of Biology
Biology 156-29 Caltech
Pasadena, CA 91125
Tel: 626/395-3360
titus@caltech.edu

Jessica Bryant
School of Natural Sciences
University of California, Merced
Tel: 209/724-2953
Fax: 209/724-2912 (fax)
jbryant@ucmerced.edu

Andrew Cajigal
California State University
San Bernardino
5500 University Pkwy
San Bernardino, CA 92407
Tel: 909/537-5405
pilyoboy@hotmail.com

Alexandra Calteau
Department of Molecular Evolution
University of Uppsala
Evolutionary Biology Center
Uppsala, 752 36 Sweden
Tel: 46-18-471-4379
Alexandra.Calteau@ebc.uu.se

Gregory L. Challis
Department of Chemistry
University of Warwick
Coventry CV2 7AL
g.l.challis@warwick.ac.uk

Casey Chen
USC School of Dentistry
925 W. 34th St.
Los Angeles, CA 90089-0641
Tel: 213/740-1075
ccchen@usc.edu

Chao-Ying Chen
Department of Plant Pathology and Microbiology
National Taiwan University
No. 1 Sec. 4 Roosevelt Rd.
Taipei, Taiwan, ROC
Tel: 886-2-33665207
Fax: 886-2-23657735
cychen@ntu.edu.tw

Ohn Chow
University of California, San Diego
School of Medicine
Cellular & Molecular Medicine East, Room 1066
9500 Gilman Drive, Mail Code 0687
La Jolla, CA 92093-0687
Tel: 858/534-9760
Fax: 858/534-5611
ochow@ucsd.edu

Jackie Chui
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
jchui@ucla.edu

Michael Cipriano
Lawrence Berkeley National Laboratory
1 Cyclotron Road Mail Stop 84-171
Berkeley, CA 94720
Tel:510/486-5692
Fax:510/486-5614
mjcipriano@lbl.gov

Aaron Darling
University of Wisconsin-Madison
425G Henry Mall, Rm 4400V
Madison, WI 53706
Tel:608/890-0172
darling@cs.wisc.edu

Charity Deluca
LiveFuels
3525 Alameda de las Pulgas Suite B
Menlo Park, CA 94025
Tel: 650/234-9810
Fax: 650-332-1514
charity@livefuels.com

Greg Davis
Sigma-Aldrich Corporation
Biotechnology R&D, Genomics
2909 Laclede Ave.
St. Louis, MO 63103
Tel: 314/771-5765 x2666
gdavis@sial.com

Ashlee Earl
Harvard Medical School
Dept. of Microbiology and Molecular Genetics
200 Longwood Ave, D-1 #219
Boston, MA 02115
Tel: 617/432-4359
Fax: 617/738-7664
Ashlee_earl@hms.harvard.edu

Aled Edwards
Banting and Best Department of Medical Research
University of Toronto
100 College St. Rm351
Toronto, Ontario
Canada M5G IL5
Tel: 416/946-3795
Fax: 416/946-0880
Aled.Edwards@utoronto.ca

Luis Eguiarte
Universidad Nacional Autonoma de Mexico
Instituto de Ecologia, CU
Coyoacan 04510, Mexico DF
Tel: 52-55-5622-9006
Fax: 52-55-5622-8996
fruns@servidor.unam.mx

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

Eric Eisenstadt
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850
Tel: 301/795-7670
eeisenst@tigr.org

Adam Feist
University of California, San Diego
Department of Bioengineering
9500 Gilman Dr. Dept 0412
La Jolla, CA 92093
Tel: 858/822-3181
Fax: 858/822-3120
afeist@ucsd.edu

Karen Flummerfelt
Instructional Lab Research Associate & Teaching Assistant
Department of Microbiology, Immunology, & Molecular Genetics
University of California, Los Angeles
1602 Molecular Science Building
609 Charles E. Young Dr. East
Los Angeles, CA 90095
karenf@microbio.ucla.edu

Maria Pilar Francino
Genomics Division
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: 925/296-5872
Fax: 925/296-5666
mpfrancino@lbl.gov

Francesca Frentiu
Ecology and Evolutionary Biology
University of California Irvine
321 Steinhaus Hall
Irvine, CA 92697
Tel: 949/824-7158
Fax: 949/824-2181
ffrentiu@uci.edu

Pauline Funchain
paulinef@gmail.com

Osnat Gillor
Department of Environmental Hydrology & Microbiology
Zuckerberg Institute for Water Research
Jacob Blaustein Institutes for Desert Research
Ben Gurion University of the Negev,
Sde Boqer Campus, 84990, Israelâ*Î
gillor@nsm.umass.edu

Michael C. Gilson
University of Wisconsin-Madison
4400V Genetics-Biothechnology
Center Building
425 Henry Mall
Madison, WI 53706
Tel: 608/890-0172
Gilson@cs.wisc.edu

Enid Gonzalez
USDA, ARS, Crops Pathology/Genetics
Research Unit
259 Hutchison Hall, Plant Pathology
One Shields Ave, UC Davis
Davis, CA 95616
Tel: 530/752-1246
Fax: 530/754-7195
ettgonzalez@ucdavis.edu

Jessica L. Green
School of Natural Sciences
University of California, Merced
Tel: 209/724-2953
Fax: 209/724-2912
jgreen@ucmerced.edu

Joe Grzymski
Division of Earth and Ecosystem
Science
Desert Research Institute
2215 Raggio Parkway
Reno, Nevada 89512
Tel: 775/673-7478
Fax: 775/673-7485
joegjoeg@gmail.com

Robert Gunsalus
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Science Bldg.
Los Angeles, CA 90095
Tel: 310-825-4099
Fax: 310-206-5231
robg@microbio.ucla.edu

Brian Hammer
Department of Molecular Biology
Princeton University
Tel: 609/258-2864
Fax: 609/258-6175
bhammer@princeton.edu

Amber Hartman
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
alh@jhu.edu

Loren Hauser
Oak Ridge National Laboratory
Life Sciences Division
P.O. Box 2008
Building 5700 Room G-229
Oak Ridge, TN 37831-6164
hoi@ornl.gov

Manzour H. Hazbon
Center for Emerging Pathogens
University of Medicine and Dentistry of
New Jersey
185 South Orange Ave. MSB A920B
Newark, New Jersey 07103
Tel: 973/972-2174
Fax: 973/972-0713
hazbonmh@umdnj.edu

Zhili He
Institute for Environmental Genomics
Stephenson Research & Technology
Center
University of Oklahoma
101 David L. Boren Blvd.
Norman, OK 73072
Phone: 405.325.6073
Fax: 405-325-7552
zhe@rccc.ou.edu

John Henkhaus
OpGen Technologies, Inc.
510 Charmany Dr., Ste 151
Madison, WI 53719
Tel: 608/441-8102
Fax: 608/441-8101
jhenkhaus@opgen.com

Christopher Herring
Mascoma Corporation
161 First St.
Cambridge, Mass 02139
Tel: 617-234-0099
Fax: 617-868-0408
cherring@mascoma.com

Emma Hill
PloS Biology
185 Berry St. Suite 3100
San Francisco, CA 94107
Tel: 415/625-1231
Fax: 415/546-4090
ehill@plos.org

Les Hoffman
Epicentre Biotechnologies
726 Post Rd.
Madison, WI 53713
Tel: 602/442-6146
Fax: 602/442-6147
Les.hoffman@epibio.com

Hanna Hong
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
hanayaah@gmail.com

Christiane Honisch
Sequenom, Inc.
Fax: 858/202-9084
chonisch@sequenom.com

Kimberly Houck
University of California Los Angeles
Life Science Core Curriculum
2847 Slichter Hall
Los Angeles, CA 90095
houckimb@ucla.edu

Julie Huber
Josephine Bay Paul Center
Marine Biological Laboratory
7 MBL St.
Woods Hole, MA 02543
Tel: 508/548-3705 x 6616
Fax: 508/457-4727
jhuber@mbl.edu

Ken-ichi Inatomi
Research Institute of Innovative
Technology for the Earth
9-2 Kizugawadai, Kizu-cho Soraku-gun
Kyoto 619-0292 Japan
Tel: 81-774-75-2308
Fax: 81-774-75-2321
kinatmi@rite.or.jp

Janet Jansson
Swedish University of Agricultural
Sciences Dept. of Microbiology
Genetics Center, Box 7025
SE-75007 Uppsala, Sweden
Tel: 46-18-673201
Fax: 46-18-673393
Janet.jansson@mikrob.slu.se

Justin Johnson
J. Craig Venter Institute
9704 Medical Center Dr.
Rockville, MD 20850
Tel: 240-268-2785
Fax: 240-268-4000
jjohnson@venterinstitute.org

Dave Jones
Live Fuels, Inc.
3525 Alameda de las Pulgas Suite B
Menlo Park, CA 94025
Tel: 650/234-9810
Fax: 650-332-1514
dave@livefuels.com

Lissa Morgenthaler-Jones
Live Fuels, Inc.
3525 Alameda de las Pulgas Suite B
Menlo Park, CA 94025
Tel: 650/234-9810
Fax: 650/332-1514
lissa@livefuels.com

Andrew Joyce
University of California, San Diego
Department of Bioinformatics
9500 Gilman Dr. Dept 0412
La Jolla, CA 92093
Tel: 858/822-3181
Fax: 858/822-3120
ajoyce@ucsd.edu

Connie Kang
University of California Berkeley
717 Potter St. Bldg 977
Berkeley, CA 94710
Tel: 510/495-2627
srhconnie@berkeley.edu

Jay Keasling
Department of Chemical Engineering
University of California, Berkeley
Berkeley, CA 94720-1462
Tel: 510/495-2620
Fax: 510/495-2630
keasling@berkeley.edu

Cheryl Kerfeld
University of California Los Angeles
Life Science Core Curriculum
2847 Slichter Hall
Los Angeles, CA 90095
Tel: 310-825-7417
Fax: 310-206-3914
kerfeld@mbi.ucla.edu

Sergey Kalachikov
Columbia University Genome Center
1150 St. Nicholas Ave. Room 408
New York, NY 10032
Tel: 212/851-5193
Fax: 212/851-5215
Sk363@columbia.edu

Patricia Kale
Joint Genome Institute
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
kale1@llnl.gov

Arthur Katz
US Department of Energy
Office of Biological and Environmental Research
1000 Independence Avenue
Washington, DC 20585-1290
Tel: 301/903-4932
Arthur.katz@science.doe.gov

Katherine Kim
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
Kkim1987@yahoo.com

Sun Chang Kim
Korea Advanced Institue of Science and Technology Dept of Biological Science
373-1 Guseong-dong, Yuseong-gu.
Daejeon Republic of Korea
Tel: 82-42-869-2619
Fax: 82-42-869-2610
sunkim@kaist.ac.kr

Lance Kizer
Dept of Chemcial Engineering
University of California Berkeley
Berkeley West Biocenter
Lawrence Berkeley National Lab
717 Potter St. Room 292
Berkeley, CA 94720
lancek@berkeley.edu

Dan Kluepfel
USDA. ARS, Plant pathology
Hutchison Hall, Plant Pathology Dept
One Sields Ave, UCD
Davis, CA 95616
dafluepfel@ucdavis.edu

Michel L. Knotek
Knotek Scientific Consulting/ DOE
10127 N. Bighorn Butte Dr.
Oro Valley, AZ 85737
Tel: 520/591-8108
Fax: 520/877-3233
M.Knotek@verizon.net

Karen Kong
California State University
San Bernardino
5500 University Pkwy
San Bernardino, CA 92407
Tel: 909/537-5405
karenkong@gmail.com

Cord D. Krishner
University of California Los Angeles
Life Science Core Curriculum
2847 Slichter Hall
Los Angeles, CA 90095
krishcor@verizon.net

Victor Kunin
DOE Joint Genome Institue
2800 Mitchell Dr.
Walnut Creek, CA 94598
Tel: 925/296-5814
vkunin@lbl.gov

Raman Lall
University of Wisconsin
Bacter Institute Rm. 6630
433 Babcock Dr.
Madison, WI 53706
Tel: 608/262-6844
Fax: 608/262-3453
lall@wisc.edu

Federico Lauro
University of California San Diego
MBRD –Scripps Inst. Of Oceanography
8750 Biological Grade Rm 4305
La Jolla, CA 92037
Tel: 858/534-5283
Fax: 858/534-7313
flauro@ucsd.edu

John H. Leamon
DNA Amplification Technologies
454 Life Sciences
20 Commercial St.
Branford, CT 06405
Tel: 203/871-2346
jleamon@454.com

Jun Hyoung Lee
Korea Advanced Institue of Science and Technology Dept of Biological Science
373-1 Guseong-dong, Yuseong-gu.
Daejeon Republic of Korea
Tel: 82-42-869-2659
Fax: 82-42-869-2610
mistygate@kaist.ac.kr

Ju Young Lee
Korea Advanced Institue of Science and Technology Dept of Biological Science
373-1 Guseong-dong, Yuseong-gu.
Daejeon Republic of Korea
Tel: 82-42-869-2659
Fax: 82-42-869-2610
windbell@kaist.ac.kr

Sang Hee Lee
Korea Advanced Institue of Science and Technology Dept of Biological Science
373-1 Guseong-dong, Yuseong-gu.
Daejeon Republic of Korea
Tel: 82-42-869-2659
Fax: 82-42-869-2610
fullhouse@kaist.ac.kr

To Hang (Shela) Lee
Department of Microbiology, Immunology, & Molecular Genetics
University of California, Los Angeles
1602 Molecular Science Building
609 Charles E. Young Dr. East
Los Angeles, CA 90095
shela1984@hotmail.com

Elliot Lefkowitz
University of Alabama at Birmingham
845 19th St. South, BBRB 277
Birmingham, AL 35294-2170
Tel: 205/934-1946
Fax: 205/934-9256
ElliotL@uab.edu

Richard E. Lenski
Michigan State University
East Lansing, MI 48824
lenski@msu.edu

Haiping Li
Western Institute for Food Safety and Security
279 Cousteau Place, Suite 100
Davis, CA 95616
Tel: 530/752-9694
hpli@wifss.ucdavis.edu

Li Li
Microbiology Service, NIH
Bldg 10 Rm 2C-306
10 Center Dr.
Bethseda, MD 20892
Tel: 301/435-8735
Fax: 301/402-1886
lli@cc.nih.gov

Michelle Lizotte-Waniewski
University of Massachusetts
Morrill III, Rm. 306
Amherst, MA 01003
Tel: 413/577-2313
mlizotte@bio.umass.edu

Philip LoCascio
Life Sciences Division
Oak Ridge National Laboratory
One Bethel Valley Rd.
P.O. Box 2008
Building 5700 Room G-217
Oak Ridge, TN 37831-6164
Tel:865/574-4567
Fax: 865/576-5332
l15@ornl.gov

Athanasios Lykidis
Joint Genome Institue
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: 925/296-5842
Fax: 925/296-5850
aLykidis@lbl.gov

Derek Lovley
Department of Microbiology
University of Massachusetts
Amherst, MA 01003
Tel: 413/545-9651
Fax: 413-545-1578
dlovley@microbio.umass.edu

Zaida (Zan) Luthey-Schulten
Department of Chemistry
University of Illinois
A544 Chemical and Life Sciences Lab
600 South Mathew Ave.
Urbana, IL 61801
Tel: 217/333-3518
Fax: 217/244-3186
zan@uiuc.edu

Tiffany Maisonet
Biological Sciences
Mount St. Mary's College
12001 Chalon Road
Los Angeles, CA 90049
tiffmais@msmc.la.edu

Hector Garcia Martin
Joint Genome Institute
2800 Mitchell Dr.
Walnut Creek, CA 94598
Tel: 510/229-8329
Fax: 925/296-5720
hectorg@gmail.com

Eric Mather
Synthetic Genomics, Inc..
11149 North Torrey Pines Road,
Suite 200
La Jolla, CA 92037
Tel: 858/754-2903
Fax: 858/754-2988
Emathur@syntheticgenomics.com

Leslie Klis McNeil
National Center for Supercomputing
Applications
University of Illinois at Urbana-
Champaign
1205 W. Clark St.
Urbana, IL 61801
Tel: 217/244-0597
Fax: 217/244-2909
lkmcneil@ncsa.uiuc.edu

Michael Miller
Stanford University School of Medicine
Beckman B239
279 Campus Dr.
Stanford, CA 94035
Tel: 650/723-7026
Fax: 650/723-1399
MCMiller@stanford.edu

Jeffrey 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
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
jhmiller@microbio.ucla.edu

Jenna Morgan
Lawrence Berkeley National Laboratory
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
JLMorgan@lbl.gov

Hirotada Mori
Graduate School of Biological Sciences
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

Helene Morlon
School of Natural Sciences
University of California, Merced
Tel: 209/724-2953
Fax: 209/724-2912
hmorlon@ucmerced.edu

Irina Morozova
Columbia University Genome Center
1150 St. Nicholas Ave. Room 406
New York, NY 10032
Tel: 212/851-5162
Fax: 212/851-5215
IM198@columbia.edu

Thuy Phuong Nguyen
University of California Los Angeles
Life Science Core Curriculum
2847 Slichter Hall
Los Angeles, CA 90095
tweezie@yahoo.com

Yummy Nguyen
Baylor College of Medicine
12503 Adirondack Dr.
Houston, TX 77089
Tel: 832/824-2230
ynguyen@bcm.tmc.edu

Victor Nizet
University of California, San Diego
School of Medicine
Cellular & Molecular Medicine East, Room 1066
9500 Gilman Drive, Mail Code 0687
La Jolla, CA 92093-0687
Tel: (858) 534-7408
Fax: (858) 534-5611
vnizet@ucsd.edu

Max Oeschger
LSU Health Sciences Center
Department of Microbiology: P6-1
1901 Perdido st.
New Orleans, LA 70012
Tel: 504/599-0246
Fax: 504/598-2918
Moesch@lsuhsc.edu

Kathleen O'Neill
NCBI, NIH
Bldg 45 Rm 6AN44K
National Institute of Health
9000 Rockville Pike
Betheseda, MD 20892
Tel: 301/402-7938
koneill@ncbi.nlm.nih.gov

Paul M. 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

Sergey Pampou
Columbia University
1150 St. Nicholas Ave Rm 408
New York, NY 10032
Sp2046@columbia.edu

Yue Pan
University of Wisconsing-Madison
Department of Computer Science
6710 Medical Science Center
1300 University Ave.
Madison, WI 53706
Tel: 608/263-7627
Fax: 608/265-7916
Ypan@cs.wisc.edu

Katya Panina
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/206-9053
epanina@ucla.edu

Alex Parker
California State University
San Bernardino
5500 University Pkwy
San Bernardino, CA 92407
Tel: 909/537-5405
Mfg4rox@att.net

Julian Parkhill
The Welcome 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

Martha Pastuszka
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
marthap@ucla.edu

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

Vasiliy Portnoy
University of California, San Diego
9500 Gilman Drive Dept. 0412
La Jolla, CA 92093-0412
Tel: 858/534-5668
Fax: 858/534-5722

Gyorgy Posfai
Biological Research Center
Hungarian Academy of Science
Temesuari Krt 62
Szeged H-6726, Hungary
Tel: 36-62-599-653
Fax: 36-62-433-506
posfaigy@brc.hu

David M. Raskin
Dept. of Microbiology & Molecular Genetics
Harvard Medical School
Armenise Bldg, Rm 411
200 Longwood Ave.
Boston, MA 02115
Tel: 617/432-2355
Fax: 617/738-7664
draskin@hms.harvard.edu

Jaques Ravel
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850
Tel: 301/795-7884
Fax: 301/838-0208
jravel@tigr.org

Claudia Reich
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
1205 W. Clark St.
Urbana, IL 61801
Tel: 217/244-4729
Fax: 217/244-2909
creich@ncsa.uiuc.edu

Susanna Remold
University of Louisville
353 Belknap Research Building
2210 S. Brook St.
Louisville, KY 40208
Tel: 502-852-0960
Fax: 502-852-7214
Susanna.remold@louisville.edu

Margaret A. Riley
Department of Biology
University of Amherst
Amherst, MA 01003
Tel: 413/545-2132
riley@bio.umass.edu

Toby Richardson
Synthetic Genomics, Inc.
11149 North Torrey Pines Road, Suite 200
La Jolla, CA 92037
Tel: 858/754-2909
Fax: 858/754-2988
trichardson@syntheticgenomics.com

German Bonilla Rossa
UNAM
Instituto de Ecologia, CU, Coyoacan
04510 Mexico
Tel: 52-55-5622-9006
Fax: 52-55-5622-8995
Germanbr@gmail.com

Edward M. Rubin
DOE Joint Genome Institute
2800 Mitchell Dr. Bldg 400-404
Walnut Creek, CA 94598
Tel: 510-486-5072
emrubin@lbl.gov

Zakee Sabree
University of Wisconsin-Madison
589A Russel Laboratories
1630 Linden Dr. Rm 580
Madison, WI 53706
Tel: 608/262-8237
zls@plantpath.wisc.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

Anurag Sethi
Department of Chemistry
University of Illinois
A544 Chemical and Life Sciences Lab
600 South Mathew Ave.
Urbana, IL 61801
Tel: 217/333-3518
Fax: 217/244-3186
Sethi2@polaris.scs.uiuc.edu

Michael Seibert
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401
Tel: 303/384-6279
Fax: 303/384-6150
Mike_seibert@nrel.gov

Katie Shaefer
Biological Sciences
Mount St. Mary's College
12001 Chalon Road
Los Angeles, CA 90049
katscha@msmc.la.edu

Jie Shao
Purdue University
915 West State St.
1392 Lilly Hall
West Lafayette, IN 47907
Tel: 765/494-8034
Fax: 765/496-1189
Shao2@purdue.edu

Alexander Shearer
SRI International
333 Ravenswood Ave.
EK229
Menlo Park, CA 94025
Tel: 650/859-5932
shearer@ai.sri.com

Jing Shi
Stanford University
Biomedical Informatics Program and
Genetics Dept
MSOB X-215
251 Campus Dr.
Stanford, CA 94305
Tel: 650/387-3825
Fax: 650/725-1536
jshi@smi.stanford.edu

Roland J. Siezen
Center for Molecular and Biomolecular
Informatics
Radboud University
PO Box 9010
6500GL Nijmegen, Netherlands
Tel: 31-318-659511
Fax: 31-318-650400
Roland.siezen@nizo.nl

Catherine Silva
University of New Orleans
2143 Easter Lane
New Orleans, LA 70114
Tel: 504/261-4829
Fax: 504/598-2918
cesilva@uno.edu

Melinda Simmons
Goor and Betty Moore Foundation
PO Box 29910
The Presidio of San Francisco
San Francisco, CA 94129-0910
Tel: 415/561-7416
Fax: 415/561-7707
Melinda.simmons@moore.org

David R. Sims
Los Alamos National Laboratory
MS 888
Los Alamos, NM 87545
Tel: 505/665-8512
Fax: 505/665-3024
dsims@lanl.gov

David Soergel
University of California, Berkeley
461 Koshland Hall
Berkeley, CA 94720-3102
Tel:650/303-5324
Fax:413/280-7813
soergel@berkeley.edu

Parag Vaishampayan
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: 925/296-5621
Fax: 925/296-5666
pavaishampayan@lbl.gov

Tina K. Van Dyk
DuPont Central Research and
Development
Experimental Station
Wilmington, DE 19880-0173
Tel: 302/695-1430
Fax: 302/695-9183
Tina.k.van-dyk@usa.dupont.com

Nathan C. VerBerkmoes
Oak Ridge National Laboratory
1 Bethel valley Rd.
Building 4500-S, Rm F-160
Oak Ridge, TN 37831
Tel: 865/576-4546
Fax: 865/576-8559
nve@ornl.gov

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@bilbo.bio.purdue.edu

Falk Warnecke
DOE Joint Genome Institute
2800 Mitchell Dr.
Walnut Creek, CA 94598
Tel: 925/296-5867
Fax: 925/296-5720
fwarnecke@lbl.gov

Cheryl Washer
University of Toronto
200 College St.
Toronto, Ontario Canada M5S 3E5
Tel: 416/978-4222
Fax: 416/978-8605
Cheryl.waher@utoronto.ca

Chris Waters
Department of Molecular Biology
Princeton University
Princeton, NJ
cwaters@princeton.edu

Marvin Whiteley
University of Texas, Austin
1 University Station A5000
Austin, TX 78712-0162
Tel: 512/471-5493
mwhiteley@mail.utexas.edu

John C. Wooley
University of California, San Diego
9500 Gilman Dr. MC 0446
La Jolla, CA 92093
Tel: 858/534-2494
Fax: 858/822-3610
jwooley@sdsc.edu

Tanja Woyke
DOE Joint Genome Institue
2800 Mitchell Dr.
Walnut Creek, CA 94598
Tel: 925/296-5839
Fax: 925/296-5850
twoyke@lbl.gov

M. Sc. Yogiara
Atma Jaya Catholic University
Faculty of Biotechnology
Yustinus Building 1st floor
Jalan Jenderal Sudirman No. 51
Jakarta 12930. Indonesia
Tel: 62-21-5703306 ext. 450
Fax: 62-21-5719060
yogiara@atmajaya.ac.id

Rotem Sorek
DOE Joint Genome Institue
2800 Mitchell Dr.
Walnut Creek, CA 94598
Tel: 925/927-2509
rsorek@lbl.gov

Valeria Souza
Universidad Nacional Autonoma de Mexico
Instituto de Ecologia, CU
Coyoacan 04510, Mexico DF
Tel: 52-55-5622-9030
Fax: 52-55-5622-8995
souza@servidor.unam.mx

Bong Hyun Sung
Korea Advanced Institue of Science and Technology Dept of Biological Science
373-1 Guseong-dong, Yuseong-gu.
Daejeon, 306-701 Republic of Korea
Tel: 82-42-869-2659
Fax: 82-42-869-2610
fullhouse@kaist.ac.kr

Curtis Suttle
Earth & Ocean Sciences, Microbiology & Immunology and Botany
University of British Columbia
1505-6270 University Blvd
Vancouver, BC Canada V6T 1Z4
Tel: 604/822-9652
Fax: 604/822-5558
csuttle@eos.ubc.ca

Eduardo Taboada
Laboratory for Foodborne Zoonoses
Public Health Agency of Canada
C/o Animal Diseases Research Institute
PO Box 640, Township Road 9-1
Lethbridge, Alberta, Canada T1J 3Z4
Tel: +1 403-382-5550
Fax: +1 403-381-1202
taboadae@inspection.gc.ca

Rie Takahashi
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
riet@ucla.edu

Cindy Tamae
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
ctamae@ucla.edu

Dominic Taylor
University of California Los Angeles
Life Science Core Curriculum
2847 Slichter Hall
Los Angeles, CA 90095
dmltaylor@aol.com

Tracy K. Teal
Mail code 156-29
California Institute of Technology
Pasadena, CA 91125
Tel: 626/395-5925
tracyt@caltech.edu

Isadora Trejo
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
itrejom@ucla.edu

Complete Sequence and Comparative Genome Analysis of *Leuconostoc citreum*, a Key Player in Kimchi Fermentation

Haeyoung Jeong1, Sang-Haeng Choi2, Soohyun Lee1, Misook Ha1, Ji-Sun Kim2, Jung-Sook Lee3, Tae Kwang Oh4,
Hong-Seog Park2 and Jihyun F. Kim1*

1Systems Microbiology Research Center, 2Genome Research Center, 3Biological Resource Center, and 421C Frontier Microbial Genomics & Applications Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), PO Box 115, Yuseong, Daejeon 305-600, Korea.

Leuconostoc citreum is known as a predominant lactic acid bacterium during kimchi fermentation process. We sequenced the genome of *L. citreum* KM20, which is capable of inhibiting the growth of several harmful bacteria. The complete genome is composed of one circular chromosome (1,796,284 bp, 39.0% G+C) and four plasmids of 100,330 bp. Among the 1,826 putative protein-coding genes, functions of 1,621 (88.8%) could be assigned. Genome analysis revealed complete gene set for heterolactic fermentation via phosphoketolase pathway with an incomplete tricarboxylic acid cycle, and their limited biosynthetic capacity for various amino acids and cofactors. Codon adaptation index analysis based on ribosomal protein references indicates genes for phosphoketolase pathway and, in addition to housekeeping genes, some phosphotransferase systems are potential highly expressed genes. A plasmid-encoded, putative cell wall-anchored protein with five mucin-binding domains suggests that *L. citreum* has the capacity to colonize on the surfaces of the gastrointestinal tract. This work will provide scientific insights of the probiotic effects and new biotechnological applications of traditional fermented foods. [Supported by the 21C Frontier Microbial Genomics and Applications Center Program, MOST, Korea]

Evolutionarily conserved RNA secondary structures within CRISPR repeats

Victor Kunin, Rotem Sorek and Philip Hugenholtz

DOE Joint Genome Institute, 2800 Mitchell Drive,Walnut Creek, 94598, CA

Clustered Regularly Interspersed Palindromic Repeats (CRISPRs) are repetitive structures in Bacteria and Archaea comprised of exact repeat sequences of approximately 21 to 48 bases long separated by unique spacers of similar length. Multiple genes appear in conjunction with these repeats, and are called CRISPR-associated sequences (CAS). The spacers are often highly similar to fragments of extrachromosomal DNA, such as phage or plasmid DNA. It was therefore suggested that the CRISPR/CAS system participates in an antiviral response probably by an RNAi-like mechanism. The gene composition of CAS systems was studied extensively, and it was proposed that they could be divided into types, according to their operon organization and gene phylogeny. However, the nature of the repeat sequences was not examined closely. Here we analyze 267 prokaryotic organisms and identify distinct groups of CRISPR repeats based on sequence conservation. We show that some groups of CRISPR repeats fold into distinct stem-and-loop patterns, as confirmed by compensatory mutations in the stem region. While the stem-loop motif is seen several groups, the actual sequence, as well as the length of the stem, its position relative to the unstructured part, and the size of the unstructured sequence varies between clusters. Despite earlier reports, no preference for folding is found in spacers. The CRISPR repeat clusters correspond to some degree to the types of CAS machineries, and structured clusters often have more homogeneous protein machinery. While some clusters are phylogenetically confined, some contain repeats from organisms as diverged as Bacteria and Archaea. In contrast, spacers are usually exclusive to narrow phylogenetic groups such as species or strains. A single case of a highly similar repeats is observed between organisms belonging to different phylogenetic classes. Our results suggest that the CRISPR system works through an RNA-intermediate of the repeats, and that RNA secondary structures might contribute to the system's function.

REGENERATION - A Next Generation Microbial Gene Modeller

Philip LoCascio, Loren Hauser, Miriam Land, Gwo-Liang Chen, Ed Uberbacher

Contact: locasciop@ornl.gov

The continued unprecedented increase in the number of sequenced genomes has made it critical to automate as accurately as possible genome annotation, with a premium put on capturing as much biological detail as possible. We present the initial findings of our novel microbial gene modeller REGENERATION, that uses the combination of a Markov scoring model, sophisticated GC content analysis and operon detection algorithms, to provide an accurate prediction of > 95% of genes when compared to manual annotations. Each genome has a unique signature that is used to build a specific scoring model of ORF density, coding potential , strand coding preference, and also start site distribution statistics. We use a expontential weighting metho to analyze changes in GC content and yield transitions in likely coding regions. Together with the Markov ORF scoring , this yields candidates that can be assembled into likely operon context, greatly reducing the complexity of producing properly align gene models.

We shall present some of the findings for typcially problematic high-GC (>65%) genomes, as well as some of the mathematical procedures utilized in acheiving our current results.

Metagenomic analysis of phosphorus removing sludge communities.

Héctor García Martín[1], Natalia Ivanova[1], Victor Kunin[1], Falk Warnecke[1], Kerrie Barry[1], Linda Louise Blackall[2], Katherine D. McMahon[3], and Philip Hugenholtz[1]

[1] DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, USA.

[2] Advanced Wastewater Management Centre, University of Queensland, St Lucia, 4072, Queensland, Australia.

[3] Civil and Environmental Engineering Department, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI, USA.

Excessive inorganic phosphate (Pi) supply to freshwater negatively affects water quality and ecosystem balance through a process known as eutrophication. Enhanced Biological Phosphorus Removal (EBPR) is a treatment process in which microorganisms remove Pi from wastewater by accumulating it as polyphosphate. Here we present a metagenomic analysis of two geographically separated lab-scale EBPR sludges dominated by the main phosphorus removal agent, the uncultured bacterium, "*Candidatus* Accumulibacter phosphatis". This analysis resolves several controversies in EBPR metabolic models and provides hypotheses explaining the dominance of *A. phosphatis* in this habitat, its lifestyle outside EBPR and probable cultivation requirements. Comparison of metagenomes from two geographically separated locations highlights recent evolutionary dynamics in the *A. phosphatis* genome that could be linked to mechanisms for environmental adaptation. In spite of an apparent lack of phylogenetic overlap in the flanking communities of the two sludges studied, common functional themes were found, at least one of them complementary to the inferred metabolism of the dominant organism. Locally, phage predation appears to be a key driver of community instability in this and probably other low complexity engineered systems. On a global scale, we postulate that diversification of anti-phage responses retained within a metapopulation of freely moving strains is a survival strategy of bacterial species experiencing phage-induced population collapse.

The present study provides a much-needed blueprint for a systems-level understanding of EBPR. Its integration with transcriptomic, proteomic and metabolomic data will produce quantitative models capable of predicting the necessary conditions for EBPR effective operation.