12th INTERNATIONAL CONFERENCE ON MICROBIAL GENOMES

PROJECT

September 26-30, 2004

UCLA CONFERENCE CENTER

LAKE ARROWHEAD

CALIFORNIA

① uncultured organism
- symbionts - pure are ideal - wolbachia

- symbionts 2 - Baumannia

- more to come

② finishing genome not always possible
- Baumannia

- s. glossa

③ F/x/ predict
- ed mentioned problem w/ predicting
Rhodoplum

- free doesn't help
- GDR
- PG profile

scaffold RNA like
F/x/ product 1/2
rRNA

9:30-10:00 am	Martin Keller Diversa **"From Genomes to Life"**
10:00-10:30 am	Kenneth H. Nealson University of Southern California **"Subfreezing Activity of Microorganisms and the Potential Habitability of Mars' Polar Regions"**
10:30-10:50 am	Break
10:50-11:20 am	Kasthuri Venkateswaran Jet Propulsion Laboratory **"Microbial Diversity of Spacecraft Associated Microbes and Their Survival Capabilities Under Martian Environmental Conditions"**
11:20-11:50 am	Steven K. Schmidt University of Colorado **"Active Microbial Communities Under Snow"**
11:50 am-12:20 pm	Patrick Schloss University of Wisconsin **"Using Functional Anchors to Study the Soil Metagenome"**
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	**Microbial Communities II: Biodiversity/Interaction/Evolution**
7:45-8:15 pm	John Heidelberg The Institute for Genomic Research **"Comparative Genomics of Yellowstone Cynabacterial Mats"**
8:15-8:45 pm	Gene Tyson University of California, Berkeley **"Genomics of Microbial Communities"**

the sequence monkey

7:47 ~ 8:12

8:17

8:47

How accurate is the bunny?

8:45-9:15 pm — Beth Lazazerra
University of California, Los Angeles
"Mechanisms of Cell-Cell Signaling by Bacillus Subtilis"

9:15-9:30 pm — Break

9:30-10:00 pm — Margaret A. Riley
Yale University
"Bacteriocin Diversity: Evolutionary Perspectives"

10:00-10:25 pm — Rachna J. Ram
University of California, Berkeley
"Key Roles for Novel "Hypothetical" Proteins in a Natural Microbial Biofilm are Revealed by Whole Community Proteomics"

TUESDAY, SEPTEMBER 28

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

Session III — **Syngenomics: Pathways/Modeling/Synthetic Biology I**

8:45-9:15 am — Bernhard Palsson
University of California, San Diego
"Using *in silico* Models of Microorganisms to Elucidate Metabolic Pathways"

9:15-9:45 am — Jay Keasling
University of California, Berkeley
"Engineering *E. coli* to Produce New Pathways"

9:45-10:15 am — Drew Endy
Massachusetts Institute of Technology
"Design of Genetic System"

10:15-10:35 am — Break

10:35-11:05 am	Donald A. Bryant Pennsylvania State University **"Biochemical Genetics in the Obligately Phototrophic Green Sulfur Bacterium Chlorobium Tepidum"**
11:05-11:35 am	Michael A. Savageau University of Michigan School of Medicine **"Constructing Genetic Circuitry in *E. coli*"**
11:35 am-12:00 pm	Huiying Li University of California, Los Angeles **"Using Bioinformatics to Detect Parallel Pathways in Organisms"**
12:00 pm- 12:25 pm	Les Hoffman Epicentre **"Molecular Biology Tools for Microbial Prospecting"**
12:30 pm	Lunch (Dining Room)
4:00-6:00 pm	Poster Session (Lakeview Room) Social/Mixer (Lakeview Room)
5:00- 6:00 pm	Workshop (Pineview Room) Peter Karp SRI International **"Making the Most of the EcoCyc Database"**
6:15- 7:45 pm	Dinner (Dinning Room)
Session IV	**Pathways/Modeling II: New Genome Studies**
7:45- 8:15 pm	Clifford Unkefer Los Alamos National Laboratory **"Stable Isotope Assisted Metabolome Analysis of Methylobacterium extorquens AM1 and *Nitrosomonas europae*"**
8:15- 8:45 pm	Michael Laub Harvard University **"Global Approaches to Dissecting the Caulobacter Cell Cycle"**
8:45- 9:00 pm	Break

Functional Anchors to Study the Soil Metagenome

Patrick D. Schloss
Jo Handelsman
Department of Plant Pathology, University of Wisconsin-Madison

The metagenome, the collective genomes of an assemblage of bacteria, has been studied by cloning DNA extracted directly from environmental samples and screening for genes of interest by probing, functional screening, and brute force sequencing. We propose the concept of a "functional anchor" to describe clones and profile entire communities. A functional anchor is a phenotype that can be detected in a library that provides a starting point for studying a single clone or all of the clones in a library that express the function. To demonstrate the utility of a functional anchor, we constructed metagenomic libraries in BAC and fosmid vectors from Alaskan soil and developed an intracellular screen that detects the ability of a metagenomic clone to induce and inhibit quorum sensing. We identified 8 clones that induce and 12 that inhibit quorum sensing. We have successfully sequenced about 170 kb of DNA from four of the inducers and found that only one of the clones contains a LuxI homologue. Using the arrangement and annotation of contiguous genes, we have been able to assign the phylogenetic origin of these clones to the phylum level with reasonable confidence. This work provides the first step in developing a functional profile of a clone library that will link quorum sensing activity with the identity of organisms that contain this activity in a culture-independent manner.

10:35-11:05 am	Donald A. Bryant Pennsylvania State University **"Biochemical Genetics in the Obligately Phototrophic Green Sulfur Bacterium Chlorobium Tepidum"**
11:05-11:35 am	Michael A. Savageau University of Michigan School of Medicine **"Constructing Genetic Circuitry in *E. coli*"**
11:35 am-12:00 pm	Huiying Li University of California, Los Angeles **"Using Bioinformatics to Detect Parallel Pathways in Organisms"**
12:00 pm- 12:25 pm	Les Hoffman Epicentre **"Molecular Biology Tools for Microbial Prospecting"**
12:30 pm	Lunch (Dining Room)
4:00-6:00 pm	Poster.Session (Lakeview Room) Social/Mixer (Lakeview Room)
5:00- 6:00 pm	Workshop (Pineview Room) Peter Karp SRI International **"Making the Most of the EcoCyc Database"**
6:15- 7:45 pm	Dinner (Dinning Room)
Session IV	**Pathways/Modeling II: New Genome Studies**
7:45- 8:15 pm	Clifford Unkefer Los Alamos National Laboratory **"Stable Isotope Assisted Metabolome Analysis of Methylobacterium extorquens AM1 and *Nitrosomonas europae*"**
8:15- 8:45 pm	Michael Laub Harvard University **"Global Approaches to Dissecting the Caulobacter Cell Cycle"**
8:45- 9:00 pm	Break

9:00- 9:30 pm	Susannah G. Tringe Lawrence Berkeley National Laboratory **"The JGI Genome Sequencing Program – New Findings"**
9:30-10:00 pm	Rick Cavicchioli University of New South Wales, Australia **"Insight into Cold Adaptation in Archaea from an Integrated Genomic/Proteomic Approach"**
10:00-10:25 pm	Pongpan Laksanalamai University of Maryland Biotechnology Institute **"A Versatile Nascent Associated Complex in the Archaea: Evidence for a combination of chaperone and protein degradation roles"**

WEDNESDAY, SEPTEMBER 29

7:45-8:30 am	Breakfast (Dining Room)
Session V	**Pathogens/Antibiotic Resistance/Biothreats**
8:45-9:15 am	Julian Davies University of British Columbia **"Antibiotic Resistance"**
9:15-9:45 am	Julian Parkhil Welcome Trust Sanger Institute, UK **"Comparative Genomics and Genome Organization of Microbial Pathogens"**
9:45-10:15 am	Carolin Frank University of Uppsala, Sweden **"Comparative Genomics of Vector-borne Pathogens"**
10:15-10:35 am	Break
10:35-11:05 am	Fredrick Blattner University of Wisconsin **"Genomic Studies of Pathogens"**

11:05-11:50 am	Paul Keim Northern Arizona University and Jacques Ravel The Institute for Genomic Research **"Analyzing Biothreat Agents with Genomic Sequencing Information"**
11:50 am-12:20 pm	Nicholas H. Bergman University of Michigan Medical School **"Functional Genomics Studies of Anthrax Pathogenesis"**
12:30 pm	Lunch (Dining Room)
Session VI A	**Extremophiles/New Genomes/Bioremediation**
4:30-5:00 pm	Allen Tsang University of Massachusetts **"New Insights into the "POWER" of Geobacter: Physiology, Regulatory Networks, and Electron Transfer from Genome Analysis"**
5:00-5:30 pm	John E. Battista Louisiana State University **"*Deinococcus radiodurans*, an Extremely Radiation Resistant Organism"**
5:30-6:00 pm	Wolfgang Liebl University of Goettingen, Germany **"Genome Sequence of *Picrophilus torridus* and its implications for life around pH 0"**
6:15-7:45 pm	Dinner
Session VI B	**Fungal Genomics/Comparative Analysis**
7:45-8:15 pm	Scott E. Baker Pacific Northwest National Laboratory **"Lessons from Comparative Genomics and Proteomics of Flamentous Fungi"**
8:15-8:45 pm	Jim Kronstad University of British Columbia **"A SAGE Approach to Fungal Pathogenesis"**

8:45-9:10 pm	Vinayak Kapatral Integrated Genomics **"Comparative Functional Genome and Evolutionary Analysis of *Wolbachia, Rickettsia, and Ehrlichia* Genomes Using ERGO Bioinformatics Suite"**
9:10 pm	Reception/Party (Iris Room)

THURSDAY, SEPTEMBER 30

7:30-8:15 am	Breakfast (Dining Room)
Session VII	**New Methods/New Discoveries**
8:30-9:00 am	Owen White The Institute for Genomic Research **"Bacterial Annotation Tools"**
9:00-9:30 am	Monica Riley Marine Biological Laboratory **"Capturing Information from Structural Elements Provides Functional Information for Many Proteins of Unknown Function"**
9:30-10:00 am	Tom Albert NimleGen Systems, Inc **"Whole Genome Analysis of Microbial Genomes Using High-Density Oligonucleotide Microarrays"**
10:00-10:30 am	Jizhong Zhou Oak Ridge National Laboratory **"Novel Random Matrix Theory-Based Approach for Identifying Gene Interaction Networks"**
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

SPEAKER ABSTACTS

Systems Biology in the Ocean

Edward F. DeLong

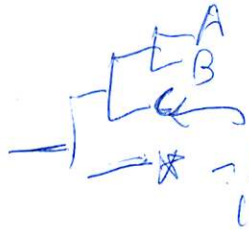
Biological interactions and biotic processes that occur within natural microbial communities are diverse and complex. Much of this complexity is encoded in the nature, structure, distribution, and dynamics of interacting genomes in the environment. This genomic information can now be rapidly and generically extracted en toto from the genomes of microbes cooccuring in natural habitats, using standard genomic technologies. By analogy, consider that in 1965, Linus Pauling predicted that individual macromoleculare sequences could serve as documents of evolutionary history. A logical extension of this view is that co-occurring genomic sequences within a population are documents of (micro)evolutionary history, population biology, and ecology for entire microbial ecosystems. The culture-independent nature of environmental genomic approaches now permits equal access to abundant, yet still uncultivated microbial species. Genome analyses of such elusive microbial species are already providing new information on their metabolic pathways, biochemistry and physiological properties. Genomic approaches are also now enabling a better appreciation of natural biogeochemical cycles, and the microbes that mediate them. Operationally, both directed sequencing of large insert clones, as well as blind shotgun sequencing surveys are both applicable, and have been shown to be effective. Each approach has their won relative merits and shortcomings. Recent examples of environmental genomics applied to oceanic plankton, and anaerobic methane-eating microbes from deep-sea methane seeps provide examples that illustrate the potential of genomic approaches to reveal the properties and natural history of indigenous microorganisms. The challenge now is to extend these approaches beyond individual organisms, and beyond single populations – to systematically extract information, and infer the interrelationships, and deduce the networks that constitute working biomes.

what was I thinking?

It's like a footsie pop!

Halobacterium have many rhodopsins

① genomics of natural microbes
② population gen distributions, speciation
③ metabolisms
④ comparative ecosystem genomics



cross-feeding
- culture inform environgenomics

The Axis of Evol: Phylogenomics, the Tree of Life, and Shotgun Sequencing Uncultured Organisms

Jonathan A. Eisen. The Institute for Genomic Research.

Shotgun sequencing of DNA from environmental samples is becoming more and more common. The samples range from nearly pure cultures of endosymbionts to extremely diverse soil and oceanic collections. I discuss here recent work from my group in analyzing such samples including *Wolbachia* and the Sargasso Sea. I discuss how these studies are can be greatly improved by having genome sequence data from a wide diversity of species and discuss our recent results on our "Tree of Life" project to sequence genomes from novel Phyla of bacteria including Chrysiogenetes, Deferribacteres, Dictyoglomus, Nitrospira, Coprothermobacter, Synergistes, Thermodesulfobacteria, and Thermomicrobium. Finally, I discuss the importance of phylogenomic analysis in the context of analyzing these samples.

Accessing microbial diversity by high throughput cultivation

Karsten Zengler[1], Greg Clark[1], Imke Haller[1], Gerardo Toledo[1], Marion Walcher[1], Michael Rapp[1], Garry Woodnut[1], Eric Mathur[1], Jay M. Short[1], and Martin Keller[1*]
[1]Diversa Corporation, San Diego, California 92121, USA; [2] University of Hawaii, Kaneohe, Hawaii 96744, USA

Even though significant advances have been made in understanding microbial diversity, most microorganisms (bacteria as well as fungi) are still only characterized by „molecular fingerprints" and have resisted cultivation. Here we describe a universal and novel method that provides access to the immense reservoir of untapped microbial diversity by cultivation (1). This technique uses microcapsules to encapsulate single cells combined with parallel microbial cultivation under low nutrient flux conditions, after which flow cytometry is used to detect microcapsules that contain microcolonies (originated from a single encapsulated cell). Microcapsules separate microorganisms from each other, while still allowing the free flow of metabolites and signaling molecules between different microcolonies. The method was applied to cultivate vast numbers of so-far-uncultured bacteria and fungi, as identified by rDNA and ITS sequence analysis. After sorting individual microcapsules into microtiter dishes containing organic-rich medium, most of the cultures grew to cell densities greater than 0.1 (OD_{600}). This high throughput cultivation can provide 10,000s of bacterial and fungal isolates per environmental sample. To deal with the tremendous number of strains isolated by this technique, a fully automated high throughput method to determine the uniqueness of the bacterial and fungal isolates has been developed. Using Fourier transform infrared spectroscopy (FT-IR) in combination
with novel spectra-comparison algorithms, 10,000 microtiter plate cultures can be analyzed per day and unique strains can be separated automatically.
In addition, this high throughput cultivation method has been demonstrated to be suitable for targeted cultivation of microorganisms belonging to bacterial phyla that until now contained very few cultivated representative (for example members of the *Acidobacteria*). In addition individual microcapsules can be analyzed by FISH and sorted based on their fluorescent signal. Thus individually sorted microcolonies can be used to amplify the whole genome to construct genomic libraries for total genome sequencing. The information obtained through the total genome sequencing will allow us to model the metabolism and confirm these by RNA-chip technologies.

(1) Zengler K. et al. 2002. Cultivating the uncultured. Proc Natl Acad Sci U S A 99:15681-15686

Part of this work was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63771.

MDA-500x cells
-5 cells still work

- consortia
- how well preserv
- how many genomes/cell

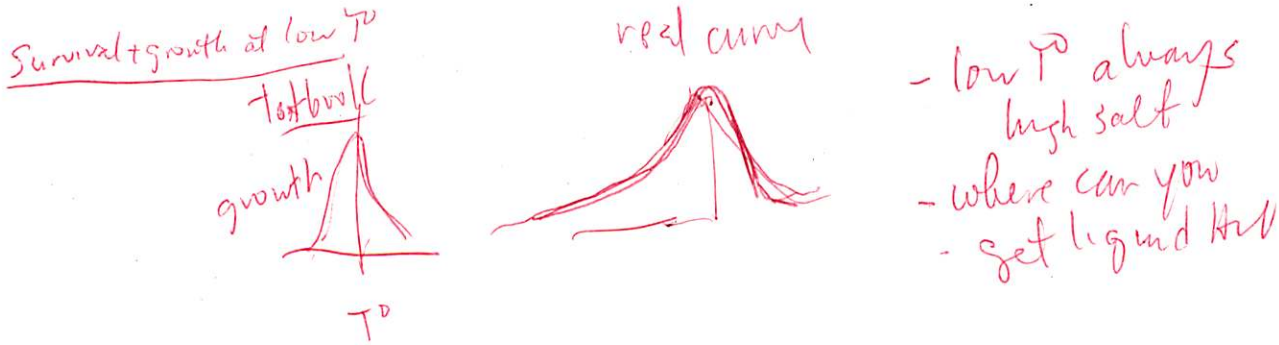
Cold

Microbial Diversity of Spacecraft Associated Microbes and Their Survival Capabilities Under Martian Environmental Conditions.

K. Venkateswaran*, M.T. LaDuc & D.A. Newcombe – Jet Propulsion Laboratory, Pasadena, CA

Phylogenetic characterization of the cultivable aerobic microorganisms associated with spacecraft revealed that 85% of the strains were identified as Gram-positive bacteria. About 65% of the total strains isolated survived heat shock protocols (i.e., 80°C for 15 min.) with members of the *Bacillus* genus being the predominant microbes among the survivors (>91%). *B. licheniformis* (25%) and *B. pumilus* (16%) were the most prevalent *Bacillus* species isolated from this facility. In addition to *Bacillus*, members of *Paenibacillus, Sporosarcina, Urebacillus, Nocardiopsis* and *Streptomyces* also survived heat shock. The modern molecular microbiological methods revealed equal representation of both Gram-positive and –negative bacteria and about >99% seems to be non-cultivable under the conditions employed.

Spore-forming microbes recovered from spacecraft surfaces and assembly facilities were exposed to simulated Mars UV irradiation. The effects of UVA, UVA+B, and the total Martian UV irradiation on the survival of microorganisms were studied. Investigation of these different wavelengths is important because dust suspended in the Martian atmosphere can attenuate UVC while allowing both UVA and UVB to reach the surface. Among 45 *Bacillus* spore lines screened, 19 isolates showed resistance to UVC irradiation after exposure to 1,000 J/m^2 at UVC (254 nm using a low-pressure mercury lamp). Separate UV bands at the Mars simulated solar constant including UVA (315 – 400 nm), UVA+B (280 – 400 nm), and full spectrum (200 – 400 nm) irradiation were created by placing filters in the light path of calibrated xenon-arc lamps. Results showed that spores of *Bacillus* species isolated from spacecraft surfaces were more resistant than a laboratory strain, *B. subtilis* 168. The fact that 90% reductions in viable spore numbers required 35 times greater exposure times to UVA+B than to the full UV spectrum is confirming that UVC (200 to 280 nm) is detrimental to the viability of spores. Although all spores tested exhibited sensitivity to UVA+B, most damages of UVA+B might be attributed to the wavelength 280 to 315 nm (UVB), since all the spores tested were resistant to UVA (315 to 400 nm). Among all *Bacillus* species tested, a strain of *B. pumilus* SAFR-032 showed highest resistance to all three UV bandwidths. LD_{90} of *B. pumilus* SAFR-032 spores under Mars UV simulated solar constant was >270 sec, about 6 times greater than *B. subtilis* 168 spores. These results show that the elevated resistance to simulated Mars UV irradiation was strain-specific rather than species-specific. The whole genome sequence of *B. pumilus* SAFR-032 is being sequenced.



Thriving microbial life beneath snow packs and glaciers: Biogeochemical and biotechnological implications.

S.K. Schmidt*, A.P. Martin, C.W. Schadt, D.A. Lipson, A.F. Meyer, D.R. Nemergut and E.K. Costello.

I will review our work exploring the under-snow world of alpine environments in Colorado and high elevation (5000 m+) glaciers in Peru. Surprisingly, there are very active microbial communities functioning in both of these environments. We have recently shown that tundra soil microbial biomass can reach its annual peak under snow, and that fungi account for most of the biomass. Phylogenetic analysis revealed a high diversity of novel microbes that constitute major new groups that are divergent at the kingdom, subphylum or class levels. An abundance of previously unknown, cold-tolerant microbes beneath the snow substantially broadens our understanding of both the diversity and biogeochemical functioning of microorganisms in cold environments.

Functional Anchors to Study the Soil Metagenome

Patrick D. Schloss
Jo Handelsman
Department of Plant Pathology, University of Wisconsin-Madison

The metagenome, the collective genomes of an assemblage of bacteria, has been studied by cloning DNA extracted directly from environmental samples and screening for genes of interest by probing, functional screening, and brute force sequencing. We propose the concept of a "functional anchor" to describe clones and profile entire communities. A functional anchor is a phenotype that can be detected in a library that provides a starting point for studying a single clone or all of the clones in a library that express the function. To demonstrate the utility of a functional anchor, we constructed metagenomic libraries in BAC and fosmid vectors from Alaskan soil and developed an intracellular screen that detects the ability of a metagenomic clone to induce and inhibit quorum sensing. We identified 8 clones that induce and 12 that inhibit quorum sensing. We have successfully sequenced about 170 kb of DNA from four of the inducers and found that only one of the clones contains a LuxI homologue. Using the arrangement and annotation of contiguous genes, we have been able to assign the phylogenetic origin of these clones to the phylum level with reasonable confidence. This work provides the first step in developing a functional profile of a clone library that will link quorum sensing activity with the identity of organisms that contain this activity in a culture-independent manner.

Community genomics-enabled investigation of ecology and evolution in a low complexity system

Gene W. Tyson, Eric E. Allen, Ian Lo, Brett J. Baker, Rachna J. Ram, Philip Hugenholtz, Michael Thelen, and Jillian F. Banfield

University of California, Berkeley CA, 94720 USA

Genomic data simultaneously record information about metabolic potential and evolutionary history. Recently, we reported the shotgun sequencing of a microbial biofilm from a low complexity community growing in a subsurface acidic ecosystem. The data were reconstructed to recover partial and near complete genomes of the five dominant members. The results provided new insights into the partitioning of key community roles and enabled the first successful cultivation of the only member capable of nitrogen fixation. It has also revealed one novel nanoarchaeal lineage that was previously undetected using conventional PCR-based approaches. The community genomic data are derived from many individuals. Thus, the data reveal the genomic heterogeneity within populations. Comparative genomics involving archaeal strains, species, and genera has been used to deduce the dominant modes and relative rates of genome evolution for these microorganisms. The results highlight the importance of recombination in shaping population structure and of phage as sources of novel genes. The initial genomic dataset has been applied to identify the abundant proteins in the first comprehensive community proteomic study (Ram *et al.*, in prep.). Thus, community genomic data enable functional analyses of microbial consortia in their natural environments.

Mechanisms of Cell-Cell Signaling by *Bacillus subtilis*

B.A. Lazazzera. University of California, Los Angeles.

Cell-cell signaling in bacteria regulates many processes, including virulence, antibiotic production, and genetic exchange, and is commonly used for quorum sensing (i.e. monitoring cell density). For quorum sensing, bacteria secrete signaling molecules that bind to cellular receptors to regulate gene expression. Gram-positive bacteria, such as *Bacillus subtilis*, secrete peptides as signaling molecules. One class of signaling peptide produced by *B. subtilis* is represented by ComX pheromone, which is a 10 amino acid, modified peptide that signals through a membrane-associated, histidine protein kinase. The other class is represented by the CSF peptide, which is an unmodified, five amino acid peptide. When CSF reaches a sufficiently high concentration in the extracellular medium, it is transported into the cell via an oligopeptide transporter. Once inside the cell, CSF binds to and inhibits the activity of RapC and RapB, which inhibit the activity of response regulators ComA and Spo0F, respectively, thus, allowing CSF to stimulate gene expression. CSF is derived from the C-terminal 5 amino acids of the *phrC* gene product, and interestingly, *phrC* is located immediately downstream of *rapC*. There are several *rap phr* gene cassettes found on the *B. subtilis*, *Bacillus anthracis*, *thuringiensis*, and *cereus* genomes. All these *phr* have in common that they encode an ~40 amino acid protein that has a signal sequence for export. The Rap proteins appear to have a tetratricopeptide-repeat (TPR) domain. To test the widespread nature of the *rap phr* cassettes, the *Clostridium acetobutylicum* genome was searched for genes encoding TPR domain proteins and then these genes were screened for those that have a small gene located immediately downstream with similar features to *phr* genes. This revealed that *C. acetobutylicum* encodes 7 putative peptide signaling cassettes and suggests that other bacterial that also encode TPR domain proteins may also participate in peptide signaling.

Microbial Toxins Promote Biodiversity in a Real Life Game of Rock, Paper, Scissors.

M.A. Riley, University of Massachusetts, Amherst.

Microbes comprise, by far, the greatest diversity of life on this planet. However, little is known about how this extraordinary diversity is generated or maintained. Recent work suggests two factors play a role in this process, spatial structure and non-transitive interactions. *In silico, in vitro* and *in vivo* studies reveal that interactions mediated by microbial toxins create non-transitive interactions, similar to those exhibited in the childhood game of rock, paper, scissors. When such ecological forces are applied in a structured setting, biological diversity is maintained.

Key roles for novel "hypothetical" proteins in a natural microbial biofilm are revealed by whole community proteomics

Rachna J. Ram, Nathan VerBerkmoes, Michael P. Thelen, Gene W. Tyson, Brett J. Baker, Robert C. Blake II, Robert Hettich, and Jillian F. Banfield

University of California, Berkeley, CA, 94720 USA

The microbial world is almost unimaginably rich and diverse, but incredibly poorly understood. Our understanding is limited because detailed analysis of physiology and genetics has been confined largely to studies of microorganisms from the few lineages for which cultivation conditions have been determined. An additional limitation of pure culture-based studies is that potentially critical community and environmental interactions are not sampled. Recent acquisition of genomic data from environmental samples is beginning to reveal the genetic potential of communities and ecosystems (e.g., Tyson et al. 2004). However, major remaining challenges are to verify predicted genes and to identify and assign functions to novel proteins, especially those that are lineage-specific or important for habitat adaptation. We have applied proteomic analyses to acidophilic biofilm samples previously characterized by cultivation-independent genomic methods. About one third of the products of predicted genes in *Leptospirillum* group II, the dominant community member, were detected. Many abundant novel proteins in the biofilm could be assigned to putative functional categories based on operon structure. However, some were encoded in blocks of novel genes, a subset of which appear to have been introduced by viral transduction and other recombination processes. These recently acquired genes likely confer critical new functions and promote species divergence. Targeted analysis of prominent novel extracellular proteins resulted in identification of cytochromes central to iron oxidation, and thus AMD formation.

Bringing Genomes to Life:
The use of genome-scale in silico models

Bernhard O. Palsson
Department of Bioengineering, University of California-San Diego, La Jolla, CA, USA

High throughput (HT) data generation in biology has led to the availability of vast amounts of chemical compositional data about cells. These developments have led to the emergence of systems biology that is widely viewed as being comprised of four steps: 1) information about cellular components, 2) reconstruction of biochemical reaction networks, 3) formulation of in silico model of network functions (i.e. phenotypes) and 4) measurement of phenotypic responses and their comparison to computed properties. Disagreement leads to an iterative model building procedure. HT phenotyping is one of the limiting steps in this process.

Reconstruction of genome-scale networks for metabolism and regulation in single cellular organisms in now possible, and efforts in reconstructing networks in human cells have begun. In silico models that characterize their function can be used to analyze, interpret and predict the genotype-phenotype relationship. Reconstructed genome-scale models for E. coli and Yeast that include metabolism, regulation and transcription/translation have been formulated. These models integrate and represent a wide variety of high-throughput data.

Genome-scale models can be used to analyze the phenotypic consequences of gene deletions, optimal growth rates, the outcome of adaptive evolution, and for design of strains for bioprocessing. Examples in all these categories will be given, with emphasis on the computational and experimental analysis of adaptive evolution. Full characterization of adaptive evolutionary processes in terms of genome-wide expression profiling and full DNA re-sequencing has been performed. Thus both the genetic and epigenetic changes underlying adaptive evolution have been measured on a genome-scale and this data can be interpreted with the genome-scale in silico models.

Synthetic Biology: Engineering Microbial Metabolism for Production of Terpenoid Drugs

Jay D. Keasling, University of California, Berkeley, CA 94720-1462

Plants and fungi produce a plethora of isoprenoids that function as aromas, flavors, and pest rejection. Many of these molecules have found use as antibacterial, anti-fungal, and anti-cancer agents, even though evolution has optimized them for function in plants. Unfortunately, many of these compounds are produced at extremely low levels. To increase their diversity and the level of production in *E. coli*, we have recruited genes from several organisms for complete reconstruction of the terpene biosynthetic pathways and have evolved sesquiterpene genes to increase the diversity of terpene olefins produced by this organism. The genes encoding the mevalonate-dependent isopentenyl pyrophosphate (IPP) biosynthetic pathway from *Saccharomyces cerevisiae* were cloned into operons and integrated into the chromosome, and their expression was optimized using laboratory evolution and promoter control to maximize the flux through the pathway. The genes encoding sesquiterpene cyclases from various plants were cloned and optimized for expression in *E. coli*. Although it was possible to detect sesquiterpenes from the culture extracts of the recombinant strains, the levels produced were very low (□g/L). Increases in the flux to the precursor (FPP) did not yield higher terpene production, indicating that poor terpene cyclase expression and/or activity was limiting. To test this hypothesis, cyclase genes optimized for expression in *E. coli* were synthesized in the laboratory. These enzymes (amorphadiene and *epi*-cedrol cyclases) did not show significantly higher yields of terpenes from endogenous levels of FPP but produced significantly higher yields (mg/L) when expressed in a host over-producing FPP from the mevalonate-dependent IPP pathway. To sample the potential terpene diversity produced from terpene cyclases, we constructed point mutation and chimeric libraries of cyclase genes and screened these libraries for the presence of sesquiterpenes. Results on the sesquiterpene diversity generated by these libraries will be presented.

Refactoring Bacteriophage T7

Leon Chan [1*], Sriram Kosuri [2*], Drew Endy [2]

* L.C. and S.K. contributed equally to this work

[1] Department of Biology

[2] Division of Biological Engineering

Massachusetts Institute of Technology

MIT 68-580, 77 Massachusetts Avenue, Cambridge, MA 02139

Email: endy@mit.edu

Natural biological systems are optimized, if at all, to continue to exist. Natural biological systems may not be optimized to enable discovery, characterization, manipulation, modeling, or redirection. Here, we redesign the 39,937 base pair genome of bacteriophage T7 in order to insulate and enable independent, extensible manipulation of primary genetic elements. We replace the left 11,515 base pairs of the wild-type genome with 12,179 base pairs of redesigned DNA. The resulting chimeric genome encodes a viable bacteriophage that grows less well than the natural isolate. Those natural biological systems that we most care about should be rebuilt in service of human understanding and intention.

Genomics-Enabled Biochemical Genetics and Genome Analyses of Green Sulfur Bacteria

D.A. Bryant*, N.-U. Frigaard, A. Gomez Maqueo Chew, J.A. Maresca, H. Li, and T. Li.
Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802 USA.

Chlorobium tepidum is a naturally transformable, obligately anaerobic, phototrophic green sulfur bacterium (GSB). The genome of *C. tepidum* was recently sequenced and found to be a circular DNA molecule of 2,154,946 bp that encodes ~2284 ORFs. A second GSB genome has now been completely sequenced. *Chlorobium chlorochromatii*, the epibiont of the phototrophic consortium "*Chlorochromatium aggregatum*," has a circular chromosome of 2,572,079 bp. Finally, draft sequences for *C. phaeobacteroides* (2.8 Mb), *Pelodictyon luteolum* (2.35 Mb), *P. phaeoclathratiforme* (2.7 Mb), and *Prosthecochloris aestuarii* (2.6 Mb) have been determined. A few results from comparative analyses with *C. tepidum* will be presented. GSB can synthesize 3 types of (bacterio)-chlorophyll ((B)Chl) in the approximate proportions: 97% BChl *c*, 3% BChl *a*, and 0.3% Chl a_{PD}. A pathway for the biosynthesis of these three Chls has been proposed and tested by targeted mutagenesis. Nearly all of the genes encoding enzymes that specifically participate in the synthesis of BChl *c* have been identified, and the properties of corresponding mutants have been analyzed biochemically and physiologically. These include the genes encoding the three unique methyltransferases specific for the C-20, C-8^2 and C-12^1 positions of BChl *c*. Physiological characterization of the mutants provides novel insights into the functional roles of BChl *c* methylation in GSB. The carotenoid biosynthetic pathway has also been elucidated. A novel lycopene cyclase, also found in some cyanobacteria, was identified by phylogenomics and complementation of an *Escherichia coli* strain that synthesizes lycopene. Mutants completely lacking carotenoids, because of a mutation in phytoene synthase, or possessing carotenoids with differing numbers of conjugated double bonds have been produced. Growth experiments suggest that carotenoids in GSB have different functions at high and low light intensity. Surprisingly, three chlorosome envelope proteins (CsmF, CsmJ, and CsmD) are not found in the genome of *C. chlorochromatii*.

DISCOVERY OF SYSTEM DESIGN PRINCIPLES AND CONSTRUCTION OF GENE CIRCUITS

Michael A. Savageau

Department of Biomedical Engineering, and Microbiology Graduate Group, The University of California, Davis, CA 95616

The ability to comprehensively and quantitatively monitor dynamic changes in gene expression, together with new genome-scale informatic methods, is enabling high-throughput characterization of genetic regulatory networks. In addition, methods of genetic engineering now allow synthetic regulatory circuits to be readily built. Attention is currently being turned towards manipulating genetic regulatory circuits for therapeutic and technological applications, which increases the need to understand the functional consequences of genetic manipulations and to discover principles that can guide the design process. This issue will be addressed by comparing and contrasting what has been learned about gene circuits in their complex natural setting and what has been learned from designing, constructing and analyzing simple synthetic gene circuits.

Using Bioinformatics to Detect Parallel Pathways in Organisms

Huiying Li*, Matteo Pellegrini, and David Eisenberg, Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Department of Chemistry and Biochemistry, University of California, Los Angeles

Parallel pathways and complexes, originating by gene duplication during evolution, are separate sets of proteins in an organism that catalyze the same or similar biochemical reactions but act on different substrates or utilize different cofactors. Parallel pathways provide versatility and complexity to organisms, and increase cellular flexibility and robustness. Detecting parallel pathways in a genome helps explain and predict the phenotypes of the organism in different environments. We developed a Four-Step Approach for genome-wide discovery of parallel pathways from protein functional linkages. From ten genomes, we identified 37 cellular systems that consist of parallel pathways. This approach recovers known parallel pathways and complexes, and discovers new ones that conventional homology-based methods did not previously reveal, as illustrated by examples of peptide transporters in *Escherichia coli* and nitrogenases in *Rhodopseudomonas palustris*. The approach untangles intertwined functional linkages between parallel pathways and expands our ability to decode protein functions from genome sequences.

Molecular Biology Tools for Microbial Prospecting

L.M. Hoffman* & B.W. Jarvis. EPICENTRE, Madison, WI

Microbes are the most abundant life form on earth. Yet, we know very little about the vast majority of microorganisms, largely due to the challenges of culturing them. New culture-independent techniques have been devised to examine the spectrum of organisms in specific environments. We have developed kits and reagents applicable to isolating, cloning, screening and sequencing rRNA genes (rDNAs) from environmental sources. These tools were used for examining the diversity of bacteria in a spring pond in southern Wisconsin. Libraries of rDNAs from pond bacteria were constructed and sequenced. In such cases, the generation of molecular mosaics between amplicons from related species may artificially increase the species diversity. The PCR "reconditioning" method of Thompson, et al. (NAR 30: 2083-2088, 2002) was used to decrease the incidence of artifactual sequences in rDNA clones. Nonetheless, the majority of species inhabiting the spring pond have not been previously described.

Various algorithms are used to identify non-cytoplasmic proteins encoded by sequenced microbial genomes. We have developed a biological screen for non-cytoplasmic protein genes, using a simplified fosmid cloning process and selectable transposon fusions. A copy-controlled, blunt fosmid vector (pCC1) allowed quick construction of a genomic library from a marine *Vibrio sp.* Random fosmid clones were "bombed" with the transposon <BlaM/ R6K_ ori> to create fusions between fosmid genes and a signal peptide-less ampicillin resistance gene (BlaM). In-frame fusions of BlaM with exported or membrane protein genes allowed selection of colonies on ampicillin plates. Several *Vibrio* genes selected in this manner were previously categorized as hypothetical or conserved hypothetical. Transposon screening for exported or membrane protein genes can therefore complement bioinformatics for bacterial gene discovery and annotation.

Global Approaches to Dissecting the *Caulobacter* Cell Cycle

Michael T. Laub, Harvard University.

Progression through the cell cycle requires the precise coordination and timing of multiple events, each critical to survival and proliferation. For *Caulobacter crescentus*, this includes DNA replication, DNA methylation, chromosome segregation, cell division, and the biogene of polar organelles such as a stalk, flagellum, and pili. Each of these events, which must occ at a specific stage of the cell cycle, requires the expression of a discrete set of genes. To comprehensively identify these sets of temporally-regulated genes, we used whole genome DNA microarrays to measure global patterns of gene expression in a population of synchronized wild type cells. Analysis of cell cycle expression profiles for nearly 3000 gene led to the identification of 553 genes whose transcripts varied in a cell cycle-dependent fashi including more than 40 regulatory genes. Expression profiling of strains bearing mutant allel of two of these regulatory genes, the response regulator *ctrA* and the histidine kinase *cckA*, reveals that nearly a quarter of all cell cycle regulated genes are controlled, directly or indirectly, by these master regulators. Using microarrays and a modified chromatin-immunoprecipitation procedure, we have also now mapped the binding sites of CtrA through the genome, allowing us to identify approximately 95 genes that are directly controlled by CtrA. These data are beginning to resolve the modular genetic network controlling cell cycle progression and differentiation in *Caulobacter*.

Including *ctrA* and *cckA*, the *Caulobacter* genome encodes 108 two-component signal transduction genes. Systematic deletion and overexpression of these genes has identified more than 20 additional two-component signaling genes that participate in the regulation of cell growth, cell cycle progression, and morphogenesis. Of these, eight are essential for viability; through a combination of phenotypic analysis, expression profiling, and a new kinase-substrate profiling technique, the role of these two-component signaling genes in cell cycle progression is being dissected.

Metagenomic shotgun sequencing of soil

S. G. Tringe* & E. M. Rubin. DOE Joint Genome Institute.

C. Chang & E. Mathur, Diversa Corp.

Obtaining the sequence of intact microbial genomes is challenging in the study of most environmental samples where the diversity of organisms present largely precludes high-coverage assembly. In these studies we have performed a metagenomic analysis of an extremely complex soil microbial community using unassembled sequencing reads, with a focus on describing the functional diversity of the predominant genes present in this environment rather than attempting to link sets of genes to particular microbial species. From 100 megabases of shotgun genomic sequencing as well as two thousand 16S rRNA sequences we were able to demonstrate that the 0.5 g soil sample contains upwards of three thousand distinct microbial species. The paucity of overlapping reads in the genomic sequence suggested that at least a gigabase of sequence data would be needed to develop even a cursory assembly of the most predominant species in the sample. Despite the organismal complexity of this community and the absence of assembled genomes, gene predictions from individual sequence reads revealed much of the functional diversity following a limited amount of sequencing. The relative allocation of proteins into different functional categories in this and other environmental samples revealed a habitat-specific functional fingerprint. These studies suggest that the repertoire of predominant proteins coded for in the sequence data, even in a complex soil community, can be captured following a moderate amount of sequencing and that the particular functions represented serve as identifiers for the environment from which the sample was collected.

INSIGHT INTO COLD ADAPTATION IN ARCHAEA FROM AN INTEGRATED GENOMIC/PROTEOMIC APPROACH

Rick Cavicchioli*, Amber Goodchild, Neil Saunders, Mark Raftery
The University of New South Wales, Sydney, Australia

Temperature is a critical environmental factor controlling the evolution and biodiversity of life on Earth. The majority of the Earth's biosphere is permanently cold. Archaea contribute significantly to biomass in cold environments, although only a handful have been isolated.

The methanogens stand out as the only group of organisms that have species capable of growth at 0°C and 110°C. Through our genome sequencing studies of two Antarctic methanogens, *Methanogenium frigidum* and *Methanococcoides burtonii*, we compared the genomes of species incrementally spanning this entire temperature range. Distinguishing features of psychrophiles from the study included structural and compositional features of proteins and tRNA. (Saunders et al, 2003, Genome Research, 13: 1580-1588).

Underpinned by the genome sequence data, we developed proteomics using two dimensional electrophoresis (2DE) and liquid chromatography-mass spectrometry (LC-MS), and have achieved the first global view of the biology of cold-adaptation in *M. burtonii*.

Proteins specific to growth at 4°C versus 23°C (T_{opt}) were identified by 2DE, and 43 differentialy expressed proteins were identified by MS. mRNA levels were also determined for all 43 genes. Cold adaptation was linked to changes in fundamental cellular processes including metabolism, transcription and protein folding, with key roles identified for the pyrrolysine-containing trimethylamine methyltransferase, RNA polymerase subunit E, a response-regulator from a two component regulatory system and peptidyl prolyl *cis/trans* isomerase. Increased levels of the heat shock protein, DnaK were observed during growth at 23 C, indicating that growth of cold-adapted organisms at apparently optimum temperatures is stressful. (Goodchild et al, 2004, Molecular Microbiology, 53: 309-321.)

Gene annotation and gene organisation from the draft genome was linked to LC-MS analysis of the expressed-proteome to define the key biological processes functioning at 4 C. 528 proteins ranging in p*I* from 3.5 to 13.2, and 3.5 – 230 kDa, were identified. Knowledge of the expressed proteins advances our level of understanding of the biology of the cell from coding potential to actual process, and highlights a mechanistic complexity that must be managed by the cell. For example, the eucaryotic-like, core RNA polymerase machinery is simultaneously interacting with bacterial-like anttiterminator and numerous response regulator proteins to enable transcription. As a second example, transposases are expressed during normal, steady-state-growth and are therefore not just dormant relics of past evolutionary events, but are active and likely to be affecting the cell's genetic composition and structure. The expression of the transposases therefore has major implications, ranging from effects on the genetic diversity and fitness of natural populations through to a previously unrealised value for the development of genetic tools in *M. burtonii*. Characteristics of the fundamental cellular processes inferred from the expressed-proteome highlight the evolutionary and functional complexity existing in this domain of life. The study also demonstrates the capacity to perform high-throughput analyses of proteins from psychrophiles. (Goodchild et al, 2004, Journal of Proteome Research, in press).

A versatile Nascent Associated Complex in the Archaea: Evidence for a combination of chaperone and protein degradation roles.

P. Laksanalamai* and F. T. Robb
Center of Marine Biotechnology, University of Maryland Biotechnology Institute
701 E. Pratt St. Baltimore, MD 21202.
Tel 410.234.8872, Email: laksanal@umbi.umd.edu

We recently noted that molecular chaperone systems in hyperthermophilic archaea are considerably less complex than those in eukaryotes[1]. In this study, we have investigated a novel chaperone, the archaeal nascent polypeptide associated complexes (NACs). Most eukaryotes express several NAC homologs and the complex contains several subunits[2]. Database searches revealed that all the archaea genomes contain a single highly conserved NAC homolog, except for the obligate symbiotic archaeon, *Nanoarchaeum equitans*, which lacks the NAC encoding genes. Further analysis of the full-length archaeal NAC proteins by Pfam and SMART revealed that all of them appear to have a highly conserved domain known as a ubiquitin associated domain (UBA domain)[3]. This arrangement occurs in rarely eukaryotes. Recent studies revealed that the archaea may have ubiquitin and database searches indicated that ubiquitin homologs occur in some archaeal genomes. Several lines of experimental evidence also indicated that most archaea possess proteasomes.

We have cloned, overexpressed and purifed *Pyrococcus furiosus* NAC (Pfu-NAC) protein and found that it has chaperone properties. The characterization of the archaeal NAC proteins based on their sequence homology, amino acid motifs and composition, domain architecture or the presence of protein degradation machines, leads us to propose possible combination roles for archaeal NAC proteins as follow: (1) they may function as "holdase" chaperones by interacting with unfolded nascent polypeptides to prevent premature aggregation (2) they may be involved in protein quality control processes by targeting unfolded proteins to appropriate protein degradation machines. This potential dual function of NAC proteins may strengthen our hypothesis that the archaea have minimal sets of protein folding chaperones.

References:

1 Laksanalamai, P. et al. (2004) Minimal protein-folding systems in hyperthermophilic archaea. *Nat Rev Microbiol* 2 (4), 315-324

2 Rospert, S. et al. (2002) Nascent-polypeptide-associated complex. *Cell Mol Life Sci* 59 (10), 1632-1639

3 Hofmann, K. and Bucher, P. (1996) The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem Sci* 21 (5), 172-173

Comparative Genomics and Genome Organization of Microbial Pathogens

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

Bordetella pertussis, *Bordetella parapertussis*, and *Bordetella bronchiseptica* are very closely related Gram-negative beta-proteobacteria that colonize the respiratory tracts of mammals. *B. pertussis* is a strictly human pathogen of recent evolutionary origin and is the primary agent of whooping cough. *B. parapertussis* can also cause whooping cough, and appears to be restricted to humans, although different strains can colonise sheep. *B. bronchiseptica* causes chronic respiratory infections in a wide range of animals. We have recently sequenced the genomes of *B. bronchiseptica* RB50, *B. parapertussis* 12822 and *B. pertussis* Tohama I.

Despite their close similarity (*B. bronchiseptica* and *B. parapertussis* are 100% identical at the 16S rRNA level, and *B. pertussis* is 99.8% identical to each of them), the genomes have remarkably different sizes; *B. bronchiseptica* is 5,338,400 bp, with 5007 predicted genes; *B. parapertussis* is 4,773,551 bp, with 4404 predicted genes; and *B. pertussis* is just 4,086,186 bp, with 3816 predicted genes. The genomic analysis indicates that *B. parapertussis* and *B. pertussis* are independent derivatives of *B. bronchiseptica*-like ancestors and the differences in size are primarily due to DNA loss in the derivative species. This DNA loss is a primarily a consequence of IS-element expansion, and subsequent recombination. The degree of IS element expansion is extreme – *B. parapertussis* contains 31 IS elements, while B. pertussis contains over 260 (1 per 15 kb), with the 1 kb IS*481* element being present in 238 copies. None of these elements is present in the sequenced *B. bronchiseptica*. In addition to DNA loss, these IS elements have also caused a large amount of genome rearrangement and gene inactivation; nearly 10% of the genes in B. pertussis are predicted to be non-functional.

The possible reasons for these IS element expansions, and their consequences for the differences in genome structure, host adaptation and virulence amongst these organisms will be discussed.

Comparative Genomics of Vector-borne Pathogens

A.C. Frank*, E.O. Karlberg, C.M. Alsmark, B-A. Legault, D.H. Ardell, A.S Eriksson, K. Näslund and S.G.E. Andersson. Department of Molecular Evolution, Uppsala University, Sweden.

We have sequenced the complete genomes of two human pathogens, *Bartonella quintana* (1.58 Mbp) and *Bartonella henselae* (1.93 Mbp). The two species are closely related, but differ in their reservoir and vector ecology. Whereas *B. quintana* is a specialist restricted to humans and transmitted by the body louse, *B. henselae* is frequently isolated both from humans and felines, and can use both fleas and ticks as vectors. The genomes are overall very similar and both show evidence of genome reduction. However, in *B. quintana* the genome degradation process is more advanced, and a prophage and phage-related islands present in *B. henselae* have been lost. A broader analysis aimed to place *B. henselae* and *B.quintana* in the perspective of alpha-proteobacterial evolution suggests that they evolved from a plant- pathogen or symbiont through an animal-associated species with more than one chromosome. Both genomes are reduced versions of chromosome I from the closely related pathogen *Brucella melitensis*. Bartonella-specific genes, some of which are most likely involved in host interaction, are concentrated to a 250-kb region in the fourth quarter of the genome. This region is flanked by two rRNA operons and shows segmental similarity to genes located on chromosome II of *B. melitensis*, suggesting that it was aquired though integration of a secondary replichon or a section thereof. The comparisom with other alpha-proteobacteria also shows a striking parallel between *Rickettsia prowazekii* and *B. quintana*, which have identical lifestyles, and are both smaller subsets of their closest relatives. This suggests an association between host- and vector range and accelerated rate of genome degradation.

A Reductionist Approach to E. coli Genomics

By Frederick R. Blattner
University of Wisconsin, Madison.

Since the complete genome of *E. coli* K-12 was determined, approximately 17 additional strains of *Enterobacteriaceae* have been sequenced. Comparison of these genome sequences reveals a remarkable division into lineage specific islands peppering a core genome whose elements are present in most or all of the genomes. The former are thought to be distributed among the species by horizontal gene transfer mechanisms while the core is thought to evolve primarily by descent. In comparing the sequences of four strains of *E. coli* that we sequenced, the core genome consists of about 3900 genes in a reasonably conserved order whereas the and the laterally transferred islands account for 500 to 1800 genes depending on the strain. When the comparison is, extended to greater evolutionary distance e.g. *Yersinia pestis, vs E. coli* the core genome is about 2500 genes and the order is less well conserved although recognizably syntenic.

In order to examine the properties of the core genome, absent the influence of horizontally transferred elements, we are systematically reducing the genome of *E. coli* K-12 by a series of scarless deletions. Current progress rests at 43 deletions removing 729 genes totaling 707,955 base pairs (15.3% of the genome). This leaves about 3500 genes in the reduced genome.

The genes removed include IS and RHS sequences, cryptic phages, pathogenicity islands, flagellar structural genes, genes for fimbria, adhesins and toxins as well as a few genes for catabolism of substances not often found in lab culture media. The construction was guided by the principle that the resulting cell grow on minimal salts medium and this was achieved.

The growth properties of MDS 43 are quite similar to wild type including the ability to be propagated to high density in a laboratory fermenter, although the doubling time is slightly longer and lag phase somewhat increased. Transformation frequencies by electroporation are normal and recombinant proteins are expressed to a slightly higher level. Point mutation rates are almost unchanged but mutation due to IS hopping is eliminated. This has resulted in elimination of cloning artifacts. For the most part gene expression profiles are normal for the genes that have not been deleted.

The use of a simplified genome will have many applications, including the possibility to simplify the systems biological modeling of this most important model organism, *E. coli.*

Formation and Composition of the *Bacillus anthracis* Endospore.

N. H. Bergman*, Hongbin Liu, Brendan Thomason, Shamira Shallom, Alyson Hazen, Joseph Crossno, David A. Rasko, Jacques Ravel, Timothy D. Read, Scott N. Peterson, John Yates, III. & Philip C. Hanna. University of Michigan Medical School, Scripps Research Institute, The Institute for Genomic Research, Naval Medical Research Center.

The endospores of *Bacillus anthracis* are the infectious particles of anthrax. Spores are dormant bacterial morphotypes able to withstand harsh environments for decades, which contributes to their ability to be formulated and dispersed as a bioweapon. We monitored gene expression in *B. anthracis* during growth and sporulation using full genome DNA microarrays, and matched the results against a comprehensive analysis of the mature anthrax spore proteome. A large portion (~36%) of the *B. anthracis* genome is regulated in a growth phase-dependent manner, and this regulation is marked by 5 distinct waves of gene expression as cells proceed from exponential growth through sporulation. The identities of more than 750 proteins present in the spore were determined by multi-dimensional chromatography and tandem mass spectrometry. Comparison of data sets revealed that while the genes responsible for assembly and maturation of the spore are tightly regulated in discrete stages, many of the components ultimately found in the spore are expressed throughout and even before sporulation, suggesting that gene expression during sporulation may be mainly related to the physical construction of the spore, rather than synthesis of eventual spore content. These findings contribute to our understanding of spore formation and function, and also highlight the complementary nature of genomic and proteomic analyses. We are currently expanding these studies to include analyses of *B. anthracis* germination and growth within the mammalian macrophage using both microarray and proteomic techniques.

Deinococcus radiodurans' transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance

M. Tanaka, A. M. Earl, & J. R. Battista*. Louisiana State University, Baton Rouge.

During the first hour after a sub-lethal dose of ionizing radiation, 72 genes are up-regulated three fold or higher in *D. radiodurans* R1. Thirty-three of these loci were also among a set of 73 genes expressed in R1 cultures recovering from desiccation. The five transcripts most highly induced in response to each stress are the same and encode proteins of unknown function. The genes (designated *ddrA*, *ddrB*, *ddrC*, *ddrD*, and *pprA*) corresponding to these transcripts were deleted, both alone and in all possible two-way combinations. Characterization of the mutant strains defines three epistasis groups that reflect different cellular responses to ionizing radiation-induced damage. The *ddrA* and *ddrB* gene products have complementary activities and inactivating both loci generates a strain that is more sensitive to ionizing radiation than strains in which either single gene has been deleted. These proteins appear to mediate efficient RecA-independent processes connected to ionizing radiation resistance. The *pprA* gene product is not necessary for homologous recombination during natural transformation, but nevertheless may participate in a RecA-dependent process during recovery from ionizing radiation-induced damage. These characterizations clearly demonstrate that novel proteins significantly contribute to the ionizing radiation resistance in *D. radiodurans* R1.

Life in hot acid – implications from the genome sequence of the extreme thermoacidophile *Picrophilus torridus* about how to thrive around pH 0

O. Fütterer[1], A. Angelov[1], H. Liesegang[1], G. Gottschalk[1], C. Schleper[2], B. Schepers[3], C. Dock[3], G. Antranikian[3], W. Liebl[1]*

[1] Institut of Microbiology und Genetics, Grisebachstr. 8, D-37077 Göttingen, Germany
[2] Institute of Microbiology and Genetics, Schnittspahnstr. 10, D-64287 Darmstadt, Germany
[3] Technical Microbiology, Kasernenstr. 12, D-21073 Hamburg, Germany

Picrophilus torridus and *P. oshimae* are euryarchaea able to grow at pH values around pH 0 at growth temperatures up to 65°C, thus they represent the most thermoacidophilic organisms known. The high specialization of the obligately aerobic heterotroph, *P. torridus*, for growth in extremely hot and acidic habitats is evident from its inability to grow at pH values above 4.0 and its low intracellular pH (4.6), and makes it a model organism to study the strategies used to cope with extreme thermoacidophilic conditions. Sequence analysis of the *P. torridus* genome yielded a 1.55 Mb large single circular chromosome wich encodes 1,536 proteins. The following features deduced from comparative genome analysis may contribute to the thermoacidophilic survival strategy of *P. torridus*: (1) The small genome of *P. torridus* with the highest coding density among thermoacidophiles indicates that genome reduction may have been important for the evolution of adaptation to the combination of extremely low pH and high temperature. (2) Comparison of the amino acid composition of *P. torridus* proteins with values deduced from other genomes suggests that perhaps an increased (surface) hydrophobicity may contribute to their acid tolerance. (3) An exceptionally high ratio of secondary over primary ATP-consuming transporters demonstrates that the high proton concentration in the surrounding medium is extensively used for solute transport. (4) It appears that certain genes which may be particularly supportive for the extreme lifestyle of *P. torridus* have been internalized into the genome of the *Picrophilus* lineage by horizontal gene transfer from the phylogenetically distant crenarchaea and bacteria. Also, it is noteworthy that thermoacidophiles from phylogenetically distant branches of the *Archaea* apparently share an unexpectedly large pool of genes, since 66 % of all *P. torridus* genes were found to have orthologs in the related euryarchaeon *Thermoplasma acidophilum* but also 58 % in the crenarchaeon *Sulfolobus solfataricus*. Finally, results on the characterization of selected *P. torridus* enzymes such as glucose dehydrogenase, the first enzyme of the non-phosphorylated Entner-Doudoroff pathway variant of this organism, will be presented.

Lessons from Comparative Genomics and Proteomics of Filamentous Fungi

S.E. Baker*, E.A. Panisko, J.K. Magnuson, Z. Dai, C.F. Wend, L.L. Lasure. Pacific Northwest National Laboratory, Richland, WA.

The explosion in genome sequence data has accelerated discoveries in fungal biology. These discoveries aid in fungal bioprocess development as well as strain and gene discovery efforts, leading to biobased products as well as a greater understanding of secondary metabolite evolution and production. The work in our laboratory is directed at understanding extracellular enzymes, especially as it relates to biomass conversion to valuable products and morphology of fungi as it relates to the hyperproductivity observed in some fungal fermentation processes. We have used genomic and global proteomic approaches to learn about morphology control in *Phanerochaete chrysosporium,* a lignin degrading basidiomycete fungus whose genome was sequenced by the US Department of Energy's Joint Genome Institute (JGI) and *Aspergillus niger,* a filamentous ascomycete fungus whose genome sequence we recently acquired.

A SAGE approach to fungal pathogenesis.

J.W. Kronstad* T. Lian, B. Steen, K. Tangen, G. Hu, S.J.M. Jones, M. Marra & C. D'Souza. University of British Columbia; Genome Sciences Centre, B.C. Cancer Agency, Vancouver.

The opportunistic fungal pathogen *Cryptococcus neoformans* (capsular serotypes A and D) causes central nervous system infections in immunocompromised people such as AIDS patients. In addition, serotype B strains are primary pathogens as demonstrated by an outbreak of infections in otherwise healthy people in British Columbia. Our goals are to develop genomic resources for *C. neoformans* and to use these to understand cryptococcosis. For the first goal, we constructed physical maps and contributed sequence data for five genome projects on strains representing the three serotypes, and we have sequenced the genome of a serotype B strain. The sequence data allowed us to address our second goal through the use serial analysis of gene expression (SAGE) to examine the transcriptome as a function of serotype, temperature, iron deprivation and infection in a rabbit model of meningitis. These data support annotation efforts and provide a foundation for exploring virulence. For example, the SAGE data revealed a striking difference in the temperature response between serotype A and D strains of different virulence, and also suggested a similarity between the transcriptomes of cells from an infection and cells grown in low iron medium. We have focused most of our analysis on the SAGE data for cells grown in low-iron versus iron-replete media. This condition increases expression of the polysaccharide capsule that is the major virulence factor of the fungus. We found that iron deprivation results in substantial remodeling of the transcriptome and abundant transcript levels for an exported mannoprotein called Cig1. Mutants defective in *cig1* fail to grow in low-iron medium and display defective iron sensing under iron-replete conditions.

Title: Comparative functional genome and evolutionary analysis of *Wolbachia*, *Rickettsia*, and *Ehrlichia* genomes using ERGO bioinformatic suite.

Vinayak Kapatral* Integrated Genomics, 2201, West Campbell Park Drive.
Chicago, Il 60612.

Email: vinayak@integratedgenomics.com

Using the ERGO bioinformatic suite developed at Integrated Genomics, we present whole genome analysis of the intracellular Gram-negative bacterial genomes such as *Wolbachia spp* (Droshophila), *Rickettsia conorii* (Malsh7), *R. prowazekii* (Madrid E), *R. sibirica* and *Ehrlichia chaffeensis* (Ark). These organisms belong to α protobacteria and are found in diverse hosts such as insect, nematode, and humans. A total of 2,151 protein clusters common to all the genomes have been identified. Protein clusters unique to each and shared amongst genomes have been calculated. A large number of genes/pathways common to all and those that were lost during co-evolution with their hosts have been identified. In general, *Wolbachia spp* appear to have lost fewer metabolic genes/pathways compared to human pathogens such as *Rickettisia spp* and *Ehrlichia spp*. For example, cell division gene *mraZ* is absent in *Wolbachia spp* but present in all others, however, all of them seem have lost other cell division genes such as *ftsE, ftsL, ftsX, zipA* Each of these bacteria seem to have acquired gene/gene clusters from other organism perhaps by type-IV secretion system (T4SS), some of which include genes coding for unique metabolic pathways (threonine aldolase and GDS L family lipase in *Wolbachia spp*), LPS (in *Rickettsia spp*), drug resistance (in *Wobachia spp, Rickettsia spp*), etc. Such acquisition gives selective advantage and survival in diverse hosts. Functional analysis of protein clusters in biological context and their evolutionary significance will be presented.

Bacterial Annotation Tools

Owen White
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850

Manatee (MANual Annotation Tool, Etc Etc) is a graphical user interface designed to manage data in a common database that allows multiple users to simultaneously operate on that information. Production annotation teams at TIGR, as well as outside collaborators, have been using this system for the past number of years to obtain essential annotation information in a user-friendly way. This architecture also serves as the basis of a system called Sybil that allows scientists to discover, evaluate, and summarize intra-species variation. Sybil permits comparative analysis in two broad areas: 1) the interface serves as web resource and may be used on-line by scientists interrogating data from TIGR resident in our database, and 2) users may install the Sybil system on their local computers to perform custom analyses of their data. These interfaces, phylogenetic analyses, global chromosomal rearrangements, and high-throughput assessment of orthologs will be shown in my presentation.

Assigning names to all predicted proteins in a complete genome, even when done with exquisite accuracy, provides only an initial layer of understanding of the activities in a cell. Genome annotation is made more complete when individual genes are placed in context of "Genome Properties" such as metabolic pathways, coordinated cellular activities or cellular structures. A Genome Property is a single assertion (a numerical value, a truth state such as "Yes" or "No", or a controlled vocabulary term such as "facultative anaerobic", "motile") for some attribute. We have defined 172 properties resulting in over 17,500 property assertions after their application to 145 completed prokaryotic genomes. They include computed values such as DNA GC content, manually set phenotypic properties such as human pathogen and rule-based properties that identify metabolic pathways. We will report on the accuracy of these methods, their application for comprehensive comparison of bacterial genomes, and their utility in identifying the function of previously uncharacterized genes.

Whole Genome Analysis of Microbial Genomes using High-Density Oligonucleotide Microarrays

Thomas Albert[1*], Jason Norton[1], Michael Molla[2,3], Rebecca Selzer[1], Todd Richmond[1], Dennis Kopecko[4], Fred Blattner[5], and Roland Green[1]
[1]NimbleGen Systems Inc., Madison, WI 53711
[2] Computer Sciences Department, University of Wisconsin-Madison, Madison, WI 53706
[3] Living Circuit L.L.C., Madison, WI, 53703
[4] Food and Drug Administration
[5] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706

High-density microarray-based resequencing is an accurate and high-throughput technique for analyzing complex biological samples. Despite the ability to resequence tens of thousands of nucleotides per array, large numbers of arrays are required to resequence whole microbial-sized genomes. In this report we describe a rapid, cost-effective, and accurate technique called comparative genomic resequencing (CGR). This strategy utilizes custom oligonucleotide microarrays and combines the techniques of high-resolution comparative genomic hybridization with resequencing to provide unprecedented resequencing throughput and accuracy directly from whole genome preparations. Here we demonstrate CGR on the 4,791,958 base-pair genome of the typhoid bacterial vaccine strain, *Salmonella* Typhi Ty21a, and compare it with traditional array-based resequencing (ABR). The CGR approach resulted in a significantly lower false positive rate when (estimated to be 1 false positive per ~400,000 bases for CGR vs. 1 false positive per ~45,000 bases for ABR), identified >95% of the estimated total number of SNPs present in the genome, and required just 6 arrays to analyze each strain vs. 101 arrays for ABR. The CGR strategy also produces a high-resolution map of sites of insertion, deletion, transversion, and amplification. CGR allows up to 1 Mb to be resequenced per array on average, allowing a single laboratory technician to generate up to 20 Mb of accurate sequence per day without automation, allowing whole genome analysis across microbial populations.

Identifying modular transcriptional network by random matrix theory

Zhou, J.-Z., F. Luo, Y. Yang, J.-X. Zhong, and H. Gao

Large-scale sequencing of entire genomes represents a new age in biology, but the grand challenge is to define the biological roles of large portions of functionally unknown genes, and genetic interaction networks consisting of many individual functional modules. Although genomic technologies such as microarrays provide powerful tools for identifying cellular interaction networks, defining such modules without ambiguity is very difficult because all current methods rely on artificially chosen thresholds and hence the results are subjective. Here we present a novel, reliable, sensitive and robust approach for automatically identifying functional modules using a mathematically defined threshold predicted by random matrix theory (RMT). Applying this approach to microarray data from yeast, human, *Escherichia coli* and *Shewanella oneidensis* demonstrated that it correctly identifies functional modules with the expected properties consistent with general network theory. Experimental validation on the predicted functions of 10 unknown genes from yeast and *Shewanella* indicated that this approach is useful for predicting the functions of unknown genes. This approach will be ideal for analyzing high throughput genomics data for modular network identification and gene function prediction.

WORKSHOP ABSTRACT

Pathway Tools Software for Creation and Analysis of Pathway/Genome Databases
P. Karp, M. Green, S. Paley, J. Pick, SRI International, USA

This talk describes many new developments in the Pathway Tools software, which provides an environment for creation, editing, analysis, and Web publishing of organism-specific Pathway/Genome Databases (PGDBs). A user community of 70 groups have created close to 40 PGDBs, making Pathway Tools an emerging standard for microbial model-organism databases.

Given an annotated genome as its input, the software creates a new PGDB that describes the genes, replicons, and proteins described in the input file. The software next computationally predicts the operons and the metabolic pathways of the organism. The software can also predict which genes code for the many enzymes that are typically missing from predicted metabolic pathways.

Editing tools allows users to update the PGDB to reflect new findings about the gene functions, metabolic pathways, and genetic network of the organism. A newly developed ontology of evidence codes allows a PGDB to clearly distinguish computational predictions from experimentally derived information. A new set of editing tools allows users to annotate protein modification sites, active sites, and domains, which are displayed graphically in the protein-display window.

The software allows publishing of a PGDB through a web site to support querying and visualization of the genome, genetic network (operons and regulons), transporters, metabolic network (including pathways, reactions, and substrates). Analysis tools include animated display of gene expression data onto the complete pathway map for the organism

PGDBs are platforms for disseminating the evolving knowledge about the organism using the Web publishing facilities of the software (which power the Web site at URL http://BioCyc.org/). Pathway Tools is freely available to academics and runs on Linux, Windows, and Sun computers; see URL http://bioinformatics.ai.sri.com/ptools/.

POSTER ABSTRACTS

Correlation between codon usage, codon adaptation index, and gene expression in *Lactobacillus acidophilus* NCFM

Eric Altermann*[1], M. Andrea Azcarate-Peril[1], Raul Cano[2], and Todd R. Klaenhammer.

1.Dept. of Food Science, Southeast Dairy Foods Research Center, North Carolina State University, Raleigh, NC, USA; 2. Environmental Biotechnology Institute, California Polytechnic State University, San Luis Obispo, USA

The genome of *L. acidophilus* NCFM, a probiotic lactic acid bacterium, was completely sequenced and automatically annotated using the GAMOLA engine. A combined approach of in *silico* and expression analysis studies is now required to identify gene sets potentially involved in probiotic characteristics. Predictions of highly expressed genes, using codon usage and codon adaptation indices, have been based mostly on gene sets commonly assumed to be highly expressed, for example genes associated with housekeeping functions. However, little information is available about the influence of varying growth conditions and environmental stress factors on the expression of genes and their *in silico* prediction. In this study we used a whole genome microarray representing 97.5% of the genome of *L. acidophilus* NCFM to analyze transcriptional patterns under different stress conditions (acid, bile, ethanol, and ammonium oxalate). Microarray data were used to analyze changes in codon usage and codon adaptation indices based on highly expressed and induced genes. Indices calculated on highly expressed genes showed a significantly lower overall adaptation index (0.3) relative to indices based on induced genes (0.44). The results indicated that codon adaptation in *L. acidophilus* has evolved to ensure a more efficient translation under stress conditions

A Genome-Scale Metabolic Reconstruction of Staphylococcus Aureus N315

S.A. Becker* and B.O. Palsson. University of California, San Diego.

Staphylococcus aureus is an important pathogen that causes a multitude of potentially life-threatening infections and has developed resistance to nearly all available antibiotics. A genome-scale metabolic model (iSB623) of *S. aureus* strain N315 has been reconstructed, accounting for 623 genes, their transcripts, and 657 intracellular reactions. The reconstruction process depended largely on genome sequence data due to the relative scarcity of available biochemical data specific to this organism.

Based on data available for the related and better-characterized organism *Bacillus subtilis*, we devised a biomass function quantifying the relative production of many of the molecules required for the growth of *S. aureus*. Using this biomass function, optimal growth phenotypes of *S. aureus* have been predicted *in silico* under diverse conditions.

Each gene in the reconstruction is linked with at least one protein and reaction through the use of GPR (gene-protein-reaction) associations. These associations, when combined with the biomass function, allowed us to predict lethal gene deletions. While the lethality of many of these deletions is intuitively obvious, some of the predictions are not apparent without a systems-level analysis. Consideration of the proteins encoded by predicted essential genes yielded potential drug targets.

Transcriptome and Proteome Analysis of a *Shewanella oneidensis etrA* Deletion Mutant

S. D. Brown*[1], H. Gao[1], Y. A. Gorby[2], M. Lipton[2], H. M. Mottaz[2], G.E. Pinchuk[2], E. Hill[2], X. Wan[3], C. Bilskis[2], J.K. Fredrickson,[2] J. Zhou[1]; [1]Oak Ridge National Laboratory, Oak Ridge, TN. [2]Pacific Northwest National Laboratory, Richland, WA. [3]University of Missouri, Columbia, MO.

Shewanella oneidensis strain MR-1 has diverse respiratory capabilities and the ability to reduce and/or precipitate a wide range of metals and radionuclides. The *S. oneidensis* EtrA has a high degree of identity to *Escherichia coli* Fnr (73.6%). Previous characterization of an *etrA* mutant in our laboratory using partial-genome microarrays showed responsive genes fell into five putative functional categories: (i) electron transport, (ii) intermediary carbon metabolism, (iii) transcription regulation, (iv) substrate transport and binding, and (v) biosynthesis, assembly, and other cellular processes. In order to better understand regulation of electron transport in *S. oneidensis* we examined the transcriptomes and proteomes of a *S. oneidensis ΔetrA* deletion mutant and the parental strain. In this study, chemostats were used to collect aerobic, microoxic and anaerobic steady state samples as well as aerobic to microxic transition phase samples. In addition the transcriptomes from the *etrA* deletion and parental strains grown in batch culture were also compared. The chemostat and batch culture samples were hybridized to a whole genome microarray of *S. oneidensis* that covered approximately 99% of genome. The combined proteome and transcriptome analysis of the *etrA* mutant is an initial step toward a better understanding of the contribution of EtrA in *S. oneidensis* regulation.

Nebulon: a system for the inference of functional relationships of gene products from the recombination of predicted operons

Sarath Chandra-Janga[1], Julio Collado-Vides[1], and Gabriel Moreno-Hagelsieb[2]*
[1]Program of Computational Genomics, CIFN-UNAM, Apdo Postal 565-A, Cuernavaca, Morelos, 62100 Mexico. [2]Department of Biology, Wilfrid Laurier University, 75 University Avenue West, Waterloo, ON, Canada, N2L 3C5.
*Presenting Author (gmorenohagelsieb@wlu.ca)

Here we introduce Nebulon, a system to build networks of predicted functional relationships of gene products based on their organization into operons in any available genome. The method is based on a previously developed method to predict operons by the distances between adjacent genes in the same strand (1), and on the high variability of operon associations across genomes that can reveal functional relationships among gene products (2). Our system uses different kinds of threshold to accept a functional relationship, from those related to the prediction of operons, to finding the association in at least a given number of non-redundant genomes. We also work by layers, meaning that we decide on the number of iterations to allow for the complementation of related gene sets. The method shows high reliability as revealed by an analysis based on Gene Ontologies (3) and on relationships as found in the KEGG metabolism database (4). We also exemplify with several known and characterized gene sets.

References

1. Moreno-Hagelsieb, G. and Collado-Vides, J. (2002) A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics*, **18 Suppl 1,** S329-336.
2. Galperin, M.Y. and Koonin, E.V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol*, **18,** 609-613.
3. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, **32 Database issue,** D258-261.
4. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res*, **32 Database issue,** D277-280.

Nebulon-NetView: a web tool for visualization and analysis of genomic networks of functional interactions.

Diaz-Peredo Edgar[1], Sánchez-Solano Fabiola[1], Chandra-Janga Sarath[1], Collado-Vides Julio[1], and Moreno-Hagelsieb Gabriel[2].

[1]Program of Computational Genomics. Center of Nitrogen Fixation. UNAM. Cuernavaca, A.P. 565-A, Morelos, 62100, México.
[2]Department of Biology. Wilfrid Laurier University. Waterloo, ON N2L 3C5

Abstract.

Nebulon-NetView is a Java applet display application based on the graphLayout software of Sun Microsystems, Inc (J2SE platform). This interactive tool graphs the complete genomic network of interactions inferred from the functional classification provided by distance-based operon predictions across genomes (Nebulome generator)[1]. The graph shows the connectivity of the genetic network from the perspective of a selected gene and gives the global view of the complexity of the interactions of this gene with the rest of the genome, as well as the modularity and density of the sets of genes working on essential tasks in the cell. The graph nodes represent genes and the connecting lines or edges represent the inferred functional interactions. The nodes can be clicked to get regulatory information from Gene Ontologies, from the Clusters of Orthologous Genes (COGs), and from RegulonDB. The tool has an attractive graphic interface where the user can select from a list of 169 genomes with the possibility to display one or two levels of functional connectivity. It is also possible to search by gene name or GI number and sort the network through successive interactive zooms. In addition, the user can select to display only the portion of a gene network representing exclusively the interactions of a regulatory type.

1 Moreno-Hagelsieb Gabriel, *et al*, J. (2004) *In preparation.*

Genomic Diversity of *Bacillus subtilis*

A. M. Earl* & R. G. Kolter. Harvard Medical School, Boston, MA.

The long held belief that members of a particular bacterial species are genetically homogeneous has been recently challenged. Advances in molecular techniques, most notably whole genome sequencing, have revealed a surprising degree of genomic plasticity among species members. These findings have revolutionized the way in which we think about bacterial genetic diversity and raise a number of intriguing questions about bacterial evolution, including the nature and dynamic of selective pressures encountered within an organism's environment. Thus far, there are only few bacterial species with more than one fully sequenced representative and until whole genome sequencing becomes routine this is not a realistic approach to rapidly assess genetic variation among different strains of the same species. We are currently utilizing genomotyping, a microarray based hybridization technique, to examine genomic diversity among multiple isolates of the ubiquitous soil-dweller, *Bacillus subtilis*. Our results suggest that unlike *B. subtilis'* close relative, *B. anthracis*, members of this species do exhibit a relatively large degree of genetic variation. We are using the hybridization profiles (gene absence versus presence) to 1) construct phylogenetic relationships among strains in an effort to understand the evolution of this bacterium, and 2) inform investigation into specific gene regions that may be important in ecological adaptation.

Applying Phylogenetically-Based Comparative Methods to Bacterial Genome Evolution

M. P. Francino* & J. L. Boore. Department of Evolutionary Genomics, DOE Joint Genome Institute, Walnut Creek CA.

By taking advantage of the availability of numerous and diverse bacterial genome sequences, we have obtained a well-resolved global bacterial phylogeny on which we are tracing the evolution of genome level characters and molecular evolution parameters. In order to determine absolute rates of evolution, we are employing maximum likelihood to estimate the times of divergence among species in the absence of a general molecular clock, and we are employing the comparative method to detect correlations between characters independent of phylogeny. Our initial analyses show that the size and GC content of a bacterial genome are negatively correlated with the rate of evolution of its encoded proteins. The implications of these correlations for genome evolution are discussed, in a context of selection, mutational load and cell economics.

An ecological perspective on the evolution of gene families and the diversification of gene functions.

M. Pilar Francino. Department of Evolutionary Genomics, DOE Joint Genome Institute, Walnut Creek CA.

In this poster, I address the subject of evolution of new gene functions. This subject is of fundamental importance in genome evolution, but, although several models have been hereto proposed, none has been fully satisfactory. I am presenting a new model, based on a thorough review and reinterpretation of the current literature. I call this model the "adaptive radiation" model of evolution of new gene functions, because it draws close parallels to phenomena observed in selection experiments in bacterial genetics, and to the patterns observed in nature during the radiation of lineages into new ecological niches. The "adaptive radiation" model postulates that new gene functions evolve in rapid, punctuated bursts when new biochemical niches appear, through large, selected amplifications of
the best preadapted genes followed by competition among the paralogous gene copies.

Comparative Community Genomics of the Gut Microbiota

J. Froula[1], J. Morgan[1], Nila Bala[1], H. Ochman[2] & M. P. Francino[1*]. [1]Department of Evolutionary Genomics, DOE Joint Genome Institute, Walnut Creek, CA and [2]Department of Biochemistry & Molecular Biophysics, University of Arizona, Tuscon.

The microbiota of the human gastrointestinal (GI) tract plays essential roles in health, including a significant contribution to the digestive process, promotion of gut maturation and integrity and modulation of the immune system. Moreover, the microbiota interacts with pathogenic agents in several complex ways. On one hand, resident bacteria exert a protective barrier effect against enteropathogens; but, on the other, they could contribute to enrich the arsenal of incoming pathogens through horizontal transmission of genes involved in host-microbe interaction or antibiotic resistance. In addition, many normally benign GI commensals have the potential to become opportunistic pathogens in compromised hosts. Elucidating the composition and coding capabilities of the GI microbiota is therefore crucial for a comprehensive analysis of infectious disease.

We are producing large-insert bacterial artificial chromosome (BAC) libraries from bacterial genomic DNA isolated directly from fecal samples. The availability of BAC libraries will provide extensive genomic sequences that will serve to elucidate the phylogenetic positions and coding capabilities of the members of this community. Given that the composition of the GI microbiota varies greatly with age and diet, we have chosen to generate BAC libraries from two very distinct stages: adult and breast-feeding infant, as represented by mother and child. To maximize our insight into the evolution and ecology of infectious disease, we will focus our sequencing efforts towards genomic regions relevant to pathogenicity and other ecological interactions, both among the microbial community members and between microbes and host.

A first pass BAC library has been constructed consisting of 50-70 Kb genomic inserts. The final library will contain 20,000 clones generated from EcoRI digested DNA and the same number generated from MboI. Furthermore, to eliminate bias of genomic coverage due to non-uniform restriction site distribution, we will supplement the BACs with a 50,000 clone Fosmid library with blunt-end cloned inserts.

Microarray transcription profiling of *Shewanella oneidensis* wilt-type and *ompR*, *envZ* mutant strains in response to osmostress

Haichun Gao*, Zamin Yang, Dorothea K. Thompson, and Jizhong Zhou. Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

Shewanella oneidensis is a facultatively aerobic Gram-negative bacterium and an important model organism for bioremediation studies. We carried out DNA microarray analysis of global gene expression in *S. oneidensis* wild-type upon an osmotic change. Gene expression profiles of *S. oneidensis* wild-type strain revealed that the stress had widespread effects affecting genes encoding proteins involved in membrane synthesis and functions. The result also showed that cells growing at high osmolarity already exhibited increased levels of σ^S (encoded by *rpoS*), a master regulator governing the expression of many stationary-phase-induced genes, during the exponential phase of growth. OmpR and EnvZ comprise a two-component system that regulates the porin genes in response to changes in osmolarity. To identify on the whole-genome scale the genes that are controlled by the system, strains carrying single in-frame deletion mutations in *ompR*, *envZ*, or both were constructed and investigated. Microarray analysis of these strains displayed a high consistency among expression profiles, suggesting a single copy of the operon and no independent role for either component under an osmolarity increase. Comparison in profiles of wild-type and mutant strains under various stress conditions strongly suggested that genes *SO1420* and *SO1557*, encoding putative outer membrane porin proteins in *S. oneidensis*, be controlled under the two component system and be the counterpart *ompC* and *ompF* of *Escherichia coli.*

Proteomic Analysis of the *Synechococcus* WH8102 CCM with Varying CO_2 Concentrations

Arlene Gonzales, Yooli K. Light, Zhaoduo Zhang, Tahera Iqbal, Todd W. Lane, Anthony Martino*; Biosystems Research Department, Sandia National Laboratories, Livermore, CA.

The genera *Synechococcus* and *Prochlorococcus* are oxygenic photoautotroph cyanobacteria. They are the most abundant picophytoplankton in the world's oceans where they form the foundation of the marine food web and are likely the largest contributors to primary production. Whole genome sequences are now available for a number of cyanobacteria including *Synechococcus* WH8102, *Prochlorococcus* MED4, and *Prochlorococcus* MIT9313. The sequences make it possible to use comparative analysis and high-throughput functional genomics and proteomics experiments to help better understand global diversity involved in carbon fixation.

Synechococcus WH8102's 2.4 Mb genome has yielded a number of interesting results regarding the carbon concentrating mechanism (CCM) in this organism. The carboxysome encoding operon in 8102 resembles that of β-proteobacteria rather than cyanobacteria. The operon most likely was acquired through horizontal gene transfer from phage. Carbonic anhydrase (CA) activity in the carboxysome shell protein csoS3 has been determined experimentally. Genome analysis indicates a putative β-CA and a ferripyochelin binding protein CA may also exist. Finally, transport of inorganic carbon in 8102 may occur through the low affinity CO_2 uptake genes ndhD4, ndhF4, and chpX. In *Prochlorococcus*, uptake genes have not been observed. Perhaps a unique transport mechanism exists in oceanic cyanobacteria.

We will present a high-throughput proteomic approach using mass spectrometry (MS), 2D gel electrophoresis, 2-hybrid analysis, and phage display to deconstruct components of the CCM and determine the effect of changing CO_2 levels in *Synechococcus* 8102. Protein identifications, expression levels, and protein-protein interactions of CCM components will be presented.

Combining Homology, Pathway, and Operon Based Data to Fill Holes in Pathway/Genome Databases

Michelle L. Green* and Peter D. Karp. Bioinformatics Research Group, SRI International.

A pathway hole occurs in a computationally predicted pathway/genome database (PGDB) when the genome used to construct the PGDB appears to lack the enzyme needed to catalyze a reaction in a pathway. We have developed a method that efficiently combines homology and pathway-based evidence to identify enzymes that fill pathway holes in PGDBs. The program uses BLAST to identify potential candidate sequences, and then combines data from multiple sources to assess the likelihood that a candidate has the required function. The method can be applied across an entire metabolic pathway network and is generally applicable to any pathway database. Filling pathway holes in a PGDB increases the completeness of the database, generating a more effective resource for experimental and computational researchers.

We have applied our pathway hole filler to the PGDBs for *Caulobacter crescentus*, *Mycobacterium tuberculosis* H37Rv, and *Vibrio cholerae*. The program identified over 200 enzymes to fill pathway holes in these three databases, thereby increasing the number of complete pathways by 42%. In the process of filling pathway holes, our program not only identifies functions for previously unannotated sequences, but also for sequences with existing non-specific annotations (e.g., thiolase family protein). The program also assigns putative additional functions that were not discovered during the original annotation process. Here we will present examples of these cases that illustrate how the pathway hole filler can increase the utility of a PGDB and generate testable hypotheses for elucidating protein function.

Probe Design Criteria Establishment and Software Development for Oligo Microarrays

Zhili He[1*], Xingyuan Li[2], Liyou Wu[1], Matthew W. Fields[1,3] and Jizhong Zhou[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, 37831
[2]Harbin Institute of Technology, Harbin, China 150001
[3]Miami University, Oxford, OH 45056

To experimentally establish the criteria for designing gene-specific and group-specific oligo probes, an oligo (PMO) array containing perfect match (PM) and mismatch (MM) probes of 50mer or 70mer of four genes was constructed. The effects of probe-target identity, contiguous stretch, mismatch position, and hybridization free energy on the specificity of oligo probes were examined using artificial targets. The results suggested that a 50mer gene-specific probe should have identity <= 85%, contiguous stretch <= 15 bases, and free energy >= -30 kcal/mol with other non-target sequences, and that a 50mer group-specific probe should have identity >= 96%, contiguous stretch >=35 bases, and free energy <= -60 kcal/mol within the group and the same criteria as a gene-specific probe outside the group. Similarly, for 70mer oligos, the criteria should be identity <= 85%, contiguous stretch <= 20 bases, and free energy >= -40 kcal/mol for a gene-specific probe, and identity >= 96%, contiguous stretch >=50 bases, and free energy <= -90 kcal/mol for a group-specific probe. Considering those three criteria together as well as other filters (self-binding, Tm and GC content), a software tool, *CommOligo* has been developed. The program uses a new global alignment algorithm to select a single or multiple gene-specific probes for each gene, and single or multiple group-specific probes for a group of highly homologous genes if a gene-specific probe is impossible. Users can modify all parameters if default values are not used. Designed oligos are output as a tab-delimited text file formatted for 96-well plates. For design of gene-specific probes, the program was evaluated using both whole-genome and highly homologous sequence data and compared with other probe design software tools. For group-specific probe selection, the program has been tested using small to medium sizes of data sets with documented phylogenetic relationships. Results showed that this program performed well and promises to select oligo probes for complex microarrays, such as functional gene array (FGA).

Streamlining the Genetic Characterization of Environmental DNA Samples

L.M. Hoffman* and B.W. Jarvis. EPICENTRE, Madison, WI

Microbial populations are commonly surveyed by isolating environmental DNA, amplifying ribosomal RNA genes (rDNA), cloning the amplicons, and sequencing individual rDNA clones. A large number of steps are involved, beginning with DNA extraction and proceeding to the bioinformatic analysis of sequence data. EPICENTRE offers several kits and methods that can facilitate many of these processes. The MasterPure DNA Isolation Kits are tailored for specific starting materials, and the purified nucleic acids are readily amplified using the FailSafe PCR system. To enable efficient cloning and downstream sequencing, the CopyControl PCR Cloning Kit can be used to construct rDNA libraries in a vector with a very low background of "empty" vector clones. DNA was extracted from several different environments with the appropriate kit and amplified using PCR with universal eubacterial 16S rDNA primers. A method ("reconditioning PCR") for preventing mosaic rDNA clones was used to decrease artifactual gene diversity estimates. The ribosomal PCR products were blunt-ended enzymatically and cloned into the fosmid pCC1. Bacterial colonies were used as templates for a second PCR without selection for clones containing rDNA insertions. The PCR products were purified by an improved PEG precipitation method and sequenced. Nearly all colonies picked contained rDNA inserts and generated rDNA PCR products, eliminating the need to screen for inserts. The ribosomal sequences were analyzed by BLAST and the results indicated that the majority of organisms found in a spring pond were uncharacterized and probably new species.

Comparing the fine structure of promoter regions in representative bacterial species.

Araceli M. Huerta[1], M. Pilar Francino[1] and Julio Collado-Vides[2].
[1]Genomics Division, Lawrence Berkeley National Lab and
Evolutionary Genomics Department, DOE Joint Genome Institute
2800 Mitchell Drive. Walnut Creek, CA 94598
[2]Laboratory of Computational Genomics. CIFN-UNAM P.O. Box 565-A.
Cuernavaca, Morelos México 62100

The *Escherichia. coli* RNAP is composed of a core complex of alpha, beta, and beta' subunits and one of a variety of sigma factors, the principal one being sigma70, which is capable of binding to the -10 (TATAAT) and -35 (TTGACA) promoter sequences and is essential for general transcription in exponentially growing cells. We have shown that regulatory regions in *E. coli* with sigma70 promoters are located within zones containing high densities of promoter-like signals, in contrast to coding regions and regions located between convergently-transcribed genes. Moreover, functional promoter sites identified experimentally are found in the subregions of highest density of signals, even when individual sites with higher binding affinity for RNAP exist elsewhere within the region. The essential primary sigma factors, which are present in all known eubacteria are closely related to sigma70 of *E. coli*. Recent data suggest that there is only one primary sigma factor present in any given eubacterial species. We have found that regulatory regions contain an excess of promoter-like signals in at least 70 different representative bacterial genomes. Only genomes with small size do not present this abundance of signals for housekeeping promoters, which is in agreement with evidence that small genomes have impoverished regulatory mechanisms. This confirms that even though the canonical picture of a bacterial DNA promoter is a 40 bp region defined by the transcription startpoint
and two conserved hexanucleotide regions upstream of the +1, bacterial promoters rather exist as a series of overlapping potentially competing RNAP interaction sites. We suggest that regulatory proteins play a key role in helping RNAP bind the correct site. We have also initiated a comparative study of regions containing high densities of promoter-like signals in different enteric species. We are analyzing conservation in these regions, in terms of sequence and information content, both globally and separately for subregions of different signal density, as well as for individual sites within subregions. Analysis of these patterns of conservation will shed light on the molecular and selective mechanisms that maintain this level of signal redundancy within regulatory regions.

The Construction of a Whole Genome Array for *Methanosarcina acetivorans* C2A.

Unmi Kim*, Kirsty A. Salmon and Robert P. Gunsalus. Dept. Microbiology, Immunology and Molecular Genetics. University of California, Los Angeles.

DNA microarray technology is a powerful tool in identifying changes in gene expression. This application has been employed in various organisms ranging from *Escherichia coli* to humans. Moreover, the development of DNA microarray technology has facilitated the identification of genes involved in regulatory networks. This report describes the construction and development of a whole genome array in a methanogenic archaeaon, *Methanosarcina acetivorans* C2A. The specific oligonucleotide primer pairs were designed for 4528 ORFs such that each PCR product (average size 300bp) did not show any homology to any other ORF in the genome. Of the 4528 ORFs, 3919 ORFs have been successfully PCR amplified. Primer pairs could not be designed for the remaining 609 ORFs due to homology issues. PCR reactions were carried out in 96-well format with an average annealing temperature of 55°C. All PCR products were verified by gel electrophoresis, and then re-amplified from the PCR product to dilute out the genomic DNA. All PCR products were purified using a commercially available kit, quantified by UV spectrophotometry, normalized and spotted onto aminopropylsilane microarray slides in duplicate. The microarray will be validated by two control experiments. The microarray will be hybridized using total RNA isolated from different cultures of the same growth conditions (biological control) or from the same cultures (self-self control). Upon validation, the microarray will be used to study the gene expression in various growth conditions and aid in identification of regulatory elements.

Isolation and Identification of a Novel Extremophile and its Application in Bioremediation

Anil Kumar* and Rita Kumar. Institute of Genomics and Integrative Biology, Delhi.

The isolation and characterization of microorganisms that are able to thrive in extreme environments has received a great deal of attention because of their potential biotechnological applications. Present work deals with the isolation strategy for extremophiles, identification and their application in bioremediation. For isolation, soil sample of high alkalinity was taken from the beverage industrial premises. Different concentrations of soil extract mixed with alkaline bacillus medium were used as a media to entrap the desired micro flora after enriching the soil in soil extract for two days at 120rpm/32^{0}C. On the basis of high pH tolerance (12.0), a rod shaped and yellow colored bacterium was screened. Partial 16S rDNA sequence shows a similarity of 99.4% to *Exiguobacterium aurantiacum* strain ' Z8' and physiological results are also different which show its novelty. In a trial to neutralize the alkaline industrial wastewater, this bacterium has been found capable to neutralize textile industrial wastewater from pH 12.2 to pH 7.5 within two hours. Alkaline bacillus medium (ABM) was selected as the suitable medium and maximum growth of one-liter culture could be achieved in eight hours. For the neutralization of textile wastewater, eight hours grown culture was centrifuged and the pellet was added to textile wastewater of pH 12.2. Lowering of pH from 12.2 to 7.5 using this bacterium could be achieved in a period of two hours.

High Pressure Motility of *Photobacterium profundum.*

F.M. Lauro[1*], D.S. Millikan[1], E.A. Eloe[1], A. Vezzi[2], S. Campanaro[2], N. Vitulo[2], F. Simonato[2], G. Valle[2], D.H. Bartlett[1]

[1]*Marine Biology Research Division, Scripps Institution of Oceanography, La Jolla CA, USA*
[2]*CRIBI Biotechnology Center, Universita' Degli Studi Di Padova, Padova, Italy*
E-mail: flauro@ucsd.edu

One of the challenges that deep-sea bacteria face is maintaining motility under high pressure. Motility is an important adaptation in the ocean environment, in particular to avoid grazing or toxic compounds and to acquire nutrients and colonize biotic and abiotic substrates. High hydrostatic pressure selects for negative volume changes, while multimerization of proteins, including that of flagellin, is generally accompanied by a positive volume change. Here we describe the use of comparative genomics, using microarray technology, to compare and contrast motility in three different strains of the marine bacterium *Photobacterium profundum* that differ with respect to their optimal pressure: (1) the sequenced deep-sea strain SS9, (2) the deep-sea strain DSJ4 and (3) the shallow water strain 3TCK. Preliminary results indicate that *P. profundum* strains SS9 and DSJ4 possess vigorous motility at high pressure, but swim very poorly at atmospheric pressure. The opposite phenotype has been observed in the pressure-sensitive strain 3TCK. Genes for motility and flagellar assembly in *P. profundum* SS9 are arranged in two large clusters, and many of the flagellar genes are duplicated. Genome comparison indicates that 3TCK (but not DSJ4) lacks all of the first flagellar cluster, while, in the shared cluster, both *flaA* (flagellin) and *flgM* (a flagellar hook-associated protein) are over-expressed by SS9 at 0.1 MPa. Among the three possible SS9 flagellin genes, the most divergent sequence is shared only by the piezophilic strains. Flagellum examination by electron microscopy showed that the single polar flagellum of SS9 is long and helical at 28 MPa, but short and deformed at 0.1 MPa.
We hypothesize that *P. profundum* SS9 has evolved a cluster of flagellar genes to function under high pressure conditions. This hypothesis is currently being tested using reverse genetics.

Sequencing, assembly and annotation of the genome of the marine unicellular cyanobacterium *Synechococcus* sp. PCC 7002.

T. Li[1,3], J. Marquardt[1], G. Shen[1], C. Nomura[1], S. Persson[1], C. Detter[4], C. Lanz[2], S.C. Schuster[2], J. Zhao[3], & D.A. Bryant[1*]. [1]*Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802 USA;* [2]*AG Genomics and Signal Transduction, Max-Planck-Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany;* [3]*College of Life Science, Peking University, Beijing, China;* [4]*DOE Joint Genome Institute, 2800 Mitchell Drive, B400, Walnut Creek, CA 94598 USA.*

The unicellular, transformable, marine cyanobacterium *Synechococcus* sp. PCC 7002 has long served as a model organism for the genetic and biochemical characterization of genes involved in photosynthesis, respiration and biosynthetic pathways. The sequencing of the genome of this organism is now nearly complete. Primary sequence data were collected from a combination of 23,279 reads from whole-genome shotgun sequencing and about 1000 reads from targeted sequencing of cosmid and BAC libraries. After assembly, the connecting of ~825 contigs was pursued through primer walking with templates from cosmid libraries and the sequencing of PCR products. Recently, paired-end sequencing of 1,536 fosmids with average inserts of 37.5 kb facilitated the scaffolding of the chromosome and the two largest plasmids. The chromosome has been assembled into a single circular molecule of 3.008 Mb. Our data confirm the presence of five of six plasmids identified by Roberts and Koths (1976): pAQ1 (4,809 bp), pAQ3 (16,073 bp), pAQ4 (32,035 bp), pAQ5 (38,516 bp), and pAQ6 (~117 kb). These sizes closely match the sizes measured by electron microscopy and gel electrophoresis: 4.6 kb, 15.9 kb, 31.0 kb. 38.6 kb, and 115.6 kb. We have shown that plasmid pAQ2 (~10 kb) is actually a dimer of pAQ1 and that trimers of pAQ1 also occur. Interestingly, our most recent assembly data suggest the existence of a sixth plasmid, which we have named pAQ7 (~128 kb). A preliminary annotation of the genome has been performed using TIGR's Annotation Engine for Prokaryotic Annotation and Analysis. The *Synechococcus* sp. PCC 7002 genome model encodes 3,498 ORFs; the coding percentage is 88% and the genome has a mol% G+C content of 49.6. Results of the preliminary analysis and some aspects of comparative genomics will be presented.

Roberts TM and Koths KE 1976. The blue-green alga *Agmenellum quadruplicatum* contains covalently closed DNA circles. Cell **9**:551-557.

Classifying bacterial species using base composition analysis.

C. Massire*, T.A. Hall, R. Sampath & D.J. Ecker. Ibis Therapeutics, Carlsbad.

Identification of bacterial species is typically carried out through PCR amplification and sequencing of the 16S rRNA gene, or through time-consuming culture and physical and biochemical characterization. The TIGER technology currently developed at Ibis Therapeutics circumvents the need for the time-consuming and labor-intensive processes by PCR amplification of small information-rich loci followed by high-performance mass spectrometry and determination of their base composition.

In this study we show how the triangulation of amplicon base composition alone, coming from 6 carefully selected primer pairs targeting different regions of the 16S and 23S rRNA, allows for the identification of unique "signatures" that can be associated with any particular phylogenetic level, in spite of the limited amount of information that is summarized in each individual base composition.

Using the publicly available 230 complete bacterial genomes (representing a significant coverage of bacterial diversity), alignments of the 6 amplicons were generated and the expected base counts were determined for each amplicon. In addition to the four single base counts, A, G, C and T, the six binary combinations, A+G, C+T, A+C, G+T, A+T and G+C were also determined, yielding a total of ten metrics per amplicon, hence sixty per organism. For each of the taxonomic levels included in the NCBI bacterial phylogeny, the minimum and maximum values of these sixty metrics are then determined. We then show how a simple scoring system, that primarily summarizes the number of metrics that fall within the range of the permissible metrics for each recognized taxon, allows the phylogenetic placement of an unknown organism at a particular phylogenetic level.

The reliability of the method is then discussed with the use of different combinations of primers pairs, as well as with alternate training and sampling sets, including the wider set of partial ribosomal sequences available in the GeneBank database.

Comparison of Genome degradation in Paratyphi A and Typhi, serovars of *Salmonella enterica* causing typhoid.

Michael McClelland[1], **Carlos Santiviago[1]**, Kenneth E. Sanderson[2], Sandra W. Clifton[3], Phil Latreille[3], Steffen Porwollik[1], Aniko Sabo[3], Rekha Meyer[3], Tamberlyn Bieri[3], Phil Ozersky[3], Mike McLellan[3], Richard Harkins[3], Chunyan Wang[3], Christine Nguyen[3], Amy Burghoff[3], Glendoria Elliott[3], Sara Kohlberg[3], Cindy Strong[3],Johar Ali[3], Mike Dante[3], Feiyu Du[3], Dan Layman[3], Shawn Leonard[3], Hui Sun[3], Liliana Florea[4], John Spieth[3], Michael Nhan[3], & Richard K. Wilson[3]

1. Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, CA 92121, USA
2. Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada
3. Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA
4. Informatics Research, Celera/Applied Biosystems, 45 W. Gude Drive, Rockville, MD 20850, USA.

Salmonella enterica serovars often have a broad-host range and some cause both gastrointestinal and systemic disease. In contrast, the serovars Paratyphi A and Typhi are restricted to humans and cause only systemic disease. It has been estimated that Typhi arose in the last few thousand years. The sequence of the Paratyphi A genome and comparative genomics among Paratyphi A strains using microarray analysis reveals close similarity to the Typhi genome but suggests an even more recent evolutionary origin. Both genomes have independently accumulated many pseudogenes; 173 in Paratyphi A and ~210 in Typhi. Only 30 genes are mutated in both serovars, all but one mutated at a different site, indicating that despite their similarity and similar niche, these genomes are evolving down distinctly different paths. The 30 shared pseudogenes include three known to be important in gastroenteritis, which does not occur in these serovars, and four for Salmonella-translocated effectors, which are normally secreted into host cells to subvert host functions. Loss of function also occurs by mutation in different genes in the same pathway, for example in chemotaxis, in iron transport and response, and in the production of fimbriae. Some of this loss may be neutral drift, revealing genes that are used for functions that are no longer needed in these more specialized pathogens, such as gastroenteritis, or infection of other hosts. Other losses may increase fitness of Paratyphi A or Typhi by permitting a new or altered response during infection. For example, chemotaxis mutants are already known to be selected during cholera infections and are known to increase the virulence of Typhimurium for enteric disease. In summary, nature has run an experiment twice, degrading the genomes of two very similar pathogens as they specialize in an almost identical niche, providing indications about which previously identified pathogenesis-associated genes are dispensable for enteric disease, at the same time as indicating which pathogenesis-associated genes have not been mutated and could be necessary for enteric disease.

Bayesian Resolution of Bacterial Phylogeny

Jenna Morgan*, Arthur Kobayashi, Joel Martin, Pilar Francino

A large number of complete genome sequences have recently become available that represent many evolutionary lineages of bacteria. We are using large-scale amino acid sequence comparisons to reconstruct the ancient evolutionary relationships among these different lineages. Here we present a phylogeny based on a concatenated alignment of 194 orthologous clusters that are pervasive and in single copy throughout 250 bacteria, with 17 archaeal sequences as outgroups. We generated the largest possible complete matrix of proteins cross species, and then added key species to create an optimal dataset from the point of view of taxonomic breadth, tolerating some amount of missing data. We are using this dataset to generate phylogenetic trees using various types of tree reconstruction methods, including maximum parsimony, maximum likelihood, neighbor-joining and Bayesian phylogenetics.

Mutagenesis with UV, cis-platin, EMS, and MNNG on *Pyrobaculum aerophilum* and its genome-wide gene expression studies.

Theresa T. Nguyen*, Huong T. Tran, Anh Nguyen, Erika Wolff, & Jeffrey H. Miller**.
University of California, Los Angeles.

The hyperthermophilic archaeon, *Pyrobaculum aerophilum*, is isolated from boiling marine water holes at Maronti Beach, Ischia, Italy. This novel rod-shaped species grows optimally up to 100°C and pH 7.0 in laboratory condition via aerobic respiration and dissimilatory nitrate reduction. *Pyrobaculum aerophilum* was exposed to various mutagens. This is the first study that looks at the mutagenicity of *Pyrobaculum aerophilum* and its resistance to mutagens. The goal of the study is to examine the response to different physical and chemical agents that damages DNA and proteins. The lethal and induced mutagenic effects of various mutagens such as ultraviolet light (UV) at 254 nm, cis-platinum (II) diamine dichloride (cis-platin), ethyl methanesulfonate (EMS), N-methyl-N'-nitro-N-nitrosoguandinine (MNNG), and methyl methanesulfonate (MMS) were determined with different dosages. The mutation frequency of *Pyrobaculum aerophilum* was obtained by plating on paromycin solid media and calculated based on the paromycin resistant (*Par*^{r}) colonies. Optimal dosages were selected to expose to *Pyrobaculum aerophilum* and its RNAs were extracted for microarray experiments in order to examine genome-wide gene expression at different recovery time point after the exposure.

In silico identification of the first _Bordetella_ effector toxin.

Ekaterina M. Panina*, Seema Mattoo, and Jeff F. Miller. University of California, Los Angeles.

Although multiple _in vitro_ and _in vivo_ phenotypes have been associated with the _Bordetella bsc_ type III secretion (TTS) system, the effector proteins responsible had not been previously identified. We performed a computational search for candidate effectors in the _Bordetella_ genome. The search utilized frequent co-localization of effector genes with the genes for their cognate chaperones. Due to the lack of sequence similarity between TTS chaperones, the latter were identified based on their biophysical characteristics, such as molecular weight, isoelectric point and potential 3D fold. The search identified two divergently-transcribed genes, _btcA_ and _bteA_, which are unlinked to the _bsc_ locus and highly conserved in _B. pertussis_, _B. parapertussis_ and _B. bronchiseptica_. BtcA meets all known characteristics of a TTS chaperone, while BteA has no homologs outside of the _Bordetella_ lineage. By RT-PCR and _lacZ_ transcriptional fusion analysis, we found that transcription of _bteA_ is co-regulated with transcription of the _bsc_ gene cluster. Western blot analysis of supernatant fractions showed that BteA is secreted by _B. bronchiseptica_ strain RB50 in a TTS-dependent manner. A pore forming hemolysis assay confirmed that BteA is not part of the TTS injectisome. The _bsc_ TTS system is required for the induction of non-apoptotic cytotoxicity in tissue culture cells _in vitro_ and for persistent tracheal colonization _in vivo_. Consistent with BteA function as a TTS effector, the _bteA_ mutant was unable to induce cytotoxicity in HeLa cells as demonstrated by LDH release assays and was cleared from the trachea of rat in our animal model of infection. These data suggest that BteA is the first identified type III secreted effector protein in _Bordetella_ subspecies.

ERIC – Enteropathogen Resource Integration Center, an NIAID Bioinformatics Resource Center for Biodefense and Emerging/Re-emerging Infectious Disease.

Guy Plunkett III*[1], Valerie Burland[1], Eric Cabot[1], Jeremy Glasner[1], Paul Liss[1], Michael Rusch[1], Thomas Hampton[2], Robin Martell[2], Matthew Shaker[2], Lori Shaull[2], John Greene[2], Frederick Blattner[1], & Nicole Perna[1]. [1]University of Wisconsin, Madison and [2]SRA International, Inc.

A partnership between the University of Wisconsin–Madison and SRA International, ERIC is intended to serve as a central repository for genomic and related data on five enteropathogens: diarrheagenic *E.coli*, *Shigella* spp., *Salmonella* spp., *Yersinia enterocolitica*, and *Yersinia pestis*. ERIC will be a web-based, pathogen-centric bioinformatics portal intended to integrate information on the genomic sequence, genome annotations, and related biological data for these five organisms. Among the major goals of the project is to provide a method for community annotation of genes of these organisms, carefully curated to provide a high quality standard, using the ASAP (A Systematic Annotation Package for community annotation) component developed at the University of Wisconsin-Madison. In addition, the system will provide tools for comparative genomics, sequence analysis and comparison, microarray analysis using the NCI microarray database system (mAdb), proteomics data and analysis, functional analysis tools such as GoMiner and MatchMiner, and text mining of the literature related to these pathogens. The annotation component for these five genomes is currently online at http://www.ericbrc.org. Our primary objective for ERIC is to provide integrated access to both data and analysis tools, using a scalable, flexible, robust system architecture. All tools and data developed under this NIAID contract will be freely available both over the web portal and for download. In addition, we will interact heavily with the enterobacterial research community, soliciting feedback on the system development and features, providing training, disseminating knowledge of the system, and facilitating development of ontologies and controlled vocabularies.

Detection of phage activity and resulting escape replication in *Salmonella* by microarray

S. Porwollik*, J. Frye, P. Cheng & M. McClelland. Sidney Kimmel Cancer Center, San Diego, CA

Salmonella enterica genomes usually contain one or more integrated prophage. However, using conventional methods it is difficult to ascertain which of these phage genomes can become active and under what conditions. Phage genome replication of the four temperate functional prophages from Typhimurium LT2 was studied by microarray analysis of total DNA after induction using peroxide or mitomycin C. All four phage were observed to be induced. A large amplification of host DNA, "escape replication", spanning several hundred genes, was detected adjacent to the integration site of the Fels-1 lambdoid phage. Regions of host genome amplification were also observed in Typhimurium strains SL1344 and 14028s indicating integration sites of phage, the genes of which were not included in the microarray. Escape replication and possibly resulting escape synthesis (changes in expression of RNA from amplified host genes) may have an impact on transduction, recombination, and host response to stresses. Microarray analysis of phage particles harvested from the supernatants of induced cultures indicated that even after escape replication, phage are mostly packaged correctly. Using mitomycin C or peroxide as an inducer, activity of at least one prophage was also detected in *S. enterica* serovar Paratyphi A SARB42 and serovar Enteritidis PT4, but not for any of the prophage genomes present in Typhi CT18. In summary, microarray analysis revealed the extent of escape replication for several phages in Typhimurium, and allowed activation of phage to be monitored in different *Salmonella.*

Comparative analysis of expressed sequence tags in the soybean rust pathogen *Phakopsora pachyrhizi*.

M.L. Posada-Buitrago*, J.L. Boore & R.D. Frederick (1). DOE-JGI-LBNL, Walnut Creek, CA, USA. (1) USDA-ARS-FDWSRU, Ft. Detrick, MD.

The obligate fungal pathogen *Phakopsora pachyrhizi* causes soybean rust, the most devastating foliar disease in many soybean production areas around the world. The disease has not been reported in the continental United States of America but recent findings in South America demonstrate that *P. pachyrhizi* is spreading to new geographic regions, becoming a major threat to U.S. soybean production. The infection process was examined by comparing approximately 16,000 expressed sequences tags (ESTs) generated from three stage-specific unidirectional cDNA libraries from *P. pachyrhizi*. The EST cDNA libraries were constructed using mRNA isolated from urediniospores germinating on a water surface for 16 h and from infected soybean leaf tissue at 6-8 and 13-15 days post inoculation. Intercellular hyphal growth and the peak of sporulation occur at 6-8 and 13-15 days, respectively. The ESTs were analyzed for redundancy within and among the three libraries, and they were classified according to putative function based on similarity to known proteins using the Blast sequence algorithms. These ESTs will provide insight into the biochemical and physiological processes that are necessary for infection, survival and reproduction.

A reduced *Escherichia coli* lacking transposable elements

György Pósfai[1,5], Guy Plunkett III[2,5], Tamás Fehér[1], Guenther Keil[3], Kinga Umenhoffer[2], Vitaliy Kolisnychenko[2], Shamik Sharma[4], Sarah Harcum[4], Frederick R. Blattner [2,5]

Institute of Biochemistry and Institute of Biophysics, Biological Research Center, H-6701 Szeged, Hungary and [2]Department of Genetics, University of Wisconsin, Madison, WI 53706, USA [3]BFAV Institute of Molecular Biology, Griesswald Insel Riems; Germany, [4]Clemson University, Clemson SC USA, [5]Scarab Genomics Madison WI

Scarless genomic surgery was used to delete 43 genomic regions from the sequenced *Escherichia coli* strain, MG1655. Regions deleted include large K-islands, prophages, phage remnants, restriction modification genes, flagellar and chemotaxis related genes and all transposable elements including IS sequences and RHS (recombination hot spot) elements. The resulting multiple deletion strain MDS43 has a 15.27% reduction in its genome size from 4,639,221 to 3,930,956 bp. It is also missing endonuclease A, the lac operon and is T1 phage resistant. Strain MDS 43 along with intermediates MDS41 and 42, were characterized for growth on standard microbiological media, and for auxotrophy using the Biolog panel of 384 substrates. Transformation efficiency, global gene expression, and mutation rates were compared with the parent MG1655. As expected for strains lacking transposable elements, the measured frequency of mutation by IS hopping dropped to zero. Point mutation rates were virtually unchanged. Unexpectedly we found that several practical properties of the strains were improved. In particular, some gene sequences that were unclonable in wild type E. coli were able to be cloned in MDS strains. The reason for this unclonability was found to be hyperactive IS hopping into the target plasmid which was, of course, not possible in our construct. Reintroduction of the IS free plasmid into DH10B or C600 resulted in immediate introduction of IS elements of a variety of types into a variety of places. We conclude that the stress of producing a toxic protein induces transposition. Consistent with this hypothesis, small increases in the IS transposition rate were measured in wild type E. coli cells subjected to heat shock, cold shock or upon induction of an expression plasmid for Cloramphenicol Acetyl Transferase. The reduced genome E. coli strains can be efficiently transformed and excellent yields of recombinant proteins have been obtained. This demonstration of strain improvement through genome reduction, provides further incentive for constructing additional reduced genome strains which will provide a "clean background" for functional genomics studies, a more efficient platform for biotechnology applications as well as a unique tool for studies of genome stability and evolution.

GGETools: Genome Gene Expression Analysis Tools for comparative analysis of transcriptome experiments in bacterial genomes.

Sánchez-Solano Fabiola[1], Diaz-Peredo Edgar[1], Huerta M. Araceli[2], Moreno-Hagelsieb Gabriel[3], and Collado-Vides Julio[1].

[1]Program of Computational Genomics. Center of Nitrogen Fixation. UNAM. Cuernavaca, A.P. 565-A, Morelos, 62100, México.
[2]Evolutionary Genomics Department. DOE Joint Genome Institute. 2800 Mitchell Drive. Walnut Creek, CA 94598

[3]Department of Biology. Wilfrid Laurier University. Waterloo, ON N2L 3C5

The advances in DNA microarray technology and the sequencing of multiple bacterial genomes have provided a unique opportunity for the integration of comparative genomics with high-throughput gene expression analysis. Now that genomics has matured into a comparative research program, the rich legacy of biological knowledge available for *Escherichia coli* can be exploited to gain knowledge about non-model bacterial species, through comparisons of their microarray data against the knowledge available in the literature and the computational predictions available for *E. coli*. Here we present Genome Gene Expression Tools (**GGETools**), a computational environment to analyze the transcriptional regulation of expression profiles from different enteric bacterial genomes, by evaluating their congruence with predicted and characterized transcription units, upstream regulatory signals and putative transcriptional factors. We used this tool to do a comparative analysis between *Shigella* and *Escherichia coli* from the data set published on *J. Bateriol* 2004 by Fukiya S. *et al.*

Development of a Functional Gene Microarray for Analyzing Microbial Genomes and Communities.

C.W. Schadt, T.J. Gentry*, J. Liebich, S.C. Chong, Z. He, H. Pan & J. Zhou. Oak Ridge National Laboratory, Oak Ridge, TN.

A functional gene microarray (FGA) containing 23,864 probes for various microbial genes, encoding functional enzymes involved in several biogeochemical cycling and bioremediation processes, has been developed for the investigation of microbial genomes and environmental communities. Initially, 19,718 genes were acquired from public databases and ORNL clone libraries. A new software program, CommOligo, was developed and used for probe design. The program uses a new global alignment algorithm to design single or multiple unique or group probes for each gene. Unique probes were designed with a maximum similarity of 90% and a maximum of 20 contiguous, matched bases with non-target sequences. Group probes were also designed for highly similar sequences with a minimum similarity of 96% and a minimum contiguous stretch of 35 bases within a group and the same parameters as unique probes outside a group. Probes were successfully designed for 73% of analyzed sequences. Selected probes were screened against GenBank for potential cross-hybridization with known sequences. Up to 3 probes with no obvious potential for cross-hybridization were selected for each gene and added to the array. The FGA contains probes for several processes including: 1) carbon degradation – 2,532 unique & 276 group; 2) carbon fixation and formyltetrahydrofolate synthesis – 584 unique & 215 group; 3) nitrogen cycling – 6,068 unique & 4,332 group; 4) sulfur cycling – 1,286 unique & 329 group; 5) organic contaminant degradation – 6,920 unique & 1,087 group; 6) metal resistance – 4,039 unique & 507 group; and 7) perchlorate remediation – 21 unique. The FGA will be a powerful tool for the simultaneous detection and quantification of thousands of genes in microbial genomes and environmental samples.

Global Identification of Protein Modularity in the *M. tuberculosis* Genome.

M. Strong*, T. Graeber, M. Pellegrini, M. Thompson & D. Eisenberg. UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles.

10 years ago, the introduction of high-throughput genome sequencing revolutionized the biological sciences. While the identification of the complete set of genes and corresponding proteins was indeed a remarkable achievement, one question remains: "How do the encoded proteins function together within their cellular environment?" Over the past decade a number of methods have been developed to answer such a question.

We have applied some of these methods to identify functionally linked genes and proteins throughout the *Mycobacterium tuberculosis* genome. Using a combination of the Rosetta Stone, Phylogenetic Profile, conserved Gene Neighbor, and Operon computational methods, we have been able to identify networks of interconnected genes in this deadly pathogen. We have also developed novel methods for the visualization of genome-wide functional linkages. Specifically, we have developed a method to visualize functional linkages on a two-dimensional matrix, which we have termed a genome-wide functional linkage map. The use of these maps has enhanced our understanding of protein connectivity in relation to genome organization, and subsequent hierarchical clustering of these maps has led to the identification of functional modules within the genome of *M. tuberculosis.* We have also developed a novel expression system, which we have employed to identify physically interacting proteins. This method mimics bacterial operon organization, and has been used to demonstrate physical interactions among *M. tuberculosis* proteins. We have also extended our methods to identify global protein domain modularity. Such a holistic approach to microbial genomics has facilitated our investigations of species-specific protein networks, and has provided new means to investigate global protein domain organization and modularity.

Method Developments for Characterizing the Complex Metaproteomes of Microbial Communities

N.C. VerBerkmoes[*1]; M. Thompson[1]; C. Shook[2]; A.P. Borole[3]; M. Shah[3]; F.W. Larimer[3]. R.L. Hettich[4]; B.H. Davison[2]
1. ORNL-UTK Genome Sciences and Technology Graduate School
2. University of Tennessee
3. Life Sciences Division, Oak Ridge National Laboratory
4. Chemical Sciences Division, Oak Ridge National Laboratory

Recently, there have been strong efforts to develop techniques for genomic sequencing and annotation of microbial communities (*metagenomics).* With the potential of partial or near complete microbial genomes obtained from environmental samples along with the rapid proliferation of isolate microbial genomes, system biology in microbial communities combining genomic, transcriptomic, proteomic and metabolic studies may be possible in the near future. Our current studies seek to develop and demonstrate *metaproteome* analysis techniques for mixed cultures, establishing the basis for "whole community proteomics". Currently our major focus is to use controlled simple microbial mixtures of four species (*E. coli, S. cerevisiae, R. palustris, and S. oneidensis*) to develop MS-based proteomics methods as well as the proteome bioinformatics tools for detailed analysis of sequenced microbial communities. Our goal is to provide "deep" and "wide" proteome measurements of complex microbial mixtures. The analytical goal of the MS-based proteomics studies is to determine current applicability and limitations of common 2D-LC-MS/MS techniques for analyzing the microbial mixture. To date, this effort has focused on varying the concentration of *R. palustris* with respect to the other microbes in the mixture. With current 2D-LC-MS/MS analysis techniques we have demonstrated deep proteome analysis of *R. palustris* at 25% and 5%, with 1% and 0.1% being the borderline for definitive analysis. We are currently testing different 3D-LC-MS/MS methodologies to develop deep analysis at the 1% and 0.1% concentrations. The proteome bioinformatics effort has focused on the effect of database size on the identification of unique peptides, which are most useful for confident identifications. We are currently testing protein databases made of 4 test species, as well as a medium database with 9 additional species and a large database with over 200 species. The results from these studies as well as efforts to use these techniques for real microbial communities will be discussed.

Research sponsored by the ORNL Laboratory Director's Research and Development Program, U.S. Department of Energy, under contract No. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.

Life at depth: genome sequencing and expression analysisof*Photobacterium profundum*strain SS9

A. Vezzi*^, M. D'Angelo^, S. Campanaro^, F. Simonato^, N. Vitulo^, F.M. Lauro§, A. Cestaro^, G. Malacrida^, N. Cannata^,C. Romualdi^, D.H. Bartlett§ & G. Valle^. ^CRIBI Biotechnology Centre, University of Padua, Italy. §Scripps Institution of Oceanography,University of California, San Diego, La Jolla CA,92093-0202.*e-mail to: sandrin@cribi.unipd.it

Around 70% of the world surface is covered by oceans that are mainly characterized by low temperature, high pressure, high proportions of recalcitrant organics and high invertebrate diversity. Nevertheless, the adaptations for living at these extreme conditions are poorly investigated.In order to investigate how micro-organisms could grow in the deep sea, we have accomplished the genome sequencing and microarray analysisof *Photobacterium profundum* strain SS9.

Photobacterium profundum belongs to the *Vibrionaceae* family and is a deep sea psychrotolerant and moderately piezophilic bacterium. It survives up to90MPa with an optimum at 28MPa, but it can also grow at atmospheric pressure. Moreover it is amenable to genetics, therefore, compared to other psychro-piezophiles, itrepresentsan ideal model system to study adaptations to the deep sea.

SS9 has a tripartite genome structure: two main circular chromosomes, 4.1Mbp and 2.2Mbp, and a plasmid, 80kbp. The genome encodes for 5446 predicted proteins.

An unusual feature for gamma-proteobacteria, derived from the analysis of the genome, is the presence of genes for amino acid fermentation via the Stickland reaction. So far this pathway has only been foundin the *Clostridiales*and *Spirochaetales*. This is the first report of Stickland reaction genes outside of these anaerobes.

We have analyzed gene expression profiles using RNA extracted from cultures grown at 28MPa, 0.1Mpa, 15°C and 4°C. Differentially expressed genes identified by microarray experiments were classified according to the gene ontology biological process.Aminoacid transport, ion transport, protein folding and glycolysis are mainly overexpressed at 0,1MPa, while cell wall biosynthesis is overexpressed at 28MPa. Moreover we found an overexpression at 28MPa of genes involved in the degradation ofcomplex carbohydrates, indicating that such processes can be regulated by pressure.

An interactive SS9 genome browser, containing both genomic and microarray data, has been implemented underthe UCSC environment.

DNA Uptake Signal Sequences in *Actinobacillus actinomycetemcomitans*.

Y. Wang*, D. Dyer, & C. Chen. University of Southern California School of Dentistry, Los Angeles.

The naturally transformable bacterium *Actinobacillus actinomycetemcomitans*, a member of the family *Pasteurellaceae* and a major causative agent of periodontitis, preferentially takes up DNA that contains a specific uptake signal sequence (USS), which is similar to that of *Haemophilus influenzae*. We showed that a 6-kb donor DNA containing one or two USSs located near the middle or at the end, transformed *A. actinomycetemcomitans* about 200-fold better than that using a similar DNA without USS. By transformation assays with in vitro constructed DNA and by genome sequence comparison, the complete USS appeared to be a 30-bp DNA (three turns of double helix), 5'-aaAAGTGCGGT nrwwwwtnn nnnyrwwttt (r=purine, w=A or T, y=T or C), containing a highly conserved 9-bp core sequence (capitalized) and two less conserved A/T rich regions. Missing of 6 or 7 bases from either end abolished the function of USS, demonstrating that the complete USS contains about 30-bp. Several single-base mutations in the highly conserved 9-bp core sequence decreased the transformation efficiency. Replacement of the 9-bp core region completely abolished the USS function, while replacement of the less conserved second and third regions reduced the DNA uptake efficiency to about half compared to that by the DNA with an intact USS. The genome of *A. actinomycetemcomitans* strain HK1651 (2105-kb) contains 1759 copies of perfect 9-bp core sequence, and 725 copies of imperfect USS with one base variation in the core region.

Phylogenetic Typing of Cultivated and Uncultivated Microflora of UCLA Soil Communities: Developing a Research Program for Widespread Undergraduate Participation

Krystle Ziebell, Maxine Karimoto, Grace Lee, Erika Wolff, Cindy Tamae, Jennifer Okada, Lavitania Bismart, and Jeffrey H. Miller
Department of Microbiology, Immunology, and Molecular Genetics and the Molecular Biology Institute, University of California, Los Angeles, CA 90095

Undergraduate research programs offer students a wide range of opportunities in one of the many areas of science. However, typical undergraduate programs require a long apprenticeship on a highly specialized project. Thus, not only is the discovery process delayed, but students are also offered only a narrow view of the entire Life Sciences field. Therefore, we are developing a program for use in universities that will enable undergraduates to participate in cutting edge research and to experience the thrill of discovery within a short time of starting. The program is multidisciplinary, exposing students to microbiology, molecular biology, bacteriology, bioinformatics, genomics, and eventually bacterial physiology, biochemistry, and organic chemistry.

We have streamlined protocols that allow starting undergraduates to isolate the DNA of microorganisms cultivated from the soil, PCR amplify 16S rRNA gene segments, and build phylogenetic trees based on the sequence. They can also isolate DNA directly from the soil, as it is estimated that greater than 99% of the microorganisms in soil have not been cultivated (Pace, 1997). In a subsequent stage, students identify antibiotic-producing microorganisms and characterize them and their products more fully. We have termed this program "**I, Microbiologist.**" First year students can go from the collection of soil samples to the generation of phylogenetic trees and the identification of antibiotic-producing microorganisms in as little as two to three weeks.

Does the three dimensional organization of the nucleoid of the *Deinococcaceae* contribute to their ionizing radiation resistance?

J. M. Zimmerman* & J. R. Battista. Louisiana State University, Baton Rouge

Transmission electron micrographs of *Deinococcus radiodurans* R1 suggest that the nucleoid of this species exists as a toroidal ring, and have led to speculation that this structure facilitates the extreme radioresistance of this species. However, little direct evidence supports this contention. Since extreme radioresistance is characteristic of all the members of the *Deinococcaceae*, we hypothesize that if nucleoid morphology contributes to radioresistance, the genomic DNA of each species should form similar structures. Using epifluorescence and deconvolution microscopy, we evaluated the nucleoid morphologies of eight of the nine validly described species of *Deinococcus*, the radioresistant bacterium *Rubrobacter radiotolerans*, and the less radioresistant *Thermus aquaticus*, a distant relative of the deinococci. Although the nucleoids of *Deinococcus murrayi, Deinococcus proteolyticus, Deinococcus radiophilus,* and *Deinococcus grandis* have structures similar to *D. radiodurans*, the nucleoids of *Deinococcus radiopugnans* and *Deinococcus geothermalis* lack specific organization. The nucleoid of *R. radiotolerans* consists of multiple highly condensed spheres of DNA. Since only five of the seven recognized deinococcal species exhibit a structurally distinct nucleoid, we conclude there is no obvious relationship between the three dimensional organization of genomic DNA and extreme radioresistance. However, the genomic DNA of all extremely radioresistance species is highly condensed relative to the more radiosensitive species examined. We have examined nucleoid structure following the introduction of DNA double strand breaks and show that the shape of the nucleoid does not demonstrably change in radioresistant species even in strains incapable of repairing strand breaks, suggesting that DNA held in this tightly packed configuration contributes to the radioresistance of these bacteria.

DIRECTORY

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

Melissa Azarraga
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
mazarraga@ucla.edu

Thomas J. Albert
NimbleGen System, Inc.
1 Science Court
Madison, WI 53711
Tel: 608/218-7613
Fax: 608/218-7601
talbert@nimblegen.com

Eric Altermann
Dept. of Food Science
North Carolina State University
301 Schaub Hall, Box 7624
Raleigh, NC 27695-7624
Tel: 919/515-3552
Fax: 919/515-7124
ealterm@unity.ncsu.edu

M. Andrea Azcarate-Peril
Food Science Department
Box 7624
North Carolina State University
Raleigh, NC 27695
Tel: 919/515-3552
Fax: 919/515-7124
maperil@ncsu.edu

Scott E. Baker
Pacific Northwest National Laboratory
MSIN: K2-12
902 Battelle Blvd
Richland, WA 99352
Tel: 509-372-4759
Fax: 509/372-4732
scott.baker@pnl.gov

John E. Battista
Dept. of Biological Sciences
202 Life Sciences Building
Louisiana State University
Baton Rouge, LA 70803
Tel: 225/578-2810
Fax: 225/578-2597
jbattis@lsu.edu

Scott Becker
University of California, San Diego
Department of Bioengineering
9500 Gilman Drive
Mail Code 0412
La Jolla, CA 92093-0412
Tel: 858/349-4440
Fax: 858/822-3120
sabecker@ucsd.edu

Nicholas H. Bergman
Dept. of Microbiology & Immunology
6620B Med Sci II
0620
Ann Arbor, MI 48109
Tel: 734/615-3707
niber@umich.edu

Lavitania Bismart
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
lbismart@microbio.ucla.edu

Fredrick Blattner
University of Wisconsin-Madison
Dept. of genetics
445 Henry Mall Rm B19
Madison, WI 53706
Tel: 608/262-2534
Fax: 608/263-7459
fred@genome.wisc.edu

Steve Brown
Microbial Functional Genomics Group
Environmental Sciences Division
Building 1505, MS 6038
Oak Ridge National Laboratory
Oak Ridge, TN 37831
Tel: 865-574-7314
Fax: 865-576-8646
brownsd@ornl.gov

Jose Manuel Bruno-Barcena
North Carolina State University
South Gardner Hall. Box 7615
Department of Microbiology
Raleigh, NC 27695-7615
Tel: 919/515-3641
Fax: 919/515-7867
jbbarcen@ncsu.edu

Donald A. Bryant
Dept. of Biochemistry and Molecular Biology
Pennsylvania State University
231S Frear Building
University Park, PA 16802
Tel: 814/865-1992
Fax: 814/863-7024
dab14@psu.edu

Rick Cavicchioli
School of Microbiology and Immunology
The University of New South Wales
Sydney, 2052
Australia
Tel: +(61) 2 9385 3516
Fax: +(61) 2 9385 2742
r.cavicchioli@unsw.edu.au

Casey Chen
USC School of Dentistry
925 West 34th Street
Room 4107
Los Angeles, CA 90089
Tel: 213/740-1075
Fax: 213-740-6778
cchen@usc.edu

Denise Chen
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
ddee@ucla.edu

George M. Church
Dept. of Genetics
Harvard Medical School
77 Ave Louis Pasteur
Genetics NRB Room 238
Boston, MA 02115
Tel: 617/432-7562
Fax: 617/432-7663
church@rascal.med.harvard.edu

Julian Davies
Dept. of Microbiology & Immunology
University of British Columbia
Vancouver, B.C.
Canada V6T 1Z4
Tel: 604-822-9308, 604-822-4737
jed@interchange.ubc.ca

Edward F. Delong
Monterey Bay Aquarium Research Institute
7700 Sandholdt Road
Moss Landing, CA 95039-0628
Tel: 831/775-1843
delong@mbari.org

Edgar Ulises Diaz-Peredo
Computational Genomics CIFN/UNAM
Av. Universidad s/n Col. Chamilpa
Cuernavaca
Morelos Mexico
CP 62100
Tel: +52/ 777-313-20-63
Fax: +52/555-622-76-94
ediaz@cifn.unam.mx

Daniel Drell
US Department of Energy
Life Sciences Division/ SC.72
Office of Biological + Environmental Research
1000 Independence Avenue
Tel: 301/903-4742
Fax: 301/903-8521
daniel.drell@science.doe.gov

Ashlee Earl
Harvard Medical School
Department of Microbiology & Molecular Genetics, 200 Longwood Avenue, D-1 #219
Boston, MA 02115
Tel: 617-432-4359
Fax: 617-738-7664
ashlee_earl@hms.harvard.edu

Jonathan Eisen
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301/838-3507
Fax: 301/838-0208
jeisen@tigr.org

Drew Endy
Biological Engineering
68-623
Massachusetts Institute of Technology
Cambridge, MA 02139-4307
617/258-5152
endy@mit.edu

M. Pilar Francino
Genomics Division
Lawrence Berkeley National Laboratory
Evolutionary Genomics Department
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
mpfrancino@lbl.gov

Carolin Frank
Uppsala University
Molecular Evolution Department
Norbyvagen 18C
Uppsala, 75236
Sweden
Tel: +(46) 18 471 43 79
Fax: +(46) 18 471 64 04
carolin.frank@ebc.uu.se

David Frisch
University of Wisconsin, Madison
Genome Center of Wisconsin
445 Henry Mall
Madison, WI 53706
Tel: 608/262-2534
Fax: 608/263-7459
dfrisch@genome.wisc.edu

Pauline Funchain
Human Cancer Genetics
The Ohio State University
480 Medical Research Facility
420 West 12th Avenue
Columbus, OH 43210
Tel: 614/688-4316
funchain.1@osu.edu

Haichun (Hubble) Gao
Microbial Functional Genomics Group
Environmental Sciences Division
Building 1505, MS 6038
Oak Ridge National Laboratory
Oak Ridge, TN 37831
Tel: 865-574-7314
Fax: 865-576-8646
gaoh@ornl.gov

Lilit Garibyan
Harvard Medical School
lilit_garibyan@student.hms.harvard.edu

Terry J. Gentry
Microbial Genomics and Ecology Group
Environmental Sciences Division
Oak Ridge National Laboratory
PO Box 2008, Bldg 1505, MS-6038,
Room 350
1 Bethel Valley Rd (for shipments)
Oak Ridge, TN 37831-6038
Phone: (865) 574-7302
Fax: (865) 576-8646
gentrytj@ornl.gov

Michelle Green
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
Tel: 650-859-5669
Fax: 650-859-3735
green@ai.sri.com

Gabriel Moreno-Hagelsieb
Department of Biology
Wilfrid Laurier University
gmorenohagelsieb@wlu.ca

Tom Hall
Ibis Therapeutics
Isis Pharrmaceruticals
1891 Rutherford Road
Carlsbad, CA 92008
Tel: 760/603-2667
Fax: 760/603-4653
thall@isisph.com

Jo Handelsman
Plant Pathology
Industrial Engineering
College of Agricultural & Life Sciences
College of Engineering
589A Russel Laboratories
1630 Linden Dr
Madison, WI 53706
Tel: 608/263-8783, 608/262-1410
joh@plantpath.wisc.edu

Loren Hauser
UT Genome Science and Technology
Genome Analysis and Systems Modelling
Life Sciences Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6164
Tel: 865/241-1960
Fax: 865/576-9618
hauserlj@ornl.gov

Zhili He
Oak Ridge national Laboratory
1 Bethel Valley Road
Oak Ridge, TN 37831
Tel: 865-576-5134
hez@ornl.gov

John F. Heidelberg
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301/838-0200
Fax: 301/838-0208
jheidel@tigr.org

Christopher D. Herring
University of California, San Diego
Bioengineering
Room 417 Bldg PFBH
9500 Gilman Drive Dept. 0412
La Jolla, CA 92093-0412
Tel: 858/822-1144
cherring@ucsd.edu

Doug Hershberger
The Dow Chemical Company
5501 Oberlin Drive
San Diego, CA 92121
Tel: 858/352-4515
Fax: 858/352-4557
cdhershberger@dow.com

Les Hoffman
Epicentre
726 Post Road
Madison, WI 53713
Tel: 608/442-6146
Fax: 608/442-6156
les.hoffman@epicentre.com

Philip Hugenholtz
DOE Joint Genome Institute
2800 Mitchell Drive Bldg 400-404
Walnut Creek, CA 94598
Tel: 925/296-5725
Fax: 925/296-5720
phugenholtz@lbl.gov

Ayuko Iverson
University of California-Los Angeles
Dept. of Microbiology & Molecular
Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
ayuko@ucla.edu

Vinayak Kapatral
Integrated Genomics Inc.
2201 West Campbell Park Drive
Chicago, Illinois 60612
Tel: 312/491-0846 Ext. 326
Fax: 312/226-9466
vinayak@integratedgenomics.com

Maxine Karimoto
University of California-Los Angeles
Dept. of Microbiology & Molecular
Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
maxinek@ucla.edu

Peter Karp
SRI International
333 Ravenswood Avenue
Room No. EK207
Menlo Park, CA 94025
Tel: 650/859-4358
Fax: 650/859-3735
karp@ai.sri.com

Jay Keasling
Department of Chemical Engineering
University of California, Berkeley
Berkeley, CA 94720-1462
Tel: 510/642-4862
Fax: 510/643-1228
keasling@socrates.berkeley.edu

Paul Keim
Northern Arizona University
Department of Biology
Flagstaff, AZ 86011-5640
Tel: 928/523-1078
Fax: 928/523-7500
Paul.keim@nau.edu

Martin Keller
Diversa Corporation
4955 Directors Place
San Diego, CA
Tel: 858/526-5000
Fax: 858/526-5551
mkeller@diversa.com

Ingrid Keseler
SRI International
333 Ravenswood Avenue
Room No. EK223
Menlo Park, CA 94025
Tel: 650/859-5872
Fax: 650/859-3735
keseler@ai.sri.com

Mandy Kim
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
Mandy177@ucla.edu

Un Mi Kim
University of California-Los Angeles
1602 MSB
609 Charles Young Dr. E.
Los Angeles, CA 90095
Tel: 310/825-4099
Fax: 310/206-5231
unmik@microbio.ucla.edu

Joel Klappenbach
Michigan State University
Center for Microbial Ecology
540 Plant and Soil Sciences Bldg
East Lansing, MI 48824
Tel: 517/353-9021
Fax: 517/353-2917
klappenb@msu.edu

Jim W. Kronstad
Biotechnology laboratory
237-6174 University Blvd
University of British Columbia
Vancouver, B.C., Canada
V6T 1Z3
Tel: 604/822-4732
Fax: 604/822-2114
kronstad@interchange.ubc.ca

Anil Kumar
506, Environmental Biotechnology
Division
Institute of Genomics and Integrative
Biology, DU North Campus
Mall Road, Delhi-110007, India
Tel: +91-11-27662133 Ext. 154
Fax: +91-11-27667471
anilcbt@yahoo.com

Nikos Kyrpides
DOE Joint Genome Institute
2800 Mitchell Drive Bldg 400-404
Walnut Creek, CA 94598
Tel: 925/296-5725
Fax: 925/296-5720
nckyrpides@lbl.gov

Pongpan Laksanalamai
Center of Marine Biotechnology
University of Maryland Biotechnology Institute
701 E. Pratt St.
Baltimore, MD 21202
Tel: 410/234-8872
laksanal@umbi.umd.edu

Frank W. Larimer
Oak Ridge National Laboratory
1 Bethel Valley Road
P. O. Box 2008, Bldg. 5700
Oak Ridge, TN 37831-6164
Tel: 865/574-1253
Fax: 865/576-5332
larimerfw@ornl.gov

Michael Laub
Bauer Center for Genomics Research
Harvard University
7 Divinity Avenue
Cambridge, MA 02138
Tel: 617/384-9647
laub@cgr.harvard.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

Beth Lazazerra
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
University of California, Los Angeles
bethl@microbio.ucla.edu

Huiying Li
University of California, Los Angeles
MBI
611 Charles E. Young Dr. E.
Los Angeles, CA 90095
Tel: 310/206-3642

Wolfgang Liebl
Wolfgang Liebl
Institute of Microbiology and Genetics
Unversity of Goettingen
Grisebachstrasse 8
D-37075 Goettingen
Germany
wliebl@gwdg.de

Anthony Martino, Ph.D.
Biosystems Research Department
P.O. Box 969, MS9951
Sandia National Laboratories
Livermore, CA 94551
Tel:925/ 294-2095
Fax: 925/ 294-3020
martino@sandia.gov

Christian Massire
Ibis Therapeutics
Isis Pharrmaceruticals
1891 Rutherford Road
Carlsbad, CA 92008
Tel: 760/603-2667
Fax: 760/603-4653
cmassire@isisph.com

Roger Milkman
University of Iowa
Marine Biological Laboratory
Iowa City, Iowa 52242
Tel: 508-548-6248
Fax: 508-540-6902
rmilkman@mbl.edu

Jeffrey H. Miller
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8460
Fax: 310/206-3088
jhmiller@mbi.ucla.edu

Araceli Huerta Moreno
Genomic Division
Lawrence Berkeley National Lab and
Volutionary Genomics Department
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: 925-296-5789
Fax: 925-296-5666
amhuerta@jgi.doe.gov

Gabriel Moreno-Hagelsieb.
Department of Biology
Wilfrid Laurier University
gmorenohagelsieb@wlu.ca

Jenna Morgan
Genomics Division
Lawrence Berkeley National Laboratory
Evolutionary Genomics Department
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
jlmorgan@lbl.gov

Garry Myers
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301-795-7860
Fax: 301-838-0208
gmyers@tigr.org

Satoshi Nakagawa
Xanagen Inc.
164-0003
2-7-14, Higashinakano, Nakano-ku
Tokyo, Japan
Tel: 81-3-5332-5182
Fax: 81-3-5332-5190
snakagawa@xanagen.com

Kenneth H. Nealson
Dept. of Earth Sciences
University of Southern California
223 Science Hall
Los Angeles, CA 90089-0740
Tel: 213/ 821-2271
knealson@wrigley.usc.edu

Therese Nguyen
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
tqnguyen@ucla.edu

Shoko Oji
National Institute of Technology and Evaluation (NITE)
2-49-10 Nishihara, Shibuya-ku
Tokyo
Japan
Tel: +81-3-3481-1972
Fax: +81-3-3481-8424
Oji-syoko@nite.go.jp

Jennifer Okada
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
jenokada@ucla.edu

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

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

Julian Parkhill
The Welcome Trust Sanger Inst
Pathogen Sequencing Unit
Genome Campus Hinxton
Cambridge, CB 10 ISA
United Kingdom
Tel: 44 1223 834244
Fax: 44 1223 494919
parkhill@sanger.ac.uk

Guy Plunkett III
University of Wisconsin, Madison
Laboratory of Genetics
445 Henry Mall
Madison, WI 53706
Tel: 608/262-2534
Fax: 608/263-7459
guy@genome.wisc.edu

Steffen Porwollik
Sidney Kimmel Cancer Center
10835 Altman Row
San Diego, CA 92121
Tel: 858/450-5990 Ext. 274
Fax: 858/410-1989
sporwollik@skcc.org

Martha Lucia Posada-Buitrago
Evolutionary Genomics Department
DOE-Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: 925/296-5745
Fax: 925/296-5666
MLPosada-Buitrago@lbl.gov

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@nucleus.szbk.u-szeged.hu

Elisabeth A. Raleigh
New England Biolabs
32 Tozer Road
Beverly, MA 01915
Tel: 978/998-7238
Fax: 978/921-1350
raleigh@neb.com

Rachna Ram
Banfield Geomicrobiology Lab
Department of Environmental Science, Policy, and Management
Hilgard Hall,
University of California, Berkeley
Berkeley, CA 94720-3100
Tel: 510/643-2225
Fax: 510/643-5098

David Rasko
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301-838-0200
Fax: 301-838-0208

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

David A. Relman
Infectious Disease and Geographic Medicine
Stanford University
Tel: 650/852-3308
FAX: 650/ 852-3291
relman@cmgm.stanford.edu

Margaret A. Riley
Ecology/Evolutionary Biology
Yale University
New Haven, CT 06520
Tel: 203/432-3875
Riley.margaret@comcast.net

Monica Riley
Marine Biological Laboratory
7 MBL St.
Woods Hole, MA 02543
Tel: 508/289-7612
Fax: 508/540-6902
mriley@mbl.edu

Edward Rubin
Lawrence Berkeley National Laboratory
One Cyclotron Rd.
Mailstop: 84-171
Berkeley, CA 94720
Tel: 510/486-5072
Fax: 510/486-4229
EMRubin@lbl.gov

Fabiola Sanchez
Computational Genomics CIFN/UNAM
Av. Universidad s/n Col. Chamilpa
Cuernavaca
Morelos Mexico
CP 62100
Tel: +527/313-20-63
Fax: +527/317-55-81
fsanchez@cifn.unam.mx

Carlos Santiviago
Sidney Kimmel Cancer Center
10835 Altman Row
San Diego, CA 92121
Tel: 858/450-5990
Fax: 858/410-1989
csantiviago@skcc.org

Michael A. Savageau
1019 Academic Surge Building
University of California, Davis
Davis, CA 95616
Tel: 530/754-7350
Fax: 530/754-5739
masavageau@ucdavis.edu

Steven K. Schmidt
Dept. of Environmental, Population and Organismic Biology
University of Colorado
Ramaley N122
Campus Box 334
Boulder, Colorado 80309
Tel: 303/492-6248
Fax: 303/492-8699
schmidts@spot.colorado.edu

Patrick D. Schloss
583 Russell Labs
Department of Plant Pathology
University of Wisconsin-Madison
1630 Linden Drive
Madison, WI
Tel: 608/262-8735
Fax: 608/265-5289
pds@plantpath.wisc.edu

Eric Selker
University of Oregon
Institute of Molecular Biology
355D Streisinger Hall
Tel: 541/ 346-5193
selker@molbio.uoregon.edu

Rekha Seshadri
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301-795-7858
Fax: 301-838-0208
rekha@tigr.org

Lynn Silver
LL Silver Consulting
3403 Park Place
Springfield, NJ 07081
Tel: 973/218-1466
Cell: 732/406-1643
Lynx46@erols.com

Michael Strong
University of California, Los Angeles
MBI
611 Charles E. Young Dr. E.
Los Angeles, CA 90095
Tel: 310/206-3642
strong@mbi.ucla.edu

Sebastian Sudek
Marine Biology Research Division
Scripps Institution of Oceanography
University of California, San Diego
9500 Gilman Drive
Mail Code 0208
La Jolla, CA 92093
Tel: 858/822-2866
Fax: 858/534-7313
ssudek@ucsd.edu

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

Katsuji Tani
Osaka University
Environmental Science and
Microbiology
Graduate School of Pharmaceutical
Sciences
1-6 Yamada-oka, Suita
Osaka 565-0871
Japan
Tel: 181/6-6879-8173
Fax: 181/6-6879-8174
tani@phs.osaka-u.ac.jp

Gerardo Toledo
Diversa Corporation
4955 Directors Place
San Diego, CA 92121
Tel: 858/526-5414
Fax: 858/525-5574
gtoledo@diversa.com

Susannah G. Tringe
Lawrence Berkeley National Laboratory
One Cyclotron Rd.
Mailstop: 84-171
Berkeley, CA 94720
Tel: 510/486-5072
Fax: 510/486-4229
925/296-5813
SGTringe@lbl.gov

Allen Tsang
Department of Microbiology
Morrill Science Center IV North
University of Massachusetts
639 North Pleasant St.
Amherst, MA 01003-9298
Tel: 203/824-0232
Fax: 413/545-1578
atsang@microbio.umass.edu

Naofumi Tsukatani
National Institute of Technology and Evaluation (NITE)
2-49-10 Nishihara, Shibuya-ku
Tokyo
Japan
Tel: +81-3-3481-1972
Fax: +81-3-3481-8424
Tsukatani-naofumi@nite.go.jp

Gene Tyson
Banfield Geomicrobiology Lab
Department of Environmental Science, Policy, and Management
Hilgard Hall, University of California, Berkeley
Berkeley, CA 94720-3100
Tel: 510/643-2225
Fax: 510/643-5098
gtyson@nature.berkeley.edu

Clifford J. Unkefer
National Stable Isotope Resource
Los Alamos National Laboratory
B-3, MS G758
Los Alamos, NM 87545
Tel: 505/667-5324
Fax: 505/667-0110
cju@lanl.gov

Kasthuri Venkateswaran
California Institute of Technology
Jet Propulsion Laboratory
Biotechnology and Planetary Protection Group
89-2, Oak Grove Drive
Pasadena, CA 91109
Tel: 818/393-1481
Fax: 818/393-4176
kjvenkat@jpl.nasa.gov

Nathan VerBerkmoes
Oak Ridge National Laboratory
P. O. Box 2008
MS6131
Oak Ridge, TN 37831-6131
Tel: 865/574-2848
Fax: 865/576-8559
nverberk@utk.edu

Alessandro Vezzi
Department of Biology
University of Padova Via Ugo Bassi 58/B
30121, Padova
Italy
Tel: +390 49 8276281
Fax: +390 49 8276280
sandrin@cribi.unipd.it

Barry L. Wanner
Purdue University
915 West State St.
Lilly Hall, Rm. 1-228
West Lafayette, IN 47907-2054
Tel: 765/494-8034 or 765/494-4954
Fax: 765-496-1189
blwanner@bilbo.bio.purdue.edu

Ying Wang
USC School of Dentistry
925 West 34th Street
Room 4107
Los Angeles, CA 90089
Tel: 213/740-4788
wang@hsc.usc.edu

Naomi Ward
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301-838-0200
Fax: 301-838-0208
nward@tigr.org

Owen White
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301-838-0200
Fax: 301-838-0208
owhite@tigr.org

Martin Wu
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Tel: 301/795-7830
Fax: 301/294-3142
mwu@tigr.org

Jizhong Zhou
Oak Ridge national Laboratory
P.O. Box 2008
Oak Ridge, TN 37831
Tel: 865/576-7544
Fax: 865/576-8646
zhouj@ornl.gov

Krystle Zeibell
University of California-Los Angeles
Dept. of Microbiology & Molecular Genetics
1602 Molecular Sciences Dr.
Los Angeles, CA 90095
Tel: 310/825-8924
Fax: 310/206-3088
kzeibell@ucla.edu

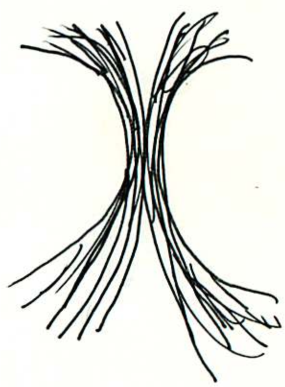
Julie Zimmerman
Louisiana State University and
A & M College
Biological Sciences
626 Life Sciences Bldg
Baton Rouge, LA 70803
Tel: 225/578-5226
jzimme5@lsu.edu

Jeff Miller

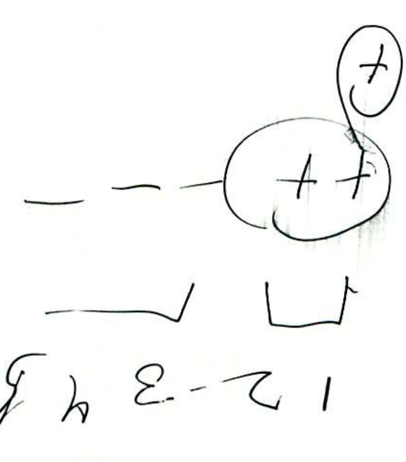
- Intro talk about relativity
- Human mind can concerned things w/ no basis in physical reality
- dinosaurs
- space-time continuum of genes + genomes
- the wormhole that allows one to tunnel into to today world
- tunnel back thru that wormhole

gene sequences

PCR gDNA



Pyrrolysine 22 a.a.
Hao et al Science 296:



can anything from earth survive on m