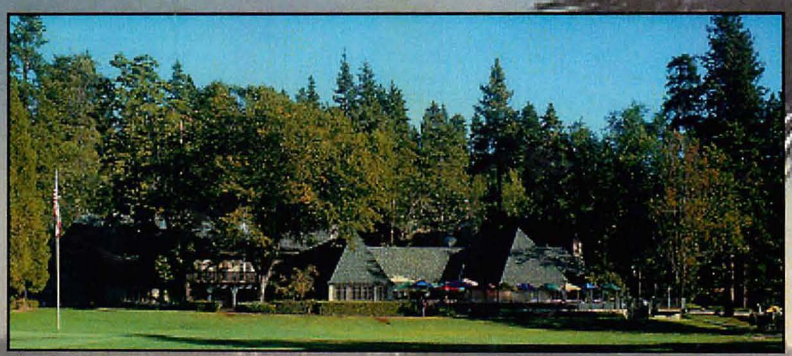
18th International Microbial Genomes Conference



September 12-16, 2010

UCLA Conference Center
Lake Arrowhead
California

SCIENTIFIC PROGRAM ORGANIZERS

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

Dr. George Weinstock,
Washington University, St. Louis, MO

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

Dr. Elisabeth Raleigh
New England BioLabs, Beverly, Massachusetts

Dr. Jizhong Zhou
University of Oklahoma, Norman, Oklahoma

Dr. Fredrick Blattner
University of Wisconsin, Madison, Wisconsin

Dr. Ashlee Earl
The Broad Institute of Harvard and MIT, Cambridge, Massachusetts

ACKNOWLEDGEMENTS

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

National Institutes of Health
National Institute of Allergy and Infectious Diseases
National Cancer Institute
National Science Foundation

UCLA-DOE Institute for Genomics & Proteomics

New England Biolabs

Proctor & Gamble

Dr. Elisabeth Raleigh

CONTACT NUMBER

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

SCIENTIFIC PROGRAM

SUNDAY, SEPTEMBER 12

4:00-6:00 pm	Arrival and Check-in at Lake Arrowhead Conference Center
6:15-7:45 pm	Dinner (Dining Room)
Opening of Meeting	(Pineview Room)
7:45-8:05 pm	Jeffrey H. Miller University of California, Los Angeles **"Welcome"**
8:05-9:00 pm	**Keynote Address** David A. Relman Stanford University School of Medicine, Stanford, CA **"Advances in Understanding the Human Microbiome"**
9:00 pm	Reception (Iris Room)

MONDAY, SEPTEMBER 13

7:45-8:30 AM	Breakfast (Dining Room)
Opening Session	**(Pineview Room)**
8:45-9:00 am	Introduction/Announcements
Session I	**Microbial Communities I: Microbiomes/Biodiversity**
9:00-9:30 am	Nina R. Salama Fred Hutchinson Cancer Research Center, Seattle, WA **"Helicobacter pylori: Diversity, interaction with the resident microbiome, and its role in gastric cancer"**
9:30-10:00 am	Huiying Li University of California and the David Geffen School of Medicine, Los Angeles, CA **"The Human Skin Microbiome: *Acne vulgaris*"**
10:00-10:30 am	Julia A. Segre National Human Genome Research Institute, Bethesda, MD **"Diversity of the Human Skin Microbiome: Eczema and Immunodeficiencies"**

SCIENTIFIC PROGRAM

SUNDAY, SEPTEMBER 12

4:00-6:00 pm	Arrival and Check-in at Lake Arrowhead Conference Center
6:15-7:45 pm	Dinner (Dining Room)
Opening of Meeting	(Pineview Room)
7:45-8:05 pm	Jeffrey H. Miller University of California, Los Angeles **"Welcome"**
8:05-9:00 pm	**Keynote Address** David A. Relman Stanford University School of Medicine, Stanford, CA **"Advances in Understanding the Human Microbiome"**
9:00 pm	Reception (Iris Room)

MONDAY, SEPTEMBER 13

7:45-8:30 AM	Breakfast (Dining Room)
Opening Session	**(Pineview Room)**
8:45-9:00 am	Introduction/Announcements
Session I	**Microbial Communities I: Microbiomes/Biodiversity**
9:00-9:30 am	Nina R. Salama Fred Hutchinson Cancer Research Center, Seattle, WA **"Helicobacter pylori: Diversity, interaction with the resident microbiome, and its role in gastric cancer"**
9:30-10:00 am	Huiying Li University of California and the David Geffen School of Medicine, Los Angeles, CA **"The Human Skin Microbiome: *Acne vulgaris*"**
10:00-10:30 am	Julia A. Segre National Human Genome Research Institute, Bethesda, MD **"Diversity of the Human Skin Microbiome: Eczema and Immunodeficiencies"**

10:30-10:50 am	Break
10:50-11:20 am	Larry Forney University of Idaho, Moscow, ID **"Temporal Dynamics of Vaginal Microbes in Reproductive Age Women"**
11:20-11:50 am	Cindy M. Liu Translational Genomics Research Institute, Flagstaff, AZ **"The Penis Microbiome: Uncovering the Impacts of Circumcision"**
11:50-12:20 pm	Ashlee Earl The Broad Institute of Harvard and MIT, Cambridge, MA **"Our Microbes: A reference Genome Collection for the Human Microbiome"**
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: Metagenomics/Biodiversity**
7:45-8:15 pm	Steve Kembel University of Oregon, Eugene, OR **"Phylogenetic Ecology and Metagenomics"**
8:15-8:45 pm	Tanja Woyke DOE Joine Genome Institute, Walnut Creek, CA **"Single Cell Genomics"**
8:45-9:00 pm	Break
9:00-9:30 pm	Shannon Williamson J. Craig Venter Istitute, San Diego, CA **"Single Virus Genomics: Changing the Landscape of Virology"**
9:30-10:00 pm	TBA

TUESDAY, SEPTEMBER 14

7:45-8:30 am	Breakfast (Dining Room)
Session III	**Pathogens/Antibiotics/Resistance**
8:45-8:55 am	Jeffrey H. Miller University of California, Los Angeles, CA **Introduction/Overview**
8:55-9:25 am	Gautam Dantas Washington University School of Medicine, St. Louis, MO **"Antibiotic Resistance Reservoirs in Human and Environmental Microbiomes"**
9:25-9:55 am	Shauna McGillivray Texas Christian University, Forth Worth, TX **"Proteases and Virulence of *Bacillus anthracis*"**
9:55-10:25 am	Michael A. Fischbach University of California, San Francisco, CA **"Small Molecules from Microbes: Streptomyces"**
10:25-10:45 am	Break
10:45-11:15 am	Lance B. Price Translational Genomics Research Institute, Flagstaff, AZ **"Multidrug Resistant *Staphylococcus aureus* from Food Animals: Whole Genome Sequence Analysis of Clonal Complex 398"**
11:15-11:45 am	Magdalene So University of Arizona, Tucson, AZ **"Comparative Genomics of Commensal and Pathogenic Neisseria"**
11:45-12:15 pm	Colin Manoil University of Washington, Seattle, WA **"Phenotyping a Non-Model Prokaryote at Genome Scale"**
12:15-12:30 pm	Michael Y. Galperin NCBI/NLM/NIH, Bethesda, MD **"Evolution of Bacterial Signal Transduction"**
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 IV	**Evolution/Extremophiles**
7:45-8:15 pm	Eric J. Alm The Broad Institute of Harvard and MIT, Cambridge, MA **"Darwin's Footprints in the Microbial World"**
8:15-8:45 pm	Jonathan Eisen University of California, Davis, CA **"A Genomic Encyclopedia of Bacteria and Archaea"**
8:45-9:00 pm	Break
9:00-9:30 pm	Frederico Lauro University of New South Wales, Sydney, Australia **"Integrative Approaches to the Studies of Natural Microbial Communities"**
9:30-10:00 pm	Valérie de Crécy-Lagard University of Florida, Gainesville, FL 32611 **"A Tale of Two Proteins: Discovering the Function of Two Protein Families of Unknown function Conserved Throughout the Tree of Life"**

WEDNESDAY, SEPTEMBER 15

7:45-8:30 am	Breakfast (Dining Room)
Session V	**COMBREX Workshop for Computational and Experimental Determination of Protein Function**
8:45-9:05 am	Richard J Roberts New England Biolabs, Beverly, MA **"Identifying Protein Function-A Call for Community Action"**

9:05-9:25 am	Peter Karp SRI International, Menlo Park, CA **"Identification of Annotation Error and Incompleteness through Network Analysis"**
9:25-9:45 am	Steven Brenner University of California, Berkeley, CA **"Function and dysfunction in protein annotation"**
9:45-10:05 am	Bruno Sobral Virginia Tech, Blacksburg, VA **"Prokaryotic Annotation Status"**
10:05-10:25 am	Break
10:30-10:50 am	Manuel Ferrer CSIC-Institute of Catalysis, Marie Curie 2, Madrid, Spain **"High Throughput Arrays for Protein Function"**
10:50-11:10 am	Alexander Yakunin University of Toronto, Toronto, Canada **"Application of Enzymatic Assays for Experimental Annotation of Unknown Proteins"**
11:10-11:30 am	John Gerlt University of Illinois, Urbana-Champaign, IL **"Predicting and Discovering New Functions in the Enolase Superfamily"**
11:30-11:50 am	Patricia Babbitt University of California, San Francisco, CA **"Annotation Error in Public Databases"**
11:50-12:10 am	Simon Kasif Boston University, Boston, MA **"Breaking the Power Law Curse: A Computational Bridge to Experiments"**
12:10-12:20 am	Iddo Friedberg **Introduction of breakout groups for afternoon discussion**
12:30 pm	Lunch (Dining Room)

Session VI	**COMBREX Workshop for Computational and Experimental Determination of Protein Function**
4:15-5:30 pm	Breakout Groups convene Annotation/DB issues
	Iddo Friedberg Miami University, Oxford, OH **Computational predictions**
	Andrei Osterman Sanford-Burnham Medical Research Institute, La Jolla, CA **HT/LT experimentation**
	Martin Steffen Boston University School of Medicine, Boston, MA **COMBREX nuts and bolts**
5:30-6:15 pm	Report Preparation
6:15-7:45 pm	Dinner (Dining Room)
7:45-7:55 pm	**Poster Awards Presentation**
7:55-8:10 pm	BG1 presentation
8:10-8:25 pm	BG2 presentation
8:25-8:40 pm	BG3 presentation
8:40-8:55 pm	BG4 presentation
8:55 pm	Richard J. Roberts New England Biolabs, Beverly, MA **Closing**
9:25 pm	Reception/Party (Iris Room)

THURSDAY, SEPTEMBER 16

7:30-8:20 am	Breakfast (Dining Room)

Session VII	**Bioenergetics/Regulatory and Metabolic Pathways/Modeling**
8:30-8:55 am	Jizhong (Joe) Zhou University of Oklahoma, Norman, OK **"Rapid Genome Evolution and Adaptation to Salt Selection"**
8:55-9:25 am	Athanasios Lykidis DOE Joint Genome Institute, Walnut Creek, CA **"The Genomics of Bacterial Lipid Metabolism"**
9:25-9:50 am	Kimmen Sjölander University of California, Berkeley, CA **"The PhyloFacts Phylogenomic Microbial Protein Family Encyclopedia"**
9:50-10:30 am	TBA
10:45 am	**Checkout**
11:00 am	**Departure of Bus for LAX**
12:00 noon	**Lunch**
1:00 pm	**Departure of Vans for LAX**

SPEAKER ABSTRACTS

Advances in Understanding the Human Microbiome

David A. Relman

Stanford University School of Medicine, Stanford, CA

Complex microbial ecosystems occupy the cutaneous and mucosal surfaces of humans. Recent advances in our understanding of these ecosystems have highlighted the tremendous diversity of these communities, features of individuality, conserved and personalized predicted functional attributes, and their importance to host physiology. Questions remain about the ecological processes that establish and maintain the human microbiota throughout life, as well as the features of community robustness that are associated with stability, and with prevention of invasion or successful competition by potential pathogens. We have examined the distal gut microbiota of three individuals over 10 months that spanned two courses of the antibiotic ciprofloxacin. The effect of ciprofloxacin on the gut microbiota was profound and rapid, with a loss of diversity and a shift in community composition occurring within 3-4 days of drug initiation. By one week after the end of each course, communities began to return toward their initial state, but the return was often incomplete. While broadly similar, community changes after ciprofloxacin varied among subjects and between the two courses within subjects. In all subjects, the composition of the gut microbiota stabilized by the end of the experiment, but was altered from its initial state. As with other ecosystems, the human distal gut microbiome at baseline is a dynamic regime with a stable average state. Antibiotic perturbation may cause a shift to an alternative stable state, the full consequences of which remain unknown. From a broad perspective, approaches that combine ecological theory and statistics, sequence-based and other molecular assessments of community structure an host response, and standardized clinical measurements may improve our understanding of health and disease within the human supra-organism.

Bacterial community dynamics in the stomach: *Helicobacter pylori* and beyond

Nina Salama
Marion Dorer, Jutta Fero, Olivier Humbert, and Sarah Talarico

Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle WA 98109

The discovery of *Helicobacter pylori* as a resident of the human stomach that causes chronic disease (peptic ulcer and gastric cancer) was radical on many levels. While the mouth and the colon were both known to host a large number and diversity of bacteria (microbiome), the stomach was thought to be sterile due to its high acidity. We now know that *H. pylori* is one of many species of bacteria that live in the stomach, although *H. pylori* seems to dominate this community when present. *H. pylori* does not behave as a classical bacterial pathogen: disease is not solely mediated by production of toxins, although certain *H. pylori* genes, including those that encode exotoxins, increase the risk of disease development. Paradoxically, some of these same "virulence" factors also correlate with protection from other diseases, particularly with respect to diseases of the esophagus. Indeed *H. pylori* associated disease risk seems to result from a complex interaction between the bacterium, the host, and the environment (which may include other constituents of the microbiota at the gastro-esophageal junction). The innate and adaptive immune defenses of the host, combined with factors in the environment of the stomach, apparently drive a continual high rate of genomic variation in *H. pylori*. Central to this genetic diversification is *H. pylori's* natural competence which is induced by DNA damage stress encountered during stomach colonization. Interestingly sequence divergence does not present a major barrier to transformation, but multiple restriction modification systems cooperate to limit the extent of DNA mosaics generated during transformation. Dense sampling of *H. pylori* populations within the human stomach suggests that subsequent to infection an *H. pylori* strain generates de novo variants and samples exogenous DNA to generate a diverse community of clones that facilitates persistence in the stomach for decades.

Bacterial community dynamics in the stomach: *Helicobacter pylori* and beyond

Nina Salama
Marion Dorer, Jutta Fero, Olivier Humbert, and Sarah Talarico

Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle WA 98109

The discovery of *Helicobacter pylori* as a resident of the human stomach that causes chronic disease (peptic ulcer and gastric cancer) was radical on many levels. While the mouth and the colon were both known to host a large number and diversity of bacteria (microbiome), the stomach was thought to be sterile due to its high acidity. We now know that *H. pylori* is one of many species of bacteria that live in the stomach, although *H. pylori* seems to dominate this community when present. *H. pylori* does not behave as a classical bacterial pathogen: disease is not solely mediated by production of toxins, although certain *H. pylori* genes, including those that encode exotoxins, increase the risk of disease development. Paradoxically, some of these same "virulence" factors also correlate with protection from other diseases, particularly with respect to diseases of the esophagus. Indeed *H. pylori* associated disease risk seems to result from a complex interaction between the bacterium, the host, and the environment (which may include other constituents of the microbiota at the gastro-esophageal junction). The innate and adaptive immune defenses of the host, combined with factors in the environment of the stomach, apparently drive a continual high rate of genomic variation in *H. pylori*. Central to this genetic diversification is *H. pylori's* natural competence which is induced by DNA damage stress encountered during stomach colonization. Interestingly sequence divergence does not present a major barrier to transformation, but multiple restriction modification systems cooperate to limit the extent of DNA mosaics generated during transformation. Dense sampling of *H. pylori* populations within the human stomach suggests that subsequent to infection an *H. pylori* strain generates de novo variants and samples exogenous DNA to generate a diverse community of clones that facilitates persistence in the stomach for decades.

Skin Microbiome in Health and Disease

Julia A. Segre
Elizabeth A. Grice[1], Heidi H. Kong[2], Sean Conlan[1], Evan Snitkin[1], Lilia Mijares[3], Patrick Murray[3], Eric D. Green[1], Maria L. Turner[2]

NISC Comparative Sequencing Program[1].[1]NHGRI, NIH, Bethesda, MD
[2]Dept. of Dermatology, CCR, NCI, NIH
[3]Dept. of Laboratory Microbiology, NIH Clinical Center, NIH.

While human skin serves as a first line of defense against pathogenic bacteria, it also provides a home to billions of symbiotic bacteria. By sequencing the DNA of bacteria collected from the skin of humans and mouse models of human disease, we investigate how these bacteria contribute to health and, conversely, how changes in the bacterial community structure might contribute to chronic skin disorders. Our analysis of 16S ribosomal RNA gene sequences obtained from 20 distinct skin sites of healthy humans revealed that physiologically comparable sites harbor similar bacterial communities and provides a baseline for studies that examine the role of bacterial communities in disease states and the microbial inter-dependencies required to maintain healthy skin. We explore the selective shift in the microbiota observed in skin disorders commonly treated with antimicrobial agents such as eczema and diabetic wound healing. For example, we show that a longitudinal selective shift in wound microbiota coincides with impaired healing in diabetic mice. We demonstrate a parallel shift in longitudinal gene expression that occurs in a cluster of genes related to the immune response. Further, we establish a correlation between relative abundance of *Staphylococcus* spp. and the expression of cutaneous defense response genes. This data demonstrate that integrating two types of global datasets lends a better understanding to the dynamics governing host-microbe interactions. Clinical management of these disorders requires better biomarkers to realize the therapeutic potential of manipulating the microbiome. Targeted therapies to maintain healthy skin might require not only inhibiting the growth of pathogenic bacteria, but also promoting the growth of symbiotic bacteria. Microbiome studies show the power of bridging genomics and clinical medicine to gain valuable insight into the human body as a complex super-organism.

Temporal Dynamics of Vaginal Microbiomes in Reproductive Age Women

Larry J. Forney
Pawel Gajer[1], Zaid Abdo[2], Rebecca M. Brotman[1], Stacey McCulle[1], Sara S.K. Koenig[1], Joyce Sakamoto[1], Sam Ma[3], Ursel Shütte[3] and Jacques Ravel[3]

[1] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; [2] Departments of Mathematics and Statistics and the Initiative for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID; [3] Department of Biological Sciences and the Initiative for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID

It is now recognized that the vaginal microbiota along with intrinsic host factors play a major role in maintaining the vaginal health of women. We have hypothesized that differences in the species composition of vaginal bacterial communities in health may affect the level of community dynamics, hence the risk of developing bacterial vaginosis or acquiring communicable diseases. We have analyzed the composition and structure of the resident bacterial community in 400 healthy women in four ethnic groups in North America and evaluated the level of community dynamics in 33 women sampled twice weekly for 16 weeks using pyrosequencing of barcoded 16S rRNA genes. Statistical analyses revealed five dominant community states; four of which had high proportions of *Lactobacillus species.* Longitudinal analysis of the vaginal microbiota showed the communities were dynamic at both at the species level and more surprisingly at the strain (within a species) level, and that continua might exist between some or all community states. Further, we observed that the dynamics of communities vary between women and changes in community composition are sometimes reflected in Nugent scores, a commonly used diagnostic for bacterial vaginosis. Analyses of the frequency of transition between different community groups in a time series reveal communities that were highly stable, while higher diversity was indicative of higher transition to another types. Menses is an important factor in affecting microbial community differences among samples (p-value 0.005) using a distance based redundancy analysis and survival analysis using Extended Cox model controlling for time dependence between samples within a women. Individual microbiome dynamics have also been observed and should be taken into consideration in designing personalized medicine treatments.

The Penis Microbiome: Uncovering the Impacts of Circumcision

Cindy M. Liu
Price LB 1, Aziz M 1, Contente-Cuomo T 1, Buchhagen J 1, Ravel J 3, Keim PS 1, Serwadda S 4, Wawer MJ 5, and Gray R 5

1 TGen North Center for Microbiomics and Human Health, Flagstaff, Arizona, 2 Center for Microbial Genetics and Genomics (MGGen), Northern Arizona University, Flagstaff, Arizona, 3 University of Maryland School of Medicine, Baltimore, Maryland, 4 School of Public Health, Makerere University, Kampala, Uganda, 5 Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

Circumcision is associated with significant reductions in HIV, HSV-2 and HPV infections among men. The biological mechanisms for these associations are not yet known, but may involve alterations in the penis microbiome. In a previous study, we reported significant changes in the microbiota of the coronal sulci associated with circumcision. In the current study, we assessed the penile microbiota in 128 HIV-negative Ugandan men taking part in a randomized control trial of circumcision. Sixty-six men from the intervention arm and 62 men from the control arm of the study were sampled at baseline and 1 year after enrolment. Microbiota were characterized using sequence-tagged 16S rRNA gene pyrosequencing targeting the V3-V4 hypervariable regions. Taxonomic classification was performed using the RDP Naïve Bayesian Classifier. Circumcision was associated with a significant change in the overall microbiota and with a significant decrease in putative anaerobic bacterial families. Specifically, two families—Clostridiales Family XI and Prevotellaceae—were uniquely abundant before circumcision. The anoxic microenvironments of the subpreputial space may support anaerobes that can activate Langerhans cells to present HIV to CD4 cells in draining lymph nodes. Thus, the reduction in putative anaerobic bacteria after circumcision may play a role in protection from HIV and other sexually transmitted diseases.

Phylogenetic Ecology and Metagenomics

Steven Kembel

University of Oregon, Eugene, OR

Patterns of phylogenetic diversity (evolutionary relatedness among co-occurring organisms) can provide important insights into the mechanisms underlying community assembly. Previous studies of microbial diversity using metagenomic data have generally quantified the structure of microbial assemblages by binning metagenomic sequences into taxonomically or functionally similar groups based on sequence similarity, and to date it has not been possible to apply phylogenetic ecology methods to these data sets. To address this challenge, we present a novel approach for inferring phylogenetic relationships among assemblages of microorganisms based on metagenomic data. This method uses fully sequenced bacterial genomes as a scaffold to enable inference of phylogenetic relationships among metagenomic sequences from multiple phylogenetic marker gene families. I present several applications of this method to understand patterns of microbial diversity and community assembly along spatial and environmental gradients.

Single Virus Genomics: Changing the Landscape of Virology

Shannon Williamson

J. Craig Venter Institute, San Diego, CA

Whole genome amplification and sequencing of single microbial cells has significantly influenced genomics and microbial ecology by facilitating direct recovery of reference genome data. However, viral genomics continues to suffer due to difficulties related to the isolation and characterization of uncultivated viruses. We report here on a new approach called 'Single Virus Genomics', which enabled the isolation and complete genome sequencing of the first single virus particle. A mixed assemblage comprised of two known viruses; E. coli bacteriophages lambda and T4, were diluted and sorted using flow cytometric methods and subsequently immobilized in an agarose matrix. Genome amplification was then achieved in situ via multiple displacement amplification (MDA). The complete lambda phage genome was recovered with an average depth of coverage of approximately 500X. The isolation and genome sequencing of uncultivated viruses using Single Virus Genomics approaches will enable researchers to address questions about viral diversity, evolution, adaptation and ecology that were previously unattainable.

Antibiotic Resistance Reservoirs in Human and Environmental Microbiomes

Gautam Dantas

Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, MO 63108

Increasing rates of multi-drug resistance in human pathogens and declining antibiotic development pipelines are precipitating a worsening global infectious disease crisis. Steady use and abuse of antibiotics over the past century in food animals, humans and the environment has provided substantial selective pressure for enrichment of resistance genotypes in their associated microbiomes. Antibiotic resistance determinants encoded on mobilizable elements can disseminate between these interacting microbial communities, which may all serve as reservoirs of resistance for pathogens. We have applied high-throughput phenotypic assays with metagenomic functional selections to characterize reservoirs of transferable antibiotic resistance genes encoded by the microbiomes of healthy humans as well as diverse soils. These include a set of phylogenetically diverse soil microbes we recently discovered with the capacity to consume antibiotics as their sole source of carbon and resist them at extremely high concentrations. We find that hundreds of resistance genes we identify from these different microbial communities are identical to resistance genes found in major human pathogens, indicating recent genetic exchange between these microbes. We also find hundreds of functionally validated resistance genes which are genetically novel, highlighting the potential that microbes from diverse environments may exacerbate the problems with clinical resistance by serving as reservoirs of novel resistance genes accessible to pathogens.

Role of the ClpXP Protease in the Virulence of *Bacillus anthracis*

Shauna McGillivray

Texas Christian University, Forth Worth, TX

Bacillus anthracis, the causative agent of anthrax, must avoid an array of host antibacterial defenses during the course of infection. Anthrax toxin and capsule, virulence factors located on two large plasmids, are known to play important roles in the pathogenesis of the disease. However, evidence indicates that chromosomal genes may also contribute to virulence. We generated a mutant library of *B. anthracis* Sterne using a Mariner-based transposon mutagenesis system and screened for attenuation of virulence-associated phenotypes. This yielded a number of candidates including a mutant with a disruption of the chromosomal *clpX* gene. ClpX is a regulatory ATPase that works in conjunction with the proteolytic subunit, ClpP, to generate the intracellular bacterial protease ClpXP. *B. anthracis* lacking ClpX have decreased proteolytic and hemolytic activity, increased susceptibility to antimicrobial peptides and severely attenuated virulence *in vivo*. In addition, loss of ClpX renders *B. anthracis* more susceptible to certain antibiotics. This susceptibility to antimicrobial peptides and antibiotics could be reproduced through treatment with a pharmacological inhibitor of the ClpXP protease. The importance of the ClpXP protease for resistance to these antimicrobial agents is not unique to *B. anthracis* as similar effects are seen using the pharmacological inhibitor with another pathogen, *Staphylococcus aureus*. Inhibition of the ClpXP system could be an attractive therapeutic target as it may affect a range of bacterial pathogens and increase bacterial susceptibility not only to host defense, but also antibiotic treatment.

Small Molecules from Microbes: Streptomyces

Michael A. Fischbach

University of California, San Francisco, CA

The discovery of natural products - small molecules from microbes often used as drugs - has been an ad hoc pursuit for almost a century. The rapidly growing database of microbial genome sequences offers new opportunities to leverage genomics and bioinformatics toward discovering natural products and characterizing their roles in mediating interspecies interactions. This lecture will describe three convergent, ongoing lines of research: our use of genomics and bioinformatics to identify biosynthetic genes and predict the structures of their small molecule products, our characterization of a new class of biosynthetic gene clusters that produce a set of heavily modified peptide antibiotics, and our new focus on small molecules produced by human-associated microbes.

Multidrug Resistant Staphylococcus aureus from Food Animals: whole genome sequence analysis of clonal complex 398

Lance B. Price
Frank Aarestrup 2, Robert Skov 3, Marc Stegger 3, Henrik Hasman 2, Lindsey Watson 1, Tania Contente-Cuomo 1, Andrew E. Waters 1, Jordan Buchhagen 1, James Schupp 1, John Gillece 1, Mia Champion 1, Elizabeth Driebe 1, Dave Engelthaler 1, and Paul S. Keim 1.

1 Translational Genomics Research Institute (TGen), Pathogen Genomics Division (TGen North), Flagstaff, Arizona, USA
2 National Food Institute, Technical University of Denmark, Lyngby, Denmark
3 Statens Serum Institut, Copenhagen, Denmark

Methicillin-resistant Staphylococcus aureus (MRSA), a common cause of nosocomial and community-associated infections, has emerged as a global life-threatening pathogen. Recently, it has been estimated that 1.5% of the U.S. population (~4.1 million persons) is colonized with MRSA, and that the number of deaths due to MRSA infections in the U.S. has eclipsed the number of deaths from any other infectious disease, including HIV/AIDS. In 2003, a new MRSA lineage, multilocus sequence type 398 (ST398) emerged in the community. Previously, MRSA ST398 was only associated with livestock farming, but it has now been shown to colonize and cause invasive disease in humans. The zoonotic potential of MRSA ST398 is of great concern, especially since livestock is thought to be the source of other emerging antimicrobial-resistant pathogens in the community. MRSA isolates belonging to the ST398 clonal lineage are difficult to characterize using currently available methods, hampering the epidemiological assessment of transmission and outbreaks. Rapid, accurate analytical methods that can identify genetic diversity within MRSA ST398 are needed to track the sources and dissemination of this global pathogen. Here, we describe identification of high-resolution single nucleotide polymorphisms (SNPs) using whole genome sequencing. Forty ST398 isolates were collected from humans and livestock from Europe, Asia and North America and sequenced on the Illumina Genome Analyzer II using a multiplex paired-end preparation. Sequences were aligned and assembled using a finished ST398 reference genome. From these sequence data, SNPs were identified that allowed for the discrimination of ST398 from other MRSA clonal lineages and for improved differentiation of strains within the ST398 clonal group. The identification of these genomic signatures will aid in more robust epidemiological investigations including the putative animal origins of human ST398 infections.

Genome Comparisons Reveal Widespread Virulence Gene Exchange Among *Neisseria* Species and the Role of the Type IV Pilus in *Neisseria* Biology

Magdalene So
P.R. Marri, D. Higashi, N. Weyand and G. Weinstock

University of Arizona, Tucson, and Washington University, St. Louis

To understand the genetic basis of virulence in *Neisseria gonorrhoeae* and *Neisseria meningitidis*, we undertook a comprehensive genome analysis of the entire *Neisseria* genus. We generated high quality draft genome sequences of eight species of commensal *Neisseria* from the human mucosa, and compared them to the 11 published genomes of the two pathogens. Commensal *Neisseria* are the oldest members of this genus; the two pathogens are the newcomers. Commensals have an extensive repertoire of virulence alleles. A large fraction of these alleles has been exchanged among the *Neisseria*. Commensals also have the genes and traits that allow them to acquire DNA from other *Neisseria*. These findings strongly suggest that commensal *Neisseria* serve as reservoirs of virulence alleles, and that members of this genus engage extensively in genetic exchange *in vivo*.

The type IV pilus (Tfp) is a structure produced by many bacteria and members of the Archaea. It is considered a virulence factor for *N. gonorrhoeae* and *N. meningitidis* because it promotes bacterial attachment to human cells. Our genome study revealed the presence of a complete set of Tfp biogenesis genes in all *Neisseria* species. Recent studies indicate that these genes are expressed in *N. elongata*, the oldest member of the *Neisseria* genus. Tfp is therefore an ancestral trait whose function is not solely linked to pathogenesis. These findings lend additional support to our cell biological observations that Tfp of *N. gonorrhoeae* promote host adaption rather than pathogenesis.

Phenotyping a Non-Model Prokaryote at Genome Scale

Colin Manoil
M. Enstrom, L. Gallagher & B. Ramage

University of Washington, Seattle

A limitation of traditional genome annotation has been its reliance on sequence similarity for predicting gene function. In this study, we developed an approach employing large-scale phenotyping to help identify the biological functions of predicted genes. The method was used to analyze a non-model prokaryote with a small genome, the Gram-negative pathogen *Francisella novicida*. The phenotyping was carried out by measuring the growth of the members of a comprehensive transposon mutant library under numerous nutritional and stress conditions, including in the presence of antibiotics. Phenotypes were observed for mutants affected in about 40% of the nonessential genes of the bacterium, corresponding to eleven carbon source utilization pathways, nine amino acid and nucleotide biosynthetic pathways, seven intrinsic antibiotic resistance networks and six other stress resistance networks. For nearly every phenotype examined, there were unexpected gene associations, and for some phenotypes, the majority of assignments were novel. The results made it possible to assign mutant phenotypes to hypothetical genes and other incompletely defined genes, to identify new metabolic and stress resistance functions, and to group phenotypes based on how frequently they were jointly altered by individual mutations.

A phylogeny driven genomic encyclopedia of bacteria and archaea and the search for the dark matter of the biological universe

Jonathan Eisen

JGI and UC Davis

Sequencing of bacterial and archaeal genomes has revolutionized our understanding of the many roles played by microorganisms, including those related to human health. Unfortunately, the available genomes have been from a highly biased subset of the known phylogenetic diversity of microbes. In order to explore the value added by choosing microbial genomes for sequencing based on their evolutionary relationships, the Joint Genome Institute in partnership with the DSMZ has sequenced the genomes of ~175 culturable species of bacteria and archaea selected to maximize phylogenetic coverage. I will present results from analysis of these genomes that demonstrate that there are pronounced benefits (compared to an equivalent set of genomes randomly selected from the existing database) in diverse areas including the reconstruction of phylogenetic history, the discovery of new protein families and novel biological properties, and the prediction of functions for known genes from other organisms. The results strongly support the need for systematic 'phylogenomic' efforts to compile a phylogeny-driven 'Genomic Encyclopedia of Bacteria and Archaea' in order to derive maximum knowledge from existing microbial genome data as well as from genome sequences to come. I will also present new results on characterization of the protein family diversity from these genomes and show how this not only improves our ability to analyze metagenomic data but also shows that there are likely at least 10 million protein families present in bacteria and archaea; most of which remain unsampled and unstudied. Finally I will present related analyses of protein families in Metagenomic data that revel a that, even for housekeeping genes we have barely scratched the surface of sampling sequence diversity and will present data consistent with the existence of many uncharacterized deep branches in the tree of life.

Integrative Approaches to the Studies of Natural Microbial Communities

F. M. Lauro
R. Cavicchioli

The University of New South Wales, Sydney, Australia

Through the use of metagenomics (shot-gun sequencing of an environmental sample) and associated functional studies, whole ecosystems have begun to be described. In an acid mine drainage system dominated by only a few species, almost complete composite genomes and metabolic reconstruction of pathways has been achieved. In contrast, the Global Ocean Survey has uncovered an enormous extent of microbial diversity and functional potential of oceanic microbial communities.
What has not yet been achieved is to determine if a structured ecosystem with distinct microbial communities can be described in its entirety. The aim is not only define the identity and functional capacity of microorganisms, but to establish the interactions between populations that fulfil nutrient cycling and shape the evolution of the microbial communities. To achieve this goal new approaches to data analysis and integration have to be employed.
As an example I will describe the development and application of a model that predicts, from genomic sequences, the adaptation of marine bacteria to nutrient levels. Marine bacteria have evolved to grow optimally at either high (copiotrophs) or low (oligotrophs) concentrations of nutrients to exploit the many different oceanic niches.
In general, the copiotrophs have larger genomes, with a higher number of rRNA operons, predicted prophages and CRISPR repeats. Their predicted proteomes also contained significantly larger percentages of non-cytoplasmic proteins. COG categories N, V, K, T are enriched in the copiotrophic genomes while I and Q in the oligotrophic.
The application of the trophic model to environmental shotgun projects and its integration within the larger framework of the physico-chemical metadata sheds new light on the microbial dynamics occurring in complex aquatic environments.

A Tale of Two Proteins: Discovering the Function of Two Protein Families of Unknown function Conserved Throughout the Tree of Life

Valérie de Crécy-Lagard
Basma El Yacoubi

University of Florida, Gainesville, FL

Deciphering both the cellular and molecular roles of universally conserved proteins of unknown function is a challenge. Of the top-ten universal protein families of unknown function awaiting characterization listed by Galperin and Koonin in 2004[1], only half have been assigned a cellular function in 2010[2]. We have previously shown that one of those , the Sua5/YrdC family (YrdC), is involved in the biosynthesis of N^6-threonylcarbamoyl adenosine (t^6A) [3], a universal modification found at position 37 of tRNAs decoding ANN codons. However, the t^6A biosynthesis pathway is yet to be fully characterized. Indeed, early studies from the seventies suggested that ATP, threonine and carbonate are required and YrdC does not seem to bind threonine. In addition, based on initial biochemical experiments and more recent failed attempts to reconstitute the pathway *in vitro*, we predicted that YrdC is not the sole enzyme required for t^6A biosynthesis.The YgjD/Kae1/Qri7 family (COG0533) family is also on the top-ten list of universally conserved enzymes of unknown function. It has also recently been used to illustrate the difficult path of "converting data into knowledge and knowledge into understanding (quote of Sydney Brenner in [2]). It has been linked to DNA maintenance in bacteria and mitochondria and transcription regulation and telomere homeostasis in eukaryotes but its actual function has never been found. Based on a comparative genomic and structural analysis, we predicted this family was also involved in t^6A biosynthesis. This was confirmed as a yeast mutant lacking Kae1 is devoid of t^6A. Yeast t^6A^- strains were also used to reveal that that this modification has a critical role in initiation codon restriction to AUG and in restricting frameshifiting at tandem ANN codons. We also showed that YaeZ, a YgjD paralog, is required for YgjD function *in vivo* in bacteria. Indeed the *B. subtilis ygjD* cannot complement the essentiality of the *E. coli ygjD* in the absence of the *B. subtilis yeaZ* gene. This work lays the foundation for understanding the pleiotropic roles of both the YrdC and YgjD universal protein families.

1. Galperin, M. Y., Koonin, E. V. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Research* **2004,** 32, 5452-63.
2. Galperin, M. Y., Koonin, E. V. From complete genome sequence to 'complete' understanding? *Trends Biotechnol* **2010,** 28, 398-406.
3. El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williamson, J. R., Schimmel, P., Swairjo, M. A., de Crécy-Lagard, V. The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. *Nucleic Acids Research* **2009,** 37, 2894-909.

COMBREX – COMputational BRidges to EXperiments

Richard J. Roberts

New England Biolabs

Bacterial and archaeal genomes are being sequenced at an ever-increasing rate leading to vast numbers of putative genes of unknown function. Despite a large investment in DNA sequencing there has been no concomitant investment in understanding the function of these genes. Indeed the current state of annotation of genomes is so heavily dependent on similarity to previously annotated genes that errors are propagating at a high rate. COMBREX is a project that aims to increase the rate at which genomes are correctly annotated. The basic idea is to produce a database of computational predictions of gene function and then to invite biochemists to test those predictions. A key feature is to recruit biochemists with expertise in the area of the prediction so that the experimental work takes advantage of existing expertise, reagents and assays. Testing new predictions can then be carried out by students under the supervision of a senior lab member at minimal incremental cost. In the US small grants are available for this purpose and we are currently seeking funding possibilities in the UK and elsewhere in the world. The current status of this project will be discussed.

Identification of Annotation Errors and Incompleteness Through Network Analysis

Peter D. Karp

Bioinformatics Research Group, SRI International, Menlo Park CA

Global analysis of a genome and its associated metabolic network and transporters can identify errors and inconsistencies in a genome annotation. The following strategies will be presented. (1) Dead-end metabolite analysis identifies metabolites that are either only produced by the metabolic network and transporters, or only consumed by them, and can identify missing reactions and transporters, or incorrect reaction directionality. (2) A by-product of the prediction of metabolic pathways from genome information is the ability to identify pathway holes, or reactions within metabolic pathways that have no associated enzyme in the genome. (3) Reachability analysis of a metabolic network determines what end products are reachable through the network from a given set of nutrients. This form of analysis can be used to identify gaps in metabolic networks. (4) Pathways in the MetaCyc pathway database [1] are annotated with the taxonomic groups in which these pathways are expected to be found. Therefore, when MetaCyc pathways are recognized in sequenced genomes during computational pathway prediction, we can use that taxonomic information to flag pathways and enzymes that are outside their expected taxonomic range and may represent genome annotation errors. (5) Given a list of universally occurring bacterial genes, we can search bacterial genomes for homologs of those genes and flag those that are not present. The Pathway Tools software [2] implements methods 1-4.

[1] R. Caspi et al, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases," *Nucleic Acids Research* 2010 38:D473-9.

[2] P.D. Karp et al, "PathwayTools version 13.0: integrated software for pathway/genome informatics and systems biology," *Briefings in Bioinformatics* 2010 11:40-79.

Function and dysfunction in protein annotation

Steven E. Brenner

University of California, Berkeley

There have been significant strides in protein function prediction since the first genomes were sequenced a decade and a half ago. However, the pace of genome sequencing has introduced challenges that arguably outstrips these advances. An ever dwindling fraction of proteins have been experimentally characterized, and reliance on automated methods for function prediction remains a fraught enterprise. For several years after the first genomes were sequenced, function prediction methods were assessed largely on how many proteins they could annotate. However, comparing different groups' annotations of the same proteins revealed many conflicts, which implied alarmingly high error rates. This presentation will illustrate the challenge with several gold-standard curated families including the Nudix hydrolases for which own manual literature search quadruples the number of proteins with experimentally characterized function, many of which revealed apparent mistakes in the GOA database, and for which we also needed to GO with many additional molecular function terms. We will present the automated approach for phylogenetically-informed protein function prediction which we have been developed and address the broader outstanding challenges in the prediction of protein function.

Prokaryotic Annotation Status

Bruno Sobral

Virginia Tech, Blacksburg, VA

Annotation of prokaryotic genomes has changed dramatically with the decrease of costs of genomic sequencing and the increasingly decentralized and democratized manner in which prokaryotic genomes are sequenced. Data generation has become cheaper than data analysis, which is unprecedented in the real world. Because of the decentralization of data generation, which is likely to continue as technologies evolve further, data are increasingly unstructured, and adequately structuring data is both unaffordable and too time consuming to be practical. Furthermore, it is likely that a very significant portion of prokaryotic genomes are not made public in a timely manner. Although this situation is intractable, I believe we have a responsibility as a scientific community to tackle it.

"Annotation" comes from the Latin "annotatio" and the verb "annotare", which means to "add a mark". Generally, in genomic databases, this means to add explanations or comments to genomic sequences. These markings can be of various types, including, but not limited to, functional aspects of the genes/proteins known or predicted to be encoded by the sequences. In the current technological environment, prokaryotic annotation is typically done by a series of algorithms that markup the genome with respect to genes, proteins, RNAs and other such features. Any annotation of course is simply a hypothesis, whether the approach includes computational experiments, wet chemistry experiments, or some mixture of the two. Given this situation, one important feature that bioinformatics resources can provide is the ability to support multiple hypotheses with respect to annotation. One example where this is achieved is at the PaThosystems Resource Integration Center (PATRIC, <http://www.patricbrc.org>www.patricbrc.org), one of NIAID's Bioinformatics Resource Centers (BRCs, <http://www.pathogenportal.org>www.pathogenportal.org) that deals with prokaryotes from the perspective of infectious diseases.

In this talk, I will show where PATRIC stands with respect to annotation efforts for prokaryotes.

High Throughput Arrays for Protein Function

Manuel Ferrer

CSIC - Institute of Catalysis, Madrid, SPAIN

Certainly many of you have followed the controversial story about the recently reported metabolic array technology. In the reactome array, approximately 2000 metabolites, each linked to a quenched fluorescent compound, are covalently attached to the array surface. Upon interaction with an enzyme in a biological sample, the quenched fluorescent compound is liberated, providing a fluorescence signal for detection purposes. Bioinformatics analyses of array data provide a global overview of the metabolic network of cells and their communities - the "reactome" in real time, at the moment of sampling.

This article has instantly attracted the great interest and many controversial comments. Here, I provide an overview of the technology, explaining what is being debated and the most innovative aspects of this analytical tool with a special emphasis in giving insights into its use for protein function prediction.

Application of Enzymatic Assays for Experimental Annotation of Unknown Proteins

Alexander F. Yakunin

Banting and Best Department of Medical Research and the Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, M5G 1L6, Canada

Global genome and metagenome sequencing efforts have identified thousands of genes encoding unknown proteins which need to be functionally annotated. The Protein Structure Initiative (PSI) Centers have developed high-throughput methods of protein purification and crystallization with the aim of providing structural information about unknown proteins. Using the protein production pipeline of the Structural Proteomics Centre in Toronto (SPiT), we have established a set of more or less general enzymatic assays to screen purified unknown proteins for enzymatic activity. The assays have relaxed substrate specificity and are designed to identify the subclass or sub-subclasses of enzymes (phosphomonoesterase, phosphodiesterase, carboxylesterase). Proteins with identified activity are further characterized using the secondary screens with specific susbtrates (substrate profiling). We have demonstrated the feasibility and merits of this approach for hydrolases and oxidoreductases and have identified enzymatic activity in over 250 proteins. Substrate profiling has revealed that many enzymes exhibit considerable substrate promiscuity and show significant activity against several similar substrates. Crystal structures of many unknown proteins have been solved providing insight into potential catalytic activity or substrate. The efficiency of functional annotation of unknown proteins can be further improved by the combination of enzymatic screening with sequence analysis or other experimental approaches.

Discovery and Prediction of Functions in the Enolase Superfamily

John A. Gerlt

Departments of Biochemistry and Chemistry, University of Illinois, Urbana, IL 61801

Determining the functions of proteins encoded by sequenced genomes is a major challenge in biology. We are implementing an integrated sequence-structure-function strategy to facilitate functional assignment by predicting the substrate specificities of unknown proteins in the mechanistically diverse enolase superfamily. The reactions are initiated by abstraction of a proton from a carbon acid substrate to generate a Mg^{2+}-stabilized enolate intermediate. We are using three approaches: 1) operon context for unknowns encoded by bacterial genomes, 2) experimental screening of libraries of potential substrates, and 3) computational prediction by *in silico* docking of libraries of potential substrates to experimentally determined structures and homology models. This lecture will highlight functional assignments using these approaches. Our successes using computational prediction establish this approach as a viable strategy to facilitate functional assignment of unknown enzymes discovered in genome projects. Supported by 2R01GM071790 and 1U54GM093342.

Annotation Error in Public Databases

Patricia C. Babbitt
A.M. Schnoes, S.D. Brown, and I. Dodevski

University of California, San Francisco

As the number of protein sequences in public databases continues to expand, correct annotation of their molecular functions is important for their effective use in achieving new biological insight. Few recent studies have examined misannotation in depth in large public databases. We have addressed this issue using as a "gold standard" for evaluation a model set of 37 enzyme families from six functionally diverse enzyme superfamilies that have been well-characterized experimentally. Comparing the known molecular functions of these families to their annotations in the GenBank NR, UniProtKB/TrEMBL, UniProtKB/Swiss-Prot, and KEGG databases, we determined the levels and types of misannotation for each. Of these, only Swiss-Prot shows consistently low levels of misannotation, with the other databases showing highly variable levels of misannotation across these families that range from 0% - 90%. The levels of misannotation we found were generally consistent across the archival Genbank and TrEMBL databases and also across the secondary database KEGG. Despite significant improvement in the sophistication of computational protocols available for annotation transfer since the early days of genome sequencing, we also found for the GenBank NR database that misannotation has increased during the period 1993-2005, indicating that transfer of incorrect annotations to new sequences remains a significant problem. Most misannotations found in these families can be ascribed to over prediction of molecular function, i.e., annotation at a greater level of functional specificity than available evidence supports. This type of error may be especially prevalent in proteins from functionally diverse enzyme superfamilies in which the constituent families of each catalyze different reactions yet share some sequence, structural (including active site) and functional features.

Breaking the Power Law Curse: A Computational Bridge to Experiments

Simon Kasif

Boston University, Boston, MA

Basic publication statistics show that the allocation of research attention to genes is far from uniform, following instead a power law distribution. Most genes have few or no citations and few genes are associated with a large number of references. This holds for human genes as well as genes in microbial organisms. While this publication bias could reflect an essential biological or medical significance associated with some genes it could also suggest that research tends to be clustered in silos and there could be benefits to breaking the "power law curse" by broadening our understanding of biology. This direction is consistent with the recent emphasis on macro systems biology where we attempt to understand increasingly complex biological systems starting from their genomes by increasing the number of proteins and their specificity captured by protein-protein interaction or regulatory networks in a large number of biological contexts.
We review the recent computational advances and many remaining challenges in computational function prediction. We then describe the initial steps we have taken to build a community of computational researchers that will participate in populating the COMBREX database with predictions that will made available to experimental testing by participating biochemists. These efforts complement the parallel effort to build a synergistic experimental community.
We also describe some of the challenges we face to produce a novel bridge between computational predictions and their experimental validation. A single experiment is obviously not sufficient to produce a computational model of gene function for a large protein family and its validation. We discuss the COMBREX project as a first of a kind community implementation of Active Learning paradigm that aims to drive predictive sciences by cascades of experiments testing hypotheses that are expected to have the highest impact on the overall accuracy of predictive models used in the field. This approach is expected to change the current culture where experimental data generation is followed by analysis and model building resulting in models that are better at explaining the data instead of maximizing their predictive power.

Introduction to Breakout Groups, COMBREX Workshop

Iddo Friedberg

Miami University, Oxford OH.

COMBREX is a project that aims to productively engage a large community towards an ambitious goal: the systematic annotation of microbial genomes, utilizing computational and experimental means. When engaging such a large number of people, the usual culture of scientific collaboration changes. Thus, for the project to bear fruit, new ideas need to be developed on subjects such as data exchange and warehousing, intellectual credit and engagement of experimental and computational biologists, among others. The goal of the working groups is to address these issues and to provide ideas as to how a project of COMBREX's size can perform well. Four working groups will be set up to address these questions. The groups will discuss: annotation/database issues; high throughput / low throughput experimentation issues; computational predictions and COMBREX nuts & bolts. Each group will discuss topics within its purview, and will report on its recommendations in the evening session.

We will provide a more detailed agenda for each group at the conference itself. The groups are open to all participants, and in fact we urge each and every one of you to contribute and join one or more of the discussions. We believe that there is something of interest in these discussions for all attendees at the Microbial Genomics conference.

COMBREX Workshop Breakout Groups

Martin Steffen - Boston University, Iddo Friedberg - Miami University, Andrei Osterman - Sanford-Burnham Institute for Medical Research

COMBREX is a new effort that aims to increase the pace of experimental determination of the function of bacterial and archaeal genes. Central to this effort is the creation of a broad community of participants interacting via a web portal <http://combrex.bu.edu/> in order to coordinate efforts and achieve the efficiencies needed to meet the daunting challenge of determining gene function at rates commensurate with gene discovery.

Participants are able to contribute to the project in multiple ways:

1) Submit bids (small proposals) for funds to cover modest costs associated with the experimental validation of gene function.
2) Submit predictions of gene function that are then candidates for experimental validation.
3) Submit annotations regarding prior experimental annotations.

A small proportion, but still a large number, of genes are labeled with functions with evidence codes of "electronic annotation" or without proper reference to the published experimental validation. A major effort of COMBREX, in collaboration with NCBI, is to catalog precisely what is experimentally known about gene function.

4) Nominate genes for "high priority status."

These genes may require either experimental validation or high-quality predictions .This provides the opportunity for community members to make the case for a gene's scientific importance or the benefit to be derived from validation. We anticipate that COMBREX will actively try to encourage bids on genes identified as high priority targets.

The broad outlines of COMBREX have been established, primarily under the leadership of Rich Roberts and Simon Kasif. However, consistent with a community enterprise, there is much to be determined in the organization and operation of COMBREX, and the purpose of the workshop and breakout groups is to provide a forum for your input and feedback, and to encourage your active participation.

Topics to be discussed by the breakout groups include:

1. How to prioritize, assess and improve computational predictions?
2. How to evaluate experimental bids?
3. How to handle non-enzymatic proteins?
4. How best to handle predictions/phenotypes from high-throughput experimentation?
5. What is sufficient to be considered a valid demonstration of biochemical function?
6. How best to incorporate a controlled vocabulary describing function?
7. Identify innovative ways to extend the reach of the program both within the US and to the rest of the world.
8. How best to identify previously experimentally validated genes?

We look forward to your participation.

Rapid genome evolution and adaptation to salt selection

Jizhong Zhou
Aifen Zhou[1,2], Zhili He[1,2], Judy Wall[1,4], Terry Hazen[1,3], and Adam Arkin[1,4]

[1]Virtual Institute of Microbial Stress and Survival,
[2]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019,
[3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720,
[4]Biochemistry and Molecular Microbiology & Immunology Departments, University of Missouri, Columbia, MO 65211
[5]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

High salinity is one of the most common environmental stressors. The responses of microorganisms to high salinity in physiological, global transcriptional as well as the accumulation of osmoprotectants have been studied, however, little is known about the molecular basis underlying the salt adaptation, the dynamics of the salt adaptation and how such adaptation affect the fitness. Here we used experimental evolution of *Desulfovibrio vulgaris*, a model sulfate-reducing bacterium, to salt stress to explore how natural selection modify the genomic structure, changes of the intracellular metabolites and the fitness when they evolve in an osmotically stressful environment. Significant improvement in salt specific fitness (250 mM NaCl) was observed after 300 generations of evolution. And stable salt resistance phenotype was observed at 1,000 generations. Whole genome sequencing of the ancestral and evolved strains under both controlled and salt stress condition revealed salt-specific mutations and deletions. Genetic evidences indicated that almost of all salt-specific mutations and deletions were beneficial. Glutamate, glutamine, alanine, aspartate, valine and threonine accumulated in evolved populations, suggesting that these amino acids maybe used as osmoprotectants in *D. vulgaris*. These results imply that biological systems can rapidly adapt to changes in environmental factors such as salt.

Genomics of bacterial lipid metabolism

Athanasios Lykidis

DOE-Joint Genome Institute, Walnut Creek, CA

Lipid metabolism is a major cellular system that generates a wide spectrum of chemical structures critical for cell survival and growth. Recent work in bacteria, algae and yeast have focused on modifying lipid and fatty acid biosynthesis as a prime target for metabolic engineering and optimization for the production of next generation biofuels. To better understand the regulation of fatty acid metabolism and lipid accumulation we study the distribution and abundance of the relevant genes and pathways in the available genomic data. The emerging picture is a dynamic system of genes and regulatory pathways revealing that, although the structure of lipid biosynthetic pathways is conserved among organisms, there is significant diversity in the corresponding regulatory mechanisms. We will also present data on the identification and characterization of novel enzymatic functions relevant to biofuel production.

Phylogenomic Inference for Functional Annotation of a Bacterial Genome: *Helicobacter pylori*

Kimmen Sjölander
R.S. Datta(1), U. Pieper(2), and A. Sali(2)

(1) University of California, Berkeley, and (2) University of California, San Francisco

Functional annotation based on annotation transfer is known to be prone to systematic error, with error rates estimated at 20% and higher. Using evolutionary reconstruction of gene families and structure information (or prediction) -- a process we call "structural phylogenomics" -- improves functional annotation accuracy.We are constructing a phylogenomic database with predicted functions and structures for genes in the genome H. pylori strain 26695, clustered into families, subfamilies and orthology groups, including homologs from hundreds of other species. Using this database, we will predict functions of H. pylori genes based on an ortholog with experimentally supported GO process and/or molecular function, or based on an ortholog with experimentally supported EC number (according to BRENDA). Where possible we will use these predictions to map H. pylori genes to KEGG pathways. We are also building comparative models for H. pylori proteins, in the cases where a suitable template structure exists. Lastly, we will use the INTREPID and Discern algorithms to predict enzyme active sites and other functionally important positions. We will supply the provenance of each of our predictions so that users can weigh them with an appropriate level of confidence. Each predicted annotation will include the set of orthologs supporting that annotation, the threshold at which the orthology prediction was made (lower thresholds indicate higher precision), and the set of PHOGs containing these orthologs. Each PHOG will be linked to the multiple sequence alignment and tree on which these predictions are based. These data and supporting evidence are part of the PhyloFacts Phylogenomic Encyclopedia of Microbial Gene Families, supported by the NSF and DOE, at http://phylogenomics.berkeley.edu/phylofacts/. We will submit these annotations for validation by COMBREX, the COMputational BRidge to EXperiments.

POSTER ABSTRACTS

Estimating linkage among short metagenomic read fragments with Bayesian phylogenetics

A.E. Darling* & J.A. Eisen

University of California, Davis

Next-generation sequencing can provide immense metagenomic datasets, but interpreting the short sequence fragments has until now remained difficult. The depth of sequence coverage provided by even a modest Illumina or SOLiD run implies that, for all but the most diverse environments, many sequence reads will come from the same species. Indeed, if a small number of dominant species exist, those species may be sequenced at high enough coverage to assemble their consensus genomes. However, many organisms in the community will not be sequenced deeply enough for assembly.

We describe a new method to identify short sequence reads that come from the same species, even when coverage of that species is so low that assembly would be impossible. The method uses a Bayesian phylogenetic model to directly estimate the phylogenetic position of reads relative to reference organisms. The key feature of our model is that reads are grouped into some unknown number of *linkage groups* (putative species), and each *linkage group* attaches to a single point on the phylogeny. Thus, reads in a linkage group share phylogenetic information with each other. A prototype MCMC sampler for this model has been implemented and we describe preliminary results on its application to simulated datasets.

Most previous phylogenetic binning methods including AMPHORA, MEGAN and PhymmBL, assign each read to a phylogenetic position independently. Our method differs by allowing reads to share information during the phylogenetic inference. Whereas binning reads independently makes sense in the case of a small number of long reads, short reads may not have enough phylogenetic information to allow accurate placement. Finally, we note that the training stage of some composition-based methods such as PhyloPythia may indirectly allow reads to share information, at least during the training phase.

Phylogenomic Inference for Functional Annotation of a Bacterial Genome: *Helicobacter pylori*

R.S. Datta*(1), U. Pieper(2), A. Sali(2), and K. Sjölander(1)

(1) University of California, Berkeley, and (2) University of California, San Francisco

Functional annotation based on annotation transfer is known to be prone to systematic error, with error rates estimated at 20% and higher. Using evolutionary reconstruction of gene families and structure information (or prediction) -- a process we call "structural phylogenomics" -- improves functional annotation accuracy.We are constructing a phylogenomic database with predicted functions and structures for genes in the genome H. pylori strain 26695, clustered into families, subfamilies and orthology groups, including homologs from hundreds of other species. Using this database, we will predict functions of H. pylori genes based on an ortholog with experimentally supported GO process and/or molecular function, or based on an ortholog with experimentally supported EC number (according to BRENDA). Where possible we will use these predictions to map H. pylori genes to KEGG pathways. We are also building comparative models for H. pylori proteins, in the cases where a suitable template structure exists. Lastly, we will use the INTREPID and Discern algorithms to predict enzyme active sites and other functionally important positions. We will supply the provenance of each of our predictions so that users can weigh them with an appropriate level of confidence. Each predicted annotation will include the set of orthologs supporting that annotation, the threshold at which the orthology prediction was made (lower thresholds indicate higher precision), and the set of PHOGs containing these orthologs. Each PHOG will be linked to the multiple sequence alignment and tree on which these predictions are based. These data and supporting evidence are part of the PhyloFacts Phylogenomic Encyclopedia of Microbial Gene Families, supported by the NSF and DOE, at http://phylogenomics.berkeley.edu/phylofacts/. We will submit these annotations for validation by COMBREX, the COMputational BRidge to EXperiments.

The Adaptive Narrative of Isogenic Confinement: Experimental Evolution of *Legionella pneumophila*

A.W. Ensminger*[2], Y. Yassin[3], A. Miron[3], and R.R. Isberg[1,2]

[1]Howard Hughes Medical Institute, [2]Tufts University School of Medicine, and [3]Dana Farber Cancer Institute, Boston, MA

We have established a broadly applicable experimental evolution system to study *Legionella pneumophila*, an accidental pathogen of human hosts. These facultative intracellular bacteria were restricted to growth within primary mouse macrophages for hundreds of generations. During passage, bacterial adaptations resulted in hyper-replication within the mammalian host cells. In most instances, these adaptations occur at the expense of replication within amoebae, the natural host. Next generation whole-genome resequencing of these bacteria identified several adaptive mutations that recapitulate these phenotypes when reintroduced into a progenitor strain. High-throughput genotyping technologies were used to measure the frequency of each mutation during the course of adaptation. In addition to providing an unprecedented glimpse into evolutionary dynamics, these studies provide unique insight into the intracellular lifestyle of *Legionella*, its nutritional requirements, and adaptive potential. Many of the adaptive mutations reside in pathways central to the recognition of *Legionella* by the innate immune system. This raises the intriguing possibility that these adaptations serve to modify the composition of pathogen-associated molecular patterns presented during infection.

Critical Assessment of Function Annotations: a community effort to assess computational protein function prediction programs

Iddo Friedberg(1)*, Michal Linial(2), Predrag Radivojac(3)

(1) Dpts of Microbiology and Computer Science & Software Engineering, Miami University, Oxford OH USA
(2) Dpt of Biological Chemistry, the Hebrew University of Jerusalem, Israel
(3) Dpt of Informatics & Computing, University of Indiana, Bloomington IA, USA

How correct are computational gene and gene product functional predictions? To date, few and limited efforts have been made to assess the veracity of protein function predictions. In contrast, the protein structure prediction community has been running a community-driven effort termed CASP (Critical Assessment of Structure Prediction) for over a decade. This effort has dramatically changed the methodology of structural predictions, as many methods were elevated (or conversely, culled), following CASP's assessment of their accuracy. We will assess the quality of current computational techniques in an unbiased manner via a set of protocols we term CAFA: Critical Assessment of (protein) Functional Annotations. We call upon the community of computational function predictors to submit their predictions of several thousand gene targets we will provide. Looking at historical trends in the genomic databases, we expect that over the period of time CAFA takes, some of the targets shall be annotated experimentally. We will then compare the predictions to the annotations. We shall use a set of commonly accepted metrics such as the Lord GO distance measure, Jaccard's coefficient, etc. We show here our pilot results in annotations, as well as an elaboration of the annotation assessment techniques we will employ. We will publicize the results at the Automated Function Prediction (AFP) meeting to be held as an ISMB SIG at the Summer of 2011. CAFA results will be a part of the broad goals of the meeting, which is to provide an opportunity for scientists working on protein function from both experimental and computational perspectives to discuss progress and pressing issues in functional annotation. We intend to raise questions and stimulate the exchange of ideas/experience regarding functional ontologies, experiments leading to functional annotations, biases or limitations introduced in the process, methodologies for predicting function, and effectiveness of these annotation algorithms.

Nitrogen Cost Minimization in Proteomes of Open Ocean Microorganisms

Joseph J. Grzymski and Alex M. Dussaq

Desert Research Institute, Division of Earth and Ecosystem Sciences Reno, Nevada

Biosynthetic cost minimization in microorganisms is a reduction in amino acid biosynthetic costs to maximize central metabolic costs and is thought to be a necessary, adaptive strategy under resource limitation in natural environments. Thus, microorganisms generally first use amino acids with lower molecular weights and lower assimilation costs especially in less critical protein domains. The need to thrive under low levels of organic and inorganic nitrogen is imprinted on the genomes of open-ocean microorganisms in a phenomenon named nitrogen cost minimization. Here we analyze over 20 million predicted proteins and protein fragments and 20 whole genomes to show that amino acid sequences from the open ocean are reduced in nitrogen side chains but increased in mass in comparison to coastal ocean microorganisms despite added energetic requirements resulting from an increase in average amino acid mass. These findings suggest that nitrogen limitation in the open ocean is a stronger selective force than biosynthetic costs based on amino acid mass.

Metagenomic Comparisons of Toxic *MICROCYSTIS* Blooms

Zhou Li[1], Loren J. Hauser[1,2], Morgan M. Steffen[3], Mike Leuze[2], Brian Dill[4], Nathan VerBerkmoes[1,3,4], Steven W. Wilhelm[1,3]

1) Genome Science and Technology Graduate School, University of Tennessee
2) Biosciences Division, Oak Ridge National Laboratory
3) Department of Microbiology, University of Tennessee
4) Chemical Sciences Division, Oak Ridge National Laboratory

The *Microcystis spp.* are cyanobacteria that form massive hepatotoxic blooms in freshwater lakes worldwide. To better understand the factors that cause *Microcystis* to bloom, samples were collected from blooms in Lake Erie (near Erie, Pa) and Lake *Taihu*, (near Wuxi, People's Republic of China). Lake Erie blooms were collected in August from part of the lake with mild urban and agricultural (wheat, corn, grapes) runoff, while the Lake Taihu blooms were collected in May from an area with heavier urban and agricultural (chicken farm) runoff. After mechanical enrichment for *Microcystis* cells, metagenomic analyses revealed that ~40% of the reads from the Lake Erie bloom mapped to the *Microcystis aeruginosa* NIES-843 genome, providing 12x coverage of 87% of the genome with approximately 99.7% identity. Only ~12% of the reads from the Lake Taihu bloom (which was substantially denser than the Lake Erie bloom) mapped to the *Microcystis* genome giving 4.5x coverage of 75% of the genome. Community analyses suggest that function was generally conserved and the *Microcystis* blooms were accompanied by a core of co-occurring heterotrophic bacteria, although site-specific members existed for each microbial community. Comparative sequence analyses of the reads that mapped to *Microcystis aeruginosa* NIES-843 and the reference genome identified potential loss of function from gene truncations resulting from frame shifts or non-synonymous SNPs in conserved domains; changes in regulation due to mutations in promoters, including transcription factor binding sites and ribosomal binding sites; and changes in expression levels influenced by variations in mRNA secondary structure resulting from synonymous SNPs. The metagenomic databases as well as proteomics data from *Microcystis aeruginosa* NIES-843 cultures have set the stage for shotgun proteomic analyses of bloom samples, increasing the potential for understanding why blooms occur at an unprecedented level of cellular physiology and biochemistry.

Diversity and Virulence Determinants of Clinical Multidrug-Resistant *Streptococcus pneumoniae*

N. Luisa Hiller*[1], Rory A. Eutsey[1], Evan Powell[1], Benjamin Janto[1], Azad Ahmed[1], Josh Earl[1], Jennifer Bennett[1], Mark Longwell[1], Meg Dahlgren[1], Darren Martin[2], Suzanne Dawid[3], Hervé Tettelin[4], Alexander Tomasz[5], Herminia deLencastre[5], Todd Hillman[1], Farrel Buchinsky[1], J. Christopher Post[1, 6], Garth D. Ehrlich[1, 6], Fen Z. Hu[1, 6]

1. Center for Genomic Sciences, Pittsburgh, USA. 2. University of Cape Town, Cape Town, South Africa. 3. University of Michigan, Ann Arbor, USA. 4. University of Maryland School of Medicine, Baltimore, USA. 5. Rockefeller University, New York, USA. 6. Drexel University College of Medicine, Pittsburgh, USA.

The bacterium *Streptococcus pneumoniae* is a major human pathogen that shows very extensive diversity with respect to gene possession among independent clinical isolates. However, the vast majority of penicillin-resistant strains belong to one of several highly-related clades that collectively display decreased genomic diversity when compared with the penicillin-sensitive strains using traditional molecular typing methods for strain discrimination. To investigate the level of conservation within one of these clonal lineages, the Spanish_USA clone or SP1, we have performed comparative genomics using the 454 Life Sciences -generated whole genome sequence of four SP1 strains isolated many years apart on different continents. Two strains, isolated twelve years apart, one in a hospital in Spain and the other in a day-care center in Portugal, are almost identical. This observation suggests that the SP1 genome can remain very stable within the population over a long period of time. In contrast, comparison with another two strains isolated in the US, revealed 18 regions on the chromosome that are variable among these three genotypes. These regions differ by >3500 single nucleotide polymorphisms and 96 differences in the possession/absence of genes. These differences are concentrated in regions coding for the capsule, an antibiotic resistance marker, a bacteriophage, and multiple cell wall proteins. Collectively they demonstrate greater variability than had been appreciated using molecular typing methods. Furthermore, animal models and epidemiological data show that there is extensive phenotypic heterogeneity among the SP1, where strains range from avirulent to highly pathogenic. Animal model experiments using mutants where the capsule has been switched between strains, have demonstrated that the type 3 capsule is a virulence determinant within this lineage. Future experiments are focused on the identification of additional SP1 pathogenic factors.

Large-scale protein purification and structure determination in service of microbial genomics

Samuel K. Handelman, Nicholson Price, Farhad Forouhar, Jon D. Luff, Nelson Tong, John K. Everett, Thomas Acton, Rong Xiao, Barry L. Wanner, Gaetano T. Montelione, and John F. Hunt

Northeast Structural Genomics Consortium, PrFEcT Consortium, Columbia University, University of Medicine and Dentistry of the State of New Jersey, Rutgers University, and Purdue University

The NIH launched the Protein Structure Initiative (PSI) in year 2000 to develop high-throughput protein purification and structure-determination capabilities to complement genome sequencing projects. Since then, the PSI has produced >4,000 crystal and NMR structures, almost entirely of bacterial proteins. While a limited number of these structures have had high impact in specific areas of biological research, most target proteins were selected based on sequence uniqueness without regard to biological interest or phylogenetic prevalence. However, the principles governing PSI target selection have recently changed to promote collaborative target selection, emphasizing the interests of practicing biologists. The Northeast Structural Genomics Consortium (NESG -- www.nesg.org), one of four large-scale structure-determination groups supported by the PSI, is seeking collaborators interested in obtaining high resolution structures of bacterial proteins. Projects can involve single targets or large sets of target families. The foundation of the PSI is the ability to purify dozens of proteins in parallel with high economic efficiency. Purified samples can be delivered to collaborating labs with complete biophysical characterization (stored online in an expression, purification, and structure-determination database). Antibody generation is also possible. NESG's experience expressing >10,000 microbial proteins has provided quantitative insight into the sequence properties influencing protein expression, solubility, and crystallization, including unexpected observations concerning the influence of codon usage on protein expression in E. coli. To support collaborations with bacteriologists and genomicists, NESG has developed a number of publicly available web resources. One is the CRSH (Clusters of Reciprocal Sequence Homologs) database of microbial proteins of likely equivalent function. The groupings in this database are superior to other online resources in their ability to predict physical and transcriptional associations, and the CRSH website serves related high-throughput genomics data along with information on the availability of purified proteins via the PSI. Another resource analyzes an entire protein family to identify members with the highest probability of yielding soluble protein and crystal/NMR structures. Finally, NESG established collaborations with experts in transcriptomic, proteomic, metabolomic, enzymologic, epistasic, and interprotein interaction profiling to generate and archive related high-throughput functional data in conjunction with structure-determination projects. The resources of this PRotein Function EluCidation Team (PrFEcT) will be described. Examples involving cyclic diguanylate signaling and toxin-antitoxin systems will be used to illustrate the kinds of biological insights that can be provided by high-throughput protein structure determination.

Exploring the Global Genomic Landscape of Enteric Pathogens: Genes, INDELs, SNPs-n-Chips

Scott A. Jackson*, Isha R. Patel, Mark K. Mammel, Joseph E. LeClerc

FDA-CFSAN, Laurel, M.D.

Escherichia coli, Salmonellae, Shigella spp, and *Vibrio cholera* are recognized as four of the major food-borne bacterial pathogens in the U.S., causing a combined total of over 3 million cases of illness annually. Recent outbreaks of such illnesses suggest a rapid capacity for these pathogens to adapt to novel environmental niches, including new food sources. From a genomic perspective, horizontal gene transfer among and between these pathotypes may account for a majority of such adaptation potential and has resulted in a mosaic genomic milieu. Indeed, the unique gene content of individual strains in concert with regulation-based adaptation, derived probably at the SNP-level, should dictate the overall character of a strain including virulence properties. To investigate the genomic diversity and evolutionary history of this class of pathogens, we utilized a comprehensive DNA microarray to characterize a broad and diverse collection of food borne isolates. Our Affymetrix microarray contains over 2.5 million features and represents over 85 whole genome sequences, 90 plasmids, and all known antimicrobial resistance and virulence genes from various public and private repositories. Gene prediction and subsequent comparative analysis revealed >545k total genes. Of these, 83k were found to be unique and were represented on our array. The novelty of this design necessitated non-traditional data analysis methods that were developed and validated in-house and utilize statistical approaches for accurate genotype calls. Using this array, genomic DNA from over 350 temporally and geographically diverse isolates of *Salmonella and E. coli* pathotypes, and *Vibrio spp* have been interrogated to date. Strains were chosen based on their importance to public health, inclusion in historical collections, and other molecular subtyping discrimination. Our cumulative findings have provided an accurate and insightful assessment of the true genomic diversity that exists within natural populations.

An FDA Bioinformatics Tool, ArrayTrack™, for Molecular Characterization of Microbial Food Borne Pathogens Using Microarrays

Joshua Xu[1], Scott Jackson[2*], Isha Patel[2], Jonathan G. Frye[3], Wen Zou[1], James Chen[1], Rajesh Nayak[1], Steven Foley[1], Carl Cerniglia[1], ArrayTrack™ Team[1], Hong Fang[1] and Weida Tong[1]

1 National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR
2 Center for Food Safety and Applied Nutrition, Food and Drug Administration, Laurel, MD
3 Agricultural Research Service, US Department of Agriculture, Athens, GA

Recent advances in microbial genomics and bioinformatics data are offering greater insights in the emergence and spread of foodborne pathogens in outbreak scenarios. The FDA has developed a genomics tool, ArrayTrack™, which provides a rich set of functionality to manage, analyze, and interpret genomic data for mammalian organisms. It has been widely adopted by the research community and used for the pharmacogenomics data review in the FDA Voluntary Genomics Data Submission program. ArrayTrack™ has been extended to support microbial genomics research using microarrays. It has been populated with bioinformatics data from the public domain related to bacterial pathogen species. Data processing and visualization tools have been enhanced with customized options to facilitate analysis of genetic profiling microarray data. These functions are particularly relevant and effective for the identification and characterization of bacterial pathogens through microarray genetic profiling data. We have demonstrated the new functionality of ArrayTrack™ on FDA-ECSG, a custom Affymetrix microarray biochip developed by FDA scientists for identifying and discriminating between different strains of *E. coli* and *Shigella*. The array consists of 23K genes from 32 diverse *E. coli* and *Shigella* whole genome sequences, and 46 related plasmid sequences. This biochip was used to investigate the genetic characteristics of *E. coli* O157:H7 outbreak strains linked to spinach and lettuce and provided genotypic discrimination between and within closely related species. The pathogen-specific bioinformatics functionality in ArrayTrack™ allows researchers to compare the genomics profile of the outbreak strain with those of known reference strains. ArrayTrack™ is also being used to analyze data from an antimicrobial resistance gene microarray developed by USDA scientists. This array has over 1,200 probes to detect genes commonly associated with antimicrobial resistance. Analysis of data from this biochip using ArrayTrack™ enables the identification of genetic elements responsible for the spread of multidrug resistance in bacterial pathogens. Application of ArrayTrack™ to different biochip platforms demonstrates its flexibility to analyze a variety of data. This enhanced ArrayTrack™ tool will provide vital support for the FDA's Food Protection Plan by improving the capabilities to rapidly identify the bacteria and their genetic traits (i.e. antimicrobial resistance, virulence, etc.) in outbreak investigations. ArrayTrack™ is freely available to the public at http://www.fda.gov/Arraytrack.

Hologenomic Analysis of the Marine Tunicate *Ecteinascidia turbinata*: Identification and Characterization of the ET-743 (trabectedin) Biosynthetic Pathway

B. Janto*, C. Rath, J. Earl, M. Dahlgren, A. Ahmed, R. Kreft, N.L. Hiller, F. Yu, R. Williams, K. Hakansson, F. Hu, G. Ehrlich, D. Sherman

Center for Genomic Sciences, Allegheny Singer Research Institute

ET-743 is an approved chemotherapeutic obtained in low abundance (1^{-4}-1^{-5}% w/w) from the tunicate *Ecteinascidia turbinata*, and manufactured for clinical use by a lengthy semi-synthetic process. We hypothesized that ET-743 is the product of a marine bacterial symbiont due to structural similarity with three previously characterized bacterial natural products. Furthermore, we expected that the ET-743 biosynthesis genes would have sequence homology to these reference biosynthetic pathways. We used a metagenomic sequencing strategy to query the tunicate hologenome (DNA encoded by both the host, *E. turbinata*, and associated microorganisms) to identify and characterize the biosynthetic pathway for ET-743. Metagenomic 16S rRNA targeted gene sequencing allowed analysis of the total population of microbial symbionts providing an important comparative analysis with the Mediterranean and Caribbean tunicate populations that produce ET-743. From the bulk sequencing data we detected and extended a 35 kb contig in which we identify 26 biosynthetic genes that represent a large part of the predicted ET-743 biosynthetic pathway. Based on codon usage and phylogenetic sequence analysis between this contig and a rRNA-containing contig, we propose that *Candidatus Endoecteinascidia frumentensis* is the producing bacterium. This work provides a foundation for accessing the drug and novel analogs through heterologous expression.

Pathway Tools: Comprehensive Pathway/Genome Informatics

Tomer Altman, Fred Gilham, Pallavi Kaipa, Markus Krummenacker, Mario Latendresse, Suzanne Paley, and Peter D. Karp

Bioinformatics Research Group, SRI International, Menlo Park CA USA.

Pathway Tools [1] provides an array of capabilities for analyzing and navigating integrated collections of genome, metabolic, and regulatory information. The software addresses several use cases in bioinformatics and systems biology. It predicts metabolic pathways, pathway hole fillers, and operons in annotated genomes. It supports querying, visualization, and web publishing of Pathway/Genome Databases (PGDBs). It includes visual tools for analysis of omics datasets. It provides tools for analysis of biological networks, for comparative analyses of PGDBs, and for metabolic engineering.

The software can be installed locally to create PGDBs for organisms of interest to an organization. Pathway Tools constructs a metabolic model of an organism from its annotated genome by recognizing known pathways from the MetaCyc database of 1,500 experimentally elucidated metabolic pathways. MetaCyc contains more than twice as much pathway data as does KEGG, and thus can recognize a broader set of pathways in genome and metagenome data.

Three omics viewers support projection of large datasets onto cellular network diagrams that provide complementary whole-cell perspectives for interpreting omics data. The first omics viewer paints omics datasets onto a diagram of metabolic pathways. The second omics viewer paints onto a diagram of the transcriptional regulatory network. The third omics viewer paints omics datasets onto the genome. Several enrichment analysis tools have recently been added.

The Pathway Tools genome browser can be applied either to a single genome, or can be applied to multiple genomes in a comparative mode; a variety of other comparative tools exist. Systems biology analyses available include detection of dead-end metabolites, and reachability analysis of metabolic networks to determine what product compounds can be synthesized under a hypothetical growth medium.

A set of interactive editing tools allow scientists to update the contents of a PGDB, including a new editor for signaling pathways.

[1] P.D. Karp et al, "PathwayTools version 13.0: integrated software for pathway/genome informatics and systems biology," Briefings in Bioinformatics 2010 11:40-79.

Genome Sequence Analysis of *Methylobacterium oryzae* CBMB20^T, a Plant Growth-Promoting Methylotroph in the Phylloshpere

Min Jung Kwak[1,2]*, Haeyoung Jeong[1], Byung Kwon Kim[1], Ju Yeon Song[1], Dong Su Yu[1], Tong-Min Sa[3], Tae Kwang Oh[1,4], and Jihyun F. Kim[1,2]

[1]Industrial Biotechnology and Bioenergy Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, Korea; [2]Biosystems and Bioengineering Program, University of Science and Technology (UST), Daejeon 305-333, Korea; [3]Department of Agricultural Chemistry, Chungbuk National University, Cheongju 361-763, Korea; [4]21C Frontier Microbial Genomics and Applications Center, KRIBB, Daejeon 305-806, Korea.

Methylobacteria are pink-pigmented, facultative methylotrophs that are widespread in nature. Several species of *Methylobacterium* have been isolated from plants and demonstrated to produce substances that promote plant growth. We performed the genome project of *Methylobacterium oryzae* CBMB20^T that was isolated from rice stem and promotes plant growth. The complete genome sequence was determined by the whole-genome shotgun strategy that involves 454 pyrosequencing and fosmid paired-end sequencing. The *M. oryzae* genome consists of one circular chromosome of 6,286,629 bp and four plasmids of 14,537~156,479 bp. 6,338 coding sequences were identified and annotated. The genome has most of the genes related to central metabolism, and genes required for methylotrophy are located in methylotrophy islands. *M. oryzae* genes responsible for promotion of plant growth include those that encode enzymes for biosynthesis of auxin or cytokine, pyrroloquinoline quinine synthesis protein, 1-aminocyclopropane-1-carboxylate deaminase, and organic or inorganic phosphate-solubilizing proteins. *M. oryzae* has a number of genes for adaptation to the surface of plants. In conclusion, we confirm the genes involved in methylotrophy and plant growth promotion, and provide clues to widespread distribution of *Methylobacterium* in diverse environments including plants. [Financial support from the 21C Frontier Microbial Genomics and Applications Center program, MEST, Korea]

Genome Sequence Analysis of the Dual Proteorhodopsin-Containing Marine Microbe *Donghaeana dokdonensis* DSW-6^T

Soon-Kyeong Kwon[1,2]*, Haeyoung Jeong[1], Jung-Hoon Yoon[1], Tae-Kwang Oh[1], and Jihyun F. Kim[1,2]

[1]Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, Korea; [2]University of Science and Technology (UST), Daejeon 305-333, Korea

A warm branch of the Kuroshio Current and a cold extension of the Liman Current meeting to get mixed and merged, the East Sea of Korea (Donghae in Korean) around Ulleungdo and Dokdo islands incubates a wealth of marine lives of both macroscopic and microscopic levels. Isolated from the sea water sampled off the coast of Dokdo, *Donghaeana dokdonensis* (homotypic synonym of *Persicivirga dokdonensis*) DSW-6 belongs to the *Cytophaga-Flavobacterium-Bacteroides* group which constitutes high percentage of marine isolates. Genome sequence analysis of *D. dokdonensis* reveals that this bacterium harbors two functional bacterial rhodopsin genes similar to the proteorhodopsin genes of various bacteria and archaea of marine origin. Proteorhodopsins convert light energy to chemical energy via the light-activated proton pumping that generates the proton-motive force across the membrane as a result. A gene cluster involved in biosynthesis of a carotenoid as well as the *blh* gene, the product of which is responsible for oxidative cleavage of the carotenoid into two retinal molecules was found. Retinal binding to rhodopsin is required for its functioning. We also present the results of a comparative genomic analysis of orthologous genes among the whole genomes in the order *Flavobacteriales*. This information provides insights into a better understanding on the evolution and environmental roles of the flavobacterial group, whose members are associated with diverse habitats and lifestyles. [Financial support from the 21C Frontier Microbial Genomics and Applications Center program, MEST, Korea]

Characterizing Protein Families of Unknown Function

M.G.I Langille* & J.A. Eisen

University of California, Davis.

Perhaps one of the most frustrating aspects of genomic and metagenomic analysis is that functional predictions for many genes cannot be identified. These "hypothetical" or "unknown" genes represent a significant fraction of the genes in most genomes or metagenomes, and this fraction will likely increase as sequencing technology continues to outpace functionally informative lab experiments. To start tackling this situation we characterized and ranked protein families with unknown function from completed genome sequences. Ranking of these families were done using several metrics such as quantity of members, presence across tree of life, presence in mostly pathogens or other habitats, etc. This ranking allows particular families of unknown function to be targeted for more in-depth analysis due to their ubiquitous nature or their role in a particular niche. In addition to ranking these families, we analyze their abundance profiles across several metagenomic studies and cluster them with families of known function in the hopes of making novel functional predictions.

Genome Invasion and Exchange: Mobile Genetic Elements in Microbial Genomes

Ming-Chun Lee*, David Robinson and Christopher Marx

Harvard University.

The increasing number of sequenced microbial genomes provides a great opportunity to understand microbial genome architecture and microbial evolution. Such analyses rely heavily on accurate annotation of genomic features, including mobile genetic elements (MGEs). Due to their movement within and across genomes, MGEs not only represent an important source of genetic variation within genomes, but also mediate horizontal gene transfer (HGT) among organisms. MGEs have been frequently found in various organisms but are often poorly annotated. To investigate the evolution of MGEs and their significance in microbial evolution, we developed an annotation program, OASIS (Optimized Annotation System for Insertion Sequences), designed specifically for the major group of MGEs in microbial genomes- insertion sequences (ISs). OASIS was applied to 1,012 completely sequenced bacterial and archeal genomes available from NCBI. The analysis shows a widespread IS distribution across all genomes, completely overturning the previous view of the constrained phylogenetic breadth of IS invasion. Furthermore, extensive HGT events were found among all groups of bacteria and archea by using both reconciliation methods and GC content with codon usage. This unusual frequency of HGT events is orders of magnitude higher than any other HGT genes that have been reported, indicating frequent HGT might be a major cause of IS prevalence. Ongoing work is estimating the time of these HGT events and the birth of each IS family. Overall, this study reveals the general features of IS invasion and provides a way of understanding the complicated co-evolutionary history of ISs and their hosts.

Discovery of non-coding elements toxic to bacteria within genomic "death valleys"

Asaf Levy*,† Aya Kimelman†, and Rotem Sorek

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel
† Equal contribution.

In the process of shotgun genome sequencing, initial assemblies usually contain gaps due to DNA fragments that cannot be successfully propagated in bacteria. We have previously discovered that many of these uncloneable gaps are caused by genes that are toxic to *E. coli*. When these genes are cloned into *E. coli*, their expression product kills the cell, and hence the inability to clone and sequence them.

We have computationally scanned 260 microbial genomes and detected over 5,000 cloning gaps harboring toxic genes. Surprisingly, these gaps included ~400 toxic sequences lying in intergenic regions. Further analysis showed that some of these regions harbor small toxic non-coding RNAs (sRNAs) that inhibit *E. coli* growth. Experimental exploration of selected novel toxic sRNAs validated their toxicity to *E. coli*. Among the set of toxic sequences, we also discovered DNA binding sites of toxic proteins. This large-scale identification of natural non-coding elements that inhibit the growth of bacteria may help in the future to develop novel antibiotics.

Biocontrol Genomic Islands in *Pseudomonas* spp. and Functional Genomic Study of a Biocontrol Pseudomonad Model under Iron Starvation Conditions

Chee Kent Lim*, Karl Hassan[1], Sasha Tetu[1], Joyce Loper[2] & Ian Paulsen[1]

[1]Department of Chemistry and Biomolecular Sciences, Macquarie University, NSW, Australia
[2]USDA-ARS Horticultural Crops Research Laboratory, Corvallis, OR, USA

Pseudomonas spp. are ubiquitious bacteria found in nature, including soil. We are interested in biocontrol pseudomonads, which during commensal associations with plants, can suppress plant pathogens, such as fungi and bacteria through the production and secretion of secondary metabolites and exoenzymes that are antagonistic to the plant pathogens. Recently, we have sequenced the genomes of seven biocontrol *Pseudomonas* spp. and have employed functional genomic techniques to investigate conditions promoting the expression of secondary metabolite biosynthesis genes.

Comparative analysis of recently sequenced genomes was performed using multiway blastp searches. These analyses, in combination with the results of atypical trinucleotide scans and repeat element searches, revealed many genomic islands with putative roles in biocontrol. For example, genomic islands with putative functions in the production of insecticidal toxin complexes, antibiotics, non-ribosomal peptides and polyketides were identified within the newly sequenced genomes.

In addition to genome sequence analysis, functional genomic studies have been performed on a model biocontrol pseudomonad, *P. fluorescens* Pf-5, using a whole genome microarray. These studies aimed to identify environmental conditions, such as micronutrient availability, that influence the expression of biocontrol traits. For example, under iron-limiting conditions we found that 180 genes were up-regulated by at least 2-fold while 121 genes were down-regulated. Prominent among the genes upregulated were several gene clusters involved in the production of two iron-chelating siderophores. These siderophores are likely to play an indirect role in biocontrol by allowing Pf-5 to out-compete potential plant pathogens for available iron resources within shared plant-associated environments. Interestingly, expression of ribosomal protein L36 was significantly increased when Pf-5 was starved of iron. Upstream analysis on the gene revealed the presence of ferric uptake regulation (FUR) binding site. Further bioinformatic analysis showed that this particular gene is lacking a zinc-binding motif which is common for this family of proteins.

Multiple Syntrophic Interactions in a Terephthalate-Degrading Methanogenic Consortium

Athanasios Lykidis*[1], Chia-Lung Chen[2], Susannah G. Tringe[1], Alice C. McHardy[3], Alex Copeland[1], Nikos C. Kyrpides[1], Philip Hugenholtz[1] and Wen-Tso Liu[2,4].

[1]Joint Genome Institute, Lawrence National Berkeley Laboratory, Walnut Creek, CA 94598, USA
[2]Division of Environmental Science and Engineering, National University of Singapore, Singapore 117576
[3]Computational Genomics & Epidemiology, Max-Planck Institute for Informatics, 66123 Saarbrücken, Germany
[4]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, IL61801, USA

Terephthalate (TA) is one of the top 50 chemicals produced worldwide. Its production results in a TA-containing wastewater that is treated by anaerobic processes through a poorly understood methanogenic syntrophy. Using metagenomics, we characterized the methanogenic consortium inside a hyper-mesophilic (i.e., between mesophilic and thermophilic), TA-degrading bioreactor.
We identified genes belonging to dominant *Pelotomaculum* species presumably involved in TA degradation through decarboxylation, dearomatization, and modified β-oxidation to H2/CO2 and acetate. These intermediates are converted to CH4/CO2 by three novel hyper-mesophilic methanogens. Additional secondary syntrophic interactions were predicted in *Thermotogae*, *Syntrophus* and candidate phyla OP5 and WWE1 populations. The OP5 encodes genes capable of anaerobic autotrophic butyrate production and *Thermotogae*, *Syntrophus* and WWE1 have the genetic potential to oxidize butyrate to CO2/H2 and acetate. These observations suggest that the TA-degrading consortium consists of additional syntrophic interactions beyond the standard H2-producing syntroph – methanogen partnership that may serve to improve community stability.

Increasing the Accuracy of Automated Protein Function Prediction: Assessing the contribution of Functional linkages

A. Medrano-Soto*, D. Pal & D. Eisenberg

UCLA-DOE Institute for Genomics and Proteomics. Howard Hughes Medical Institute. Bioinformatics Center, Indian Institute of Science

Functional linkages predicted by methods based on genomic context often unveil functional relationships between non-homologous proteins that can be used both to improve the quality of existing annotations, and to gain valuable insights into the biological role of proteins with unknown function. Motivated by this fact, we assessed the extent to which functional linkages may contribute to enhance the performance of bioinformatics methods that infer the molecular function of proteins. In our approach (Medrano-Soto et al, 2008) we upgraded the ProKnowfunction metapredictor to include the information on functional relationships contained in the ProLinks database. Then we benchmarked ProKnow by first including andthen excluding functional linkages from the inference process. Our results indicate that when we take functional linkages into account, the accuracy of function prediction improves by nearly 10%, and the detail of Gene Ontology functional descriptions for inferred functions is increased in more than 34% of top assignments. Furthermore, we found support in the biochemical literature for more than 80% of our inferences for previously unannotated proteins.

Medrano-Soto, A., Pal, D. and Eisenberg, D. (2008). "Inferring molecular function: contributions from functional linkages." *Trends Genet* **24**(12): 587-90.

"My God, it's full of polyamines!"...cue Also Spracht Zarathustra.

Anthony J. Michael

Dept. of Pharmacology, Univeristy of Texas Southwestern Medical Center, Dallas

Polyamines are small, flexible, linear organic polycations. They are found in almost all cells of the bacteria, archaea and eukarya and are mostly found bound to RNA but are also incorporated into many specialised metabolites such as siderophores, antibiotics, toxins and alkaloids. Polyamine metabolism has been studied intensively in humans because of its essential role in cell proliferation and cancer, however, it has received relatively scant attention elsewhere in the biological cosmos. In many ways polyamine metabolism exemplifies the problems facing comprehension of genomes and metagenomes. Previous knowledge of microbial polyamine biosynthesis was based on model species such as *Escherichia coli* and *Saccharomyces cerevisiae*. However, in the wider biological universe there are different ways of making the same polyamines, the same pathway can be used to make different polyamines, new pathways are still being discovered, and biosynthesis is an enigma in some important species. Evolution of polyamine metabolism is based on the well established principles of pathway evolution: gene duplication and functional divergence, gene fusion, acquistion of broad substrate specificity enzymes by recruitment from within the genome and by horizontal and endosymbiotic gene transfer. Most polyamine biosynthetic enzymes belong to large, mechanistically diverse enzyme superfamiles of closely related sequence, resulting in almost random annotation of function in microbial genomes. Progress in the deep exploration of microbial polyamine metabolism, through the use of comparative and functional genomics, biochemistry and structural biology will be presented.

Identification of a Novel *Campylobacter upsaliensis* Subspecies Through Comparative Genomics of *C. upsaliensis* and *C. helveticus* Isolates

W.G. Miller*, E. Yee & R.E. Mandrell

U.S. Dept. of Agriculture, Agricultural Research Service, Albany, CA

Members of the genus *Campylobacter*, a taxon within the Epsilonproteobacteria, have been isolated from a wide variety of environments, as well as multiple avian and mammalian hosts. *Campylobacter* spp. have been implicated in disease in both livestock and humans. The primary human pathogen within the genus is the thermotolerant campylobacter *C. jejuni*; however, many other *Campylobacter* spp. have been isolated from human clinical samples and are considered emerging human pathogens. *Campylobacter upsaliensis* and *C. helveticus* are catalase-negative/weakly catalase-positive thermotolerant campylobacters that are associated primarily with canids and felines. *Campylobacter upsaliensis* is also a human pathogen that is isolated, albeit infrequently, from human diarrheal stool samples; to date, no *C. helveticus* strains have been isolated from humans.

Multilocus sequence typing (MLST) of *C. upsaliensis* and *C. helveticus* isolates identified two distinct clades within *C. upsaliensis*, termed 'Cluster I' and 'Cluster II'. To provide further insights into the *upsaliensis/helveticus* group of organisms, the genomes of two human clinical *C. upsaliensis* isolates, representing both MLST clusters, and a feline *C. helveticus* isolate were sequenced to completion. Besides a large core proteome, the genomes of all three strains had multiple features in common, e.g. unlinked 16S and 23S ribosomal rRNA genes, non-clustered lipooligosaccharide (LOS) and capsular biosynthetic genes and a large number (>70) of hypervariable G:C tracts. However, a surprising amount of variation (~96% amino acid identity) exists between the two *C. upsaliensis* core proteomes. Additionally, significant differences in the outer surface structures of the two *C. upsaliensis* clusters were identified: Cluster I and II strains encode genes involved in the addition of sialic acid (associated in *C. jejuni* with GBS) or phosphorylcholine, respectively, onto the LOS. These and other phenotypic differences between the two clusters suggest that the two *C. upsaliensis* clusters represent two genotypically, phenotypically and perhaps clinically distinct subspecies.

Meta-Assembly: An Integrative Approach for *de novo* Assembly of Complete Microbial Genomes From Short Reads

Harish Nagarajan[*1], Jessica Butler[2], Anna Klimes[2], Yu Qiu[1], Karsten Zengler[1], Joy Ward[2], Nelson D. Young[2], Barbara A. Methé[3], Bernhard Ø. Palsson[1], Derek R. Lovley and Christian L. Barrett[1]

[1]University of California, San Diego, CA, USA
[2]University of Massachusetts, Amherst, MA, USA
[3]J.Craig Venter Institute, Rockville, MD, USA

Background: *De novo* sequencing using next-generation technologies have necessitated the development of new algorithms for assembling these short and more error-prone reads into complete genomes. Despite recent reports that indicate significant progress towards this problem by integrating Illumina and 454 technologies, *de novo* assembly of complete microbial genomes from only short reads and without aid from Sanger sequencing still remains an unsolved challenge. However, for confident and complete genome characterization, it is a necessary prerequisite to obtain a single contiguous sequence of the complete genome.
Methods: As a solution to this challenge, we present a four phase approach (Meta-Assembly) for *de novo* assembly of complete microbial genomes from short reads only.
Results: Meta-Assembly is a bi-level integrative approach since we adopt an early integration of Illumina & 454 reads to assemble them into a few scaffolds; and further leverage the different and complementary results provided by multiple assembly programs in order to bridge scaffolds and possibly resolve any degenerate positions. In the third phase, we employ a PCR-based search strategy in order to obtain the correct relative ordering and orientation of the scaffolds. Finally, we corrected for indels and any errors introduced during the earlier phases by aligning the Illumina reads to obtain the finished genome. We first applied Meta-Assembly using 50X Illumina GA1 singleton reads and 16X 454 GS-FLX paired-end sequencing reads for an enhanced electricity-producing *Geobacter* variant (KN400). This resulted in the complete genome of KN400 of length 3714259bp. We have also applied Meta-Assembly on two clinical isolates of *Acinetobacter baumannii* and these genomes are about to be completed.
Conclusions: To our knowledge, Meta-Assembly on KN400 is the first reported de novo assembly of a complete bacterial genome using next generation sequencing technologies Our strategy will add significant value to many current and future sequencing efforts and will impact the field of microbial genomics by accelerating the increase in the number of complete prokaryotic genomes.

Diversity-Generating Retroelements Promote Accelerated Evolution of Target Genes in Bacteria and Phage and Have Broad Implications for Protein Engineering

Angela Nieh*, Huatao Guo, Longping V. Tse, Mari Gingery, Diego Arambula, Minghsun Liu, Sabrina Oukil, and Jeff F. Miller

David Geffen School of Medicine at UCLA, Los Angeles, CA

Diversity-generating retroelements (DGRs) introduce vast amounts of targeted diversity into protein-encoding DNA sequences. During mutagenic homing, adenine residues are converted to random nucleotides in a unidirectional retrotransposition process from a donor template repeat (TR) to a recipient variable repeat (VR). Using the *Bordetella* phage DGR as a signature, we have discovered DGRs not only in bacteriophages, but also in a multitude of plasmids and bacterial genomes from diverse environments. DGRs are present in human pathogens (*Treponema, Legionella spp.*), human commensals (*Bacteroides, Bifidobacterium spp.*), green sulfur bacteria (*Chlorobium, Prosthecochloris spp.*), cyanobacteria (*Trichodesmium, Nostoc spp.*), magnetotactic bacteria (*Magnetospirillum spp.*) and many other species. Further studies confirmed DGR activity in *Legionella,* demonstrating the functionality of bacterial genome-encoded DGRs. Moreover, the target protein of the *Legionella* DGR appears to be a cell surface protein, indicating potential physiological functions for DGRs. Using the *Bordetella* phage DGR as a model system, we showed that DGR homing occurs through a unique target DNA-primed reverse transcription (TPRT) mechanism. Deletion analysis showed that homing requires a ~24 bp sequence downstream of VR that could potentially form a hairpin structure. Subsequent analysis showed that base pairing, but not the specific sequence, is required for mutagenic homing. Although the wild type hairpin configuration appears to be optimal, various hairpin stem lengths and distances from VR are tolerated. Using the tetracycline resistance gene as a reporter, we showed that the BPP-1 DGR can be engineered to target virtually any heterologous gene, suggesting broad applications for protein engineering.

Characterization of the microbial populations associated with the nematode

Jean-Paul Baquiran1, Brian Thater2, Sammy Sedky3, David E. Crowley2, Paul De Ley3 & Paul M. Orwin4

1Environmental Toxicology Program, University of California, Riverside
2Environmental Sciences Department, University of California, Riverside
3Department of Nematology, University of California, Riverside
4Biology Department, California State University at San Bernardino

Microbial symbiosis has long been an essential survival feature for bacteria and their eukaryotic hosts. In this research, we examined the potential symbiotic relationships of bacteria with the free-living bacteriovorus nematode, Acrobeloides maximus. Large numbers of A. maximus cultured on E. coli were introduced into rhizosphere soil samples for over 24 hours, extracted from the soil using a high salt solution, and washed to remove superficially associated bacteria. The microbial population associated with A. maximus, was analyzed using both culture-independent and culture-dependent techniques. 16S rDNA clone libraries were constructed and sequencing results indicate that there are two main organisms that may act as bacterial symbionts, an alpha-proteobacterium most closely related to Ochrobactrum tritici and a previously undescribed sphingobacterium, Pedobacter sp. In addition, long term co-culture plate assays, where the nematodes extracted from the soil were allowed to grow for extended periods of time, confirmed the dominance of both organisms. A number of bacterial species were isolated from worms from these studies, including both symbionts as well as others favored by agar plate growth conditions. Parallel studies with another representative nematode, Caenorhabdits elegans and the original soil samples indicate that the symbiont colonization is specific to Acrobeloides maximus. Fluorescence in situ hybridization (FISH) was used to visualize the bacterial symbionts, which support their close association to the nematode. We hypothesize that several potential positive interactions may be taking place in the soil. The bacterial symbiont may be defending the nematode from pathogenic bacteria or fungi, it may be aiding in the digestion of certain bacterial prey, or the symbiont may be providing an essential metabolic precursor for worm growth. Overall, our analysis revealed that specific microbial populations associate with A. maximus, which may function as an ecosystem similar to the microbiomes of other metazoans.

Spatial Clustering Used to Investigate Operon Evolution

D. C. Ream*(1) M.D. Wohlever(2) & I. Friedberg(1,3)

1. Dpt. Of Microbiology

2. Dpt. of Mathematics & Statistics

3. Dpt. of Computer Science and Software Engineering

Miami University, Oxford OH USA

Correspondence: i.friedberg@muohio.edu

Operons are important features in bacterial genomes, containing related genes under common regulatory control. Several mechanisms were suggested for operon evolution, those include whole-entity horizontal gene transfer know as the "selfish operon" model and several models where operons are created or altered from single genes or small gene clusters. Here we establish a set of metrics based on the chromosomal clustering of the gene components of an operon to help determine evolutionary mechanisms in different operons.

We investigated all experimentally determined operons found in RegulonDB that contained at least five genes. We then located proteobacteria that have all or part of these operons within their genomes, and determine how closely related to each is to our reference organism, *E.coli*. Next we developed the set of the metrics mentioned above, which allow us to quantitatively classify operon evolution. The evolutionary events we measure in operons are their spreading out, breaking up, rearranging, and gene loss. Our preliminary results support that in general the less related an organism is to the gold standard organism, the greater the divergence in its operonic structure. However, this is not absolute, and phylogenetic distance alone cannot explain the patterns of operon de-structuring observed. It appears that operons can be preserved in distant species, while broken up in closely related ones. These initial finding support horzontal gene transfer as a major determinant in operon evolution. When more operons are included in the study we hope to show that there are varying degrees of divergence, with some operon types being highly conserved, while other types show less conservation.

The organismal distribution and substitutions in key residues suggest functional change in Cld, a important enzyme in microbial detoxification of (per)chlorate

Kathryne G. Byrne-Bailey, Susanna Repo*, Eunice Moon-Lim, Antinea H. Chair, Anna Engelbrektson, Charlotte Carlström, Cameron J. Trash, Steven E. Brenner and John D. Coates

Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA.

The dismutation of chlorite into chloride and oxygen, mediated by the enzyme chlorite dismutase (Cld), is encoded by the *cld* gene. Cld is a key enzyme in the respiratory pathway by which (per)chlorate is completely reduced to innocuous chloride. This metabolic activity has been identified in environmental bacterial isolates from across the Proteobacteria. A recently created Pfam family (PF06778) identifies homologs of *cld* gene in a large number of the bacterial and archaeal species not experimentally confirmed to have perchlorate reduction activity. Moreover, most genomes with the *cld* homologs were not found to have homologs of another key enzyme in the perchlorate reduction pathway. This suggests that some of the *cld* homologs may not have chlorite dismutase activity, and instead have some other functional role. To investigate the functional evolution of the Cld and the homologs further, we inferred a phylogenetic tree of the family. In the tree, the Clds of known (per)chlorate reducing bacteria ((P)CRB) separated into a distinct clade from non/untested-(P)CRB bacteria. We validated experimentally the functions of *cld* homologs outside the known (P)CRB clade, and these proteins were not found to have chlorite dismutase activity. Based on the sequence and 3D-structure analyses we performed on all known Cld and *cld homologs*, we suggest residues and regions crucial for the chlorite dismutase activity. We conclude that probably not all the recently identified *cld* homologs have chlorite dismutase function, but another or possibly several yet unknown functions.

The Enzyme Misannotation Resource

A.M. Schnoes*, F. Lauck, S.D. Brown, P.C. Babbitt

University of California, San Francisco

The massive amounts of sequence data determined by the many genome sequencing projects has resulted in a commensurate need by the scientific community to know the biological functions of all of these sequences. However, this rate of sequence deposition into public databases completely outpaces the ability of current scientific methodology to experimentally characterize each individual protein product. As a result, automated prediction is the mechanism by which almost all sequences in public databases have been annotated with functional designations and is a critical area of active research. The accuracy of these functional annotations is important to verify and recently we computationally evaluated the predicted functions of proteins in several functionally diverse enzyme superfamilies. We specifically chose these types of enzyme superfamilies for analysis because they pose difficult problems for function prediction; the sequences, structures and conserved active sites of proteins in any one such superfamily are similar yet their overall chemical reactions, substrates and products vary widely. The misannotation analysis utilized as a gold standard the expertly curated sequence-structure-function data available in the Structure-Function Linkage Database (SFLD, http://sfld.rbvi.ucsf.edu) using an analysis protocol consisting of calculated sequence thresholds and manual curation. Querying functional annotations for these enzyme families in the NCBI NR, UniProtKB/TrEMBL, UniProtKB/Swiss-Prot and KEGG protein sequence databases, we found that misannotation is a greater problem than has previously been described and that the problem of misannotation appears to be worsening over time. This has motivated the development of a community resource for misannotation identification, prediction and curation. This Enzyme Misannotation Resource (EMR) is integrated with the SFLD and aims to allow users to perform their own misannotation analyses using SFLD data and network visualization.

Analysis of Bacterial Community at the Human Conjunctiva

Valery I. Shestopalov

University of Miami Miller School of Medicine

The ocular surface (OS) microbiota has been shown to contribute to the pathogenesis of infectious and immune-mediated diseases of the eye. However, the microbial diversity at the OS remains poorly investigated remains poorly investigated due to the limits of traditional cultivation techniques. To fill this knowledge gap we performed comprehensive survey of the human ocular surface microbiota using microbial ecology techniques based on the 16S rRNA classification.
Microbial community analysis of the conjunctiva revealed an average of 224 distinct bacterial phylotypes per individual (n=4). This amount exceeds what has been identified by traditional culture-based techniques over 20-year time period by approximately two fold. Among 55 different bacterial genera identified in this study, 22 have never been previously detected in the eye. Our analysis revealed a diverse population with significant variability between individuals. Individual variation (Shannon diversity index) averaged 2.70+0.69 across all samples. Remarkably, eight "core" genera were present in all volunteers, with six of them accounting to 68% of the bacterial population within the analyzed group.
Our first global survey of bacterial community at the conjunctiva revealed well known ocular pathogens, abundant commensal and environmental species, and detected significant number of previously uncharacterized bacteria.

Genome-scale reconstruction and integration of the ArgR and Lrp regulatory networks in Escherichia coli

S. Federowicz*, Y.S. Park, B.O. Palsson & B.K. Cho

University of California, San Diego.

To determine the global topology of the transcriptional regulatory network for the arginine repressor protein (ArgR) and the leucine-responsive protein (Lrp) a series of high-resolution ChIP-chip and expression profiling experiments were performed. ChIP-chip experiments revealed a total of 64 and 143 unique and reproducible binding regions found under different conditions for ArgR and Lrp respectively. These binding regions were then extracted from the genome and used to computationally derive binding motifs which agreed precisely with previously determined motifs for ArgR and Lrp. Binding motifs were then used to rescan the ArgR and Lrp binding regions to determine the extent of multiple binding events and operator sites. Expression data was then integrated with these results to determine patterns of activation and repression in response to position specific binding. This allowed for the construction of a global transcriptional regulatory network encapsulating both the ArgR and Lrp regulons. Analysis of this network reveals the individual and shared regulatory networks of these key global regulators in amino acid biosynthesis and metabolism.

Investigation of Type III Secretion and effectors proteins in pathogenic *Burkholderia*

I.Toesca Spears, C. Todd French, L. Huang, S. Moghadam, W. Wong, M. Liu, Jeff F. Miller

University of California Los Angeles, Los Angeles, CA

Burkholderia pseudomallei is the agent of the infectious disease melioidosis, and is considered by the CDC to be a potential biological weapon. B. pseudomallei employs type III secretion systems (T3SS) to deliver protein effectors into host cells, thus facilitating its intracellular survival. Currently, few effectors have been identified and characterized in B. pseudomallei. Our research focuses on the identification and the characterization of such effectors. *In silico* analysis based on colocalization of effectors with T3SS chaperones and sequence homology to other effectors from animal and plant pathogens led to the identification of 19 putative Burkholderia effector protein (Bep) candidates. Using transient transfection of Bep proteins in eukaryotic cells, we show that BepA exhibits perinuclear localization and co-immunoprecipitates with cellular actin and the Fam38a endosomal protein. BepC was found to bundle actin, while BepD decorates microtubules and the spindle apparatus of mitotic cells. We have also shown that BepA is secreted in a T3SS-dependent manner from B. pseudomallei, using the BopE T3SS effector as positive control. Experiments using a Cre/LoxP recombination system are currently in progress to verify that potential Bep proteins are indeed transclocated into mammalian cells. Previously, T3SS-3 has been posited to be an important factor required for Burkholderia invasion and in maintenance of the intracellular life cycle. Here, we demonstrate that T3SS-3 null mutants fail to escape the endosome following infection of non-phagocytic cells, but are fully capable of polymerizing actin and spreading between cells if delivered directly into the cell cytoplasm via plasmonic photothermal injection.

Comparative Genomic Analysis of Mycobacterium Phage Bruin by UCLA Undergraduates

W. Villella[1]*, A. Sahakyan[1], P. Nim[1], F. Miller[1], R. Liang[1], C. Chang[2], T. Chen[2], D. Cheng[2], C. Chu[2], P. Eyermann[2], S. Gustafson[2], S. Horvath[2], S. Ji[2], M. Kim[2], Y. Lee[2], S. Lyu[2], V. Minassian[2], F. Muqri[2], D. Nguyen[2], C. Shek[2], T. Sheu[2], J. Sugiyama[2], A. Tran[2], L. Tran[2], K. Wu[2], F. Xing[2], Z. Qiao[2], J. Ku[2], J. McCracken[2], L. Feldman[2], I. Atanasov[3], H. Zhou[3], and E. Sanders[1,2].

[1] Genome Annotation team (MIMG 189-Spring 2010), [2] Phage Discovery teams (MIMG 189-Winter 2008), Department of Microbiology, Immunology, & Molecular Genetics, University of California, Los Angeles; [3] Electron Imaging Center for Nanomachines (EICN), California Nanosystems Institute (CNSI), University of California, Los Angeles.

Bacteriophages (phages) represent the most abundant biological entities in the biosphere and only a fraction of which have been purified and characterized. Investigations of phage diversity and comparative analysis of phage genomes is important because such studies reveal information about the role of phage in the environment and in mediating horizontal gene transfer, their interactions with bacterial communities, and their potential as a vehicle for therapies against bacterial infections. Mycobacterium Phage Bruin was isolated in 2008 from a compost heap and mulch pile on the UCLA campus using the bacterial strain *Mycobacterium smegmatis* mc^{2}155 (ATCC 700084). As participants in a phage discovery course at UCLA, undergraduates worked collaboratively on the morphological description of Phage Bruin. This phage forms predominantly large, clear plaques. Negative stain electron microscopy reveals Phage Bruin has an isometric head and a long noncontractile tail, consistent with the virion morphotype *Siphoviridae*. In collaboration with the HHMI Science Education Alliance and the DOE Joint Genome Institute, the genome of Phage Bruin was sequenced. Classified as a double-stranded DNA virus, the genome length of Phage Bruin is approximately 74.2 kbp with a 63% GC content. Following automated annotation, the genome is estimated to contain 131 protein coding genes and 2 tRNA genes. Comparative analysis suggests Phage Bruin gene products are most closely related to other mycobacteriophages such as Cjw1, Porky, and Kostya. Only 18 protein coding genes have a functional prediction. Manual annotation by undergraduates in a genome annotation course at UCLA proposed new annotations for several of the 113 remaining gene calls. The results of this collaborative initiative will be presented here.

Studying the Human Oral Microbiome with Illumina Sequencing of 16S rDNA Fragments

Katrine Whiteson*[aΔ], Vladimir Lazarevic[aΔ], Susan Huse[b], David Hernandez[a], Laurent Farinelli[c], Magne Østerås[c], Jacques Schrenzel[a], Patrice François[a]

[a] University of Geneva Hospitals, Switzerland, [b]Marine Biological Laboratory, Woods Hole, MA, [c]Fasteris, Switzerland. [Δ]Both authors contributed equally to this work.

Improvements in speed, cost and depth of sequencing in the recent past have led to a revolution in the study of microbial communities. Previously only the 1% of bacteria that can grow in laboratory conditions were accessible; now biologists are sequencing DNA directly from human and environmental samples, and learning about community composition in the context oftime, place, sickness, and health. Our assumptions about how infectious diseases arise are being turned upside-down - a "pathogen" could be a disrupted microbial community rather than an individual species.

In this context, we undertook the first study of a microbial community relying on Illumina sequencing technology, in order to establish what microbial communities in the human mouth look like. Following an *in silico* comparison of 16S rDNA segments from species known to reside in the human mouth, we identified the V5 region as the most promising for identification when limited to short reads (~75bp). We sequenced the V5 region of a pooled sample of saliva and throat swabs from 3 adults using Illumina technology, and identified 135 genera or higher taxonomic ranks from the resulting 1,373,824 sequences. While the abundances of the most common phyla (*Firmicutes, Proteobacteria, Actinobacteria, Fusobacteria* and*TM7*) are largely comparable to previous studies on a smaller scale, *Bacteroidetes* were less present. Potential sources for this difference include classification bias in this region of the 16S rRNA gene, human sample variation, sample preparation and primer bias. Using an Illumina sequencing approach, we achieved a much greater depth of coverage thanprevious oral microbiota studies, allowing us to identify several taxa not yet discovered in this type of sample.

Metabolically Active Bacteria in the Atmosphere Revealed by RNA-based Community Composition.

A. M. Womack*, B. J. M. Bohannan, J. L. Green.

University of Oregon

Microorganisms are ubiquitous and diverse in the atmosphere, and the air has long been recognized as an important dispersal conduit for microbial life. Recent studies have underscored the important role that airborne microbes play in atmospheric processes, by affecting climate (e.g. as ice nucleators) and atmospheric chemistry (e.g. through biological transformations of atmospheric compounds). Despite the importance of airborne microbes, we do not know the extent of microbial biodiversity in the atmosphere, or what proportion of this biodiversity is metabolically active, and we understand very little about how microbial biodiversity affects the functioning of the atmosphere. Here, we bridge this knowledge gap by applying RNA-based analyses to aerial bacterial communities. RNA-based approaches can be used to differentiate the metabolically active portion of a sampled community from the total community, which includes dormant, dead, or otherwise inactive cells. Total DNA and RNA were extracted from an air sample collected over a period of six hours. Total community diversity based on 16S rDNA was compared to active community diversity based on 16S rRNA. Significant differences were observed between the RNA-based and DNA-based community composition, indicating that a significant portion of the aerial microbial community was metabolically active.

Studying the Human Oral Microbiome with Illumina Sequencing of 16S rDNA Fragments

Katrine Whiteson*[aΔ], Vladimir Lazarevic[aΔ], Susan Huse[b], David Hernandez[a], Laurent Farinelli[c], Magne Østerås[c], Jacques Schrenzel[a], Patrice François[a]

[a] University of Geneva Hospitals, Switzerland, [b]Marine Biological Laboratory, Woods Hole, MA, [c]Fasteris, Switzerland. [Δ]Both authors contributed equally to this work.

Improvements in speed, cost and depth of sequencing in the recent past have led to a revolution in the study of microbial communities. Previously only the 1% of bacteria that can grow in laboratory conditions were accessible; now biologists are sequencing DNA directly from human and environmental samples, and learning about community composition in the context oftime, place, sickness, and health. Our assumptions about how infectious diseases arise are being turned upside-down - a "pathogen" could be a disrupted microbial community rather than an individual species.

In this context, we undertook the first study of a microbial community relying on Illumina sequencing technology, in order to establish what microbial communities in the human mouth look like. Following an *in silico* comparison of 16S rDNA segments from species known to reside in the human mouth, we identified the V5 region as the most promising for identification when limited to short reads (~75bp). We sequenced the V5 region of a pooled sample of saliva and throat swabs from 3 adults using Illumina technology, and identified 135 genera or higher taxonomic ranks from the resulting 1,373,824 sequences. While the abundances of the most common phyla (*Firmicutes, Proteobacteria, Actinobacteria, Fusobacteria* and*TM7*) are largely comparable to previous studies on a smaller scale, *Bacteroidetes* were less present. Potential sources for this difference include classification bias in this region of the 16S rRNA gene, human sample variation, sample preparation and primer bias. Using an Illumina sequencing approach, we achieved a much greater depth of coverage thanprevious oral microbiota studies, allowing us to identify several taxa not yet discovered in this type of sample.

Metabolically Active Bacteria in the Atmosphere Revealed by RNA-based Community Composition.

A. M. Womack*, B. J. M. Bohannan, J. L. Green.

University of Oregon

Microorganisms are ubiquitous and diverse in the atmosphere, and the air has long been recognized as an important dispersal conduit for microbial life. Recent studies have underscored the important role that airborne microbes play in atmospheric processes, by affecting climate (e.g. as ice nucleators) and atmospheric chemistry (e.g. through biological transformations of atmospheric compounds). Despite the importance of airborne microbes, we do not know the extent of microbial biodiversity in the atmosphere, or what proportion of this biodiversity is metabolically active, and we understand very little about how microbial biodiversity affects the functioning of the atmosphere. Here, we bridge this knowledge gap by applying RNA-based analyses to aerial bacterial communities. RNA-based approaches can be used to differentiate the metabolically active portion of a sampled community from the total community, which includes dormant, dead, or otherwise inactive cells. Total DNA and RNA were extracted from an air sample collected over a period of six hours. Total community diversity based on 16S rDNA was compared to active community diversity based on 16S rRNA. Significant differences were observed between the RNA-based and DNA-based community composition, indicating that a significant portion of the aerial microbial community was metabolically active.

StressChip: A Powerful Tool to Assess the Stress Responses in Microbial Communities

Aifen Zhou*, Zhili He, Yujia Qin, Zhenmei Lu, Jianping Xie, Ye Deng & Jizhong Zhou.

University of Oklahoma, Norman

Understanding stress responses is critical for our ability to control microbial growth, survival and adaptation in the context of ecology, biotechnology, or pathogenesis. We have developed a functional gene-based microarray, termed StressChip to analyze stress responses of microbial communities. Sigma factors, transcription factors and functional genes involved in specific and general stress responses were included in the StressChip. Eleven functional categories including 45 functional genes involved in responses to temperature, salt or nutrient limitation were covered in the StressChip. StressChip with a total of 21,560 probes covering 55,425 gene sequences has been evaluated with two sets of environmental samples. First, an analysis of oil plume samples taken from the Gulf of Mexico showed that the signal intensities of primary sigma factor σ^{70}, heat stress sigma factor σ^{24} and the general stress response sigma factor σ^{38} were significantly increased compared to control samples, suggesting the presence of increased general stress responses. Also, the increase of signal intensities of stress response genes involved in limitations of oxygen, phosphate and nitrogen was detected. These data suggested that marine ecosystems were under stress conditions with oil spill, and that these functional genes were important for microbial growth, survival and adaptation under oil spill related environmental stresses. Second, a study on the effect of warming (+2°C) on soil microbial communities demonstrated the significantly increased signal intensities of heat shock sigma factors σ^{24} and σ^{32} and sigma factor σ^{70} as well as function genes involved in heat shock, stringent response, responses to limitation of phosphate, nitrogen or oxygen, strongly helping our understanding of the effect of global warming on soil microbial communities. In summary, the StressChip will be a powerful tool to investigate stress responses of microbial communities to environmental stresses and provide insights into our understanding of microbial growth, survival and adaptation in stressed environments.

Evaluation of Pulsed-field Gel Electrophoresis Profiles for the Identification of *Salmonella* Serotypes

Wen Zou,[1*] Wei-Jiun Lin,[1] Steven L. Foley,[2] Chun-houh Chen,[3] Rajesh Nayak,[2] and James J. Chen[1]

[1] ^{1}Division of Personalized Nutrition and Medicine, [2]Division of Microbiology, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas 72079 and [3]Institute of Statistical Science, Academia Sinica, Taipei, 11529, Taiwan

Pulsed-field gel electrophoresis (PFGE) is a standard typing method for isolates from *Salmonella* outbreaks and epidemiological investigations. Eight hundred sixty six (866) *Salmonella enterica* isolates from eight serotypes, including Heidelberg (n = 323); Javiana (n = 200), Typhimurium (n = 163), Newport (n = 93); Enteritidis (n = 45); Dublin (n = 25); Pullorum (n = 9); and Choleraesuis (n = 8), were subjected to PFGE and their profiles analyzed by Random Forest classification, and compared to the conventional hierarchical cluster analysis to determine potential predictive relationships between PFGE banding patterns and particular serotypes. Cluster analysis displayed only the underlying similarities and relationships of the isolates from the eight serotypes. However, for serotype prediction of a non-serotyped *Salmonella* isolate from its PFGE pattern, Random Forest classification provided better accuracy than conventional cluster analysis. Discriminatory DNA band class markers were identified for distinguishing serotype Heidelberg, Javiana, Typhimurium and Newport isolates.

Using genome-scale metabolic network models to predict environments that induce microbial cross-feeding

Niels Klitgord(1)*, Daniel Segrè(1,2)

(1) Bioinformatics Program, (2) Department of Biology and Department of Biomedical Engineering, Boston University, Boston.

Interactions between microbial species can be mediated by the exchange of small molecules, secreted by one species and metabolized by another. Both one-way (commensal) and two-way (mutualistic) interactions may contribute to complex networks of interdependencies. Understanding these interactions constitutes an open challenge in microbial ecology, with applications ranging from the human microbiome to environmental sustainability. We approach this challenge by identifying environments that induce symbiotic interactions between pairs of microbes. Specifically, we use genome-scale stoichiometric models of metabolism to identify media that can sustain growth for a pair of species, but fail to do so for one or both individual species, thereby inducing symbiotic interactions. We first tested our approach on two previously studied mutualistic pairs, and on a pair of highly curated model organisms, showing that our algorithms successfully recapitulate known interactions, robustly predict new ones, and provide novel insight on exchanged molecules. We then applied our method to all possible pairs of seven microbial species, and found that it is always possible to identify putative media that induce commensalism or mutualism. Our analysis also suggests that symbiotic interactions may arise more readily through environmental fluctuations than genetic modifications. We envision that our approach will help generate microbe-microbe interaction maps useful for understanding microbial consortia dynamics and evolution, and for exploring the full potential of natural metabolic pathways for metabolic engineering applications.

DIRECTORY

Eric J. Alm
The Broad Institute of MIT and Harvard
Parsons Laboratory
Room 48-317
15 Vassar Street
Cambridge, MA 02139
Tel: (617) 253-2726
E-mail: ejalm@mit.edu

Brian Anton
New England Biolabs
240 County Road
Ipswich, Massachusetts 01938-2723
Tel: (978) 927-5054
Fax: (978) 921-1350
E-mail: anton@neb.com

Patricia Babbit
Sciences and Pharmaceutical Chemistry
University of California San Francisco
The California Institute for Quantitative
Biomedical Research (QB3)
1700 4th Street, 5th Floor
San Francisco, CA 94134
Tel: (415) 476-3784
Fax: (415) 476-2744
Email: babbitt@cgl.ucsf.edu

Jean-Paul Baquiran
Dept. of Environmental Sciences
University of California, Riverside
5500 University Pkwy
San Bernardino, CA 92407
Tel: (909) 537-5405
E-mail:jean-
paul.baquiran@email.ucr.edu

Elinne Becket
University of California, Los Angeles
Molecular Biology Institute
Jeffrey H. Miller Lab
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 825-8924
E-mail: efini@ucla.edu

Nikolaj Blom
Novozymes A/S
Krogshoejvej 36
2880 Bagsvaerd Denmark
Tel: (45) 4-446-0880
E-mail: nblm@novozymes.com

Steven E. Brenner
University of California Berkeley
461A Koshland Hall #3102
Berkeley, CA 94720-3102
Tel: (510) 643-9131
Fax: (510) 666-2505
Email: brenner@compbio.berkeley.edu

Ilana L. Brito
Earth Institute
Columbia University
Tel: (617) 817-8715
E-mail: ilb2111@columbia.edu

Roger L. Chang
Systems Biology Research Group
9500 Gilman Drive
University of California San Diego
La Jolla, CA 92093-0412
Tel: (858) 822-1144
E-mail: rlchang@ucsd.edu

Casey Chen
USC school of Medicine
925 W. 34th St.
Los Angeles, CA 90089
Tel: (213) 740-2538
Fax: (213) 740-7965
Email: ccchen@usc.edu

Valérie de Crécy-Lagard
Department of Microbiology and Cell
Science

P.O. Box 110700
University of Florida
Gainesville, FL 32611-0700
Tel: (352) 392-9416
Fax: (352) 392-5922
E-mail: vcrecy@ufl.edu

Elizabeth Czornyj
California State University of Northridge
Tel: (818) 267-4590
E-mail: elizabethczornyj0@hotmail.com

Aaron Darling
UC Davis Genome Center
451 Health Sciences Dr.
Davis, CA 95616
Tel: (530) 752-8409
E-mail: aarondarling@ucdavis.edu

Gautam Dantas
Washington University School of Medicine
Center for Genome Sciences
4444 Forest Park Ave, Rm 6202-A
Campus Box 8510
Saint Louis, MO 63108
Tel: (314) 362-7238
Fax: (314) 362-2156
Email: dantas@wustl.edu

Ruchira S. Datta
324D Stanley Hall
Department of Bioengineering
California Institute for Quantitative Biosciences (QB3)
University of California
Berkeley, CA 94720-3220
Tel: (510) 642-6642
Email: ruchira@berkeley.edu

Catherine Davis
The Proctor & Gamble Company
6110 Center Hill Avenue
Cincinnati, OH 45224
Tel: (513) 634-1866
E-mail: Davis.cc@pg.com

Armand E Dichosa
Los Alamos Laboratory, Bioscience Division
TA-43-0001, MS-M888
B6:Genome Science-JGI
Los Alamos National Laboratory
Los Alamos, NM 87545
Tel: (505) 606-0783
Fax: (505) 665-3024
Email: armand@lanl.gov

Ashlee Earl
Genome Sequencing and Analysis Program
The Broad Institute of Harvard and MIT
Rm. NE30-7028C, 7 Cambridge Center
Cambridge, MA 02142
Tel: (617) 324-3878
Fax: (617) 324-4987
Email: aearl@broad.mit.edu

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

Mallory Embree
417 Powell-Focht Bioengineering Hall
9500 Gilman Drive
University of California San Diego
La Jolla, CA 92093-0412
Tel: (858) 822-1144
E-mail: mallory.embree@gmail.com

Alex W. Ensminger
Tufts University School of Medicine
145 Harrison Avenue
Boston, MA 02111
E-mail: Alexander.Ensminger@tufts.edu

Steve Federowicz
Systems Biology Research Group
9500 Gilman Drive MC 0419
University of California San Diego
La Jolla, CA 92093-0419
Tel: (858) 822-1144
E-mail: sfederow@gmail.com

Tamas Feher
Biological Research Center
Of the Hungarian Academy of Sciences
62 Temesvari Krt.
H6726 Szeged
Hungary
Email: fehert@brc.hu

Manuel Ferrer
CSIC-Institute of Catalysis
Marie Curie 2
Madrid, Spain
Email: mferrer@icp.csic.es

Michael Fischbach
University of California, San Francisco
Department of Bioengineering and
Therapeutic Sciences
1700 4th Street, Byers Hall 308C
San Francisco, CA 94158
Tel: (415) 514-9435
Fax: (415) 514-9736
Email: fischbach@fischbachgroup.org

Michael Fitzsimons
Los Alamos National Laboratory
Mailstop M888
Bioscience Division
Los Alamos, NM 87545
Tel: (505) 665-4059
Fax: (505) 665-3024
E-mail: msfitz@lanl.gov

Jessica Forbester
Biology Department
CSU San Bernardino
5500 University Pkwy
San Bernardino, CA 92407
Tel: (909) 537-5405
E-mail: jf285@bath.ac.uk

Larry J. Forney
Department of Biological Sciences
Life Sciences South, Rm. 441A
University of Idaho
Moscow, Idaho, USA 83844-3051
Tel: (208) 885-6011
Fax: (208) 885-7905
E-mail: lforney@uidaho.edu

Kevin Forsberg
Washington University in St. Louis
6th Floor, Center For Genome Sciences
Campus Box 8510
4444 Forest Park Ave
St. Louis, MO, 63108
Tel: (314) 362-2763
E-mail: kjforsbe@wustl.edu

Kevin Frew
University of California, Los Angeles
Molecular Biology Institute
Jeffrey H. Miller Lab
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 825-8924
E-mail: k.f@ucla.edu

Iddo Friedberg
Dpt. of Microbiology, 32 Pearson Hall
700 E. High St.
Miami University Oxford, OH 45056
Tel: (513) 529-3182
E-mail: idoerg@gmail.com

Michael Galperin
NCBI, National Institute of Health
8000 rockville Pike
Bldg 38A, Room 507, MSC3830
Bethesda, MD 20894
Tel: (301) 435- 5910
Fax: (310) 435- 7793
Email: galperin@nih.gov

John Gerlt
Department of Biochemistry
University of Illinois
301A Roger Adams Lab
600 South Mathews Avenue
Urbana, IL 61801
Tel: (217) 333-3945
Email: j-gerlt@illinois.edu

Garry Gippert
Novozymes A/S
Krogshoejvej 36
2880 Bagsvaerd Denmark
Phone: (45) 4-446-0032
E-mail: gpgi@novozymes.com

Robyn Goacher
87 Kingston Place
Buffalo NY 14210
Email: rgoacher@gmail.com

Patrick R. Gonzales
Washington University in St. Louis
4444 Forest Park Bl, Campus Box 8510,
St. Louis, MO 63108
Tel: (314) 303-2411
Fax: (314) 362-2156
Email: prgonzales@go.wustl.edu

Joseph J. Grzymski
Division of Earth and Ecosystem
Science
Desert Research Institute
2215 Raggio Parkway
Reno, Nevada 89512
Tel: (775) 673-7478
Fax: (775) 673-7485
E-mail: joeg@dri.edu

Lucia Guisado
Biological Sciences, specializing in
Biotechnology, at the Universidad
Complutense de Madrid (Spain)
Tel: (310) 825-8924
E-mail: luciaguisado@gmail.com

Loren Hauser
Computational Biology and
Bioinformatics Group
Biosciences Division
Oak Ridge National Laboratory
PO Box 2008, MS 6420
Tel: (865) 241-1960
E-mail: hauserlj@ornl.gov

Luisa Hiller
Center for Genomic Sciences
320 E. North Ave
South Tower, 11th floor
Pittsburgh, PA 15212
Tel: (412) 359-6214
Fax: (412) 359-6995
E-mail: luhiller@gmail.com

Shiao-Wei Huang
Institute of Zoology, National Taiwan
University
No1, Sec4, Roosevelt Road
Taipei 10617, Taiwan, R.O.C
Tel: (886) 2-33662456
Fax: (886) 2-23638179
E-mail: d2205002kimo@yahoo.com.tw

John F. Hunt
Columbia University
702A Fairchild Center, MC 2434
212 Amsterdam Avenue
New York, New York 10027
Tel: (212)-854-5443
fax: (212)-865-8246
E-mail: jfhunt1@gmail.com

Scott A. Jackson
US FDA-CFSAN
Division of Molecular Biology
8301 Muirkirk Rd.
MOD1 Rm 3706
HFS-025
Laurel, MD 20708
Tel: (301) 210-7807
Fax: (301) 210-7976
E-mail: Scott.Jackson@fda.hhs.gov

Benjamin Janto
Center for Genomic Sciences
Drexel College of Medicine,
Microbiology & Immunology
E-mail: BJANTO@wpahs.org

Dezhi Kang
UCLA, Pharmacology
1288 Barry Ave. # 1
Los Angeles, CA 90025
Tel: (310) 840-0432
Email: days369@gmail.com

Manzhu Kang
Department of Biological Sciences
California State University, L.A.,
5151 State University Dr., BS-143
Los Angeles, CA 90032-8201
Tel: (323) 343-2188
E-mail: mkang8@calstatela.edu

Gregory Kanzaki
Department of Biological Sciences
California State University, L.A.,
5151 State University Dr., BS-143
Los Angeles, CA 90032-8201
Tel: (323) 343-2040
E-maiL:gkanzak@calstatela.edu

Peter Karp
SRI International
333 Ravenswood Avenue
Room AE206
Menlo Park, CA 94025-3493
Tel: (650) 859-4358
Fax: (650) 859-3735
Email: pkarp@AI.SRI.com

Simon Kasif
Bioinformatics Research
Department of Biomedical Engineering
Boston University
44 Cummington St. Boston, MA 02215
Tel: (617) 358-1845
Email: kasif@bu.edu

Gabriela Kasimatis
Molecular and Medical
Pharmacology/UCLA
Tel: (310) 983-3212
E-mail: gkasimatis@mednet.ucla.edu

Steve Kembel
Center for Ecology and
Evolutionary Biology
335 Pacific Hall
5289 University of Oregon
Eugene, Oregon 97403
Tel: (541) 346-4812
Email: skembel@uoregon.edu

Jihyun F. Kim
Korea Research Institute of Bioscience
and Biotechnology (KRIBB)
University of Science and Technology
(UST)
111 Gwahangno, Yuseong-gu, Daejeon
305-806, Republic of Korea
Tel: (82) 42-860-4412
Fax: (82) 42-879-8595
E-mail: jfk@kribb.re.kr

William Klimke
NCBI/NLM/NIH
6A5-430-4a, Bldg 45
Bethesda, MD 20894
Tel: (301) 496-4869
Fax: (301) 480-2484
E-mail: klimke@ncbi.nlm.nih.gov

Niels Klitgord
Boston University, Dept of
Bioinformatics
24 Cummington Street, 9th Floor
Boston, MA 02215
Tel: (510)-847-2192
Fax: (617) 353-4814
Email: niels@bu.edu

Daniel A. Kluepfel
USDA, Agricultural Research Service

Crops Pathology/Genetics Research Center
284 Hutchinson Hall Plant Pathology Department, UC Davis
Davis, CA 95616
Tel: (530) 752-1137
Fax: (530) 754-7195
E-mail: dakluepfel@ucdavis.edu

Min Jung Kwak
University of Science and Technology (UST)
111 Gwahangno, Yuseong-gu, Daejeon 305-806, Republic of Korea
Tel: (82) 42-860-4414
Fax: (82) 42-879-8595
E-mail: mjinvers@kribb.re.kr

Soon-Kyeong Kwon
University of Science and Technology (UST)
111 Gwahangno, Yuseong-gu, Daejeon 305-806, Republic of Korea
Tel: (82) 42-879-8215
Fax: (82) 42-879-8595
E-mail: skkwon@kribb.re.kr

Christopher Lam
Department of Biological Sciences
California State University, L.A.,
5151 State University Dr., BS-143
Los Angeles, CA 90032-8201
Tel: (323) 343-2040
E-mail:ibillgates@yahoo.com

Morgan G.I. Langille
UC Davis Genome Center
451 East Health Sciences Drive
Davis, CA 95616
Tel: (530) 752-8409
E-mail: mlangille@ucdavis.edu

Federico M. Lauro
School of Biotechnology and Biomolecular Sciences
The University of New South Wales
Sydney NSW 2052, Australia
E-mail: flauro@unsw.edu.au

Ming-Chun Lee
Department of Organismic and Evolutionary Biology
Harvard University
Biological Laboratories, Room 3079
16 Divinity Avenue
Cambridge, MA 02138
Tel: (617) 495-9137
E-mail: mclee@fas.harvard.edu

Asaf Levy
Rotem Sorek Lab
Weizmann Institute of Science
Tel: (972) 8-934-6530
E-mail: asaf.levy@weizmann.ac.il

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

Chee Kent Lim
Department of Chemistry and Biomolecular Sciences
Macquarie University
NSW, Australia
E-mail: chee.lim@mq.edu.au

Cindy M. Liu
Translational Genomics Research Institute
3051 West Shamrell Blvd., Suite 106
Flagstaff, AZ 86001
Tel: (928) 226-6364
Fax: (928) 226-6360
Email: cindy.liu@nau.edu

Cindy M. Liu
Center for Microbial Genetics and Genomic (MGGen)
Northern Arizona University
Applied Research and Development Building
Bldg 56, 3rd Floor
Flagstaff, Arizona 86011-4073
Tel: (928) 523-1943
Fax: (928) 523-4015
Email: cindy.liu@nau.edu

Hsiao-Pei Lu
Institute of Zoology and Department of Life Science
National Taiwan University
No1, Sec4, Roosevelt Road
Taipei 10617, Taiwan, R.O.C
Tel: (886) 2-33662456
Fax: (886) 2-23638179
E-mail: par1206@hotmail.com

Athanasios Lykidis
Genome Biology Program
DOE-Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: (925) 296-5842
Email: alykidis@lbl.gov

Colin Manoil
University of Washington
Department of Genome Sciences
Foege S243D, Box 355065
1705 NE Pacific St
Seattle WA 98195-5065
Tel: (206) 543-7795
Fax: (206) 543-0754
Email: manoil@u.washington.edu

Varun Mazumdar
Boston University, Dept of Bioinformatics
24 Cummington Street, 9th Floor
Boston, MA 02215
Tel: (617) 777-4860
Fax: (617) 353-4814
Email: mazumdar@bu.edu

Shauna McGillivray,
Winton Scott Hall, Room 432
2800 South University Drive
Texas Christian University
Fort Worth TX 76129
Tel:(817) 257-6178
Email: s.mcgillivray@tcu.edu

Diane Meas
Department of Biological Sciences
California State University, L.A.
5151 State University Dr., BS-143
Los Angeles, CA 90032-8201
Tel: (323) 343-2040
Email:dianemeas@gmail.com

Tony Michael
University of Texas Southwestern Medical School
Dept. of Pharmacology
T. Boone Pickens Biomedical Research Building
6001 Forest Park Rd.
Dallas, TX 75390-9041
Tel: (214) 648-4170
Email:
Anthony.michael@southwestern.edu

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

Jeffrey H. Miller
University of California, Los Angeles
Molecular Biology Institute
Jeffrey H. Miller Lab
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 825-8924

E-mail: jhmiller@microbio.ucla.edu

William Miller
USDA, ARS, PWA, WRRC
800 Buchanan St.
Albany, CA 94710
Tel: (510) 559-5992
Fax: (510) 559-6162
E-mail: William.miller@ars.usda.gov

Harish Nagarajan
Systems Biology Research Group
9500 Gilman Drive MC 0419
University of California San Diego
La Jolla, CA 92093-0419
Tel: (858) 822-1144
E-mail: nh@ucsd.edu

Parwez Nawabi
Genome Biology Program
DOE-Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: (925) 927-2570
E-mail: pnawabi@lbl.gov

Angela Nieh
University of California Los Angeles
Dept. of Microbiology
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 206-7926
E-mail: anieh@ucla.edu

Max Oeschger
LSUHSC
Dept. of Microbiology
1901 Perdido St.
New Orleans, LA 70112
Tel: (504) 220-2616
Fax: (504) 568-2918
E-mail: moesch@lsuhsc.edu

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

Andrei Osterman
Bioinformatics and Systems Biology
Sanford-Burnham Medical Research Institute
La Jolla, CA
Tel: (858) 646-3100 x5296
Fax: (858) 795-5249
E-mail: osterman@sanfordburnham.org

Donald Petrey
Columbia University
Dept. of Biochemistry & Molecular Biophysics
Room 812B
Tel: (212) 851-4656
Fax: (212) 851-4650
E-mail: dsp18@columbia.edu

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

Lance Price
TGen Center for Microbiomics and Human Health
Translational Genomics Research Institute
Division of Pathogen Genomics
3051 West Shamrell Blvd., Suite 106
Flagstaff, AZ 86001
Tel: (928) 226-6371
Fax: (928) 226-6360
Email: lprice@tgen.org

Elisabeth A. Raleigh
New England Biolabs
240 County Road
Ipswich, Massachusetts 01938-2723
Tel: (978) 380-7238
FAX: (978) 921-1350
E-mail: raleigh@neb.com

David C. Ream
Dept. of Microbiology
Miami University
501 E. High St.
Oxford, OH 45056
E-mail: reamdc1@muohio.edu

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

Susanna Repo
Department of Plant and Microbial Biology
University of California, Berkeley, CA 94720-3102
E-mail: srepo@compbio.berkeley.edu

Richard J. Roberts
New England Biolabs
240 County Road
Ipswich, MA 01938-2723
Tel: (978) 380-7405
Fax: (978) 380-7406
E-mail: roberts@neb.com

Kenn Rudd
Department of Biochemistry and Molecular Biology (R-629), P.O. Box 016129 University of Miami School of Medicine
Tel: (305) 243-6055
Fax: (305) 243-3065
E-mail: krudd@med.miami.edu

Surya Saha
Cornell University
334 Plant Science Building
236 Tower Road
Ithaca, NY 14853
Tel: (662) 312-3227
Email: ss2489@cornell.edu

Nina R. Salama
Fred Hutchinson Cancer Research Institute
Human Biology Division
University of Washington
P.O.Box 19024 Seattle
Washington 98109-1024
Tel: (206) 667-1540
Fax: (206) 667-6524
Email: nsalama@fhcrc.org

Sahar Salehi
University of California, Los Angeles
Molecular Biology Institute
Jeffrey H. Miller Lab
1602 Molecular Science Building
Los Angeles, CA 90095
Tel: (310) 825-8924
E-mail: saharsalehi@ucla.edu

Erin R. Sanders
Department of Microbiology, Immunology & Molecular Genetics
UCLA
1602 Molecular Science Building
609 Charles E. Young Dr. East
Los Angeles, CA 90095-1489
Office: 2839 Slichter Hall
Tel: (310) 825-1783
Email: erinsl@microbio.ucla.edu

Alexandra Schnoes,
University of California, San Francisco
Byers Hall MC 2550

1700 4th St.
San Francisco, CA 94158-2330
Tel: (415) 502-1248
Fax: (415) 514-9656
Email: alexandra.schnoes@ucsf.edu

Julia A. Segre
National Human Genome Research Instititue, NIH
49 Convent Dr.
Bldg 49, Room 4A26, MSC 4442
Bethesda, MD 20892-4442
Tel: (301) 402-2314
Fax: (301) 402-4929
E-mail: jsegre@nhgri.nih.gov

Valery I. Shestopalov
University of Miami Miller School of Medicine
Bascom Palmer Eye Institute
Department of OphthalmologyDepartment of Anatomy and Cell Biology
McKnight Research Bldg- D880; 1638 NW 10th Ave, Miami, FL 33136
Tel: (305) 547-3680
Email: vshestopalov@med.miami.edu

Isba Silva
Department of Biological Sciences
California State University, L.A.
5151 State University Dr., BS-143
Los Angeles, CA 90032-8201
Tel: (323) 343-2040
E-mail: isbasilva@yahoo.com

Kimmen Sjolander
University of California, Berkeley
Dept. of Bioengineering
308C Stanley Hall # 1762
Berkeley, CA 94720-1762
Tel: (510) 642-9932
Fax: (510) 642-5835
E-mail: kimmen@berkeley.edu

Magdalene So
Dept of Immunology
BI05 Institute
University of Arizona
Keating Bldg, Rm 224
1657 E. Helen St.
Tucson, AZ 85721
Tel: (520) 626-3097
Fax: (520) 626-2661
Email: somaggie@gmail.arizona.edu

Bruno Sobral
Virginia Bioinformatics Institute at Virginia Tech
Blacksburg, VA
Tel: (540) 231-2317
Email: sobral@vt.edu

Yuli Song
The Proctor & Gamble Company
8700 Mason Montgomery Rd.
MBC DV2-5K1
Mason, OH 45040
Tel: (513) 622-4861
Fax: (513) 945-2158
E-mail: Song.y.7@pg.com

Martin Steffen
Proteomics Core Facility
Dept of Pathology and Biomedical Engineering
Boston University School of Medicine
670 Albany Street, Rm 512
Boston MA, 02118
Tel: (617) 826-9486
Fax: (617) 638-4085
E-mail:steffen@bu.edu

Justina Tain
Department of Biological Sciences
California State University, L.A.
5151 State University Dr., BS-143
Los Angeles, CA 90032-8201
Tel: (562) 304-8755
E-mail: jtain@calstatela.edu

Isabelle Toesca Spears
University of California, Los Angeles
Dept. of Microbiology, Immunology and Molecular Genetics
BSRB Room 254
615 Charles E. Young Dr. South
Los Angeles, CA 90095
Tel: (310) 825-0986
E-mail: toesca@hotmail.fr

Tatiana Tatusova
NCBI/NLM/NIH
2000 Rockville Pike
Bethesda, MD 20892
Tel: (301) 435-5756
Email: Tatiana@ncbi.nlm.nih.gov

Shuta Tomida
UCLA David Geffen School of Medicine
570 Westwood Plaza, Building 114
CNSI, Room 4338, UCLA
Los Angeles, CCA 90095-7227
Tel: (310) 936-6256
E-mail: shuta@ucla.edu

William Villella
Department of Microbiology, Immunology & Molecular Genetics
University of California Los Angeles
Los Angeles, CA 90095
Tel: (310) 825-1783
Email:willez@ucla.edu

Barry L. Wanner
Purdue University
Biological Sciences
915 W. Stat St.
West Lafayette, IW 47907-2054
Tel: (765) 494-8034
Fax: (765) 496-9028
Email: blwanner@purdue.edu

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

Katrine L. Whiteson
University of Geneva Hospital
Geneva, Switzerland
E-mail: katrine.whiteson@genomic.ch

Ann Womack
Center for Ecology and Evolutionary Biology
335 Pacific Hall
5289 University of Oregon
Eugene, Oregon 97403
Tel: (541) 346-4812
E-mail: womack.ann@gmail.com

Tanja Woyke
Department of Genome Sciences
Lawrence Berkeley National Laboratory/
DOE Joint Genome Institute
Walnut Creek, California
Tel: (925) 296-5839
Email: twoyke@lbl.gov

Meng Wu
Washington University in St. Louis
Center for Genome Sciences
4444 Forest Park Blvd. Room 5515,
Gordon Lab
Tel: (314) 362-3963
Fax: (314) 362-2156
E-mail: mwu@artsci.wustl.edu

Howard Xu
Department of Biological Sciences
California State University, L.A.
5151 State University Dr., BS-143
Los Angeles, CA 90032-8201
Tel: (323) 343-2188
E-mail: hxu3@calstatela.edu

Alexander Yakunin
C.H. Best Institute
112 College St. Room 72

Toronto, Canada
Tel: (416) 978-4013
Fax: (416) 978-8528
Email: a.iakounine@utoronto.ca

Aifen Zhou
University of Oklahoma
Department of Botany & Microbiology
SRTC2030, 101 David L. Boren Blvd.
Norman, Ok 73019
Tel: (405) 325-3052
Fax: (405) 325-7552
Email: Aifen.Zhou-1@ou.edu

Jizhong (Joe) Zhou
Stephenson Research & Technology
Center
University of Oklahoma
101 David L. Boren Blvd.
Norman, OK 73072
Tel: (405) 325.6073
Fax: (405) 325-7552
E-mail: jzhou@ou.edu

Wen Zou
National Center for Toxiolopical
Research/FDA
3900 NCTR Rd.
Jefferson, AR 72079-9502
Tel: (870) 543-7301
Fax: (870) 543-7662
Email: wen.zou@fda.hhs.gov