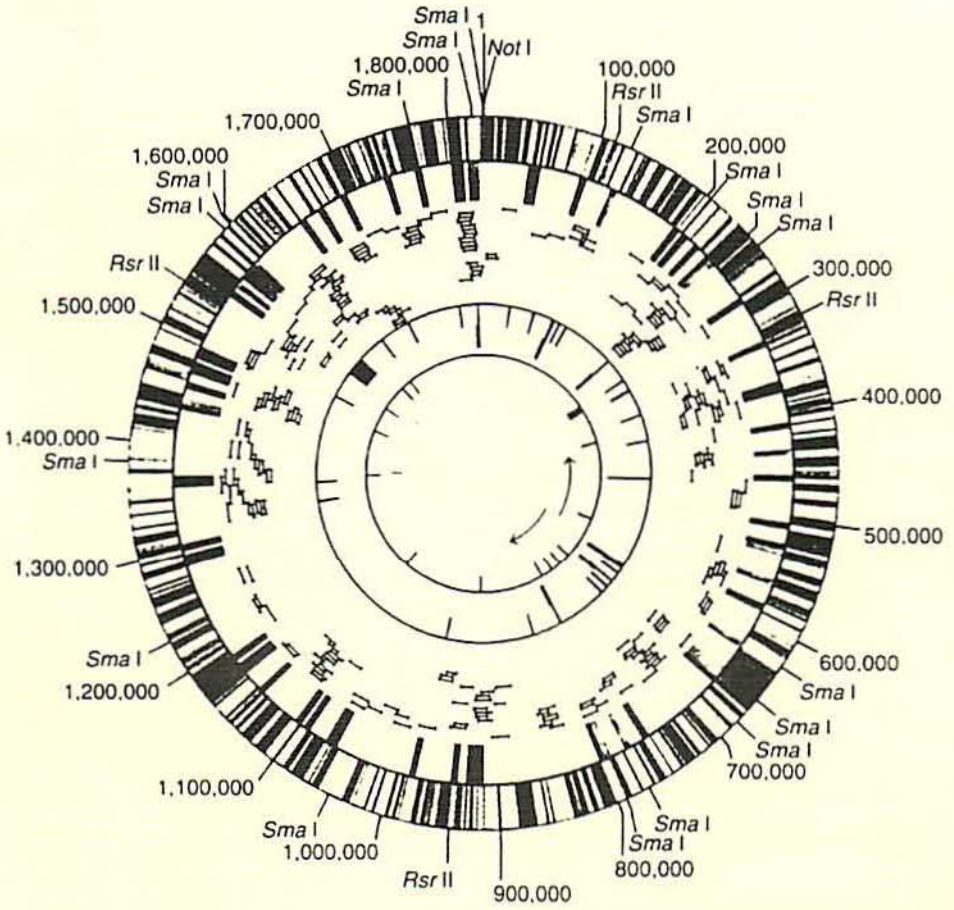
8th INTERNATIONAL CONFERENCE ON

Small Genomes



September 24-28, 2000

UCLA CONFERENCE CENTER

LAKE ARROWHEAD
CALIFORNIA

SCIENTIFIC PROGRAM ORGANIZERS

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

Dr. George Weinstock
University of Texas-Houston Health Sciences Center

Dr. Jizhong Zhou
Oak Ridge National Laboratory

Dr. Monica Riley
Woods Hole

Dr. Elisabeth Raleigh
New England Biolabs

Dr. Theresa (Terry) Gaasterland
Rockefeller University

ACKNOWLEDGEMENTS

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

Amgen

Diversa Corp.

DNASTAR Inc.

Dupont De Nemours

Genencor

Integrated Genomics, Inc.

Merck

New England Biolabs

Pharmacia & Upjohn Co.

Proctor & Gamble

SmithKline Beecham

Oak Ridge National Laboratory

National Science Foundation

Department of Energy

National Institute of Allergy and Infectious Diseases, National Institutes of Health

CONTACT NUMBER

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

SCIENTIFIC PROGRAM

SUNDAY, SEPTEMBER 24

4:00-6:00 pm	Arrival and Check-in at Lake Arrowhead Conference Center
6:15-7:45 pm	Dinner (Dining Room)
	Opening of Meeting (Pineview Room)
7:45-8:10 pm	Jeffrey H. Miller University of California, Los Angeles *Welcome*
8:10-9:00 pm	**Keynote Address** Julian Davies TerraGen Discovery, Inc. **"Evolution of Microbial Resistance to Antibiotics"**
9:00 pm	Reception (Iris Room)

MONDAY, SEPTEMBER 25

7:45-8:30 am	Breakfast (Dining Room)
Session I	**Genomes of Pathogens** (Pineview Room)
8:45-9:00 am	Introduction/Announcements
9:00-9:35 am	Stephen Bentley The Sanger Centre **"Sanger Center Pathogen Genome Projects"**

9:35-10:10 am	Najib El-Sayed The Institute for Genomic Research **"Analysis of Protozoan Pathogen Genomes at TIGR"**
10:10-10:45 am	George Weinstock University of Texas Medical School ***"Spirochete Genomes:* What Have We Learned?"**
10:45-11:05 am	Break
11:05-11:40 am	Richard Stephens University of California, Berkeley **"The *Chlamydia pneumoniae* Genome"**
11:40 am- 12:15 pm	Frederick Blattner University of Wisconsin-Madison **"Comparative Studies of Enteric Pathogens"**
12:30	Lunch (Dining Room)
4:00-6:00 pm	Poster Session (Pineview Room) Social/Mixer (Lakeview Room)
6:15-7:45 pm	Dinner (Dining Room)
Session II	**Genomes of Pathogens; Analysis of Genomes** (Pineview Room)
7:45-8:20 pm	Antonello Covacci IRIS Chiron Spa **"Analysis of the Meningococcal Genome"**
8:20-9:05 pm	Siv Andersson Uppsala University **"Analysis of the *Bartonella henselae* Genome"**

9:05-9:40 pm	Al Ivens The Sanger Centre **"Analysis of Protozoan Pathogen Genomes"**
9:40-9:55 pm	Break
9:55-10:20 pm	Peter Myler Seattle Biomedical Research Institute **"Trypanosomes"**
10:20-10:50 pm	Nancy L. Craig HHMI/Johns Hopkins School of Medicine **"Using Tn7 to Dissect Genomes"**

TUESDAY, SEPTEMBER 26

7:45-8:30 am	Breakfast (Dining Room)
Session III	**Bioremediation**
8:45-9:20 am	Terry Hazen Lawrence Berkeley National Laboratory **"Bioremediation: The Hope and the Hype"**
9:20-9:55 am	Frank Larimer Oak Ridge National Laboratory **"Genomic Survey of Key Autotrophic Microorganisms in the Natural Carbon Cycle"**
9:55-10:30 am	Frank Robb University of Maryland Biotechnology Institute **"A DNA Analysis Approach to Bioremediation and Microbial Physiology"**
10:30-10:50 am	Break

10:50-11:25 am	Margaret Romine Pacific Northwest National Laboratory *"Sphinogomonas"*
11:25 am- 12:00 pm	John Heidelberg The Institute for Genomic Research **"Genomic Sequence of *Shewanella oneidensis* MR-1"**
12:00-12:20 pm	To be announced
12:30 pm	Lunch (Dining Room)
4:00-6:00 pm	Poster Session (Pineview Room) Social/Mixer (Lakeview Room)
6:15-7:45 pm	Dinner (Dining Room)
Session IV	**Evolution of Biodiversity**
7:45-8:00 pm	Jeffrey H. Miller University of California, Los Angeles **"The Mismatch Repair System and Horizontal Transfer"**
8:00-8:35 pm	Martin Keller Diversa Corp. **"Microbial Biodiversity"**
8:35-9:10 pm	Edward F. DeLong Monterey Bay Aquarium Research Institute **"Genomic Windows into the Natural Microbial World"**
9:10-9:45 pm	Sophie Courtois Aventis Pharma **"Exploring Uncultivated Microorganisms for Natural Products Drug Discovery"**

9:45-9:55 pm Break

9:55-10:30 Jonathan Eisen
The Institute for Genomic Research
"The Evolution of DNA Repair Systems"

10:30-10:55 Monica Riley
Woods Hole
"Mechanisms of Divergence from Ancestral Enzymes"

WEDNESDAY, SEPTEMBER 27

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

Session V **Genomic Technology, Bioinformatics, and Functional Genomics**
(Pineview Room)

8:45-9:20 am Terry Gaasterland
Rockefeller University
"Integration of Genome Annotation and Microarray Data"

9:20-9:55 am Owen White
The Institute for Genomic Research
"Bacterial Annotation in a High-throughput Sequencing Environment"

9:55-10:30 am Paul Warrener
PathoGenesis Corporation
"The *Pseudomonas aeroginosa* Genome Project: Building the Pipeline from Annotation to Target Discovery"

10:30-10:50 am Break

10:50-11:25 am Gary Anderson

10:50-11:25 am	Gary Anderson Lawrence Livermore National Laboratory **"Use of Suppression Subtractive Hybridization as a Tool for Whole Genome Comparisons"**
11:25 am- 12:00 pm	Tim Palzkill Baylor College of Medicine **"Developing Phage Display for Functional Genomics"**
12:00-12:25-pm	Michael Fonstein Integrated Genomics, Inc. **"Comparative Genomics"**
12:30 pm	Lunch (Dining Room)
4:00-6:00 pm	Social/Mixer (Iris Room); *Hors d'oeuvres*
6:15-7:45 pm	Dinner (Dining Room)
Session VI	**Proteomes and Cellular Pathways**
7:45-7:55 pm	Discussion of Future Meetings
7:55-8:30 pm	Marc Vidal Massachusetts General Hospital Cancer Center **"Protein Interaction Mapping: Its Use in *C. elegans*"**
8:30-9:05 pm	Alex Bateman The Sanger Centre **"Protein Function Analysis through the PFAM Data Base"**
9:05-9:40 pm	Richard Smith Pacific Northwest Laboratory **"New Technology for the Rapid and Precise Study of Proteomes: Global Views for the Understanding of the Cellular Processes"**

9:40-9:55 pm Break

9:55-10:15 pm Edward Marcotte
Protein Pathways, and UCLA
"Genome-wide Prediction of Protein Function"

10:15-10:50 pm Ross Overbeek
Integrated Genomics
"Filling in the Missing Pieces of Core Metabolisms"

THURSDAY, SEPTEMBER 28

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

Session VII **Extremophiles**

8:30-9:05 am J. Martin Odom
Dupont
"New Insights into the Biology of an Obligate Methanotroph"

9:05-9:40 am John R. Battista
Louisiana State University and A&M College
"The *Deinococcus radiodurans* Genome and its Application to Understanding Extraordinary Resistance to Radiation"

9:40-10:10 am Sorel Fitz-Gibbon
University of California, Los Angeles
"The Fully Annotated Genome of *Pyrobaculum aerophilum*"

10:10-10:40 am Patrick Dennis
University of British Columbia
"Small Nucleolar RNAs in Archaea"

10:45 am	Check-out
11:15 am	Departure of 1st Conference Bus for LAX
12:00 pm	Lunch (Dining Room)
1:15 pm	Departure of 2nd Conference Bus for LAX

SPEAKER ABSTRACTS

Evolution of Microbial Resistance to Antibiotics

Julian Davies*. TerraGen Discovery, Inc.

Antibiotic resistance is biochemically and genetically a diverse and complex process. Two distinct genetic mechanisms can lead to a resistance phenotype: mutation or acquisition of genes. In addition, multiple mutations may be required to establish a stable resistance phenotype in either case.

Studies of antibiotic resistance development in the laboratory are useful but do not accurately reflect the situation in clinical situations. For one thing, the role played by horizontal gene transfer and the extensive range of this process is poorly understood.

What can be done about the problem of antibiotic resistance? Is there any solution other than stopping the use of antibiotics completely? Will compounds designed to be inhibitors of microbial virulence functions be a viable approach? A more complete understanding of the genetic ecology of resistance will aid in designing approaches to delay the inevitable.

Sanger Centre Pathogen Genome Projects

Stephen Bentley*. The Sanger Center

The Pathogen Sequencing Unit at the Sanger Centre has a large and expanding portfolio of microbial sequencing projects. The majority of our efforts are directed towards human pathogens, although we do have an interest in veterinary pathogens and some model organisms. At present we have completed, are sequencing, or are funded to sequence, twenty-one bacterial genomes; we have recently published the sequences of Campylobacter jejuni and Neisseria meningitidis, and completed the sequencing of Mycobacterium leprae and Salmonella typhi. The majority of our funding comes from the Wellcome Trust, via its Beowulf Genomics initiative, and all of the data from these projects is made immediately available via the World Wide Web (see http://www.sanger.ac.uk/Projects/Microbes). An update of current progress on these projects, and some interesting features of the genomes will be presented.

ANALYSIS OF PROTOZOAN PATHOGEN GENOMES AT TIGR

El-Sayed, N., Song, J., Ghedin, E., Kaul, S., Hou, L., Zhao, H., Wanless, D., Larkin, C, Gardner, M., Nene, V., Cummings, L., Peterson, J., Ullu, E., 2Melville, S., 3Donelson, J., White, O., Adams, M. and Fraser, C.

The Institute for Genomic Research, Rockville, MD, 1Yale University, New Haven, CT, 2Cambridge University, UK, 3University of Iowa, Iowa City, IA

The Institute for Genomic Research (TIGR) is a major participant in international efforts to sequence the genomes of three protozoan parasites: African trypanosomes (Trypanosoma brucei), the causative agents of African sleeping sickness in humans and cattle, American trypanosomes (Trypanosoma cruzi) that cause a fatal illness called Chagasí disease, and the malarial parasite Plasmodium falciparum, responsible for one the most dangerous infectious diseases affecting human populations. We have also initiated the sequencing of the genome of Theileria parva, a tick-borne protozoan parasite found in sub-Saharan Africa that causes a fatal disease in cattle called East Coast Fever. T. parva invades the host lymphocytes and induces the malignant transformation of the infected cells.

This talk will focus on our efforts to characterize the T. brucei genome. T. brucei has attracted much attention from the scientific community at large and become a model organism for examining a variety of general biological properties. Phenomena of fundamental biological importance first discovered in T. brucei include antigenic variation of surface glycoproteins, glycophosphatidylinositol (GPI) anchors of membrane proteins, eukaryotic polycistronic transcription, trans-splicing of precursor RNAs, mitochondrial RNA editing, unique metabolic products and pathways, and distinctive organelles such as the kinetoplast.

The sequencing of chromosome II of T. brucei is nearly completed. A preliminary annotation of the ~1.25 Mbase pairs sequence indicates that it encodes ~320 predicted genes. The genes are organized in 13-15 main clusters on one strand or the other, with as many as 100 genes lined up one after another over a 300 kb region. Strikingly, an analysis of skewed oligomers reveals strand compositional asymmetries that coincide with the distribution of the protein-coding genes, suggesting that these asymmetries may be the result of transcription coupled repair on coding vs. non-coding strand. The completion of the sequence of this chromosome provides, among other things, some new insights in regulation of gene expression and transcription by pol II promoters in T. brucei.

The *Chlamydia* Genomes

Richard S. Stephens
Division of Infectious Diseases
School of Public Health
University of California, Berkeley, CA

Chlamydia are obligate intracellular pathogens that are widely distributed in nature. Different chlamydial species infect humans, birds, numerous mammals and some invertebrates. Phylogenetically these organisms are deeply separated from other bacteria and make up a separate taxonomic division that has evolved within eukaryotic cells for millions of years. The genome sequences of six chlamydial strains have been completed. Each genome consists of approximately 1-1.2 million base pairs. Most of the differences in genome size are attributable to paralogous genes. Comparison of these genomes reveals common metabolic pathways, unexpected mechanisms of signal transduction and gene regulation, unusual DNA-binding proteins likely involved with developmental stage-specific chromatin structure, numerous novel outer membrane proteins and encoded proteins that are unique to each strain. New virulence determinants including a family of phospholipase D, type III secretion system, protein kinases and phosphatases and a toxin were discovered. Genome comparisons have revealed that the differences between strains and species that account for tissue and host tropism are due to minor differences in coding capacity and gene function. The synteny of gene organization between genomes was remarkable with most of the polymorphism localized to one region near the chromosome replication terminus. The genome sequences of chlamydiae have provided a comprehensive framework for understanding and investigating chlamydial biology that previously did not exist because of the limited growth of chlamydiae *in vitro* and the inability to genetically manipulate these organisms. This framework is now being exploited to pose hypotheses and direct experimental design that should result in a new era of productive research.

THE CAG PATHOGENICITY ISLAND OF HELICOBACTER PYLORI AND EVOLUTION OF VIRULENCE Antonello Covacci-IRIS, Chiron SpA, Via Fiorentina 1-53100 Siena, Italy

The evolution and dissemination of virulence traits between bacteria is based on mobilization of DNA by a) phage conversion, b) distant >events of horizontal transfer that promoted the pathogenicity island assembly and c)plasmid mobilization. Type III and Type IV secretory apparatuses are often associated with an increased virulence of the bacterial specialist. Type I strains of Helicobacter pylori possess the cag pathogenicity island that encodes a specialized Type IV secretion machinery activated during infection. The cag organelle is involved in cellular responses like induction of pedestals, secretion of interleukin-8 and phosphorylation of proteic targets. It has previously been reported that co-cultivation of epithelial cells with Helicobacter pylori triggers signal transduction and tyrosine-phosphorylation of a 145 kDa putative host cell protein. We and others have recently demonstrated that this protein is not derived from the host but rather is the bacterial immunodominant antigen CagA, a virulence factor commonly expressed during infection in patients of peptic ulcer disease and thought to be orphan of a specific biological function. CagA is delivered into the epithelial cells by the cag type IV secretion system where it is phosphorylated on tyrosine residues by an as yet unidentified host cell kinase and wired to eukaryotic signal transduction pathways and cytoskeletal plasticity.

Analysis of the *Bartonella henselae* genome

Siv G.E. Andersson*, Cecilia Alsmark, Asa Sjogren, Caroline Frank, Ann-Sofie Eriksson, Kristina Näslund. Department of Molecular Evolution, University of Uppsala, S-751 36 Uppsala, Sweden.

Bartonella species are reemerging pathogens with a broad animal host range, including cats, dogs, coyotes, panthers, rabbits, elk, deer and cattle. We are studying *Bartonella henselae*, the causative agent of cat-scratch disease (CSD) as well as *Bartonella quintana*, the causative agent of trench fever. The clinical syndrome of CSD is a persistent, necrotising inflammation of the lymph nodes and endocarditis. Immunocompromised patients infected with *B. henselae* or *B. quintana* develop more severe clinical manifestations, such as bacillary angiomatosis (BA) and bacillary pelioses (BP), i.e. vasoproliferative lesions resulting from a process of pathogen-stimulated angiogenesis. *B. henselae* and *B. quintana* belong to the α-proteobacteria and are closely related to the typhus causing agent *Rickettsia prowazekii*, the genome of which we have sequenced previously. Here, we present an analysis of the *B. henselae* genome with a specific focus on a comparison of the *B. henselae* and the *R. prowazekii* genomes. A first, comparative analysis of the *B. henselae* and *B. quintana* genomes will also be discussed.

The genome sequence of *B. henselae* is 1.9 Mb and the genomic G+C content is 39%. As expected, the biosynthetic and regulatory capacity is much broader in *B. henselae* than in *R. prowazekii.* Furthermore, *B. henselae* has a complete set of glycolytic genes that are not present in *R. prowazekii.* Another difference is that repetitive sequences are prominent in the *B. henselae* genome. For example, there are 2 copies of the *virB* gene operon as well as a repeated segment of 15 kb encoding proteins such as hemolysins and hemaglutinines which are of presumptive importance for virulence and invasion. A bacteriophage has also been identified in this genome. Based on the combined genetic repertoires of *Rickettsia* and *Bartonella* we have reconstructed putative gene functions of the common ancestor of the α-proteobacteria. We discuss a model for how one of the daughter lineages of this ancestor was converted into a mitochondrion.

USING TN7 TO DISSECT GENOMES

Nancy L. Craig, Ph.D.* Howard Hughes Medical Institute

Transposons have long been used in vivo to analyze gene structure and function. in vitro transposition systems are now available that can be used for this purpose. Several years ago we used in vitro Mu transposition to make insertions into a plasmid encoding the Tn7 transposition genes that allowed us to identify each gene by screening for loss of Tn7 transposition activity and insertion localization (1). Transposon insertions in plasmids generated by in vitro transposition also provide convenient substrates for priming for DNA sequencing (2). in vitro mutagenesis of whole genomic DNA in vitro can also be used to generate chromosomal insertions by transformation and recombination of the insertion into the chromosomal via homologous recombination. We used this strategy to generate a large battery of mutation in H. influenza that allowed identification of a particular competence gene (3). We are now using the strategy of in vitro mutagenesis of S. cerevisia genomic DNA. (4)

1 = Waddell, C. and N.L.Craig (1988). Tn7 Transposition: Two Transposition Pathways Directed by Five Tn7-encoded Genes. Genes & Devel. 2:137-149.

2 = Oppon JC, Sarnovsky RJ, Craig NL, Rawlings DE. A Tn7-like transposon is present in the glmUS region of the obligately chemoautolithotrophic bacterium Thiobacillus ferrooxidans. J Bacteriol. 1998 180:3007-12.

3 = Gwinn ML, Stellwagen AE, Craig NL, Tomb JF, Smith HO. In vitro Tn7 mutagenesis of Haemophilus influenzae Rd and characterization of the role of atpA in transformation. J Bacteriol. 1997 179 :7315-20.

4 = Bachman, N., M. Biery, N.L. Craig and J. Boeke

BIOREMEDIATION: THE HOPE AND THE HYPE

Dr. Terry C. Hazen, Head, Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Bioremediation has proven to be one of the most cost effective and environmentally sound remediation technologies available at sites where it will work. Though the petroleum industry has been using bioremediation to handle oil sludges (petroleum land farming) for more than 50 years, and a patent was issued for in situ bioremediation of gasoline spills in 1974, this technology is perceived as being "new". Indeed, the first patent on life, a precedence setting court case, was an oil degrading bacteria patented by GE. A plethora of new strategies have shown that chlorinated solvents, PAHs, PCBs, UXO, metals, and radionuclides can be bioremediated, biotransformed, or bioimmobilized. These techniques include passive and active aeration, injection of various electron donors and acceptors, slow oxygen releasing compounds, chelating agents, surfactants and coupling with intrinsic processes (natural attenuation). In fact, a number of companies are importing contaminant biodegraders from Russia and other countries. Several companies and institutes are also actively engaged in research and development on genetically modified microorganisms to be used as primary agents and sensors for in situ bioremediation. A number of issues are emerging that have implications for use of bioremediation to environmental cleanup, eg. release of non-indigenous species, release of genetically modified organisms, horizontal and vertical gene transfer, etc.

Genomic Survey of Key Autotrophic Microorganisms in the Natural Carbon Cycle

Frank Larimer
Oak Ridge National Laboratory
Oak Ridge, TN 37831
larimerfw@ornl.gov

The DOE Joint Genome Institute has established a program to obtain the complete genome sequence of microorganisms that may significantly impact global climate. This program supports the new DOE Global Carbon Management and Sequestration initiative, which funds basic research aimed at understanding factors that contribute to global warming and effective ways to manage carbon (particularly carbon dioxide) in soil and ocean cosystems. The goal of this effort is to explore the role of diverse microorganisms in carbon cycling by elucidating their genetic content to identify metabolic pathways that allow these organisms to adapt to their respective niches. These specialized processes include nutrient-uptake systems, pathways that contribute to nitrogen fixation and carbon cycling in soils, and pathways that regulate photosynthesis. Genome sequencing is focused initially on five microorganisms: *Nitrosomonas europaea, Rhodopseudomonas palustris, Nostoc punctiforme*, and two marine cyanobacteria, *Prochlorococcus marinus* and *Synechococcus*. The common trait shared by these microbes is that all are autotrophic (i.e., they fix C02 as their sole carbon source), are fairly numerous within their respective ecosystems, and contribute materially to carbon cycling or biomass production (with the exception of *N. europaea*). Draft sequence and annotation can be accessed at http://spider.jgi-psf.org/JGI_microbial/html/ and http://genome.ornl.gov/microbial/

A DNA analysis approach to bioremediation and microbial physiology

Frank T. Robb
Center of Marine Biotechnology, University of Maryland.

The recent publication of multiple complete microbial genomes has made it possible to contemplate an environmental genomic approach to bioremediation. In this paper, the use of unique sequences as identifiers of microbial strains will be described.. A database of sequences derived from the 16SrRNA and 16S-23S rRNA intergenic spacer region (ISR) has been assembled by amplifying bacterial DNA from contaminated Hanford groundwater (contaminated with carbon tetrachloride) and Lawrence Livermore (tricholoroethylene and dichloroethylene). ISR sequence distribution indicates that most sequences/species are unique to the given stage of remediation. Hybridization assays with immobilized oligonucleotide probes can detect sequences indicative of bioremediation after PCR amplification of ISR sequences. Bioremediation using acetate and nitrate injection caused a radical shift in community composition, and therefore a small (4 X 4) array is adequate to distinguish this effect, indicating that the ISR are useful for devising probes. Species-specific oligonucleotide probes can potentially be used to map the extent of underground pollution and of bioremediation. The sequence database from samples at Lawrence Livermore National Lab Site 300 has been expanded, and a new detection method, bead hybridization, has been applied to increase the precision of measurement and throughput of sample processing.

The predictions of microbial physiology that follow on from phylogenetic analysis is relatively unsatisfactory. The ongoing analysis of the genome of a carbon monoxide oxidizing bacterium, Carboxydothermus hydrogenoformans, provides an example of a Gram negative bacterium with a significant genetic content apparently acquired from methanogenic Archaea. The disconnection of phylogeny from function serves as a warning that the way forward in predicting the metabolic potential and the effectiveness of microbial population in bioremediation is though extensive sequence analysis coupled with functional genomic analysis of physiology.

Uniqueness of Aromatic Catabolic Gene Organization in *Sphingomonas aromaticivorans* F199

M.F. Romine.* Pacific Northwest National Laboratory, Richland.

We recently reported the complete sequence of a 184-Kb conjugative plasmid, pNL1, from *Sphingomonas aromaticivorans* F199 that encodes genes necessary for the catabolism of numerous aromatic compounds over roughly one-third of its circumference. The remainder encodes genes that function in replication, conjugation, and gene evolution (DNA rearrangement and recombination). The arrangement of the aromatic catabolic genes is unique in that they are distributed over approximately 21 transcriptional units and in many instances occur together with genes associated with different aromatic catabolic pathways. Complete degradation of aromatic substrates known to support growth of F199 are predicted to require expression of at least six to eight different transcripts.

Similarly large aromatic catabolic regions from surface *Sphingomonas* isolates have been sequenced and found to encode genes that possess extensive homology to pNL1-encoded genes. Remarkably, these similar genes are all in the same order and on the same DNA strand as their counterparts on pNL1. These regions differ from pNL1 in that portions have been deleted and in some instances replaced by insertion sequences. These findings suggest that these related aromatic catabolic regions have arisen through genetic rearrangement of DNA in an ancestral strain whose genetic composition is closely related to F199.

S. aromaticivorans F199 was isolated from Middendorf aquifer sediments collected at a depth of 410 meters below the ground surface near DOEs Savannah River Site. The age of these sediments and permeating groundwater is predicted to be approximately 100 million and 4,000 years old, respectively. This added with low nutrient conditions and the slow rates of *in situ* metabolism and growth suggests that these organisms have not undergone significant genetic rearrangement since the time of sediment deposition. We hypothesize that the current gene arrangement in strain F199 is genetically stable and is optimized for catabolism of a wide range of aromatic compounds that occur naturally (e.g., from lignite) in it's native environment.

Genomic windows into the natural microbial world.

Edward F. DeLong, Monterey Bay Aquarium Research Institute Moss Landing Ca. 95039

Cultivation-independent molecular surveys have recently revealed a vast array of diverse, novel, and environmentally abundant microbial species new to science. But the biology of these microbes is not known, because they cannot be cultivated. Genomic approaches are now providing a window into this

world, and yielding significant insight into genome organization, structure, and content of uncultured microorganisms. Environmental Bacterial Artificial Chromosome (BAC) libraries now provide the raw material necessary for dissecting the genomes and reconstructing the biochemical pathways of naturally-occurring,
uncultured microorganisms. Preliminary examination of genome fragments from such mixed population genome libraries has revealed the presence of new genes in bacteria, for instance a new bacterial rhodopsin. Functional analysis is possible now for these uncultivable microbes, and we now show that this marine bacterial rhodopsin is a light driven proton pump that is operational when expressed in host cells. The provide evidence for a novel type of phototrophy in the sea. Global analysis of environmental BAC libraries has potential to reveal higher order population level genomic trends, and provides comparisons between populations occurring in different environmental contexts.

Exploring Uncultivated Microorganisms for Natural Products Drug Discovery.

Sophie Courtois (1,2)*, Paul R. August (2,3), Carmela M. Cappellano (2), Trudy H. Grossman (3), Pascale Jeannin (2), Kara. A. Loiacono (3), Berkley A. Lynch (3), Ian A. MacNeil (2,3), S. Narula (3), Marcia S. Osburne (2,3), Jean-Luc Pernodet (4), Pascal Simonet (1), Choi Lai Tiong Yip (2,3), A. Tyler (5). Laboratory of Microbial Ecology, CNRS- University of Lyon 1, Villeurbanne, France (1) ; Aventis Pharma, Cambridge MA and Vitry-sur Seine, France (2) ; Ariad Pharmaceuticals Inc, Cambridge MA (3) ; Laboratory of Molecular Biology and Genetic, CNRS- University of Paris Sud, Orsay (4) ; Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA (5)

The rich diversity of terrestrial microorganisms has been a prolific source of clinically useful natural products. Over the last decade it has become apparent that the vast majority of microorganisms in environmental niches such as soil have not been cultivated, providing an enormous potential resource for novel natural product drug discovery. As a pilot program to further explore the utility of this untapped resource, we created several *E. coli* libraries containing soil DNA shotgun-cloned into various vectors. These recombinant libraries were analyzed for a number of properties, including microbial diversity, insert size, and expression of soil-derived genes in the heterologous host strain. Our results, to be presented, confirm that soil DNA derived from local sources contains spectacular microbial diversity. Further, we show that some soil-derived genes are expressed in *E. coli*, enabling the detection of heterologous enzymes and small molecules. Our data support the idea that this approach to accessing the genomes of uncultivated microorganisms has the potential to greatly enhance natural product drug discovery efforts.

Nucleotide Sequence of the Biphenyl/4-Chlorobiphenyl Catabolic Transposon Tn*4371*. Ariane Toussaint (1,3), D.Springael (2), C.Merlin (3), M-e-T.Hassan (1), C.Wyndham (4), S.Mouz (3) and M.Mergeay* (5,2).(1) ULB,Brussels, Belgium; (2)Vito, Mol, Belgium; (3) Université J.Fourier,Grenoble, France;(4) University of Toronto,Canada;(5)SCK-CEN,Laboratory for Microbiology, Mol, Belgium.

Tn*4371* is a 55 kb transposable element carrying a gene cluster (*bphSEGFA$_1$A$_2$A$_3$BCDA$_4$R*) involved in the metabolism of biphenyl and 4-chlorobiphenyl. The mobile element was originally identified in strain *Ralstonia eutropha* A5, a biphenyl utilizing isolate from PCB contaminated sediments of a Tennessee lake. In contrast with other catabolic transposons, Tn*4371* does not belong to the class I or class II transposons; it rather appeared as a composite transposon combining an enteric phage-like integration system, RP4/Ti-like conjugation genes and *Pseudomonas*-like catabolic genes . We now report the complete sequence of the 55 kb transposon that is a good system to study the mechanisms involved in the dissemination of the *bph* genes.

On one side of the *bph* gene cluster we identified two *orf's* whose conceptually translated products were similar to *E. stewardii* plasmid pSW100 RepA and *P. alcaligenes* plasmid pRA2 ParA proteins. These were followed by an *orf* whose product would be similar to RK2 plasmid TraF. A large cluster of genes whose products showed significant similarity with conjugative transfer genes (TraG/VirD4 and some the Trb proteins of the Tra2/VirB core from plasmids RP4/Ti) was found on the opposite side of the *bph* genes and constituted Tn*4371* right end. It thus appears that the *bph* genes may have inserted into a "conjugative transfer operon". A weird "mosaic" *orf* that appeared as a combination of an *E. coli* phage tail fiber protein, the MocB protein of *Bacteroides fragilis* conjugative transposon Tn*4399* and the Gp9A protein of bacteriophage phi-C31 laid between the *traF*-like *orf* and the first *bph* gene. It could be the remnant of a phage genome which also contributed the previously identified *int* and adjacent genes (including equivalents of *recF* and *uvrD*) located at the Tn*4371* left end.

Comparative Expression Analyses Between Distantly Related Archaeal Hyperthermophiles

T.M. Lowe*, S. Fitz-Gibbon (2), J.H. Miller (2), J. DiRuggiero (3) P.O. Brown* & D. Botstein*

*Stanford University, Stanford, CA; (2) University of California, Los Angeles; (3) Center of Marine Biotechnology, University of Maryland, Baltimore, MD.

We have begun a project to measure global mRNA expression of two archaeal hyperthermophilic species using DNA microarrays. The two species, *Pyrococcus spp.*, a euryarchaeote, and *Pyrobaculum aerophilum*, a crenarchaeote, are distantly related, yet both have adapted to life at near-boiling temperatures. Structural biologists have studied individual proteins from hyperthermophiles for clues to their thermostability, although much less is known about the global complement of proteins and RNAs involved in survival at high temperatures. We wish to identify genes important in hyperthermophily by identifying the intersection of genes between *Pyrococcus* and *Pyrobaculum* whose expression is modulated by changes in growth temperature, DNA damage, and other stresses associated with the extreme growth conditions. We initiated this project by constructing test microarrays containing several hundred spotted cDNAs geared towards detection of two aspects of hyperthermophily: (1) small RNA genes which we suspect confer ribosome thermostability by methylation of ribosomal RNA, and (2) DNA repair genes. The initial experiments involve growing both species at 5 degree increments, between 80-105 degrees Celsius, and comparing expression changes of small RNAs, of which over 50 have been identified in each genome. For DNA repair, *Pyrococcus furiosus* will be irradiated with 1000 and 2500 Gy gamma radiation and a mixture of 150 known or predicted DNA repair mRNAs will be surveyed. Once we have successfully collected data from these pilot experiments, we will construct full-genome microarrays and proceed to test a more diverse set of growth challenges.

Bacterial Genome Structure as a Phylogenetic Marker

Shu-Lin Liu*, Gui-Rong Liu, Kenneth E. Sanderson, and Randal N. Johnston. University of Calgary, Calgary, AB, Canada

Classification of bacteria based on their phylogenetic relationships has been the goal of many scientists for over a century and is still a challenging issue. During the last two decades, rRNA sequences have been used for phylogenetic studies of prokaryotes and have revolutionized this field of study. However, the usefulness of rRNA sequences is limited by their small information contents. As a result, rRNA loses resolution at the levels of genus and below and often can barely distinguish bacteria that diverged no more than 100 million years ago. In addition, rRNA sequences cannot reveal the evolutionary events that may have contributed to the speciation of the bacteria. We are making systematic genomic comparisons by mapping among different bacteria and found: 1. bacterial genome structure is relatively stable in evolution as exemplified by *Salmonella, E. coli* and *Klebsiella*; 2. closely related bacteria have similar genome structures, as exemplified by *Salmonella* spp. and distantly related bacteria have dissimilar genome structures as exemplified by *Pasteurella* spp.; 3. genome structure reveals possible evolutionary events that may have contributed to bacterial speciation as exemplified by *S. typhi*; 4. genome structure has a "clear-cut" nature, distinguishing phylogenetic groups even among very closely related bacteria as exemplified by the 8 subgenera of *Salmonella*; 5. genome structure directly reveals genome size as exemplified by *Klebsiella* (6200 kb), *Salmonella* (4800 kb) and *Morganella* (3800 kb), which is convenient for genome size evolution studies; and 6. bacterial genome structure is easily determined by physical mapping, e.g., genome structure of over 100 bacterial strains could be determined within a week by one researcher using one PFGE machine. Therefore, bacterial genome structure may be used as a reliable and convenient phylogenetic marker for studies of bacterial evolution and speciation and, most importantly, for the eventual establishment of a phylogeny-based bacterial taxonomy system.

Discovery of Fur Binding Sites at the *Escherichia coli fhuF* Promoter by an Information Theory Model.

Karen A. Lewis* [1,2], Bernard Doan[3], Ming Zheng[3,4], Gisela Storz[3], Thomas D. Schneider[1,5].

Eleven *Escherichia coli* Fur binding sites, experimentally confirmed by DNase I footprinting, were used to create an information theory model of Fur binding. The model assigns a bit value to each base at each position according to its frequency in the data set. This weight matrix can then be scanned across any DNA sequence to evaluate the information content of the sequence. To test the model, the weight matrix was used to scan the 11 footprinted sequences. Sequence walkers, which are visual depictions of predicted binding sites, appeared in every region. In 8 out of the 11 regions, sequence walkers appeared in clusters, with individual walkers within the clusters being separated from its neighbors by exactly 6 bases. In the other 3 regions, only one sequence walker appeared. Overall, the sequence walkers fit the published footprint data closely, indicating that the model could accurately predict Fur binding. The *E. coli* genome was then scanned with the weight matrix to find additional sequences to which the Fur protein would most likely bind. 38 sites were found with information content greater than 15.0 bits; 33 of these displayed the clustering of sequence walkers as observed in the footprints, while 5 regions contained a single sequence walker. The promoter region of *fhuF* contained the walker with the highest information content in the genome. Previous studies have indicated Fur binding to this promoter, but the range of binding had not been established. We found that the binding sites predicted by sequence walkers were completely within the DNase I region protected by Fur. Thus, our footprinting experiments confirmed the predictions made by molecular information theory.

[1]National Cancer Institute, Frederick Cancer Research and Development Center, Laboratory of Experimental and Computational Biology, P. O. Box B, Frederick, MD 21702-1201 (301) 846-5581 (-5532 for messages), fax: (301) 846-5598.

[2]Catoctin High School, 14745 Sabillasville Rd, Thurmont, MD 21788; present address: 28 Westhampton Way, University of Richmond, VA 23173.

[3]National Institute of Child Health and Human Development, Cell Biology and Metabolism Branch, Building 18T, Room 101, Bethesda, MD 20892-5430 (301) 402-0968, fax: (301) 402-0078.

[4]present address: Dupont Central Research and Development, Experimental Station E328-B31, P.O. Box 80328, Wilmington, DE 19880-0328, (302) 695-7136.

[5]Corresponding author. toms@ncifcrf.gov; http://www.lecb.ncifcrf.gov/~toms/

The *Rhodobacter capsulatus* genome project.

Alla Lapidus*, Mikhail Mazur#, Yakov Kogan#, Randal Cox#, Ross Overbeek#, Evgeni Selkov#, Pavel Ulbrich$, Jan Paces$, Robert Haselkorn, Michael Fonstein#. Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58 Street, Chicago, IL 60637 USA. $ Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingova 2, CZ-16637 Prague 6, Czech Republic. #Integrated Genomics, Inc., 2201 W. Campbell Park Drive, Chicago, IL 60612 USA.

The genome sequencing project was started with an organized library, which consisted of 192 cosmids covering the 3.7-Mb chromosome and the 0.13-Mb plasmid of *Rhodobacter capsulatus SB1003*, a purple, non-sulfur photosynthetic bacterium. The ease of plating and generation of mutations, together with convenient systems for cloning and genetic analysis, has made *Rhodobacter capsulatus* a popular model system for studies of photochemical reaction centers and nitrogen fixation.

The sequencing project was organized as a combination of random cloning and primer walking using an ordered cosmid encyclopedia, in which different areas of the chromosomal map were represented by individual sub-projects. A set of contiguous DNA sequences of different quality, totaling nearly 3 Mb, was generated during the first phase of the project. The current goal was to complete the genome sequence and to establish a framework for its functional analysis. Nearly 1 Mb of new DNA sequence was generated during the course of the latest project. This sequence, together with the rest of the *Rhodobacter capsulatus* genome, has been loaded into the WIT-Pro system for genome analysis (**http://wit.IntegratedGenomics.com/IGwit/**). Sequence similarities, relative gene positions on the chromosome and placement of gene products in metabolic pathways were all used for functional gene assignments and for the development of a functional overview. Examples of gene assignments together with their verification by independent approaches will be presented.

A survey of global gene expression of *Escherichia coli* under antibiotic stress.

Hongfan Jin*, Bernard Wisblum, Jeremy Glasner, Robert Mau, Yu Qiu, Chris Herring, Mingzhu Liu & Frederick Blattner. University of Wisconsin-Madison

The rapid global expansion of bacteria resistant to antimicrobials has become a major medical and public problem. The emergence of resistance is likely the result of the widespread use or overuse of antimicrobial agents. A better understanding of the molecular basis of the interaction between the antibiotic and the-microorganism may aid the development of new drugs with longer lasting antimicrobial activity. The availability of whole genome sequences and the advent of DNA microarray technology allows analysis of the consequences of antibiotic action at the genomic scale. We used a DNA microarray containing 97% of the open reading frames (ORFs) predicted from the complete *E. coli* sequence to monitor changes in *E. coli* genes' transcripts abundance in response to the antibiotics Erythromycin and Chloramphenicol, which both inhibit 50S ribosomal function during protein synthesis. These antibiotics showed similar expression profiles in spite of their different structures. We found members of *pho* regulon, *pho*RB, and *pst*SCAB are highly induced, possibly involved in active efflux of the drugs. Additional induced transcripts include IS and phage encoded ORFs, likely due to general stress response, as well as a number of ORFs that encode "hypothetical" proteins. Interestingly, chemtaxis and flagella related genes appear repressed, including *che*BZRY, *fli* and *flg* operons. In addition, several genes encoding heat shock proteins are repressed, including *mop*AB, *dna*J, and *hsl*U. Many aspects of basic cellular metabolism are repressed including amino acid biosynthesis, energy metabolism and the biosynthesis of cofactors. Insights gained from this approach will lead to greater understanding of the nature of drug activity and the mechanisms of antibiotic resistance.

Functional Genomics of *Pseudomonas putida* KT2440: iron deficiency.

S. Heim*[1], H. Hilbert[2], K. E. Nelson[3], K. N. Timmis[1]. [1]GBF-German Research Centre for Biotechnology, Braunschweig, Germany; [2]Qiagen GmbH, Hilden, Germany; [3]The Institute for Genome Research (Tigr), USA.

Pseudomonas putida is an ubiquitous, metabolically and physiologically extremely variable soil bacterium and can be used as a biocontrol agent for plant pathogens. The sequencing of the 6.1 megabase *Pseudomonas putida* KT2440 genome, as the central part of a BMBF- and D.O.E.-supported project, is nearly completed and a computer-based sequence data analysis is in progress to identify the genetic elements and to classify the coding sequences according to their functions, protein families, motifs and homologies. At this stage, no significant homologies for more than 35 % of all putative coding regions to known sequences were found. The functions of these orphan genes will be elucidated by phenotypic analysis of Tn*5*-mediated transposon mutants in combination with proteomic studies of *P. putida* in defined media and under exposure to various stress conditions.

P. putida is able to recruit iron from its environment. As a good coloniser of plant roots this ability may be advantageous for the plants. On the one hand, possible plant pathogens are deprived of iron, which could result in protection of the plants. Additionally some siderophores exhibit antifungal activities. On the other hand, *P. putida* may make iron more easily available for the plants which should promote plant growth. Therefore, we chose the influence of iron limitation as first approach for the functional analysis of the *Pseudomonas putida* KT2440. The consequences for protein synthesis and expression, respectively, were investigated by 2D-electrophoresis combined with radioactive protein labelling with ^{35}S-methionine. Proteins of interest were analysed by MALDI-TOF peptide mass fingerprint. For the identification of the proteins we used an in-house peptide mass database which was built-up based on the genomic sequence data.

The Complete Sequence of the Symbiotic Plasmid of *Rhizobium etli*

V. González*, P. Bustos, A. Medrano, J. Hernández, V. Quintero, MA Cevallos, O. Rodríguez, ML Girard, J. Collado, R. Palacios & G. Dávila. Centro de Investigación Sobre Fijación de Nitrógeno, UNAM. AP. 565-A. Cuernavaca, Mor., México 62210.

Biological nitrogen fixation is a property of free-living and symbiotic microorganisms. Although ***nif*** genes are scattered among different eubacterial and archeal subdivisions, symbiotic associations are only performed by species belonging to the Rhizobeacea family of the alpha-proteobacteria. ***Rhizobium etli***, the symbiont of bean, has a megaplasmid (pSym) which contains the great majority of ***nif*** and ***nod*** genes (those involved in root interactions). Previously, a physical map of this plasmid was obtained. Eleven overlapping cosmids that cover the entire replicon were purified and shotgun libraries were constructed by random shearing of each one. Fragments between 700 bp, to 2000 bp were selected and cloned either in M13 or pUC19 Automatic sequencing was made in an ABI-373 apparatus (Applied Biosystems) using the Big-Dye terminator method. The total assembly, based on CONSED software package, consists of 372 259 nucleotides obtained through 6 216 sequence readings of 450 bases on average. Gaps between contigs were closed with walking-primers and PCR products. An error rate of 1 nucleotide over 10000 nucleotides was taken as a quality parameter and the coverage was estimated to be 7-fold on average.

The open reading frames (ORFs) were predicted using GLIMMER, trained with different sets of known genes of ***R. etli*** and ***Rhizobium*** NGR234. 380 common ORFs to all the predictions were obtained by cross-comparison among them. BLAST searches were used to identify homologous genes in the nonredundant data base. Until now, we have assigned 51 unambiguously similar genes. This group includes the basic genetic machinery for the synthesis of the nod factor (13), the nitrogen fixation genes (17) and others related to the maintenance and transfer of the plasmid (7). In addition we have found a cluster of ORFs closely related to *vir* and nopaline utilization genes of ***Agrobacterium***, a homologue to the chaperonin ***groEL***, and a copy of the sigma factor ***rpoN***. A comparison between the pNGR234 and the ***R. etli*** pSym plasmid, showed that at least 30% of the gene content is shared but the gene order is not conserved. These findings alow us to propose a possible evolutionary pathway for the origin and diversification of these plasmids.

Genome of the uncultivated bacterial endosymbiont from the deep-sea hydrothermal vent giant tubeworm Riftia pachyptila

Robert A. Feldman*1, Horst Felbeck2, S. Craig Cary3, Shellie Bench1, Bill Prasad1, Kathy Caldwell1, Jeffrey L. Stein4, 1Molecular Dynamics, Inc., part of Amersham Pharmacia Biotech, Sunnyvale CA, robert.feldman@am.apbiotech.com, 2Univ. of Cal. San Diego, Scripps Inst. of Oceanography, 3Univ. of Delaware, Lewes DE; 4Quorex Pharmaceuticals, Carlsbad CA.

The power of high-throughput genomic technology has yet to be applied to ecological and evolutionary studies. To date, the massive public and private genomics efforts have concentrated solely on sequencing complete genomes of cultivated, isolated organisms. While these studies are leading a revolution in biology, they will do little to further our understanding of the inter-connections between organisms. Work in our laboratories is geared toward generating large scale sequence data from, and conducting genomic analyses of uncultivated microbes and sequencing complete microbial genomes from organisms living in extreme environments.

The deep-sea hydrothermal vent ecosystems are potentially ancient remnants of earth's earliest ecosystems and have evolved largely in isolation from photosynthetically driven surface ecosystems. Powered by geo-thermal energy sources and bacterial chemoautotrophic food production, this environment may provide important clues in the search for extraterrestrial life. We are pioneering the environmental genomic study of deep-sea hydrothermal vent ecosystems. Current studies involve genomic variation of the giant tubeworm bacterial endosymbionts and genome scanning of the microbial communities associated with deep-sea hydrothermal vent polychaetes living in the highest temperature environments yet observed for eukaryotes.

The giant tubeworm of the East Pacific deep-sea hydrothermal vents, Riftia pachyptila, and its uncultivated chemoautotrophic bacterial endosymbiont exhibit a tightly coupled symbiotic relationship in which the host is completely dependent on the symbiont for nutrition. The bacterial symbionts found within Riftia and other similar tubeworms show little genetic variability across geographic and host species ranges and may exist as monocultures within a given host animal. For this reason, total bacterial DNA extracted from a given host animal should show little to no genetic variability making production of a genome sequence from this microbial community a feasible undertaking.

Modification of the IrrE Protein Sensitizes *Deinococcus radiodurans* R1 to the Lethal Effects of UV and Ionizing Radiation

Ashlee M. Earl* and John R. Battista, Department of Biological Sciences, Louisiana State University and A & M College, Baton Rouge, LA 70803

IRS24 is a strain of *Deinococcus radiodurans* carrying mutations in two loci, *uvrA* and *irrE*, rendering it sensitive to the lethal effects of UV and ionizing radiation. In order to determine the contribution that the *irrE* gene product makes to IRS24's sensitivities, we restored wild type *uvrA* and *irrE* to IRS24 using natural transformation creating three new strains, AE1011, AE2011 and AE3011, and evaluated the survival of these strains relative to their parent, IRS24, and wild-type *D. radiodurans*. Restoration of either locus restores IRS24 to near wild-type levels of UV and ionizing radiation resistance. This suggests functional overlap between the *uvrA* and *irrE* gene products. Either the *irrE* gene product catalyzes some type of excision repair or the *uvrA* gene product of *D. radiodurans* plays a previously uncharacterized role in DNA damage tolerance.

The restoration of *irrE* to radioresistance was accomplished by transforming IRS24 with any one of five clones isolated from a R1 cosmid library. Sequence alignment of these clones was useful in localizing *irrE* to an approximately 20 kb region. Utilizing the *D. radiodurans* genomic sequence available through The Institute of Genomic Research (TIGR), we were able to localize the site of the mutation further by using restriction digests of the 20 kb region. Large linear fragments obtained following restriction of the 20 kb overlap were purified from an agarose gel and then dotted directly onto IRS24 and screened for their ability to restore radioresistance. This analysis allowed us to localize the mutation to a 970 bp region containing one putative open reading frame (ORF), DR0167, and 179 bp of sequence upstream. Restoration was confirmed by dot transforming IRS24 with PCR amplified fragments derived from the wild-type genome. Subsequent sequence analysis of the *irrE* allele in IRS24 revealed a transition mutation at codon 111 resulting in an arginine to cysteine amino acid substitution.

DR0167 is the first ORF in what appears to be a four gene operon containing three homologues of genes in the folate biosynthetic pathway (*folP*, *folk* and *folB*). IRS24 is, however, capable of growing in the absence of folate suggesting that the mutation in DR0167 has little effect on the downstream folate genes. Blast search analysis of DR0167 reveals only minimal similarity to proteins currently available in the databases. A "weak" helix-turn-helix (HTH) motif was identified within the protein that may indicate a capacity to bind DNA and, perhaps, a potential role for IrrE in gene regulation.

Towards the genome of the psychrophilic archaeon, *Methanogenium frigidum*

Rick Cavicchioli*[1], Torsten Thomas[1], Neil Saunders[1], Paul Curmi[2], Elizabeth Kuczek[3], Rob Slade[3], John Davis[3] and John Mattick[3]. 1. School of Microbiology and Immunology, The University of New South Wales, Sydney, Australia; 2. School of Physics, The University of New South Wales, Sydney, Australia; 3. The Australian Genome Research Facility, Brisbane, Australia.

Despite the knowledge that low-temperature adapted (psychrophilic or psychrotolerant) archaea are abundant and are likely to have important ecological roles in low-temperature environments, little is known about their molecular and physiological means of adaptation. We have initiated genome sequencing of *Methanogenium frigidum*, a methanogenic, psychrophilic archaeon originally isolated from Ace Lake in the Vestfold Hills region of the Australian Antarctic Territory where the *in situ* temperature is a constant 1-2°C. It has the lowest, upper growth temperature limit (18°C) of any isolated archaea and is unique in the field. To date, matches to over 700 genes or ORFs from genome sequences have been obtained. We are presently seeking funds/partnerships to complete the genome sequence.

Genome sequences presently exist for a hyperthermophilic methanogen (*Methanococcus jannaschii*), a thermophilic methanogen (*Methanobacterium thermoautotrophicum*) and are being completed for mesophilic methanogens (*Methanococcus maripaludis & Methanosarcina mazei*). The genome of the psychrophile will provide a key link for investigating cellular adaptation in a set of metabolically and phylogenetically similar organisms that cover the temperature spectrum from 0C to above 80C.

The genome sequence provides the opportunity for "mining" genes and their proteins that have significant commercial value. An advantage of psychrophilic proteins is that they are fully functional at low temperature and yet can be inactivated by raising the temperature. Potential applications include lipases, proteases, cellulases & amylases as biodegradable detergents; lipases, proteases and β-galactosidase for modifying food products; anti-freeze and ice-nucleating proteins for cryogenic processes; alkaline-phosphatase for molecular biology. In addition, information about the structures of cold active proteins can facilitate protein engineering to construct enzymes with specific catalytic properties plus the ability to function in the cold. The number and type of applications are numerous and diverse; many of which will only be realised after the genome sequence is completed.

The vision of the research is not limited to the genome sequence itself but extends to biochemical characterisation, functional genomics (including proteomics) and structural genomics. In addition, The Australian Centre for Extremophiles is presently being formed to reflect a consolidated effort by a number of Australian scientists to realise the potential of Australias natural biota and to enhance the focus of our international research collaborative efforts.

Analysis of the Mitochondrial Genome of *Tetrahymena thermophila.*

C. F. Brunk*, A. Tran & J. Li. Biology Department, University of California, Los Angles, CA

The mitochondrial genome of the ciliate, *Tetrahymena thermophila,* is a linear molecule 47,460 base pairs in length (exclusive of the telomers). There are 22 putative proteins (the *nad 9* gene, is duplicated in tandem and the *nad 1* gene is split) and 22 additional open reading frames (orfs). This mitochondrial genome contains a large and small ribosomal RNA (both of which are split into two rearranged segments) and 7 unique t RNAs (the leucine t RNA is duplicated). There are inverted repeats at each end of the genome that include the large ribosomal RNA genes and the leucine tRNA. These repeats are virtually identical, indicating concerted evolution. The genome is very compact with putative proteins, orfs and RNA genes comprising over 95% of the sequence. The codon UGA codes for tryptophan and UAA is the sole termination codon (UAG is not used).

The *T. thermophila* mitochondrial genome has a high degree of sequence similarity with the *T. pyriformis* mitochondrial genome and the gene organization is virtually identical (the only exception is the *nad 9* duplication). As is generally true of ciliate genes, the mitochondrial genes have diverged substantially from those of other organisms. In order to identify orfs that truly do code for proteins, the nucleotide substitution rate for the various orfs were computed by comparing the sequences from *T. thermophila* with those from *T. pyriformis.* Clearly the whole genomes are homologous, but if an orf codes for a protein, nucleotide substitutions which do not change the amino acid (synonymous substitutions) should predominate over nucleotide substitutions that change the amino acid (nonsynonymous substitutions). On this basis several of the orfs appear not to code for proteins. One of the largest orfs probably contains the origin of DNA replication.

Towards Robust Bacterial and Universal Trees. J.R. Brown, M.J. Italia, C. Douady, W. Marshall and M.J. Stanhope. Anti-Infectives Bioinformatics Group, SmithKline Beecham Pharmaceuticals, Collegeville PA.

Ribosomal RNA (rRNA) has been the mainstay marker molecule for microbial phylogenies for nearly 30 years. However, universal and bacterial phylogenetic trees based on a single molecule either rRNA or more recently, conserved proteins, have been consistently plagued by poor resolution, low statistical support and conflicting branching orders of major nodes. In contrast, phylogenies based on multiple protein genes have been used to resolve evolutionary relationships among rapidly evolving animal groups, such as orders of mammals, with considerable success. Herein, we describe the construction and phylogenetic analyses of concatenated amino acid alignments of over 60 orthologous proteins collected across 31 bacterial species. In addition, universal trees were constructed using over 30 protein orthologs from Archaea, Bacteria and eukaryotes. As expected, the universal tree showed strong support for the separate monophyly of the Archaea, Bacteria and eukaryotes. Phylogenetic trees of bacterial species, based on mega-alignments of over 25,000 amino acids, show strong statistical support for nearly all major nodes thereby allowing for the development of more rigorous hypotheses about the mode and tempo of bacterial evolution.

Physical and genetic mapping of the *Magnetospirillum magnetotacticum* genome.
Elizabeth Bertani* & Juliani Weko. California Institute of Technology, Pasadena.

M. magnetotacticum synthesizes single-domain crystals of the iron mineral, magnetite (Fe_3O_4), which it stores in membrane-enclosed chains called magnetosomes. Biologically it is of interest as a possible model for the process of biomineralization and for its role in the evolution of the magnetotactic response in higher organisms; geologically, for its contribution to the magnetization of sediments and for its potential as a geobiological tracer, since it leaves a detectable fossil remain; and commercially, because of the exceptionally fine quality of its single-domain magnetite crystals. Pulsed-field gel analysis of MS-1 DNA indicates that the genome is a single, circular structure of about 4.3 Mb. A few genes have been identified by sequence similarity. They include putative genes involved in DNA (*dnaA*), RNA (*rpoA*) and protein synthesis (*rpl6, rrn*), as well as cell division (*ftsH*); a gene encoding a subunit of the protein involved in dissimilatory nitrate reduction (*napA*); and two *nifL* analogs , encoding components of a nitrogen-fixation system. In addition, five (*stp*) out of the twenty-seven genes appear to be components of sensory transduction systems. These genes have been localized and arranged in a map with *dnaA*, indicating the presumed origin of replication. There are at least two rRNA operons. In addition, rRNA genes are found on a 40 kb, possibly extrachromosomal, structure. Genes thought to be involved in magnetite synthesis, *bfr* and *magA*, are located at a maximum of 17% of each other on the genome.

Characterization of *Shewanella oneidensis* MR-1 *etrA* mutant using functional genomic approach

Alex Beliaev[1]*, Dorothea Thompson[1], Gary Li[1], Carol Giometti[2], Doug Lies[3], Kenneth Nealson[3] & Jizhong Zhou[1]. [1] Oak Ridge National Laboratory†, Oak Ridge, TN; [2] Argonne National Laboratory, Chicago, IL; [3] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA.

The wide distribution and prevalence of *Shewanella* species in various natural environments can be attributed to the unique metabolic properties of these bacteria. The range of electron acceptors utilized by *S. oneidensis* MR-1 include organic compounds, such as fumarate, glycine, trimethylamine N-oxide and dimethyl sulfoxide, as well as inorganic acceptors such as Fe(III), Mn(IV), nitrate, nitrite, thiosulfate, sulfite and elemental sulfur. Despite the significant amount of data collected on the physiology of *S. oneidensis* respiration, little is known about the genes regulating the expression of the different electron-transport chain components. So far a single regulatory gene of *S. oneidensis* MR-1, *etrA*, responsible for the switch from aerobic to anaerobic growth has been identified. It was shown previously that EtrA exhibits 73.6% identity to the Fnr regulator of *Escherichia coli*. To further investigate the function of the *etrA* gene, we generated an insertional mutation in this gene. Phenotype analysis of the resulting EtrA$^-$ strain revealed no detectable differences in the utilization of all the electron acceptors compared to the wild-type, MR-1. For more detailed analysis the expression profiles of the wild-type and EtrA$^-$ strains grown under different conditions were studied using 2-dimensional (2D) gel electrophoresis and partial DNA microarrays. The wild-type and the mutant were grown under aerobic, fumarate-, nitrate- and iron-reducing conditions. Our preliminary results indicated differences in the 2D gel expression patterns of the wild-type and EtrA$^-$ mutant. Microarray studies are underway to determine the identities and changes in expression levels of the affected genes.

†Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract number DE-AC05-000R22725.

POSTER ABSTRACTS

DIVERSITY OF AUTOTROPHIC AMMONIA OXIDIZING BACTERIA

Ågot Aakra, Janne B. Utåker, and Ingolf F. Nes.
Lab. of Microbial Gene Technology, Agricultural University of Norway, P.O. Box 5051, N-1432 Ås Norway.

The chemoautotrophic ammonia oxidizing bacteria (AOB) are an important group in the nitrogen cycle, since they are the only organisms oxidizing ammonia to nitrite in significant amounts. Except for two marine strains, all AOB belong to the beta-subgroup of the *Proteobacteria.* Due to obligate chemoautotrophy, with ammonia as energy source and CO_2 as the sole carbon source, the AOB grow very slowly. Nowadays, there are three genera of AOB, *Nitrosococcus*, *Nitrosomonas* and *Nitrosospira* (formerly *Nitrosospira, Nitrosolobus*, and *Nitrosovibrio*).

A major problem when studying the phylogeny of AOB using 16S rRNA gene sequence data is low resolution resulting from the high similarity values of the 16S rDNA sequences. The need for alternative, supplementary techniques is obvious, and we have performed ribotyping (RFLP of rRNA genes) and phylogenetic analyses of the 16S-23S rDNA intergenic spacer region (ISR) and the ammonia monooxygenase gene (*amoA*). Our studies have 5 isolates of *Nitrosococcus*, 15 *Nitrosomonas* isolates and 24 *Nitrosospira* isolates.

Ribotyping revealed one copy of the *rrn* operon per genome, which has not been found in any other chemoautotrophic bacteria. Even though the ribotyping patterns had too low complexity to be used for future typing purposes, all AOB studied had unique ribotyping patterns, indicating different relative location of the *rrn* operon in their genomes. ISR-analysis, however, appears to be a promising tool in the study of AOB. The phylogeny of AOB based on the ISR sequences complements the 16S rDNA based phylogeny, but gives a much higher resolution (lower similarity levels). Furthermore, the rDNA-based phylogenies appear to be consistent with species affiliations of the AOB analysed. The *amoA*-based phylogeny is, in general, complementary to the rDNA-based phylogeny, with some important exceptions. Probably, the *amoA* is an example of gene that does not follow the "average" phylogenetic history of the AOB.

Homologs of small nucleolar RNAs in Archaea

Patrick P. Dennis[1*], Arina Omer[1], and Todd Lowe[2]
[1]Department of Biochemistry and Molecular Biology, University of British Columbia Vancouver, BC, V6T 1Z3, Canada
[2]Department of Genetics, Stanford University, Stanford, CA, 94305-5120, USA

Archaea contain genes encoding homologs of two nucleolar proteins, fibrillarin and NOP56, and the rRNA of *Sulfolobus* contains more than 60 sites of ribose methylation. To determine if Archaea have sno-like RNAs to guide methylations in rRNA, we purified the Fib and NOP56 proteins from *Sulfolobus acidocaldarius* and generated polyclonal antibodies. The antibodies were used to immunoprecipitate RNP particles from fractionated cell extracts and the coprecipitated RNAs were used to construct a cDNA library. Twenty eight clones that possessed all the hallmark features of eukaryotic C/D box snoRNAs were recovered. The presence of many of these sRNAs in total cellular RNA was confirmed by primer extension and/or Northern hybridization analysis and a number of sites of methylation within rRNA predicted by guide-target complementarity were confirmed by primer extension pause assays. To identify sRNA genes in other archaeal species, the length and sequence features of these *S. acidocaldarius* sRNAs were used to retrain a search program that had been used to identify eukaryotic C/D box RNA. The program identified from a few to more than 50 candidate C/D box sRNAs in each of seven separate species. Analysis of these genes shows the following. First, sRNA genes are generally dispersed within noncoding regions of the genomes and in only a few cases are genes in close proximity; about 25% overlap the 3' ends of protein ORFs and in one widespread instance, an sRNA gene is located within the intron of the intron-containing tRNA-Trp gene. Second, the number of the sRNA genes in a genome correlates with the optimum growth temperature -- the higher the growth temperature, the easier it was to identify candidate sRNA genes. Third, guide regions in many sRNAs exhibit sequence complementarity to various regions within tRNAs and occasionally other stable RNAs. Fourth, the guide regions within sRNAs tend to diverge much more rapidly than the corresponding rRNA targets -- this has a profound influence on sRNA gene divergence and provides clues about the roles of sRNAs in ribosome biogenesis.

The Fully Annotated Genome of *Pyrobaculum aerophilum*

S.T. Fitz-Gibbon*[12], L. Garibyan[1], V. Gabrovsky[1], U.-J. Kim[3], K.O. Stetter[4], M.I. Simon[3], J.H. Miller[1]

[1.] Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles.

[2.] IGPP Center for Astrobiology, University of California, Los Angeles, CA, 90095-1567

[3.] Biology, California Institute of Technology.

[4.] University of Regensberg, Germany

Pyrobaculum aerophilum is an extremely thermophilic (T_{max} = 104°C, T_{opt} = 100°C) and metabolically versatile member of the predominantly hyperthermophilic sulfur metabolizers, the crenarchaeota . Unlike most extreme thermophiles, *Pyrobaculum aerophilum* can withstand the presence of oxygen, growing efficiently in microaerobic conditions, thus making it relatively easy to work with in the laboratory. Unlike most of its phylogenetic neighbors, the growth of *Pyrobaculum aerophilum* is inhibited by the presence of elemental sulfur, but it grows well anaerobically using nitrate reduction.

The genome is 2,222,877 bp and codes for approximately 2600 proteins. A phase of detailed, careful annotation has recently been completed. This involved manually inspecting all open reading frames (orfs) larger than 300 bases and numerous smaller orfs if they showed high coding potential. Decisions were made for each orf based on nucleotide and amino acid alignments to other pyrobaculum orfs and public database seqeunces, as well as GeneMark coding potentials, hidden Markov model matches, and orf size and position. If the orf seemed to code for a protein, alternative start sites were considered and potential frameshifts were considered at the raw sequence data level. The most likely major function of the protein and the associated category was determined by inspection of database matches and database abstracts associated with the matches. Regions of synteny between the *P.aerophilum* genome and other completed genomes (at the protein coding level) were identified and used to help specify functions and functional categories. 43 tRNAs and approximately 50 homologs of small nucleolar RNAs have been identified. *P.aerophilum* has a particularly complex nucleotide level repeat structure. For example, one 5600 base contiguous region of the genome matches 75 other scattered regions of the genome

The *Deinococcus radiodurans* Genome, and Its Application to Understanding Extraordinary Resistance to Radiation

John R. Battista*, Louisiana State University and A & M College, Baton Rouge, LA 70803

The *D. radiodurans* genome encodes essentially the entire ensemble of DNA repair proteins found in *E. coli*. With the exception of alkylation transfer and photoreactivation, all of the major prokaryotic DNA repair pathways are represented. This observation is significant, not because it says anything about why *D. radiodurans* is radioresistant, but because it confirms something long suspected; *D. radiodurans* possesses unique mechanisms for dealing with ionizing radiation-induced DNA damage. Clearly, the collection of repair proteins identified in *D. radiodurans*, in and of itself, is not sufficient to confer radioresistance. If it were, *E. coli* would be as radioresistant. *D. radiodurans* must encode novel DNA repair proteins or, alternatively, it must use the DNA repair proteins it encodes much more efficiently than more radiosensitive prokaryotes. Either possibility suggests that there are unprecedented mechanisms facilitating this species recovery following exposure to ionizing radiation.

Assuming that the identified repair proteins encoded by *D. radiodurans* perform the same functions as their *E. coli* homologues, it seems reasonable to expect that any novel proteins, critical for ionizing radiation resistance, will promote reassembly of the genome post-irradiation. It is, after all, this ability that most clearly distinguishes *D. radiodurans* from other species. Of the 3187 open reading frames identified in *D. radiodurans* R1, only 1493 could be assigned a function based on similarity to other gene products found in the protein databases. Of the 1694 proteins of unknown function, 1002 are, at present, unique to *D. radiodurans*, showing no database match. The secret to understanding the radioresistance of *D. radiodurans* is presumably found among these proteins of unknown function.

To achieve the goal of defining the proteins necessary for the radioresistance of *D. radiodurans*, a collection of 50 ionizing radiation sensitive strains was generated using chemical mutagenesis. In addition, the genetic tools were developed that allow the rapid identification of the locus responsible for this phenotype, and that permit convenient disruption of any *D. radiodurans* gene. In characterizing the mutant collection, at least one novel gene product (encoded by locus DR0167) has been identified. This protein bears no similarity to any gene product previously associated with DNA repair. We have also identified at least six mutant strains that carry defects in genes that encode DNA repair proteins in other species. Of these, defects in the *pol* gene are the most interesting. Phenotypes associated with the *pol* mutant suggests that the *pol* gene product plays a role in ionizing radiation and UV resistance, but that UV resistance is strongly influenced by the *uvrA* gene product.

New Insights into the Biology of an Obligate Methanotroph.

J. Martin Odom*, Jean-Francois Tomb, Rick Ye, Kelley Norton, Andreas Schenzle, Shiping Zhang. Central Research & Development Department, E. I. DuPont De Nemours & Co. Wilmington, Delaware 19880

Obligately methanotrophic bacteria are highly specialized aerobic eubacteria capable of utilizing only methane, and in some cases, methanol as sole source of carbon and energy. The distinguishing biochemical feature of these organisms is their ability to activate methane with oxygen under ambient conditions. The presentation will provide the first genome-level characterization of biochemical functionality and metabolic patterns for the assimilation and dissimilation of carbon and nitrogen in an environmental isolate designated Methylomonas 16a.. 16srRNA sequence data and other biochemical parameters type this isolate as a gram-negative, pink-pigmented Type I methanotroph within the genus Methylomonas. Methylomonas 16a possesses an extensive biochemistry for internal interconversion of sugars and polysaccharides yet is incapable of growth on sugars. Genes involved in the tetrahydromethanopterin- dependent pathway of formaldehyde oxidation to carbon dioxide are present as are the genes for formaldehyde and formate dehydrogenases and cyclic pentose oxidation. This strain is shown to catalyze reduction of nitrate or nitrite to nitrous oxide via multiple forms of nitrite and nitric oxide reductases. The physiological role of these enzymes will also be discussed.

Developing Phage Display for Functional Genomics. Timothy Palzkill. Department of Molecular Virology and Microbiology., Baylor College of Medicine, Houston, TX 77030.

Proteomics refers to the large-scale study of proteins, often by biochemical methods. One branch of proteomics is the large-scale analysis of protein-protein and protein-ligand interactions. The function of many genes cannot be deduced from sequence similiarity, and biochemical methods are usually required. The identification of binding partners by protein-ligand interaction studies can provide clues as to the function of unknown gene products. Protein-ligand interactions can be identified by affinity purification methods such as phage display. The display of proteins on the surface of filamentous phage has been shown to be a powerful method to select variants of a protein with desired binding properties from large combinatorial libraries of mutants. Phage display has also been used to study protein-ligand interactions on a genome wide scale. A genomic phage display library has been constructed for *E. coli* MG1655. The system has been tested by using the library to map dominant binding epitopes for an anti-RecA protein polyclonal antibody sera. This approach is now being used to study genome-wide protein-ligand interactions of *Treponema pallidum*, the causative agent of syphilis. To facilitate the study, a recombination-based cloning strategy is being used to clone and express all 1024 open reading frames of *T. pallidum*. The set of clones can be rapidly converted to His6 and GST-fusion protein expression plasmids by recombination. This will enable the rapid expression and purification of *T. pallidum* proteins as targets for phage display. In addition, the phage display vector has been incorporated into the recombination cloning system and is being used for the systematic fusion of each of the 1024 open reading frames of *T. pallidum* to the gene III protein of M13 filamentous bacteriophage. These experiments will create a normalized phage display library of the entire set of *T. pallidum* gene products. The combination of large-scale protein expression and phage display will be used for mapping protein-ligand interactions on a genome wide scale.

"Bacterial Annotation in a High-throughput Sequencing Environment"

Owen White, The Institute for Genomic Research

One challenge presented by large-scale genome sequencing efforts is effective display of uniform information to the scientific community. We have an annotation system and annotation standards that are uniformly applied to all microbial genomes sequenced at TIGR. This system will be described. The Comprehensive Microbial Resource (CMR) contains robust annotation of all complete microbial genomes and allows for a wide variety of data retrievals. Retrievals can be based on sequence searches, protein properties such as molecular weight or hydrophobicity, GC-content, functional role assignments, and taxonomy. The database contains extensive curated data as well as pre-run homology searches to facilitate data mining. Special emphasis has been placed on the ability to traverse across genomes via genes that have the same function. The infrastructure underlying the CMR includes complete bacterial genome sequences, a structured methodology for data curation, and a single relational database representing the data. The CMR content, methodology, and web presentation will be discussed.

Mechanisms of divergence from ancestral enzymes

Monica Riley, Alastair Kerr, Laila Nahum
Marine Biological Laboratory
Woods Hole, MA 02543

The molecular mechanisms of divergence can be elucidated by studying the kinds of similarities and differences present in proteins that have diverged from a common ancestor. Most proteins related by amino acid sequence have clear similarities in function in the sense of binding same or similar substrates/ligands or operating with the same reaction mechanism, or both. Even greater divergence can be detected that gives information on how divergence leads ultimately to new functions. Two examples will be presented: the SDR family and some of the enzymes of arginine metabolism in E. coli.

Use of Suppression Subtractive Hybridization as a Tool for Microbial Genome Comparisons.

G. L. Andersen*, L. Radnedge, and P. G. Agron. Lawrence Livermore National Laboratory, Livermore, California.

Suppression subtractive hybridization (SSH) has proven to be an efficient method for the selection of unique genomic regions among related microorganisms. Unique oligonucleotide adaptors are ligated to the ends of genomic restriction fragments and selective PCR amplification enriches for only those fragments present in one genome and absent in the other. We are developing an automated, high-throughput approach to SSH that will increase the likelihood of uncovering all genomic differences greater than 200-bp for related microbial genomes. With a combination of parallel processing in a microtiter plate format and the use of multiple restriction enzymes to increase coverage, SSH has the potential of being a cost-effective alternative to whole genome sequencing of organisms related to those for which complete sequence information already exists. To rigorously examine this approach, two sequenced strains of *Helicobacter pylori* (J99 and 26695) were used as a model system, as this allows rapid determination and mapping of difference products based on the available sequence. A comparison of the two strains (Alm et al., 1999 Nature 397: 176-180) shows significant differences such as 89 unique open reading frames (ORFs) out of 1,495 total for strain J99 and 117 unique ORFs out of 1,552 total for strain 26695. To increase the likelihood of amplifying difference products from any given region, four restriction enzymes, *Sau*3a, *Dra*I, *Apo*I and *Alu*I were used in separate SSH experiments. The difference products from each comparison were cloned, sequenced and then mapped by comparing the data to the *H. pylori* genome database to assess coverage, redundancy, and reproducibility of SSH. The minimum number of cloned difference products required to identify all unique ORFs was determined when using single or multiple restriction enzymes. The overall efficacy of the use of subtractive hybridization for pangenomic microbial comparisons will be discussed.

High-Throughput Transposon Insertion Mapping (HTTIM) in Pseudomonas aeruginosa: A Strategy for Identification of Essential Genes.

Paul Warrener, Richard Boykin, Michelle Lagrou, Lettie Goltry, Ernesto J. Tolentino, Ya Qi Zhu, Bryce E. Mansfield, Kim R. Folger, and C. Kendall Stover.
PathoGenesis Corporation, 201 Elliott Ave. West, Seattle, WA 98119

To help identify novel antimicrobial targets in Pseudomonas aeruginosa, we mapped the insertion sites of more than 12,000 transposon mutants on the recently completed P. aeruginosa PAO1 genome. Given that transposon-disruption of a gene indicates that gene is not required for the survival of this organism on rich media, we identified by elimination those genes with a high likelyhood of being essential for the survival of this microorganism. The 6.26 Mb PAO1 genome contains 5570 predicted open reading frames (ORFs). The function of the encoded proteins has been demonstrated for a small fraction of the total, and specific functions have been predicted with moderate certainty for approximately 25%. The result is that selection of gene products for further investigation as antimicrobial targets is a daunting task. Recovery of a mutant containing transposon insertions in a given gene indicates that the gene is not required for survival on rich media. Therefore, our approach of extensive transposon insertion mapping identified, by a process of elimination, those genes with a high likelihood of being essential for the survival of this microorganism. Our transposon system employs a mini-Tn5 transposon containing a tetracycline resistance marker and a modified transposase gene. Analyzed mutants were selected at random from several independent libraries, each of which was estimated to contain 40,000-50,000 independent insertional events. A high-throughput PCR-rescue technique was used for direct determination of transposon insertion site sequence. An automated algorithm for sequence analysis and insertion site mapping was developed. This strategy identified disruptions in 70% of the PAO1 ORFs, thereby focusing our interest on the remainder as putative essential genes. We are further further analyzing these genes using targeted inactivation and conditional expression approaches. In addition, transposon mutants exhibiting a viable but growth-impaired phenotype have also been characterized. These mutants, which make up approximately 1% of the mutant population, directly implicate the affected genes as being important to growth of P. aeruginosa, and include disruptions of genes shown to be essential in other organisms. These two approaches provide valuable insight into pathways which may be targeted by novel anti-microbial compounds.

Construction of new removable cassettes for *Escherichia. coli* gene deletion

C. Merlin*, S. McAteer and M. Masters. ICMB, University of Edinburgh, Scotland

Although DNA micro-arrays can yield gene expression data on a genome-wide scale, there remain many situations when information about the expression of a single gene is desired and the expense and complex analysis which accompany micro-array use is not possible. We have constructed a number of cassettes which can be used to replace genes of interest and which, because they are flanked by recombination sites (FRT sites, resolvable by FLP recombinase), can later be removed to leave an in-frame deletion. We have completed the construction of *aph* (Kan^R), *lacZ aph* and *lacZ cat plac*; *lacZ aph plac* is currently under construction. *aph* and *cat* offer removable selective markers, *lacZ* can act as a reporter for the expression of the deleted gene while *plac* provides a promoter for any downstream gene in the target operon. Each *frt* flanked cassette is further flanked by restriction sites which facilitate its use in the crossover PCR deletion vector, pKO3 (Link *et al.*, 1997, J. Bact. 179:6228-6237), or in pUC18. The cassettes can also be used as templates in the linear DNA replacement methods described by Datsenko and Wanner (2000, Proc. Natl. Acad. Sci. USA, 97:6640-6645) and by Yu *et al.* (2000, Proc. Natl. Acad. Sci. USA, 97:5978-5983).

The poster will describe details of cassette construction and their use for the separate deletion and expression analysis of *ruvA* and *ruvB*.

Individual Recombinational Replacements and Mutational Substitutions in the Recent Evolutionary History of Eleven Closely Related *E. coli* Strains.

R. D. Milkman, R. McBride, J. Harrington, M. Thompson & E. Wright. The University of Iowa, Iowa City.

Individual recombinational replacements in ECOR 8, in ECOR 11, and in the most recent ancestor of K12 and ECOR 11 have lengths of ~136 kb, ~91 kb and ~10 kb, respectively. Colinear 11-kb stretches of the ECOR 8 and ECOR 11 replacements were compared with 34 other ECOR strains: each is a mosaic of diverse identities and near-identities (≤ 0.3% difference) to one or another of these strains. Evidently a small number of ancient clonal frames are represented in fragments produced by successive DNA transfers in the ancestors of each respective replacement donor. – Comparison of eleven strains (K12, ECOR 1, 2, 3, 5, 8, 9, 10, 11, 12, 25) indicated that they are closely enough related to permit the construction of a tree based on a set of consistent mutational and DNA-transfer events in the non-hypervariable 85% of the *E. coli* chromosome. Thus a mutation or replacement shared between K12 and (for example) ECOR3, but not ECOR 11, is not expected, or does not imply a single ancestral event.

Relatedness among microorganisms assessed by gene specific DNA:DNA hybridizations to DNA microarrays. A. E. Murray*[1], D. P. Lies[2], J. C. Boles[1], G. Li[3], K. H. Nealson[2], J. Zhou[3], J. M. Tiedje[1]. [1] Center for Microbial Ecology, 540 PSSB, Michigan State University, East Lansing, MI, 48824-1325, [2]Department of Geology and Planetary Sciences, Jet Propulsion Laboratory, and California Institute of Technology, Pasadena, CA 91125-2300 [3] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

Central to the study of microbial evolution and ecology is having an understanding of the relationships between different organisms. Though informative from a phylogenetic perspective, the SSU rRNA gene provides few clues regarding functional relationships between different microorganisms, and little information as to what makes phylogenetically related organisms distinct ecological species. We have undertaken a project to evaluate whether a DNA microarray designed with genes from one organism (*Shewanella oneidensis*, MR1), can be used to assess homology between the reference strain (MR1) and another strain (DLM7) or species in the *Shewanella* genus. A pilot microarray was constructed with 192 PCR-amplified genes. Most genes on the array were selected because of homology to genes with roles in energy metabolism and electron transfer. Organisms in the *Shewanella* genus utilize a diverse repertoire of terminal electron acceptors when respiring, and can be found in a broad range of environments. The results suggest that there is a high correspondence between the relationships predicted from both SSU rRNA and DNA gyrase (gyrB) gene phylogenies and the relatedness predicted from hierarchical cluster analysis of microarray hybridization data between different *Shewanella* species. The results of parametric and nonparametric methods of data analysis were compared. Regardless of the algorithm used, both universally conserved and universally heterogeneous gene clusters were consistently identified with the same core genes in each cluster. The results suggest that there are strain specific differences between MR1 and DLM7 in the genes involved in anaerobic respiration. Among all species investigated, several enzymes with general housekeeping function such as genes coding for ATP synthase, a ribosomal protein, and lysyl-tRNA synthetase were universally conserved. This approach to studying microbial relatedness provides an information-rich level of understanding of the specific genomic content between closely related microorganisms, and should prove to be quite useful from evolutionary, ecological, and taxonomic perspectives.

Potential for Divergence of Function in Sequence-Related Groups of Escherichia coli Enzymes

L.A. Nahum* & M. Riley. Marine Biological Laboratory, Woods Hole.
(laanahum@mbl.edu & mriley@mbl.edu)

Groups of sequence-related proteins of Escherichia coli have been assembled that seem likely to have arisen by duplication and divergence of genes in the ancestral genomes over time, some arising recently, some in early evolutionary times. The process of duplication and divergence is believed to have generated groups of proteins of similar sequence that share features of binding site specificity and/or mechanisms of action. We have tested this outlook by close examination of pairs or triplets of sequence-related proteins in E. coli. The proteins have been selected as enzymes of small molecule metabolism and include representatives of amidotransferases, phosphotransferases, decarboxylases, and others. The enzymes have been characterized in terms of sequence similarity (DARWIN, Gapped BLAST), protein domains (Pfam), and functional features (EcoCyc, GenProtEC) using different databases. Our data indicate that the similarity found in these enzymes is in some cases related to binding site specificity, but in most cases is related to the chemistry of the reaction catalyzed. We have also looked for more distant relatives of these groups of enzymes by PSI-BLAST searches, trying to identify larger families in E. coli. The occurrence of the homologs in the microbial complete genomes was investigated as well. Grouping by distant sequence relatedness allows us to collect together proteins that differ, but whose molecular specificity, binding sites and/or mechanisms of action are similar, revealing commonalities that probably reflect common ancestry. These sequence-related groups give us a view of molecular evolution from the point of view of divergence in action. As we continue this analysis, we will be contributing to an understanding of the mechanisms of protein evolution.

Financial Support: FAPESP (S„o Paulo ñ Brazil)

Pathogenic capability has been lost at least twice during the evolution of the genus *Listeria*.

Authors: Eva Ng[1], Michael Schmid[2], Michael Wagner[2], Werner Goebel[1]

[1]Lehrstuhl für Mikrobiologie, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. [2]Lehrstuhl für Mikrobiologie, Technische Universität München, Am Hochanger 4, D-85350 Freising, Germany.

We undertook to resolve the phylogeny of the genus *Listeria* with the goal of ascertaining insights into the evolution of pathogenic capability of its members. The phylogeny of *Listeriae* had not yet been resolved due to a scarcity of phylogenetically informative characters within the 16S and 23S rRNA molecules.

The genus *Listeria* contains 6 species: *L. monocytogenes, L. ivanovii, L. innocua, L. seeligeri, L. welshimeri*, and *L. grayi;* of these, *L. monocytogenes* and *L. ivanovii* are pathogenic. Pathogenicity is enabled by a 10-15Kb virulence gene cluster found in *L. seeligeri, L. monocytogenes* and *L. ivanovii*. There are no remarkable GC content differences between the virulence gene clusters and their respective genomes. The genetic contents of the virulence gene cluster *loci*, as well as some virulence-associated internalin *loci* were compared among the 6 species.

Phylogenetic analysis based on a data set of nucleic acid sequences from *prs, ldh, vclA, vclB, iap*, 16S and 23S rRNA genes identified *L. grayi* as the ancestral branch of the genus. This is consistent with previous 16S and 23S rRNA findings. The remainder 5 species formed 2 groupings. One lineage represents *L. monocytogenes* and *L. innocua*, while the other contains *L. welshimeri, L. ivanovii* and *L. seeligeri*, with *L. welshimeri* forming the deepest branch within this group. This implies that the virulence gene cluster was present in the common ancestor of *Listeria monocytogenes, innocua, ivanovii, seeligeri* and *welshimeri;* and that pathogenic capability has been lost in two separate events represented by *L. innocua* and *L. welshimeri*.

THE ECOCYC AND METACYC DATABASES.

J. Collado, UNAM, Cuernavaca, Mexico; P.D. Karp, SRI International, Menlo Park, CA; I. Paulsen, TIGR, Rockville, MD; A. Pellegrini-Toole*, Marine Biological Laboratory, Woods Hole, MA; M. Riley, Marine Biological Laboratory, Woods Hole, MA; M. Saier, UC San Diego, San Diego, CA.

EcoCyc is a model-organism database for Escherischia coli K-12. It describes the complete genome of E. coli, and provides detailed functional annotations for the genome including descriptions of enzymes, metabolic pathways, two-component signal-transduction pathways, transporters, and transcription units.

Transcriptional regulatory information is a recent addition to EcoCyc. Data from the RegulonDB database has recently been incorporated into EcoCyc, including descriptions of 617 E. coli promoters, 83 transcription factors, and 617 transcription-factor binding sites. EcoCyc version 5.5 describes 526 transcription units, and provides a unique resource for interpretation of gene-expression data in the context of known transcriptional regulatory mechanisms.

MetaCyc is a meta-metabolic pathway database describing pathways from many different organisms. MetaCyc describes metabolic pathways, reactions, enzymes, and substrate compounds. The MetaCyc data were gathered from a variety of literature and on-line sources. Each pathway is labeled with the organism(s) in which it is known to occur based on laboratory experimentation. Each pathway references the source from which it was obtained. MetaCyc was initialized to contain all metabolic pathways of EcoCyc. Many additional pathways from other organisms were then added to the database. MetaCyc employs the same database schema and software visualization tools as does EcoCyc.

EcoCyc and MetaCyc are available at URL http://ecocyc.doubletwist.com/

References

Karp, P.D., Riley, M., Saier, M., Paulsen, I., Paley, S. and Pellegrini-Toole, A., "The EcoCyc and MetaCyc Databases," Nucleic Acids Research 28(1):56-59 2000.

Karp, P.D., Krummenacker, M., Paley, S., and Wagg, J., "Integrated pathway/genome databases and their role in drug discovery," Trends in Biotechnology 17(7):275-281 1999.

Salgado H., Santos-Zavaleta A., Gama-Castro S., Millán-Zárate D., Blattner F.R., and Collado-Vides J. (2000) "RegulonDB (version 3.0): Transcriptional Regulation and Operon Organization in Escherichia coli" Nucleic Acids Research 28: 65-67

Complete Sequence Analysis of dsRNAs from *Discula destructiva*.

R. Rong*, S. Rao, S. W. Scott and F. H. Tainter. Clemson University, Clemson, SC.

Complete dsRNAs sequences were determined in *Discula destructiva* isolate 247 from South Carolina. Four dsRNAs (ca. 1.8, 1.6, 1.2 and 0.3 kb) were extracted from subcultures on membrane potato-dextrose media. Linkers 4-14-1 (GAGGGATCCAGTTTAAAATCCTCAGAGGA) and 4-14-pp (p-TCCTCTGAGGATTTTAAACT-p) and MuLV reverse trancriptase were used for cloning dsRNAs. Full-length sequences revealed that the first band consisted of 1787 bp. An ORF, from nt. 65 to 1684 within this band, has the potential to code for RNA-directed RNA polymerase. The second band contained 1584 bp, putatively encoding capsid protein. The third band contained 1180 bp with an ORF from nt. 259 to 1014 in one strand and the fourth band is 307 bp in length. Common terminal sequences appeared in all these dsRNA sequences. A total of 18 different dsRNA banding patterns were detected in the 108 Alabama, South Carolina and Idaho isolates examined. The numbers of dsRNA bands in each isolate varied from 0 to 7 and ranged in size from 0.3 to 12 kb. Sequence homology of corresponding dsRNA bands of the same size was indicated by Northern hybridization among 23 randomly chosen isolates. By continuing cloning and sequencing, the interrelationships and functions of dsRNAs among different isolates will be characterized.

Anatomy of *Escherichia coli* Ribosome Binding Sites

Ryan K. Shultzaberger [*1], R. Elaine Bucheimer[2], Kenneth E. Rudd[3] and Thomas D. Schneider [4, 5]

During the process of translational initiation in prokaryotes the 3′ end of the 16S rRNA binds to a region just upstream of the initiation codon. The relationship between this 'Shine and Dalgarno' (SD) region and the binding of ribosomes to translation start points has been well studied, but a unified mathematical connection between the SD, the initiation codon and the spacing between them has been lacking. Using information theory we constructed a simple model that treats these three components uniformly by assigning to each a strength in bits of information. To do this, we first aligned the SD region by maximizing the information content there. The ease of this process confirmed the existence of the SD in nearly 4000 *Escherichia coli* genes in the EcoGene 12 database. This large data set allowed us to graphically show by sequence logos that the spacing between the SD and the initiation region affects both the SD site conservation (in bits) and its pattern. We used the aligned SD, the spacing, and the initiation region to model ribosome binding and to identify gene starts that do not conform to the SD model, either because they are anomalous starts, or because they are errors made by start site annotation programs. These steps were easily visualized with sequence walkers. The model obtained by this process is similar to a model made from 569 experimentally proven gene starts, suggesting that an effective ribosome model can be constructed without the tedious work of generating a complete, biochemically supported database. This information theory based technique can be used to refine the prediction of gene start annotation in whole genome sequences.

[1]University of Maryland, College Park, Maryland, 20742 and National Cancer Institute, Frederick Cancer Research and Development Center, Laboratory of Experimental and Computational Biology, P. O. Box B, Frederick, MD 21702-1201.

[2]Univ. of Virginia School of Medicine Charlottesville, VA 22908

[3]Department of Biochemistry and Molecular Biology (R-629), University of Miami School of Medicine, P. O. Box 016129, Miami, FL 33101-6129

[4]National Cancer Institute, Frederick Cancer Research and Development Center, Laboratory of Experimental and Computational Biology, P. O. Box B, Frederick, MD 21702-1201. (301) 846-5581 (-5532 for messages), fax: (301) 846-5598, email: toms@ncifcrf.gov, http://www-lecb.ncifcrf.gov/~toms/

[5]*Corresponding author*

DEVELOPMENT OF THERMOFIDELASE AND FIMERS FOR GENOMIC SEQUENCING

Alexei Slesarev*, Andrei Malykh, O. Malykh, N. Polouchine, S. Kozyvkin
Fidelity Systems, Inc., 7961 Cessna Avenue, Gaithersburg, MD 20879

ThermoFidelase has a unique combination of activities that are not found in other proteins. As a thermostable topoisomerase, it unlinks and denatures topologically constrained DNA, as a DNA binding protein, it accelerates primer annealing, helps polymerase to pass through secondary structures and protects DNA from thermal degradation. We have developed several formulations of ThermoFidelase for various types of sequencing reactions. To overcome limitations of using standard oligonucleotides as primers in sequencing off genomic DNA we have generated and screened combinatorial libraries of chemically modified primers. The method is based on our proprietary monomers containing MOX or SUC reactive moieties that are incorporated in primer precursors. After synthesis, libraries of chemically modified primers ("Fimers") are prepared by treating precursors with a variety of compounds. Fimers prevent non-specific amplification and premature truncation at secondary structures, increase sensitivity and quality of genomic sequencing with small quantities of templates. During the development we have established a number of records: sequencing through very long hairpin (T_m>90°C) and >1.5 kb simple nucleotide (C,T) repeat region, sequencing unique regions off 5 Mb microbial genomic DNA and repeated regions off 3 Gb human genomic DNA, highly sensitive sequencing off 10 ng BAC and 100-300 ng genomic DNA, direct sequencing off 1 ul microbial cultures without any steps of DNA isolation or purification.

We have used ThermoFidelase with Fimers in several high-throughput projects: rapid finishing of 'working-draft' human genome project (10 BACs from BCM-HGSC), closing gaps in bacterial genome projects, shotgun sequencing of unpurified plasmids, rapid identification of gene knock-outs in transposon-modified microbial clone library, and for rapid identification of bacterial strains without PCR.

Leaderless Transcripts of the Crenarchaeal Hyperthermophile *Pyrobaculum aerophilum*.

Malgorzata M. Slupska[1], Angela G. King[1], Sorel Fitz-Gibbon[1], John Besemer[2], Alex Lomsadze[2], Mark Borodovsky[2] and Jeffrey H. Miller[1]

[1]University of Los Angeles, CA 90095

[2]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

We mapped transcription start sites for eight unrelated protein-encoding *Pyrobaculum aerophilum* genes by primer extension and S1 nuclease mapping. All of the mapped **transcripts** start at the computationally predicted **translation** start codons. A whole genome computational analysis of the regions from -50 to +50 nucleotides around the predicted translation starts codons revealed a clear upstream pattern matching the consensus sequence of the archaeal TATA box located unusually close to the translation starts. For genes with the TATA boxes that best matched the consensus sequence, the distance between the TATA box and the translation start codon appears to be shorter than 30 nucleotides. Two other promoter elements distinguished were also found unusually close to the translation start codons: a transcription initiator element with significant elevation of C and T frequencies at the -1 position and a BRE element with more frequent A's at position – 29 to – 32 (counting from the translation start site). We also show that one of the mapped genes is transcribed as the first gene of an operon. For a set of genes likely to be internal in operons the upstream signal extracted by computer analysis was a Shine-Dalgarno pattern matching the complementary sequence of *P. aerophilum* 16S rRNA. Together these results indicate that the translation of proteins encoded by single genes or genes that are first in operons in the hyperthermophilic crenarchaeon *P. aerophilum* proceeds mostly, if not exclusively, through leaderless transcripts. Internal genes in operons are likely to undergo translation via a mechanism that is facilitated by ribosome binding to the Shine-Dalgarno sequence.

Generation and Analysis of a *fur* (Ferric Uptake Regulator) Mutant of *Shewanella oneidensis* MR-1 Using Genomic and Proteomic approaches.

D. K. Thompson[a]*, A. Beliaev[a], G. Li[a], C. S. Giometti[b], S. Tollaksen[b], D. Lies[c], K. H. Nealson[c], & J. Zhou[a]. [a]Environmental Sciences Division, Oak Ridge National Laboratory[†], Oak Ridge, TN; [b]Argonne National Laboratory, Argonne, IL; [c]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA.

Shewanella oneidensis MR-1 is capable of utilizing a variety of compounds as terminal electron acceptors, including ferric iron. However, little is known about the genetic basis and regulatory mechanisms controlling iron metabolism in this bacterium. To determine whether the putative *fur* gene is involved in iron reduction, a *fur* mutant of MR-1 was generated by insertional mutagenesis and characterized by DNA microarrays and two-dimensional (2D) PAGE. In *E. coli* and other organisms, the Fur protein represses transcription in the presence of high iron by binding to specific sequences in the promoters of iron-regulated genes. Comparative sequence analysis revealed that MR-1 Fur exhibits a high degree of sequence identity (>70%) at the amino acid level to Fur proteins of *E. coli* and *Vibrio cholerae*. The MR-1 *fur* gene was disrupted by cloning a 179-bp internal *fur* fragment into the suicide vector pKNOCK-Km and then transferring the construct into MR-1 cells. Physiological studies indicated that the *fur* mutant was similar to wild-type MR-1 when compared for anaerobic growth and reduction of various electron acceptors. To define genes regulated by Fur, we used partial microarrays containing approximately 1,000 MR-1 genes thought to be involved in energy metabolism, transcriptional regulation, environmental stress, and siderophore production. Preliminary results suggested that disruption of the *fur* gene affected the expression of siderophore-related genes as well as other genes. This agreed with the finding that the *fur* mutant produced 3-fold higher levels of siderophore than the wild-type strain. Analysis of the *fur* mutant by 2D-gel electrophoresis indicated that at least 4 major proteins increased significantly ($P<0.005$) in abundance in *fur* mutant cells relative to MR-1 cells. Efforts are underway at identifying these proteins. Although MR-1 Fur is not directly involved in metal reduction, it does appear to play an important regulatory role in pathways leading to iron acquisition.

[†]Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract number DE-AC05-000R22725.

Pattern and context searching for subtilases in microbial genomes.
Roland Siezen1,2, Bernadet Renckens1,2, Johan van Hylckama Vlieg*1, & Jack Leunissen2. 1NIZO food research, Ede, and 2Center for Molecular and Biomolecular Informatics, Nijmegen, the Netherlands.

Subtilases are a very diverse family of subtilisin-like serine proteases found in archaebacteria, eubacteria, fungi, yeasts and higher eukayotes (1). They are generally exported outside the cell, and are known to play a role in either nutrition (providing peptides and amino acids for cell growth), precursor processing (e.g. hormones, toxins, bacteriocins), or host invasion. Their sequence homology can be very low, with only the segments surrounding the 3 catalytic residues Asp, His and Ser showing some sort of conservation throughout the entire family. We have now employed a multiple pattern searching method to identify low homology members of this family
in 55 complete or nearly complete microbial genomes, using the updated PROSITE patterns:
Subtilase-Asp:
[STALIV]-x-[STLIVMF]-[LIVMFNA]-D-[STAIQEND]-[GNAP]-[LIVMFYC]
Subtilase-His:
H-G-[STAMVQE]-x-[STAVICM]-[STAGCIL]-[STAGN]-x-[STAPLIVM]
Subtilase-Ser:
G-[STANLI]-S-x-[SA]-x-[SAGNP]-x-[TALIVFC]-[STAGVYC]-[TAGN]

Individually these patterns find mostly false positives, but in combination there are no false positives, and only a few false negatives in 5 genomes. In 34 microbial genomes no subtilase genes were found. In 10 genomes we found 1 subtilase gene, and in 11 genomes more than 1 subtilase gene. Three of the 21 subtilase genes found in these genomes were not originally annotated as members of the subtilase family. The genomes of Mycobacterium tuberculosis, Bacillus subtilis and Deinococcus radiodurans contain 5, 7 and 9 subtilase genes, respectively.
Further analysis predicted that most of these subtilases have a signal peptide and are probably secreted, while several have predicted membrane anchors. In an attempt to predict the cellular function of these subtilases, we performed context searches by screening the upstream and downstream genes. Most subtilase genes do not appear to be functionally linked to other genes. An exception is formed by the 5 genes encoding subtilases in Mycobacterium tuberculosis (and other mycobacteria) which all occur in conserved gene clusters that may be involved in production of virulence factors (N. Gey van Pittius et al, manuscript in preparation).

RJ Siezen & JAM Leunissen (1997) Subtilases: the superfamily of subtilisin-like serine proteases. Protein Science 6, 501-523.

Microarray analysis reveals *scoC* as a major pleiotropic modulator of *Bacillus subtilis* physiology. Randal R. Maile, Stuart C. Causey, Walter Weyler*, Robert Caldwell, Ron Sapolsky, Eugenio Ferrari. Genencor International, 925 Page Mill Road, Palo Alto, California 94304.

We analyzed the effect of the *scoC4* mutation, of *Bacillus subtilis* by genome wide transcriptome analysis with Sigma-Genosys Nylon membrane DNA arrays. This mutation is a transvertion A to C leading to the amino acid change, A21E. In excess of 500 genes, including previously reported genes, showed significant alteration in expression level when the mutant strain was compared to its isogenic control. From these results we can conclude that the scope of the regulatory role of *socC*, or rather ScoC, is probably larger than is currently recognized. Clear effects are observed for regulatory proteins and transport systems as well as a number of intracellular and extracellular enzymes. While a large number of effects are noted near the transition state, many genes undergo transcriptional changes during exponential and stationary phase. Increases and decreases in expression levels are observed. The *scoC* mRNA is evident over the entire course of the culture and the physiological changes observed between the wild type and the mutant suggest that *scoC* must play a role over the whole life cycle. This poster will provide specific examples of regulation of genes sorted by functional group and followed over culture time.

Functional Gene Arrays for Microbial Community Analyses

Liyou Wu1, 2, Dorothea K. Thompson1, Guangshan Li1, Richard A. Hurt1, James M. Tiedje2, and Jizhong Zhou1, 2

1Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 38831. 2Center for Microbial Ecology, Michigan State University, East Lansing, MI, 48824.

While microarrays are proven tools for whole-genome expression analysis, the technology has not been systematically tested for microbial community studies using diverse environmental samples. To assess the potential of array-based methods for monitoring bacterial populations, DNA microarrays were constructed using selected genes involved in nitrogen cycling: heme- and copper-containing nitrite reductase genes (nirS and nirK, respectively) and ammonia monooxygenase (amoA)-like genes from pure cultures and those cloned from marine sediments. Microarray fabrication and hybridization were optimized in terms of fluorescence intensity by evaluating different glass slides, DNA deposition buffers, rehydration and denaturation times, and probe concentrations. Specific hybridizations were obtained for the different target genes at high stringency (65(C). The limit of detection was approximately 1 ng with pure genomic DNA and 25 ng with soil community DNA. A strong linear quantitative relationship was observed between signal intensity and target DNA concentration within a range of 1 to 100 ng. However, sequence divergence and probe size had significant effects on hybridization intensity. The applicability of functional gene arrays for microbial community analysis was demonstrated by measuring the distribution of nir and amoA genes in marine sediment and surface soil environments. Our results show that microarrays can be used as a specific, sensitive, quantitative, and parallel tool for characterizing microbial community composition and structure in natural environments.

Chipping the Salmonella typhimurium genome

Steffen Porwollik1, David Boyle1, Rick Wilson2, Sandy Clifton2, Ken Sanderson3, and Michael McClelland1*,
1 Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, Ca 92121, USA
2 Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108
3 University of Calgary, Calgary, Alberta, Canada T2N 1N4
* for correspondence: ph 858 450 5990 ext 280, fax 858 550 3998, email: mmcclelland@skcc.org

The gram-negative bacterium Salmonella enterica serovar Typhimurium infects a broad spectrum of mammalian hosts and causes gastroenteritis in humans. It is a major model organism in microbial pathogenicity and motility research. Sequencing of the genome of Salmonella typhimurium will be completed in October 2000 (http://genome.wustl.edu/gsc/bacterial/salmonella.shtml).
We are arraying the complete open reading frame set of this bacterium by spotting PCR amplified products on glass slides. Primers are designed to allow for easy cloning of the amplified gene into a wide range of commercially available expression vectors. At present, more than 2200 ORFs have been successfully amplified and spotted. The complete S. typhimurium chip, which will also accommodate genes unique to S. typhi and S. paratyphi, will be available to the research community in the winter of 2000.

Transcript profiling of virulence factor genes of the enteric pathogen Campylobacter jejuni.

E.N. Taboada*, C.M. Szymanski, & J.H.E. Nash.
Institute for Biological Sciences, National Research Council of Canada.

Campylobacter jejuni is an important human pathogen and is the world's leading cause of food-borne enteric disease. Unlike other enteric pathogens such as *Salmonella* and *Listeria*, the processes leading up to *Campylobacter* infection are not well understood. Similarly, the mechanism through which *C. jejuni* induces disease in human hosts remains poorly characterized. The completion of the *C. jejuni* genome sequencing effort by The Sanger Centre has opened the possibility of applying genomics techniques to the study of *Campylobacter* infection and disease. We are interested in investigating *C. jejuni* gene expression during colonization, invasion, and infection in order to establish genes that are important in these various processes. To this end, we are currently developing microarrays that contain all 1654 ORFs in the genome of *C. jejuni* NCTC 11168. We present here preliminary results obtained on a smaller array comprising 162 different known and putative virulence factor-encoding genes. We have used this array to monitor changes in gene expression in *C. jejuni* cells harvested from *in vitro* invasion assays performed on cultured human epithelial cells.

Genetic Relationships of Escherichia coli O157:H7 subpopulations isolated from Australia and the United States

Jaehyoung Kim*, Jingliang Ju & Andrew K. Benson. University of Nebraska-Lincoln, Nebraska

Escherichia coli O157:H7 is a geographically disseminated clonal pathogen that is the primary etiologic agent of foodborne hemorragic colitis in the United States, Canada, Japan, and the United Kingdom. Comparative genome analysis of Escherichia coli O157:H7 strains by Octamer Based Genome Scanning has recently demonstrated the existence of two genetically distinct lineages, I and II, that are disseminated throughout the United States. Lineage II strains are overrepresented among bovine isolates and underrepresented among human clinical isolates that have been tested, suggesting that this lineage may be inefficiently transmitted to humans from bovine sources or that it is less virulent. Using Octamer Based Genome Scanning, we have now examined the relationship of lineage I and II strains from the United States to O157 strains from Australia, where other serotypes of enterohemorragic E. coli are more frequently isolated from hemorragic colitis patients despite the presence of O157:H7 in cattle. Phylogenetic analysis indicates that subpopulations of lineage I and lineage II strains exist in Australia with lineage II being more prevalent among the isolates that were tested. Included was a large clade of non-motile lineage II O157 isolates and a smaller clade of non-motile lineage I isolates. The distributions of polymorphisms detected by OBGS and motility phenotypes further indicate phylogeographic variation and selection among the O157 subpopulations that is independent of ancestry. These results suggest a model in which common genetic events have occurred in each lineage subsequent to geographic transmission.

High-Throughput Mutagenesis and Functional Genomics in *E. coli*

C.D. Herring*, J.Glasner, C. Richmond, H. Jin, R. Mau, Y. Qiu, M. Liu, F. Blattner. University of Wisconsin, Madison.

The complete sequence of the *Escherichia coli* genome was published in 1997. A putative function can be assigned to most ORFs by sequence comparison, but this approach is of limited reliability and no match can be found for nearly 40% of all ORFs. The goal of our lab is to systematically investigate the function of all genes in *E. coli*. We have approached this task in two ways; 1) by studying changes in gene expression under different experimental conditions using microarrays, and 2) by mutating and overexpressing individual ORFs and then looking for an effect on gene expression or growth phenotype. While linear recombination may prove the quickest method for the mutation of most genes, a plasmid system has been developed to quickly make suppressible amber mutations in essential genes. This plasmid utilizes a tet-controlled lambda origin and I.SceI meganuclease counterselection.

Extremely diversified bacterial defense systems, the restriction-modification systems, in *Helicobacter pylori* revealed by comparative genome studies

L.-F. Lin, J. Posfai, R. J. Roberts, & H. Kong*. New England Biolabs.

Helicobacter pylori is a gram-negative bacterial pathogen with a small genome of 1.64-1.67 Mb. However, more than twenty putative DNA restriction-modification (R-M) systems were identified in the two completely sequenced *H. pylori* strains, 26695 and J99, based on sequence similarities. In this study, we have investigated the biochemical activities of 14 Type II R-M systems in *H. pylori* 26695. Less than 30% of the Type II R-M systems in 26695 are fully functional, similar to the results obtained from strain J99. Although close to 90% of the R-M genes are shared by the two *H. pylori* strains, different sets of the fully active R-M were functional in each strain. The strain specific R-M genes are much more active than the shared ones. This suggests the notion that strain specific genes have been acquired more recently through horizontal transfer from other bacteria and thus are less likely to be inactivated by random mutations. Our results also show that *H. pylori* has extremely diversified R-M systems in different strains and the diversity may be maintained by constantly acquiring new R-M systems, and by inactivating and deleting old ones.

Finding Modules Present In Homologous Proteins Encoded By Completely Sequenced Microbial Genomes

S. Le Bouder, K. Abou-Merhy, H. Arfaoui, A. Bazureau, I. Montaland, R. De Rosa and B. Labedan*, Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris-Sud, 91405 Orsay Cedex, France

We have previously shown that the homology which can be detected between proteins is often limited to long structural segments. We have called modules such structural segments of homology (Riley and Labedan, 1997 - Labedan and Riley, 1999). We describe here a suite of automatic programs allowing to catalogue the whole set of modules present in microbial proteomes. First, we detect successively homologous proteins using stringent thresholds for evolutionary distance and minimal length of aligned segments and then modules they are containing. Secondly, after grouping these homologous modules in families we further group those families which are related by a chain of neighbouring unrelated homologous modules. The analysis of these groups allows us to break up into their component parts many fused modules and/or to deduce by logic more distant modules. Then, all detected and deduced modules are assembled in refined families. To make automatic these two steps, we have further designed another program. Finally, we describe how we check the validity of this module approach using three independent and complementary tests. This experimental approach is illustrated by the data obtained in the comparison of *Campylobacter jejuni, Escherichia coli, Haemophilus influenzae* and *Helicobacter pylori*. We feel that our method could retrieve all modules in homologous proteins with the underlying idea that such modules will be the signatures of ancient gene duplication and gene fusion events.

Labedan, B et M. Riley. (1999) pp 311-329 *In* "Organization of the Prokaryotic Genome" (Robert Charlebois, editor) ASM Press, Washington.

Riley, M. and B. Labedan. (1997). *Journal of Molecular Biology*, **268**:857-868

LIST OF PARTICIPANTS

Ågot Aakra
Agricultural University of Norway
Dept. of Chemistry and Biotechnology
Lab of Microbial Gene Technology
P.O. Box 5051
N-1432 ÅS
Phone: (+47) 64 94 94 63
Fax: (+47) 64 94 94 63
Email: agot.aakra@ikb.nlh.no

Gary Andersen
Biology&Biotechonology National Research Program
Lawrence Livermore National Laboratory
P.O.Box 808 L- 452
Livermore, CA 94550
Phone: (925) 423-2525
Fax: (925) 422- 2282
Email: andersen2@llnl.gov

Siv Andersson
Molecular Evolution
Uppsala University
Norbyuagen 18C
Uppsala, 7523y
Sweden
Phone: (+46) 18 471 43 79
Fax: (+46) 18 455 77 23
Email: siv.andersson@molbio.uu.se

Linda Banerjei
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Phone: (301) 315-2524
Fax: (301) 838-0208
Email: banerjei@tigr.org

Alex Bateman
The Sanger Centre
Wellcome Trust Genome Campus
Hinxton, Cambs CB 10 1SA UK
Phone: (+44) 1223-494950
Fax: (+44) 1223-494919
Email: agb@sanger.ac.uk

John R. Battista
Dept. of Biological Sciences
Louisiana State University and A&M College
508 Life Sciences Bldg.
Baton Rouge, LA 70803
Phone: (225) 388-2810
Fax: (225) 388-2597
Email: jbattis@lsu.edu

Alex S. Beliaev
Oak Ridge National Laboratory
Environmental Sciences Division
Building 1505, Rm 390
Oak Ridge, TN 37831-6038
Phone: (865) 574-7379
Fax: (865) 576-8646
Email: baj@ornl.gov

Stephen Bentley
Pathogen Sequencing
The Sanger Centre
Wellcome Trust Genome Campus
Hinxton, Cambs CB 10 1SA UK
Phone: (+47) 1223 494761
Fax: (+47) 1223 494761
Email: sdb@sanger.ac.uk

Giuseppe Bertani
Jet Propulsion Laboratory
California Institute of Techonology
JPL 125-224
975 Dale St,
Pasadena, CA 91109
Phone: (818) 354-4239
Fax: (818) 393-4057
Email: gbertani@lalc.k12.ca.us

L. Elizabeth Bertani
California Institute of Technology
Biology 156-29
1200 E. California Blvd.
Pasadena, CA 91125
Phone: (626) 395-4917
Fax: (626) 449-0756
Email: Lebert@cco.caltech.edu

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

James Brown
Bioinformatics
UP1345
SmithKline Beecham Pharmaceuticals
1250 South Collegeville Rd.
Collegeville, PA 19426-0989
Phone: (610) 917-6374
Fax: (610) 917-7901
Email: james_r_brown@sbphrd.com

Cliff Brunk
Organismic Biology, Ecology & Evolution
2203 Life Science Building
405 Hilgard Ave
Los Angeles, CA 90095
Phone: (310) 825-3114
Fax: (310)
Email: cbrunk@ucla.edu

Stuart C. Causey
Genencor International
925 Page Mill Road
Palo Alto, CA 94303
Phone: (650) 846-5847
Fax: (650) 621-7847
Email: scausey@genencor.com

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

Julie Ju-Huei Chiang
Dept. of Microbiology and Molecular Genetics
University of California, Los Angeles
1602 Molecular Sciences Building
Los Angeles, CA 90095
Phone: (310) 825-8924
Fax: (310) 206-3088
Email:jchiang@mbi.ucla.edu

Ludmila Chistoserdova
University of Washinton
Box 351750 Benson Hall
UW Seattle, WA 98195
Phone: (206) 543-6683
Fax: (206) 616-5721
Email: milachis@u.washington.edu

Wendy Clendenin
Dept. of Microbiology and Molecular Genetics
University of California, Los Angeles
1602 Molecular Sciences Building
Los Angeles, CA 90095
Phone: (310) 825-8924
Fax: (310) 206-3088
Email: wluther@mbi.ucla.edu

Sophie Courtois
Aventis Pharma
Cambridge Genomics Center
26 Landsdowne Street
Cambridge MA 02139
Phone : (617)523 5043
Fax :(617)374 8811
Email : sophie.courtois2@aventis.com

Antonello Covacci
IRIS Chiron Spa
Via Fiorentina 1
53100 Siena
Italy
Phone: (+39) 0577-243235
Fax: (+39) 0577-243564
Email: antonello_covacci@biocine.it

Nancy Craig
HHMI/John Hopkins University
School of Medicine
725 North Wolfe St 601 PCTB
Baltimore, Maryland 21042
Phone: (410) 955-3933
Fax: (410) 955-0831
Email: ncraig@jhmi.edu

Julian Davies
TerraGen Discovery, Inc.
Suite 300-2386 East Mall-UBC
Vancouver, B.C., Canada V6T 1Z3
Phone: (604) 221-8896
Fax: (604) 221-8881
Email: JED@interchange.ubc.ca

Edward F. Delong
Monterey Bay Aquarium Research Institute
7700 Sandholdt Rd.
Moss Landing, CA 95039
Phone: (831) 775-1843
Fax: (831) 775-1645
Email: delong@mbari.org

Wen Deng
Dept. of Genetics
University of Wisconsin
445 Henry Mall
Madison, WI 53706
Phone: (608) 262-2534
Fax: (608) 263-7459
Email: wdeng@genome.wisc.edu

Patrick Dennis
Dept. of Biochemistry and Molecular Biology
University of British Columbia
2146 Health Sciences Mall
Vancouver, B.C. Canada V6T 1Z3
Phone: (604) 822-5975
Fax: (604) 822-5227
Email: patrick.p.dennis@ubc.edu

Joyce Duan
Molecular Biology Institute
University of California Los Angeles
611 Charles Young Drive
Los Angeles, CA 90095
Phone: (310) 206-3907
Fax: (310) 206-3914
Email: joyce@mbi.ucla.edu

Ashlee M. Earl
Louisiana State University
202 Life Science Building
Baton Rouge, LA 70803
Phone: (225) 388-5226
Fax: (225) 388-2597
Email: aearl@lsu.edu

Jonathan Eisen
The Institute of Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Phone: (301) 838-3507
Fax: (301) 838-0208
Email: jeisen@tigr.org

Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute of Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Phone: (301) 838-3525
Fax: (301) 838-0208
Email: nelsayed@tigr.org

Robert Feldman
Molecular Dynamics/Amersham Pharmacia Biotech
928 E. Arques Ave
Sunnyvale, CA 94086
Phone: (408) 737-3179
Fax: (408) 737-4887
Email: robert.feldman@am.apbiotech.com

Greg Ferry
The Pennsylvania State University
Dept. of Biochemistry and Molecular Biology
205 South Frear Building
University Park, PA 16802
Phone: (814) 863-5721
Fax: (814) 863-6217
Email: jgf3@psu.edu

Sorel Fitz-Gibbon
Dept. of Microbiology and Molecular Genetics
University of California, Los Angeles
1602 Molecular Sciences Building
Los Angeles, CA 90095
Phone: (310) 825-8924
Fax: (310) 206-3088
Email:Sorel@ewald.mbi.ucla.edu

Michael Fonstein
Integrated Genomics, Inc.
2201 W. Cambell Park. Dr.
Chicago, Il 60612
Phone: (312) 491-0846
Fax: (312) 491-0856
Email: Michael@integratedgenomics.com

Jean Francois-Tomb
Dupont Company. Experimentary Station
E328/B45
Wilmington, Delaware 19880
Phone: (302) 695-1166
Fax: (302) 695-1829
Email:
jean-francois.tomb@usa.dupont.com

Thomas Fulton
R80Y-235
Microbioal Vaccine Research
Merck Research Laboratories
P.O. Box 2000
126 E. Lincoln Ave,
Rahway, NJ 07065-0900
Phone: (732) 594-5082
Fax: (732) 594-1399
Email: tom_fulton@merck.com

Pauline Funchain
Dept. of Microbiology and Molecular Genetics
University of California, Los Angeles
1602 Molecular Sciences Building
Los Angeles, CA 90095
Phone: (310) 825-8924
Fax: (310) 206-3088
Email:pfunchain@yahoo.com

Theresa (Terry) Gaasterland
The Rockefeller University
1230 York Avenue
New York, NY 10021
Phone: (212) 327-7755
Fax: (212) 327-7765
Email: gaasterl@cleo.rockefeller.edu

Emilio Garcia
Lawrence Livermore National Laboratory
L0452, 7000 East Avenue,
Livermore, CA 94550
Phone: (925)422-1746
Fax: (925) 422-2282
Email: garcia12@llnl.gov

Vanessa Gabrovsky
Dept. of Microbiology and Molecular Genetics
University of California, Los Angeles
1602 Molecular Sciences Building
Los Angeles, CA 90095
Phone: (310) 825-8924
Fax: (310) 206-3088
Email: bisous@ucla.edu

Gerhard Gottschalk
Georg-August University
Inst. of Microbiology and Genetics
Grisebachstr 8
37077 Gottingen, Germany
Phone: (+49) 551-393781
Fax: (+49) 551-393808
Email: ggottsc@gwdg.de

Robert Gunsalus
Dept. Microbiolgy and Molecular Genetics
University of California Los Angeles
1602 Molecular Sciences Building
405 Hilgard Ave
Los Angeles, CA 90095
Phone: (310) 825- 5409
Fax: (310) 206-5231
Email: robg@microbio.ucla.edu

Terry Hazen
Lawrence Berkerly National Laboratory
One Cyclotron Rd
MS 70A-3317
Berkeley, CA 94720
Phone: (510) 486-5234
Fax: (510) 486-7152
Email: tchazen@lbl.gov

John Heidelberg
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850
Phone: (301) 315-2528
Fax: (301) 838-0208
Email: jheidel@tigr.org

Sabina Heim
German Research Centre For Biotechnology
Mascheroder Weg 1
D-38124 Braunschweig
Germany
Phone: (+49) 531 6181405
Fax: (+49) 531 6181411
Email: she@gbf.de

Wolfram Hemmer
GAIC GmbH
Jakob-Stadler Platz 7
D-78467 Konstanz
Germany
Phone: (+49) 7531-816012
Fax: (+49) 7531-816081
Email: wehemmer@gatc.de

Christopher D. Herring
University of Wisconsin
Dept. of Genetics
445 Henry Mall
Madison, WI, 53706
Phone: (608) 262-2534
Fax: (608) 263-7459
Email: cdherring@students.wisc.edu

Sarah Highlander
Baylor College of Medicine
One Baylor Plaza MS 280
Houston, TX 77030
Phone: (713) 798-6311
Fax: (713) 798-7375
Email: sarahh@bcm.tmc.edu

Helge Holo
Agricultural University of Norway
Department of Chemistry and Biotechnology
Lab of Microbial Gene Technology
P.O. Box 5051
N-1432 ÅS
Phone: (+47) 64 94 94 63
Fax: (+47) 64 94 14 65
Email: Helge.Holo@ikb.nlh.no

David J. Holmes
SmithKline Beecham
FP0825
P.O. Box 13681
Philadelphia, PA 19101-3681
Phone: (610)917-6453
Fax:
Email:

Al Ivens
The Sanger Centre
Wellcome Trust Genome Campus
Hinxton
Cambs. CB10 1SA
Phone: (+47) 1223 49 48 51
Fax: (+47) 1223 49 49 19
Email: alicat@sanger.ac.uk

Ethel Jackson
Dupont Company, Experimentary Station
E328/B45
Wilmington, Delaware 19880
Phone: (302) 695-4170
Fax: (302) 695-1829
Email: ethel.n.jackson@usa.dupont.com

Hongfan Jin
University of Wisconsin
Dept. of Genetics
445 Henry Mall
Madison, WI, 53706
Phone: (608) 262-2534
Fax: (608) 263-7459
Email: hongfan@genome.wisc.edu

Jingliang Ju
University of Nebraska-Lincoln
329 Food Industry Complex
Lincoln, NE 68583-0919
Phone: (402) 472-9177
Fax: (329) 472-1693
Email: ju09@hotmail.com

Udo Kaempf
Aventis Research
65929 Frankfurt
Bruenungstr. 50

Martin Keller
Diversa Corporation
10665 Sorrento Valley Road
San Diego, CA 92121
Phone: (858) 623-5162
Fax: (858) 623-5120
Email: mkeller@diversa.com

Jaehyoung Kim
University of Nebraska-Lincoln
329 Food Industry Complex
Lincoln, NE 68583-0919
Phone: (402) 472-0290
Fax: (402) 472-1693
Email: jkim7@bigred.unl.edu

Sun Chang Kim
KAIST
Dept. of Biological Sciences
373-1 Kusung-dong, Yusong-ku
Taejon, Korea
Phone: (+82) 42 869 2619
Fax: (+82) 42 869 2610
Email: sckim@mail.kaist.ac.kr

Sara Klink
University of Wisconsin
Dept. of Genetics
445 Henry Mall
Madison, WI, 53706
Phone: (608) 262-2534
Fax: (608) 263-7459
Email:sara@genome.wisc.edu

Eugene Kolker
Institute of Systems Biology
4225 Roosevelt Way,
Seattle, WA 98105-6099
Phone: (206) 732-1278
Fax: (206) 732-1299
Email: ekolker@systemsbiology.org

Huimih Kong
New England Biolabs
32 Tozer Rd
Beverly, MA 01915
Phone: (978) 927-5054
Fax: (978) 921-1350
Email: kong@neb.com

Konstantinos Konstantinidis
Michigan State University
Ctr for Microbial Ecology
540 Plant and Soil Science Bldg
East Lansing, MI 48824
Phone: (517) 353-9021
Fax: (517) 353-2917

Kristen Koretke
Bioinformatics
UP1345
SmithKline Beecham Pharmaceuticals
1250 South Collegeville Rd.
Collegeville, PA 19426-0989
Phone: (610) 917-7034
Fax: (610) 917-7901
Email: kristin_k_koretke@sbphrd.com

Bernard Labedan
Institut de Genetique et Microbiologie
Universite Paris-Sud, Bat 409
91405 Orsay Cedex, France
Phone: (+33) 1 69 15 57 18
Fax: (+33) 1 69 15 78 08
Email: labedan@igmors.u-psud.fr

Alla Lapidus
The University of Chicago
Dept. of Molecular Genetics and Cell Biology
920 East 58th Street
Chicago, IL 60637
Phone: (773)702-1081
Fax: (773)702-3172
Email: alapidus@midway.uchicago.edu

Frank W. Larimer
Joint Genome Institute
Microbial Sequencing Program

P.O. Box 2008
Oak Ridge, TN 37831-6480
Phone: (865) 574-1253
Fax: (865) 241-1965
Email: fwl@ornl.gov

Karen A. Lewis
National Cancer Institute, NIH
FCRDC, Bldg. 469, Rm 144
Miller Drive
Frederick, MD 21702-1201
Phone: (301) 846-5581
Fax: (301) 846-5598
Email: lewiska@ncifcrf.gov

Xiaoying Lin
Celera Genomics
45 West Gude Dr
Rockville, MD 21029
Phone: (240) 453-3695
Fax: (240) 453-3303
Email: linxj@celera.com

Gui-Rong Liu
Dept. of Microbiology and Infectious Diseases
University of Calgary
3330 Hospital Dr.
NW, Calgary, AB Canada
Phone: (403) 220-3799
Fax: (403) 270-0834
Email: grliu@ucalgary.ca

Jia Liu
Parke-Davis Pharmaceutical Research
2800 Plymouth Rd.
Ann Arbor, MI 48105
Phone:(743)622-5127
Fax:(743)622-7158
Email: jia.liu@wl.com

Shu-Lin Liu
Dept. of Microbiology and Infectious Disease
University of Calgary Faculty of Medicine
3330 Hospital Dr. NW
Canada
Calgary, AB T2N 4N1
Phone: (403)220-8661
Fax: (403)270 0834
Email: slliu@ucalgary.ca

Todd Lowe
Dept. of Genetics
Stanford University Medical School
300 Pasteur Drive, M322
Stanford, CA 94305
Phone: (650) 498-6012
Fax: (650) 723-7016
Email: Lowe@genome.stanford.edu

Jin Lu
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, Indiana, 46285
Phone: (317) 277-5672
Fax: (317) 277-0778
Email: lu_jin@lilly.com

Andrei Lupas
Bioinformatics
UP1345
SmithKline Beecham Pharmaceuticals
1250 South Collegeville Rd.
Collegeville, PA 19426-0989
Phone: (610) 917-7312
Fax: (610) 917-7901
Email: Andrei_N_Lupas@sbphrd.com

Edward Marcotte
Protein Pathways
1145 Gayley Ave, Suite 304
Los Angeles, CA 90024
and
Molecular Biology Institute 229
University of California Los Angeles
Los Angeles, CA 90095-1570
Phone: (310) 825-9847
Fax: (310) 206-3914
Email: marcotte@mbi.ucla.edu

Bill Marshall
Bioinformatics
UP1345
SmithKline Beecham Pharmaceuticals
1250 South Collegeville Rd.
Collegeville, PA 19426-0989
Phone: (610) 917-5388
Fax: (610) 917-7901
Email: bill_marshall-1@sbphrd.com

Partrick Martin
Microcide Pharmaceuticals, Inc.
850 Maude Ave
Mountain View, CA 94043
Phone: (650) 428-3510
Fax: (650) 428-3550
Email: martin@microcide.com

Millicent Masters
Edinborgh University
Institute of Cell and Molecular Biology
Darwin Building, Kings Building,
Mayfield Road,
Edinborgh EH9 3JR
Phone: (+44) 131 650 5355
Fax: (+44) 131 650 8650
Email: m.masters@ed.ac.uk

Max Mergeay
Lab Microbiologym Radioactive Waste &
Clean-up Division, SCK/CEN
Boeretang 200,
B-2400-MOL-Belgium
Phone: (+32) 14379189
Fax: (+32) 14372488
Email: mmergeay@sckcen.be

Christophe Merlin
Edinborgh University
Institute of Cell and Molecular Biology
Darwin Building, Kings Building,
Mayfield Road,
Edinborgh EH9 3JR
Phone: (+44) 131 650 5356
Fax: (+44) 131 650 8650
Email: c.merlin@ed.ac.uk

Jeffrey H. Miller
Dept. of Microbiology and Molecular
Genetics
University of California, Los Angeles
1602 Molecular Sciences Building
Los Angeles, CA 90095
Phone: (310) 825-8460
Fax: (310) 206-3088
Email: Jhmiller@mbi.ucla.edu

Roger Milkman
Biological Sciences
University of Iowa
138 Biology Bldg
Iowa City, IO 52242-1324
Phone: (319) 335-1077
Fax: (319) 335-1069
Email: roger-milkman@uiowa.edu

Vladimir Motin
Lawrence Livermore National Laboratory
L-452, 7000 East Avenue
Livermore, CA 94550
Phone:(925) 422-1746
Fax: (925) 422-2282
Email: motin1@llnl.gov

John P. Mueller
Pfizer Global R&D - Groton
RAIIID Biology - Infectious Diseases
Bldg. 118D Rm D309A
Mailstop MS8118D-3087
Eastern Point Road
Groton, CT 06340
Phone: (860)441-3880
Fax: (860)715-8062
Email: muellerjp@groton.pfizer.com

Ellen Murphy
Wyeth-Ayerst Research
401 N. Middletown Road
Pearl River, NY 10965
Phone: (845) 732-4411
Fax: (845) 732-5671
Email: murphye3@war.wyeth.com

Alison E. Murray
Center for Microbial Ecology
540 PSSB
Michigan State University
East Larsing, MI 48840
Phone: (517) 353-9021
Fax: (517) 353-2917
Email: murraya@msu.edu

Peter Myler
Dept. of Pathobiology
University of Washington
Seattle, Wa 98195
Phone: (206) 284-8846 x332
Fax: (206) 284-0313
Email: mylerpj@sbri.org

Laila Alves Nahum
Marine Biological Laboratory
7 MBL Street
Woods Hole, MA 02543-1015
Phone: (508) 289-7388
Fax: (508) 289-7388
Email: laanahum@mbl.edu

John Nash
Institute for Biological Sciences
National Research Council of Canada
Room 3135, 100 Sussex Dr.
Ottawa, Ontario Canada, K1A 0R6
Phone: (613) 990-6236
Fax: (613) 952-9092
Email: john.nash@nrc.ca

Ingolf F. Nes
Agricultureal University of Norway
Dept. of Chemistry and Biotechnology
Lab of Microbial Gene Technology
P.O. Box 5051
N-1432 ÅS
Phone: (+47) 64 94 94 63
Fax: (+47) 64 94 14 65
Email: ingolf.nes@ikb.nlh.no

Eva Ng
Biozentrum der Uni-wuerzburg
Lehrstuld fuer Mikobiologie
Am Hubland
97074 Wuerburg, Germany
Phone:(+49) 931-888-4413
Fax: (+49) 931-888-4402
Email: eva@biozentrum.uni-wuerghurg.de

Steven Norris
Dept. of Pathology and Lab Medicine
University of Texas Houston Medical School
6431 Fannin, MSB 2.292
Houston, TX 77030
Phone: (713) 500-5338
Fax: (713) 500-0730
Email: steven.j.norris@uth.tmc.edu

James Martin Odom
Dupont Company
P.O. Box 80328
Wilmington, DE 19880-0328
Phone: (302) 695-3573
Fax: (302) 695-1577
Email: J-Martin.Odom@usa.dupont.com

Arina Omer
University of British Columbia
Department of Biochemistry and Molecular Biology
Faculty of Medicine
2146 Health Science Mall
Vancouver BC, V5T 1Z3
Phone: (604) 822-5245
Fax: (604) 822-5227
Email: arina@interchange.ubc.ca

Ross Overbeek
Integrated Genomics, Inc.
2201 W. Campbell Park Drive
Chicago, Il 60612
Email: ross@integratedgenomics.com

Timothy Palzkill
Baylor College of Medicine
Houston, Texas 77030
Phone: (713) 798-5609
Fax: (713) 798-7375
Email: timothyp@bcm.tmc.edu

Norbert Peekhaus
Merck & Co, Rahway
P.O. Box 2000
Rahway, NJ 07065
RY80Y-205
Phone: (732) 594-0593
Fax: (732) 594-1399
Email: norbert_Peekhaus@merck.com

Alida Pellegrini-Toole
Marine Biological Laboratory
7 MBL Street
Woods Hole, MA 02543
Phone: (508) 548-3705
Fax: (508) 540-6902
Email: ptoole@mbl.edu

Nicole Perna
Dept. of Genetics
University of Wisconsin
445 Henry Mall
Madison, WI 53706
Phone: (608) 262-2534
Fax: (608) 263-7459
Email: nicole@genome.wisc.edu

Guy Plunkett III
Dept. of Genetics
University of Wisconsin
445 Henry Mall
Madison, WI 53706
Phone: (608) 262-2534
Fax: (608) 263-7459
Email: guy@genome.wisc.edu

Steffen Porwollik
Sidney Kimmel Cancer Center
10835 Altman Row
San Diego, CA 92121
Phone: (858) 450-5990 x 274
Fax: (858) 550-3998
Email: sporwollik@skcc.org

György Pósfai
Biological Research Center of the Hungarian Academy of Sciences
Temesvari Krt 62
Szeged, H 6726
Hungary
Phone: (+36) 62 432-232
Fax: (+36) 62 433 506
Email: posfaigy@nucleus.szbk.u-szeged.hu

Elisabeth A. Raleigh
New England Biolabs, Inc.
32 Tozer Road,
Beverly, MA 01915
Phone: (978) 927-5054 ext 238
Fax: (978) 921-1350
Email: raleigh@neb.com

Tom Ramseier
The Dow Chemical Company
5501 Oberlin Dr.
San Diego, CA 92121
Phone: (858) 352-4357
Fax: (858) 352-4557
Email: tmramseier@dow.com

Monica Riley
Marine Biological Laboratory
7 MBL Street
Woods Hole, MA 02543
Phone: (508) 548-3705
Fax: (508) 540-6902
Email: mriley@mbl.edu

Frank T. Robb
Center for Marine Biotechnology
University of Maryland
Biotechonology Institute
701 E. Pratt Street,
Baltimore, MD 21202
Phone: (410) 234-8870
Fax: (410) 234-8896
Email: robb@umbi.umbd.edu

Margaret Romine
Environmental Microbiology Group
Pacific Northwest National laboratory
PO Box 999 MS P7-50
Richland, WA 99352
Phone: (509) 376-8287
Fax: (509) 376-1321
Email: Margie.Romine@pnl.gov

Rui Rong
Clemson University
Dept. of Forest Resources
261 Lehotsky Hall
Clemson, SC 29634-0331
Phone: (864) 656-4400
Fax: (864) 656-3304
Email: rrong@clemson.edu

Junichi Ryu
Dept. of Microbiology and Molecular Genetics
Loma Linda University
Loma Linda, CA 92350
Phone: (909) 558-1000 ext. 42773
Fax: (909) 558-4035
Email: jryu@som.llu.edu

Neil Saunders
School of Microbiology and Immunology
The University of New South Wales
Sydney, UNSW 2052
Australia
Phone: (+61) 2 9385 2093
Fax: (+61) 2 9385 1591
Email: neil.saunders@unsw.edu.au

Ryan K. Shultzaberger
National Cancer Institute, NIH
FCRDC, Bldg. 469, Rm 144
Miller Drive
Frederick, MD 21702-1201
Phone: (301) 846-5581
Fax: (301) 846-5598
Email: Shultzab@ncifcrf.gov

Richard Siehnel
Anti-Infectives Research
Procter & Gamble Pharmaceutical Company
8700 Mason-Montgomery Road
Mason, OH 45040
Phone: (513) 622-3904
Fax: (513) 622-0085
Email: siehnel.rj@pg.com

Alexel Slesarev
Fidelity Systems, Inc.
7961 Cessna Avenue,
Gaithersburg, MD 20879
Phone: (301) 527-0804
Fax: (301) 527-8250
Email: alex@fidelitysystems.com

Gosia Slupska
Dept. of Microbiology and Molecular Genetics
University of California Los Angeles
439 Boyer Hall
405 Hilgard Ave
Los Angeles, CA 90095
Phone: (310) 825-8924
Fax: (310) 206-3088
Email: gosia@mbi.ucla.edu

Richard Smith
Environmental Molecular Sciences Laboratory
Pacific Northwest National laboratory
Mail Stop K8-98
Richland, WA 99352
Phone: (509) 376-0723
Fax: (509) 376-7722
Email: dick.smith@pnl.gov

Michael Stanhope
Bioinformatics
UP1345
SmithKline Beecham Pharmaceuticals
1250 South Collegeville Rd.
Collegeville, PA 19426-0989
Phone: (610) 917-6577
Fax: (610) 917-7901
Email: michael_j_stanhope@sbphrd.com

Sergei Stoliar
Institute for Systems Biology
4225 Roosevelt Way, NE STE 2000
Seattle, WA 98105-6099
Phone: (206) 732-1278
Fax: (206) 732-1299
Email: sstoliar@systemsbiology.org

Charles K. Stover
Pathogenesis Corporation
201 Elliott Ave West
Suite 150
Seattle, WA 98199
Phone: (206) 270-3309
Fax: (206) 282-5065
Email: Kstover@pathogenesis.com

Michael Sträetz
German Research Centre For Biotechnology
Mascheroder Weg 1
D-38124 Braunschweig
Germany
Phone: (+49) 531 6181408
Fax: (+49) 531 6181411
Email: msr@gbf.de

Clemens Suter-Crazzolara
LION Bioscience AG
Waldhofer Str. 98
69123 Heidelberg
Germany
Phone: (+49) 6221 4038 365
Fax: (+49) 6221 4038 290
Email: suter@lionbioscience.com

Richard Stephens
School of Public Health
University of California, Berkeley
235 Earl Warren Hall
Berkeley, CA 94720
Phone: (510) 643-9900
Fax: (510) 643-1537
Email: rss@uclink4.berkeley.edu

Eduardo Taboada
Institute for Biological Sciences
National Research Council of Canada
Room 3135, 100 Sussex Dr.
Ottawa, Ontario Canada, K1A 0R6
Phone: (613) 990-6236
Fax: (613) 952-9092
Email: ed.taboada@nrc.ca

Dorothea K. Thompson
Oak Ridge National Laboratory
Environmental Sciences Division
Building 1505, MS 6038
Oak Ridge, TN 37831
Phone: (865) 574-4815
Fax: (865) 576-8646
Email: thompsondk@ornl.gov

Marc Vidal
Dept. of Genetics
Dana-Farber Cancer Institute
Harvard Medical School
Smith Bldg, 858A
44 Binney Street
Boston, MA 02115
Phone: (617) 632-5180
Fax: (617) 632-5739
Email: Marc_Vidal@dfci.harvard.edu

Hylckema Vlieg
Flavour and Natural Ingredients
Kernhemseweg 2
P.O. Box 20
6710 BA Ede
The Netherlands
Phone: (+31) 318 659 554
Fax: (+31) 318 650 400
Email: hylckama@nizo.nl

Craig Voelker
Bioinformatics
UP1345
SmithKline Beecham Pharmaceuticals
1250 South Collegeville Rd.
Collegeville, PA 19426-0989
Phone: (610) 917-6267
Fax: (610) 917-7901
Email: craig_2_volker@sbphrd.com

Guido Volckaert
University of Leuven
Lab of Gene Techonology
Kardinaal Mercierlaan 92
B-3001 Leuven, Belgium
Phone: (+32) 76-329662
Fax: (+32) 76-321965
Email: gvolckaert@agr.kuleuven.ac.bc

Veronika Vonstein
Integrated Genomics, Inc.
2201 W. Campbell Park Drive
Chicago, Il 60612
Phone: (312) 491-0846
Fax: (312) 491-0856
Email: Veronika@integratedgenomics.com

Patrick Warren
Bioinformatics
UP1345
SmithKline Beecham Pharmaceuticals
1250 South Collegeville Rd.
Collegeville, PA 19426-0989
Phone: (610) 917-6388
Fax: (610) 917-7901
Email: patrick_v_warren@sbphrd.com

Paul Warrener
Pathogenesis Group
Seattle, WA 98119
Phone: (206) 664-6165
Fax: (206) 282-5065
Email: pwarrener@pathogenesis.com

George Weinstock
Dept. of Microbiology and Molecular Genetics
University of Texas Medical School
6431 Fannin Street
Houston, TX 77030
Phone: 713-500-6083
Fax: 713-500-0652
Email:
georgew@utmmg.med.uth.tmc.edu

Walter Weyler
Genencor International
925 Page Mill Rd
Palo Alto, CA 94304
Phone: (650) 846-7539
Fax: (650) 621-7939
Email: wweyler@genencor.com

Owen White
The Institute of Genomic Reseach
9712 Medical Center Drive
Rockville, MD 20850
Phone: (301) 838 3534
Fax: (301) 838 0209
Email: owhite@tigr.org

Kelly Winterberg
Dept. of Biochemistry
University of Wisconsin
433 Babcock Dr.
Madison, WI 53706
Phone: (606) 262-4550
Fax: (606) 265-2603
Email: kmwinterberg@excite.com

John Wise
University of Nebraska
Dept. of Food Science and Technology
143 Filly Hall
Lincoln, NE 68583-0919
Phone: (402) 972-5637
Fax: (402) 972-1693
Email: jwise@cse.unl.edu

Kai Wu
Affymetrix Inc.
3380 Central Expressway, RM 618
Santa Clara, CA 95051
Phone: (408) 731-5510
Fax: (408) 481--435
Email: kai_wu@affymetrix.com

Yan Xu
New England Biolabs
32 Tozer Road
Beverly, MA 01915
Phone: (978) 927-5054
Fax: (978) 927-1350
Email: xuy@neb.com

Robert Zagursky
Wyeth-Lederle Vaccine
211 Bailey Road
West Henrietta, NY 14586-9728
Phone: (716) 273-7619
Fax: (716) 273-7515
Email: zagursrj@war.wyeth.com

Jizhong Zhou
Oak Ridge National Lab
P.O. Box 2008
Bethel Valley Road
Oak Ridge, TN 37831-6038
Phone: (865) 576-7544
Fax: (865) 576-8646
Email: zhou@ornl.gov

Anthony Zuccarelli
Dept. of Microbiology and Molecular
Genetics
AHBS 115
Loma Linda University
Loma Linda, CA 92350
Phone: (909) 558-4480
Fax: (909) 558-4035
Email: azuccarelli@som.llu.edu

Preliminary Program

All speakers confirmed except where indicated by an asterisk (*)

Sunday, September 24th.

Arrival Registration. 4:00 to 6:00 pm

Dinner 6:15 pm

Evening

Keynote Speaker - Julian Davies, TerraGen Discovery, Inc.
"Evolution of Microbial Resistance to Antibiotics"

Reception following lecture.

Monday morning, September 25th

Session 1 - Genomes of Pathogens, and the Evolution of Pathogenicity

Stephen Bentley, The Sanger Centre
"Sanger Centre Pathogen Genome Projects"

Najib El-Sayed, The Institute for Genomic Research
"Analysis of Protozoan Pathogen Genomes at TIGR"

George Weinstock, University of Texas Medical School
"Spirochete Genomes: What We Have Learned"

Richard Stephens, University of California, Berkeley
"The *Chlamydia pneumoniae* Genome"

Frederick Blattner, University of Wisconsin-Madison
"Comparative Studies of Enteric Pathogens"

Evening

Session 2 - Genomes of Pathogens (cont.)

Antonello Covacci, IRIS Chiron Spa
"Analysis of the Meningococcal Genome"

Siv Andersson, Uppsala University
"Analysis of the *Bartonella henselae* Genome"

Al Ivens, The Sanger Centre
"Analysis of Protozoan Pathogen genomes"

Peter Myler, Seattle Biomedical Research Institute
"Trypanosomes"

Nancy L. Craig, HHMI/Johns Hopkins School of Medicine
"Using Tn7 to Dissect Genomes"

Tuesday Morning, September 26th

Session 3 - Bioremediation

Terry Hazen, Lawrence Berkeley National Laboratory
"Bioremedation: The Hope and the Hype"

Frank Larimer, Oak Ridge National Laboratory
"Genomic Survey of Key Autotrophic Microorganisms in the Natural Carbon Cycle"

Frank T. Robb, University of Maryland Biotechnology Institute
"A DNA Analysis Approach to Bioremediation and Microbial Physiology"

Margaret Romine, Pacific Northwest National Labratory
"*Sphingomonas* "

John Heidelberg, The Institute for Genomic Research
"Genomic Sequence of *Shewanella oneidensis* MR-1"

Evening

Session 4 - Evolution of Biodiversity

Jeffery H. Miller, UCLA
"The Mismatch Repair System and Horizontal Transfer"

Martin Keller, Diversa Corp.
"Microbial Biodiversity"

Edward F. DeLong, Monterey Bay Aquarium Research Institute
"Genomic Windows into the Natural Microbial World"

Sophie Courtois, Aventis Pharma
"Exploring Uncultivated Microorganisms for Natural Products Drug Discovery"

Jonathan Eisen, The Institute for Genomic Research
"The Evolution of DNA Repair Systems"

Monica Riley, Woods Hole
"Mechanisms of Divergence from Ancestral Enzymes"

Wednesday morning, September 27th

Session 5- Genomic Technology, Bioinformatics, and Functional Genomics

Theresa (Terry) Gaasterland, Rockefeller University
"Integration of Genome Annotation and Microarray Data"

Owen White, The Institute for Genomic Research
"Bacterial Annotation in a High-throughput Sequencing Environment"

Paul Warrener, Pathogenesis Corporation
"The *Pseudomonas aeroginosa* Genome Project: Building the Pipeline from Annotation to Target Discovery"

Gary Anderson, Lawrence Livermore National Laboratory
"Use of Suppression Subtractive Hybridization as a Tool for Whole Genome Comparisons"

Tim Palzkill, Baylor College of Medicine
"Developing Phage Display for Functional Genomics"

Michael Fonstein, Integrated Genomics, Inc.
"Comparative Genomics"

Evening

Session 6 - Proteomes and Cellular Pathways

Marc Vidal, Massachusetts General Hospital Cancer Center
"Protein Interaction Mapping: Its Use in *C. elegans* "

Alex Bateman, Sanger Center, United Kingdom.
"Protein Function Analysis through the PFAM Data Base"

Richard Smith, Pacific Northwest National Laboratory
"New Technology for the Rapid and Precise Study of Proteomes: Global Views for the Understanding of the Cellular Processes"

Edward Marcotte, Protein Pathways
"Genome-wide Prediction of Protein Function"

Ross Overbeek, Integrated Genomics
"Filling in the Missing Pieces of Core Metabolisms"

Thursday Morning, September 28th

Session 7- Extremophiles

J. Martin Odom, Dupont
"New Insights into the Biology of an Obligate Methanotroph"

John R. Battista, Louisiana State University and A&M College
"The *Deinococcus radiodurans* Genome, and Its Application to Understanding Extraordinary Resistance to Radiation"

Sorel Fitz-Gibbon, UCLA
"The Fully Annotated Genome of *Pyrobaculum aerophilum* "

Patrick Dennis, University of British Columbia
"Small Nucleolar RNAs in Archaea"

Lunch

HOME | **PROGRAM** | REGISTRATION | POSTERS | CONFERENCE SITE