ASM Conference on

Small Genomes

September 20-24, 1998
Lake Arrowhead, California

American Society for Microbiology

ASM Conference on

Small Genomes

September 20-24, 1998
Lake Arrowhead, California

CONTENTS:

Program Information 1

Scientific Program 2

Speaker Abstracts 7

Poster Abstracts 21

Index 33

American Society for Microbiology

PROGRAM INFORMATION

SCIENTIFIC PROGRAM ORGANIZERS

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

George Weinstock, Co-chair
University of Texas-Houston Health Science Center

Nancy Craig
Johns Hopkins University, Baltimore, Maryland

Elizabeth Raleigh
New England BioLabs, Beverly, Massachusetts

Monica Riley
Marine Biological Laboratories, Woods Hole, Massachusetts

John Roth
University of Utah, Salt Lake City

ACKNOWLEDGMENTS

The American Society for Microbiology and the Scientific Program Organizers gratefully acknowledge the contributions from the following organizations in support of this conference:

DNA Star

National Human Genome Research Institute (NHGRI), NIH

National Institute of Allergy and Infectious Diseases (NIAID), NIH

National Science Foundation

New England BioLabs, Inc.

PathoGenesis Corporation

SmithKline Beecham Pharmaceuticals

REGISTRATION

The conference registration desk is located outside the Pineview Room in the conference center. ASM staff will be available to assist you with any questions or requests during session hours.

GENERAL SESSIONS

All general sessions will be held in the Pineview Room, located in the conference center.

POSTER SESSIONS

Posters will be displayed throughout the meeting in the Pineview Room. Presenters will attend their poster boards during scheduled poster session hours on Monday and Tuesday.

MEALS

All meals will be served in the Dining Room during scheduled times as detailed in the Program. Meals are for registered participants or guests paying the full-package fee only.

GUEST ACCOMMODATIONS

Due to space constraints, accommodations for guests are extremely limited. Adult guests of registered conference participants may attend meals only if they have registered as a guest of the conference center and paid the $88 daily fee in advance.

SCIENTIFIC PROGRAM

SUNDAY, SEPTEMBER 20

3:30-6:00 pm	Arrival and Check-in at Lake Arrowhead Conference Center
5:15-7:15 pm	Reception (Iris Room)
7:30-9:00 pm	Dinner (Dining Room)
9:00-11:00 pm	Mixer (Iris Room)

MONDAY, SEPTEMBER 21

7:45-8:30 am	Breakfast (Dining Room)
Session I	**Completed Genomes, Genome Comparisons, and Evolution** (Pineview Room)
8:45-9:15 am	Jeffrey H. Miller, UCLA Dept. of Microbiology & Molecular Genetics *Welcome/Introduction*
9:15-9:50 am	Nicole Perna, Univ. of Wisconsin-Madison *" Comparative Genomics of E. coli K-12, O157:H7 and Related Enterobacterial Pathogens."*
9:50-10:15 am	Andy Benson, University of Nebraska *"High-Resolution Genomic Scanning of Epidemiologically Characterized E. coli O157:H7 Isolates."*
10:15-10:50 am	Robert Haselkorn, University of Chicago *"The Rhodobacter capsulatus Genome Project."*
10:50-11:05 am	Break
11:05-11:40 am	Sorel Fitz-Gibbon, UCLA Dept. of Microbiology & Molecular Genetics *"Whole Genome Sequence of Pyrobaculum aerophilum."*
11:40 am-12:15 pm	Elizabeth Kutter, Evergreen State College *"T-even Phages: Genomics, Evolution and Therapeutic Potentials."*
12:30 pm	Lunch (Dining Room)
4:00-6:15 pm	Poster Session (Pineview Room)
4:00-6:15 pm	Social/Mixer (Lakeview Room)
6:30-8:00 pm	Dinner (Dining Room)
Session II	**Regulatory Networks, Proteomes, and Annotation** (Pineview Room)
8:00-8:35 pm	George M. Church, Harvard Medical School *"Whole Genome Approaches to Regulatory Networks and Phenotypes."*
8:35-9:10 pm	Terry Gaasterland, Argonne National Laboratory & Univ. of Chicago *"Automated Genome Annotation Through Whole Genome Comparisons."*
9:10-9:45 pm	Monica Riley, Marine Biological Lab , Woods Hole, MA *"What We Know About the so-called Unknown Orfs in Sequenced Microbial Genomes."*
9:45-10:00 pm	Break
10:00-10:35 pm	Bernard Labedan, Universite Paris-sud *"Use of the Comparison of Families of Paralogous Proteins to Trace Back the Evolution of*

Proteins and of Procaryotic Genomes."

10:35-10:55 pm Hannes Loferer, Geneva Biomedical Research Institute, Geneva, Switzerland
"A Genome-Based Approach to Identify Novel Essential Bacterial Genes."

TUESDAY, SEPTEMBER 22

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

Session III **Chromosome Structure, Inteins, and Integrons** (Pineview Room)

8:45-9:20 am Didier Mazel, U.P.M.T.G.- Dept des Biotechnologies Institut Pasteur
" Super-integrons in Vibrio Genomes."

9:20-9:55 am Pat Higgins, University of Alabama at Birmingham
" Chromosomal Domain Structure in Salmonella typhimurium. "

9:55-10:30 am Elisabeth Raleigh, New England BioLabs, Inc., Beverly, MA
"Genomic Carpetbaggers: Restriction-Modification Genes in Prokaryotes."

10:30-10:45 am Break

10:45-11:20 am John Reeve, The Ohio State University
"Archaeal Histones and Nucleosomes."

11:20-11:55 am Maurice W. Southworth, New England Biolabs, Inc.
"Protein Splicing."

11:55 am-12:15 pm I.B. Zhulin, Loma Linda University
"Chemotaxis Operons in Microbial Genomes: Comparison to What is Known."

12:30 pm Lunch (Dining Room)

4:00-6:00 pm Poster Session (Pineview Room)

4:00-6:00 pm Social/Mixer (Lakeview Room)

6:15-7:45 pm Dinner (Dining Room)

Session IV **Completed Genomes and Genomics** (Pineview Room)

7:45-8:20 pm Karen E. Nelson, The Institute for Genomic Research
" The Thermotoga Maritima Msb8 Genome."

8:20-8:55 pm George Weinstock, Univ. of Texas Med. School at Houston
"Biological Systems in Treponema pallidum: A Black Box Opened by Genomics "

8:55-9:20 pm Michael Fonstein, Integrated Genomics Inc.
"Metabolic Reconstruction of Zymomonas mobilis and Thiobacillus ferrooxidans Based on Their Gapped Genome Sequencing."

9:20-9:35 pm Break

9:35-10:10 pm Chris Lee, UCLA Dept. of Chemistry and Biochemistry
"Automated Microbial Proteome Annotation and Data-Mining."

10:10-10:30 pm Frank Robb, Center of Marine Biotechnology, University of Maryland
"Lateral Gene Transfer and Genome Divergence Inferred from the Genomic Sequences of the Closely Related Hyperthermophilic Archaea, Pyrococcus furiosus , Thermococcus litoralis and P. horikoshii."

WEDNESDAY, SEPTEMBER 23

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

Session V **Recombination and Repair** (Pineview Room)

8:45-9:20 am Jeffrey H. Miller, UCLA Dept. of Microbiology & Molecular Genetics
"DNA Repair: From Bacteria to Humans, to Archaea."

9:20-9:55 am John Battista, Louisiana State Univ.
"DNA Repair in Deinococcus radioudrans: Adapting to Life in Dry Environments."

9:55-10:30 am Alvin J. Clark, University of California, Berkeley
Jonathan Eisen, Stanford University
"DNA Repair Genes in Archaea and Bacteria."

10:30-10:45 am Break

10:45-11:10 am Jocelyne DiRuggiero, Center of Marine Biotechnology, Univ. of MD
"DNA Repair Mechanisms In Hyperthermophilic Archaea."

11:10-11:45 am Thomas A. Cebula, Division of Molecular Biological Research and Evaluation, FDA, Wash., D.C.
"Mutators Among Escherichia coli and Salmonella enterica: Adaptation and Emergence of Bacterial Pathogens "

11:45 am-12:10 pm Alexei Slesarev, UCLA Dept. of Molecular, Cell & Dvlpmnt. Biology
"Studies of Proteins Regulating DNA Topology in Hyperthermophile Methanopyrus kandleri."

12:30 pm Lunch

4:00-6:15 pm Social/Mixer (Iris Room)

6:30-8:00 pm Dinner (Dining Room)

Session VI **General Genomics and Molecular Genetics** (Pineview Room)

8:00-8:35 pm Patrick Forterre, IGM
" Molecular Phylogeny in Crisis: the Case of the Universal Tree of Life."

8:35-9:00 pm Karl A. Reich, Abbott Labs
"Genome Scanning in Haemophilus influenzae: in vitro Transposition and PCR Analysis for the Identification of Essential Genes."

9:00-9:20 pm Igor Yu Goryshin, University of Wisconsin-Madison
"Tn5 Transposition: A Simple Efficient Tool for Manipulating DNA Molecules in vitro."

9:20-9:35 pm Break

9:35-9:55 pm Matthew Biery, Howard Hughes Medical Institute at Johns Hopkins
Nancy Craig, Howard Hughes Medical Institute at Johns Hopkins
"A High Efficiency, Low Target Site Specificity in vitro Transposition Reaction: A Powerful Tool for in vitro Random Mutagenesis and Sequencing."

9:55-10:15 pm Sherwood Casjens, University of Utah Medical Center
"Pseudogenes on the Borrelia Linear Plasmids."

THURSDAY, SEPTEMBER 24

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

Session VII **Short Talks** (Pineview Room)

8:45-9:05 am R.L. Charlebois, University of Ottawa
"The NeuroCadgets Inc./Univ. of Ottawa Bioinformatics Web Site."

9:05-9:25 am T-T Tseng, University of California, San Diego
M.H. Saier Jr., University of California, San Diego
"Recent Advances in Characterization of the Major Facilitator Superfamily (MFS) Superfamily (MFS)."

9:25-9:45 am	Patrick V. Warren, SmithKline Beecham Pharmaceuticals *"Gleanings from a Comparative Genomics Study: Another Look at DNA Replication."*
9:45-10:20 am	Craig Richmond, University of Wisconsin, Madison, WI *"Genome Wide Expression Analysis and Comparative Genomics in E. coli K-12."*
10:20-10:55 am	Meeting Summary; Planning Next Meeting
11:00 am	Check-out
11:30 am	Departure of 1st Conference Bus for LAX
12:00 pm	Lunch (Dining Room)
1:15 pm	Departure of 2nd Conference Bus for LAX

ABSTRACTS FOR PRESENTATION BY SPEAKER

Monday, September 21
Session I - Completed Genomes, Genome Comparisons, and Evolution

SA-01. Comparative Genomics of *E. coli* K-12, O157:H7 and Related Enterobacterial Pathogens.
N.T. Perna*, V. Burland, G. Plunkett III, J. Gregor, G.F. Mayhew, D.J. Rose, Y. Shao, F. R. Blattner.
University of Wisconsin, Madison, WI.

We have nearly finished sequencing the complete genome of Escherchia coli O157:H7 strain EDL933. A comparison of this genome with our published sequence from E. coli K-12 provides insight into what distinguishes the deadly pathogen from the harmless laboratory strain. Sequence data from a whole-genome shotgun is initially compared to the complete E. coli K-12 sequence to identify elements comprising a common "backbone." Insertions, deletions, substitutions, and rearrangements between the two strains are noted as deviations from this framework. The O157:H7 genome is about 5,463,000 bp. Of this, 4,171,000 bp are shared with E. coli K-12, and approximately 500,000 bp of the K-12 genome are absent from O157:H7. The 1,292,000 bp unique to O157:H7 cluster into about 200 regions, ranging in size from less than 100 bp to 61676 bp. The base composition of many of these regions is atypical of E. coli, suggesting acquisition via horizontal transfer. Three large genetic elements of particular relevance to virulence, LEE, bacteriophage 933W and the large plasmid pO157, have been characterized more completely. We are extending this study to include a number of additional enterobacterial pathogens, and have collected sequence data from a uropathogenic E. coli strain, Shigella flexneri, Salmonella typhi, and Yersinia pestis. These sequences begin to define a "pathosphere", the set of virulence determinants in the enterobacterial gene pool.

SA-02. High-resolution Genomic Scanning of Epidemiologically Characterized *E. coli* O157:H7 Isolates.
JAEHYOUNG KIM[1], JOSEPH NIETFELDT[1], ABRAHAM OOMMEN[2], and ANDY BENSON[1]*
[1]Univ. of Nebraska, Lincoln, NE and [2]Li-Cor, Inc., Lincoln, NE.

The Enterohemorraghic *E. coli* (EHEC) have emerged as a major public heath concern. Although these strains can be isolated from cattle only at a low frequency, longitudinal studies have indicated that certain genotypes of the O157:H7 serotype can be consistently isolated (persistent genotypes) from cattle herds over time, despite the appearance and disappearance of other O157:H7 genotypes (periodic genotypes) that are apparently less fit to propagate or survive in this environment. Understanding the relationship of these persistent and periodic genotypes to those genotypes isolated from human cases may provide information about the pathogenic potential of these strains and insight into the nature of selective forces that may be imposed in the herd environment. To assess these relationships, we have developed a high-resolution genome scanning technique that affords full genome coverage at single-nucleotide resolution. The strategy is based on the use of PCR primers to amplify short segments of DNA between frequently occurring sequences and subsequent resolution of the products on an automated DNA sequencer. Based on the sum of the band sizes from representative reactions, approximately 5% of the genome can be amplified and resolved in a single reaction and thus, genome coverage can theoretically be achieved with 20 primer combinations. Using this approach, we have compared a set of 30 distinct PFGE subtypes of O157:H7 strains consisting of human and cattle (periodic and persistent genotypes) isolates from a limited geographical region. Binary files representing the presence/absence of reaction products were created by image analysis and used to determine the relatedness of the strains determined by PAUP. These results indicate that most of the cattle isolates form a cluster that is distinct from the human isolates. This apparent dimorphism was readily observed by the presence of several dimorphic bands from the genome scanning reactions. Based on the average frequency of occurrence of priming sites and the fact that the dimorphic bands can be observed in several primer combinations, it is likely that they originate from a single region of the chromosome that is of significant size.

SA-03. The *Rhodobacter capsulatus* Genome Project.
R. Haselkorn, Y. Kogan, J. Paces, A. J. Milgram, D. Rebrekov, R. Cox and M. Fonstein.
Dept of Molecular Genetics & Cell Biology, The University of Chicago, 920 East 58 St., Chicago IL.

The current status of the sequencing project is as follows: about one half of the genome has been sequenced and annotated completely; about one-fourth is being annotated; the remaining fourth is more than 90% sequenced. At present there are five gaps in the alignment, all of which are being closed by using lambda libraries or PCR on chromosomal DNA. These results are reported in detail on our web site, which is updated regularly. Here are the basic site addresses:

http://rhodol.uchicago.edu/rhodo/map/status.html
http://rhodol.uchicago.edu/rhodo/server/nph-arrows.cgi
http://rhodol.uchicago.edu/rhodo/server/nph-database.cgi
http://rhodol.uchicago.edu/rhodo/P/final/fastahom/map
http://capsulapedia.uchicago.edu

The data are so extensive that they can only be appreciated by opening the web sites and examining the sequence with the tools available on the sites. The major unexpected results so far are the discovery of a number of cryptic phages in the genome, duplications of operons of unknown value (such as the transporter of polyamines), and the combination of both the aerobic and anaerobic pathways for vitamin B12 synthesis in the same region of the chromosome. The fraction of unidentifed orfs is about the same as for other bacterial genomes.

From the beginning of the project, we have been interested in strain comparisons for the glimpse it should provide of bacterial chromosome evolution and genome-changing events. A study of the physical maps for three strains, the type strain SB1003 and two variants, St. Louis (SL) and 2.3.1, were completed first. This study showed that for a 2-Mb part of the chromosome, restriction sites were either highly conserved or, in adjacent chromosomal regions, highly polymorphic. Sequencing was undertaken to expand this mosaic model of the chromosome. About 200 kb from "conserved" regions of the SL and 2.3.1 strains were compared with the corresponding region of SB1003. The major result is that by comparing strains it is relatively easy to determine where orfs start and stop, because the DNA sequence is highly conserved (>90% identity) in coding regions and much less conserved (~75%) in non-coding, intergenic regions.

SA-04. Whole Genome Sequence of *Pyrobaculum aerophilum*.

S. T. FITZ-GIBBON, U. J. KIM*, E. CONZEVOY, M. YAMAHA, G. PARK*, K. S. STETTER+, M. I. SIMON*, J. H. MILLER.
Univ. California Los Angeles. *California Inst. Technology, Pasadena, CA. +Univ. Regensburg, Germany.

Pyrobaculum aerophilum is a hyperthermophilic archaeon, isolated from a boiling marine water hole, that is capable of growth at 104'C. This microorganism can grow microaerobically, unlike most of it's thermophilic relatives, making it amenable to a variety of experimental manipulations and a candidate as a model organism for studying archaeal and thermophilic microbiology. We have sequenced the entire genome using a random shotgun approach (3.5X genomic coverage) followed by oligonucleotide primer directed sequencing, guided by our fosmid map. The 2.2 Mb genome codes for more than 2000 proteins, 30% of which have been identified by their sequence similarities to proteins of known function. Only 15% of the *Pyrobaculum aerophilum* proteins have related high resolution structures. In collaboration with the DOE/UCLA Laboratories and Los Alamos National Laboratories we have initiated a project to express, and purify proteins for structure determination by NMR or xray diffraction. The three dimensional structures of the *Pyrobaculum aerophilum* proteins will give one the power to understand and manipulate protein function and are crucial to fully exploiting the information in the genome. At this time several proteins have been cloned, expressed in *E.coli*, purified and crystals which diffract to high resolution have been obtained.

SA-05. T-even Phages: Genomics, Evolution and Therapeutic Applications.

E M KUTTER, E THOMAS, J. CUSHING, M MUELLER, P LIPPINCOTT, B ANDERSON, AND J NEITZEL; F ZUCKER AND T HUNKAPILLAR.
The Evergreen State Coll., Olympia, WA; Univ. of WA, Seattle, WA.

We still have no specific functional clues for almost half of T4's 270 probable protein-encoding genes. Only about 40 enzymes have clear database matches, so we are also using other approaches to look at functionality. For example, 13 T4 proteins are predicted to be integral membrane proteins (cf. Boyd et al 1998, Protein Science 7:201-205). These include imm, ac and t, as expected, and several interesting clusters of small ORFs. However, surprisingly, rII A and B, ndd and gp46 probably are not "integral membrane proteins". Most of the clear homologies involve enzymes of nucleic acid metabolism. There are still virtually no indications as to the origins of the large number of host-lethal proteins responsible for the transition from host to phage metabolism, or of the phage structural proteins. The latter have been shown by Krisch et al. to be shared among phages of T4 morphotype that infect a range of gram-negative bacteria and are quite different in other parts of their genomes. With the exception of exchange among T-even-type phages and tail-fiber gene swapping, there is no evidence for acquisition of any genes since well before the split between Hemophilis and E. coli. In some cases, the divergence predates the separation of eubacteria and eucaryotes, and some are more like genes of eucaryotes or their viruses. We are working on a system to integrate and manipulate results from numerous runs of such programs as BLAST, FASTA and Smith-Waterman using various parameters, queries and databases and applying it to try to tease out meaning from weak T4 homologies. We are also trying to better

understand phage-host interactions in nature, and finding surprising complexities. For example, in contrast to the total shutoff of host replication, transcription and translation seen when T4 infects exponential-phase bacteria, we find that infection of nutrient-starved cells leads to a sort of state of hibernation. When nutrients are then added, both host and phage proteins are made, with the host syntheses gradually shutting off as the cell prepares a small burst of phage. We have also tested nearly 100 phages of the T-even morphotype against a range of hosts: the ECOR collection, pig pathogens, O157, Shigella. Anywhere from 1 to 43 phages infect each, with surprisingly complex patterns of infectivity; the mean is about 20 phages/strain. The implications of these various findings for medical antibacterial applications will be discussed. T-even-like phages also infect bacteria such as Pseudomonas and Vibrio, and form important components of the therapeutic cocktails used for decades for phage therapy in Eastern Europe, particularly through the Eliava Bacteriophage Institute, Tbilisi, Georgia, our collaborators in these explorations. For further information, see www.evergreen.edu/user/T4/PhageTherapy/phagethea.html.

Session II - Regulatory Networks, Proteomes, and Annotation

SA-06. Whole Genome Approaches to Regulatory Networks and Phenotypes.
George M. Church.
Harvard Medical School, Boston, MA.

The success of whole genome sequencing provides both the information and the inclination to collect data comprehensively for whole networks accurately and automatically. Models and comparisons provide checks and extensions of both positive and negative results. We are developing methods for quantitating RNAs, proteins and metabolites from wild-type and growth rates of in-frame mutants under a variety of environmental conditions. The design of oligonucleotide chips, electrospray ionization mass spectrometry, homologous recombination engineering, and computational resources for modeling and motif analysis make this an exciting community effort. Relevant publications, databases, and protocols are available at: http://arep.med.harvard.edu.

SA-07. Automated Genome Annotation Through Whole Genome Comparisons.
Terry Gaasterland.
Argonne National Laboratory and Univ. of Chicago.

Genome interpretation is an on-going iterative process in which each successive pass incorporates previously gathered data into a new decision process. In genome annotation, every potential coding region in a genome must be compared with each protein sequence in public curated databases, including all other fully sequenced genomes. Similarity at the sequence level translates into putative function assignments. To reinforce sequence alignment information, DNA patterns, e.g. promoter and terminator sites, can be deduced and associated with coding regions. However, no functional assignment is sure until it has been confirmed through biological experimentation.

A system that carries out automated genome analysis must be capable of reasoning about the genomic data in the context of this uncertainty. An important part of such a reasoning process is to reinforce putative and even suspected assignments based on subsequent deductions. Just as important is the visual presentation to users of evidence about decisions made by the systems. The MAGPIE system (joint work with Christoph Sensen, IMB-NRC, Halifax, Canada) has been designed to meet these requirements.

To compare genomes, every coding region in a genome is aligned with every coding region in every other fully sequenced genome. We have devised a system to parse the alignment data into genomic and phylogenetic signatures for every coding region in every genome. Collectively, those signatures provide a phylogenetic overview for an entire organism. If we consider the phylogenetic and genomic signatures for a functionally defined subset of coding regions from multiple genomes (e.g. all gene products involved in energy metabolism or all gene products categorized as global regulatory proteins), we can deduce allowable losses, gains, and alterations of function. As with annotation, visualization of comparative genomic data helps users to gain insights and intuition about the genomes.

We have used the MAGPIE system as the data collection engine to gather cross-genome analysis data for 23,971 open reading frames (ORFs) in 10 genomes (Aquifex aeolicus, E.coli, H.infl.,M.genit.,M.pneu., Synechocystis sp, M.janna., M.thermo., S.solfataricus, and S.cerev.). The amino acid sequences from each coding region in each genome were compared via BLAST, FASTA, CLUSTALW, and PHYLIP coordinated via MAGPIE. We used a new suite of programs (joint work with Mark Ragan, IMB-NRC, Halifax, Canada) to generate genomic signatures and derive cross-genome analyses from the collected data.

We use the genomic signatures to further define the following concepts: genomically universal proteins (proteins that have a counterpart in every fully sequenced genome); proteins characteristic of phylogenetic subsets,

including proteins characteristic of bacteria (proteins that have a counterpart in every fully sequenced bacterial genome and NO detectable counterpart in any other genome), of archaea, of prokaryotes, of both bacteria and eukaryote, of both archae and eukaryote.

This study lays a foundation for systematic comparison of multiple whole genomes. It also demonstrates how to include partially sequenced genomes in ``cross-genome" profiles. With 10 microbial genomes, our system has confirmed and qualified common observations from the literature. It has also led to new insights into genomic evolution at a protein level. Future work will include the next cohort of fully sequenced genomes, which include pathogens, non-archaeal extremophiles, and a putatively ancient bacteria. It will also include the available predicted protein coding regions from C. elegans and human.

SA-08. Use of the Comparison of Families of Paralogous Proteins to Trace Back the Evolution of Proteins and of Procaryotic Genomes.

B. LABEDAN*, C. POULOU and R. DE ROSA.
Institut de Génétique et Microbiologie, Université Paris-Sud, 91405 Orsay Cedex, France.

This work derives directly from studies made in collaboration with Monica Riley (Labedan & Riley, 1995ab - Riley & Labedan, 1997 - Riley & Labedan, this meeting) in which we used the whole set of E. coli proteins to study the mode of evolution of a bacterial chromosome and to get information about the processes of protein evolution. Our previous results have shown that a large proportion of E. coli proteins are paralogues, many of them corresponding to one long segment of homology - that we called a module - and a significant minority corresponding to the fusion of at least two evolutionarily unrelated modules. These paralogous modules may be further grouped in families, each family descending from one ancestral module. We have now extended this approach to several (seven bacteria and three archaea) completely sequenced procaryotic genomes. The comparison of these families of paralogous modules present inside each of these organisms helped us to define two kinds of families : families which are present in all studied organisms which we call homologous families, and families which are unique to some organisms that we call signature families. Both kinds of families allows one to check the validity of the branch topology of the tree of life : signature families are an obvious tool but examination of genealogical trees of homologous families is also another indirect means to examine branch topology. Indeed, the timing of the events of paralogy (gene duplication) and of speciation must be consistent from one tree to another using the same set of organisms. Moreover, characterizing the homologous families is crucial in tracing back from present-day proteins to very ancestral modules. Such an approach will progressively disclose a large part of the genetic panoply of the last universal common ancestor. Our first results show already a large redundancy in this genetic panoply, strongly suggesting that complex events of paralogy of primitive modules followed by selective losses and fusion events of unrelated modules occurred well before the emergence of this last universal common ancestor.

Labedan, B. & Riley, M. (1995a and b). J. Bacteriol. 177, 1585-1588. - Mol. Biol. Evol. 12, 980-987. - Riley, M. & Labedan, B. (1997). J. Mol. Biol. 268, 857-868

SA-09. A Genome-based Approach to Identify Novel Essential Bacterial Genes.

F. Arigoni, F. Talabot, M. Peitsch, M. D. Edgerton, E. Meldrum, E. Allet, R. Fish, T. Jamotte, M.-L. Curchod, and H. Loferer.
Geneva Biomedical Research Institute, Glaxo Wellcome Research and Development S.A., 14, chemin des Aulx, CH-1228 Plan-les-Ouates, Geneva, Switzerland.

We have used comparative genomics to identify a list of 26 E. coli open reading frames which are both of unknown function (hypothetical open reading frames, "y-genes") and are also conserved in the compact genome of M.genitalium. Not surprisingly, these genes are broadly conserved in the bacterial world. Reasoning that proteins involved in essential cellular functions would be included among such evolutionary conserved genes, we used a marker-less knock-out strategy to screen for E. coli genes essential for growth on LB. To verify this phenotype, we constructed conditional mutants in genes for which no null mutants could be obtained. In total we identified 6 genes which are essential for E. coli. The respective orthologs were also shown to be essential in B. subtilis. This low number of essential genes was rather unexpected and might be due to a phenomenon in the versatile genomes of E. coli and B.subtilis which is comparable to the "non-orthologous gene displacement". The gene ygjD is of considerable interest. It was eliminated from a "minimal genome" derived from the in-silico comparison of the H.influenzae and M.genitalium genomes as being specific for the interaction with the host based on its annotation as a secreted sialoglycoprotease. We show here that ygjD and its ortholog ydiE are essential in E. coli and B. subtilis, respectively. Thus, we propose to include this gene into a "minimal genome". This study systematically integrates comparative genomics and targeted gene disruptions to identify broadly conserved bacterial genes of unknown function required for survival on complex media. The novel essential genes identified are attractive targets for future broad-spectrum antibiotics.

Tuesday, September 22
Session III - Chromosome Structure, Inteins, and Integrons

SA-10. Super-integrons in *Vibrio* Genomes.
D. MAZEL, A. DYCHINCO*, N. KAIM, V.A. WEBB* and J.E. DAVIES*.
Institut Pasteur, Paris, France and * Univ. of British Columbia, Vancouver, B.C., Canada.

Integrons are naturally-occurring gene expression structures relying on a site-specific integration mechanism for the tandem insertions of open reading frames (gene cassettes). Since their characterization in the 1980's these genetic elements have been widely implicated in the spread of antibiotic resistance genes and in the evolution of multidrug-resistance among the Gram negative rods that has occured in the last fifty years. We have observed a striking homology between the recombination element of the integron cassette encoding the carbenicillinase CARB-4 and a family of repeated sequences found in *Vibrio cholerae*, the VCR. Analysis of *V. cholerae* published sequences harbouring VCR copies showed that the "gene - VCR" organization was similar to the resistance gene cassette array of contemporary integrons. We have demonstrated that they were definitely integron cassettes. We have characterized a gene (*intI4*), associated to the VCR clusters, which encodes an integrase showing a specific recombinase activity on the VCR cassettes. The overall structure of the locus harbouring *intI4* and the VCR cassette clusters is identical to the antibiotic resistance integrons, but is ten times larger and could be defined as a super-integron. Moreover, VCR cassettes are found in a number of *vibrio* sp. including a strain of *V. metchnikovii* isolated in 1888, demonstrating that this mechanism of heterologous gene acquisition predated the antibiotic era. Currently, we have undertaken the characterization of super-integrons harboured by the *V. mimicus* and *V. metchnikovii* genomes in order to establish their evolutionary relationship. Several of the VCR cassette genes identified in *V. cholerae* encode proteins whose functions are associated with pathogenicity (heat-stable toxin, haemagglutinin); we hypothesize that the VCR's and their associated genes could have been cassettes for pre-antibiotic resistance integrons to provide a generalized system for the entrapment and spread of adaptive functions.

SA-11. Chromosomal Domain Structure in *Salmonella typhimurium*.
N. Patrick Higgins*, Dipankar Manna*, Pawel Staczek*#.
*Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35294-2170. # Department of Genetics of Microorganisms, University of Lodz, Banacha 12/16, 90-237 Lodz, Poland.

The gamma delta resolution system was used to study the in vivo domain structure and supercoil mobility in S. typhimurium. Three surprising results are: 1). Plectonemic duplex DNA-DNA synapsis occurs over intervals of at least 100 kb, and in some circumstances, over intervals of more than 500 kb (1). 2) The most common supercoil barrier is random, occurring with a 30% probability over each 20 kb interval in cells growing exponentially. 3). Domain barriers disappear when cells enter stationary phase, and the domain size is modulated by the activities of two essential enzymes -- DNA gyrase and Topoisomerase IV (2). The efficiency of DNA supercoil movement in vivo fits measurements in vitro. Oram et al. (3) found all sites in a 5 kb plasmid become intertwined with all other sites at a frequency of once per second. Domain barriers may modulate the stability and duration of protein-protein interactions involving two different segments (DNA loops). To see if other protein-protein interactions obey domain rules, we analyzed transposition immunity, a phenomenon in which the presence of one transposon limits the probability of a second identical sequence integrating into sequences nearby. In phage Mu, transposition immunity involves interactions between the Mu A transposase, which binds specifically to Mu ends, and the B protein, which is a non-specific DNA binding protein that chooses insertion targets. Mu transposition immunity fit the domain rules. A Mu A binding site blocked secondary Mu insertions and this inhibition decayed gradually, disappearing at 20 kb, which is the average distance to the first supercoil barrier. This unexpected mechanism, controlled by topoisomerases, may modulate many critical responses like transcription initiation with activator proteins and special types of DNA repair, replication, and recombination.

1. N. P. Higgins, X. Yang, Q. Fu, J. R. Roth, J. Bacteriol. 178, 2825-2835 (1996).
2. P. Staczek, N. P. Higgins, Mol. Micro. (In Press) (1998).
3. M. Oram, J. F. Marko, S. E. Halford, J. Mol. Biol. 270, 396-412 (1997).

SA-12. Archaeal Histones and Nucleosomes.
J.N. REEVE, K.A. BAILEY, W-T. LI, S.L. PEREIRA, K. SANDMAN and D.J. SOARES.
Dept. Microbiology, Ohio State Univ., Columbus, OH.

Studies designed to determine the mechanisms of genome compaction and stabilization in *Methanothermus fervidus*, a hyperthermophilic *Archaeon*, led to the

discovery of HMfA and HMfB, two small polypeptides with sequences in common with the histone fold regions of the eukaryal nucleosome histones H2A, H2B, H3 and H4 (1). NMR analyses of recombinant (r) versions of HMfA, HMfB and of a closely-related protein rHFoB from a mesophilic methanogen confirmed that these proteins are histones (2,3). They form both homodimers and heterodimers in solution, and dimerize through canonical histone folds. Results will be reported that document that HMfA and HMfB assemble into structures, designated archaeal nucleosomes (4), that protect ~60 bp of DNA from micrococcal nuclease digestion, contain an archaeal histone tetramer, and exhibit positioned assembly consistent with direct participation in gene expression.

All the euryarchaeotal genomes sequenced to date contain archaeal histone encoding genes, and now with over 20 very similar, but nevertheless different archaeal histone sequences available, roles for both conserved and non-conserved regions have been predicted and evaluated by comparative studies and site-specific mutagenesis. Residues predicted to be responsible for histone dimer and tetramer formation, for DNA binding, and for fold stabilization at high temperatures have been changed and the results of structure-function-stability assays of the resulting variants will be presented. They are consistent with archaeal histones binding to DNA and forming structures very similar to that formed by the histone (H3+H4)-(H3+H4) tetramer at the center of the eukaryal nucleosome (5). Residues and structures responsible for conferring resistance to temperature-induced unfolding will also be discussed.

Recently, histone folds have been identified as domains within several different eukaryal transcription factors (6), and MJ1647 in the *Methanococcus jannaschii* genome appears to encode an archaeal histone fold with a C-terminal extension (7). Studies with rMJ1647 purified from E. coli have revealed DNA binding properties consistent with an archaeal histone and demonstrated that removal of the C-terminal extension reduces resistance to thermal unfolding.

1. Reeve, J.N., et al. (1997) Cell, 89:999.
2. Starich, M.R.,et al.(1996) J.Mol.Biol.,255:187.
3. Zhu, W., et al. (1998) Biochem., 37:10573.
4. Pereira, S.L., et al. (1997) PNAS., 94:12633.
5. Luger, K., et al., (1997) Nature, 389:251.
6. Struhl, K., et al., (1998) Cell, 94:1.
7. Bult, C.J., et al., (1996) Science, 273:1058.

SA-13. Protein Splicing.

M W SOUTHWORTH, F B PERLER.
New England Biolabs, Beverly, MA.

More than 70 putative inteins (protein splicing elements) have been identified in eubacteria, archaea and eucarya. Genome sequencing projects have identified more than half of these inteins, with 19 found in *Methanococcus jannaschii*, 14 in *Pyrococcus horikoshii* and at least 9 in *Pyrococcus furiosus*. However, many small genomes such as *Archaeoglobus fulgidus*, *Helicobacter pylori*, *Haemophilus influenzae*, *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Borrelia burgdorferi* and *Escherichia coli* do not contain inteins, whilst *Methanobacterium thermoautotrophicum*, *Aquifex aeolicus*, *Bacillus subtilis* and *Saccharomyces cerevisiae* genomes have only 1 intein. There are 3 inteins in *Mycobacterium tuberculosis* and 4 in *Synechocystis sp.* The high number of inteins found in some species may be indicative of a higher rate of genetic exchange, possibly an advantage to organisms living in extreme environments. Inteins range in size from 134 to 548 amino acids and are found in a wide range of proteins involved in DNA replication, recombination, transcription, translation, metabolism, proteolysis and ATPases. The intein plus the first downstream amino acid (always Ser, Thr or Cys) are sufficient for protein splicing in a foreign context. Inteins have a conserved Ser or Cys at their N-terminus and a conserved Asn or Gln at their C-terminus. The autocatalytic splicing pathway consists of 4 steps. An N-O (N-S) acyl shift at the intein N-terminus is followed by attack by a downstream Ser, Thr or Cys thiol/hydroxyl group resulting in the formation of a branched intermediate. The intermediate is resolved by Asn (Gln) cyclization followed by a spontaneous O-N (S-N) acyl shift resulting in intein excision and formation of a peptide bond between the exteins. Many inteins contain homing endonuclease sequences similar to those observed in self-splicing introns. The bifunctional intein has the ability to efficiently move its coding sequence from an intein containing protein to an inteinless one by gene conversion initiated by double strand break repair. Endonuclease activity is independent of splicing activity. The splicing and homing endonuclease functions are present in two distinct structural domains, although there is a DNA recognition region attached to the Sce VMA intein splicing domain. Inteins have been modified to aid in the purification of proteins, ligate peptides, control enzyme activity and partially label a protein, providing a robust workhorse for protein engineering.

SA-14. Chemotaxis Operons in Microbial Genomes: Comparison to What is Known.

S.A. BULLOCH and I.B.ZHULIN.
Dept. Microbiol. & Mol. Genet., Sch. Medicine, Loma Linda Univ., Loma Linda, CA.

We have analyzed completely sequenced genomes of bacteria and archaea that were available through public

databases for the presence of chemotaxis operons. Chemotaxis operons appeared to be present in *H. pylori* (Proteobacteria), *B. subtilis* (Low G+C Gram-positive bacteria), *B. burgdorferi* (Spirochaeta), and *A. fulgidis* (Archaea). A putative chemotaxis operon was found in *Synechocystis* sp. (Cyanobacteria). No chemotaxis operons or any chemotaxis genes were found in two *Mycoplasma* species, *H. influenzae* (all Bacteria), *M. jannaschii, M. thermoautotrophicum* (both Archaea), *S. cerevisiae* (Eucarya). Functional organization of the identified putative operons was compared to that of *E. coli, S. typhimurium, P. putida, M. xanthus, R. meliloti, R. sphaeroides* and *R. centenum* (all Proteobacteria), *T. pallidum* (Spirochaeta) and *H. salinarum* (Archaea). In two species, *R. sphaeroides* and *B. burgdorferi*, two chemotaxis operons have been identified, whereas other species appear to have only one chemotaxis operon. Chemotaxis operons consist of three to ten genes. The *cheA, cheW* and *cheY* analogues of known *E. coli* genes are present in all identified chemotaxis operons. Proteins that are encoded by these genes constitute a phosphorylation cascade, which is the central pathway for signal transduction in bacterial chemotaxis. Other chemotaxis genes in the operons encode for CheR methyltransferase and CheB methylesterase, transducing proteins, and proteins that have no analogues in *E. coli*. Phylogenetic analysis of the chemotaxis genes revealed that they are conserved in Bacteria and Archaea, indicating an ancient origin of bacterial behavior.

Session IV - Completed Genomes and Genomics

SA-15. The *Thermotoga Maritima* Msb8 Genome.

Karen E. Nelson*, R.Clayron, R.D. Fleischmann, J.A. Malek, K.D.Linher, M.M. Garrett, A.R. Stewart, L. McDonald, T. Utterback, G. Sutton, C.M. Fraser, H.O. Smith and J.C. Venter.
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD.

The number of microbial genome sequencing projects at TIGR continued to increase this year with the initiation of sequencing on nine new genomes. Among these are the pathogens*Staphylococcus areus, Mycobacterium avium* and*Chlamydia trachomatis*.. The whole genome sequencing project of*Thermotoga maritima* is the latest to be completed.

For the *Thermotoga maritima* genome project, random sequences were generated from a small insert library to ~ 7 X coverage. The sequences when assembled resulted in 50 groups of assemblies separated by physical gaps closed primarily by a combination of direct sequencing of genomic DNA and sequencing PCR products amplified from the respective gaps. Approximately 30, 140 good sequences were used to generate the final assembly of 1.86 Mbp.

The genomic sequence of *Thermotoga maritima*, the most extreme thermophilic organotrophic bacterium known and one of the earliest branching bacteria, will provide greater insight into natural mechanisms of biopolymer degradation, and allow us to address existing questions on phylogeny and evolution. The organism also stands to provide genes whose thermostable products could be useful for industrial processes. The findings from an initial analysis of the completed genome sequence will be presented.

SA-16. Biological Systems in *Treponema pallidum*: A Black Box Opened by Genomics.

G. M. WEINSTOCK, S. J. NORRIS, M. MCLEOD, G. MYERS, T. BRETTIN, E. SODERGREN, J. HARDHAM.
Univ. Texas Medical School, Houston and Los Alamos Natl Laboratory.

The spirochete *Treponema pallidum* is the causative agent of syphilis. Because this bacterium cannot be continuously cultured in the laboratory, knowledge of its biochemical and genetic systems has been sparse.

Recently we determined the complete genomic sequence of *T. pallidum* (Science 281:375-88, 1998) in collaboration with TIGR. The genome is 1,113,011 bp and contains 1041 predicted protein coding sequences. Approximately 55% of these putative coding sequences were assigned tentative functions based on database matches. About one-third of the remaining putative coding sequences matched hypothetical proteins from other organisms while the remainder were unique to *T. pallidum*.

Although functions could only be assigned to slightly more than half of the predicted coding regions, this results in a huge increase in our knowledge of this bacterium. It provides a clearer view of the basic housekeeping functions that are present (or notably absent) in this organism, such as central dogma pathways, protein export, and metabolism. This is significant since the spirochetes have not been studied extensively yet are a more ancient class of bacteria than many better-studied organisms. As an example, *T. pallidum* lacks many biosynthetic pathways, similar to other bacteria who have minimal genomes. Yet although it is missing virtually all de novo amino acid synthesis enzymes, it contains a complete pathway for proline biosynthesis. This may reflect a use for proline as an osmoprotectant. Another example is the presence of

enzymes similar to *E. coli's* recF pathway of recombination and repair, but an absence of the enzymes from the recBCD pathway. This is observed in a number of other sequenced prokaryotes. It is interesting to note that in the only other spirochete to be sequenced, Borrelia burgdorferi, the recBCD pathway is present but not recF enzymes. It can be argued from these observations that the recF pathway is an ancient mechanism while the recBCD pathway may be more specialized.

In addition to such observations of housekeeping functions, a roster of possible virulence factors and other interesting open reading frames can be constructed. These include putative virulence functions widely distributed among other organisms, others which are spirochete-specific, and other candidates that resemble host functions. These various biological systems, the personality of *T. pallidum*, will be summarized.

SA-17. Metabolic Reconstruction of Zymomonas mobilis and Thiobacillus ferrooxidans Based on the Gapped Genome Sequencing.

M. Fonstein, L. Chu, R. Haselkorn, Y. Kogan, Y. Nikolsky, R. Overbeek, E. Selkov, V. Vonstein. Integrated Genomics Inc, 2201 W. Campbell Park Drive, Chicago, IL 60612, University of Chicago, 920 E. 58th St.., Chicago, IL 60637 MCS, Argonne Natl Lab, Argonne IL 60439.

Bacterial genomics is characterized by the following trends: 1. The majority of sequenced microorganisms are human pathogens or extremophiles . Most of the industrially and environmentally important microorganisms are not included in the genomic race. 2. Complete polished DNA genome sequence without gaps is a standard in academic community. However, growth of the databases and development of analytical tools allow to extract comparable metabolic information from the much cheaper "gapped" genomes. 3. Sequence annotation is mainly limited to the assignment of functions to genes, whereas integration of these assignments into a system of metabolic reconstructions permits their substantial refinement and leads to the modeling of the biochemical processes of the living cell. We report the results of sequencing and metabolic reconstruction of Zymomonas mobilis, a bacterium known for it's efficient ethanol production. Plasmid and cosmid libraries of Z. mobilis genomic DNA were constructed and sequenced using automated ABI 377 machines. A total of 15,000 runs with an average cut-off of 600 bases have been assembled into ~1000 contigs (0.6 Kb - 40 Kb range). The total length of the assembled DNA is about 1.93 Mb which corresponds to >90% of the 2.1 Mb genome. A summary of the metabolic can be summarized as:

Asserted Pathways	285
Assigned Functions	1875
Connected Assignments	632
Missing Assignments	135
Missing Sequences	52.

We present the metabolic model of *Z. mobilis* aminoacid biosynthesis based entirely of its genome sequencing. A similar project studying Thiobacillus ferrooxidans is currently at the stage of contig assembly. Results of the data analysis should be ready by the time of the conference.

SA-18. Lateral Gene Transfer and Genome Divergence Inferred from the Genomic Sequences of the Closely Related Hyperthermophilic Archaea, Pyrococcus furiosus , Thermococcus litoralis and P. horikoshii.

F. T. ROBB*, R. B.WEISS', D. DUNN', Y. KAWARABAYASI^, J. DIRUGGIERO*, D. L. MAEDER*, J. CHATARD*, M. STUMP', J. CHERRY'. *Ctr. of Marine Biotechnoloiv. of Maryland. 'Department of Genetics, Univ. of Utah. ^ Biotechnology Center, Natl. Inst. of Technology and Evaluation, Nishihara 2-49-10, Shibuya-ku, Japan.

Archaea of the genus Pyrococcus consist of several species that grow rapidly at temperatures exceeding 100°C, and their genomic sequences provide a significant new resource for comparative studies of hyperstable enzymes. We report here the closure of the 1.92 mbp circular genomic sequence from the hyperthermophilic archaeon, Pyrococcus furiosus. The comparison of this sequence to the closely related sequence of P. horikoshii, which has an average of 81% DNA homology to P. furiosus provides insights concerning the rearrangements and insertion/deletion events accompanying divergence of this group of Archaea. For example, the pre-insertion sites of three inteins, as well as divergence of intein coding regions can be detected. From this analysis, it is clear that the divergence of the genus Pyrococcus which is represented by P. furiosus, isolated from the shoreline in Italy, and P. horikoshii, isolated from the Okinawa Trough, in the north Pacific Ocean at a depth of 1390 m, has been punctuated by insertion or deletion events, and extensive genomic rearrangement. The 37 bp tandem, repetitive sequence elements which are conserved in both of the Pyrococcus genomes, have diverged both in copy number and in the position and number of clusters: the P. furiosus genome has five clusters whereas P. horikoshii has two clusters. A 17.8 kbp region in the P. furiosus genome contains a bacteria-like operon encoding maltose/trehalose uptake and catabolism which is absent in P. horikoshii, which is unable to utilize these sugars. The region, which is flanked by putative transposons containing transposase homologs, apparently inserted

into the P. furiosus genome by a lateral transfer event. Mapping and sequence of the homologous region in the P. furiosus genome and in T.litoralis, which was isolated from the same site in Italy as P. furiosus, indicates that the sequence is 100% homologous in both hyperthermophiles and does not resemble any sequences in the genome of P. horikoshii. We propose that this lateral transfer event occurred relatively recently, presumably after the divergence of P. furiosus and P. horikoshii.

Wednesday, September 23
Session V - Recombination and Repair

SA-19. DNA Repair in *Deinococcus radiodurans*: Adapting to Life in Dry Environments.
JR BATTISTA.
Louisiana State Univ. and A&M Coll.

Available evidence indicates that the radioresistance of *Deinococcus radiodurans* is a direct result of its ability to efficiently repair the DNA damage generated during irradiation. In other words, the extreme ionizing radiation resistance of this species -- and presumably the other deinococci -- appears to be the result of an evolutionary process that selected for organisms that could tolerate massive DNA damage. The reasons for *D. radiodurans'* ionizing radiation resistance are obscure, however. It cannot be an adaptation (i.e., an evolutionary modification of a character under selection) to ionizing radiation, because there is no selective advantage to being ionizing radiation resistant in the natural world. There are no terrestrial environments that generate a high flux of ionizing radiation. It must, therefore, be assumed that deinococci's radioresistance is a chance use of the cell's DNA repair capability. Forty-one ionizing radiation sensitive strains of *D. radiodurans* were evaluated for their ability to survive six weeks desiccation. All exhibited a substantial loss of viability upon rehydration when compared with wild type *D. radiodurans*. Examination of chromosomal DNA from desiccated cultures revealed a time dependent increase in DNA damage as measured by an increase in DNA double strand breaks. This evidence suggests that *D. radiodurans'* ionizing radiation resistance is incidental, a consequence of this organism's adaptation to a common physiological stress, dehydration.

SA-20.

DNA Repair Genes in Archaea and Bacteria.
A. J. CLARK*. LBNL, Berkeley, CA. J. A. EISEN*. Stanford University, Stanford, CA, and S. J. SANDLER, U. Mass., Amherst, MA

RadA and RadB proteins in Archaea are similar in amino acid sequence to RecA proteins of Bacteria. Phylogenetic analysis of amino acid sequences encoded by PCR fragments of the radA genes from 11 cultivated archaeal species and a monoculture archaeal species showed a cladal distribution which mimicked to a large extent that obtained by a similar analysis of archaeal 16S rRNA sequences from the same species. An analysis of PCR fragments of 15 radA genes from uncultivated natural sources reveals several clades with affinity to the putative RadA proteins of several species of Crenarcheota.

The sequences of all proteins with well established roles in DNA repair have been compared to all the complete genomes as well as to all sequences in Genbank using a variety of database searching algorithms. Proteins and open reading frames with significant similarity to any known repair protein were compiled. The presence and absence of homologs of known repair genes across the main domains of life, and a molecular phylogenetic analysis of the proteins have been used to construct an evolutionary history of DNA repair proteins and DNA repair processes Specific repair genes likely to have been present in the ancestor of all living organisms can be identified. In addition, the presence and absence of specific repair genes from certain species can be used to predict the likely repair capabilities of these species.

SA-21. DNA Repair Mechanisms In Hyperthermophilic Archaea.
*DIRUGGIERO J., *I. MICHOUX, *F. CHAUSSADE, *F.T. ROBB AND #R. WEISS.
*Ctr. of Marine Biotechnology, Univ. of Maryland Biotechnology Institute, Baltimore MD and #University of Utah, Salt Lake City, UT.

Very few experimental studies have addressed repair systems in Archaea, and the question as to how the hyperthermophiles maintain the integrity of their genetic material at the structural and sequence information level remains largely unanswered. In recent studies we showed that the 2 Mb chromosome of Pyrococcus furiosus was fragmented into pieces from 500 to 30 kb after irradiation at a dose of 2,500 Gy (75% survival). The chromosome was subsequently fully reassembled upon incubation of the irradiated cultures at 95°C for several hours, demonstrating an unusual capacity for double-strand break repair. Homologs of eukaryotic and bacterial repair genes have been identified from the complete genome sequences of several hyperthermophilic and extreme thermophilic Archaea and include genes involved in: base excision repair, nucleotide excision repair, SOS mutagenesis, mutator

phenotype, SOS repair, mismatch repair and recombination. The annotation of most of these genes is based on amino acid sequence homologies with bacterial and/or eukaryal repair proteins present in databases. In most cases the archaeal proteins are more closely related to the eukaryal than to the bacterial homologs, with exceptions such as DinF and DinG, whose functions remain unknown in E. coli, or the bacterial type of Uvr proteins involved in excision repair. Two archaeal homologs of recA/RAD51-like genes have been described in several archaeal species including P. furiosus. Phylogenetic analysis of the archaeal recombinase protein sequences showed that both recombinases are more closely related to the eukaryotic Rad51/DMC1 proteins than to the bacterial RecA proteins. RadA proteins are about 100 amino acids longer than RadB and in P. furiosus the identity between the two Rad proteins is about 35%. Of the RecA-like sequences known so far, only the archaeal RadB proteins consist solely of the Domain II of E. coli RecA. This domain, which contains specific ATP binding motifs, may be the only sequence element common to all recombinases. Domain III contributes to protein-protein interactions and is absent in all RadB homologs. We have overexpressed and purified both the RadA and RadB proteins from P. furiosus and a preliminary characterization of these recombinants proteins is in progress. In addition, using nuclease protection assay we have investigated the regulation of the radA and radB gene expression in P. furiosus cells grown under different conditions and following exposure to DNA damaging agents.

SA-22. Mutators Among *Escherichia coli* and *Salmonella enterica*: Adaptation and Emergence of Bacterial Pathogens.

T.A. CEBULA, B. LI, W.L. PAYNE and J.E. LE CLERC.
U.S. Food and Drug Administration, 200 C Street, SW, Washington, D.C. 20204 USA.

Genetic change (mutation) and exchange (recombination) provide the genetic diversity upon which selection works to establish a specific microbe in its particular niche. How successful a microbe is at surviving the diverse challenges of an ever-changing environment ultimately rests upon the relative diversity within the microbial population at large. Under adverse conditions, a high mutation rate might be anticipated since it would increase the chances of spawning the rare mutant needed to survive--be it to escape immune surveillance, to elude therapeutic intervention, or to evade the manifold barriers meant to keep microbial populations in check. In this vein, we discuss here the importance of particular mutators, those defective in methyl-directed mismatch repair (MMR), that we found relatively frequently (> 1%) among natural isolates of *Escherichia coli* and *Salmonella enterica*. We weigh the role of MMR mutators in microbial evolution, for such mutators are not only hypermutable (*i.e.*, they exhibit high mutation rates) but promiscuous (*i.e.*, they show an increase in recombination) as well. Moreover, unselected populations of *E. coli* (Mao *et al.*, J Bacteriol **179**:417-422, 1997) and *S. typhimurium* (LeClerc *et al.*, Mutat Res **400**:89-97, 1998) sport MMR mutators at a frequency of 1-10 x 10^{-6}. We question whether these MMR subpopulations may be the "mixing pools" where horizontal transfer of genes from similar or disparate species takes place.

SA-23. Studies of Proteins Regulating DNA Topology in Hypethermophile Methanopyrus kandleri.

A. Slesarev, G. Belova, S. Kozyavkin, D. Musgrave, R. Prasad, S. Wilson, J. Lake.
A. Slesarev and G. Belova, Inst. of Bioorganic Chemistry, Moscow, Russia; S. Kozyavkin, Fidelity Systems, Inc., Gaithersburg, MD; R. Prasad and S. Wilson, NIEHS, NIH, Research Triangle Park, NC; A. Slesarev and J. Lake, Univ. of California, Los Angeles,

Methanopyrus kandleri uniquely possesses striking combination of prokaryotic and eukaryotic traits. For example, M. kandleri is the only organism in which both a major bacterial-like type 1A topoisomerase (reverse gyrase) and a major eukaryotic-like type 1B topoisomerase (topoisomerase V) co-exist (1,2). In addition it was found that M. kandleri contains a histone (MkaH) which is, as we established recently, more closely related to the eukaryotic histones, than any other prokaryotic histones (3). Eukaryotic topo I and histones play major roles in the solenoidal organization of eukaryotic chromatin, which lacks superhelical stress. Apparently there is a parallel in the control of chromosome topology between eukaryotes and M. kandleri since the latter has both histone and eukaryotic-like Topo V. We have shown recently using a Topo V-based topology assay that MkaH as well as other prokaryotic histones wrap DNA in negative constrained supercoils in physiologically relevant conditions (high temperature and high salt). This provides an additional evidence of a similarity in DNA packaging between eukaryotes and some prokaryotes. Topo V is an abundant and extremely powerful topoisomerase activity in vitro. The rate of DNA unlinking (separation of complementary strands) at 108C is about 4 cycles/s/enzyme monomer or 16 times faster than the rate of DNA relaxation at 90C, and much faster than the rates of other known topoisomerases. This is a real puzzle since it suggests that topo V would denature M. kandleri DNA around 100C in a matter of minutes unless its unlinking activity is inhibited by unknown mechanism(s). A clue to other

possible functions of topo V in M. kandleri could provide our recent finding that Topo V has two activities that are important for DNA repair: it cleaves the phosphodiester bond at apurinic/apyrimidinic (AP) sites and removes 2-deoxyribose 5-phosphate (dRP) containing termini. Both topo V isolated from M. kandleri and recombinant one reveal these activities. It is intriguing why DNA topoisomerase and DNA repair activities reside in one protein.

1. Slesarev, A.I., Stetter, K.O., Lake, J.A., Gellert, M., Krah, R., & Kozyavkin, S.A. (1993). Nature 364: 735-737.
2. Kozyavkin, S. A., Krah, R., Gellert, M., Stetter, K. O., Lake, J. A., and Slesarev, A. I. (1994). J Biol Chem 269, 11081-9.
3. Slesarev AI, Belova GI, Kozyavkin SA, Lake JA (1998). Nucleic Acids Res 26, 427-430.

Session VI - General Genomics and Molecular Genetics

SA-24. Molecular Phylogeny in Crisis : the Case of the Universal Tree of Life.

Patrick FORTERRE and Herve PHILIPPE.
Univ. Paris-Sud, ORSAY, France.

For the last ten years, the most popular version of the universal tree of life, based on rRNA and elongation factors phylogenies, featured the grouping of eucaryotes and archaea in the same clade and a hyperthermophilic last universal cellular ancestor (LUCA). This traditional tree has been recently shaken by data from comparative genomics and by several well documented protein phylogenies. In particular, the misplacement of microsporidiae in both rRNA and elongation factor trees has emphasised the artefacts induced by the phenomenon of long branch attraction (LBA). LBA is especially critical when sequence data sets are saturated with multiple substitutions and when outgroups used to root trees have long branches. In particular, LBA can explain why the rooting of the tree of life is in the bacterial branch, despite the absence of valid phylogenetic signal in the data sets used (1). This strongly questions the use of traditional methods of molecular phylogeny to infer ancient evolutionary events. The best way to retrieve some signal from sequence alignments is presently to use insertion/deletion to check critical nodes in phylogenetic tree, another possibility might be to discriminate between good and bad positions according to their variability (2). In addition to LBA, various other pitfalls can explain discrepancies between different phylogenies. Emphasis is presently put on massive lateral gene transfer to explain the mixture of features from two domains in the third one, but comparative genomics shows that other factors have play a significant role in shaping microbial evolution, such as paralogy, gene loss or else non orthologous displacement (3). From these considerations, we propose new hypotheses to explain recent data obtained by comparative genomics, a novel topology of the universal tree, and an evolutionary scenario from a complex LUCA to simpler but more efficient procaryotes.

PHILIPPE, H. and FORTERRE, P. J. Mol. Evol. in press
LOPEZ, P., FORTERRE, P. and PHILIPPE, H. J. Mol. Evol. in press
FORTERRE, P. Current Opinion Gen. Dev., 7, 764-770 [1997)

SA-25. Genome Scanning in *Haemophilus influenzae*: *in vitro* Transposition and PCR Analysis for the Identification of Essential Genes.

KARL A. REICH*, LINDA E. CHOVAN AND PAUL HESSLER, Abbott Laboratories, Abbott Park IL.

The genomic sequence of a number of micro-organisms are now available and their analysis reveals many hundreds of open reading frames of unknown function. Using the assumption that essential genes are those that cannot tolerate a loss of function mutation, we have developed a method for identifying 'essential genes'. Using a small (~975 bp) insertional element containing an antibiotic resistance marker and an in vitro transposition system, we have constructed a saturated mutant library in *H. Influenzae*. Our mutant bank has >300,000 inserts randomly distributed throughout the genome (~ 1 insert/6 nucleotides). The position of these inserts relative to the deduced open reading frames of *H. Influenzae* can be determined by PCR and Southern analysis. Essential genes are identified two methods: 'mutant exclusion' or 'zero-time analysis'. Mutation exclusion consists of growing the complete mutant library and identifying open reading frames that do not contain insertional elements: in a growing population of bacteria, insertions in essential genes are excluded. Zero-time analysis consists of saturation mutagenesis of a defined chromosomal segment and following the fate of individual insertions after transformation in a growing culture: the *loss* of inserts in essential genes are followed over time. Both methods of analysis permit the identification of genes required for bacterial survival. The list of essential genes changes with the environment tested: rich media *vs.* minimal media *vs.* animal infection model. The construction of the mutant bank and examples of mutant exclusion and zero-time analysis are presented. The method is transferable to any naturally competent bacterial species whose genomic sequence is available.

SA-26.

Tn*5* transposition: A simple efficient tool for manipulating DNA molecules *in vitro*. I. Yu. GORYSHIN, D. YORK, M. ZHOU, A. BHASIN, L. M. BRAAM AND W. S. REZNIKOFF. Department of Biochemistry, University of Wisconsin, Madison, WI.

DNA transposition is a process catalyzed by transposase in which DNA sequences defined by short specific DNA sequences are "moved" from one site to a second site. For Tn*5*, transposition is a "cut and paste" process. We describe an *in vitro* Tn*5* transposition system which utilizes hyperactive transposase and hyperative end DNA sequences. The system is simple requiring only one protein (transposase), a transposon DNA defined by inverted 19 bp sequences, a target DNA, and a buffer containing Mg^{++}. The system is efficient; more than 50% of the transposon is metabolized in a typical 3 hr incubation. The insertions are random. We will describe the use of a Tn*5* intramolecular transposition system to generate nested deletion templates for sequencing from both directions, and the use of Tn*5* intermolecular transposition to tag large molecules. Other applications of this system will also be described.

SA-27. A High Efficiency, Low Target Site Specificity in vitro Transposition Reaction: A Powerful Tool for in vitro Random Mutagenesis and Sequencing.

M.C. BIERY, A.S. STELLWAGEN, B.E. SLATKO, E.A. RALEIGH, AND N.L. CRAIG.
Biery-Howard Hughes Medical Inst./Johns Hopkins Univ. School of Medicine, Baltimore, MD.

An adapted in vitro transposition reaction utilizing transposition proteins from the bacterial transposon Tn7 has been developed as an efficient tool for in vitro random mutagenesis of a variety of DNA targets. In vivo, Tn7 relies on one of two target selection mechanisms to direct insertions, employing either the high frequency highly target site specific reaction mediated by DNA binding protein TnsD, or the low frequency, low target specificity reaction mediatied by an alternative protein, TnsE. If TnsA, TnsB, and TnsC are present in the cell in the absence of both of the targeting proteins, then no transposition occurs. A previous screen for mutants of TnsC that relieve the requirement for TnsD and TnsE yielded several gain-of-function mutants capable of activating transposition in the absence of these targeting proteins. A particular mutant, TnsCA225V, demonstrated significant levels of transposition in vivo and in vitro. PCR and sequencing analyses of transposition products isolated from in vitro reactions have shown that this mutant is capable of directing insertions of a variety of miniTn7 elements to many sites on target molecules with low sequence specificity. Because of the high efficiency and low target site specificity of the TnsA+TnsB+TnsCA225V reaction, it has become an excellent tool for random mutagenesis of DNA in vitro. The miniTn7Kan element has been cloned into a counterselectable plasmid harboring an R6K-gamma origin of replication to facillitate exclusive recovery of simple insertions into competent replicons upon transformation into E. coli.

SA-28.

Pseudogenes on the *Borrelia* linear plasmids

SHERWOOD CASJENS*, WAI MUN HUANG, MARGARET ROBERTSON, GRANGER SUTTON[1] and CLAIRE FRASER[1] Dept. of Oncological Sciences, U. of Utah School of Medicine, Salt Lake City, UT and The Institute for Genomic Research, Rockville, MD[1]

The complete sequence of the genome of *Borrelia burgdorferi* strain B31 reveals that it carries 12 linear and 9 circular plasmids. Analysis of the sequence has shown that there are many paralogous families of sequences scattered about the linear plasmids. Analysis of these groups of similar sequences shows that the complex relationships among the plasmids within this strain apparently represent recombinational events of the following types: (i) integration of one plasmid into another, (ii) transposition, (iii) insertion, (iv) deletion, (v) inversion, (vi) amplification of short direct repeat tracts, and (vii) exchange of telomeres. Non-homologous and non-reciprocal recombination events appear to have occurred. In addition, we have found a reassortment of sequences between chromosome and plasmid telomeres. One result of this "wild and crazy" recombination among the linear plasmids is the presence of many apparently nonfunctional fragments of genes or pseudogenes. These gene fragments account in part for the observation that the predicted fraction of protein-encoding DNA on these linear plasmids varies from 83% down to 32%, compared to about 90% for the *Borrelia burgdorferi* chromosome and circular plasmids as well as in other completely sequenced bacterial genomes.

Thursday, September 24
Session VII - Short Talks

SA-29. The NeuroGadgets Inc. / University of Ottawa Bioinformatics Web Site

R. L. CHARLEBOIS.
Univ. Ottawa, Ottawa, ON, Canada.

A web site has been developed which facilitates the study and comparison of sequenced microbial genomes. Available queries include: listing taxon-specific genes within a genome; listing, sorting and plotting predicted protein characteristics; finding gene families; comparing genomic inventories; exposing genetic mosaicism; finding rapidly-evolving genes; determining the extent of gene order conservation; constructing whole-organism phylogenies; and more. The user is given great flexibility in setting query parameters. As an example, one can easily find all of the predicted proteins within M. jannaschii which are in the top acidic 15% of ORFs 200-300 aa long, whose best match outside of the Archaea is to Cyanobacteria. Future plans for the site include analyses dealing with the positional context of genes, implementation of neural network pattern recognition, and linking to the Sequence Retrieval System (SRS).

SA-30. Recent Advances in Characterization of the Major Facilitator Superfamily (MFS) Superfamily (MFS).

TSENG T-T; JAHN PS; HUANG S-C; JACK D; SAIER MH JR..
Department of Biology, Univ. of California, San Diego, La Jolla, CA.

The major facilitator superfamily (MFS) is a very old, large and diverse superfamily that includes several hundred sequenced members. It is one of the two largest families of membrane transporters found on Earth. They catalyze uniport, solute:cation (H^+ or Na^+) symport and/or solute:H^+ or solute:solute antiport. They exhibit specificity for sugars, polyols, drugs, neurotransmitters, Krebs cycle metabolites, phosphorylated glycolytic intermediates, amino acids, peptides, osmolites, nucleosides, organic anions, inorganic anions, etc. They are found ubiquitously in all three kingdoms of living organisms. We here describe the recent advances on the characterization of the major facilitator superfamily. Results from PSI-BLAST indicated the possible connection between the macrolide-efflux protein (MEF) family and the MFS transporters. The MEF family is a moderately sized family with only a few functionally characterized members. The latter proteins appear to function in macrolide, multiple drug or heavy metal ion efflux. These proteins exhibit 11 or 12 putative transmembrane alpha-helical spanners (TMSs) and 400-500 amino acyl residues in length. Their largest inter TMS loop is in the middle of the proteins between putative spanners 6 and 7. In these respects, proteins of the MEF family resemble the proteins of the MFS. Another recent addition to the MFS is the PucC family. Two members in the family are characterized as bacteriochlorophyll synthase. The members of the PucC family exhibit 12 putative TMSs. With the above evidence, we propose the addition of the PucC family into the major facilitator superfamily.

SA-31. Gleanings from a Comparative Genomics Study: Another Look at DNA Replication.

PATRICK V. WARREN, JAMES R. BROWN, MICK GWYNN, EARL MAY.
SmithKline Beecham Pharmaceuticals, Collegeville, PA.

Suppositions about the evolutionary history of protein families, metabolic pathways and multi-subunit protein complexes can now be proposed in light of the ever increasing number of complete and partial genomic datasets. By organizing all possible open reading frames and their pairwise similarities into a relational database we can rapidly access a protein's presence or absence in a particular species and potentially reconstruct how it evolved through time. We have chosen DNA replication as a case study to examine how well our biochemical and genetic knowledge of the model replication system in Escherichia coli translates to replication systems in other bacteria. By examining over 33 complete and partial genomes from Bacteria, Archaea and Eucarya we show evidence for significant variation from the well characterized E. coli model. We also present the minimal set of common proteins involved in DNA replication and a possible link between bacterial and eukaryotic DNA replication.

SA-32. Genome Wide Expression Analysis and Comparative Genomics in E. coli K-12.

Craig Richmond*, Jeremy Glasner, Robert Mau and Frederick Blattner.
University of Wisconsin, Madison, WI.

High density arrays of PCR products corresponding to greater than 95% of all open reading frames (ORFs) of E. coli K-12 (MG1655) were used to access global changes in gene expression under varying growth conditions. Conditions tested in this work include induction with IPTG, heat shock at 50oC, osmotic shock and growth on rich vs. minimal media. Induction with IPTG showed the expected response with genes of the lac operon induced to high levels and an additional operon, melAB, also induced to significant levels. Global changes in gene expression due to heat shock showed induction of a group of genes known to be part of the heat shock regulon. A set of previously uncharacterized genes were also found to be induced to significant levels, adding to the list of members of the heat shock regulon. This study also identified a group of genes that were specifically down-regulated in response to increased temperature. These results illustrate the utility of whole genome expression analysis in E. coli. Data using Affymetrix GeneChip technology to perform comparative genomic analysis in E. coli will also be presented. A set of

experiments was carried out to compare the E. coli K-12 genome with that of the pathogenic E. coli strain O157:H7. For this study genomic DNA from both strains was biotin labeled and hybridized with two separate E. coli K-12 specific GeneChips. Comparison of hybridization signal from both chips allows global genomic mapping to be carried out in a single, rapid experiment.

ABSTRACTS FOR POSTER PRESENTATION

PA-01. Topo IV and Gyrase Mutations Can Alter The Structure of Bacterial Chromosome In Vivo.
PAWEL W. Staczek and N. PATRICK Higgins.
Univ. of Lodz, Lodz, Poland, Univ. of Alabama at Birmingham, Birmingham, Alabama.

Our previous results utilizing the γδ resolution system showed that the supercoil „domain" encompassing the min. 43-45 segment of the *Salmonella typhimurium* chromosome is dynamic. We found no evidence for existence of stable barriers organizing the chromosome into sequence specific 50-100 kb DNA loops in this sector of the genome. Now we analyzed reaction kinetics for γδ site-specific recombination in six chromosomal intervals ranging in size from 14-90 kb. We found the evidence for stochastic behaviour in which the stationary barriers vary in number and/or position either from moment to moment or from cell to cell. To test the biochemical nature of domain barriers, a genetic screen was used to find mutants with small domains. Rare temperature sensitive mutants of DNA gyrase (*gyrA, gyrB*) and Topoisomerase IV (*parC*) increase the apparent frequency of domain barriers throughout the tested region of the chromosome. From the obtained results we propose that „random"topological tangling of the interwound DNA strands accounts for a significant fraction of the domain behaviour in bacterial chromosomes.

PA-02. Correlation of PAS Domains with Electron Transport-Associated Proteins in Microbial Genomes.
I.B. ZHULIN and B.L. TAYLOR.
Dept. Microbiol. & Mol. Genet., Sch. Medicine, Loma Linda Univ., Loma Linda, CA.

PAS domains are signal transduction modules in a variety of sensor proteins from both eukaryotes and prokaryotes that sense light, oxygen and redox (1, 2). There is an asymmetric distribution of PAS domains in the microbial genomes. Of eleven analyzed genomes, PAS domains are absent from five: *M. genitalium, H. influenzae, H. pylori, B. burgdoferi* and *M. Jannaschii*. PAS domains are abundant in two genomes, from *synechocystis* sp. and *A. Fulgidis*. Some proteins contain more than one PAS domain. Both PAS orthologs and paralogs have been identified. There was no correlation between the size of the genome and the number of PAS domains within a given species. There was however a positive correlation ($r = 0.925$; $P < 0.01$) between the number of PAS domains and electron transport-associated proteins. A relationship of PAS domains with aerobic and sulfate respiration, and photosynthesis is presented.

1. Zhulin IB & BL Taylor. 1997. Trends Biochem. Sci. 22:331.
2. Ponting C & L Aravind. 1997. Curr. Biol.7:674.

PA-03. Genomic Structure And Phylogeny Of *Brucella* Strains Isolated From Marine Mammals.
Boschiroli (1), Bourg G.(1), Jumas-Bilak E.(1), Macmillan A.(2), O'Callaghan D.(1) and Ramuz M.(1).
(1) Institut national de la sante et de la recherche medicale, Unité 431, Faculte de medecine, Nîmes France and (2) FAO/WHO Collaborating Centre for Reference and Research on Brucellosis, Central Veterinary Laboratory, Addelstone, UK.

The genomic structure and the restriction maps were studied in 24 Brucella strains isolated from marine mammals: seals, porpoises, dolphins, an otter and a mink whale. On the basis of the SpeI restriction data, the strains could be ascribed to three groups, each corresponding to a specific host. The otter isolate has a pattern identical to that of seal isolates and the pattern of the whale isolate is identical to that of porpoises. This suggest that the marine strains form clones selected by host restricted virulence. Restriction maps were obtained for three strains representative of the three groups and compared to the already published map of the type strain, B. melitensis 16M. These maps are rather similar, with the presence of two chromosomes, as in the type strain. However, differences in restriction fragment length could be observed. They are due to small deletions or insertions, presence or absence of restriction sites, together with a large inversion affecting the large chromosome of one strain. We have identified a 62 kb SpeI fragment specific to the small chromosome of seal isolates. A phylogenetic tree was obtained from restriction data.

PA-04. A Rapid Method for Identifying Genes Involved in DNA Damage Resistance in *Deinococcus radiodurans* R1.
N. C. SHANK AND J. R. BATTISTA.
Louisiana State Univ. and A & M Coll., Baton Rouge, LA.

The availability of nearly complete genomic sequence information for the wild type strain of *Deinococcus radiodurans* R1 has permitted the development of a method for rapidly identifying defective loci in randomly generated ionizing radiation sensitive (IRS) mutants of this organism. Initially, a cosmid library is screened for clones that restore ionizing radiation resistance to an IRS

strain. (*D. radiodurans* R1 is naturally transformable and competent throughout exponential growth. This species can be efficiently transformed with closed circular or linear fragments of DNA as small as 1000bp.) Once a cosmid is found, the DNA sequence of approximately 300 bp of the ends of the genomic insert is determined. This sequence is matched to the *D. radiodurans* R1 genomic sequence database, defining the region of the *D. radiodurans* R1 chromosome that restores radioresistance. Candidates for the gene of interest are chosen by examining a translation of all open reading frames within the cosmid insert. After likely candidates are identified, the sequence available from the database is used to design primers that permit amplification of these genes by PCR. An attempt is made to restore ionizing radiation resistance by transforming the IRS strain with the amplified sequence. This approach has been employed to establish the identity of alleles that result in a DNA damage sensitive phenotype in several strains of *D. radiodurans* R1. For example, the *irrB* locus of strain IRS18 is now recognized as a homolog of the *uvrD* gene of *E. coli*.

PA-05. Orthologous Enzymes in Microbial Genomes: Using COG Database for Analysis of Metabolic Pathways and Identification of Missing Enzymes.

M. Y. GALPERIN, R. L. TATUSOV, and E. V. KOONIN.
National Center for Biotechnology Information, NIH, Bethesda, Maryland.

The COG system identifies orthologous proteins in each of the completely sequenced genomes based upon all-against-all comparisons and is largely unaffected by the differences in their relative evolution rates. Clusters of orthologous proteins (www.ncbi.nlm.nih.gov/COG) delineated by this approach allow the use of the information collected about better studied organisms (*E. coli, B. subtilis*, yeast) in the analysis of the genomes of poorly studied microbes. By analyzing the patterns of distribution of homologous enzymes among various genomes it becomes possible to identify the metabolic pathways that are present or absent in a given organism.

By superimposing COGs on the biochemical pathways map, we were able to identify the missing enzymes for purine, pyrimidine, amino acid and cofactor biosynthesis pathways in 16 organisms with completely sequenced genomes. The results obtained led to re-evaluation of the role of glycolysis in *H. pylori*, uncovering of a certain degree of metabolic plasticity of biosynthetic pathways in human parasites (*H. influenzae, H. pylori*), and establishment of non-orthologous gene displacement as a major source of enzyme diversity in various microorganisms. In certain cases, (phosphoglycerate mutase, riboflavin synthetase) the use of the COG system resulted in confident prediction of archaeal homologs of the enzymes that could hardly be identified by other approaches.

Cross-genome comparison implemented in the COG system offers a way to unify functional annotation in all species with completely sequenced genomes. The COG system can be helpful in identifying and resolving the cases where the same genes (operons) from different organisms (e.g., *E. coli* and *B.subtilis* or yeast) have been assigned different names. Finally, the COG approach can be used to design algorithms for automated functional annotation of newly sequenced genomes.

PA-06. Phylogenetic Analysis of Known and Putative Sigma Factor Genes From Selected Bacteria, Including Over 20 Identified in the Genome of *Pseudomonas aeruginosa* PAO1.

F. S. L. BRINKMAN AND R. E. W. HANCOCK.
Univ. of British Columbia, Vancouver, British Columbia, Canada.

Within the *Pseudomonas aeruginosa* PAO1 genome sequence, we have identified a large number of putative sigma factors genes. Twenty two have been identified in total, with eighteen of them sharing highest similarity with sigma factors of the ECF (extracellular factors) class. In order to gain insight into the evolution of these genes, we have performed a comprehensive phylogenetic analysis of all putative sigma factor gene sequences from completed bacterial genomes, with a focus on genes we have identified from *P. aeruginosa* genomic sequence. Some additional sequences from other selected bacteria were also included. For *P. aeruginosa*, each new gene was PCR amplified and its sequence verified against the preliminary genome project sequence. Trees were constructed using both distance matrix and parsminony methods. While non-ECF class sigma factors generally had a branching order and degree of sequence divergence consistent with orthologous evolution, the ECF sigma factor gene sequences were much more divergent and branching order was not necessarily consistent with organism relationships. Some of the *P. aeruginosa* ECF sigma factors genes formed clusters, and these clustered genes had similar flanking genes, however these clusters were deeply rooted, and so have not likely evolved from recent gene duplication events. In fact, none of the *P. aeruginosa* ECF sigma factors seemed to have evolved as a result of recent gene duplications. Branching order and branch lengths suggested that many of these genes may have horizontal origins. The origin of these genes, and the implications for ECF sigma factor and *P. aeruginosa* genome evolution in general, is discussed.

PA-07. Fluorescence-Based Isolation of *Bartonella henselae* Genes Differentially Expressed Within Host Cells.

A. SEUBERT, C. LANZ AND C. DEHIO.
Max-Planck-Institut f. Biologie, Abt. Infektionsbiologie, Tuebingen, Germany.

In this study we report on the application of Differential Fluorescence Induction (DFI, for review see Valdivia and Falkow, Trends Microbiol. 5, 360-363) to isolate differentially expressed genes which play a role in invasion and intracellular maintenance of Bartonella henselae within host cells. Infection by *B. henselae* may cause a variety of clinical manifestations, e.g. bacillary angiomatosis, bacillary peliosis, cat scratch disease and persisting bacteremia. Due to the various specific host cell interactions mediated by this pathogen, a subtle regulation of bacterial gene expression can be expected. The transcriptional fusion of a library of subgenic DNA fragments of the *B. henselae* chromosome to a promoterless gfp gene encoding the green fluorescent protein (GFP) as a convenient and sensitive expression marker should allow to isolate differentially regulated promoters by the aid of fluorescence-activated cell sorting (FACS sorting). To this end we constructed an expression vector based on a broad-host-range plasmid carrying a promoterless gfp gene, and subgenic (0.2 - 0.8 kb) fragments of the *B. henselae* chromosome were cloned upstream of the gfp gene. We have infected endothelial cells with the resulting library of fluorescent and non-fluorescent bacteria. Following the killing of extracellular bacteria by gentamicin and the lysis of host cells, fluorescent intracellular bacteria could be recovered by FACS sorting. These fluorescent bacteria were plated on blood agar and agar grown bacteria were subsequently sorted for weak fluorescence. We are currently cloning genes corresponding to the promoters isolated from these populations. Development of DFI as a high-throughput screening for differentially expressed genes in *B. henselae* should represent a first approach to functional genomics in this emerging pathogen.

PA-08. Microbial GRAIL Gene-finding Systems.

F. LARIMER, R. MURAL, M. SHAH, A. SUBRAMANIAN, AND E. UBERBACHER.
Computational Biosciences, Life Sciences Div., Oak Ridge National Laboratory, Oak Ridge, TN.

Open reading frame assignment in microbial genomes is relatively straight forward because of the generally simple gene structure and genome compactness. Generating reasonable gene models and annotation proceeds by identifying homologs in various sequence databases. Novel genes, a substantial fraction of microbial genomes, represent a more difficult problem. The GRAIL gene finding program combines multiple lines of evidence from various statistical properties of DNA sequence, pattern recognition, and rules based on the biology of the organism, using techniques of machine learning and othe AI methods to predict the portions of a DNA sequence which have protein coding potential. Our goals in applying GRAIL to microbial sequences are 1) to provide consistent ORF assignment to new genomes as a prerequisite to annotation, 2) to determine if taxon-specific GRAILs can be utilized to simplify gene identification in new genomes, and 3) to automate the process of building microbial GRAIL systems. (Research sponsored by the Office of Health and Environmental Research, USDOE under contract number DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp.)

PA-09. Comparison of a 200-kb Chromosomal Region from Three Strains of Rhodobacter capsulatus.

Y. Kogan, R. Cox, V. Cheung, R. Haselkorn and M. Fonstein.
Univ. of Chicago, Chicago, IL.

As a part of the genome project of the purple photosynthetic bacterium Rhodobacter capsulatus we are comparing chromosomal structures of three closely-related strains of R. capsulatus in order to understand a nature of genome changing events.

High-resolution physical maps derived from the ordered cosmid libraries were aligned [1,2]. This alignment shows numerous large and small translocations in the chromosomes of these three strains. Regions of 15-80 kb in which restriction sites are highly polymorphic are interspersed with regions in which the positions of restriction sites are highly conserved producing the mosaic structure of the Rhodobacter chromosome.

The genomic sequence of strain SB1003 has been recently determined (http://rhodol.uchicago.edu/rhodo). A 200-kb chromosomal regions from the other two R. capsulatus strains were chosen as having varying degrees of restriction map polymorphism. Studying these particular DNA regions, we found an average level of identity of all three DNA sequences to be 92%. In addition to large scale rearrangements and transposable elements, small chromosomal regions with a very high level of polymorphism were found. Correlation between positions of these highly polymorphic DNA stretches and distribution of different unusual DNA motifs has been shown.

Different level of conservation of DNA sequence in the coding and non-coding regions has been demonstrated. It has been used as a tool to analyze hypothetical ORFs existing in all of these strains.

PA-10. Phylogenetic Analysis of TRAP-transporters - A New Type of Periplasmic Solute Transporters.
RABUS R1, JACK D1, KELLY D2, SAIER JR MH1.
1Univ. of California San Diego, La Jolla, California 2Univ. of Sheffield, Sheffield, United Kingdom.

The TRAP (for tripartite ATP-independent periplasmic) transporters represent a new type of periplasmic secondary transporters for the C4-dicarboxylates malate, succinate and fumarate. This transport system was first discovered in *Rhodobacter capsulatus*. There are three genes (*dctPQM*), one encoding the periplasmic C4-dicarboxylate-binding protein DctP and two encoding the membrane integral proteins DctQ and DctM. Transport via the Dct-system was found to be driven by pmf and to be independent of ATP (1). We conducted phylogenetic studies of the three protein constituents of the TRAP-system (DctP, DctQ and DctM). The DctP-family comprises 9 members, the DctQ characterizes the smallest family with only 3 members and DctM the largest family with 11 members. We found constituents of the Dct-system in the Eubacteria *Escherichia coli*, *Haemophilus influencae*, *Salmonella typhimurium*, *Rhizobium* sp. and *Synechocystis* sp. and in the Archaea *Archaeoglobus fulgidus*. (1) Forward et al. 1997. J. Bacteriol. 179:5482-5493.

PA-11. Characterization of a Superfamily: RND-ATP-SecDF.
TSENG T-T; GRATWICK K; SAIER MH Jr.
Department of Biology, Univ. of California, San Diego, La Jolla, CA.

The antibiotic transport-associated protein (ATP) family consists of many proteins, all of high GC Gram-positive bacterial origin, that range in reported size from about 700 amino acyl residues to about 1200 residues. None of these proteins is functionally well characterized, but one of them, the ActII-3 protein of *Streptomyces coelicolor*, has been implicated in antibiotic export. We here characterize this superfamily by multiply aligning their sequences and define their evolutionary relationships between the members of the families. Average hydropathy, average similarity and average amphipathicity plots have allowed us to identify 12 putative TMSs with the homologues in *Mycobacterium tuberculosis*. The first TMS is near the N-terminus of the protein. This is followed by a hydrophilic domain of about 120-160 residues, then a hydrophobic domain of six putative TMSs, then a hydrophilic domain and finally, a hydrophobic domain. A similar arrangement has been observed in the resistance-nodulation-cell division (RND) family and these two families are proposed to be related, based both on their similar topologies and PSI-BLAST results. The transmembrane regions of the SecD and SecF protein components of the protein secretory apparatus of the IISP family also exhibit significant sequence similarity to proteins of both the RND family and the ATP family. With evidence upon their similar topology and sequence similarity, we propose that the RND, ATP and SecDF families should be classified within the same superfamily.

PA-12. Low-resolution Physical Map of the Magnetospirillum magnetotacticum Genome.
E. BERTANI, B. HAY, J. KIRSCHVINK, and S. QUAKE.
Caltech, Pasadena, CA.

M. magnetotacticum synthesizes single-domain, membrane-enclosed crystals of the iron mineral magnetite (Fe3O4), which it stores in chains called magnetosomes. Our goal is to isolate and identify the genes that are involved in this process, with the idea of eventually constructing an in vitro system for the synthesis of magnetite. We have begun by determining the physical structure of the genome. The DNA, which is 62% GC, was digested with restriction enzymes, SwaI and PmeI, that recognize 8 bp-AT-rich sequences, yielding 12 fragments from 40-900 kb, giving a total genome size of almost 4.5 Mb. Undigested DNA does not enter a pulsed-field gel, suggesting that the genome is a circular structure. So far, we have not detected any plasmids. Using random clones as probes of Southern blots prepared from pulsed-field gels, we have been able to arrange the fragments in a circular chromosome, although the positions of the smaller fragments are not yet exact. We have mapped some previously identified genes such as bfr, sodB, dnaA and the 16S ribosomal clusters and are currently isolating and mapping genes involved in iron uptake and utilization.

PA-13. The Amino Acid/Auxin:Proton Symport Permease Family.
YOUNG, GB; JACK, DJ; SMITH, DW; SAIER JR, M.
Univ.of California San Diego, La Jolla, CA.

Amino acids and their derivatives are transported into and out of cells by a variety of permease types which comprise several distinct protein families. Preliminary evidence from a systematic phylogenetic analyses is presented suggesting that proteins of the amino acid/auxin permease (AAAP) family are distantly related to proteins of the large ubiquitous amino acid/polyamine/choline (APC) family, as well as to those of two small bacterial amino acid transporter families. We have examined these amino acid specific families for sequence, topological and functional similarities and have found that some of them exhibit similarities that convince us of a probable common evolutionary origin.

PA-14. Pigment Formation is Not a Figment of *E. coli*'s Imagination.

A. PATEL, C. CANTIN, Y. CHANE-YENE, D. FOURNIER, AND P.C.K. LAU.
Natl. Res Council Canada, Biotechnology Res Inst., Montreal, Quebec, Canada.

E. coli is known to turn blue by the introduction of an exogenous source of genes such as the naphthalene/toluene dioxygenase complexes or it can become black by a cloned tyrosinase gene. In this study we examined the function of a set of genes (tentatively named *phdRC1C2BDA*, ASM News 64:1, 1998) present in the *E. coli* chromosome at about 57 centisomes (Blattner et al. Science, 277:1453-1462, 1997). The predicted amino acid sequences of PhdC1C2BA are counterparts of the toluene (TodC1C2BA) dioxygenase and related complexes; PhdD is a homolog of the *cis* - dihydrodiol dehydrogenase, eg. TodD. The divergently transcribed *phdR* gene is a member of the LysR-type of transcriptional activators. A cloned *phdRC1C2BDA* or *phdC1C2BDA* cluster on a pGEM
plasmid enabled the *E. coli* host to impart a diffusible brown pigmentation when cultivated on Luria broth or LB plate. Addition of indole-2-carboxylate intensified the pigmentation more than indole-3-carboxylate or indole. But the colonies never turned blue contrary to known dioxygenases or some monooxygenase systems. The brown pigmentation is
reminiscent of allomelanin (catechol melanin) formation by some bacteria. That a catechol structure is a possible precursor for the coloration was supported by disruption of the *phdD* gene in *phdR*C1C2BDA which led to loss of the phenotype. The brown pigmentation did not change upon incubating the cells in the presence of copper or L-tyrosine, hence dismissing tyrosinase activity. When supplied in *trans*, *phdR* alone can activate a number of *E. coli* hosts by producing a brown culture. Presumably *phdR* acts as a transcriptional activator on a set of well conserved genes in a number of *E. coli* K-12 strains. We have determined the frequency of occurrence of *phd* genes as well as the associated *mhp* operon responsible for the metabolism of 3-hydroxyphenylpropionate in the ECOR collection, a sample of 72 *E. coli* isolates which provides a reasonable breadth of genetic variation of the species.

PA-15. Functional Analysis Of *Deinococcus radiodurans* Genome By Targeted Mutagenesis.

LM MARKILLIE, WB CHRISLER, KK WONG.
Pacific Northwest Natl. Lab., Richland, WA.

Deinococcus radiodurans, previously known as *Micrococcus radiodurans*, strains R1, has extreme resistance to genotoxic chemicals, oxidative damage, high levels of ionizing and UV radiation, and desiccation. The ability to survive such extreme environments is attributed in part to a unique DNA repair system in combination with its chromosome copy number and structure, as well as factors affecting the survival of other cellular components. There is evidence suggesting that the carotenoids which cause red pigmentation in *D. Radiodurans* may act as free radical scavengers, thus increasing resistance to DNA damage byhydroxyl radicals. High levels of two oxygen toxicity defense enzymes, superoxide dismutase and catalase, are also found in *D. Radiodurans*. In addition, the Deinococcal outer membrane lipids are complex and distinct from those found in the rest of the bacterial world and it has been suggested that they, together with the plasma membrane, may also be involved in stress resistance. However, the genetic basis for these stress resistance is still not very clear. With the genomic sequence information of *D. Radiodurans* R1, we have developed a simple, and general targeted mutagenesis method to perform a genome-wide analysis of putative genes involved in the stress resistance. We have generated mutations in *katA* (catalase) and *sodA* (superoxide dismutase). Both *katA* and *sodA* mutants are shown to be required for the extreme ionizing radiation resistance. Several other mutants have been generated and are being analyzed for their survival under different stress conditions.

PA-16. Microsatellite Variability in *Escherichia coli.*

D METZGAR, E THOMAS, D FIELD, C.L. DAVIS, C WILLS.
METZGAR: Univ. of California at San Diego, La Jolla, CA; THOMAS: Evergreen State Coll., WA; FIELD, DAVIS, WILLS:Univ. of California at San Diego, La Jolla CA.

Our lab has been invesitigating microsatellite polymorphisms in a variety of microorganisms and viruses (S. cerevisiae, C. albicans, cytomegalovirus). We have found that even very short reiterative sequences can be length polymorphic in many of these organisms, providing both useful molecular strain markers and manageable model systems for laboratory studies of microsatellite mutational dynamics. We are currently extending this work to E. coli. Our basic approach involves utilizing published genome sequences to identify the longest microsatellites in a species, then using PCR to amplify the identified loci from a wide variety of wild isolates, and acrylamide gel electrophoresis to screen for polymorphic sites. The genome of E. coli is very streamlined, and contains very little repetitive DNA, though it does contain microsatellites equal in length to the shortest microsatellites found to be polymorphic in our previous work with cytomegalovirus. We will compare the microsatellite length distributions seen in prokaryotic

and eukaryotic microorganisms, and consider underlying evolutionary forces and consequences. Results will be presented.

PA-17. Replication Slippage in a Viral Genome: New Tools for Genome Evolution and Epidemiology.
C.L. DAVIS, D. FIELD, D. METZGAR, C.WILLS
Univ. of California at San Diego.

In this study, we examined the nature of short tandem repeats in a large viral DNA genome and show that extremely short tandem repeats (iterations of 1-3 bases) are a surprising source of variability among clinical isolates. Human cytomegalovirus (HCMV) is a global pathogen of clinical significance, especially in immunosupressed individuals. This genome is only 230kb and due to its economical size lacks long iterations (microsatellites) predicted to be sites prone to replication slippage in eukaryotes. Strikingly though, comparisions show an excess of observed versus expected numbers of mono- through trinucleotide repeat loci, and PCR amplification confirms that these are sites prone to the accumulation of replication slippage mutations. 30 of the 38 loci examined (including all triplet-repeats of four iterations or greater in the Genbank reference genome) displayed between 2 and 5 length-variable alleles in a screening of 12 clinical isolates. The maximum lengths of the loci tested were only 12, 5, and 5 repeats for mono-, di- and trinucleotide repeat loci respectively. 8 loci were found to be monomorphic. The finding that these loci show significant instability has implications for understanding genome evolution in this large DNA virus. It also provides a novel panel of molecular markers interspersed throughout the HCMV genome that should greatly aid evolutionary and epidemiological studies of this virus.

PA-18. Large-Scale Identification of *Staphylococcus aureus* Genes Contributing To Virulence in Murine Models of Infection.
B. BENTON, S. BOND, R. C. BURKE, J. BUYSSE, T. CHRISTIAN, H. FANG, , H. LYONS, K. M. WINTERBERG AND J. P. ZHANG.
Microcide Pharmaceuticals, Inc., Mountain View, CA, USA.

We have developed a genetic system for identifying microbial virulence genes in chemical or transposon induced mutants. The method employs unique DNA size tags to monitor the fate of up to fifty individual strains comprising a pooled inoculum. DNA size tags are stably integrated into the chromosome of each pool member, permitting multiplex PCR tag amplification, followed by fractionation on high resolution polyacrylamide gels and detection by silver staining. Pools of mutant pathogens are evaluated for growth attenuation by comparing the input pool composition to the recovery pool following in vivo passage. This size-marker identification technology (SMIT) is amenable to genetically intractable organisms for which marked mutagenesis methods, such as transposons, are not available. We applied the SMIT methodology to identify S. aureus virulence genes required for establishing infections in a murine host. A collection of > 10,000 S. aureus mutants was generated and screened in either a murine abscess or a murine systemic infection model for loss of virulence. Several different mutagens were employed, including transposons Tn551 and Tn917lac, and diethyl sulfate (DES). Over five hundred in vivo attenuated mutants were identified and isolated.

PA-19. Cell Cycle Control of Gene Expression by the CtrA Response Regulator in *Caulobacter crescentus* in Caulobacter crescentus.
A. Reisenauer, K. Quon and L. Shapiro.
Stanford University, Stanford, CA.

In its role as a global response regulator, CtrA controls the transcription of a diverse group of genes at different times in the *Caulobacter* cell cycle. To understand the differential regulation of CtrA-controlled genes, we compared the expression of the *fliQ* flagellar gene and the *ccrM* DNA methyltransferase gene. Despite their similar promoter architecture, these genes are transcribed at distinct times in the cell cycle. P*fliQ* is activated earlier than P*ccrM*. Phosphorylated CtrA (CtrA~P) bound the CtrA recognition sequence in both promoters, but had a 10- to 20-fold greater affinity for P*fliQ*. This difference in affinity correlates with temporal changes in cellular levels of CtrA. Our data indicate that differences in the affinity of CtrA~P for P*fliQ* and P*ccrM* govern, in part, the temporal expression of these genes. However accessory proteins may contribute to the specificity of CtrA~P for different promoters at different stages of the cell cycle. In cells expressing a stable CtrA derivative, *fliQ* transcription was prolonged while the precise timing of *ccrM* transcription remained unchanged. Moreover disrupting a unique inverted repeat element in the *ccrM* promoter significantly reduced promoter activity but not CtrA~P binding, suggesting that the inverted repeat plays an additional role in the regulation of *ccrM* transcription. We postulate that changes in the cellular concentration of CtrA~P and its interaction with accessory proteins influences the temporal expression of *fliQ*, *ccrM* and other key cell cycle genes and ultimately the regulation of the cell cycle.

PA-20. The *Trypanosoma cruzi* Genome Initiative: Expressed Sequence Tags and TcruziDB, an integrated database.
W. Degrave, A. Brandao and A. B. de Miranda.
DBBM - IOC Fiocruz, Rio de Janeiro, RJ, Brazil.

Trypanosoma cruzi, belonging to the order Kinetoplastidae, is the causative agent of Chagas disease, which affects around 18 million people in the Americas. Due to its particular biochemical and molecular biological characteristics, and the fact that no effective treatment is available thusfar, a *T. Cruzi* Genome Initiative has been launched, involving 15 Laboratories in 7 countries. The project is supported by WHO/TDR, amongst others, within the Parasite Genome framework. Random single pass sequencing of cDNA fragments, or EST sequencing, has been highly successful in the study of the gene content of more complex organisms, with the objective to identify new genes and targets for disease control and prevention, and to generate mapping probes. We have generated and deposited about 1000 EST sequences for the *T. Cruzi* genome project, partly from a CL Brener non-normalyzed library, partly from a normalyzed library, contributing to the total of around 6000 deposited EST's for this parasite. Only around 30% of the sequences showed similarity with Genbank and dbEST databases. In-depth analysis and functional classification is ongoing. The public domain software AceDB has been chosen as the common basis for parasite genome databases, and TcruziDB, the integrated *Trypanosoma cruzi* genome database, is available via ftp from ftp://iris.dbbm.fiocruz.br/pub/ genomedb/TcruziDB or through our WWW site at http://www.dbbm.fiocruz.br/genome/ tcruzi/tcruzi.html, where also information about the organization and progress of the genome initiative can be found. - Sponsored by WHO/TDR, CABBIOS, PRONEX/MCT and FIOCRUZ.

PA-21. Long Distance Genetic Communication among dsDNA Bacteriophages.

ROGER W. HENDRIX.
Univ. of Pittsburgh, Pittsburgh, PA.

The majority of genomes in the biosphere belong to the dsDNA bacteriophages, and we are sequencing representatives of those genomes to learn about the population structure of phages and to infer mechanisms of phage evolution.

Genomic sequence comparisons of the lambdoid group of bacteriophages confirm what has been known since the DNA heteroduplex experiments 30 years ago-that these phages are genetic mosaics with a common gene order but with multiple, often very different alleles for genes or groups of genes combined in different combinations in individual phages. This argues for extensive horizontal exchange of genetic material among the phages in this group. Similar comparisons in other groups of phages-for example, the mycobacteriophages-allow similar mosaicism and therefore horizontal exchange to be inferred for these groups as well.

Evidence of various sorts is beginning to emerge to argue that all of the dsDNA phages share common ancestry, even though their primary DNA and protein sequences have often diverged past the point of being recognizably related. The evidence includes biochemical similarities in the absence of sequence similarity, such as a translational frameshift that occurs in expression of certain tail genes, and similarities of gene organization, especially in the head and tail genes. For some pairs of phages that have no detectable sequence similarity it has been possible to find 'Rosetta stones'-phages or prophages that match both phages and provide a bridge between them. It appears, from these still fragmentary results, that the entire population of dsDNA phages is connected in a 'phylogenetic reticulum' or a 'World Wide Web of bacteriophages', and that they are in genetic communication through chains of local horizontal exchange events.

For the most part it appears that phages can be classified into families in which members share a common gene order and among which there is relatively frequent genetic exchange. The lambdoid, T-even, and T7 families of E. coli are good examples. Our recently completed sequence of E. coli phage N15 suggests that the actual population structure for dsDNA phages is actually not so simple. N15 has head and tail genes that place it smack in the
middle of the lambdoid family, but its early genes look very little like any known phage. Just as comparisons among the lambdoid phages allow inferences about how new lambdoid phages are built by gene exchange, comparison of phages like N15 with others is beginning to suggest ways in which exchange between phages of different families can give rise to new phage genome organizations.

PA-22. The *pacIR* Gene Resides Within a Potential Mega-Integron in the *Pseudomonas alcaligenes* Genome.

R. VAISVILA, R.D. MORGAN, E.A. RALEIGH.
New England Biolabs, Beverly, MA.

The genes for *PacI* and *PmeI* were obtained by methods based on the protein sequence; closely linked methyltransferase genes have not been found. This is unusual: type II restriction endonucleases are typically encoded by genes tightly linked to those for the corresponding modification methyltransferase(s). We are interested in the chromosomal context of these genes and the light it may shed on evolutionary mechanisms. DNA sequences flanking both *pmeIR* and *pacIR* revealed repeat sequences similar to those characteristic of

integrons. An integron consists of a site-specific integrase, an adjacent attachment site, and an array of drug resistance genes separated by DNA repeats related to the attachment site (Hall et al., *Mol Microbiol.* 5:1491). Recently, a mega-integron has been identified in the *Vibrio cholerae* genome (Mazel at al., *Science* 280:605). This mega-integron is chromosomally encoded, is much larger than standard integrons, and contains genes with a variety of functions separated by repeat sequences called VCR (*Vibrio cholerae* repeats).The *Pseudomonas alcaligenes* and *Pseudomonas mendocina* genomes contain repeat sequences, designated PAR and PMR, that are similar to the VCR. A 20 kb region of *Pseudomonas alcaligenes* DNA containing PAR sequences has been cloned and is being sequenced. The PAR repeats are arranged in clusters, similar to those of the *Vibrio cholerae* mega-integron gene cassette array. The PARs are a family of 70- to 75-bp sequences of imperfect dyad symmetry; the 40 examples sequenced thus far show an overall identity of about 90%. The sequence and analysis of (part of) the potential mega-integron region will be presented.

PA-23. Morphology Revisited: the Phylogeney of Shape Determining Genes.

J.L. Siefert, M. Kimmel, and G.E. Fox.
Rice Univ., Houston, TX and Univ. of Houston, Houston, TX.

Two separate strategies to combat osmotic pressure exist in extant life, an outer, stress-bearing fabric which lends rigidity and shape to the organism (prokaryotic) an internal framework of membranes and cytoskeletal components (eukaryotic). Integral to these strategies is the ability to divide. As peptidoglycan is the exclusive cell wall constituent in the Bacterial Domain, it has been proposed that its ubiquitous presence dictated certain aspects of bacterial morphology at the point of the bacteria's last common ancestor (Siefert and Fox, Microbiology, in press). Therefore, tracing genes involved in peptidoglycan synthesis and cell division of the Bacterial domain as well as genes identified as related to these proteins by previous investigations, will provide clues to cellular evolution. Using genomic information from 14 complete genomes publicly available as well as partial genomes, we have investigated the evolution of shape determining/cell division genes across the three domains. Six genes, *mre, ftsA, dnaK, ftsZ, rodA,* and *ftzW* were tracked and their presence or absence in each organism noted. Parsimony and distance algorithms to deduce phylogenetic relationships were applied to the aligned data sets of each gene. The results can be explained by a hypothesis that posits a series of gene duplications. The ATPase/peptide binding family (HSP70/DnaK), which exhibits homology to murein synthesizing genes (*mre*), is most likely the result of a gene duplication of an ancestral *dnaK* -like gene, occurring after the Bacterial/Archaeal split, hence its ubiquity yet unique presence in the Bacteria. Likewise, we argue a second gene duplication of this mreB gene, early in Bacterial evolution but subsequent to the Archaeal/Bacterial split, resulting in the cell division protein ftsA. This explains both its absence in the Archaea, which has been bothersome (Doolittle, Nature, 392:339-342, 1998), and its intimate association with a morphological system based on peptidoglycan in place in the Bacteria at this time.

PA-24. Identification and Characterization of Small RNAs in Prokaryotes.

A. ZHANG, K. M. WASSARMAN, and G. STORZ.
NICHD, NIH, Bethesda, MD.

Bacterial cells contain a number of small, stable RNAs. Among the well-characterized small RNAs in *E. coli* are the 93-nucleotide MicF RNA that inhibits *ompF* translation, the 377-nucleotide M1 RNA required for tRNA processing, and the 363-nucleotide 10Sa RNA (also called tmRNA) that encodes a tag peptide which targets proteins for degradation. We recently reported the discovery of the 109-nucleotide OxyS RNA that is induced by oxidative stress. The OxyS RNA acts as a pleiotropic regulator, leading to decreased or increased expression of multiple genes, and as an antimutator, protecting the cells from spontaneous and chemically-induced mutagenesis. There are also small RNAs, such as the 6S and Spot 42 RNAs, whose functions are yet-to-be determined. We will present our efforts to characterize the mechanisms of OxyS action, to elucidate the function of 6S RNA, and to identify additional small RNAs, particularly by cross species comparisons.

PA-25. The *Escherichia coli* K-12 Metabolic Genotype; Its Definition, Characteristics, and Capabilities.

Edwards, JS and Palsson, BO.
Univ. of California-San Diego, La Jolla, CA.

The complete genomic sequence for *Escherichia coli* K-12 is now available. Its annotation and known biochemical information led to the definition of the *E. coli* K-12 metabolic genotype containing 720 metabolic reactions operating on 436 metabolites. The resulting stoichiometric matrix and flux-balance analysis was used to determine the systems characteristics, the metabolic capabilities, and the effects of gene deletions on this metabolic genotype. The ability of the in silico *E. coli* strain to grow on glucose minimal media was not lost by the deletion of 85% of individual enzyme functions in the central metabolic pathways. This ability was due to the possible redistribution of metabolic fluxes with little

or no effect on cellular growth. Similarly, 70% of all possible double mutants, and 56% of all possible triple mutants in the central metabolic pathways retained growth potential. The central metabolic pathways thus exhibit a high degree of redundancy under growth on glucose. Comparison to metabolic behavior of existing mutants showed that the in silico strain accurately predicts growth abilities in 61 of 66 cases examined. For the experimentally characterized atp mutant, the characteristics of the in silico strain were representative of its in vivo counterpart. Thus, it is demonstrated that the synthesis of in silico metabolic genotypes from sequenced genomes is possible, and that systems analysis methods are available that successfully analyze, interpret, and predict phenotypic metabolic behavior from defined genotypes.

PA-26. Defining the Underlying Pathway Structure of Metabolic Genotypes.

C.H. SCHILLING & B.O. PALSSON.
Univ. of California, San Diego.

Small genome sequencing and annotations are leading to definitions of complete metabolic genotypes in an increasing number of organisms. Proteomics is beginning to give insights into the use of the metabolic genotype under given growth conditions. These data sets give the basis for systemically studying the genotype-phenotype relationship. In particular, the functional definition of biochemical pathways and their role in the context of the whole cell is now possible. It is first shown how the mass balance constraints that govern the function of biochemical reaction networks translates this problem into the realm of linear algebra. The functional capabilities of biochemical reaction networks, and thus the choices that cells can make, are reflected in the null space of their stoichiometric matrix, which can be spanned by a finite number of basis vectors. Beginning with the entire integrated metabolic network of the cell, we can first divide metabolism into subsystems of cellular metabolism, (i.e. amino acid metabolism, nucleotide metabolism). For each subsystem the proper reaction network is constructed based on the internal reactions which constitute the system and the matching of input and output fluxes which operate between subsystems. We present an algorithm based on the theory of convex analysis for the synthesis of a set of vectors used to span the null space of the stoichiometric matrix, where these vectors represent the underlying biochemical pathways that are fundamental to the corresponding biochemical reaction network. In other words, all possible flux distributions achievable by a defined set of biochemical reactions are represented by a non-negative linear combination of these pathways. These basis pathways thus represent the underlying pathway structure of the defined biochemical reaction network. Through this conceptual framework we will provide a general classification scheme of pathways based on systemic function as opposed to historical discovery, which will be important in defining, characterizing, and studying biochemical pathways from the rapidly growing information on cellular function. In addition the relevance of this approach to the regulation of cellular metabolic networks will be illustrated.

PA-27. Reconstruction of Ancestral Genomes.

WATANABE, H.
NCBI/NLM/NIH.

In order to get insight into species evolution with molecular evolutionary studies, we conduct genome evolutionary studies, that is, reconstruction of ancestral genomes. Using the complete genome sequence data of 14 prokaryotes, in which four archaeal species are included, the gene compositions and gene order in ancestral genomes are inferred from the results of molecular evolutionary studies. For the inference of gene composition, ortholog analyses are conducted. With the data of orthologous gene pairs identified between different genomes, it is also possible to know the conservation of gene order in the genomes and infer gene clusters in a common ancestral genome. So far, about 600 orthologous gene groups have been identified between Archaea and Bacteria. Among them about 100 genes have been found to form several gene clusters, in each of which gene order is conserved in both archaeal and bacterial genomes. Since these conserved regions are found in both archaeal and bacterial genomes, it is very likely that they were also part of the genome of the common ancestor of Archaea and Bacteria. Moreover, if we accept the Woesian tree, in which Archae and Eukarya are evolutionarily more closely related to each other than to Bacteria, it is possible to infer that the genome of the last common ancestor of all the modern species (cenancestor) had very similar structures to the gene clusters identified. Based on these results, backtrack characterizations of ancestral species would be possible. Several evolutionary tendencies found by looking at the variations in the conserved regions and the evolution of several genes in the conserved regions will be discussed.

PA-28. Use of Differential Display RT-PCR to Identify Conditionally Expressed Genes in *Lactobacillus acidophilus.*

M.J. KULLEN and T.R. KLAENHAMMER.
Southeastern Dairy Foods Research Center and Dept. Food Sci., Raleigh NC.

Lactobacillus acidophilus is an important member of the indigenous microflora of the gastrointestinal tract of man and animals. It remains a challenge to understand the

intestinal roles and activities of *L. acidophilus*. An important element in this regard is the determination of which characteristics are important for the survival and success of this organism in the gastrointestinal tract. We have selected acidity, mucin and bile as factors of interest and used differential display to asses the impact of these stimuli on gene expression in *L. acidophilus* ATCC 700396. DNase-treated RNA from the different treatments was reverse-transcribed and amplified by PCR using 10-11 bp primers designed to specifically avoid homology with high abundance structural RNA species (rRNA, tRNA). The amplicons were resolved by denaturing PAGE, silver-stained, excised and eluted from the polyacrylamide matrix, reamplified, cloned and sequenced. The DNA sequences were compared to those present in Genbank and homologies to bacterial genes were identified. Upon acid treatment, a fragment of the *atp* operon, which comprises the genes coding for the subunits of the proton-translocating ATPase, was revealed by differential display, and Northern blot analysis offered confirming evidence of the acid-inducibility of this gene. The remainder of the *atp* operon of *L. acidophilus* was cloned and is currently be sequenced and characterized. The use of differential display offers a functional assessment of expression from the genome in response to various stimuli.

PA-29. The *Pseudomonas aeruginosa* Genome Sequence and Postgenomic Studies on Environmentally Regulated Genes.

W. HUFNAGLE[1], S. COULTER[1], L. GOLTRY[1], M. LAGROU[1], L. LOCKWOOD[1], B. MANSFIELD[1], S. WADMAN[1], K. FOLGER[1], C. K. STOVER[1], R. GARBER[1], S. LORY[2], M. OLSON[3] and THE
1PathoGenesis Corp, 2U. of Washington Dept. of Microbiology and 3U. of Washington Genome Center. Seattle, Washington.

The gram-negative pathogen, *Pseudomonas aeruginosa*, is the leading cause of serious chronic pulmonary infections in Cystic Fibrosis and bronchiectasis. The complete genome sequence of *P. aeruginosa* strain PAO1 is being determined to facilitate studies aimed at understanding this pathogen's biochemistry and drug resis-tance mechanisms, with the ultimate aim of identifying new molecular targets for novel antipseudomonal therapies. Approximately 99% of the 6 megabase PAO1 genome sequence has been determined by a random shotgun cloning approach and is available for use via the world wide web (www.pseudomonas.com). The genome is currently in the closure stage, with ~150 physical gaps remaining to be closed. In addition to directed sequencing, we are sequencing the ends of a cosmid library to help orient the contigs on the physical map. Although the PAO1 genome is 50% larger and more GC-rich than any other completed gram-negative pathogen genome, a significant number of the *P. aeruginosa* genes exhibit close homology and analogous operon organization to that of *E. coli*. As has been observed with several of the other completed genomes, approximately one third of the putative *P. aeruginosa* genes are unique and non-homologous to the current Genbank database. Hybridization arrays were employed to evaluate differential gene expression under environmental conditions poten-tially mimicking the environment of the infected lung. These studies have provided leads for the identification of molecular mechanisms and functions, which may be essential for *P. aeruginosa* persistence in the infected lung.

PA-30. Short Leader Sequences in the Transcripts of the Hyperthermophile *Pyrobaculum aerophilum*.

M. M. SLUPSKA, A. G. KING, C. BAIKALOV, S. T. FITZ-GIBBON, J. H. MILLER.
Univ. California, Los Angeles.

As part of the continuing study of the hyperthermophilic archaeon *Pyrobaculum aerophilum*, 10 unrelated protein-encoding genes with strong homology to known proteins were cloned, sequenced and analyzed. As putative translation start codons we have designated the AUG or GUG codons near the 5'end of an open reading frame and close to the sequence resembling the consensus sequence for the archaeal TATA box. Our predictions were also supported by GenMark analysis (in 9 out of 10 cases). We then conducted the primer extension reactions and obtained the signal for the cDNA in 7 cases (for 3 genes we did not obtain the signal in 5 independent experiments). Surprisingly all mapped transcripts started one base before putative translation start codons. Of interest is also fact that the distance between the TATA box and the translation start codon is only 18 - 24 bases and that the more frequent start codon is GUG (in 6 out of 10 genes).

PA-31. *E. coli* Functional Genomics: Open Reading Frame Cloning and Mutagenesis.

Jeremy D. Glasner*, Craig Richmond, and Frederick R. Blattner .

Determination of the cellular roles of gene products requires a combination of strategies. We have developed approaches that allow high throughput cloning and characterization of E. coli open reading frames. Gene sequences are amplified by the polymerase chain reaction and ligated into a plasmid "archive" vector. An expression vector was constructed that allows rapid, directional, in-frame subcloning of ORFs from the archive vector. Overexpression of cloned ORFs was confirmed by SDS-PAGE following IPTG induction. To facilitate genetic analysis of E. coli genes we developed

methods for constructing conditional mutations in genes of interest. Cloned ORF sequences are used as templates for megaprimer PCR to generate amber stop codons in the middle of gene sequences. Amber mutant alleles are cloned into the suicide vector pKO3, electroporated into MG1655, and recombinants are selected that replace the wild-type copy of the ORF. To confirm gene replacement with the amber allele, target loci are PCR amplified using gene specific primers and digested with a restriction enzyme that cuts at the amber stop codon. To isolate mutants, even for essential genes, we constructed a Charon 26 lysogen of MG1655 that harbors an efficient amber suppressor tRNA (ala-2) under the control of the tetracycline repressor/promoter system. Addition of heat inactivated chlortetracycline to cultures of the lysogen induces su-tRNA expression and suppresses the amber mutations.

PA-32. 133-kb Plasmid of Rhodobacter capsulatus SB1003.

Jan Paces, Yakov Kogan, Gordon D Push, Robert Haselkorn, Sheeba Thomas, Michael Fonstein.

Rhodobacter capsulatus is a non-sulfur purple bacterium, used for studies of photosynthesis, nitrogen fixation and utilization of organic substrates. About 95% of the Rhodobacter genome sequence has been assembled in the ongoing cooperative sequencing project at the University of Chicago and at Institute of Molecular Genetics, Prague, supported by the bioinformatic effort at the Argonne Natl. Lab. Combination of end-sequencing of plasmid subclones and primer-walking has been used to determine the complete 132 960 nt long sequence of the plasmid pRCB133, which is present in the R. capsulatus SB1003. We have found 140 ORFs using different ORF-searching software products, like Genemark, Glimmer and Critica, together with the CodonUse program developed in our lab. One hundred and two (73%) ORFs were identical for at least two algorithms. Ambiguous start positions were rectified by locating of the adjacent SD-sites and by multiple alignments with homologous proteins. These ORF searches and preliminary functional analysis are shown at our WWW server (http://rhodol.uchicago.edu/rhodo). Tools allowing flexible graphical output of the results with various search capabilities are part of this site. This analysis also revealed 6 frame-shifts and one internal stop codon interrupting otherwise meaningful protein sequences, which are present in the original plasmid sequence. Functions for 89 (63%) of the ORFs have been assigned and integrated into an overview of the Rhodobacter metabolism created in WIT/EMP genome investigation environment. Eleven ORFs (8%) have strong homology to potential genes with unknown function. We have not found essential genes, which agrees with early physiological studies. However, a broad variety of metabolic genes are present in this plasmid, which is unusual for such replicons.

PA-33. Phylogenetic Analysis of the Proteins of Bacteriophage T4.

Elizabeth Thomas, Frank Zucker# and Elizabeth Kutter.
The Evergreen State College, Olympia, WA
University of Washington, Seattle, WA

In 1989 Bernstein and Bernstein published a minireview on the possible evolutionary origins of known T4 proteins. They found proteins with apparent strongest homologies to eukaryotes, bacteria, or both nearly equally, with a fairly even distribution among the three categories. The restructuring of the kingdoms of life, the determination of the complete genomic sequences for a number of organisms, and the new analytical tools make it high time to revisit this phylogenetic analysis and try to get a better understanding of the evolutionary history of T4's various components.

We have applied a variety of different comparison protocols to the entire set of T4 genes and probable protein-encoding ORFs. We find that only about 40 of T4's predicted 271 ORFs have clear homologies to anything yet in the databases (other than proteins from related phages). Most of these homologies involve enzymes that function in nucleotide and nucleic acid metabolism; there are still virtually no indications as to the origins of the phage structural proteins, or of the host-lethal proteins responsible for the transition from host to phage metabolism. The aerobic and anaerobic ribonucleotide reductases, tk, uvsX, and the exonuclease subunits 46/47 are most closely related to the enzymes of gram-negative bacteria, although the evidence suggests that they diverged well before the split between *E. coli* and *H. influenzae*. In contrast, td, frd and the topoisomerase subunits all appear to have diverged before the division between bacteria and eucaryotes, with some signature sequences specific to each.

There are some interesting homologies with viral proteins. T4's UV damage-specific DNA glycosylase-apyrimidine lyase enzyme, endonuclease V, has a striking similarity to a protein of the very large Chlorella virus PBCV-1, as reported by Furuta et al. (1997), with 41% identity. T4's pseT (5' polynucleotide kinase - 3' phosphatase) and gp63 (RNA ligase) are similar in amino acid sequence to the two halves of ORF 86 in the *Autographa californica* nucleopolyhedrovirus (AcNPV) (Durantel 1998). In the T4 genome, the two genes are separated by alc, which is responsible for blocking transcription of cytosine-containing DNA, and three small, uncharacterized ORFs. One of the few possible homologues found for alc in a BLAST/Smith Waterman search is another AcNPV ORF, ORF 2. This homology

is weak, but suggests the possible scenario that the kinase and ligase genes brought into the ancestor of T4 were already seperated and carried between them a gene which later evolved into alc, producing a product which now could interact with RNA polymerase to block the transcription of all cytosine-containing DNA.

The complex findings support earler suggestions that the T-even phages are very ancient in origin–perhaps as ancient as their hosts–and that substantial horizontal transfer has occurred over the course of their development.

PA-34. The Three Distinct Systems for Cytochrome *c* Biogenesis: Environmental Regulation and Rationale.
B.S. Goldman, K.A. Karberg, and R.G. Kranz.
Washington Univ., St. Louis, MO.

Cytochromes *c* are ubiquitous electron carrying proteins found in all three kingdoms of life. These proteins are fundamental in processes ranging from energy conversion to detoxification to apoptosis. Although the chemistry, structures, and functions of *c*-type cytochromes are well characterized, their biogenesis is only beginning to be understood at the molecular level. The hallmark of *c*-type cytochrome synthesis is the transport of heme and apocytochrome through a bilayer membrane and their subsequent covalent ligation (to two reduced cysteinyl residues of the apocytochrome at a CysXxxYyyCysHis signature motif). Our lab has used genetic and biochemical tools to study cytochrome *c* biogenesis in *Rhodobacter capsulatus* and *Escherichia coli*. We have recently reviewed the genomics and genetics of cytochrome c biogenesis in prokaryotic and eukaryotic organisms, leading to the proposal that three distinct systems have evolved in nature to assemble *c*-type cytochromes (Kranz et al, Mol. Microbiol. **29**:383-396, 1998). System I requires nine proteins specific for cytochrome *c* biogenesis and is used by many gram-negative bacteria and all plant mitochondria. System II requires four proteins specific for cytochrome *c* biogenesis and is used by all gram-positive bacteria, cyanobacteria, and chloroplasts. Both systems I and II are predicted to have proteins dedicated to heme delivery, apocytochrome ushering, and thio-reduction. The third system has evolved specifically in the mitochondria of fungi, invertebrates, and vertebrates.

Comparative genomic analysis has shown that two proteins (HelC and CclI) are found in all organisms and organelles that use system I for cytochrome *c* biogenesis. The CcsA protein is found in all organisms and organelles that use system II for cytochrome *c* biogenesis. These three proteins share regions of homology and are believed to facilitate transmembrane heme delivery in their respective systems. The complete topology of the HelC and CclI proteins (from *R. capsulatus*), and the CcsA protein (from *Mycobacterium leprae*) was experimentally determined (Goldman et al, Proc. Natl. Acad. Sci. USA **95**: 5003-5008, 1998). Key histidinyl residues and a conserved tryptophan-rich region (designated the WWD domain) are positioned at the site of the cytochrome *c* assembly for all three proteins. These histidinyl residues were shown to be essential for two of the proteins. We propose that these histidines protect heme from oxidation. These results and the specific environmental regulation of key proteins of System I will be presented.

INDEX OF SPEAKER ABSTRACTS BY AUTHOR

(INCLUDES CO-AUTHORS; PRESENTING SPEAKERS ARE PRINTED IN BOLD)

ALLET, E. SA-09
ANDERSON, B. SA-05
ARIGONI, F. SA-09
BAILEY, K.A. SA-12
BATTISTA, J.R. **SA-19**
BELOVA, G. SA-23
BENSON, A. **SA-02**
BIERY, M.C. **SA-27**
BLATTNER , F. R. SA-01, SA-32
BRETTIN, T. SA-16
BROWN, J.R. SA-31
BULLOCH, S.A. SA-14
BURLAND, V. SA-01
CASJENS, S **SA-28**
CEBULA, T.A. **SA-22**
CHARLEBOIS, R. L. **SA-29**
CHATARD, J. SA-18
CHAUSSADE, F. SA-21
CHERRY, J. SA-18
CHOVAN, L.E. SA-25
CHU, L. SA-17
CHURCH, G.M. **SA-06**
CLARK, A.J. **SA-20**
CLAYRON, R. SA-15
CONZEVOY, E. SA-04
COX, R. SA-03
CRAIG, N.L. SA-27
CURCHOD, M.-L. SA-09
CUSHING, J. SA-05
DAVIES, J.E. SA-10
DE ROSA, R. SA-08
DIPANKAR MANNA SA-11
DIRUGGIERO, J. **SA-18, SA-21**
DUNN, D. SA-18
DYCHINCO, A. SA-10
EDGERTON, M.D. SA-09
EISEN, J.A. SA-20
FISH, R. SA-09
FITZ-GIBBON, S. T. **SA-04**
FLEISCHMANN, R.D. SA-15
FONSTEIN, M. **SA-03, SA-17**
FORTERRE, P. **SA-24**
FRASER, C.M. SA-15
GAASTERLAND, T. **SA-07**
GARRETT, M.M. SA-15
GLASNER, J. SA-32
GORYSHIN, I. **SA-26**
GREGOR, J. SA-01
GWYNN, M. SA-31
HARDHAM, J. SA-16
HASELKORN, R. **SA-03, SA-17**
HESSLER, P. SA-25
HIGGINS, N.P. **SA-11**
HUANG, S-C. SA-30
HUNKAPILLAR, T. SA-05
JACK, D. SA-30
JAHN, P.S. SA-30
JAMOTTE, T. SA-09
KAIM, N. SA-10
KAWARABAYASI, Y. SA-18
KIM, J. SA-02
KIM, U. J. SA-04
KOGAN, Y. SA-03, SA-17
KOZYAVKIN, S. SA-23
KUTTER, E M **SA-05**
LABEDAN, B. **SA-08**
LAKE, J. SA-23
LE CLERC, J.E. SA-22
LI, B. SA-22
LI, W-T. SA-12
LINHER, K.D. SA-15
LIPPINCOTT, P. SA-05
LOFERER, H. **SA-09**
MAEDER, D.L. SA-18
MALEK, J.A. SA-15
MAU, R. SA-32
MAY, E. SA-31
MAYHEW, G.F. SA-01
MAZEL, D. **SA-10**
MCDONALD, L. SA-15
MCLEOD, M. SA-16
MELDRUM, E. SA-09
MICHOUX, I. SA-21
MILGRAM, A. J. SA-03
MILLER, J. H. SA-04
MUELLER, M. SA-05
MUSGRAVE, D. SA-23
MYERS, G. SA-16
NEITZEL, J. SA-05
NELSON, K.E. **SA-15**
NIETFELDT, J. SA-02
NIKOLSKY, Y. SA-17
NORRIS, S. J. SA-16
OOMMEN, A. SA-02
OVERBEEK, R. SA-17
PACES, J. SA-03
PARK, G. SA-04
PAYNE, W.L. SA-22
PEITSCH, M. SA-09
PEREIRA, S.L. SA-12
PERLER, F.B. SA-13
PERNA, N.T. **SA-01**
PHILIPPE, H. SA-24
PLUNKETT III, G. SA-01
POULOU, C. SA-08

PRASAD, R. SA-23
RALEIGH, E.A. SA-27
REBREKOV, D. SA-03
REEVE, J.N. **SA-12**
REICH, K.A. **SA-25**
RICHMOND, C. **SA-32**
ROBB, F.T. **SA-18, SA-21**
ROSE, D.J. SA-01
SAIER JR, M.H. SA-30
SANDMAN, K. SA-12
SELKOV, E. SA-17
SHAO, Y. SA-01
SIMON, M. I. SA-04
SLATKO, B.E. SA-27
SLESAREV, A. **SA-23**
SMITH, H.O. SA-15
SOARES, D.J. SA-12
SODERGREN, E. SA-16
SOUTHWORTH, M.W. **SA-13**
STACZEK, P. SA-11
STELLWAGEN, A.S. SA-27
STETTER, K. S. SA-04
STEWART, A.R. SA-15
STUMP, M. SA-18
SUTTON, G. SA-15
TALABOT, F. SA-09
THOMAS, E. SA-05
TSENG, T-T **SA-30**
UTTERBACK, T. SA-15
VENTER, J.C. SA-15
VONSTEIN, V. SA-17
WARREN, P.V. **SA-31**
WEBB, V.A. SA-10
WEINSTOCK, G.M. **SA-16**
WEISS, R. B. SA-18, SA-21
WILSON, S. SA-23
YAMAHA, M. SA-04
ZHULIN, I.B. **SA-14**
ZUCKER, F. SA-05

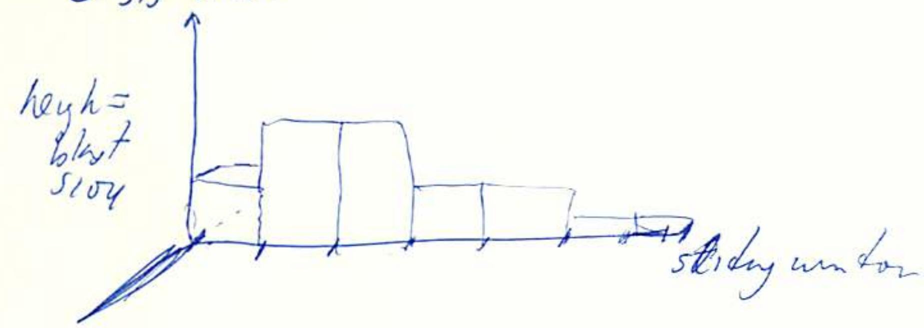
INDEX OF POSTER ABSTRACTS BY AUTHOR

(INCLUDES CO-AUTHORS)

BAIKALOV, C. PA-30
BATTISTA, J.R. PA-04
BENTON, B. PA-18
BERTANI, E. PA-12
BLATTNER, F.R. PA-31
BOND, S. PA-18
BOSCHIROLI PA-03
BOURG, G. PA-03
BRANDAO, A. PA-20
BRINKMAN, F. S. L. PA-06
BURKE, R. C. PA-18
BUYSSE, J. PA-18
CANTIN, C. PA-14
CHANE-YENE, Y. PA-14
CHEUNG, V. PA-09
CHRISLER, W.B. PA-15
CHRISTIAN, T. PA-18
COULTER, S. PA-29
COX, R. PA-09
DAVIS, C.L. PA-16, PA-17
DE MIRANDA, A.B. PA-20
DEGRAVE, W. PA-20
DEHIO, C. PA-07
EDWARDS, J.S. PA-25
FANG, H. PA-18
FIELD, D. PA-16, PA-17
FITZ-GIBBON, S. T. PA-30
FOLGER, K. PA-29
FONSTEIN, M. PA-09, PA-32
FOURNIER, D. PA-14
FOX, G.E. PA-23
GALPERIN, M.Y. PA-05
GARBER, R. PA-29
GLASNER, J.D. PA-31
GOLDMAN, B.S. PA-34
GOLTRY, L. PA-29
GRATWICK, K. PA-11
HANCOCK, R. E. W. PA-06
HASELKORN, R. PA-09, PA-32
HAY, B. PA-12
HENDRIX, R.W. PA-21
HIGGINS, N.P. PA-01
HUFNAGLE, W. PA-29
JACK, D.J. PA-13
JACK, D. PA-10
JUMAS-BILAK, E. PA-03
KARBERG, K.A. PA-34
KELLY, D. PA-10
KIMMEL, M. PA-23
KING, A. G. PA-30
KIRSCHVINK, J. PA-12
KLAENHAMMER, T.R. PA-28
KOGAN, Y. PA-09, PA-32
KOONIN, E.V PA-05
KRANZ, R.G. PA-34
KULLEN, M.J. PA-28
KUTTER, E. PA-33
LAGROU, M. PA-29
LANZ, C. PA-07
LARIMER, F. PA-08
LAU, P.C.K. PA-14
LOCKWOOD, L. PA-29
LORY, S. PA-29
LYONS, H. PA-18
MACMILLAN, A. PA-03
MANSFIELD, B. PA-29
MARKILLIE, L.M. PA-15
METZGAR, D. PA-16, PA-17
MILLER, J. H. PA-30
MORGAN, R.D. PA-22
MURAL, R. PA-08
O'CALLAGHAN, D. PA-03
OLSON, M. PA-29
PACES, J. PA-32
PALSSON, B.O. PA-25, PA-26
PATEL, A. PA-14
PUSH, G.D. PA-32
QUAKE, S. PA-12
QUON, K. PA-19
RABUS, R. PA-10
RALEIGH, E.A. PA-22
RAMUZ, M. PA-03
REISENAUER, A. PA-19
RICHMOND, C. PA-31
SAIER JR, M.H. PA-10, PA-11, PA-13
SCHILLING, C.H. PA-26
SEUBERT, A. PA-07
SHAH, M. PA-08
SHANK, N.C. PA-04
SHAPIRO, L. PA-19
SIEFERT, J.L. PA-23
SLUPSKA, M. M. PA-30
SMITH, D.W. PA-13
STACZEK, P.W. PA-01
STORZ, G. PA-24
STOVER, C. K. PA-29
SUBRAMANIAN, A. PA-08
TATUSOV, R.L. PA-05
TAYLOR, B.L. PA-02
THOMAS, S. PA-32
THOMAS, E. PA-16, PA-33
TSENG, T.-T. PA-11
UBERBACHER., E. PA-08
VAISVILA, R. PA-22
WADMAN, S. PA-29
WASSARMAN, K. M. PA-24

WATANABE, H. PA-27
WILLS, C. PA-16, PA-17
WINTERBERG, K. M. PA-18
WONG, K.K. PA-15
YOUNG, G.B. PA-13
ZHANG, A. PA-24
ZHANG, J.P. PA-18
ZHULIN, I.B. PA-02
ZUCKER, F. PA-33

① 3D blast



② blast to tree

① blast or PSI
② pull out high hits

Phyloblast → Distance matrix

align (SAPS?) → Distance matrix

Distance matrix → tree

Talk

① Trees are ONLY way to determine orthology
② TGaasterland -- Aquifex
③ apologize for broad repair stuff
④ focus on methods
⑤ web site

- why phylogenomics
- MutH -- refer to L. Raley
- Moxon SSM in H.pylori
- Igor - novel chemotaxis pathway in Aquifex

AMERICAN
SOCIETY FOR
MICROBIOLOGY

1325 Massachusetts Avenue, NW
Washington, DC 20005-4171

JIM LAKE

Some informational genes are resistant to lateral transfer

Some operational genes can be transferred

CHIMERAS

SOGIN

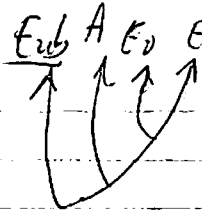
ZILLIG

GUPTA

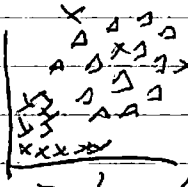
ROGERLITTLE

WFDOOLITTLE

BOOTDING



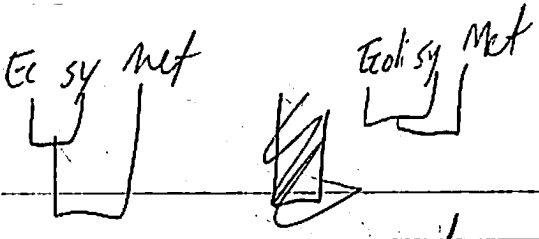
Δ = UNIQ ORFs
x = 10 ORFs



Dist from Meth. to Syn

Distance from Methan → Yeast

Informational genes = all euk. ones have ARCH ancestry



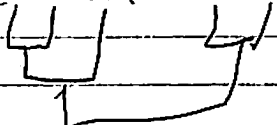
Information

Operational

Why long branch to Methanogens

Rooting the tree

- find paralogs
- regard them to be orthologs



Information rooted like this

Operational basically possible roots.

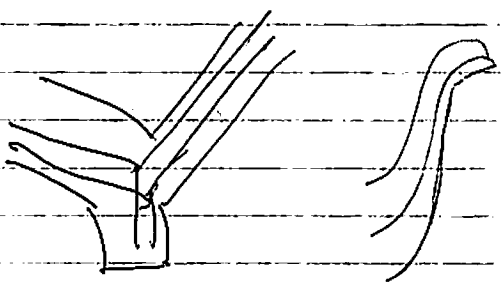
Does he believe all
these trus are
accurate

Veronika Vorstein

Zymomonas
Thiobacillus ferrooxidans

3 Dimension of ORF annotation
① in phylog. space
② in cellular metabolism
③ in the chromosome

6
102



Fusobacterium

E A Raleigh - Genomic carpetbaggers

Restriction Systems

- type II R+M systems evolve independently

Carpetbaggers
outsider
parasitic
bring new skills
generate new ideas

Outsiders
- Horizontal transfer
- Incongruency of trees

- absence in some species is NOT evidence for lateral transfer

- conservation of DNA sequence used as evidence for transfer

Paras. tes

- one can some genes - ✓ cannot lose them

modification dependent restriction
- methylation leads to cutting

NEW SKILLS

Opens its defenses

Restriction + Homol. Recombination

- reassortment of alleles
- Sugg. intergenic recomb is rare
- Sugs. intragenic reco. is high

[illegible]
[illegible]

[illegible]
[illegible]

[illegible]
[illegible]
[illegible]
[illegible]

J Reeve - Nucleosomes

Genome structure in M. thermoautotrophicum

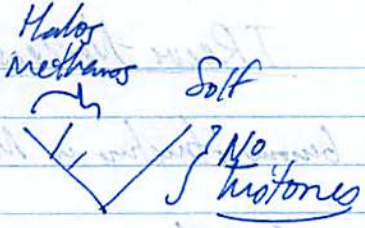
M. fervidus
- 33% GC
- 83°C growth 7°

- purified DNA binding protein
- added back to -- got regular structure
- what does this mean?

HMf - is a dimer
- it is not histone-like

Archaea
- histone fold = 3 helix fold
- always dimers

Archaea



23 Histone sequences from Archaea
- one face of helix very highly conserved -- other face NOT

- seq of regions conserved in Archaeas are also conserved in Euks

Gel acceleration assays

Can boil 5 hrs

What about nucleosome altering protein

Nuclear Ø

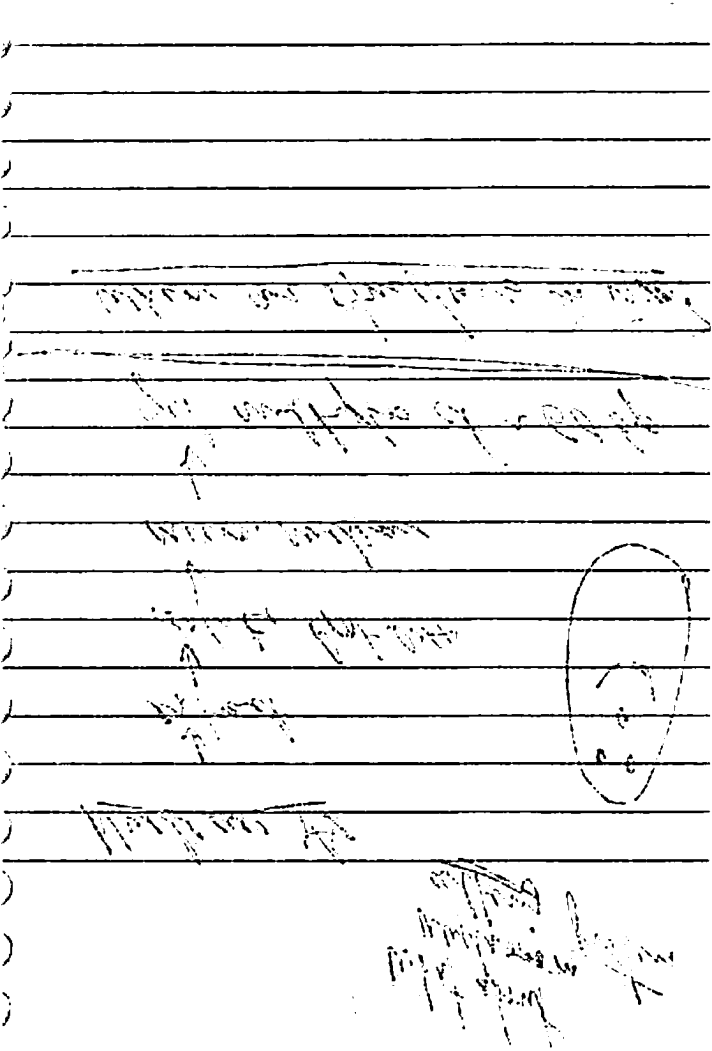
↓ extract

↓ isolate Ppt and

↓ micro. nuclease

↓ get multiples of ~60 bp

Where are structures in vivo?

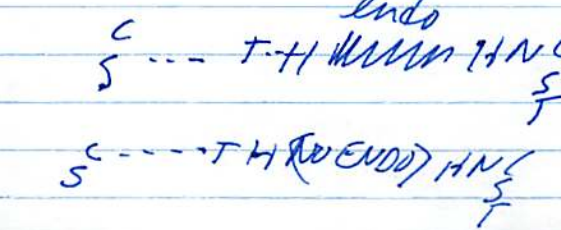


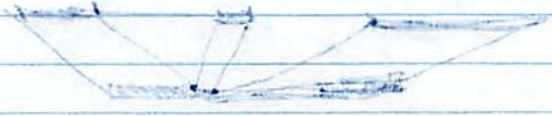
Protein Splicing

Vent Pol Gene - T. litoralis
- ORF too large

Purifying precursor

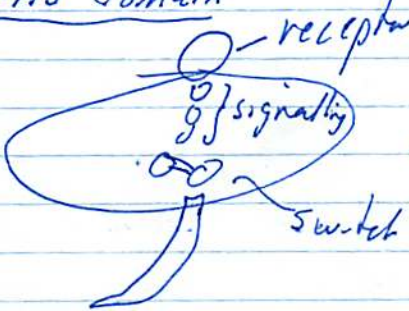
- intermediate at high no/wt
-





Igor Zhulin Chemotaxis in Coli

PAS domain



Genomes

Ecoli - some cases all genes coded w/in one operon

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Craig Richmond

Genome wide expression

- Gene expression profiling using radioactive spot blots

~~- Gene chips~~

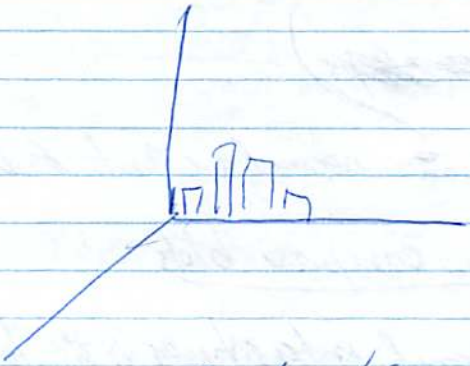
Don't normalize amts of DNA

How compare blots

- signal intensity as % of total
- or use yeast spots

Do they have sets of genes w/ correlated expression in diff conditions

Genome map w/ gene expression info



Cross-hybridization of EBGA w/ LACZ

Affy chip

- 15 oligos/gene
- chosen by computer
- 300,000 oligos (25mers)

I'll blow one up later...
Or I'd like to blow one up

- not enough signal yet
- 90% of RNA = rRNA
- how enrich for mRNA

Label Maltose/trehalose

Lateral transfer of
Thermococcus litoralis
+
pyrococcus furiosus

Frank Robb

Pyrococcus furiosus - med. sea
Thermococcus litoralis
Pyrococcus horikoshii - deep sea vent

SNYDER

BAR	TAM
BUS	NY
CHI	AN
KC	UA
CL	TEX
MIN	SEA

[illegible]

[illegible] [illegible]

[illegible] [illegible]

[illegible] [illegible]

Miller

growing mutators for many generations

Steps of selection pressure leads to mutators taking over population

Batista - D. radiodurans

- many mesophiles
- some thermophiles
- etc.

$D_{37} = 6500 gy$

Joe Dirruggiero

GC content NOT correlated w/ hyperthermophily

How protect DNA
- high intracellular salt
- histone + histone-like proteins
- topology

RAD A, RAD B, not damage inducible

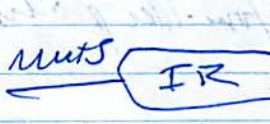
TOM CEBULA

most of mut- are deletions w/ flanking repeats

Why have hypermutable phenotype?

MMR⁻ = promiscuous

MutM:



great variability in length of this region

MutS }
MutL }
MutH - does not regulate recombination

Non-repair activities

1. RecA = tx req
2. PS3?
3. Ad q
4. PHR
5. TFII H

Some genes in nuclear genes have lost gene order

[illegible]

(١) [illegible]

(٢) [illegible]

(٣) [illegible]

(٤) [illegible]

(٥) [illegible]

[illegible]

[illegible]

[illegible]

R. Charlebois - Univ. of Ottawa

NCBI

ANALYSIS UPDATED NIGHTLY

Types of Analysis

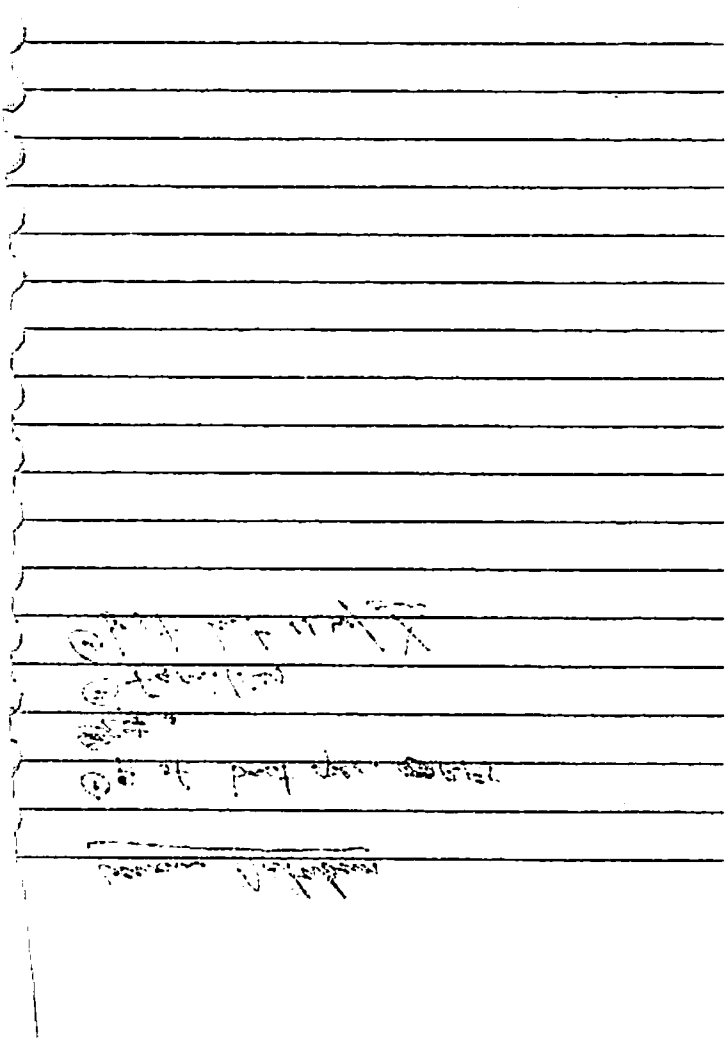
NOTE

ANALYSIS WEIGHTED AVERAGES

Types of Analysis

Genome Databases

① # of bact spec. ORFS
② #
③ families
④ best hit is w/ X



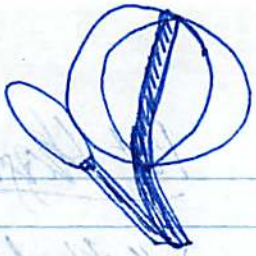
Prof Hagop Warren

Smith-Kline Beech
- ~75 people in bioinformatics

Criteria for selection
Spectrum + specificity
Essential + expressed in vivo
Cellular location
Novelty
High throughput selection

Collect sequence
Blast2P
Located results into relational tables
Superfamilies
Align + tree

Pol III Complex



POLC - grant
DNAE - grant
DNAE - grant

DNAQ

POLC-DNAQ motif

- RFC2 - in arks/archs
similar to DNA Z

Major facilitator superfamily (MFS)

e.g. ~~lac~~ LacY

RND superfamily

www-biology.ucsd.edu/

~msaier/transport/ht tpgyc.html

~ipaulsen/transport/ht tpgyc.html

Didier Maizel

① Intro to repair.

Evolutionary Mol. Bio?

① Intro
② Evol mol bio
③ DNA repair
④ Diff in repair
⑤ Comp genomics
⑥ Two stages of uses of evolution
⑦ SNP2
⑧ Muts
⑨ All genomes -- quick tables
⑩ How make sense --

PG(fix) prediction
EXAMPLES
- SNP2
- Muts

① why and working
② examples - theory
③ examples - real

12. Organic - some examples

~~13.~~

13. Fuel

14. Problems

15. Halophiles

Intro = 10	4	although actually
Methods = 10	14	about specific
SNF2 = 4	28	what I am
Tests = 10	28+10=38	really interested
Repair 15	53	in is apply and
HV = 10	63	methods