6th Annual Sequencing, Finishing, Analysis in the Future Meeting

Santa Fe, New Mexico

June 1st – 3rd, 2011

Los Alamos NATIONAL LABORATORY EST. 1943

JGI

Office of Science U.S. DEPARTMENT OF ENERGY

Contents

The 2011 "Sequencing, Finishing and Analysis in the Future" Organizing Committee:

** Chris Detter, Ph.D., JGI- LANL, Genomics Center Director, LANL*
** Johar Ali, Ph.D., Cancer Genomics Team Leader, OICR*
** Patrick Chain, Metagenomics Team Leader, LANL*
** Michael Fitzgerald, Finishing Manager, Broad Institute*
** Bob Fulton, M.S., Sequence Improvement Group Leader, WashU*
** Darren Grafham, Sequence, Analysis, Genome Enhancement Coordinator, Sanger Institute*
** Jessica Hostetler, Genome Finishing and Analysis Manager, JCVI*
** Alla Lapidus, Ph.D., Director Bioinformatics, IPM, FCCC*
** Donna Muzny, M.S., Director of Operations, BCM*

Next-Generation Sequencing

From adaptors for shotgun sequencing to controls, IDT can provide the products you need for your next-generation sequencing projects.

Adaptors for Shotgun Sequencing

- Y adaptors with terminal modifications for nuclease protection are also available.
- Adaptors for shotgun sequencing can be ordered with barcodes or molecular ID tags to aid in multiplexing samples.

Fusion Primers for Amplicon Sequencing

- Fusion Primers for amplicon sequencing, with the 3′ segment of the primer complementary to the target of interest and the 5′ segment complementary to oligos coupled to beads or substrates.
- Fusion Primers can be ordered with barcodes or molecular ID tags to aid in multiplexing samples.

Sequencing Controls

- Ultramer™ Oligos are single-stranded DNA molecules up to 200 bases that can be attached to substrates or beads for sequencing controls.
- Double-stranded DNA controls in plasmids are available through IDT's gene-synthesis platform.

THE CUSTOM BIOLOGY COMPANY

WWW.IDTDNA.COM

IDT®

INTEGRATED DNA TECHNOLOGIES

06/01/2011 - Wednesday				
Time	**Type**	**Abstract #**	**Title**	**Speaker**
7:30 - 8:30am	Breakfast	x	La Fonda Breakfast Buffet	Sponsored by NEB
8:30 - 8:45	Intro	x	Welcome Intro from Los Alamos National Laboratory	Nan Sauer
x	Session Chair	x	Session Chairs	Chair - Johar Ali Chair - Mike Fitzgerald
8:45 - 9:30	Keynote	FF0048	Challenges in Cancer Genome Sequencing	Dr. John McPherson
9:30 – 9:50	Speaker 1	FF0115	Using NGS in Cancer Research: Bioinformatics Challenges	Yuriy Fofanov
9:50 - 10:10	Speaker 2	FF0010	LifeTech - Use Of Next Generation Sequencing Technology for Targeted Resequencing	Michael Rhodes
10:10 – 10:30	Speaker 3	FF0034	454 - Comprehensive Transcriptome Sequencing with the Genome Sequencer FLX System	Clotilde Teiling
10:30 – 10:50	Break	x	Beverages and snacks provided	Sponsored by Covaris
10:50 – 11:10	Speaker 4	FF0058	illumina - Improvements in Next Generation Sequencing	Haley Fiske
11:10 – 11:30	Speaker 5	FF0122	Ion Torrent - Semiconductor Sequencing for Life	Andy Felton
11:30 – 11:50	Speaker 6	FF0068	Pacific Biosciences - Tongue Twisters and Picture Puzzles: The Advantages of Single Molecule Real Time Sequencing in Difficult Assembly Problems	Steve Turner
11:50 - 12:40	Panel Discussion	x	Next Generation Sequencing Technology Panel Discussion	Chair - Bob Fulton
12:40 – 2:00pm	Lunch	x	Coronado Lunch Buffet	Sponsored by illumina
x	Session Chair	x	Session Chairs	Chair - Darren Grafham Chair - Patrick Chain
2:00 – 2:20	Speaker 7	FF0035	Who are your customers? Using genomic data in education	Sandra Porter
2:20 – 2:40	Speaker 8	FF0025	Tephritid Fruit Fly Genomics at the Pacific Basin Agricultural Research Center	Scott Geib
2:40 – 3:00	Speaker 9	FF0021	Optical Mapping and Whole Genome Finishing Elucidate Unusual Chromosomal Architecture in Some Vibrio cholerae Isolates	Kim Bishop-Lilly
3:00 – 3:20	Speaker 10	FF0138	Comparative Gene Expression of the Caldicellulosiruptor genus using RNAseq	Loren Hauser
3:20 – 3:40	Break	x	Beverages and snacks provided	Sponsored by Covaris
3:40 - 5:20pm	Genome Center Updates (10 min each) with question panel	FF0155	Genome Sequencing Center at NCGR	Jimmy Woodward
		FF0151	Genome Sequencing Center at DOE JGI	Feng Chen
		FF0080	Genome Sequencing Center at Sanger	Darren Grafham
		FF0112	Genome Sequencing Center at Broad	Mike Fitzgerald
		FF0110	Genome Sequencing Center at Baylor	Donna Muzny
		FF0123	Genome Sequencing Center at WashU	Bob Fulton
		FF0103	Genome Sequencing Center at JCVI	Nadia Fedorova
		FF0201	Genome Sequencing Center at OICR	Johar Ali
5:30 – 7:00pm	Posters - even #s Meet & Greet Party	EVEN #s	Poster Session with Meet & Greet Party (Sponsorted by Roche) Food & Drinks	Sponsored by Roche 5:30pm-9:00pm
7:00 - 8:30pm	Posters - Odd #s Meet & Greet Party	ODD #s	Poster Session with Meet & Greet Party (Sponsorted by Roche) Food & Drinks	Sponsored by Roche 5:30pm-9:00pm

06/02/2011 - Thursday				
Time	**Type**	**Abstract #**	**Title**	**Speaker**
7:30 - 8:30am	Breakfast	x	Santa Fe Breakfast Buffet	Sponsored by NEB
8:30 - 8:45	Intro	X	Welcome Back from DOE	Dan Drell
x	Session Chair	x	Session Chairs	Chair - Alla Lapidus Chair - Mike Fitzgerald
8:45 - 9:30	Keynote	FF0107	Science of Sequencing Process Development for High Technical Replicate R2 Analyses	Dr. Nels Olson
9:30 - 9:50	Speaker 1	FF0146	Long Reads and Hybrid Assemblies	Jim Knight
9:50 -10:10	Speaker 2	FF0056	Consed and Phaster for Next-gen Sequencing	David Gordon
10:10 - 10:40	Break	x	Beverages and snacks provided	Sponsored by Isilon
10:40 -11:00	Speaker 3	FF0092	Of Parrots and Pathogens – Hybrid Assembly and Benchmarking of Long and Short Read Sequencing	Adam Phillippy
11:00 - 11:20	Speaker 4	FF0214	BioPig: Hadoop-based Analytic Toolkit for Next-Generation Sequence Data	Zhong Wang
11:20 - 11:40	Speaker 5	FF0142	Automated High-Throughput de novo Genome Assembly of Microbial Genomes Using Illumina Data	Bruce Walker
11:40 - 12:00	Speaker 6	FF0072	Efficient Graph Based Assembly of Short-Read Sequences on a Hybrid-Core Architecture	George Vacek
12:00 - 1:20pm	Lunch	x	New Mexican Lunch Buffet	Sponsored by PacBio
x	Session Chair	x	PacBio Assembly Workshop Session	Chair - Steve Turner
1:20 - 3:10pm	Hybrid De Novo Assembly Workshop (15 min each) with discussion panel	FF0204	Using AHA (A Hybrid Assembler) for genome finishing of V. cholerae	Aaron Klammer
		FF0200	Leveraging long single molecule PacBio reads for de novo genome assembly	Todd Michael
		FF0141	The Long and Short of Microbial Hybrid Assembly Generation	Aaron Berlin
		FF0197	Initial Results and Future Direction in Hybrid Assembly with Illumina and Pacific Biosciences Data	Richard McCombie
		FF0216	Using Pacbio for Sample Characterization under Real-time Conditions	Patrick Chain
x	Session Chair	x	Session Chairs	Chair - Donna Muzny Chair - Johar Ali
3:10 - 3:30	Break	x	Beverages and snacks provided	Sponsored by CLC
3:30 - 5:30pm	Tech Time Talks (15 min each)	FF0040	Creating Probe Maps with Solid-State Nanopores	John Oliver
		FF0182	Avadis NGS Analysis Software	Jean Jasinski
		FF0205	High-Speed, High-Reliability Focus Optimization for Genomics Applications	Scott Jordan
		FF0206	Novel Improvements to the Illumina TruSeq Indexed Library Construction, Amplification and Quantification Protocols for Optimized Multiplexed Sequencing	Eric van der Walt
		FF0177	Accelerating the Impact of Targeted Resequencing with SureSelectXT	Scott Happe
		FF0186	A New Method for Long Range Scaffolding of Large Complex Genomes using the Argus™ Optical Mapping System	Nick Xiao
		FF0231	Isilon Delivers Cost-Effective, Scalable Data Storage in Support of TGen's Biomedical Research Initiatives	Chris Blessington
		FF0077	Advancements in Focused Acoustics for Use in NGS Sample Preparation	Hamid Khoja
5:45 - 7:45	Happy Hour	x	Happy Hour at Cowgirls Cafe - Sponsored by LifeTech - Map Will be Provided	Sponsored by LifeTech
7:45 - bedtime	on your own	x	Dinner and night on your own - enjoy	x

06/03/2011 - Friday				
Time	**Type**	**Abstract #**	**Title**	**Speaker**
7:30 - 8:30am	Breakfast	x	Breakfast Buffet	Sponsored by NEB
8:30 - 8:45	Intro	x	Welcome Back	Chris Detter
x	Session Chair	x	Session Chairs	Chair - Patrick Chain Chair - Darren Grafham
8:45 - 9:30	Keynote	FF0032	The Study of the Human Microbiome	Dr. Granger Sutton
9:30 – 9:50	Speaker 1	FF0165	Dining on Driftwood: a genomic view of a wood-eating bacterial endosymbiosis in the shipworm Bankia setacea	Dan Distel
9:50 – 10:10	Speaker 2	FF0101	High throughput single cell genomics pipeline for microbiology	Ramunas Stepanauskas
10:10 – 10:30	Break	x	Beverages and snacks provided	Sponsored by OpGen
10:30 – 10:50	Speaker 3	FF0131	Manipulating bacterial growth for single cell genomics	Armand Dichosa
10:50 – 11:10	Speaker 4	FF0050	Preparation of Nucleic Acid Libraries for Next Generation Sequencing with an Automated Molecular Biology Platform	Ken Patel
11:10 – 11:30	Speaker 5	FF0116	Application of deep sequencing to diversity library analysis and selection	Andrew Bradbury
11:30 – 11:50	Speaker 6	FF0031	What's a referenceable reference?	Todd Smith
11:50 – 12:10	Speaker 7	FF0156	Finishing in the era of NGS: A user's perspective	Kostas Mavrommatis
12:10 - 12:30	Closing Discussions	x	Closing Discussions - discuss next year's meeting	Chair - Chris Detter
12:30 - 2:00pm	Lunch & Close of meeting	x	La Fiesta Plaza Lunch	Sponsored by Agilent

06/01/2011 - Wednesday				
Time	**Type**	**Abstract #**	Title	**Speaker**
7:30 - 8:30am	**Breakfast**	**x**	**La Fonda Breakfast Buffet**	**Sponsored by NEB**
8:30 - 8:45	Intro	**x**	Welcome Intro from Los Alamos National Laboratory	**Nan Sauer**
x	Session Chair	**x**	Session Chairs	Chair - Johar Ali Chair - Mike Fitzgerald
8:45 - 9:30	**Keynote**	**FF0048**	**Challenges in Cancer Genome Sequencing**	**Dr. John McPherson**
9:30 – 9:50	Speaker 1	**FF0115**	Using NGS in Cancer Research: Bioinformatics Challenges	**Yuriy Fofanov**
9:50 - 10:10	Speaker 2	**FF0010**	LifeTech - Use Of Next Generation Sequencing Technology for Targeted Resequencing	**Michael Rhodes**
10:10 – 10:30	Speaker 3	**FF0034**	454 - Comprehensive Transcriptome Sequencing with the Genome Sequencer FLX System	**Clotilde Teiling**
10:30 – 10:50	**Break**	**x**	**Beverages and snacks provided**	**Sponsored by Covaris**
10:50 – 11:10	Speaker 4	**FF0058**	illumina - Improvements in Next Generation Sequencing	**Haley Fiske**
11:10 – 11:30	Speaker 5	**FF0122**	Ion Torrent - Semiconductor Sequencing for Life	**Andy Felton**
11:30 – 11:50	Speaker 6	**FF0068**	Pacific Biosciences - Tongue Twisters and Picture Puzzles: The Advantages of Single Molecule Real Time Sequencing in Difficult Assembly Problems	**Steve Turner**
11:50 - 12:40	**Panel Discussion**	**x**	**Next Generation Sequencing Technology Panel Discussion**	**Chair - Bob Fulton**
12:40 – 2:00pm	**Lunch**	**x**	**Coronado Lunch Buffet**	**Sponsored by illumina**
x	Session Chair	**x**	Session Chairs	Chair - Darren Grafham Chair - Patrick Chain
2:00 – 2:20	Speaker 7	**FF0035**	Who are your customers? Using genomic data in education	**Sandra Porter**
2:20 – 2:40	Speaker 8	**FF0025**	Tephritid Fruit Fly Genomics at the Pacific Basin Agricultural Research Center	**Scott Geib**
2:40 – 3:00	Speaker 9	**FF0021**	Optical Mapping and Whole Genome Finishing Elucidate Unusual Chromosomal Architecture in Some Vibrio cholerae Isolates	**Kim Bishop-Lilly**
3:00 – 3:20	Speaker 10	**FF0138**	Comparative Gene Expression of the Caldicellulosiruptor genus using RNAseq	**Loren Hauser**
3:20 – 3:40	**Break**	**x**	**Beverages and snacks provided**	**Sponsored by Covaris**
3:40 - 5:20pm	**Genome Center Updates (10 min each) with question panel**	**FF0155**	**Genome Sequencing Center at NCGR**	**Jimmy Woodward**
		FF0151	**Genome Sequencing Center at DOE JGI**	**Feng Chen**
		FF0080	**Genome Sequencing Center at Sanger**	**Darren Grafham**
		FF0112	**Genome Sequencing Center at Broad**	**Mike Fitzgerald**
		FF0110	**Genome Sequencing Center at Baylor**	**Donna Muzny**
		FF0123	**Genome Sequencing Center at WashU**	**Bob Fulton**
		FF0103	**Genome Sequencing Center at JCVI**	**Nadia Fedorova**
		FF0201	**Genome Sequencing Center at OICR**	**Johar Ali**
5:30 – 7:00pm	**Posters - even #s Meet & Greet Party**	**EVEN #s**	**Poster Session with Meet & Greet Party (Sponsorted by Roche)** **Food & Drinks**	**Sponsored by Roche 5:30pm- 9:00pm**
7:00 - 8:30pm	**Posters - Odd #s Meet & Greet Party**	**ODD #s**	**Poster Session with Meet & Greet Party (Sponsorted by Roche)** **Food & Drinks**	**Sponsored by Roche 5:30pm- 9:00pm**

NOTES

Speaker Presentations (June 1st)

Abstracts are in order of presentation according to Agenda

FF0048

Keynote

John McPherson

The Ontario Institute for Cancer Research (OICR), Toronto, Ontario M5G 0A3, Canada

Challenges in Cancer Genome Sequencing

NOTES

FF0115

Using NGS in Cancer Research: Bioinformatics Challenges

Yuriy Fofanov

Center for Biomedical and Environmental Genomics, University of Houston, Houston, TX, USA

Over last 4 years UH CBMEG have been involved in variety of projects focused on the analysis of cancer genomics data generated by four major of sequencing platforms (454, Illumina, SOLiD, and Ion Torrent). This talk will overview bioinformatics solutions and challenges associated with using NGS data for: (a) copy number variation; and (b) large scale methylation pattern analysis; as well as (c) detection of heteroplasmy and rare sequence variants in mitochondria; and (d) identification of non-host associated genomic material in tumor tissues.

FF0010

Use of Next Generation Sequencing Technology for Targeted Resequencing

Dr. Michael Rhodes

Sr. Manager Sequencing Portfolio, Life technologies, 5791 Van Allen Way, Carlsbad, CA 92008, USA

The advent of Next Generation Sequencing has led rapid advances in sequencing throughput with output of sequencing runs described in 100s of gigabases compared to the previous generation of production capillary sequences where it was measured in kilobases. As a consequence of this a single researcher now has the ability to sequence two human genomes a week using the SOLiD™ 5500xl. This has opened up a wide range of whole genome based sequencing applications. Results from previous studies including GWAS and linkage studies, lead many researchers to want to sequence much smaller segments of the genome on many samples. In order to do this a number of approaches to targeted resequencing have been developed, based on hybridization pullout or PCR amplification of the desired regions. One of the biggest challenges with targeted resequencing has been automating the “front end” of the process. In this presentation, the latest improvements in automating the workflow, especially with regards to automation of the library construction process and automation of analysis pipelines will be covered as will improvements to the sequencing system (accuracy, usability and flexibility).

The power of targeted resequencing will be highlighted by reference to some of the latest published data, with special attention to exome resequencing (for example Haack et al 2010, Hoeschien et al 2010, Krawitz et al 2010). A slew of recent publications have demonstrated that the causative mutations for individuals can be identified by resequencing the exome, demonstrating the ability to use a next generation sequencing approach for population based analysis or for a personalized analysis of disease mutations.

FF0034

Comprehensive Transcriptome Sequencing with the Genome Sequencer FLX System

Clotilde Teiling

Roche Applied Science, 9115 Hauge Road, Indianapolis, IN 46250, USA

The Genome Sequencer FLX System offers a powerful combination of long sequencing reads (400 to 500 base pairs) and dedicated GS Assembler software, enabling the sequencing, assembly, and analysis of messenger RNAs (mRNA) that represent a comprehensive transcriptome of an organism.

Experimental designs can be developed to explore and detect disease-causing mechanisms

FF0058

Improvements in Next Generation Sequencing

Haley Fiske

Illumina, Inc., 25861 Industrial Blvd, Hayward, CA, 94545, USA

The primary focus in next gene sequencing improvements has been towards higher and higher throughput systems. With the introduction of the MiSeq the use of sequencing-by-synthesis is now available for smaller scale projects such as targeted resequencing and small genomes. MiSeq applications and data will be discussed as well as improvements to existing methods using next generation sequencing platforms.

FF0122

Semiconductor Sequencing for Life

Andy Felton, Jonathan M. Rothburg, Jason Myers, and Ion Torrent Team
Ion Torrent, 7000 Shoreline Court, Suite 201 South San Francisco, CA 94080, USA

Ion Torrent has invented the first device—a new semiconductor chip—capable of directly translating chemical signals into digital information. The first application of this technology is sequencing DNA. The device leverages decades of semiconductor technology advances, and in just a few years has brought the entire design, fabrication and supply chain infrastructure of that industry—a trillion dollar investment—to bear on the challenge of sequencing. The result is Ion semiconductor sequencing, the first commercial sequencing technology that does not use light, and as a result delivers unprecedented speed, scalability and low cost.

All of these benefits are a result of applying a technology that is massively scalable, as proven by Moore's Law, to a task that has traditionally used optics-based solutions, which work in a linear fashion: increasing capacity requires increasing the number of signals that must be read resulting in longer run times, higher capital costs and ever more sophisticated optics. By contrast, Ion Torrent semiconductor technology can provide increases in chip capacity without impacting capital costs or runtime. Ion Torrent sequencing uses only natural (label-free) reagents and takes place in Ion semiconductor microchips that contain sensors which have been fabricated as individual electronic detectors, allowing one sequence read per sensor. We will show how the technology has scaled in just a few months from ~1 million sensors in the first-generation Ion 314 chips to ~7 million sensors in the second-generation Ion 316 chips—all while maintaining the same 1- to 2-hour runtime. We will also demonstrate that Ion semiconductor sequencing provides exceptional accuracy, long read length and scalability on a single, affordable bench-top sequencing platform.

FF0068

Tongue Twisters and Picture Puzzles: The Advantages of Single Molecule Real Time Sequencing in Difficult Assembly Problems

Stephen Turner

Pacific Biosciences, 1380 Willow Rd, Menlo Park, CA 94025, USA

The recent dramatic advances in sequencing technology have propelled attempts to sequence novel genomes at a pace nearly unimaginable five years ago. Yet while raw sequencing costs have plummeted, the cost of truly finishing a genome sequence remains high and genome finishing has proven to be less susceptible to the advances enabled by second-generation DNA sequencing. The third-generation single molecule real time sequencing approach from Pacific Biosciences provides several key advantages that enable new progress in rapidly and affordably finishing novel genomes. The sample preparation process for our platform exhibits less sequence-composition bias than other platforms, and we demonstrate here low coverage bias across a wide range of G+C%. We are also able to sequence through traditionally difficult templates, and we provide several examples of this aspect in the context of cDNA sequencing and triplet repeat expansions. SMRT™ sequencing produces much longer reads than second-generation technologies, with the top 5% of reads longer than 3,000bp, and we routinely employ this capability to span long repeats and difficult gaps in unfinished genomes. We have enabled this hybrid assembly approach of combining long-read third-generation sequence data with high depth-of-coverage second-generation sequence data by providing an algorithms pipeline for scaffolding and filling in assemblies of microbial genomes. Several examples are provided including our recent work to produce a 99.9% finished assembly of the Haitian strain of *V. cholerae*. The reduction of these finishing steps to algorithmic pipelines means that we can now approach true algorithmic autofinishing with fewer library constructions and sequencing time. Finally, the low bias in SMRT sequencing and its ability to span difficult repetitive regions are essential attributes for charting the remaining dark matter in the human genome and other poorly finished genomes, such as many plant and livestock genomes.

NOTES

NOTES

Lunch

12:40 – 2:00pm

Sponsored by

NOTES

FF0035

Who Are Your Customers? Using Genomic Data in Education

Sandra Porter[1], Linnea Fletcher[2]

[1]Digital World Biology, Seattle, WA 98107, USA; [2]Austin Community College, Austin, TX 78701, USA

Over the past ten years, we have introduced college instructors across the county to data resources and tools in the public repositories. We anticipated at the beginning that instructors would come up with creative ways to use these data to engage greater numbers of students in undergraduate research. Instead, what we learned is that while biology instructors are impressed with the amount of data available, they're also overwhelmed and unsure what to do with it. From our perspective, there's no shortage of data to work with or tools for analysis; the limiting step is connecting the people with project ideas to the students who could do the work.

Over the past few years, projects such as GEBA (Genome Encyclopedia of Bacterial and Archea) and CACAO (Community Assessment of Community Annotation with Ontologies) have emerged to engage the community in genome annotation. These projects are notable for allowing students to participate as researchers in the genomics community. The downside to these projects is that they present a narrow view of the kinds of projects that people undertake, and do not adequately prepare students for careers in the life-science workforce.

We are proposing to use this wealth of public data in a different way. We are developing a course where students can engage in data-initiated investigations on diverse "back-burner" problems solicited from companies and research institutions. For example, we know a researcher who recently obtained a set of microarray data. We would have students identify statistically significant differences in gene expression, then find and summarize data from overrepresented ontologies and pathways. Not only would the researcher benefit from diverse sets of eyes, the students would benefit by working on a real-world problem.

Our model also includes a plan for professional development. Since we anticipate working on multiple types of problems, we are working to build an interdisciplinary team of faculty mentors, based at multiple colleges across the U.S. This faculty team will gain new insights into the kinds of bioinformatics projects that take place in industry and learn from each other's expertise by working together to mentor students on diverse projects.

FF0025

Tephritid Fruit Fly Genomics at the Pacific Basin Agricultural Research Center

Scott Geib and Tom Walk

USDA-ARS Pacific Basin Agricultural Research Center, Hilo, HI, USA

Fruit flies are among the most serious economic agricultural pests worldwide. In Hawaii, fruit flies limit development of a diversified fruit and vegetable industry, require export fruits to undergo expensive quarantine treatments, and provide a reservoir for pest introduction into the mainland United States. In California where the total value of the fruit and vegetable industry has been estimated to be more than $14 billion annually, it has been estimated that an established infestation could cost $1.4 billion during the first year of establishment. The oriental fruit fly is a tephritid fly that is a serious agricultural pest. Despite the economic importance of tephritids, very little molecular information is available for this family of insects. To address this, a large scale sequencing effort is being performed on the oriental fruit fly, *Bactrocera dorsalis* that includes 1. whole genome shotgun sequencing, 2. transcriptome sequencing/RNA-Seq analysis, and 3. genome-wide comparative analysis (RAD-Tag) of the *Bactrocera dorsalis* species complex. WGS sequencing and assembly was completed with a hybrid 454/Illumina approach to produce ~40 X coverage. A variety of GMOD tools were used for genome annotation, analysis, and curation. In addition, RNA-Seq analysis of a variety of life stages of this insect was performed to both improve genome annotation and to identify genes which could be targeted for fruit fly control. *De novo* transcriptome assembly was performed with the ABySS/trans-ABySS pipelines and compared to reference based mapping. A current focus at USDA is in performing comparative genomics analysis of *B. dorsalis* populations from throughout southeastern Asia, Hawaii, and invasive populations in the mainland US, to develop tools to understand the relationship of populations of this insect.

FF0021

Optical Mapping and Whole Genome Finishing Elucidate Unusual Chromosomal Architecture in Some *Vibrio cholerae* Isolates

K. A. Bishop-Lilly[1], C. Chapman[1], J. Awosika[1], M. Henry[1], H. Tsang[1], M. Patel[1], A. Butani[1], R. Ptashkin[2], A. Briska[2], P. Chain[3], C. Han[3], S. Johnson[3], T. Wagner[2], C. Rajanna[4], A. Sulakvelidze[4], C. Detter[3], and S. Sozhamannan[1]

[1]Naval Medical Research Center, Biological Defense Research Directorate, BDRD Annex, 12300 Washington Ave, Rockville, MD, 20852; [2]OpGen Inc., 708 Quince Orchard Road, Gaithersburg, MD, 20878; [3]Los Alamos National Laboratories, Los Alamos, NM, 87545, University of Florida, Gainesville, FL 32610, USA.

Vibrio cholerae is a Gram-negative bacterium responsible for an estimated 3–5 million cholera cases and 100,000–120,000 deaths every year. Cholera is a major concern in hygiene-poor parts of the world especially during mass migration of populations due to natural or man-made disasters, such as the recent earthquake in Haiti. Of the more than 200 known serogroups of *V. cholerae,* cholera epidemics and pandemics are caused only by certain serogroups (O1 and O139). Although the major virulence factors of cholera such as cholera toxin, toxin coregulated pilus and ToxR present in O1 and O139 strains have been shown also to be present in many other serogroup strains, these non-O1/non-O139 strains have not been associated with any major outbreaks. Thus, other factors may play a role in the epidemicity of *V. cholerae* O1. A factor that has been overlooked in cholera research is the serogroup diversity encoded by the O-antigen biosynthesis cluster. NextGen sequencing technologies have enabled the generation of whole genome sequences of a large collection of bacterial strains inexpensively and rapidly; however, deciphering the complete genome architecture requires genome finishing, and more investment of time and resources. We hypothesized that *V. cholerae* has a high degree of diversity as a species, especially in the O-antigen biosynthesis region. To explore this hypothesis, we adopted an automated, high-resolution microbial whole-genome optical restriction mapping approach using the *V. cholerae* Shimada serogroup set, which contains all the 206 currently typed serogroup strains. Preliminary optical mapping data support the hypothesis that *V. cholerae* is a more diverse species than once thought, and the diversity is not fully conveyed by the strains that are currently sequenced. Additionally, optical mapping identified an unusual chromosomal architecture in two strains, which we have sought to validate by using NextGen sequencing technologies and LANL's high-throughput finishing pipeline.

FF0138

Comparative Gene Expression of the *Caldicellulosiruptor* Genus Using RNAseq

Loren J. Hauser[1*], Sara Blumer-Schuette[2], Ira Kataeva[3], Sung-Jae Yang[3], Farris Poole[3], Daniel Quest[1], Inci Ozdemir[2], Andrew Frock[2], Erika Lindquist[4], Tanya Woyke[4], Bob Cottingham[1], Michael W. W. Adams[3], Robert M. Kelly[2]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]North Carolina State University, Raleigh, NC; [3]University of Georgia, Athens, GA; [4]Joint Genome Institute, Walnut Creek, CA., *presenter email: hauserlj@ornl.gov

All known members of the *Caldicellulosiruptor* genus grow optimally between 65°C to 80°C and can anaerobically degrade plant biomass using various and complementary strategies. They are prime candidates for use in an industrial consolidated bioprocessing facility to produce second generation biofuels from complex plant material such as switchgrass. In collaboration with the Department of Energy Joint Genome Institute (JGI) we have recently completed sequencing and annotating the genomes of eight members of this genus. In addition, we have generated RNAseq data from four members grown on a variety of carbon sources including, glucose, maltose, cellobiose, starch, crystalline cellulose (Avicel), and dilute acid pre-treated switchgrass. Two of the primary advantages of RNAseq are its dynamic range and sensitivity. Greater than 98.5% of all protein coding genes had some detectable expression in all growth states and varied in expression level up to 10^6 fold. The expression levels of some genes, when grown on different carbon sources, varied by over 10^3 fold. As expected, the genes encoding ABC sugar transporters, cellulases and other glycosyl hydrolases were amongst the genes with the greatest changes in expression levels when grown on sugars versus complex carbon sources such as switchgrass. However, there were a number of other genes, such as members of a CRISPR cluster and some genes involved in fatty acid metabolism, that had unexpected changes in expression when grown on different carbon sources. We are developing an analysis pipeline to process and visualize the data and will also compare them with the results from DNA microarray analyses. RNAseq analyses will also include identifying the 5' end of transcription units, defining operons, identifying co-regulated genes and operons, and predicting transcription factor binding sites. Preliminary analysis has identified putative promoters embedded in genes, which allows the definition of unconventional operons and regulons. A thorough analysis will undoubtedly reveal additional unique biological phenomenon.

This project has 2 goals. The first is to compare the gene expression profiles, using RNAseq, of a series of related high growth temperature bacteria from the genus Caldicellulosiruptor grown using both simple and complex carbon sources. The second is to develop a set of analysis tools to process large amounts of RNAseq data.

FF0155

NCGR Genome Sequencing Center & Informatics Core

Jimmy Woodward, Peter Ngam, Ritu Bharti, Sarah Nelson, Andrew Farmer, John Crow, Brandon Rice, Callum Bell, Faye Schilkey, Ryan Kim, and Greg May

National Center for Genome Resources, Santa Fe, NM, USA

The National Center for Genome Resources (NCGR) is a non-profit research institute whose mission is to improve human health and nutrition through genome sequencing and analysis. The NCGR Genome Sequencing Center offers next generation sequencing, genotyping, and analysis services. The core houses PacBio, SOLiD, HiSeq, and GAIIx instruments for sequencing plus Illumina's BeadExpress and iScan instruments for genotyping. Sequencing services include whole genome and transcriptome shotgun sequencing, ChIP, small RNA, whole or targeted exome, and methylome sequencing. NCGR is an early-access user of the PacBio sequencing system and has been running the RS system in production since Q4 2010. The system features very long nucleotide read lengths that provide fidelity useful for de novo sequencing and resequencing projects.

NCGR has a longstanding reputation for developing effective bioinformatics tools and performing cutting-edge scientific research evident by two recent papers*: Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing,* Sci Transl Med. 2011 Jan 12;3(65):65ra4 [PMID 21228398], and *Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis,* Nature 2010 Apr 29; cover [PMID: 20428171], as well as past awards including 2009 Bio-IT World Best Practices in Basic Research: Schizophrenia and 2009 Laureate by Computerworld for Alpheus® a web-based analysis pipeline for variant and expression detection of next generation sequencing data. Alpheus was integral in both studies above.

FF0151

DOE Joint Genome Institute Center Update

Feng Chen

DOE Joint Genome Institute, Walnut Creek, CA USA

Next generation sequencing platforms have ushered in a new era of high throughput sequencing with a substantial change in the type and volume of projects now underway at genome centers. These instruments allow for innovative and cost effective strategies for both de novo sequencing and improving existing draft assemblies as well as whole genome and targeted re-sequencing for mutation discovery. We will discuss the focus of our institution and provide pipeline overviews.

F0080

Sanger Sequencing Center Update

Darren Grafham

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

The Wellcome Trust Sanger Institute (WTSI) has been a world leader in genomic sequencing for over 16 years and supports over 35 faculty researchers at the institute. The institute currently has 4 platforms for sequencing: 3730 capillary, 454 FLX, Illumina HiSeq, and PacBio which provides a diverse range of projects and needs. A brief overview of the variety of projects supported along with the lims, software and pipelines available will be presented.

FF0112

Broad Center Update

Michael Fitzgerald

Broad Institute of MIT and Harvard, Cambridge, MA, USA

Next generation sequencing platforms have ushered in a new era of high throughput sequencing with a substantial change in the type and volume of projects now underway at genome centers. These instruments allow for innovative and cost effective strategies for both de novo sequencing and improving existing draft assemblies as well as whole genome and targeted re-sequencing for mutation discovery. We will discuss the focus of our institution and provide pipeline overviews.

FF0110

Baylor Center Update

Donna Muzny

Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030, USA

Next generation sequencing platforms have ushered in a new era of high throughput sequencing with a substantial change in the type and volume of projects now underway at genome centers. These instruments allow for innovative and cost effective strategies for both de novo sequencing and improving existing draft assemblies as well as whole genome and targeted re-sequencing for mutation discovery. We will discuss the focus of our institution and provide pipeline overviews.

FF0123

WashU Center Update

Bob Fulton

The Genome Center at Washington University, 444 Forest Park Blvd., Saint Louis, MO 63108, USA

Next generation sequencing platforms have ushered in a new era of high throughput sequencing with a substantial change in the type and volume of projects now underway at genome centers. These instruments allow for innovative and cost effective strategies for both de novo sequencing and improving existing draft assemblies as well as whole genome and targeted re-sequencing for mutation discovery. We will discuss the focus of our institution and provide pipeline overviews.

FF0103

JCVI Center Update

Nadia Fedorova

The J Craig Venter Institute, Rockville, MD, USA

Next generation sequencing platforms have ushered in a new era of high throughput sequencing with a substantial change in the type and volume of projects now underway at genome centers. These instruments allow for innovative and cost effective strategies for both de novo sequencing and improving existing draft assemblies as well as whole genome and targeted re-sequencing for mutation discovery. We will discuss the focus of our institution and provide pipeline overviews.

FF0201

OICR Center Update

Johar Ali

The Ontario Institute for Cancer Research (OICR), Toronto, Ontario M5G 0A3, Canada

Next generation sequencing platforms have ushered in a new era of high throughput sequencing with a substantial change in the type and volume of projects now underway at genome centers. These instruments allow for innovative and cost effective strategies for both de novo sequencing and improving existing draft assemblies as well as whole genome and targeted re-sequencing for mutation discovery. We will discuss the focus of our institution and provide pipeline overviews.

Panel Discussion Notes

Panel Discussion Notes

Meet and Greet Party

530pm – 900pm, June 1st

Sponsored by Roche Diagnostics

Enjoy!!!

Poster Presentations (June 1st)
Even #'s 5:30-7:00pm, Odd #'s 7:00-8:30pm

FF0012

Improving De novo Assemblies for Single Cell Genomic Data.

Karen Davenport, A. Zeyton, H. Daligault, W. Gu, C. Munk, C. Han, and C. Detter

Los Alamos National Lab, Los Alamos, NM, USA

Sequencing data generated with DNA amplified from a single cell currently presents several challenges for a complete genome assembly; a strong bias is introduced during the amplification and chimeric reads are created that could induce misassemble. Therefore, a complete genome from a single cell isolated from uncultivable populations continues to be elusive. The amplified single cell genome can be sequenced at high coverage (200-2000x) and our goals are to increase the percentage of the genome covered and to improve the quality of the assembly by increasing contiguity with this data. Velvet has given the most positive results in our assembly efforts with minor coverage differences from another assembler such as ALLPATHS. We are working on combining the best assemblies generated with more than one assembly program so that the raw data can be utilized to our best advantage with regard to accuracy, contiguity, and percent coverage.

FF0030

BioPASS: Biothreat Pathogen Analytic Support System -- A Single Portal Analytic System for Pathogen Identification and Characterization

Helen Cui, Craig Blackhart, Chris Stubben, Ben McMahon, Bob Funkhouser, Carla Kuiken, Jennifer Harris, Chen He, Patrick Chain, Chris Detter, Amanda Minnich, Gary Resnick

Los Alamos National Laboratory, Los Alamos, NM, USA; Email: hhcui@lanl.gov

Los Alamos National Laboratory is developing a Biothreat Pathogen Analysis Support System (BioPASS), for pathogen identification and characterization. Vast volumes of biological data continue to grow, including those about pathogens, host-pathogen interactions, and epidemiological spread. Effectively utilize the data, analytic methods and growing knowledge is essential for decision-making support and countermeasure development. The main objective for BioPASS development is to establish a single portal analytic platform that utilizes existing and developing data analysis methods for rapid and credible microorganism identification and characterization, analytical classification of early indicators of possible proliferation activities, predicting impact of threat on public health and warfighters, and evaluating possible mitigation strategies. The long-term vision is an interagency core capability that is user friendly for different analytic requirements and adaptable to increasing data and methods.

The automated single portal analysis system under the development includes the following components: (1) organism identification interface, for rapid identification of a newly sequenced micro-organism and the phylogenetic placement highlighting anomalies, by the comparison with existing sequences; (2) identification of key microbial virulence factors such as adhesion factors, secretion systems, and antimicrobial resistance genes, and a screen of the genomes for drug sensitivity and potential resistance mechanisms; (3) infectious disease progression and epidemic spread potential; (4) microblogging information aided disease outbreak tracking; (5) biothreat pathogen and infectious disease knowledge integration, systematically exploring the novel pathogen that carries out each of its necessary pathogenic duties, such as attachment to host tissues and cells, evasion of the immune response, replication, and its lifecycle outside of the human host. This system will serve the dual role of identifying well-known pathogenic mechanisms and highlighting potentially novel attributes of the emerging pathogens.

Acknowledgement: The BioPASS project is sponsored by the US Department of State

FF0036

Optical Mapping as an Adjunct to Bacterial Genome Sequence Finishing

Kyle Hubbard[1,2], Stacey M. Broomall[1], Pierce Roth[1,3], Michael D. Krepps[1,2], Lauren A. McNew[1], Joseph M. Insalaco[1,4], Mark Karavis[1], Henry S. Gibbons[1], C. Nicole Rosenzweig[1]

[1]BioSciences Division, Research & Technology Directorate, Edgewood Chemical Biological Center, APG, MD, USA; [2]Excet, Inc, Springfield, VA; [3]Optimetrics Inc, Abingdon MD; [4]Science Applications International Corporation, APG, MD, USA

While next generation sequencing has made great advances in providing draft level genomic sequences in fewer than 48 hours, optical restriction maps can be prepared in less than 24 hours from sample receipt, supplying preliminary bacterial identification down to the subspecies level based solely on the unique pattern of restriction fragments. This inexpensive, rapid analysis can be applied to unknown samples and offer information that can be used for detection, characterization, and/or forensics.

This technology is considered orthogonal to shotgun sequencing in that it not only provides information on chromosomal structure and large-scale genomic architecture analysis but also provides information on the movement of mobile elements, CpG islands, and other DNA movements within the chromosome and plasmids. The optical maps can also be used to complement whole genome sequencing (WGS) for the identification of biological agents and validation of WGS draft assembly. Beyond draft level finishing, optical mapping may have a potential role in silver level sequence finishing by improving scaffolding for the orientation and alignment of sequenced *de novo* contigs, leading to more efficient gap identification and closure.

In addition to 454 shotgun sequencing, 454 paired-end sequencing, and Illumina sequencing, the Edgewood Chemical Biological Center (ECBC) is currently evaluating the utility of incorporating optical mapping technology to improve draft consensus sequences. A comparison of finishing improvements with and without the inclusion of Optical Mapping was undertaken. This evaluation has involved writing several bioinformatic scripts to facilitate file conversions among software without data loss and modifying and testing current software for functionality with combined procedures, including those developed by JGI and LANL. Subsequent to the *Bacillus* genome used to develop the pipelines, final evaluation of pipeline improvements are through the analysis of *Yersina*, *Brucella*, *Klebsiella*, and *Bacillus*. *In silico* restriction maps developed from the finished genomes can be used to verify the architecture observed in the optical maps produced for samples with no available reference genome.

FF0037

Finishing Genomes with the Illumina Platform Only

Olga Chertkov, Ahmet Zeytun, Wei Gu,Chris Munk, Karen Davenport Linda Meinke, Chris Detter, and Cliff Han.

Genome Group, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87505, USA

Recent improvements of the Illumina platform such as length of sequence reads (120 bp vs. 36 bp), coverage of genome (200-2000x vs. 25-50x)), and amount of data produced by one run (200 Gb per run) enable us to sequence almost all of a microbial genome. However, bacterial genomes are rich in near identical repeats and all parts are not represented equally in the library; both situations cause misassemblies and gaps.

Our team has recently tested different assembly programs (Soap De Novo, Allpaths, Abyss, and Velvet) using the Illumina data of six finished genomes. It was found that Velvet is the best for creating less contigs/scaffolds and fewer misassemblies. We further optimized the Velvet by changing k-mer size, clone coverage, coverage cut-off, etc., using six recently finished genomes that had both the illumina and 454 data. We compared our assemblies with reference sequences of the finished genomes and found that the same set of data can produce very different assemblies with different parameters. A process to find the parameters of the best assembly will be developed and automated. To resolve misassemblies further we converted the Velvet assembly into a phrap assembly. With a little bit of human interaction we can bring the assembly to very manageable size: less than 10 scaffolds.
Our team has been constantly improving the assembly procedures and developing additional bioinformatics tools to close the gaps due to repeats without additional sequence data.

FF0038

DNA Sequencing Facility in the GBT at Pfizer

Xiaohong Liu; Mostafa Ait-Zahra; Don Koffman; Tony Li; Jeffrey Tetrault; Jan Kieleczawa

Pfizer, Inc., 35 CPD, Cambridge, MA 02140, USA

The DNA sequencing at Global BioTherapeutics Technologies (GBT) department at Pfizer provides comprehensive services for scientists in almost all BioTherapeutics R & D units. Scientists submit their requests through our in-house developed 4D LIMS. Based on the DNA sample information, we assign appropriate primers from our primer inventory or design new ones if needed using our proprietary software; sequencing reactions are assembled either automatically (Hamilton) or manually.
Upon completion of a specific project, edited data is submitted to a centralized database (WyseCat) and 4D emails notification to a scientist about data availability. We provide full spectrum of DNA sequencing services from clone sequence confirmation (range from 0.5 to 20 kbp) to more high-through put for antibody library screening. We strive to finish each project within 1-2 days.

FF0042

Using Genomic Analysis to Understand the Evolution of Clinical Isolates of *Acinetobacter baumannii*

Robert Blakesley[1], Evan Snitkin[2], Sean Conlan[2], Jyoti Gupta[1], Brian Schmidt[1], Adrian Zelazny[3], Clemente Montero[3], Frida Stock[3], Lilia Mijaresa[3], NISC Comparative Sequence Program[1], Gerard Bouffard[1], Patrick Murray[3], Julia Segre[2]

[1]NIH Intramural Sequencing Center, NHGRI, Rockville, MD, USA, [2] Epithelial Biology Section, National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH), Bethesda, MD, USA, [3]Department of Laboratory Medicine, NIH Clinical Center, Bethesda, MD, USA

Acinetobacter baumannii is an emerging human pathogen and a significant cause of infections amongst hospital patients worldwide. The recent emergence of pan-drug resistant strains underscores the urgency to understand how *A. baumannii* evolves in hospital environments, such that effective surveillance and treatment protocols can be implemented.

Here, we undertook a genomic study of an outbreak of multi-drug resistant *A. baumannii* at the NIH Clinical Center. Standard laboratory typing techniques revealed that three closely related strain types were present during the outbreak, but provided limited insight into how the three strains came to NIH and whether the genetic variation among them was likely to be of clinical significance. To address these questions we sequenced the complete genomes of a single representative of each of the three outbreak strain types to a minimal coverage of 15X using the 454 Ti platform. Sequenced genomes were annotated to find the locations and putative functions of all protein coding genes, and the three genomes were then compared to identify both single nucleotide variants and changes in gene content.

Comparisons of the complete sets of nucleotide variants among the three genomes enabled us to conclude that the strains diverged prior to their arrival to NIH, with subsequent analysis allowing us to track down their likely origin. Additional comparative genomic analyses allowed us to uncover that much of the variability among the outbreak genomes was due to several large horizontal transfer events, encompassing roughly 20% of the outbreak genomes. Extending our analysis to include other international clinical isolates revealed a broad phenomenon whereby *A. baumannii* is able to rapidly vary potential immune targets by horizontal transfer. These findings demonstrate how sequencing of clinical bacterial isolates can contribute to our understanding of the spread and evolution of hospital pathogens.

FF0045

Whole-Genome Sequencing with Positional Hybridization Data

Peter Goldstein

NABsys Inc., 60 Clifford St., Providence, RI 02903, USA

The use of DNA sequencing for personalized medicine will require the development of a sequencing technology that has high-throughput and that is highly scalable. NABsys is developing an electronic, solid-state, single-molecule approach to DNA sequencing that satisfies both criteria. In this approach, referred to as Hybridization Assisted Nanodetector Sequencing (HANS), target DNA is fragmented and the fragments are hybridized with a probe of known sequence. Solid-state detectors are used to locate the position of probes hybridized to each fragment. The positional hybridization information is used to assemble high-resolution probe maps of the target sequence that retain long-distance information. By producing maps for a library of probes, the sequence of the target can be reconstructed.

We will present our recent work on the development of algorithms and software that reconstruct DNA sequences from hybridization data that contains positional error. Under current simulations, we are able to assemble genomic DNA fragments into map contigs averaging 100 megabases. The speed of this comparison is over three orders of magnitude faster than pairwise comparison of all fragments.

FF0050

Preparation of Nucleic Acid Libraries for Next Generation Sequencing with an Automated Molecular Biology Platform.

Kamlesh D. Patel, Ph.D., Hanyoup Kim, Michael S. Bartsch, and Ronald F. Renzi

Sandia National Laboratories, Livermore, CA, USA

While DNA sequencing technology is advancing at an unprecedented rate, sample preparation technology still relies primarily on manual bench-top processes, which are slow, labor-intensive, inefficient and often inconsistent. Automation of sample preparation using microfluidic techniques is well-suited to address these limitations. We have designed, fabricated, and characterized a digital microfluidic (DMF) platform to function as a central hub for interfacing multiple lab-on-a-chip sample processing modules towards automating the preparation of clinically-derived DNA samples for next gen sequencing (NGS). The automated molecular biology platform (AMB) is designed to interface directly with NGS to detect unknown pathogens by enriching informative nucleic acids sequences (those derived from the pathogen) and suppressing background DNA (those from the host) to maximize the sensitivity of state-of-the-art NGS. The AMB platform will be able to carry out a diverse series of benchtop-like steps at a scale adapted to handling small, but precious, samples for DNA manipulations, but with far greater speed and efficiency than at the benchtop.

We will present our recent developments on the core architecture of the AMB platform, the DMF central hub, and demonstrate its flexibility in coupling droplet-based microfluidics with continuous-flow microchannel devices to prepare DNA samples for NGS. The strength of combining these two different, but complementary, fluid processing methods enables the manipulation of nanograms to picograms of DNA with precise temporal and spatial control. We will discuss our results for collecting fractions of nanogram amounts of normalized DNA in discrete 1-uL droplets on the DMF device. Additionally, we will also present the integration of magnetic beads for clean-up and concentrate the DNA. Fragmented DNA is analyzed in real-time with microchip-based gel electrophoresis separation interfaced to the sample droplet for the correct size distribution for eventual cluster generation and high-throughput sequencing to discover the pathogen by its genomic sequence.

FF0052

Sequencing against Biothreats: Viral Hemorrhagic Fever Genomics and Mock Release Exercises

Shannon Johnson, Jennifer Price, Hajnalka Daligault, Matthew Scholz, Chien-Chi Lo, Lance Green, Patrick Chain, Chris Detter

Los Alamos National Laboratory, Genome Science Group (B-6), Los Alamos, NM, USA

The term viral hemorrhagic fever (VHF) refers to a variety of diseases caused by phylogentically diverse viruses. These viruses cause irregular, sporadic outbreaks of disease and, with few exceptions, lack successful prophylaxis. Mortality during outbreaks of VHFs can be as high as 90%, yet due to their high rates of mutation and difficulty in culturing, few genomic sequences have been made available for comparative study. The goals of this project are to increase the number of viral genome sequences available for study and to provide a database for comparisons of the genomes and coding regions.

As viruses, each must be grown in a host cell line, generally primate (monkey or human). Of those, 13 genomes have been finished and the rest are in progress. Sequencing of these genomes may require final extraction and cDNA generation (from Trizol extracts), shotgun sequencing using Illumina and/or Roche 454 platforms, filtering to remove host sequence data, and assembly followed by finishing or targeted reactions to complete the sequences.

Member groups funded by DTRA CBD-X also participate in bi-annual capability exercises. The goal of these exercises is to establish our abilities and growth areas in rapid sequencing and identification of target pathogens in mixed samples. In a recent exercise we utilized 3rd-generation sequencing technology (PacBio RS) to identify one target in a mixed sample only 20 hours after it was received.

The combination of sequencing viral strains and participation in capability exercises has improved our ability to rapidly filter out non-pathogen sequence data and analyze the remainder for virulence genes and deviations from known reference genomes. The combination of reduced sequencing time and increased understanding from analysis should help to protect our warfighters, should they become the victims of biowarfare.

FF0053

Staphylococcus epidermidis Whole Genome Sequencing Reveals Diversity in Commensal Skin and Nosocomial Catheter Isolates

Sean Conlan[a], Lilia Mijares[b], NISC Comparative Sequence Program[c], Heidi Kong[d], Patrick Murray[b] and Julia Segre[a]

[a]Epithelial Biology Section, GMBB, NHGRI, Bethesda, MD, USA; [b]Department of Laboratory Medicine, NIH Clinical Center, NIH, Bethesda, MD, USA; [c]NIH Intramural Sequencing Center, NIH, Rockville, MD, USA; [d]Dermatology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA

While *Staphylococcus epidermidis* is commonly isolated from healthy human skin, it is also the most frequent cause of infection on indwelling medical devices, such as catheters. No current microbiological assay or genetic markers exist to inform the crucial clinical decision of whether a coagulase negative staphylococcus, detected in a blood culture, is a commensal contaminant or a pathogen. The reference genomes for *S. epidermidis* are limited, consisting of a single catheter derived strain (RP62A) and a laboratory strain (ATCC12228). We investigated the genetic diversity of *S. epidermidis* by sequencing 21 isolates, directly cultivated from 15 sites on the skin, and comparing them with 3 nosocomial isolates from venous catheters. The commensal isolates demonstrate great genetic diversity as assayed by genome architecture and gene content. The nososcomial isolates, from three patients at three different times, but of the same dominant hospital-acquired lineage share much greater clonality. These data not only represent the first sequence of this important hospital strain, but raises questions such as the forces driving diversity amongst *S. epidermidis* commensals and the evolutionary fixation of nosocomial isolates.

FF0059

Recent Advances in High-Throughput, High-Reliability Focus Optimization for Genomics Applications

Jim Gareau

Physik Instrumente, Auburn, MA USA

The genomics community is entering an era of escalating economic stakes in which throughput and uptime dominate cost metrics. Such considerations have long been the hallmark of semiconductor manufacturing tool industry, so it is not a surprise that solutions for today's challenges in industrial genomics engineering are emerging, in large part, in the form of mechanisms and technologies familiar from mission-critical semiconductor tooling.

Top among these is piezoelectric actuators, which after decades of utilization in semiconductor manufacturing are now the heart of the fast, compact and ultra-reliable focusing and sample-positioning mechanisms being deployed in noteworthy genomics instruments and research applications. Layered, solid-state ceramic structures of exquisite precision in their own right, these actuators drive the sophisticated flexure mechanisms which define the instrument's capabilities for resolution and throughput.

The field of fast piezo focusing instrumentation has enjoyed a particularly rich vein of innovation in recent months. Advancements in mechanisms and controls have included:

o Piezo mechanisms capable of nanoscale resolution with up to 1mm travel and high axial stiffness;

o Cost-effective compact digital nanopositioning controllers;

o High-speed focus sensors with native controller integration;

o Novel servo capabilities which include bumpless transitioning between focus-sensor and position-sensor control.

Here we review these and related advancements which are enabling the next generation of genomics applications and integrated systems. With an emphasis on fast focus optimization and tracking, we begin with a review of theoretic piezoelectric fundamentals relevant to the field, proceed through common and emerging mechanical and controls technologies, discuss recent advancements in throughput enhancing technologies, and conclude with a discussion of important applications metrics and performance metrology techniques.

FF0066

The EvolScope Project: An Extension of the MicroScope Platform to Study the Evolution of Bacterial Polymorphism from High-Throughput Sequencing Data

David Roche[1], Béatrice Chane-Woon-Ming[2], David Vallenet[1], Claudine Médigue[1] and Stéphane Cruveiller[1]

[1]CEA/DSV/genoscope & CNRS UMR8030, Laboratoire d'Analyses Bioinformatiques pour la Génomique et le Métabolisme, Evry, France ; [2]IBMC/CNRS UPR9002, Architecture et réactivité de l'ARN. Strasbourg, France

Recent improvements in sequencing technologies largely contributed to the comparative and evolutionary genomics rebirth as now crude complete genome sequences (i.e. thousands to millions short pieces of DNA sequence called reads) can be produced in very short time with an accuracy never/hardly reached with the Sanger technology. Consequently, with Next Generation Sequencing technologies (NGS), micro events, SNPs and insertions/deletions (indels) can be confidently identified and confirmed by several hundred reads.

These events detected via resequencing and comparative analysis of several highly related organisms constitutes a key source of information for researchers interested in evolutionary aspects of prokaryotic life. To address this specific issue, a set of dedicated tools have been developed and integrated into our Microscope platform [1] giving thus rise to the EvolScope project. EvolScope comprises three components : a pipeline to compute variations (SNiPer, Cruveiller *et al.*, unpublished), a relational database and a graphical interface.

Though extremely powerful, NGS are not biases/errors free. Hence, discriminating between true mutations that have occurred during evolution and sequencing errors (SNPs and non-SNPs) remains a challenging task. Our current micro events detection pipeline (SNiPer I) relies on SSAHA2 [2] package developed at the Sanger Institute. It's an algorithm for very fast matching and alignment of DNA sequences and thus is used to identify regions of high similarity. Events calling/filtering is then done according fixed criteria (alignments, coverage, qualities, strand bias, …) and significant events are subsequently stored in a dedicated relational database called EvolGenomes. This latter is plugged to PkGDB which constitutes the central database of the MicroScope platform where are stored, among other, annotations of completely sequenced genomes [1].
Evolution data are accessible through a web interface which offers the possibility of performing three kinds of analyzes: a “comparative” mode allowing the retrieval of mutations present in some organisms and absent from others and their genomic and functional context, a “parallelism” mode easing polymorphic sites detection and a “graphical” mode reporting the distribution of one clone's mutations along the reference chromosome and hence detecting potential mutational hot spots.
We will present several concrete examples of the use of these interfaces in the context of collaborative projects [3] [4].

References

[1] Vallenet D., Engelen S., et al. MicroScope: A platform for microbial genome annotation and comparative genomics. Database 2009:bap021 (2009) [2] Ning Z, Cox AJ and Mullikin JC. SSAHA: a fast search method for large DNA databases. Genome research 2001;11;10;1725-9 Marta Marchetti et al. Experimental Evolution of a Plant Pathogen into a Legume Symbiont PLoS Biol. 2010 January; 8(1): e1000280. Published online 2010 January 12. doi: 10.1371/journal.pbio.1000280.[3] Experimental Evolution of a Plant Pathogen into a Legume Symbiont.Marta Marchetti, Delphine Capela, Michelle Glew, Stéphane Cruveiller, Béaatrice Chane-Woon-Ming, Carine Gris, Ton Timmers, Véréna Poinsot, Luz B. Gilbert, Philipp Heeb, Claudine Médigue, Jacques Batut, and Catherine Masson-BoivinPLoS Biol. 2010 January; 8(1): e1000280. Published online 2010 January 12. doi: 10.1371/journal.pbio.1000280. [4] Barrick J.E., Yu D.S., Yoon S.H., Jeong H., Oh T.K., Schneider D., Lenski R.E., and Kim J.F. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 461, 1243-1247.

FF0077

Advancements in Focused Acoustics for Use in NGS Sample Preparation

H. Khoja, G. Durin, and J. Laugharn

Covaris Inc., 14 Gill Street, Unit H, Woburn, MA 01801, USA

Rapidly advancing sequencing technologies have continued the demand for advances in nucleic acid fragmentation technologies. The increase in sensitivity and throughput of these platforms has placed a greater demand on the accuracy, reproducibility, and throughput of the of the library preparation steps which begins with shearing of DNA, RNA, or chromatin.

Technologies such as nebulization, unfocused sonication, hydrodynamic shearing, and enzymatic digestion have remained virtually unchanged in decades. The limited utility they had with first generation, and some next generation platforms are quickly being exhausted. The increase in sensitivity and coverage of NGS platforms has exposed and amplified some of these inherent limitations which include thermal and sequence specific biased fragmentation, thermal degradation, precious sample loss, user-dependent variability, or automation incompatibility.

In contrast, the Covaris Adaptive Focused Acoustics (AFA) has advanced with sequencing technology advancements. AFA's engineered closed vessel, non-contact, isothermal technology has since become capable of offering a wide range of fragment sizes from 100bp to 5kb, and is now considered the gold standard for DNA fragmentation used with all NGS platforms in many labs including all the large sequencing centers around the world.

In this meeting we will present Covaris' technological and application advancements and supporting data. These include our award-winning LE220 instrument for ultra-high throughput labs capable of parallel processing (e.g., 96 samples to 300bp in less than 15 minutes), and our new "220" generation electronics with finer power resolution and broader dynamic range, and our new updated software. We will also introduce Covaris AFA-certified reagent kits. For example, chromatin shearing protocols for cultured cells and tissues with unprecedented control for ChIP-Seq applications enabling higher retention of epitope integrity. Our simple and reproducible RNA (total and mRNA) shearing protocol is also a beneficial replacement for the heat and chemical cleavage methods currently utilized in RNA-Seq library preparation.

FF0082

Assembly and Genome Enhancement of *Clostridium difficile* Reference Strains

Hilary Browne

Wellcome Trust Sanger Institute (WTSI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Clostridium difficile is a major cause of antibiotic associated diarrhoea and pseudomembranous colitis. Recent outbreaks have been attributed to numerous virulent strains evolving independently of each other (He *et al.* 2010). Described here is the assembly and genome enhancement of 6 of the most prevalent strains currently associated with *C. difficile* infection in the UK. These reference genomes will contribute to elucidating the genetic traits behind *C. difficile* virulence and emergence. All 6 strains were taken to 'an improved high quality draft' standard as described by Chain *et al.* 2009. Sequence data was produced using Illumina GAII (paired-end 108bp read length, 300-400bp insert size) and 454/Roche (paired-end 400bp read length, 3kb insert size) technologies. These were assembled using Velvet and Newbler respectively. A merged 454/Illumina assembly was then created using Newbler. Computational gap closure using IMAGE (Iterative Mapping and Assembly for Gap Elimination) (Tsai *et al.* 2010) was followed by a round of ABACAS (algorithm-based automatic contiguation of assembled sequences) (Assefa *et al.* 2009) to design PCRs. Subsequent to this ordering and orientation of the dataset was achieved by integrating the Newbler 454Scaffolds.fna files and then by ordering the scaffolds using a combination of ABACAS and ACT (Artemis Comparison Tool) (Carver *et al.* 2005). This was confirmed by a round of PCR. This process is described in detail by van Tonder and Grafham 2011. The Wellcome Trust Sanger Institute is the first Institute in Europe to acquire the Argus Optical Mapping system. This was used to verify the order and orientation of the final dataset. iCORN (Iterative Correction of Reference Nucleotides) (Otto *et al.* 2010) was then used to correct sequence errors. The results of this process in addition to the biological results obtained will be presented. The effectiveness of utilising Optical Map technology will also be presented.

FF0084

PRINT® Particles and Well Arrays for Next Generation Sequencing

Trevor Knutson, Ben Maynor, Ash Nijhawan

Liquidia Technologies, 419 Davis Drive, Suite 100, Durham, NC 27713, USA

While next generation sequencing technologies use different platforms and sequencing principles, many of the approaches require using micro-beads with precise control of particle size, chemistry, surface functionality and porosity and/or high density micro- or nano-wells with precise control of size and shape. Currently, micro-beads are typically made by emulsion polymerization and are limited in their capability to control all of these critical design parameters. Additionally, micro and nano-well arrays are typically fabricated using costly wet etching methods. Liquidia Technologies has recently developed the PRINT® technology platform for manufacturing monodisperse particles and nanopatterned surfaces with tailored functionality, and in this work we demonstrate its applicability to sequencing applications.

Micro- and nano-particles (100nm to 100μm in dimension) with different compositions were prepared by PRINT technology. Particles produced were truly mono-disperse and showed excellent batch to batch consistency. Prepared from a molding based process, the size and shape of particles are independently tunable from their composition. We also demonstrate unprecedented control of surface functionality, rigidity and porosity of PRINT particles. These particles offer vast compositional and structural flexibility for use as DNA capture and amplification substrates.

In addition to particles, PRINT technology was also used to manufacture high density, non-fouling micro- and nano-well arrays with high fidelity, exquisite uniformity and low cost. This technology combines the precision of a lithographic process with the scale and economics of roll to roll manufacturing. Additionally, well arrays were selectively filled with functionalized materials, leading to addressable, discretized islands of reactive material in nonreactive matrices. Empty and functional well arrays may be used as bead or free DNA capture sites for existing and future sequencing platforms.

FF0090

Detailing ‘Genome Standards’, for Whole-Genome Data Sets, as Defined by the Wellcome Trust Sanger Institute Sequencing Department

Heidi Hauser

Wellcome Trust Sanger Institute (WTSI), Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA, UK

Since the release of the community-defined standards for next generation sequencing technologies, as described in Chain et al. (2009), WTSI have refined these, largely, loosely defined categories by documenting specific workflows for each of these standards so as to define specific minimum requirements that need to be achieved in order for data sets to be released under a specified category. Transparency of these processes are vital to maintain and ensure that genomic standards are upheld, on an international level, as more New Sequencing Technologies (NST) sequence data becomes publically available (Field et al. 2008). Workflows for these defined standards have been investigated and formalized through iterative processes and will be open to continual change in the future as advances are made in bioinformatics and ‘third-generation’ sequencing technologies develop.

Workflows, with examples, for, ‘High-Quality Draft’, ‘Improved High Quality Draft’ and ‘Finished’ will be presented. Details of minimum assembly statistics together with other criteria associated with sequence and/or physical gaps, misassemblies, base quality and read-depth coverage will be shown. These workflows are designed to be rigid enough so as to provide open and meaningful public information on data quality and quantity but flexible enough so as to meet individual requirements for the scientific projects that are serviced at WSTI Sequencing Department.

FF0091

Bacterial Artificial Chromosome End Sequencing (BES) of Whole DNA Libraries Evolves with Illumina and DNA Sudoku

Richard Clark

Wellcome Trust Sanger Institute (WTSI), Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA, UK

Bacterial Artificial Chromosome (BAC), End Sequencing (ES) is a highly effective method used to map BAC clone libraries to an assembly. Forward and reverse sequencing reads are produced from each end of the BAC insert, using the Sanger method. The paired reads are aligned to an assembly, mapping the start and stop position of each BAC within the library.

A strategy currently under examination at the WTSI is the DNA Sudoku method (Yaniv Erlich et al 2009) on the Illumina platform, to sequence part of the CHORI-103 library. The CHORI-103, Schistosoma mansoni library, (Le Paslier, M. C. et al. 2000) contains nearly 37,000 BAC clones and typically costs ~£20,000 to BES. The Illumina platform provides ample capacity to sequence the whole library over 18 lanes, reducing this cost by ~66%.

BES problems can also occur when sequences fail to map or map to multiple loci. BES reads typically cover <1% (1.5Kb) of the DNA insert. The Illumina, DNA Sudoku method generates the whole BAC sequence (170Kb) and as such reduces the number of non-mapping or multiple mapping BAC clones

DNA Sudoku uses DNA barcodes not associated with single samples but rather combinatorial pooling strategies. This radically increases the number of samples to be collectively sequenced. Here we present 143 BAC clones from the CHORI-103 library and the progress of sequencing using 24 DNA barcodes, the pooling strategy and the assembly strategy.

FF0093

Single Cell Sequencing Reveals Metabolic Specialization of Bacteria and Archaea in the Dark Ocean

Swan, Brandon K., Manuel Martinez-Garcia, Nicole J. Poulton, E. Dashiell P. Masland, Monica Lluesma Gomez, Michael E. Sieracki and Ramunas Stepanauskas

Bigelow Laboratory for Ocean Sciences, 180 McKown Point Road, P.O. Box 475
West Boothbay Harbor, Maine 04575-0475, USA

The subphotic or dark ocean (ca. > 200m) contains an active and metabolically diverse microbial assemblage that significantly contributes to marine biogeochemical cycles. However, the paucity of representative pure cultures so far has hindered experimental and genomic studies of the metabolic potential of the majority of microorganisms from deep water masses.

We employed a combination of fluorescence-activated cell sorting, multiple displacement amplification, and DNA sequencing to obtain genetic information from individual microbial cells from mesopelagic (ca. 500-1,000m) samples collected in the South Atlantic and North Pacific tropical gyres. This approach links phylogenetic (e.g. SSU rRNA) and metabolic marker genes by sequencing them from the same cell obtained directly from its environment, without the need for cultivation.

We obtained genomic DNA of 502 individual bacterioplankton cells representing most of the globally distributed subphotic lineages**.** Multiple cells of the subphotic-associated *Proteobacteria* groups SAR324, ARCTIC96BD-19, Agg47 and *Oceanospirillales* contained RuBisCO and several genes involved in sulfur oxidation pathways, indicating these groups may be capable of chemoautotrophy. Marine Group I (MG-I) *Crenarchaeota* comprised 89-96% of all identified archaeal cells. The majority of MG-I were found to harbor genes coding for ammonia monooxygenase (*amoA*) and nitrite reductase (*nirK*), which mediate the oxidation of ammonia.

Our study provides the first direct evidence for the presence of chemolithoautotrophy genes in several uncultured proteobacterial lineages that are ubiquitous in the dark ocean and therefore may play an important role in global biogeochemical cycles. We also found that the majority of mesopelagic MG-I cells contain *amoA* and *nirK* genes, supporting previous findings that these *Archaea* are among the predominant chemoautotrophs in the mesopelagic. Our results demonstrate that single cell sequencing is a powerful tool for investigating the metabolic potential of uncultivated microbial lineages.

FF0098

Evaluation of NGS Assemblers and Scaffolding Tools for Assembling the Wolbachia Endosymbiont of *Diaphorina citri* (wACP)

Surya Saha[1], Wayne B. Hunter[2] and Magdalen Lindeberg[1]

[1] Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY, USA; [2] USDA-ARS, US Horticultural Research Lab, Fort Peirce, FL, USA

The Asian citrus psyllid (*D. citri* Kuwayama or ACP) is host to 7+ bacterial endosymbionts and is the insect vector of Ca. liberibacter asiaticus (Las), causal agent of citrus greening. To gain a better understanding of endosymbiont and pathogen ecology and develop improved detection strategies for Las, DNA from D. citri was sequenced to 108X coverage to produce a metagenome with paired-end and mate-pair Solexa read libraries. Initial analyses have focused on Wolbachia, an alpha-proteobacterial primary endosymbiont typically found in the reproductive tissues of ACP and other arthropods. The metagenomic sequences were mined for wACP reads using BLAST and 4 sequenced Wolbachia genomes as bait. Putative wACP reads were then assembled using velvet and MIRA3 assemblers over a range of parameter settings. The resulting wACP contigs were annotated using the RAST pipeline and compared to Wolbachia endosymbiont of *Culex quinquefasciatus* (wPip). MIRA3 uses quality values during assembly that led to a qualitatively better assembly but one with a higher N50 value than the velvet assembly. MIRA3 was able to reconstruct a majority of the wPip CDS regions and was selected for scaffolding with Minimus2, SSPACE and SOPRA using large insert mate-pair libraries. The wACP scaffolds were compared to wPip as reference genome using the Mauve contig realigner to orient and order the contigs. Gaps will be identified in the assembly and PCR reactions will be performed to resequence the missing regions to close and finish the genome. The relative merits of different assembly strategies will be discussed.

FF0099

Improving the Human Genome Reference Sequence

Tina Graves and the Genome Reference Consortium

The Genome Institute at Washington University, 4444 Forest Park Blvd., St. Louis MO 63108, USA

The production of a high quality human genome reference sequence marked a significant scientific milestone. It has provided a foundation for genome-wide studies of genome structure, human variation, evolutionary biology and human disease and will continue to provide a platform to further our understanding of human biology. Many of these studies have also revealed, however, that there are regions of the human reference genome that are not represented optimally. It was known at the time the reference genome was completed that there were some regions recalcitrant to closure with the technology and resources available at that time. It was not clear, however, the degree to which structural variation affected our ability to produce a truly representative genome sequence at some loci. It has now become apparent that to sufficiently represent some highly variable loci, multiple sequences are needed to capture all of the sequence potentially available at these loci.

In order to improve the representation of the human reference genome, the Genome Reference Consortium (GRC) was created. The goal of this group is to correct the small number of regions in the reference that are currently misrepresented, to close as many of the remaining gaps as possible, as well as produce alternative assemblies of structurally variant loci when necessary. Many of these regions in the genome, particularly the structural variant loci, tend to be associated with repetitive sequences. One resource that is currently being utilized for these problematic regions is the Hydatidiform Mole BAC library (CHORI-17), which is a single haplotype library. Selecting a clone path from this library allows us to discriminate between different haplotypes and repeat copies. To date, we have resolved several of these loci by using this resource and have many other regions under investigation. Where possible, we are incorporating next generation sequencing strategies in order to greatly decrease the cost of these efforts. The efforts of the GRC will hopefully result in a more complete view of the human reference, which will better facilitate continued progress in understanding and improving human health. This presentation highlights many of the efforts underway to improve the reference and enhance the understanding of the human genome.

FF0102

Rapid Finishing of *Yersinia pestis* KIM D27 at JCVI

D. Radune, J. Hostetler, M. Kim, J. Varga, L. Losada, W. Nierman

The J Craig Venter Institute, Rockville, MD, USA

Yersinia pestis KIM D27 strain is an avirulent strain of *Yersinia pestis*, the etiologic agent of plague. The KIM D27 strain, a derivative of the fully virulent strain KIM 10, contains a 100kbp deletion that encodes several genes including genes for iron acquisition. The KIM D27 strain sequenced at JCVI was isolated from a lab where a researcher died from an accidental infection by the D27 strain. JCVI was tasked to rapidly sequence and analyze the KIM D27 strain and to investigate any differences between KIM D27 and KIM 10 that may have contributed to the sudden virulence and death.

The strain was sequenced with a full plate 454 Titanium 8 kb paired-end run and a single lane of 100 base paired-end Illumina. The average coverage of the chromosome was 400x. In addition, Sanger sequences were generated to fill the gaps and produce a final finished molecule as well as to confirm differences between the D27 strain and the KIM 10 reference (NCBI gi|22002119|gb|AE009952.1|). A number of genome assemblers were first used with the 454 only data with variable results. Most of the assemblies were very fragmented with a high degree of discrepant bases relative to the KIM 10. Illumina sequencing was added to correct consensus errors and improve the overall assembly quality. A total of 57 chromosomal SNP/INDELs and one plasmid SNP were identified relative to the KIM 10 reference and these were verified by sequencing of PCR products. In addition, several repeat areas were verified to be different in length and repeat number compare to the KIM 10 sequence.

In order to verify that the detected repeat and SNP differences new in the D27 strain, the *Y. pestis* KIM10+ DNA was obtained from ATCC and all SNP and repeat PCRs were run on this DNA. This poster will highlight the JCVI finishing process of *Y. pestis* KIM D27, the results of the comparative analysis between KIM D27 and KIM 10 strains and some challenges that finishing group faced during this process. It will also review the consequence of the 454 homopolymer induced sequencing errors in performing this analysis.

FF0103

JCVI Viral Finishing Pipeline: A Winning Combination of Advanced Sequencing Technologies, Software Development and Automated Data Processing

Nadia Fedorova, Danny Katzel, Tim Stockwell, Jessica Hostetler, Peter Edworthy, Rebecca Halpin, David Spiro, and David Wentworth

The J Craig Venter Institute, Rockville, MD, USA

JCVI is now in its sixth year of viral genomics projects. Over ten different viral projects are in progress supported by the NIAID Genomic Sequencing Center for Infectious Disease (GSCID). These projects represent many viruses including influenza, coronavirus, rotavirus, paramyxovirus, adenovirus, arbovirus, measles, mumps, rubella, and norovirus. The viral sequencing and finishing pipeline at JCVI combines amplicon-based Sanger and next generation sequencing technologies with automated data processing. This allowed us to complete over 2600 viral genomes in the last 12 months, and over 8100 genomes since 2005.

Our NextGen pipeline, which utilizes SISPA-generated genomic libraries with Roche/454 and Illumina sequencing, enables us to complete traditionally challenging samples (e.g. avian influenza and previously unknown viruses). The automated next-gen assembly pipeline employs CLC command-line tools and JCVI cas to ace assembly format conversion tool, called cas2consed. Finishing work is performed using the widely available Consed editor. Our highly automated Sanger pipeline, initially developed for influenza viruses, is now integrated into all viral projects. Its primary component is VAPOR, a completely automated suite of assembly tools. Both NextGen and Sanger pipelines incorporate the quality control and validation software called autoTasker. Our goal is to link all pipeline components into a single, integrated software suite for rapid and efficient viral genome sequencing.

To streamline our viral pipelines, we are adapting JIRA for sample tracking. Our goal is to create a semi-automated tracking interface that follows the progress of viral samples from acquisition through to NCBI submission. The combination of highly optimized sequencing technologies and automated software tools allows for large volumes of sample processing with limited manual interaction. These new developments will increase production capacity, effectively decreasing costs, manual labor, and sample completion time.

FF0106

Improvement and Finishing of the Draft *Medicago truncatula* Genome

Haibao Tang, Vivek Krishnakumar, Shelby Bidwell and Christopher D. Town

J. Craig Venter Insitute, 9704 Medical Center Drive, Rockville, MD, 20850, USA

The *Medicago truncatula* sequencing project began in 2003 with an initial goal to decipher the sequences of the euchromatic portion of the medicago genome. We previously constructed the pseudomolecules of the 8 chromosomes of *M. truncatula* based on a tiling path that contains 2536 BACs, with an additional 146 BACs not yet placed on the chromosomes. This BAC-based assembly, termed “Mt version 3.5”, currently contains a total of 329 million bases with scaffold L50 of 1.33Mb. Since the initial BAC sequencing, additional whole genome shotgun sequences (WGS) were obtained using 454 and Illumina technology with a combined coverage of ~80X. A separate whole genome assembly based on these shotgun reads is available. The WGS assembly has a much smaller contig L50 of 6.0Kb, but contains sequences (mostly from the heterochromatic portion) that do not yet exist on the BAC-based assembly. Our goal is to improve and finish the BAC-based assembly by incorporating these new sequences. There are two major components of our finishing procedures. First, there are 832 BACs of phase 1 or phase 2 (draft status) that we can improve and finish *in silico*. The contigs within the phase 1 BACs are oriented and ordered when possible. Phase 2 BACs are improved by performing gap closures by bridging the gap-spanning WGS reads or contigs. Second, we anchor the WGS contigs as well as previously unplaced BACs onto the chromosomes using sequence overlaps, mate-pair linkages and optical map alignments. Herein, we describe our detailed finishing strategy and report the improvements in terms of contiguity and coverage for the upcoming release of the medicago genome (Mt version 4).

FF0109

Sequence Validation of the Ion Torrent Personal Genome Machine

Christian Buhay, Michael Holder, Huyen Dinh, Tittu Matthew, Yuan-Qing Wu, Mark Wang, Irene Newsham, Donna Muzny and Richard Gibbs

Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030, USA.

Since the beginning of 2011, the BCM-HGSC has been evaluating Life Technologies' Ion Torrent Personal Genome Machine sequencing platform. This next generation platform uses a large parallel array of semiconductor sensors to perform direct real-time measurement of the hydrogen ions produced during DNA replication. The HGSC is investigating the platform as a means to produce sequence data for a variety of applications with a relatively short turn-around time.

Different genome types were chosen to demonstrate performance under various conditions. A set of 3 microbial genomes— *S. aureus USA 300* (GC content: 32%), *E. coli* MG1655 (GC content 50%), and *Rhodobacter spheroides* (GC content 68%) representing a wide range of GC content, were chosen to investigate sequence recovery and error rates. A set of four rat BAC, previously finished to Phase 3 or Gold Standard were chosen to perform similar coverage and error analysis using a more complex genome without having to produce large amounts of sequence. Additionally, two microbes from the Human Microbiome project in HGSC's finishing pipeline were also sequenced to determine ease of data manipulation, assembly, and finishing. Many of the organisms above have also been sequenced on the Roche 454 and Illumina Genome Analyzer II. A comprehensive comparative cross-platform analysis is ongoing. Furthermore, HGSC is assessing the utility of the PGM for regional capture of small targets (1-2Mb designs) and amplicon sequencing applications. Sequencing and analysis efforts are pending.

FF0112

HMP Reference Bacterial Genomes

FitzGerald, M, Earl, A, Priest, M, Arachchi, HM, Macdonald, P, Imamovic, A, Lui, A, Abouelleil, A, Gearin, G, Montmayeur, A and Birren, B.

The Broad Institute of MIT and Harvard, Cambridge, MA, USA

The Human Microbiome Project (HMP) seeks to identify and characterize the microbial content of healthy humans and that in disease states. The proposal to generate 3000 bacterial reference genome assemblies is an important part of the HMP. This data set will help us to understand the functional significance of microbes living in and on the human body and will also provide an important reference for human microbiome metagenomic data. The Broad Institute will generate approximately 25% of these reference genomes, 15% of which will be improved. Genome finishing and improvement efforts are targeting reference genomes that represent both novel and important organisms found within the human microbiome. We will provide an update on the Broad HMP reference genome set, detail our genome improvement plans and provide current results.

FF0113

Process Development for Closure of Illumina Based Genome Assemblies

Abouelleil, A, Arachchi, H, Macdonald, P, Lui, A, Imamovic, A, Priest, M, Gearin, G, Montmayeur, A, Walker, B, Birren, B and FitzGerald, M.

The Broad Institute of MIT and Harvard, Cambridge, MA, USA

Completed genome sequence remains an important part of our pathogen research program. Finished sequence facilitates rapid elucidation of drug resistance and virulence mechanisms and provides a template for mapping related strains.

Illumina based sequencing and related assembly software is leading to a rapid ramp in our ability to generate high quality pathogen genome sequence. Shotgun data for 24 organisms can easily be generated from a single Illumina lane. Rather than being simplified, the path to finished genome sequence has been complicated by these advances. The most recent assembly software generation is based on k-mers and does not track the shotgun read relative location as required for existing finishing systems. The lack of aligned reads forms a significant roadblock to detection and correction of improperly assembled regions. Illumina shotgun sequence coverage can be substantially deeper, yet is comprised of shorter reads compared with the more familiar Sanger and 454 data, leading to numerous issues with read placement and standard assembly viewers. Read pairing is also substantially different, being comprised of jumping fragments rather than physically linked clones. Finally, the low cost of Illumina shotgun data makes the expense of directed finishing reads seem enormous. We will present our finishing process as applied to challenging pathogen genomes like *Mycobacterium tuberculosis*, where we will seek to exploit the additional coverage and reduced bias of these assemblies, while limiting the stated liabilities. Part of this solution may reside in application of multiple shotgun data sets.

This project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900006C.

FF0114

Case Studies on Closure of Bacterial Pathogen Genome Assemblies

Imamovic, A, Arachchi, H, Macdonald, J, Abouelleil, A, Lui, A, Priest, M, Gearin, G, Montmayeur, A, Dunbar, C, Abed, L, Birren, B and FitzGerald, M.

The Broad Institute of MIT and Harvard, Cambridge, MA, USA

Next generation DNA sequencing technologies have revolutionized pathogen research. The increased efficiency, deeper coverage and reduced bias have allowed us to rapidly increase the production of high quality draft sequence. While the production of finished sequence has not seen a similar productivity increase, complete genome sequence remains an important part of pathogen research as it is a powerful tool for the assessment of drug resistance and virulence mechanisms. As the Roche-454 system became a routine data production platform we started a finishing pilot including a number of pathogen genomes. The genomes were selected to contain a range of features like GC and repetitive content. The genomes include *Staphylococcus aureus* 55/2053 and *Lactobacillus paracasei* 8700:2 which moved rapidly to complete sequence. Genomes holding more significant repetitive content (*Enterococcus faecium* C68, *Neisseria gonorrhoeae* FA19 and *Neisseria gonorrhoeae* MS11), or sequence recalcitrant to our available Sanger and 454 sequencing technologies (*Enterocuccus casseliflavus* 899205) have been more challenging. The opportunities of 2nd generation sequence data are tempered with the reduced genome closure tool kit that can be applied. The lack of physical clones and limitation to short jump links are both significant issues. We will detail the process applied to these genomes along with relevant results, highlighting strengths and opportunities.

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900006C.

FF0117

Methods for Studying TB Global Phylogeny and Resistance Patterns

Shihai Feng[1], Karina Yusim[1], Prashini Moodley[2], Tanmoy Bhattachary[1], A Willem Sturm[1] and Bette Korber[1]

[1]Los Alamos National Laboratory, USA; [2]University of KwaZulu-Natal, South Africa

Drug resistant tuberculosis is an increasing threat to the public health. Multi-drug resistant TB and extensively drug resistant TB are associated with 40% mortality and 85% mortality respectively. Finding the mechanisms of the drug resistance of TB is extremely important. Here, we present methods for phylogenetic and computational analysis of 60 nearly complete TB genomes, 5 of them are new isolates from KwaZulu-Natal, South Africa. We developed suite of tools to codon align multiple genomes, correct for sequence and alignment problems, statistically analyze frame shifting mutations and SNPs both from protein coding and promoter regions. An artificial SNP sequence contained a concatenated version of all variable positions in which at least 1 of 60 sequences was different from other were made. A molecular parsimony tree was generated based on this artificial SNP sequence. Strong phylogenetic relationships between TB samples from different geographic regions were found.

FF0118

Marine Microbial Eukaryote Transcriptome Sequencing Project

Arvind K. Bharti, Neil A. Miller, John A. Crow, Robin Kramer, Thiru Ramaraj, Alex G. Rice, Ken A. Seal, Gregory D. May and Callum J. Bell

National Center for Genome Resources (NCGR), Santa Fe, NM 87505 USA

Marine microbial eukaryotes comprise a vast array of single-celled, nucleated microorganisms, including diatoms, dinoflagellates, prasinophytes, chlorophytes, radiolarians, foraminferia, and amoeba. These organisms fill numerous ecological roles, such as photosynthetic primary producers at the base of marine food webs, heterotrophic consumers of pre-formed organic compounds, parasites and predators. In addition, as with some dinoflagellates, some microeukaryote groups are endosymbionts of marine animals, like corals, and play a critical role in reef ecosystems. Despite their great abundance and importance, the gene content of oceanic microbial eukaryotes has not been studied intensively because their genomes can be structurally complex and extremely large. Therefore, one of our major goals is to generate a catalogue of expressed genes of these microbes in order to improve the research community's understanding of the ecology and evolution of these highly diverse organisms.

We have already received 806 nominations, of which transcriptomes of ~750 microbes will be sequenced, assembled, and annotated at NCGR. RNA libraries (300-800 bp) will be sequenced from both ends (2x100-nt reads) on the Illumina Hi-Seq 2000 platform. The first batch of multi-plexed sequencing run is already complete and data analysis is under way. Each assembly will be annotated with protein motifs, sequence similarity search results and GO annotations. All sequence reads, assemblies, annotations and metadata will be deposited with CAMERA (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis). For more details, please visit the project web site MarineMicroEukaryotes.org.

FF0121

Metatranscriptomic Inventory of Rhizosphere Soils in an Arid Grassland Under Global Environmental Change Scenario

Amy Jo Powell[1], Donald O. Natvig[2], Andrea Porras-Alfaro[2,6], Joanna Redfern[2], Miriam Hutchinson[2], Kylea Odenbach[1], Susannah Tringe[3], Edward Kirton[3], Eric Ackerman[1], Blake Simmons[1,4], Scott Collins[2], Robert Sinsabaugh[2], Diego A. Martinez[8], Chris Detter[3,5], Ralph A. Dean[7], Jon Magnuson[9], Randy Berka[10]

Sandia National Laboratories (Albuquerque, NM, USA)[1], University of New Mexico/Sevilleta Long Term Ecological Research program (Albuquerque, NM USA)[2], The Joint Genome Institute (JGI; Walnut Creek, CA, USA)[3], The Joint BioEnergy Institute (JBEI; Emeryville, CA, USA)[4], Los Alamos National Laboratory (Los Alamos, NM, USA)[5], Western Illinois University (Macomb, Illinois, USA)[6], North Carolina State University/Center for Integrated Fungal Research (Raleigh, NC. USA)[7], The Broad Institute (Cambridge, MA, USA)[8], Pacific Northwest National Laboratory[9], Novozymes[10]

We report results of a pilot metagenomic/transcriptomic inventory of rhizosphere soils in a native Chihuahuan Desert grassland dominated by the long-lived perennial C4 grass blue-grama (*Bouteloua gracilis)*. The study site is located at the Sevilleta National Wildlife Refuge (**SEV**) in central New Mexico and is part of the Sevilleta Long-Term Ecological Research (**LTER**) program. The goal of these experiments is to assess microbial community composition and metabolic potential using high throughput sequencing starting with total RNA. We compared control and treatment samples taken from plots that are part of a long-term, multi-factorial, global change experiment. The experimental sample was derived from a pool of subsamples from the rhizospheres of blue-grama plants subjected to three combined treatments: fire, increased nighttime temperatures, and nitrogen addition. The control sample was derived from a pool of subsamples taken from blue-grama plants in an adjacent untreated site. In this pilot experiment we analyzed a sample from combined treatments to maximize potential differences between experimental and control data points. Our results reveal both similarities and compelling differences between experimental and control samples. At the kingdom level, sequences from both samples are dominated by eubacteria, approximately 80% of the total. Among the eukaryotes, fungi were the most highly represented (~58%), followed by sequences from the Viridiplantae (~13%) and Metazoans (~8%); approximately 10% of the eukaryotic reads received no taxonomic assignment. Firm conclusions regarding differences between sample types in the current study cannot be made until we have analyzed replicates. Nevertheless, there are intriguing differences. Most notably, sequences from the cyanobacteria are underrepresented in the experimental sample.

The fungal data from this total-RNA study can be compared with our previous work employing PCR-based analyses of fungal communities in this same grassland. The previous analyses employed fungal-specific primers to the rDNA region and resulted in sequences dominated by Ascomycota (>80%). In contrast with these previous results, in the current study Basidiomycota account for more than 50% of fungal sequences. These findings indicate that current PCR-based methods for characterizing fungal communities are biased against the Basidiomycota. Among the Ascomycota, our results do reveal consistencies across the two different experimental methods (i.e., total RNA vs. fungal-specific PCR) in that both approaches resulted in members of the Pleosporales being the most abundant Ascomycota group.

In summary, our preliminary results indicate that this total-RNA metatranscriptomics approach is a credible tool for studying the effects of environmental change on microbial communities in rhizosphere soils.

FF0122

Semiconductor Sequencing for Life

Andy Felton, Jonathan M. Rothburg, Jason Myers, and Ion Torrent Team

Ion Torrent 7000 Shoreline Court, Suite 201 South San Francisco, CA 94080, USA

Ion Torrent has invented the first device—a new semiconductor chip—capable of directly translating chemical signals into digital information. The first application of this technology is sequencing DNA. The device leverages decades of semiconductor technology advances, and in just a few years has brought the entire design, fabrication and supply chain infrastructure of that industry—a trillion dollar investment—to bear on the challenge of sequencing. The result is Ion semiconductor sequencing, the first commercial sequencing technology that does not use light, and as a result delivers unprecedented speed, scalability and low cost.

All of these benefits are a result of applying a technology that is massively scalable, as proven by Moore's Law, to a task that has traditionally used optics-based solutions, which work in a linear fashion: increasing capacity requires increasing the number of signals that must be read resulting in longer run times, higher capital costs and ever more sophisticated optics. By contrast, Ion Torrent semiconductor technology can provide increases in chip capacity without impacting capital costs or runtime. Ion Torrent sequencing uses only natural (label-free) reagents and takes place in Ion semiconductor microchips that contain sensors which have been fabricated as individual electronic detectors, allowing one sequence read per sensor. We will show how the technology has scaled in just a few months from ~1 million sensors in the first-generation Ion 314 chips to ~7 million sensors in the second-generation Ion 316 chips—all while maintaining the same 1- to 2-hour runtime. We will also demonstrate that Ion semiconductor sequencing provides exceptional accuracy, long read length and scalability on a single, affordable bench-top sequencing platform.

FF0128

CLC bio Assembly and Genome Finishing

Dan Conway and Jannick D. Bendtsen

CLC bio, Inc. 10 Rogers Street, Suite 101, Cambridge, MA 02142, USA

Assembly and genome finishing is becoming increasingly important with the massive amounts of data being generated by next generation sequencing. The increased amounts of sequencing data by new sequencing methods still leaves room for Sanger sequencing data in finishing projects for validation and closure and hybrid data sources can complement each other to generate high quality assemblies.

CLC bio is a commercial software company with a strong focus on Next Generation Sequencing and has a robust portfolio in desktop and enterprise solutions. We here present an integrated software package which adds functionality to the CLC Genomics Workbench and CLC Genomics Server to aid in the process of genome finishing. This module will seamlessly integrate into the CLC Genomics Workbench and will provide additional functionality for aligning contigs to reference, tools to analyze assembly coverage and paired end data. Moreover, the module includes tools for automated primer design, gap closure, contig extension. Splitting of misassembled contigs, manual and automated joining of contigs is also possible.

Upcoming additions to the CLC Genomics workbench include a new packed view of reads, functionality to detect structural variations and CNVs as well as better transcript assemblies, splice site detection and much more.

FF0129

The Cost of Finishing: A Model for Estimating, Tracking, and Controlling Costs Based on Genome Project Standards

Janine C. Detter[1], David Bruce[1], Karen Davenport[1], Kostas Mavrommatis[2], Chris Detter[1]

[1]Genome Science Group, Los Alamos National Laboratory, Los Alamos, NM, USA; [2]Joint Genome Institute, Walnut Creek, CA, USA

In 2009, Chain et al. published Genome Project Standards in a New Era of Sequencing proposing a new set of genome sequence categories of standards including Standard Draft, High-Quality Draft, Improved High-Quality Draft, Noncontiguous Finished, and Finished. The Genome Sciences Group (B-6) at Los Alamos National Laboratory (LANL) has been following these standards in finishing.

LANL's Project Management team works to provide tools for estimating, tracking, and controlling costs while providing the best product for sponsors and collaborators. While a truly Finished project will always be the gold standard in genome finishing, it is important to understand the costs of all products and technologies.

This poster will outline the relative costs of draft sequencing and finishing using Genome Project Standards with Sanger-based finishing. In addition, costs of 3rd Generation (Illumina-Only) finishing is included as LANL looks to incorporate the newest technologies in their best practices.

FF0134

Next Generation Genome Finishing/Improvement of Microbial Genomes

Ahmet Zeytun, Karen Davenport, Olga Chertkov, Linda Meincke, Britney Held, Chris Detter, and Cliff Han

Genome Group, Bioscience Division, Los Alamos National Laboratories, Los Alamos, NM 87505, USA

Whole-genome sequencing is the most powerful approach to understand the genomic diversity of microorganisms, to catch evolution occurring in genomes for millions year, and to identify the basis of pathogenicity of microorganisms by comparing genomic sequences of closely related species. The Genome Group of LANL, as a partner of JGI, has been finishing the microbial genomes to a high standard in a time- and cost-effective fashion through continuously improving its finishing process.

In 2003, when the Sanger is the only sequencing platform than covered genome for ~10x, our group finished a genome in 385 days. After 2005, following introduction of the next sequencing platforms, our group has developed and implemented computational tools to resolve repeats with drafting data, to identify finishing target and request reactions, and automated wet lab processes to generate new sequence data that close gaps. As results, a genome was finished in 206 days in gold standard and last year alone 150 genomes were completed (only three genome were completed in 2003).

Recently, the length of sequence reads and coverage of genome sequenced with Illumina have been significantly improved (120 bp vs. 36 bp and 200-2000x vs. 25-50x) and we believe that such high coverage and longer single and paired-end reads could close most gaps through optimization of assembly programs and implementation of tools for repeat resolution and gap closing. Furthermore, we have ongoing R&D efforts to enrich DNA fragments in the underrepresented regions of genomes resulting in gaps, which will be done with combination of 40-50 genomes in a single massive process. Furthermore we have additional plans to close the single cell genome isolated from metagenomic population.

FF0141

Generating and Analyzing High Quality Draft Assemblies of Large Genomes from Massively Parallel DNA Sequence Data

Aaron Berlin, Terrance P. Shea, Sarah Young, Sean Sykes, David Heiman, Iain MacCallum, Sakina Saif, Carsten Russ, David Jaffe, Bruce Walker, and Chad Nusbaum

The Broad Institute of MIT and Harvard, Cambridge, MA, USA

Massively parallel sequencing (MPS) technologies are revolutionizing genomics by generating large quantities of reads at very low cost. However, MPS data pose challenges to generating high-quality *de novo* assemblies of large, repeat-rich vertebrate genomes both because the reads are short and because the enormous data sets greatly increase the computational resources required for assembly and data analysis. Until now, such assemblies have fallen far short of those obtained with a capillary-based sequencing approach.

We developed a new algorithm for genome assembly from MPS data, ALLPATHS-LG, along with new analysis methods to evaluate assemblies. We applied these to >15 vertebrate genomes sequenced on the Illumina platform. The resulting draft genome assemblies perform well in terms of accuracy, short-range contiguity, long-range connectivity and coverage of the genome. In particular, base accuracy (≥ 99.95%) and scaffold sizes (*e.g.* N50 size = 11.5 Mb for human and 17.4 Mb for mouse) are similar to those obtained with capillary-based sequencing.

Our analysis tools allow us to rapidly evaluate input data quality before assembly, to assess accuracy of output assemblies at the level of bases, contigs and scaffolds, and to investigate any assembly anomalies. The combination of new sequencing technology and new computational methods now make it possible to increase dramatically the *de novo* sequencing and analysis of large genomes.

Work is ongoing to improve closing of gaps and generating consensus of hard-to-sequence genomic regions. For example, in addition to algorithmic developments, we are exploring the potential of Pacific Biosciences data to help with these issues. At current sequencing yields, PacBio already provides a practical approach for small genomes, where sample prep costs dominate. Using a modified version of ALLPATHS-LG, we demonstrate how hybrid assemblies (including Illumina and PacBio data) of bacterial genomes can overcome some assembly challenges, yielding longer scaffolds and capturing more of the genome.

The ALLPATHS-LG program is available at:
http://www.broadinstitute.org/software/allpaths-lg/blog

FF0144

Bacterial Bodiversity and Function in a Cold Desert Ecosystem

Cristina Takacs-Vesbach[1], David Van Horn[1], Lydia Zeglin[1,2], Shannon FitzPatrick[1], Michael Gooseff[3], and John Barrett[4]

[1]Department of Biology University of New Mexico Albuquerque, NM 87131, USA; [2]Department of Crop and Soil Science Oregon State University, Corvallis, OR 97331, USA; [3]Department of Biological Sciences Virginia Tech Blacksburg, VA 24061, USA; [4]Department of Civil & Environmental Engineering Pennsylvania State University, University Park, PA 16802, USA

For many decades the soils of the McMurdo Dry Valleys, Antarctica were thought to be essentially sterile. We now know that this is an ecosystem that is dominated by microorganisms, however, early cultivation efforts largely failed to detect the biodiversity of the region's poorly weathered, low organic carbon soils. Initial surveys of microbial diversity using 16S rRNA gene sequencing has revealed a surprising bacterial richness, including representatives from at least ten different phyla, including a high proportion of unique and rare sequences. Yet, a thorough and exhaustive survey of microbial diversity has not been conducted and little is known about the function of the detected microorganisms. Furthermore, given the low in situ microbial activity and decomposition rates, the question of whether this richness is illusory has been raised. We have conducted an exhaustive survey of the microbial richness, function, and activity of soil bacteria across gradients of moisture and salinity using pyrosequencing of 16S rRNA bacterial tag-encoded FLX amplicons (bTEFAP) and environmental DNA (metagenomics). These data will be used to identify the extent of active bacteria in dry valley soils and elucidate potential microbial function with the ultimate goal of understanding the role of bacteria in cold arid soils.

FF0172

Use of Gel Microdrop Technology to Identify Synergistic Microbial Interactions and to Sequence the Human Microbiome

Fitzsimons MS, Dichosa AEK, Snook JP, Weston LL, Han CS

Los Alamos National Laboratory, Genome Science Group, Los Alamos, NM, USA

Gel microdrops (GMDs) are small agarose-based spheres in which one can encapsulate one or a few cells allowing high-throughput analysis via flow cytometry. We are currently using this technology to identify positive interactions among microorganisms in a complex community and to culture fastidious microorganisms from a single cell for genomic sequencing. Below we present data demonstrating the ability of this technique to identify a known positive interaction between two bacteria in a model bacterial community; we demonstrate that beta-lactamase producing *Escherichia coli* promote the growth of ampicillin sensitive *E. coli* when encapsulated in the same gel microdrop. This is a model interaction, which represents many types of positive interactions where the growth of one microorganism is promoted by a product produced by another. We plan to use this technology in more complex microbial communities such as the human gut where we hope to uncover those microorganisms that promote or inhibit the growth of human pathogens.

We are also using this technology to aid the sequencing of the human microbiome. To achieve this goal we encapsulate single cells within GMDs and incubate them collectively allowing the growth of fastidious microorganisms, which require the presence of other microorganisms in order to grow. The porous nature of the GMDs allows these critical molecules to diffuse within the system. The GMDs are then separated via flow cytometry in order to perform whole genome amplification and genomic sequencing of individual strains and species.

FF0176

Defining the Maize Transcriptome *de novo* Using Deep RNA-Seq

Jeffrey Martin, Stephen Gross, Cindy Choi, Tao Zhang, Erika Lindquist, Chia-Lin Wei, Zhong Wang

DOE Joint Genome Institute, Walnut Creek, California, USA

de novo assembly of the transcriptome is crucial for functional genomics studies with bioenergy crops, since many of them lack high quality reference genomes. In a previous study we successfully *de novo* assembled simple eukaryote transcriptomes exclusively from short Illumina RNA-Seq data. However, extensive alternative splicing, present in most of the higher eukaryotes, poses a significant challenge for current short read assembly processes. Gene duplications retained from ancestral polyploidization events, common in plant genomes, also present challenges in assembly of individual transcripts from distinct genes. Here we present preliminary results which greatly improved the assembly of the maize transcriptome, using combined experimental and informatics strategies to resolve transcript variants.

We chose the maize transcriptome as a test case since the reference genome can be used for assessing the quality of the assembled transcript variants. Our experimental strategies include ultra-deep sequencing and multiple libraries with various insert lengths. We generated 127 gigabases (1.1 billion reads) of both stranded and non-stranded RNA-Seq data by sequencing three libraries made from a seedling mRNA sample. The first library was a 180bp insert library; the second was a 250bp tight-insert library and was sequenced 2x151bp and subsequently the two read pairs were joined to form 250bp reads; while the third library was a 500bp tight-insert library to provide long-range connectivity. We further improved our published *Rnnotator* pipeline to assemble the reads from all libraries into transcripts. By comparing these *de novo* assembled transcripts to the reference-based gene models we evaluated the performance of our transcriptome annotation strategy for its accuracy, completeness and resolution of transcript variants and transcripts from duplicate genes.

In summary, we expect our strategies will be generically applicable to many plant transcriptome studies. The maize gene models derived in this study can serve as a valuable resource for the maize research community.

FF0177

Expanding Targeted Sequencing Applications with SureSelect^{XT}

Scott Happe[1], Bahram Arezi[3], Angelica Giuffre[1], Carlos Pabón-Peña[2], Swati Joshi[1], Joseph Ong[1], Harini Ravi[1], Marc Visitacion[2], Barbara Novak[2], Micah Hamady[2], Francisco Useche[2], Eric Lin[2], Doug Roberts[2], and Emily LeProust[2]

Agilent Technologies, Cedar Creek, TX[1]; Santa Clara, CA[2]; La Jolla, CA[3], USA

Even as the cost of whole-genome sequencing declines, targeted approaches simplify analysis and increase throughput necessary to perform large studies. Agilent's SureSelect portfolio is expanding to address these needs. SureSelect is based on hybrid-capture (Gnirke, et al.) using long, high-quality biotinylated RNA probes complementary to regions of interest. The SureSelect^{XT} system now combines library preparation with target enrichment and indexing/barcoding for Illumina and SOLiD platforms. SureSelect is also compatible with the 454 platform using third-party library preparation kits. We show performance data on a growing number of pre-defined catalog probe sets, including comprehensive All-Exon designs for human and mouse, Kinome, and new novel content. Both custom designs (<0.2 to 34 Mb) and catalog products show high performance as measured by specificity, coverage depth, uniformity, and library complexity. We also demonstrate the effectiveness of spiking custom content into catalog products to create a personalized All-Exon or Kinome "Plus" design. In addition to DNA applications, we illustrate the use of SureSelect with RNA-Seq to identify novel transcript variants within defined regions of interest. For optimal sequencing capacity, we introduce enhanced methods for indexing/barcoding samples for multiplexed sequencing in a single analysis lane. Using the integrated SureSelect^{XT} system, we accurately detect SNPs and indels of varying sizes while maintaining allelic balance. Based on our results, we propose a novel approach for variant discovery: using SureSelect catalog designs for variant discovery, followed by the design of focused custom libraries for SNP validation and region profiling in larger cohorts. To increase throughput, simply workflow, and enhance reliability, we illustrate the performance of SureSelect^{XT} on the Automated Bravo Liquid Handling Platform. Significant time savings and process efficiencies are realized while maintaining high performance characteristics. Altogether, our data demonstrate a highly robust, accurate, and flexible SNP discovery/validation pipeline for efficient, cost-effective analysis of genomic variation.

FF0182

Avadis NGS Analysis Software

Jean Jasinski

Agilent Technologies, Cedar Creek, TX[1]; Santa Clara, CA[2]; La Jolla, CA[3], USA

Avadis NGS analysis software is capable of analyzing aligned data (SAM, BAM, or Eland output) from any next generation sequencing platform. Workflows include ChIP-Seq for transcriptional regulation studies, RNA-Seq for digital gene expression, SNP identification and discovery of novel genes/exons/fusions/splice variants, and DNA-Seq for SNP, Indel, and structural variant analyses. The GeneSpring NGS module will incorporate RNA-Seq and DNA-Seq workflows plus special filters and QC statistics for Agilent's SureSelect Target Enrichment system. The built-in Genome Browser, GO Analysis, Pathway and Network tools help you assign biological meaning to your data. An overview and how to obtain a trial license will be presented.

FF0183

Use of the Argus™ Optical Mapping System to Validate Finished Microbial Genomes

T. K. Wagner, E. A. Meudt, E.B. Zentz

OpGen Inc., Gaithersburg, MD 20878, USA

Whole-genome DNA sequencing of microbes is becoming easier, and sequencing throughput has reached 200 Gb per run thereby making it less expensive. However, while the quantity of DNA sequence being produced and published is increasing exponentially, there is no strict quality or validation requirement for submitting a finished whole-genome assembly to GenBank or a peer-reviewed journal.

Whole-genome *de novo* assembled Optical Maps were generated using the Argus™ Optical Mapping System from 16 microbes ordered from ATCC that had finished whole-genomes deposited within GenBank. Twelve of the 16 finished genomes were published within peer-reviewed journals. The finished whole-genome DNA sequence for each microbe were obtained from GenBank, imported into MapSolver™ software to generate an *in silico* restriction map of each genome, and compared to the *de novo* assembled Optical Maps.

Multiple finished whole-genomes were identified to contain discrepancies due to either sequence assembly errors or the divergence of the ATCC isolate. Some of the discrepancies in the finished DNA sequences included large-scale differences such as missing an entire ~1.3 Mb region of genetic content, missing an entire ~375 Kb repetitive region, and inverting a ~1.9 Mb region.

The data in this study support the ability of the Argus™ Optical Mapping System to be used as a sequence independent technology to validate finished whole-genomes before submission to GenBank and peer-reviewed journals. The types and scale of the discrepancies identified suggest that increasing the quantity of DNA sequencing alone cannot eliminate all discrepancies and cannot guarantee that a finished genome is completely accurate. Unvalidated finished whole-genomes could miss coding regions during gene discovery and introduce errors into microarrays, PCR primers, or probes designed from them. Furthermore, unvalidated finished whole-genomes could also introduce errors into subsequent resequencing projects. The conclusions of this study highlight the risk of using unvalidated finished whole-genomes for further analysis.

FF0189

Exploring Applications for the PacBio RS in the Sequencing Workflow at Los Alamos National Laboratory

Krista Reitenga and the Genome Science Group (B-6), Bioscience Division

Los Alamos National Laboratory, Los Alamos, NM, USA

The Genome Science Group in the Bioscience Division at Los Alamos National Laboratory (LANL) works with a variety of LANL-internal and external collaborators and sponsors to address a variety of genomic challenges, ranging from sequencing microbial and eukaryotic genomes, to single cell genomics, to RNAseq and metagenomics. Depending on the specifics of the project, we utilize capillary Sanger sequencing, 454 pyrosequencing, Illumina (GA or HiSeq) sequencing, or more recently, PacBio single molecule sequencing. Since the installation of our PacBio RS instrument in March, we have worked to evaluate strategies to take advantage of the PacBio's unique technology and integrate this new data type to improve current multi-next-gen platform workflows. One of LANL's sequencing strengths is the closure of microbial genomes, a process for which we hope to evaluate and capitalize on methods including "strobe" sequencing and generation of long reads to both fill and scaffold gaps and repeats between contigs in sequence assemblies. Our first attempt at using PacBio long reads on repeat-rich genomes for the resolution of repetitive gaps suggests that this strategy may indeed help close gaps, although improvement of both chemistry and informatic processing will be required. In a DTRA-sponsored exercise designed to simulate a potential biothreat outbreak, we have also tested the capability of PacBio to help identify and characterize target pathogens present at low-levels within complex samples (air filter and blood). Our experience using PacBio to sequence these metagenomic samples suggests that the greatest advantage of the PacBio over other next-gen platforms is the speed with which sequence data can be produced to rapidly help identify target organisms. In order to make precise strain determinations of targets present at very low abundances within a sample with the PacBio, the trade-off in speed for throughput and readlength for accuracy may require optimization.

FF0198

Mammalian Y Chromosome Finishing Projects

Shannon Dugan-Rocha[1], Yan Ding[1], Alicia Hawes[1], Christian J. Buhay[1], Ziad Khan[1], Michael E. Holder[1], Qiaoyan Wang[1], Wen Liu[1], Jennifer Hughes[2], Helen Skaletsky[2], Donna Villasana[1], Lynne Nazereth[1], David Page[2], Donna M. Muzny[1] and Richard A. Gibbs[1]

[1]Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030, USA; [2]Howard Hughes Medical Institute, Whitehead Institute, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

The BCM-HGSC is currently working in collaboration with the Page laboratory at the Whitehead Institute on a project which specifically targets mammalian Y chromosomes including Macaca mulatta (rhesus monkey), Bos taurus (bovine), and Rattus norvegicus (rat) for additional sequencing and sequence finishing. Due to the highly repetitive nature of the Y chromosome, traditional Sanger BAC based sequencing with a large insert library of 5-8kb has been employed. To date, approximately 444 bovine BACs, 62 macaque BACs, and 757 rat BACs have been prepped, sequenced and assembled in the BCM-HGSC pipeline. Of these, 415 bovine BACs, 62 macaque BACs and 604 rat BACs totaling over 165Mb of unique sequence have been completed at the highest quality or “gold standard”.

Assembly and finishing issues stemming from large repeats/duplicated regions have been aided by clone selection from the smaller insert Amplicon Express BAC library.

The difficulty of finishing these BACs is largely dependent on the number of copies of these repeats present in each clone, and must be manually sorted using read pair information. Closure of all remaining gaps and low quality regions is accomplished using direct sequencing of BAC DNA amplified with the GE TempliPhi Sequence Resolver Kit. Transposon bombing or specialized shotgun libraries have also been applied for difficult regions such as larger tandem repeats or GC rich regions. Finally, each completed BAC is validated by at least two restriction digestions to confirm assembly contiguity and accuracy. Additional direct BAC sequencing data and results of these strategies will be presented.

FF0199

Method of Detecting DNA Methylation Using Hidden Markov Modeling of Single-Molecule, Real-Time (SMRT™) DNA Sequencing Data.

Adam English, Yi Han, Sandy Lee, Tittu Matthews, Donna Muzny, Jeffrey Reid and Richard A Gibbs

[1]Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030, USA

DNA methylation has profound biological consequences and plays an important role in gene expression, and aberrant DNA methylation has been associated with disease states such as cancer. With the recent advances in single molecule DNA sequencing, the detection of methylation signals in large numbers of long, contiguous DNA sequences is now possible with Pacific Biosciences' Single Molecule Real Time (SMRT) sequencing data.

SMRT sequencing technology uses DNA polymerase and fluorescently modified nucleotides to directly sequence a single DNA molecule in real-time. This process involves measuring the emission intensity from incorporated fluorophores over time. During SMRT sequencing, there is a detectible difference between the incorporation dynamics for methylated bases compared to unmethylated bases. Thus, the methylation status of each nucleotide can be assessed by measurement of the arrival times and durations of the resulting fluorescence pulses.

Previous work (Flusberg, *et. al.* 2010) has detected DNA methylation in SMRT sequencing data by comparing the inter-pulse duration (IPD) between methylated sequences and an unmethylated control sequence. Our approach uses this IPD relationship to build and train a Hidden Markov Model (HMM). This HMM does not require control sequences to find indicators of the probability of methylation state. By training our model on artificially methylated synthesized sequences with a wide variety of properties, we capture the tendencies of SMRT data for methylation while building a model that avoids local maxima in finding the most probable state path. Additionally, by taking advantage of multi-core computing, we can run multiple HMM decoding algorithms to find the most probable methylation state path without a substantial increase in computation time. Detecting methylation alongside the genomic information contained in a sequencing experiment allows epigenetic factors to be incorporated into hypothesis testing.

BA Flusberg, DR Webster, JH Lee, KJ Travers, EC Olivares, TA Clark, J Korlach, and SW Turner. "Direct detection of DNA methylation during single-molecule, real-time sequencing. " (2010) Nature Methods 7:461–465.

FF0202

High Efficiency 40 kb Paired-end Sequencing for Next Generation Platforms

Chengcang Wu, Ronald Godiska, Rosa Ye, Svetlana Jasinovica, Megan Wagner, David Mead

Lucigen Corp., Middleton WI 53562, USA

Next-generation sequencing (NGS) platforms are fundamentally altering genetic and genomic research by providing massive amounts of data in a low-cost, high-throughput format. The main drawback of existing technologies is the short sequence read lengths they produce. New tools that bridge the gap between massively parallel short read sequencing technologies (35-500 bases) and the need for both medium (~40 kb) and large (100 kb or larger) scaffolds to assemble a genome are clearly needed. Existing methods to generate large paired end sequence reads are extremely inefficient and produce only 20 kb or smaller paired ends. As a result, de novo assembly of daunting genomes is still impossible and resequencing and assembly of human genomes is a significant challenge when analyzing complex genomic regions and copy number variations. A new "front end" to NGS is being developed to construct paired-end libraries from medium and large randomly sheared DNA fragments in the 40-300 kb range. As the first result, we have successfully developed a fosmid based paired-end sequencing strategy for NGS. Our results indicate that the success rate of ~40 kb paired ends is about 85% using this tool. Details of the methods and results will be presented.

FF0206

Engineered DNA Polymerases Enable Decreased Amplification Bias and Improved Coverage in Illumina Sequencing Workflows

Eric van der Walt[1], Gavin J. Rush[1], Jacob Kitzman[2], Liesl Noach[1], Zayed Albertyn[3], Colin Hercus[3], Charlie Lee[2], Jay Shendure[2], Michael A. Quail[4], John F. Foskett III[1], Paul J. McEwan[1]

[1]Kapa Biosystems, 600 West Cummings Park, Suite 2250, Woburn, MA 01801, USA; [2]University of Washington, Department of Genome Sciences, Foege Building S-250, Box 355065, 3720 15th Ave. NE, Seattle WA 98195, USA; [3]Novocraft, C-23A-5, Two Square, Section 19, 46300 Petaling Jaya, Selangor, Malaysia; [4]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton Cambs CB10 1SA, United Kingdom

High fidelity PCR is used to selectively enrich library fragments carrying appropriate adaptor sequences and to amplify the amount of DNA prior to sequencing. During PCR enrichment of libraries, a subset of library molecules are amplified with reduced efficiencies, introducing bias and resulting in uneven sequence coverage. GC content is known to be an important factor in next-generation sequencing library amplification bias, and different PCR enzymes and buffer formulations are likely to show individual strengths and weaknesses in this regard. Furthermore, such biases – along with other artifacts such as PCR-induced errors, adaptor dimers, PCR duplicates, and chimeras – are exacerbated by over-amplification, while under-amplification results in insufficient yields.

We assessed the performance of a variety of DNA polymerases for library amplification in terms of GC bias and uniformity of coverage depth. Here we present a high fidelity, real-time PCR method for rapid enrichment of DNA library templates. The benefits of this approach include: 1) automatable workflows, 2) built-in quality metrics for each enriched DNA library, eliminating expensive and time-consuming post-enrichment gel electrophoresis, 3) precise control over the number of PCR cycles required for optimal amplification, 4) a quality control metric for identifying inconsistencies in library preparation, and 5) seamless integration with qPCR-based library quantification.

FF0206

Novel Improvements to the Illumina TruSeq Indexed Library Construction, Amplification and Quantification Protocols for Optimized Multiplexed Sequencing

Eric van der Walt[1], Gavin J. Rush[1], Jacob Kitzman[2], Liesl Noach[1], Zayed Albertyn[3], Colin Hercus[3], Charlie Lee[2], Jay Shendure[2], John F. Foskett III[1], Paul J. McEwan[1]

[1]Kapa Biosystems, 600 West Cummings Park, Suite 2250, Woburn, MA 01801, USA; [2]University of Washington, Department of Genome Sciences, Foege Building S-250, Box 355065, 3720 15th Ave. NE, Seattle WA 98195, USA; [3]Novocraft, C-23A-5, Two Square, Section 19, 46300 Petaling Jaya, Selangor, Malaysia

Dramatic improvements in commercial Next Generation Sequencing (NGS) platforms have resulted in spectacular reductions in the cost-per-base of DNA sequencing. Until recently, the primary focus for innovation has been on improvements to the core sequencing technologies, with optimization of sample preparation playing a secondary role. The exponential gains in sequencing capacity have simultaneously led to higher sample throughput, placing increasing emphasis on the importance of improved library construction protocols for multiplexed sample sequencing. While major commercial NGS systems all require the construction of similar libraries via analogous workflows, some protocols and/or reagents offer significant advantages over others, and end-users must choose among numerous alternative methods and reagents for sample preparation. We re-sequenced the *Staphylococcus aureus*, *Escherichia coli*, and *Mycobacterium tuberculosis* genomes to compare the standard Illumina TruSeq sample preparation reagents and workflow with a number of innovative improvements including: alternative library preparation reagents and protocols; automated fragment size selection; real-time library amplification; amplification-free sequencing; and accurate qPCR library quantification for sample pooling and multiplexed sequencing.

FF0207

From Pathogen Screening to Nucleic Acid Output: Design of a High-Throughput BSL3 Screening & Extraction System

J. Foster Harris[1], Tracy Erkkila[1], Craig Blackhart[1], Alexander Roth[2], Lee Borenstein[2], Angela Gray[3], Chris Pacheco[3], Ira Hoffman[3], Hilary Godwin[2], J. Chris Detter[1]

[1]Los Alamos National Laboratory, Los Alamos, NM, [2]UCLA School of Public Health, Los Angeles, CA, [3]HighRes Biosolutions, Woburn, MA, USA

Global pathogen surveillance is one of the most important approaches to combat spread of disease. High throughput solutions allow for more samples to be identified, sequenced, and characterized in less time than sample characterization by hand. Therefore, we have designed a robotic approach to pathogen screening and nucleic acid extraction for organisms that need to be handled at the BSL2 and BSL3 level. The Screening & Extraction System uses magnetic bead extraction chemistry, Real-time PCR, and a liquid handling system to extract samples, confirm and quantify the presence of pathogens, and reformat extracted samples for input into downstream sequencing processes. The system is enclosed in a custom biosafety cabinet that allows for manual access where needed, while providing maximum personnel protection. Sample tracking and data management are achieved with a customized version of StarLIMS. Status updates, notification, reporting, and data analysis can be achieved without entering the biosafety containment. The robotic system it is currently being built and will be ready for validation in Fall 2011. Once validated, the system will provide the capacity and tools to quickly respond to pandemics, extract highly pathogenic organisms, and provide nucleic acid input for large-scale sequencing projects. This research and response capacity far exceeds what is available today. Information that can be identified and tracked using this versatile robotic system include pathogen mutation rate, markers for virulence and contagion, and markers of drug resistance. The rapid, large data output also provides foundational information for biothreat forensics and attribution analyses. This system will be implemented at UCLA in their BSL3 Global Bio Lab. Future generations of this technology could be placed around the world for public health protection and surveillance.

FF0215

Discovery of Novel sRNAs in Yersinia pestis

Nick Beckloff

Los Alamos National Lab, Los Alamos, NM, USA

The discovery of small RNAs (sRNAs) and their important function in gene regulation has dramatically increased due in part to the implementation of next generation sequencing (NGS) technologies. sRNAs are widespread effectors of post-transcriptional gene regulation in bacteria. Currently minimal information exists on the expression of sRNAs in *Yersinia pestis*, the causative agent of plague. We are interested in their potential role in pathogenicity, thus are exploring their expression using RNA-Seq under conditions mimicking its natural lifecycle within human host (37°C and infection of macrophages) and within the flea vector (26°C). Illumina libraries were prepared from total RNA isolated from *Y. pestis* infected macrophages and sequenced using the Illumina HiSeq2000. Several of the libraries were prepared from enriched mRNA and size selected sRNA protocols. The reads were mapped to reference genomes and the effect of the various protocols ascertained based on mapping statistics. The sRNAs were predicted using a computational model consisting of analysis of intergenic read coverage and gene and transcript expression data. Here, we will describe the project, and resulting identification of putative novel sRNAs that are expressed under conditions mimicking mammalian host infection.

FF0226

Whole-Genome Tiled-Amplicon Deep-Sequencing of HIV-1 in the First Weeks, Months, and Years of Infection

Will Fischer, Peter Hraber, Shuyi Wang, Hui Li, Tanmoy Bhattacharya, Thomas Leitner, Mira Dimitrijevic, Nilu Goonetilleke, Andrew McMichael, Brandon Keele, Barton Haynes, Beatrice Hahn, George Shaw, Bette Korber

University of Missouri, USA

By deep-sequencing samples from a single subject, we obtained a broad picture of HIV-1 intra-patient evolution from acute infection until mid-chronic infection (3.25 years post-infection). We used 35 overlapping amplicons to cover essentially the entire HIV-1 genome. Five sample intervals ranged from 18 days to 3.25 years following onset of symptoms (DFOSx); genome-wide coverage was ~10,000X for each time point.

We developed analytical ‘cleaning’ methods to address characteristic 454 sequencing errors, and a processing pipeline for rapid data set assembly.

Extensive genome coverage allowed us to compute genome-wide changes in positional base frequencies as well as rates of reversion toward the B-clade consensus. Comparison of the 454 data with earlier single-genome sequences extend and in some cases altered previous interpretations. The availability of immunological data enabled quantitative assessment of immune-driven escape at both T-cell and neutralizing antibody epitopes.

FF0236

Isilon Delivers Cost-Effective, Scalable Data Storage in Support of TGen's Biomedical Research Initiatives

Alex Long

Isilon, Seattle, WA, USA

Challenge
In the course of their investigative work, TGen's scientists and researchers generate enormous data sets from precise sequencing. The volume of genetic data quickly breached the limitations of TGen's original data storage system, forcing TGen's IT managers to expend excessive amounts of time dealing with inefficient data movement among storage silos, performance problems and file system disruption.

Solution
TGen needed a storage solution that required less management resources and allowed TGen to repurpose its previous system as a backup archive. TGen to unified its workflow onto a single, highly scalable, cost effective Isilon IQ 36NL cluster, accelerating time-to-results for its mission-critical research. By using data compression technology from Isilon partner Ocarina, TGen is now using its previous storage area network (SAN) system combined with network file system (NFS) servers as the backup target for its Isilon IQ cluster.

Poster Session Notes

Poster Session Notes

06/02/2011 - Thursday				
Time	**Type**	**Abstract #**	Title	**Speaker**
7:30 - 8:30am	**Breakfast**	x	**Santa Fe Breakfast Buffet**	**Sponsored by NEB**
8:30 - 8:45	Intro	X	Welcome Back from DOE	**Dan Drell**
x	Session Chair	x	Session Chairs	Chair - Alla Lapidus Chair - Mike Fitzgerald
8:45 - 9:30	**Keynote**	**FF0107**	**Science of Sequencing Process Development for High Technical Replicate R2 Analyses**	**Dr. Nels Olson**
9:30 - 9:50	Speaker 1	**FF0146**	Long Reads and Hybrid Assemblies	**Jim Knight**
9:50 -10:10	Speaker 2	**FF0056**	Consed and Phaster for Next-gen Sequencing	**David Gordon**
10:10 - 10:40	**Break**	x	**Beverages and snacks provided**	**Sponsored by Isilon**
10:40 -11:00	Speaker 3	**FF0092**	Of Parrots and Pathogens – Hybrid Assembly and Benchmarking of Long and Short Read Sequencing	**Adam Phillippy**
11:00 - 11:20	Speaker 4	**FF0214**	BioPig: Hadoop-based Analytic Toolkit for Next-Generation Sequence Data	**Zhong Wang**
11:20 - 11:40	Speaker 5	**FF0142**	Automated High-Throughput de novo Genome Assembly of Microbial Genomes Using Illumina Data	**Bruce Walker**
11:40 - 12:00	Speaker 6	**FF0072**	Efficient Graph Based Assembly of Short-Read Sequences on a Hybrid-Core Architecture	**George Vacek**
12:00 - 1:20pm	**Lunch**	x	**New Mexican Lunch Buffet**	**Sponsored by PacBio**
x	**Session Chair**	x	**PacBio Assembly Workshop Session**	**Chair - Steve Turner** **Chair - Donna Muzny**
1:20 - 3:10pm	**Hybrid De Novo Assembly Workshop (15 min each) with discussion panel**	**FF0204**	Using AHA (A Hybrid Assembler) for genome finishing of V. cholerae	**Aaron Klammer**
		FF0200	Leveraging long single molecule PacBio reads for de novo genome assembly	**Todd Michael**
		FF0141	The Long and Short of Microbial Hybrid Assembly Generation	**Aaron Berlin**
		FF0197	Initial Results and Future Direction in Hybrid Assembly with Illumina and Pacific Biosciences Data	**Richard McCombie**
		FF0216	Using Pacbio for Sample Characterization under Real-time Conditions	**Patrick Chain**
x	Session Chair	x	Session Chairs	Chair - Donna Muzny Chair - Johar Ali
3:10 - 3:30	**Break**	x	**Beverages and snacks provided**	**Sponsored by CLC**
3:30 - 5:30pm	**Tech Time Talks (15 min each)**	**FF0040**	Creating Probe Maps with Solid-State Nanopores	**John Oliver**
		FF0182	Avadis NGS Analysis Software	**Jean Jasinski**
		FF0205	High-Speed, High-Reliability Focus Optimization for Genomics Applications	**Scott Jordan**
		FF0206	Novel Improvements to the Illumina TruSeq Indexed Library Construction, Amplification and Quantification Protocols for Optimized Multiplexed Sequencing	**Eric van der Walt**
		FF0177	Accelerating the Impact of Targeted Resequencing with SureSelectXT	**Scott Happe**
		FF0186	A New Method for Long Range Scaffolding of Large Complex Genomes using the Argus™ Optical Mapping System	**Nick Xiao**
		FF0231	Isilon Delivers Cost-Effective, Scalable Data Storage in Support of TGen's Biomedical Research Initiatives	**Chris Blessington**
		FF0077	Advancements in Focused Acoustics for Use in NGS Sample Preparation	**Hamid Khoja**
5:45 - 7:45	**Happy Hour**	x	**Happy Hour at Cowgirls Cafe - Sponsored by LifeTech - Map Will be Provided**	**Sponsored by LifeTech**
7:45 - bedtime	**on your own**	x	**Dinner and night on your own - enjoy**	x

Poster Session Notes

NOTES

Speaker Presentations (June 2nd)

Abstracts are in order of presentation according to Agenda

FF0107

Keynote

Nels Olson

Johns Hopkins University, Baltimore, MD, USA

Why Sequencing Changed Scientific Pursuit and Made, At Least Me, Feel Dumb Over and Over Again

Science of Sequencing Process Development for High Technical Replicate R2 Analyses

In this talk I will present advantages and caveats in the implementation of: Optics and Imaging, Dyes, Surfaces (both a challenge and an advantage), Parameter control and not knowing what the measured metrics should be and finally, Intellectual property; all as they apply to the biotechnology field of study.

NOTES

FF0146

Long Reads and Hybrid Assemblies

Jim Knight

Roche Applied Science, 9115 Hauge Road, Indianapolis, IN 46250, USA

This talk presents updates to the GS FLX system and the Newbler assembler to improve their ability to generate cost effective de novo assemblies. This includes the new GS FLX+ system upgrade, capable of generating Sanger-like readlengths (700-800bp), the new repeat region and diploid genome handling algorithms in the Newbler assembler, along with assembler support for FASTQ input files, BAM output files and hybrid assemblies of 454 and Illumina HiSeq reads.

FF0056

Consed and Phaster for Next-gen Sequencing

David Gordon and Phil Green

University of Washington, Seattle, WA, USA

Our sequence editor consed can now handle assemblies with millions of reads of mixed types (Illumina, 454, Sanger). New features include:

1) automatically generating a list of potential joins. Each potential join is annotated with whether the join is recommended or why it is not recommended
2) automatically *making* all of the recommended joins

3) taking an assembly with depth of coverage in the thousands or more, and generating a small assembly with a representative list of reads. This is useful if the assembly must be manipulated in consed.

4) improved automated fixing of the consensus at contig ends (especially useful when new reads are added)

5) fixing repeated base errors in the consensus

6) exporting scaffolds, each as a single sequence

7) batch changing the consensus (you supply a list of the positions to change and the bases to change it). Useful, for example, for changing vector to X prior to submission

8) batch complementing a list of contigs (useful for fixing contig orientation in scaffolds)

9) searching for highly discrepant regions

10) user-friendly graphical features helpful for high depth of coverage
'Phaster' performs ultrafast gapped quality-aware alignment of arbitrary-length reads to a SNP-annotated reference genome, using a simple word-frequency based strategy. In our tests it is twice as fast and superior in sensitivity to bwa (Li and Durbin, 2009). Phaster can find gapped and partial matches, allowing in particular the detection of segmental variants and spliced (RNASeq) alignments.

FF0092

Of Parrots and Pathogens – Hybrid Assembly and Benchmarking of Long and Short Read Sequencing

Adam M. Phillippy[1,2]

[1]National Biodefense Analysis and Countermeasures Center, Frederick, MD, USA; [2]Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD, USA

The emergence of high-throughput, short-read sequencing has sparked a renewed interest in genome assembly. Traditional overlap-based algorithms have been abandoned in favor de Bruijn approaches, which are better equipped to handle the computational complexity of short read data. However, with short reads now reaching 100 bp lengths, and affordable long reads over 1 Kbp imminent, overlap-based assembly is poised to regain favor. The introduction of affordable long reads will admit a new sequencing paradigm, where long reads can provide a low-coverage assembly backbone to which high-quality short reads can be added to boost coverage and accuracy. Overlap-based algorithms are best suited for this combination.

The promise of this approach is supported by two examples of hybrid assembly, utilizing 454 and Illumina sequencing. The first example involves the recent sequencing of the 1.1 Gbp Budgerigar (parakeet) genome to a depth of 11x using 454 and 7x using Illumina. The resulting 18x hybrid assembly is of quality rivaling the Sanger and BAC-based Zebra Finch genome, with contig and scaffold N50 sizes of 55.6 Kbp and 11.2 Mbp respectively. Assembly continuity is critical for success of the project, which aims to reveal the regulatory traits of vocal learning. The second example involves the sequencing of a *Bacillus anthracis* strain isolated from an intravenous drug user in Scotland, who was likely infected via contaminated heroin. To precisely identify the contaminating strain, genome sequences were generated using both 454 and Illumina at high depth. The hybrid assembly exceeds the quality of either technology in isolation, and the resulting high-quality consensus was vital to the construction of strain---specific SNP assays for microbial forensic screening. The success of these two projects is contrasted with benchmarking and validation results for the current state-of-the-art short read assemblers Velvet, SOAPdenovo, and ALLPATHS on both eukaryotic and microbial reference genomes.

FF0214

BioPig: Hadoop-based Analytic Toolkit for Next-Generation Sequence Data

Karan Bhatia, Zhong Wang

DOE Joint Genomics Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Here we introduce the BioPig sequence analysis toolkit that scales to next-generation sequence data and computation. BioPig is built upon the Apache's Hadoop map-reduce and Pig query language to provide an easily accessible data parallel programming environment. We illustrate how BioPig scripts concisely implement common sequence analysis tasks, ranging from simple ones like k-mer statistics to more complex ones like metagenomic gene discovery. For each application illustrated, we provide baseline (unoptimized) application performance as the datasizes scale and provide some analysis as to the bottlenecks at the system layer. Where possible, we provide performance comparisons with alternative methods.

FF0142

Automated High-Throughput *de novo* Genome Assembly of Microbial Genomes Using Illumina Data

Bruce J. Walker, Sarah Young, Terrance P. Shea, Aaron Berlin, Sean Sykes, David Heiman, Iain MacCallum, Sante Gnerre, Filipe Ribeiro, Dariusz Przybylski, Carsten Russ, David Jaffe, and Chad Nusbaum

The Broad Institute, Genome Sequencing and Analysis Program, Cambridge, MA, USA

Short-read sequencing technologies now enable users to generate massive amounts of DNA sequence data from large numbers of organisms at low cost. Now, our challenge is to rapidly and efficiently convert this mountain of raw data into high quality *de novo* genome assemblies. To address this, we have implemented an automated genome assembly pipeline currently capable of assembling and analyzing over one hundred microbial genomes per week from Illumina sequence data. The pipeline operates in three stages: 1) analysis of input reads; 2) *de novo* assembly using ALLPATHS-LG and/or other assembly software; 3) post-process analysis of output assemblies for quality assurance.

Analysis of input reads: We apply a set of automated analysis steps to all raw read data prior to assembly. We begin by generating basic quality metrics, including distribution of base quality scores, to ensure that the input data meet minimum standards. We then apply techniques that tell us more about the genome we are assembling, including GC content profiling to identify potential sequencing bias issues and contamination, and K-mer profiling to compute estimates of genome size, repeat content, and true sequencing error rates. We have also developed visualization techniques to quickly identify problems in the sequence data, reducing the manual effort required to diagnose poor quality data post-assembly.

De novo assembly: Information gained from the read analysis helps set the parameters used in the assembly process (For example, input read coverage may be altered for assembly of genomes with extreme GC content). We have implemented ALLPATH-LGs for current production use, after careful evaluation of the range of available tools. However, the pipeline is flexible, and capable of operating with most existing short read assembly tools. We will continually evaluate the state-of-the-art through direct comparisons of assemblers on standard input data.

Assembly analysis: We assess and compare quality of assemblies with a series of metrics, including counts, averages, and N50 lengths of both contigs and scaffolds, as well as gap statistics. Qualitative analysis of contigs in terms of GC composition, sequence coverage, and similarity to other sequenced genomes can help identify contamination and other anomalies. Additionally, for genomes with an available reference, we are able to assess assembly completeness, assembled base quality, and mis-assemblies. This automated data generation has reduced the manual effort of assessing assemblies to mere minutes, allowing us to dedicate human work to diagnosing problem assemblies.

As an initial pilot of this pipeline we sequenced and assembled twenty-four bacterial genomes over a range of size and base composition. Assembly output quality exceeds the results obtained by sequencing by 454 and assembling with Newbler, and importantly, at a fraction of the cost.

This project has been funded by the NIAID/NIH/DHHS under Contract No. HHSN272200900018C.

FF0072

Efficient Graph Based Assembly of Short-Read Sequences on a Hybrid-Core Architecture

George Vacek[1], Alex Sczyrba[2,3]

[1]Convey Computer Corporation, Richardson, TX, USA; [2]DOE Joint Genome Institute, Walnut Creek, CA, USA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Advanced architectures can deliver dramatically increased throughput for genomics and proteomics applications, reducing time-to-completion in some cases from days to minutes, and tackling problem sizes that are unattainable with commodity computing systems. One such architecture, hybrid-core computing, marries a traditional x86 environment with a reconfigurable coprocessor, based on field programmable gate array (FPGA) technology. In addition to higher throughput, the increased performance from such application-specific platforms fundamentally improves research quality by allowing more accurate, previously impractical approaches.

This presentation will discuss the approach used by Convey's de Bruijn Graph Constructor for short-read, *de novo* assembly. Bioinformatics applications that have random access patterns to large memory spaces, such as graph-based algorithms, experience memory performance limitations on cache-based x86 servers. Convey's highly parallel memory subsystem allows application-specific logic to simultaneously access 8192 individual words in memory, significantly increasing effective memory bandwidth over cache-based memory systems. Many algorithms, such as Velvet and other de Bruijn graph based, short-read, *de novo* assemblers, greatly benefit from this type of memory architecture. Furthermore, small data type operations (four nucleotides can be represented in two bits) make more efficient use of logic gates than the data types dictated by conventional programming models.

Results for a variety of research problems will be presented comparing the performance of Convey's Graph Constructor with Velvet and other assemblers on both synthetic and real data. Memory usage savings and run time reduction will be shown for various data sets with different sizes, from small microbial and fungal genomes to plant and vertebrate genomes to very large metagenomes. For larger datasets, performance improvements facilitate multiple runs with different kmer lengths. This allows the optimal kmer length to be chosen, resulting in higher quality assemblies.

.

NOTES

NOTES

Lunch

12:00 – 1:20pm

Sponsored by

Notes

Hybrid Assembly Workshop

FF0204

Using AHA (A Hybrid Assembler) for Genome Finishing of V. cholera

Aaron Klammer

Pacific Biosciences, 1380 Willow Rd, Menlo Park, CA 94025, USA

FF0200

Leveraging Long Single Molecule PacBio Reads for de novo Genome Assembly

Todd Michael

Monsanto Company, Chesterfield, MO, USA

FF0141

The Long and Short of Microbial Hybrid Assembly Generation

Aaron Berlin, Terrance P. Shea, Sarah Young, Sean Sykes, David Heiman, Iain MacCallum, Sakina Saif, Carsten Russ, David Jaffe, Bruce Walker, and Chad Nusbaum

Broad Institute of MIT and Harvard, Cambridge, MA, USA

We are evaluating Pacific Biosciences' (PacBio) long read data for use in generating high quality bacterial genome assemblies. Massively parallel sequencing (MPS) technologies generate large quantities of short reads at very low cost, however, these data pose challenges to fulfilling the demand for "perfect" assemblies. Although the current short-read assembly standard is very high, these assemblies are not perfect and can only be improved through the costly manual finishing process. This approach is impractical for the large number of microbial genomes currently being sequenced. We are looking to PacBio to fill the gap between assembly and manual finishing on our quest to generate perfect assemblies.

PacBio provides a practical approach for small genomes, where sample prep costs dominate. PacBio and MPS have complementary properties, making them ideal partners for hybrid assembly. We evaluated two methods for hybrid assembly using a series of microbes, selected to represent a variety of issues that tend to plague sequencing and assembly, from extreme GC levels to large amounts of low complexity regions: *Bifidobacterium bifidum*, *Escherichia coli*, *Klebsiella oxytoca*, *Rhodobacter sphaeroides*, *Eubacterium*, *Streptococcus pneumoniae* and *Plasmodium falciparum*. First, we used the tools provided by PacBio in the SMRTanalysis package. Second, we modified a version of the ALLPATHS-LG assembler to utilize the PacBio data. Using these methods we demonstrate how a hybrid assembly approach can overcome some bacterial genome assembly challenges, yielding longer scaffolds and capturing more of the genome than the current standard.

In addition to the ongoing work to create the best assemblies using hybrid assembly, we are also evaluating *de novo* assembly of Fosmids for targeted finishing. We demonstrate *de novo* PacBio assemblies of Fosmids containing difficult regions of *Streptomyces roseosporus* and *Mycobacterium tuberculosis*.

ALLPATHS-LG is available at www.broadinstitute.org/software/allpaths-lg/blog.

FF0197

Initial Results and Future Direction in Hybrid Assembly with Illumina and Pacific Biosciences Data

Dick McCombie

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

Short read sequencing platforms provide a vast amount of data very inexpensively. However, these data yield far less than optimal assemblies due to the very short read lengths. We have been doing initial studies mixing Pacific Biosciences reads with Illumina reads to improve assembly quality. We will present results which show enhanced assembly by using PacBio reads. We will also present future directions that we will pursue to improve the impact of hybrid data types on *de novo* assemblies.

FF0216

Using Pacbio for Sample Characterization under Real-time Conditions

Patrick Chain

Los Alamos National Lab, Los Alamos, NM, USA

Panel Discussion Notes

FF0040

Creating Probe Maps with Solid-State Nanopores

John S. Oliver, Maryam Jouzi, Peter Ianakiev, Heidi Geiser, Benjamin Stone, Debra Dederich, Dona Hevroni, Hsu-Yi Lee

NABsys Inc., 60 Clifford St., Providence, RI 02903, USA

NABsys is developing a single-molecule, electronic DNA sequencing method that utilizes solid-state nanopores. In our proposed implementation, the nanopore locates the positions of oligonucleotide probes that have hybridized to long DNA fragments. The fragments are assembled *de novo* into chromosome length maps based on the pattern of hybridization on each fragment. Collections of these maps are used to reconstruct the sequence of genomes.

We will present data that demonstrates the detection of probes on target DNA and the determination of distance between probes with positional error that is low enough to allow for mapping and sequence reconstruction of genome length sequences. The detectors have sufficient resolution to determine the variation in the velocity of translocation of DNA constructs during translocation. Assembly algorithms have been developed that allow data to be rapidly assembled into contigs with N50 sizes in excess of 100 Mb. The results validate the utility of nanopores as detectors for the determination of the position of hybridization of oligonucleotide probes to target DNA.

FF0182

Avadis NGS Analysis Software

Jean Jasinski

Agilent Technologies, Cedar Creek, TX[1]; Santa Clara, CA[2]; La Jolla, CA[3], USA

Avadis NGS analysis software is capable of analyzing aligned data (SAM, BAM, or Eland output) from any next generation sequencing platform. Workflows include ChIP-Seq for transcriptional regulation studies, RNA-Seq for digital gene expression, SNP identification and discovery of novel genes/exons/fusions/splice variants, and DNA-Seq for SNP, Indel, and structural variant analyses. The GeneSpring NGS module will incorporate RNA-Seq and DNA-Seq workflows plus special filters and QC statistics for Agilent's SureSelect Target Enrichment system. The built-in Genome Browser, GO Analysis, Pathway and Network tools help you assign biological meaning to your data. An overview and how to obtain a trial license will be presented.

FF0205

High-Speed, High-Reliability Focus Optimization for Genomics Applications

Scott Jordan

Director, Nano/Automation Technologies, PI (Physik Instrumente) L.P., Auburn, MA

High-speed, high reliability focus optimization plays an important role in high-throughput sequencing by capturing better images faster and enabling real-time tracking approaches. Piezo-actuator driven focus mechanisms offer the high-speed and high reliability required for next generation sequencing. Piezo actuator driven mechanisms are capable of sub-millisecond response and can keep pace with high speed focus detection technologies. Piezo driven mechanisms can improve the throughput economics in sequencing as they have in applications like semiconductor lithography, semiconductor metrology, and optical metrology.

Here we review:

- Four types of piezo actuators
- Reliability and speed capability of piezo actuator driven focus mechanisms
- Focus detection technologies often used with piezo mechanisms
- Examples of piezo deployment for high speed focus in semiconductor and other industries
- Key metrics for evaluating and selecting focusing technologies

FF0206

Novel Improvements to the Illumina TruSeq Indexed Library Construction, Amplification and Quantification Protocols for Optimized Multiplexed Sequencing

Eric van der Walt[1], Gavin J. Rush[1], Jacob Kitzman[2], Liesl Noach[1], Zayed Albertyn[3], Colin Hercus[3], Charlie Lee[2], Jay Shendure[2], John F. Foskett III[1], Paul J. McEwan[1]

[1]Kapa Biosystems, 600 West Cummings Park, Suite 2250, Woburn, MA 01801, USA; [2]University of Washington, Department of Genome Sciences, Foege Building S-250, Box 355065, 3720 15th Ave. NE, Seattle WA 98195, USA; [3]Novocraft, C-23A-5, Two Square, Section 19, 46300 Petaling Jaya, Selangor, Malaysia

Dramatic improvements in commercial Next Generation Sequencing (NGS) platforms have resulted in spectacular reductions in the cost-per-base of DNA sequencing. Until recently, the primary focus for innovation has been on improvements to the core sequencing technologies, with optimization of sample preparation playing a secondary role. The exponential gains in sequencing capacity have simultaneously led to higher sample throughput, placing increasing emphasis on the importance of improved library construction protocols for multiplexed sample sequencing. While major commercial NGS systems all require the construction of similar libraries via analogous workflows, some protocols and/or reagents offer significant advantages over others, and end-users must choose among numerous alternative methods and reagents for sample preparation. We re-sequenced the *Staphylococcus aureus*, *Escherichia coli*, and *Mycobacterium tuberculosis* genomes to compare the standard Illumina TruSeq sample preparation reagents and workflow with a number of innovative improvements including: alternative library preparation reagents and protocols; automated fragment size selection; real-time library amplification; amplification-free sequencing; and accurate qPCR library quantification for sample pooling and multiplexed sequencing.

FF0177

Accelerating the Impact of Targeted Resequencing with SureSelect^{XT}

Scott Happe[1], Bahram Arezi[3], Angelica Giuffre[1], Carlos Pabón-Peña[2], Swati Joshi[1], Joseph Ong[1], Harini Ravi[1], Marc Visitacion[2], Barbara Novak[2], Micah Hamady[2], Francisco Useche[2], Eric Lin[2], Doug Roberts[2], and Emily LeProust[2]

Agilent Technologies, Cedar Creek, TX[1]; Santa Clara, CA[2]; La Jolla, CA[3], USA

Targeted resequencing has become state-of-the-art for discovery of genome variation. The Agilent SureSelect in-solution hybrid capture methodology has identified variants associated with Mendelian and complex disease in over 50 peer-reviewed publications. As researchers scale up investigations to validate initial hits and screen larger populations, the SureSelect portfolio is expanding to meet the needs of these studies. Available for the Illumina, SOLiD, and 454 NGS platforms, SureSelect is a highly robust, customizable, scalable system that focuses analyses on specific genomic loci. Agilent has also introduced SureSelect^{XT} for Illumina and SOLiD, which combines gDNA prep, library prep, and target enrichment reagents in one complete kit. This presentation will focus on the latest enhancements in the SureSelect portfolio. We highlight the utility of SureSelect using pre-designed catalog products targeting cancer gene sets, sequences encoding the kinome, and human and mouse All Exon content. In addition, user-defined custom designs up to 34 Mb can be easily developed using the Agilent eArray software, manufactured on-demand, and even combined with catalog content using the Plus option. For challenging targets, we discuss approaches for performance enhancement, including probe design and sample processing adjustments. We propose using SureSelect catalog designs to uncover candidate variants, followed by design of focused custom libraries for variant validation and region profiling. By pooling many samples together per lane/slide, SureSelect^{XT} kits for Illumina and SOLiD multiplexing enable validation across large sample cohorts with substantial cost savings. Post target enrichment, accurate sample pooling is facilitated by the Agilent Bioanalyzer and QPCR NGS Library Quantification kits to ensure equal representation across samples. Further efficiencies are realized using the Bravo Automated Liquid Handling Platform for parallel preparation of multiplexed libraries. Through these advancements, the SureSelect^{XT} system is accelerating genetic variation discovery by providing a means to efficiently validate findings and expand the scope of research studies.

.

FF0186

A New Method for Long Range Scaffolding of Large Complex Genomes Using the Argus™ Optical Mapping System

Nianqing (Nick) Xiao, Ryan N. Ptashkin, Thomas S. Anantharaman, Bin Zhu, Deacon Sweeney, John K. Henkhaus.

OpGen, Inc., 708 Quince Orchard Rd, Gaithersburg, MD 20878, USA

Despite the continued improvements in DNA sequencing technologies, whole genome sequencing of large complex organisms remains a significant bioinformatics challenge, particularly when trying to discern the order and orientation of the hundreds or thousands of sequencing scaffolds typical in these projects. Optical Mapping is a single molecule technology that generates *de novo*, ordered, high-resolution restriction maps. Whole genome Optical Maps are assembled from collections of single molecule restriction maps, and are routinely used in comparative genomics, sequence assembly and sequence finishing of microorganisms. Recent improvements in data throughput and data quality suggest that the Argus™ Optical Mapping System can be practically applied to large complex genomes such as plants and animals. Proof-of-concept studies have recently been conducted on human and other animal genomes with the size of up to 3 Giga-bases. Millions of single molecule restriction maps typically of the length from 250 kb up to 1 Mb were generated rapidly with Argus Optical Mapping System in an automated fashion. Using an iterative extension-joining approach, we have recently developed an analytical pipeline that is able to join sequence scaffolds that are separated by a few hundred kilo-bases. The relative position and orientation of the scaffolds are inferred based on alignment between optical maps and sequence scaffolds. Using simulated gaps (10 ~ 200 kb) on real sequence scaffolds as testing data, we have achieved gap joining rates of over 90%, with error rate of less than 2%. These results indicate that the Argus Optical Mapping System can be used to facilitate sequence improvement and finishing of large complex genomes, reducing the bioinformatics burden and improving the overall sequence quality.

FF0231

Isilon Delivers Cost-Effective, Scalable Data Storage in Support of TGen's Biomedical Research Initiatives

Chris Blessington

Isilon, Seattle, WA, USA

Challenge

In the course of their investigative work, TGen's scientists and researchers generate enormous data sets from precise sequencing. The volume of genetic data quickly breached the limitations of TGen's original data storage system, forcing TGen's IT managers to expend excessive amounts of time dealing with inefficient data movement among storage silos, performance problems and file system disruption.

Solution

TGen needed a storage solution that required less management resources and allowed TGen to repurpose its previous system as a backup archive. TGen to unified its workflow onto a single, highly scalable, cost effective Isilon IQ 36NL cluster, accelerating time-to-results for its mission-critical research. By using data compression technology from Isilon partner Ocarina, TGen is now using its previous storage area network (SAN) system combined with network file system (NFS) servers as the backup target for its Isilon IQ cluster.

FF0077

Advancements in Focused Acoustics for Use in NGS Sample Preparation

H. Khoja, G. Durin, and J. Laugharn

Covaris Inc., 14 Gill Street, Unit H, Woburn, MA 01801, USA

Rapidly advancing sequencing technologies have continued the demand for advances in nucleic acid fragmentation technologies. The increase in sensitivity and throughput of these platforms has placed a greater demand on the accuracy, reproducibility, and throughput of the of the library preparation steps which begins with shearing of DNA, RNA, or chromatin.
Technologies such as nebulization, unfocused sonication, hydrodynamic shearing, and enzymatic digestion have remained virtually unchanged in decades. The limited utility they had with first generation, and some next generation platforms are quickly being exhausted. The increase in sensitivity and coverage of NGS platforms has exposed and amplified some of these inherent limitations which include thermal and sequence specific biased fragmentation, thermal degradation, precious sample loss, user-dependent variability, or automation incompatibility.

In contrast, the Covaris Adaptive Focused Acoustics (AFA) has advanced with sequencing technology advancements. AFA's engineered closed vessel, non-contact, isothermal technology has since become capable of offering a wide range of fragment sizes from 100bp to 5kb, and is now considered the gold standard for DNA fragmentation used with all NGS platforms in many labs including all the large sequencing centers around the world.

In this meeting we will present Covaris' technological and application advancements and supporting data. These include our award-winning LE220 instrument for ultra-high throughput labs capable of parallel processing (e.g., 96 samples to 300bp in less than 15 minutes), and our new "220" generation electronics with finer power resolution and broader dynamic range, and our new updated software. We will also introduce Covaris AFA-certified reagent kits. For example, chromatin shearing protocols for cultured cells and tissues with unprecedented control for ChIP-Seq applications enabling higher retention of epitope integrity. Our simple and reproducible RNA (total and mRNA) shearing protocol is also a beneficial replacement for the heat and chemical cleavage methods currently utilized in RNA-Seq library preparation.

Tech Time Notes

Tech Time Notes

Happy Hour (x2)

Cowgirls BBQ

505.982.2565 319 S. Guadalupe St Santa Fe, NM

See map on next page!

5:45pm – 7:45pm, June 2nd

Drink tickets (margaritas, beer, sodas) will be provided

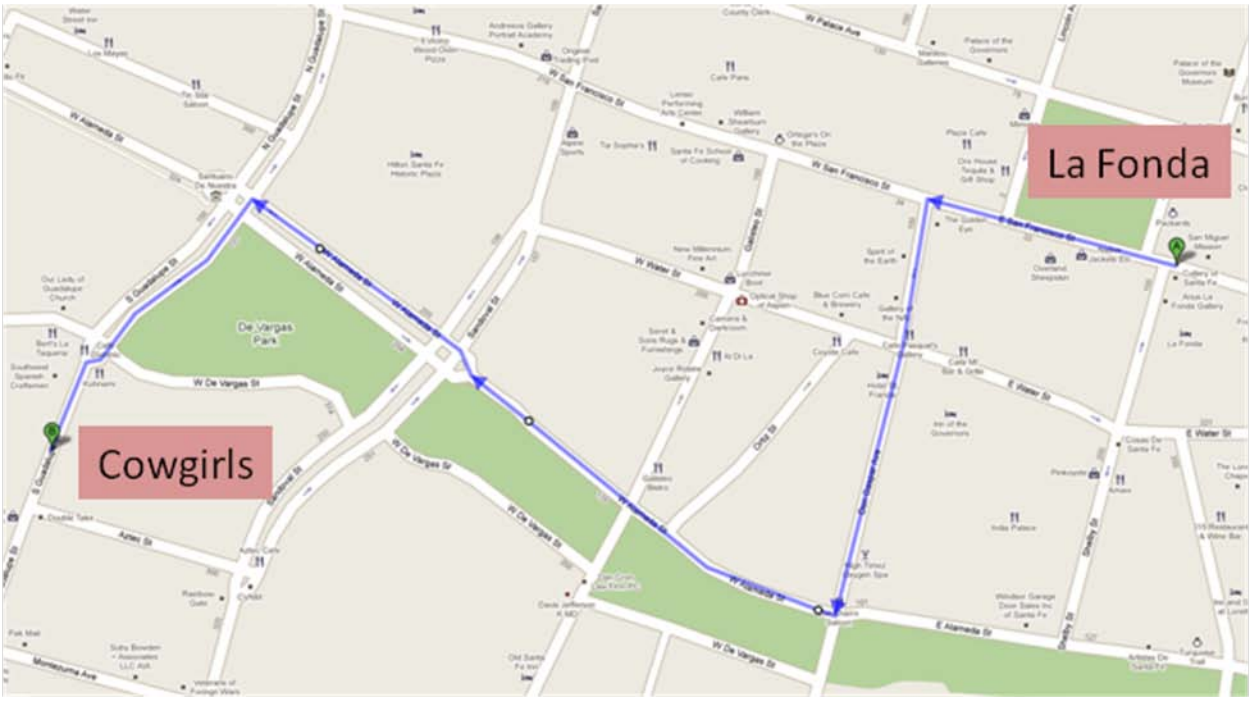
Sponsored by Life Technologies

Enjoy!!!

life technologies™

Map to Cowgirls BBQ

505.982.2565 319 S. Guadalupe St Santa Fe, NM



Total Walking Distance

0.5 miles, 10 minutes

The Legend...

Many years ago, when the cattle roamed free and Cowpokes and Cowgirls rode the range, a sassy young Cowgirl figured out that she could have as much fun smokin' meats and baking fine confections as she could bustin' broncs and rounding up outlaws. So she pulled into the fine bustling city of Santa Fe and noticed that nobody in town was making Barbeque the way she learned out on the range. She built herself a Texas-style barbecue pit and soon enough the sweet and pungent scent of mesquite smoke was wafting down Guadalupe street and within no time at all folks from far and near were lining up for heaping portions of tender mesquite-smoked brisket, ribs and chicken. Never one to sit on her laurels, our intrepid Cowgirl figured out that all those folks chowing down on her now-famous BBQ need something to wash it all down with. Remembering a long-forgotten recipe from the fabled beaches of Mexico, she began making the now-legendary Frozen Margarita and the rest, as we say, is History. Before you could say "Tequila!" the musicians were out playing on the Cowgirl Patio and the party was in full swing.

06/03/2011 - Friday				
Time	**Type**	**Abstract #**	Title	**Speaker**
7:30 - 8:30am	**Breakfast**	**x**	**Breakfast Buffet**	**Sponsored by NEB**
8:30 - 8:45	Intro	**x**	Welcome Back	**Chris Detter**
x	Session Chair	**x**	Session Chairs	Chair - Patrick Chain Chair - Darren Grafham
8:45 - 9:30	**Keynote**	**FF0032**	**The Study of the Human Microbiome**	**Dr. Granger Sutton**
9:30 – 9:50	Speaker 1	**FF0165**	Dining on Driftwood: a genomic view of a wood-eating bacterial endosymbiosis in the shipworm Bankia setacea	**Dan Distel**
9:50 – 10:10	Speaker 2	**FF0101**	High throughput single cell genomics pipeline for microbiology	**Ramunas Stepanauskas**
10:10 – 10:30	**Break**	**x**	**Beverages and snacks provided**	**Sponsored by OpGen**
10:30 – 10:50	Speaker 3	**FF0131**	Manipulating bacterial growth for single cell genomics	**Armand Dichosa**
10:50 – 11:10	Speaker 4	**FF0050**	Preparation of Nucleic Acid Libraries for Next Generation Sequencing with an Automated Molecular Biology Platform	**Ken Patel**
11:10 – 11:30	Speaker 5	**FF0116**	**Application of deep sequencing to diversity library analysis and selection**	**Andrew Bradbury**
11:30 – 11:50	Speaker 6	**FF0031**	What's a referenceable reference?	**Todd Smith**
11:50 – 12:10	Speaker 7	**FF0156**	Finishing in the era of NGS: A user's perspective	**Kostas Mavrommatis**
12:10 - 12:30	**Closing Discussions**	**x**	**Closing Discussions - discuss next year's meeting**	**Chair - Chris Detter**
12:30 - 2:00pm	**Lunch & Close of meeting**	**x**	**La Fiesta Plaza Lunch**	**Sponsored by Agilent**

NOTES

Speaker Presentations (June 3rd)

Abstracts are in order of presentation according to Agenda

FF0032

Keynote

Granger Sutton

Professor, The J. Craig Venter Institute, Rockville, Maryland, USA

The Study of the Human Microbiome

The human body is host to a multitude of microbial species and communities that are estimated to outnumber the number of host somatic cells. These microbial species have been implicated in various health and disease conditions, and we are now in a position to address the populations and the diversity that exists within these populations. This has been made possible through advances in metagenomic studies whereby we are now able to generate sequence data from entire environments without first using a culturing step. Metagenomics tools and techniques have evolved primarily from approaches that have been developed to study the diversity in the oceans, and have been made more accessible through the newer generation of sequencing technologies.

The studies on characterizing the human microbiome have gone from studies focused on one or two individuals to large-scale worldwide initiatives focused on major disorders and involving hundreds of participants. Questions focus on whether or not there is a core human microbiome, what are the correlations if any between microbial species and various disease conditions, and what are the technological and bioinformatics needs in order support the advances in data generation. The NIH funded Human Microbiome Project (HMP) includes the sequencing of at least 3000 bacterial reference genomes, and significant metagenomic sequencing to characterize the microbial communities from 15-18 body sites in 300 consented individuals. It is clear that the advent of metagenomics holds significant promise for increasing our understanding of many microbial diseases associated with the human body, inclusive of those that are yet to be characterized.

NOTES

FF0165

Dining on Driftwood: A Genomic View of a Wood-Eating Bacterial Endosymbiosis in the Shipworm *Bankia setacea*.

Distel, D.L.[1*], Fung, J.[1], Sharp, K.[1], Henrissat, B.[2], Altamia, M.[3], Lamkin, E.[4], O'Neil, C.[4], Benner, J.[4], Malmstrom, R.[5], Lee, J.[5], Tringe, S.[5] and Woyke, T.[5]

[1]Ocean Genome Legacy, Ipswich, MA, USA; [2]CNRS & Université de la Méditerranée,

Marseille, France; [3]University of the Philippines, Marine Science Institute, Manila, Philippines; [4]Division of Chemical Biology, New England Biolabs, Ipswich, MA, USA; [5]Department of Energy, Joint Genome Institute, Walnut Creek, CA, USA

Shipworms, wormlike wood-boring marine clams of the family *Teredinidae,* are the most prolific consumers of wood in marine environments and are responsible for billions of dollars in damage to wooden ships, piers, and fishing equipment. They are thought to employ a system of wood (lignocellulose) digestion that differs anatomically, functionally, and phylogenetically from all others described to date. The most obvious distinction between the shipworm system and those of termites, ruminants and other previously described lignocellulose-degraders, is that shipworms contain few microbial cells in their digestive tract. Instead, they harbor dense populations of intracellular bacterial endosymbionts within specialized cells of their gills, an organ located far from the gut. It has been proposed that enzymes encoded in and synthesized by the genomes of these gill symbionts are transported to the gut where they contribute to lignocellulose degradation. We have examined the lignocellulolytic system in the shipworm *Bankia setacea* using a combination of highly parallel genomic and metagenomic sequencing and proteomic (2D LC MS/MS) analysis. We provide evidence (1) that the shipworm gill symbiont community, though phylogenetically simple, encodes a rich and diverse variety of lignocellulose-active proteins, including many that are structurally novel, and (2) that a subset of these proteins are selectively transported to the shipworm gut. Thus, the shipworm system appears to be a natural analog of industrial biomass conversion systems that employ separate enzyme synthesis and saccharification. The simplicity of this system, as compared to those of termites or ruminants, makes it highly amenable to genomic and proteomic analyses and genetic manipulation and therefore a potentially informative model for the bioenergy industry.

FF0101

High Throughput Single Cell Genomics Pipeline for Microbiology

Ramunas Stepanauskas

Bigelow Laboratory for Ocean Sciences, 180 McKown Point Road, West Boothbay Harbor, ME, 04575, USA

The vast majority of microbial taxa remains uncultured and is therefore inaccessible to classical microbiology methods. Microbial community DNA analyses, such as metagenomics, is effective for culture-independent gene discovery but has limited capacity to reconstruct discrete genomes or organismal interactions. The promise of DNA sequencing from individual cells is to circumvent these methodological limitations and to focus genetic studies on the most fundamental level of biological organization.
Single cell genomics relies on the physical separation of individual cells, followed by their lysis, whole genome amplification, and subsequent DNA sequencing. Bigelow Laboratory for Ocean Sciences pioneered the development of single cell genomics technology and established the first shared-user Single Cell Genomics Center (SCGC), which integrates fluorescence-activated cell sorting and high-throughput molecular biology tools. During its first year in operation, SCGC contributed to cutting-edge research at over 20 organizations around the globe. Over 150,000 individual cells have been analyzed by SCGC so far, providing unique access to genomic DNA, without cultivation biases, from microorganisms representing over 60 phyla of bacteria, archaea and protists. Projects performed at SCGC span genomic studies of the uncultured prokaryotes, protists and viruses from marine, freshwater, subsurface, organismal and other environments. Research highlights include: identification of novel, ubiquitous chemoautotrophs in the dark ocean; deciphering trophic interactions of uncultured marine protists; and identification of predominant photoheterotrophs in freshwater environments. Thus, single cell genomics is emerging as a transformative research approach in diverse areas of microbiology, including ecology, evolution, and biotechnological applications.

FF0131

Manipulating Bacterial Growth for Single Cell Genomics

Armand E K Dichosa, Michael S Fitzsimons, Lea L Weston, Lara G Preteska, Jeremy P Snook, Chien-Chi Lo, Xiaojing Zhang, Wei Gu, Kim McMurry, Lance D. Green, Patrick S Chain, J Chris Detter, and Cliff Han

Bioscience Division: B-6 Genome Science – JGI, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

Genomic amplification can be accomplished through multiple displacement amplification (MDA), an isothermal reaction involving the Φ29 DNA polymerase and a circular chromosomal template. This process has allowed researchers to yield micrograms of genomic amplicons from a single chromosome copy. As such, MDA with current approaches to single cell isolation and much sequencing efforts have made it possible to determine bacterial/archael phylotypes and metagenomic profiles from an environmental sample *in situ* without the need for cultivation. However, inherent issues surrounding MDA when using a single chromosomal template make it difficult to obtain an accurate genomic profile. In particular, low genomic coverage and gaps observed in the assembled genome are due to uneven amplification and biases. To complete the genomic assembly, additional genomic template and labor are needed.

Our prior investigations have shown that increasing the number of chromosomal template greatly reduces the observed gaps in the assembled genome, thereby producing a more complete genomic profile. However, isolating an adequate number of the exact species/strain from an environmental sample without cultivation is in itself difficult. Consequently, we hypothesized that simultaneously inhibiting bacterial cells from completing cytokinesis while maintaining viability will phenotypically result in cells that are larger-than-normal and possess at least two complete copies of its chromosome. To test our hypothesis, we used PC190723 to prevent *Bacillus subtilis* ATCC 6633 from completing cytokinesis. PC190723 is an antimicrobial compound known to inhibit FtsZ, the bacterial and euryarchaeal protein that is largely implicated in cell division. Cytographic data via flow cytometry have shown that *B. subtilis* responds to PC190723 treatment with increased cell size, while microscopy evidence verifies the treated cells to be at least double in size to the untreated controls. Our qPCR analyses of sorted cells quantified more genomic content and four times less amplification bias from the inhibited cells. Ultimately, *de novo* genomic assemblies suggest that a single polyploid *B. subtilis* cell contributes 17% more genomic coverage than an untreated cell. Our study implies that similar FtsZ-inhibitors, whether singly or in combination, can be used to induce artificial polyploidy in a microbial community, and, when utilized with high-throughput, FACS-based analyses, a more complete genomic and/or metagenomic profile can be achieved with greater efficiency.

FF0050

Preparation of Nucleic Acid Libraries for Next Generation Sequencing with an Automated Molecular Biology Platform

Kamlesh D. Patel, Ph.D., Hanyoup Kim, Michael S. Bartsch, and Ronald F. Renzi

Sandia National Laboratories, Livermore, CA, USA

While DNA sequencing technology is advancing at an unprecedented rate, sample preparation technology still relies primarily on manual bench-top processes, which are slow, labor-intensive, inefficient and often inconsistent. Automation of sample preparation using microfluidic techniques is well-suited to address these limitations. We have designed, fabricated, and characterized a digital microfluidic (DMF) platform to function as a central hub for interfacing multiple lab-on-a-chip sample processing modules towards automating the preparation of clinically-derived DNA samples for next gen sequencing (NGS). The automated molecular biology platform (AMB) is designed to interface directly with NGS to detect unknown pathogens by enriching informative nucleic acids sequences (those derived from the pathogen) and suppressing background DNA (those from the host) to maximize the sensitivity of state-of-the-art NGS. The AMB platform will be able to carry out a diverse series of benchtop-like steps at a scale adapted to handling small, but precious, samples for DNA manipulations, but with far greater speed and efficiency than at the benchtop.

We will present our recent developments on the core architecture of the AMB platform, the DMF central hub, and demonstrate its flexibility in coupling droplet-based microfluidics with continuous-flow microchannel devices to prepare DNA samples for NGS. The strength of combining these two different, but complementary, fluid processing methods enables the manipulation of nanograms to picograms of DNA with precise temporal and spatial control. We will discuss our results for collecting fractions of nanogram amounts of normalized DNA in discrete 1-uL droplets on the DMF device. Additionally, we will also present the integration of magnetic beads for clean-up and concentrate the DNA. Fragmented DNA is analyzed in real-time with microchip-based gel electrophoresis separation interfaced to the sample droplet for the correct size distribution for eventual cluster generation and high-throughput sequencing to discover the pathogen by its genomic sequence.

FF0116

Application of Deep Sequencing to Diversity Library Analysis and Selection

Sara D'Angelo, Fortunato Ferrara, Nileena Velappan, Leslie Naranjo,
Tiziano Gaiotto, Csaba Kiss & Andrew Bradbury

Los Alamos National Lab, Los Alamos, NM, USA

The construction and selection of clones with desirable properties from in vitro display libraries of antibodies or protein domains is usually carried out blind. The diversity of such libraries is traditionally determined by counting the number of clones obtained after ligation and transformation, while the progression of selections is assessed by counting the number of surviving clones after selection and testing a small number (usually 96). The availability of deep sequencing, particularly 454, now allows the direct analysis of selections as assessed by the enrichment of clone sequences. The ranking of selected clones on the basis of their sequencing frequency reveals that many of the clones most frequently sequenced are not necessarily discovered by testing a microtiter plate full of clones. Sequencing information can be used to isolate corresponding clones by inverse PCR. This represents a new application for deep sequencing.

FF0031

What's a Referenceable Reference?

Todd Smith[1], Jeffrey Rosenfeld[2], Christopher Mason[3]

[1]Geospiza Inc. Seattle, WA 98119, USA; [2]Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA; [3]Weill Cornell Medical College, New York, NY 10021, USA

The goal behind investing time and money into finishing genomes to high levels of completeness and accuracy is that they will serve as a reference sequences for future research. Reference data are used as a standard to measure sequence variation, genomic structure, and study gene expression in microarray and DNA sequencing assays. The depth and quality of information that can be gained from such analyses is a direct function of the quality of the reference sequence and level of annotation. However, finishing genomes is expensive, arduous work. Moreover, in the light of what we are learning about genome and species complexity, it is worthwhile asking the question whether a single reference sequence is the best standard of comparison in genomics studies.

The human genome reference, for example, is well characterized, annotated, and represents a considerable investment. Despite these efforts, it is well understood that many gaps exist in even the most recent versions (hg19, build 37) [1], and many groups still use the previous version (hg18, build 36). Additionally, data emerging from the 1000 Genomes Project, Complete Genomics, and others have demonstrated that the variation between individual genomes is far greater than previously thought. This extreme variability has implications for genotyping microarrays, deep sequencing analysis, and other methods that rely on a single reference genome. Hence, we have analyzed several commonly used genomics tools that are based on the concept of a standard reference sequence, and have found that their underlying assumptions are incorrect. In light of these results, the time has come to question the utility and universality of single genome reference sequences and evaluate how to best understand and interpret genomics data in ways that take a high level of variability into account.

1. Kidd, JM. *et. al.* Characterization of Missing Human Genome Sequences and Copy-number Polymorphic Insertions. Nat Methods. 2010 May: 7(5): 365–371.

FF0156

Finishing in the Era of NGS: A User's Perspective

Kostas Mavrommatis[1], Land M[2], Brettin T[2], Quest D[2], Clum A[1], Goodwin L[3], Lapidus A[1], Copeland A[1], Woyke T[1], Cottingham R[2], Kyrpides NC[1]

[1]DOE Joint Genome Institute, Walnut Creek, CA, USA; [2]Oak Ridge National Laboratory, Oak Ridge, TN, USA; [3]Los Alamos National Laboratory, Bioscience Division, Los Alamos, NM, USA

Scientific value of the assembled genomes depends on the quality of sequencing itself as well as the quality of the downstream processes (assembly and annotation), which in their turn depend on the limitations of the bioinformatics tools used to perform these steps. We compared the Draft and Finished versions of 134 microbial genomes sequenced using different combinations of sequencing technologies during the last 7 years at the DOE-JGI and evaluated the quality of the assembly as well as the quality of the final genome annotation. The observations on different sequencing technologies and their effects in the downstream processes, as well as future directions will be discussed.

Discussion Notes

Discussion Notes

2011 Attendees

FF #	Name		Affiliation	email
FF0001	Chris	Detter	Los Alamos National Laboratory (LANL)	cdetter@lanl.gov
FF0002	Joann	Mudge	National Center for Genome Resources (NCGR)	jm@ncgr.org
FF0003	Tim	Hunkapiller	Discovery Bio	tim@discoverybio.com
FF0004	David	Bruce	Los Alamos National Laboratory (LANL)	dbruce@lanl.gov
FF0005	Ernie	Retzel	National Center for Genome Resources (NCGR)	efr@ncgr.org
FF0006	Fiona	Stewart	New England Biolabs	stewart@neb.com
FF0007	Take	Ogawa	RainDance Technologies, Inc.	OGAWAT@raindancetech.com
FF0008	Chrisinta	Chiu	RainDance Technologies, Inc.	chiuC@raindancetech.com
FF0009	Keith	Brown	RainDance Technologies, Inc.	brownK@raindancetech.com
FF0010	Michael	Rhodes	Life Technologies	Michael.Rhodes@lifetech.com
FF0011	John	Havens	Integrated DNA Technologies	jhavens@idtdna.com
FF0012	Karen	Davenport	Los Alamos National Laboratory (LANL)	kwdavenport@lanl.gov
FF0013	Nan	Sauer	Los Alamos National Laboratory (LANL)	nsauer@lanl.gov
FF0014	Alfredo Lopez	De Leon	Novozymes, Inc.	ALLO@novozymes.com
FF0015	Teri	Rambo Mueller	Roche Applied Science	teri.mueller@roche.com
FF0016	Ian	Watson	Defense Threat Reduction Agency (DTRA)	Ian.Watson@dtra.mil
FF0017	Gary	Qiao	Defense Threat Reduction Agency (DTRA)	Guilin.Qiao@dtra.mil
FF0018	Ken	Taylor	Integrated DNA Technologies	ktaylor@idtdna.com
FF0019	Isaac	Meek	Caliper Life Sciences	Isaac.Meek@caliperls.com
FF0020	Tara	Bennink	EdgeBio	Tbennink@edgebio.com
FF0021	Kim	Bishop-Lilly	Naval Medical Research Center	kim.bishop-lilly@med.navy.mil
FF0022	Shanmuga	Sozhamannan	Naval Medical Research Center	shanmuga.sozhamannan@med.navy.mil
FF0023	Malin	Young	Sandia National Laboratories	mmyoung@sandia.gov
FF0024	Ron	Walters	Pacific Northwest National Laboratory (PNL)	ra.walters@pnl.gov
FF0025	Scott	Geib	USDA - ARS	Scott.Geib@ARS.USDA.gov
FF0026	Mike	Lafferty	Life Technologies	Mike.Lafferty@lifetech.com
FF0027	Thomas	Walk	USDA - ARS	tom.walk@ars.usda.gov
FF0028	Cynthia	Hendrickson	New England Biolabs	hendrickson@neb.com
FF0029	Keven	Stevens	Integrated DNA Technologies	kstevens@idtdna.com
FF0030	Helen	Cui	Los Alamos National Laboratory (LANL)	hhcui@lanl.gov
FF0031	Todd	Smith	Geospiza, Inc.	todd@geospiza.com
FF0032	Granger	Sutton	J. Craig Venter Institute (JCVI)	GSutton@jcvi.org
FF0033	x	x	x	
FF0034	Clotilde	Teiling	Roche Applied Science	Clotilde.Teiling@roche.com
FF0035	Sandra	Porter	Digital World Biology & Austin Community College	sandra@digitalworldbiology.com, sandy@geospiza.com
FF0036	Kyle	Hubbard	US Army Edgewood Chemical Biological Center	kyle.hubbard@us.army.mil
FF0037	Olga	Chertkov	Los Alamos National Laboratory (LANL)	ochrtkv@lanl.gov
FF0038	Xiaohong	Liu	Pfizer, Inc.	xiaohong.liu@pfizer.com
FF0039	Peter	Vander Horn	Life Technologies	Peter.VanderHorn@lifetech.com
FF0040	John	Oliver	NABsys Inc.	oliver@nabsys.com
FF0041	Mark	Nadel	NABsys Inc.	nadel@nabsys.com
FF0042	Robert	Blakesley	National Human Genome Research Institute, NIH	rblakesl@nhgri.nih.gov
FF0043	Jyoti	Gupta	National Human Genome Research Institute, NIH	jyotig@mail.nih.gov
FF0044	Stacey	Broomall	US Army Edgewood Chemical Biological Center	stacey.broomall@us.army.mil
FF0045	Peter	Goldstein	NABsys Inc.	goldstein@nabsys.com
FF0046	Brian	Schmidt	National Human Genome Research Institute, NIH	schmidtbr@mail.nih.gov
FF0047	Gerry	Bouffard	National Human Genome Research Institute, NIH	bouffard@mail.nih.gov
FF0048	John	McPherson	Ontario Institute for Cancer Research (OICR)	john.mcpherson@oicr.on.ca
FF0049	Elisa	La Bauve	Sandia National Laboratories	elabauv@sandia.gov
FF0050	Kamlesh (Ken)	Patel	Sandia National Laboratories	kdpatel@sandia.gov
FF0051	Bharat	Patel	Griffith University	b.patel@griffith.edu.au
FF0052	Shannon	Johnson	Los Alamos National Laboratory (LANL)	shannonj@lanl.gov
FF0053	Sean	Conlan	National Human Genome Research Institute, NIH	conlans@mail.nih.gov
FF0054	Michael	Rey	Novozymes, Inc.	MWR@novozymes.com
FF0055	Amanda	Castle	Illumina, Inc.	acastle@illumina.com
FF0056	David	Gordon	University of Washington	dgordon@u.washington.edu
FF0057	Janine	McMurdie	Life Technologies	Janine.McMurdie@lifetech.com
FF0058	Haley	Fiske	Illumina, Inc.	hfiske@illumina.com
FF0059	Jim	Gareau	Physik Instrumente	jimg@pi-usa.us
FF0060	Andrey	Grigoriev	Rutgers University	agrigoriev@camden.rutgers.edu
FF0061	Louise	McConnell	Life Technologies	Louise.McConnell@lifetech.com

2011 Attendees

FF #	Name		Affiliation	email
FF0062	Graham	Threadgill	Beckman Coulter Inc.	gjthreadgill@beckman.com
FF0063	Robert	Huffman	Defense Threat Reduction Agency (DTRA)	Robert.Huffman@dtra.mil
FF0064	Robin	Kramer	National Center for Genome Resources (NCGR)	rsk@ncgr.org
FF0065	Ingrid	Lindquist	National Center for Genome Resources (NCGR)	iel@ncgr.org
FF0066	David	Roche	Genoscope (French National Sequencing Center)	droche@genoscope.cns.fr
FF0067	Eric	Pelletier	Genoscope (French National Sequencing Center)	ericp@genoscope.cns.fr
FF0068	Steve	Turner	Pacific Biosciences	sturner@pacificbiosciences.com
FF0069	Craig	Blackhart	Los Alamos National Laboratory (LANL)	blackhart@lanl.gov
FF0070	Julie	Poulain	Genoscope (French National Sequencing Center)	Poulain@genoscope.cns.fr
FF0071	Sophie	Layac-Mangenot	Genoscope (French National Sequencing Center)	mangenot@genoscope.cns.fr
FF0072	George	Vacek	Convey Computer Corporation	gvacek@conveycomputer.com
FF0073	Erick	Suh	Kapa Biosystems	erick.suh@kapabiosystems.com
FF0074	Beverly	Parson-Quintana	Los Alamos National Laboratory (LANL)	bapq@lanl.gov
FF0075	Chad	Locklear	Integrated DNA Technologies	clocklear@idtdna.com
FF0076	M.J.	Rosovitz	National Biodefense Analysis and Countermeasures Center (NBACC)	rosovitzmj@nbacc.net
FF0077	Hamid	Khoja	Covaris Inc.	Hkhoja@covarisinc.com
FF0078	Jim	Laugharn	Covaris Inc.	Jlaugharn@covarisinc.com
FF0079	George	VanDegrift	Convey Computer Corporation	gvandegrift@conveycomputer.com
FF0080	Darren	Grafham	Wellcome Trust Sanger Institute	dg1@sanger.ac.uk
FF0081	Kevin	Wuest	EdgeBio	Kwuest@edgebio.com
FF0082	Hilary	Browne	Wellcome Trust Sanger Institute	hb4@sanger.ac.uk
FF0083	Trevor	Knutson	Liquidia Technologies Inc.	Trevor.Knutson@liquidia.com
FF0084	Ben	Maynor	Liquidia Technologies Inc.	Ben.Maynor@liquidia.com
FF0085	Ash	Nijhawan	Liquidia Technologies Inc.	Ash.Nijhawan@liquidia.com
FF0086	Len	Pennacchio	Joint Genome Institute	LAPennacchio@lbl.gov
FF0087	Wenyu	Lin	Massachusetts General Hospital and Harvard Medical School	wlin1@partners.org
FF0088	Jim	Woynerowski	Covaris Inc.	jwoynerowski@covarisinc.com
FF0089	Brian	Paras	Covaris Inc.	bparas@covarisinc.com
FF0090	Heidi	Hauser	Wellcome Trust Sanger Institute	hch@sanger.ac.uk
FF0091	Richard	Clark	Wellcome Trust Sanger Institute	rcc@sanger.ac.uk
FF0092	Adam	Phillippy	National Biodefense Analysis and Countermeasures Center (NBACC)	phillippya@nbacc.net
FF0093	Brandon	Swan	Bigelow Laboratory for Ocean Sciences	bswan@bigelow.org
FF0094	Tim	Minogue	USAMRIID	timothy.minogue@us.army.mil
FF0095	Erin	Field	Bigelow Laboratory for Ocean Sciences	efield25@gmail.com
FF0096	David	Emerson	Bigelow Laboratory for Ocean Sciences	demerson@bigelow.org
FF0097	Indresh	Singh	J. Craig Venter Institute (JCVI)	Isingh@jcvi.org
FF0098	Surya	Saha	Cornell University	suryasaha@gmail.com
FF0099	Tina	(Graves) Lindsay	The Genome Institute at Washington University	tgraves@genome.wustl.edu
FF0100	Jingping	Li	Plant Genome Mapping Laboratory at University of Georgia	jingpingli@gmail.com
FF0101	Ramunas	Stepanauskas	Bigelow Laboratory for Ocean Sciences	rstepanauskas@bigelow.org
FF0102	Diana	Radune	J. Craig Venter Institute (JCVI)	Dradune@jcvi.org
FF0103	Nadia	Fedorova	J. Craig Venter Institute (JCVI)	NFedorova2@jcvi.org
FF0104	Lori	Peterson	Caldera Pharmaceuticals, Inc.	court@cpsci.com
FF0105	Nicole	Touchet	Caldera Pharmaceuticals, Inc.	touchet@cpsci.com
FF0106	Haibao	Tang	J. Craig Venter Institute (JCVI)	Htang@jcvi.org
FF0107	Nels	Olson	Johns Hopkins University	nels.olson@jhuapl.edu
FF0108	Martha	Ofelia Perez Arriga	GAITS	marperez@cs.unm.edu
FF0109	Christian	Buhay	Baylor College of Medicine	cbuhay@bcm.edu
FF0110	Donna	Muzny	Baylor College of Medicine	donnam@bcm.edu
FF0111	Michael	Holder	Baylor College of Medicine	mholder@bcm.edu
FF0112	Michael	FitzGerald	Broad Institute	fitz@broadinstitute.org
FF0113	Amr	Abouelleil	Broad Institute	amr@broadinstitute.org
FF0114	Alma	Imamovic	Broad Institute	imamovic@broadinstitute.org
FF0115	Yuriy	Fofanov	Center for Biomedical and Environmental Genomics, University of Houston	yfofanov@bioinfo.uh.edu
FF0116	Andrew	Bradbury	Los Alamos National Laboratory (LANL)	amb@lanl.gov
FF0117	Shihai	Feng	Los Alamos National Laboratory (LANL)	sfeng@lanl.gov
FF0118	Arvind	Bharti	National Center for Genome Resources (NCGR)	akb@ncgr.org
FF0119	Patrick	Minx	The Genome Institute at Washington University	pminx@watson.wustl.edu
FF0120	Matt	Davenport	Department of Homeland Security (DHS)	Matthew.Davenport@dhs.gov
FF0121	Amy Jo	Powell	Sandia National Laboratories	ajpowel@sandia.gov
FF0122	Andy	Felton	Life Technologies - Ion Torrent	Andy.Felton@lifetech.com

2011 Attendees

FF #	Name		Affiliation	email
FF0123	Robert	Fulton	Washington University School of Medicine	bfulton@genome.wustl.edu
FF0124	Abhishek	Pratap	Joint Genome Institute	apratap@lbl.gov
FF0125	Mark	Wolcott	USAMRIID	Mark.wolcott@us.army.mil
FF0126	Ben	Allen	Los Alamos National Laboratory (LANL)	bsa@lanl.gov
FF0127	Don	Natvig	University of New Mexico	dnatvig@gmail.com
FF0128	Dan	Conway	CLC Bio, LLC	dconway@clcbio.com
FF0129	Janine	Detter	Los Alamos National Laboratory (LANL)	janined@lanl.gov
FF0130	Joe	Salvatore	CLC Bio, LLC	jsalvatore@clcbio.com
FF0131	Armand	Dichosa	Los Alamos National Laboratory (LANL)	armand@lanl.gov
FF0132	Martina	Siwek	JPM - Chemical Biological Medical Systems	Martina.Siwek.ctr@us.army.mil
FF0133	Alla	Lapidus	Fox Chase Cancer Center	Alla.Lapidus@fccc.edu
FF0134	Ahmet	Zeytun	Los Alamos National Laboratory (LANL)	azeytun@lanl.gov
FF0135	Paula	Imbro	BioWatch Program, Tauri Group	paula.imbro@taurigroup.com
FF0136	Patti	Wills	Los Alamos National Laboratory (LANL)	wills@lanl.gov
FF0137	Kim	McMurry	Los Alamos National Laboratory (LANL)	kmcmurry@lanl.gov
FF0138	Loren	Hauser	Oak Ridge National Laboratory (ORNL)	hauserlj@ornl.gov
FF0139	Mary	Campbell	Los Alamos National Laboratory (LANL)	mcampbell@lanl.gov
FF0140	Sarah	Young	Broad Institute	stowey@broadinstitute.org
FF0141	Aaron	Berlin	Broad Institute	amberlin@broadinstitute.org
FF0142	Bruce	Walker	Broad Institute	bruce@broadinstitute.org
FF0143	Linda	Meincke	Los Alamos National Laboratory (LANL)	meincke@lanl.gov
FF0144	Cristina	Takacs-Vesbach	University of New Mexico	cvesbach@gmail.com
FF0145	Faye	Schilkey	National Center for Genome Resources (NCGR)	fds@ncgr.org
FF0146	Jim	Knight	Roche Applied Science	james.knight@roche.com
FF0147	Wei	Gu	Los Alamos National Laboratory (LANL)	wgu@lanl.gov
FF0148	Lee	Poeppelman	US Air Force Research Laboratory	Lee.Poeppelman@wpafb.af.mil
FF0149	John	Schlager	US Air Force Research Laboratory	john.schlager@wpafb.af.mil
FF0150	Sarah	Hicks	University of New Mexico	garlicscape@gmail.com
FF0151	Feng	Chen	Joint Genome Institute	fchen@lbl.gov
FF0152	Miriam	Land	Oak Ridge National Laboratory (ORNL)	landml@ornl.gov
FF0153	Scott	Remine	Defense Threat Reduction Agency (DTRA)	scott.remine@dtra.mil
FF0154	Chris	Munk	Los Alamos National Laboratory (LANL)	cmunk@lanl.gov
FF0155	Jimmy	Woodward	National Center for Genome Resources (NCGR)	jew@ncgr.org
FF0156	Kostas	Mavrommatis	Joint Genome Institute	mavrommatis.konstantinos@gmail.com
FF0157	Eric	Ackerman	Sandia National Laboratories	eackerm@sandia.gov
FF0158	Dhwani	Batra	SRA International	bun3@cdc.gov
FF0159	Julia	Scheerer	Defense Threat Reduction Agency (DTRA)	julia.scheerer_bna@taurigroup.com
FF0160	Brittany	Held	Los Alamos National Laboratory (LANL)	bheld@lanl.gov
FF0161	Peter	Pesenti	Department of Homeland Security (DHS)	Peter.Pesenti@dhs.gov
FF0162	Thiru	Ramaraj	National Center for Genome Resources (NCGR)	tr@ncgr.org
FF0163	Maggie	Werner-Washburne	University of New Mexico	maggieww@unm.edu
FF0164	Diana	Northup	University of New Mexico	dnorthup@unm.edu
FF0165	Dan	Distel	Ocean Genome Legacy	distel@oglf.org
FF0166	Megan	Lu	Los Alamos National Laboratory (LANL)	meganlu@lanl.gov
FF0167	Hazuki	Teshima	Los Alamos National Laboratory (LANL)	hazuki@lanl.gov
FF0168	Karina	Yusim	Los Alamos National Laboratory (LANL)	kyusim@lanl.gov
FF0169	Kirsten	McLay	The Genome Analysis Centre	Kirsten.McLay@bbsrc.ac.uk
FF0170	Kyle	O'Connor	Life Technologies	Kyle.O'Connor@lifetech.com
FF0171	Mike	Walsh	Beckman Coulter Inc.	mlwalsh@beckman.com
FF0172	Michael	Fitzsimons	Los Alamos National Laboratory (LANL)	msfitz@lanl.gov
FF0173	Norman	Doggett	Los Alamos National Laboratory (LANL)	doggett@lanl.gov
FF0174	John	Barnes	Center for Disease Control (CDC)	fzq9@cdc.gov
FF0175	Tracy	Erkkila	Los Alamos National Laboratory (LANL)	terkkila@lanl.gov
FF0176	Jeffrey	Martin	Joint Genome Institute	jamartin@lbl.gov
FF0177	Scott	Happe	Agilent Technologies	scott.happe@agilent.com
FF0178	Jane	Hutchinson	Roche Applied Science	jane.hutchinson@roche.com
FF0179	Roxanne	Tapia	Los Alamos National Laboratory (LANL)	rox@lanl.gov
FF0180	David	Cleary	Defense Science and Technology Laboratory (DSTL)	dwcleary@dstl.gov.uk
FF0181	Phil	Rachwal	Defense Science and Technology Laboratory (DSTL)	parachwal@dstl.gov.uk
FF0182	Jean	Jasinski	Agilent Technologies	jean@strandsi.com
FF0183	Trevor	Wagner	OpGen	twagner@opgen.com

2011 Attendees

FF #	Name		Affiliation	email
FF0184	Margaret (Peggy)	Rogers	OpGen	progers@opgen.com
FF0185	Randy	Stratton	OpGen	rstratton@opgen.com
FF0186	Nianqing (Nick)	Xiao	OpGen	nxiao@opgen.com
FF0187	Gary	Resnick	Los Alamos National Laboratory (LANL)	resnick@lanl.gov
FF0188	Andrew	Barry	Caliper Life Sciences	andrew.barry@caliperls.com
FF0189	Krista	Reitenga	Los Alamos National Laboratory (LANL)	reitenga@lanl.gov
FF0190	David	Hirschberg	Columbia University	david.hirschberg@columbia.edu
FF0191	Christina	Chiu	RainDance Technologies, Inc.	ChiuC@raindancetech.com
FF0192	Cathy	Cleland	Los Alamos National Laboratory (LANL)	ccleland@lanl.gov
FF0193	Julianna	Fessenden-Rahn	Los Alamos National Laboratory (LANL)	julianna@lanl.gov
FF0194	Cliff	Han	Los Alamos National Laboratory (LANL)	Han_Cliff@lanl.gov
FF0195	Penny	Hitchcock	BioWatch Program, Tauri Group	penny.hitchcock@taurigroup.com
FF0196	Catherine	Campbell	Center for National Security and Intelligence, Noblis	Catherine.Campbell@noblis.org
FF0197	Richard	McCombie	Cold Spring Harbor Lab	mccombie@cshl.edu
FF0198	Alicia	Hawes	Baylor College of Medicine	ahawes@bcm.edu
FF0199	Adam	English	Baylor College of Medicine	english@bcm.edu
FF0200	Todd	Michael	Monsanto Company	todd.p.michael@monsanto.com
FF0201	Johar	Ali	Ontario Institute for Cancer Research (OICR)	Johar.Ali@oicr.on.ca
FF0202	Cheng-Cang (Charles)	We	Lucigen Corporation	cwu@lucigen.com
FF0203	Cheryl	Gleasner	Los Alamos National Laboratory (LANL)	cdgle@lanl.gov
FF0204	Aaron	Klammer	Pacific Biosciences	aklammer@pacificbiosciences.com
FF0205	Scott	Jordan	Physik Instrumente	scottj@pi-usa.us
FF0206	Eric	van der Walt	Kapa Biosystems	eric.vanderwalt@kapabiosystems.com
FF0207	Jennifer	Foster Harris	Los Alamos National Laboratory (LANL)	jfharris@lanl.gov
FF0208	Alisa	Jackson	Beckman Coulter Inc.	acjackson@beckman.com
FF0209	Mary	Blair	Beckman Coulter Inc.	MEBlair@beckman.com
FF0210	Joanna	Redfern	University of New Mexico	jredfern@unm.edu
FF0211	Omayma	Al-Awar	Illumina, Inc.	oalawar@illumina.com
FF0212	Yuliya	Kunde	Los Alamos National Laboratory (LANL)	y.a.kunde@lanl.gov
FF0213	Gary	Simpson	Placitas Consulting Group	garyl.simpson@comcast.net
FF0214	Zhong	Wang	DOE Joint Genomics Institute	zhongwang@lbl.gov
FF0215	Nicholas	Beckloff	Los Alamos National Laboratory (LANL)	beckloff@lanl.gov
FF0216	Patrick	Chain	Los Alamos National Laboratory (LANL)	pchain@lanl.gov
FF0217	Dan	Drell	US Department of Energy (DOE)	Daniel.Drell@science.doe.gov
FF0218	Caleb	Phillips	Texas Tech University	caleb.phillips@ttu.edu
FF0219	Robert	Baker	Texas Tech University	rjbaker@ttu.edu
FF0220	Jun	Hang	Walter Reed Army Institute of Research, Division of Viral Diseases	jun.hang.ctr@us.army.mil
FF0221	Alex	Hutcheson	Pacific Biosciences	ahutcheson@pacificbiosciences.com
FF0222	Jonathan	Bingham	Pacific Biosciences	jbingham@pacificbiosciences.com
FF0223	Liz	O'Hara	OpGen	lohara@opgen.com
FF0224	Annette	Sobel	University of Missouri	sobelan@missouri.edu
FF0225	Beth	Fisher	University of Missouri	Beth.Fisher@umsystem.edu
FF0226	Will	Fischer	Los Alamos National Laboratory (LANL)	wfischer@lanl.gov
FF0227	Nicole	Rosenzweig	Edgewood Chemical Biological Center	nicole.rosenzweig@us.army.mil
FF0228	Peter	Hraber	Los Alamos National Laboratory (LANL)	phraber@lanl.gov
FF0229	Jason	Aulds	National Center for Medical Intelligence	jaulds@ncmi.detrick.army.mil
FF0230	Yael	Berstein	Cold Spring Harbor Laboratory	yberstei@cshl.edu
FF0231	Chris	Blessington	Isilon	cblessington@isilon.com
FF0232	Matt	Martin	Isilon	matt.martin@isilon.com
FF0233	Kirt	Karl	Isilon	kirt.karl@isilon.com
FF0234	Joe	Alessi	Illumina, Inc.	JAlessi@illumina.com
FF0235	John	Foskett	Kapa Biosystems	john.foskett@kapabiosystems.com
FF0236	Alex	Long	Isilon	longat@isilon.com

2011 Attendees

FF #	By Last Name		Affiliation	email
FF0113	Amr	Abouelleil	Broad Institute	amr@broadinstitute.org
FF0157	Eric	Ackerman	Sandia National Laboratories	eackerm@sandia.gov
FF0211	Omayma	Al-Awar	Illumina, Inc.	oalawar@illumina.com
FF0234	Joe	Alessi	Illumina, Inc.	JAlessi@illumina.com
FF0201	Johar	Ali	Ontario Institute for Cancer Research (OICR)	Johar.Ali@oicr.on.ca
FF0126	Ben	Allen	Los Alamos National Laboratory (LANL)	bsa@lanl.gov
FF0229	Jason	Aulds	National Center for Medical Intelligence	jaulds@ncmi.detrick.army.mil
FF0219	Robert	Baker	Texas Tech University	rjbaker@ttu.edu
FF0174	John	Barnes	Center for Disease Control (CDC)	fzq9@cdc.gov
FF0188	Andrew	Barry	Caliper Life Sciences	andrew.barry@caliperls.com
FF0158	Dhwani	Batra	SRA International	bun3@cdc.gov
FF0215	Nicholas	Beckloff	Los Alamos National Laboratory (LANL)	beckloff@lanl.gov
FF0020	Tara	Bennink	EdgeBio	Tbennink@edgebio.com
FF0141	Aaron	Berlin	Broad Institute	amberlin@broadinstitute.org
FF0230	Yael	Berstein	Cold Spring Harbor Laboratory	yberstei@cshl.edu
FF0118	Arvind	Bharti	National Center for Genome Resources (NCGR)	akb@ncgr.org
FF0222	Jonathan	Bingham	Pacific Biosciences	jbingham@pacificbiosciences.com
FF0021	Kim	Bishop-Lilly	Naval Medical Research Center	kim.bishop-lilly@med.navy.mil
FF0069	Craig	Blackhart	Los Alamos National Laboratory (LANL)	blackhart@lanl.gov
FF0209	Mary	Blair	Beckman Coulter Inc.	MEBlair@beckman.com
FF0042	Robert	Blakesley	National Human Genome Research Institute, NIH	rblakesl@nhgri.nih.gov
FF0231	Chris	Blessington	Isilon	cblessington@isilon.com
FF0047	Gerry	Bouffard	National Human Genome Research Institute, NIH	bouffard@mail.nih.gov
FF0116	Andrew	Bradbury	Los Alamos National Laboratory (LANL)	amb@lanl.gov
FF0044	Stacey	Broomall	US Army Edgewood Chemical Biological Center	stacey.broomall@us.army.mil
FF0009	Keith	Brown	RainDance Technologies, Inc.	brownK@raindancetech.com
FF0082	Hilary	Browne	Wellcome Trust Sanger Institute	hb4@sanger.ac.uk
FF0004	David	Bruce	Los Alamos National Laboratory (LANL)	dbruce@lanl.gov
FF0109	Christian	Buhay	Baylor College of Medicine	cbuhay@bcm.edu
FF0139	Mary	Campbell	Los Alamos National Laboratory (LANL)	mcampbell@lanl.gov
FF0196	Catherine	Campbell	Center for National Security and Intelligence, Noblis	Catherine.Campbell@noblis.org
FF0055	Amanda	Castle	Illumina, Inc.	acastle@illumina.com
FF0216	Patrick	Chain	Los Alamos National Laboratory (LANL)	pchain@lanl.gov
FF0151	Feng	Chen	Joint Genome Institute	fchen@lbl.gov
FF0037	Olga	Chertkov	Los Alamos National Laboratory (LANL)	ochrtkv@lanl.gov
FF0008	Chrisinta	Chiu	RainDance Technologies, Inc.	chiuC@raindancetech.com
FF0191	Christina	Chiu	RainDance Technologies, Inc.	ChiuC@raindancetech.com
FF0091	Richard	Clark	Wellcome Trust Sanger Institute	rcc@sanger.ac.uk
FF0180	David	Cleary	Defense Science and Technology Laboratory (DSTL)	dwcleary@dstl.gov.uk
FF0192	Cathy	Cleland	Los Alamos National Laboratory (LANL)	ccleland@lanl.gov
FF0053	Sean	Conlan	National Human Genome Research Institute, NIH	conlans@mail.nih.gov
FF0128	Dan	Conway	CLC Bio, LLC	dconway@clcbio.com
FF0030	Helen	Cui	Los Alamos National Laboratory (LANL)	hhcui@lanl.gov
FF0012	Karen	Davenport	Los Alamos National Laboratory (LANL)	kwdavenport@lanl.gov
FF0120	Matt	Davenport	Department of Homeland Security (DHS)	Matthew.Davenport@dhs.gov
FF0014	Alfredo Lopez	De Leon	Novozymes, Inc.	ALLO@novozymes.com
FF0001	Chris	Detter	Los Alamos National Laboratory (LANL)	cdetter@lanl.gov
FF0129	Janine	Detter	Los Alamos National Laboratory (LANL)	janined@lanl.gov
FF0131	Armand	Dichosa	Los Alamos National Laboratory (LANL)	armand@lanl.gov
FF0165	Dan	Distel	Ocean Genome Legacy	distel@oglf.org
FF0173	Norman	Doggett	Los Alamos National Laboratory (LANL)	doggett@lanl.gov
FF0217	Dan	Drell	US Department of Energy (DOE)	Daniel.Drell@science.doe.gov
FF0096	David	Emerson	Bigelow Laboratory for Ocean Sciences	demerson@bigelow.org
FF0199	Adam	English	Baylor College of Medicine	english@bcm.edu
FF0175	Tracy	Erkkila	Los Alamos National Laboratory (LANL)	terkkila@lanl.gov
FF0103	Nadia	Fedorova	J. Craig Venter Institute (JCVI)	NFedorova2@jcvi.org
FF0122	Andy	Felton	Life Technologies - Ion Torrent	Andy.Felton@lifetech.com
FF0117	Shihai	Feng	Los Alamos National Laboratory (LANL)	sfeng@lanl.gov
FF0193	Julianna	Fessenden-Rahn	Los Alamos National Laboratory (LANL)	julianna@lanl.gov
FF0095	Erin	Field	Bigelow Laboratory for Ocean Sciences	efield25@gmail.com
FF0226	Will	Fischer	Los Alamos National Laboratory (LANL)	wfischer@lanl.gov

2011 Attendees

FF #	By Last Name		Affiliation	email
FF0225	Beth	Fisher	University of Missouri	Beth.Fisher@umsystem.edu
FF0058	Haley	Fiske	Illumina, Inc.	hfiske@illumina.com
FF0112	Michael	FitzGerald	Broad Institute	fitz@broadinstitute.org
FF0172	Michael	Fitzsimons	Los Alamos National Laboratory (LANL)	msfitz@lanl.gov
FF0115	Yuriy	Fofanov	Center for Biomedical and Environmental Genomics, University of Houston	yfofanov@bioinfo.uh.edu
FF0235	John	Foskett	Kapa Biosystems	john.foskett@kapabiosystems.com
FF0207	Jennifer	Foster Harris	Los Alamos National Laboratory (LANL)	jfharris@lanl.gov
FF0123	Robert	Fulton	Washington University School of Medicine	bfulton@genome.wustl.edu
FF0059	Jim	Gareau	Physik Instrumente	jimg@pi-usa.us
FF0025	Scott	Geib	USDA - ARS	Scott.Geib@ARS.USDA.gov
FF0203	Cheryl	Gleasner	Los Alamos National Laboratory (LANL)	cdgle@lanl.gov
FF0045	Peter	Goldstein	NABsys Inc.	goldstein@nabsys.com
FF0056	David	Gordon	University of Washington	dgordon@u.washington.edu
FF0080	Darren	Grafham	Wellcome Trust Sanger Institute	dg1@sanger.ac.uk
FF0099	Tina	Graves Lindsay	The Genome Institute at Washington University	tgraves@genome.wustl.edu
FF0060	Andrey	Grigoriev	Rutgers University	agrigoriev@camden.rutgers.edu
FF0147	Wei	Gu	Los Alamos National Laboratory (LANL)	wgu@lanl.gov
FF0043	Jyoti	Gupta	National Human Genome Research Institute, NIH	jyotig@mail.nih.gov
FF0194	Cliff	Han	Los Alamos National Laboratory (LANL)	Han_Cliff@lanl.gov
FF0220	Jun	Hang	Walter Reed Army Institute of Research, Division of Viral Diseases	jun.hang.ctr@us.army.mil
FF0177	Scott	Happe	Agilent Technologies	scott.happe@agilent.com
FF0090	Heidi	Hauser	Wellcome Trust Sanger Institute	hch@sanger.ac.uk
FF0138	Loren	Hauser	Oak Ridge National Laboratory (ORNL)	hauserlj@ornl.gov
FF0011	John	Havens	Integrated DNA Technologies	jhavens@idtdna.com
FF0198	Alicia	Hawes	Baylor College of Medicine	ahawes@bcm.edu
FF0160	Brittany	Held	Los Alamos National Laboratory (LANL)	bheld@lanl.gov
FF0028	Cynthia	Hendrickson	New England Biolabs	hendrickson@neb.com
FF0150	Sarah	Hicks	University of New Mexico	garlicscape@gmail.com
FF0190	David	Hirschberg	Columbia University	david.hirschberg@columbia.edu
FF0195	Penny	Hitchcock	BioWatch Program, Tauri Group	penny.hitchcock@taurigroup.com
FF0111	Michael	Holder	Baylor College of Medicine	mholder@bcm.edu
FF0228	Peter	Hraber	Los Alamos National Laboratory (LANL)	phraber@lanl.gov
FF0036	Kyle	Hubbard	US Army Edgewood Chemical Biological Center	kyle.hubbard@us.army.mil
FF0063	Robert	Huffman	Defense Threat Reduction Agency (DTRA)	Robert.Huffman@dtra.mil
FF0003	Tim	Hunkapiller	Discovery Bio	tim@discoverybio.com
FF0221	Alex	Hutcheson	Pacific Biosciences	ahutcheson@pacificbiosciences.com
FF0178	Jane	Hutchinson	Roche Applied Science	jane.hutchinson@roche.com
FF0114	Alma	Imamovic	Broad Institute	imamovic@broadinstitute.org
FF0135	Paula	Imbro	BioWatch Program, Tauri Group	paula.imbro@taurigroup.com
FF0208	Alisa	Jackson	Beckman Coulter Inc.	acjackson@beckman.com
FF0182	Jean	Jasinski	Agilent Technologies	jean@strandsi.com
FF0052	Shannon	Johnson	Los Alamos National Laboratory (LANL)	shannonj@lanl.gov
FF0205	Scott	Jordan	Physik Instrumente	scottj@pi-usa.us
FF0233	Kirt	Karl	Isilon	kirt.karl@isilon.com
FF0077	Hamid	Khoja	Covaris Inc.	Hkhoja@covarisinc.com
FF0204	Aaron	Klammer	Pacific Biosciences	aklammer@pacificbiosciences.com
FF0146	Jim	Knight	Roche Applied Science	james.knight@roche.com
FF0083	Trevor	Knutson	Liquidia Technologies Inc.	Trevor.Knutson@liquidia.com
FF0064	Robin	Kramer	National Center for Genome Resources (NCGR)	rsk@ncgr.org
FF0212	Yuliya	Kunde	Los Alamos National Laboratory (LANL)	y.a.kunde@lanl.gov
FF0049	Elisa	La Bauve	Sandia National Laboratories	elabauv@sandia.gov
FF0026	Mike	Lafferty	Life Technologies	Mike.Lafferty@lifetech.com
FF0152	Miriam	Land	Oak Ridge National Laboratory (ORNL)	landml@ornl.gov
FF0133	Alla	Lapidus	Fox Chase Cancer Center	Alla.Lapidus@fccc.edu
FF0078	Jim	Laugharn	Covaris Inc.	Jlaugharn@covarisinc.com
FF0071	Sophie	Layac-Mangenot	Genoscope (French National Sequencing Center)	mangenot@genoscope.cns.fr
FF0100	Jingping	Li	Plant Genome Mapping Laboratory at University of Georgia	jingpingli@gmail.com
FF0087	Wenyu	Lin	Massachusetts General Hospital and Harvard Medical School	wlin1@partners.org
FF0065	Ingrid	Lindquist	National Center for Genome Resources (NCGR)	iel@ncgr.org
FF0038	Xiaohong	Liu	Pfizer, Inc.	xiaohong.liu@pfizer.com
FF0075	Chad	Locklear	Integrated DNA Technologies	clocklear@idtdna.com

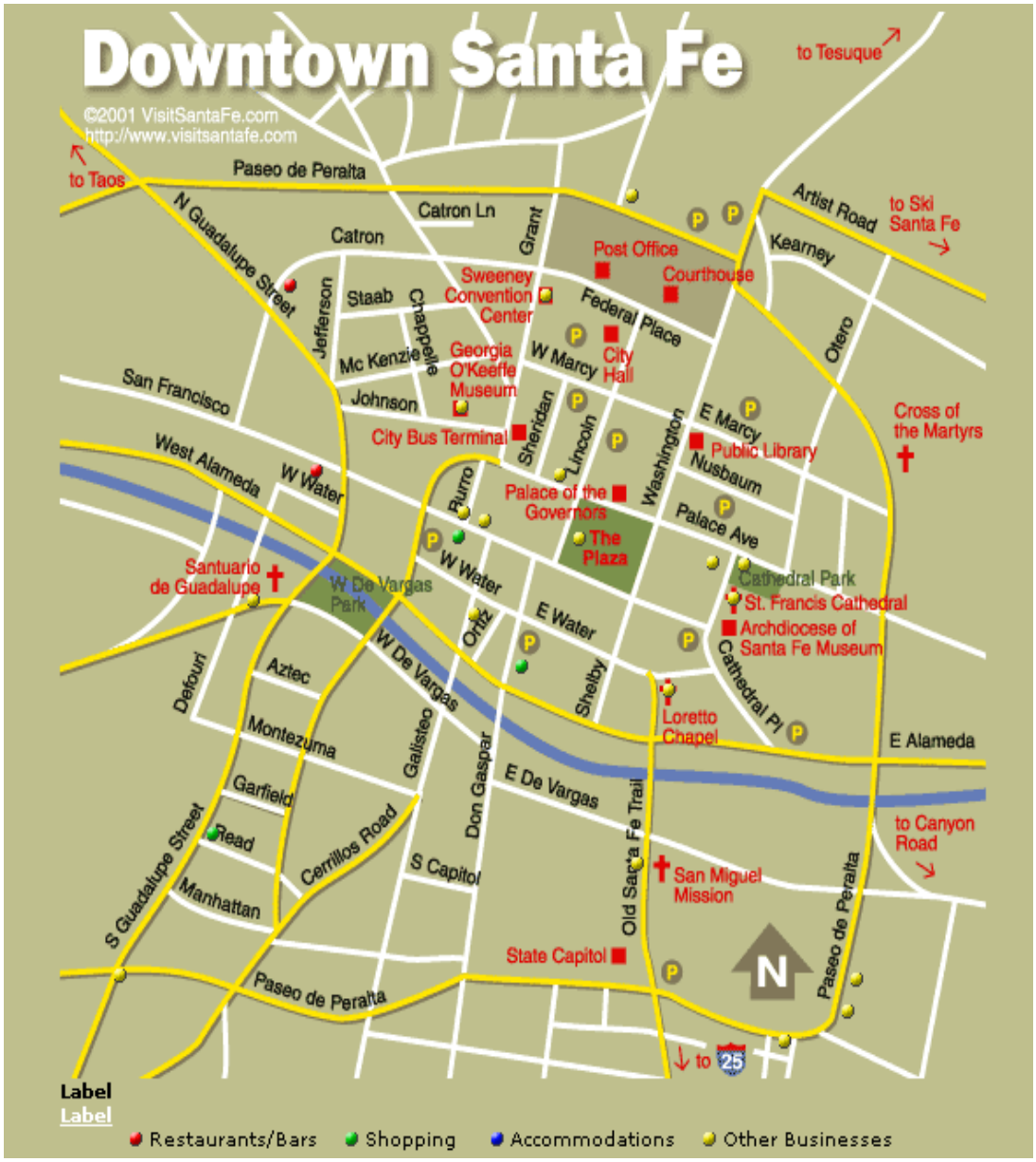
2011 Attendees

FF #	By Last Name		Affiliation	email
FF0236	Alex	Long	Isilon	longat@isilon.com
FF0166	Megan	Lu	Los Alamos National Laboratory (LANL)	meganlu@lanl.gov
FF0176	Jeffrey	Martin	Joint Genome Institute	jamartin@lbl.gov
FF0232	Matt	Martin	Isilon	matt.martin@isilon.com
FF0156	Kostas	Mavrommatis	Joint Genome Institute	mavrommatis.konstantinos@gmail.com
FF0084	Ben	Maynor	Liquidia Technologies Inc.	Ben.Maynor@liquidia.com
FF0197	Richard	McCombie	Cold Spring Harbor Lab	mccombie@cshl.edu
FF0061	Louise	McConnell	Life Technologies	Louise.McConnell@lifetech.com
FF0169	Kirsten	McLay	The Genome Analysis Centre	Kirsten.McLay@bbsrc.ac.uk
FF0057	Janine	McMurdie	Life Technologies	Janine.McMurdie@lifetech.com
FF0137	Kim	McMurry	Los Alamos National Laboratory (LANL)	kmcmurry@lanl.gov
FF0048	John	McPherson	Ontario Institute for Cancer Research (OICR)	john.mcpherson@oicr.on.ca
FF0019	Isaac	Meek	Caliper Life Sciences	Isaac.Meek@caliperls.com
FF0143	Linda	Meincke	Los Alamos National Laboratory (LANL)	meincke@lanl.gov
FF0200	Todd	Michael	Monsanto Company	todd.p.michael@monsanto.com
FF0094	Tim	Minogue	USAMRIID	timothy.minogue@us.army.mil
FF0119	Patrick	Minx	The Genome Institute at Washington University	pminx@watson.wustl.edu
FF0002	Joann	Mudge	National Center for Genome Resources (NCGR)	jm@ncgr.org
FF0154	Chris	Munk	Los Alamos National Laboratory (LANL)	cmunk@lanl.gov
FF0110	Donna	Muzny	Baylor College of Medicine	donnam@bcm.edu
FF0041	Mark	Nadel	NABsys Inc.	nadel@nabsys.com
FF0127	Don	Natvig	University of New Mexico	dnatvig@gmail.com
FF0085	Ash	Nijhawan	Liquidia Technologies Inc.	Ash.Nijhawan@liquidia.com
FF0164	Diana	Northup	University of New Mexico	dnorthup@unm.edu
FF0170	Kyle	O'Connor	Life Technologies	Kyle.O'Connor@lifetech.com
FF0108	Martha	Ofelia Perez Arriga	GAITS	marperez@cs.unm.edu
FF0007	Take	Ogawa	RainDance Technologies, Inc.	OGAWAT@raindancetech.com
FF0223	Liz	O'Hara	OpGen	lohara@opgen.com
FF0040	John	Oliver	NABsys Inc.	oliver@nabsys.com
FF0107	Nels	Olson	Johns Hopkins University	nels.olson@jhuapl.edu
FF0089	Brian	Paras	Covaris Inc.	bparas@covarisinc.com
FF0074	Beverly	Parson-Quintana	Los Alamos National Laboratory (LANL)	bapq@lanl.gov
FF0050	Kamlesh (Ken)	Patel	Sandia National Laboratories	kdpatel@sandia.gov
FF0051	Bharat	Patel	Griffith University	b.patel@griffith.edu.au
FF0067	Eric	Pelletier	Genoscope (French National Sequencing Center)	ericp@genoscope.cns.fr
FF0086	Len	Pennacchio	Joint Genome Institute	LAPennacchio@lbl.gov
FF0161	Peter	Pesenti	Department of Homeland Security (DHS)	Peter.Pesenti@dhs.gov
FF0104	Lori	Peterson	Caldera Pharmaceuticals, Inc.	court@cpsci.com
FF0092	Adam	Phillippy	National Biodefense Analysis and Countermeasures Center (NBACC)	phillippya@nbacc.net
FF0218	Caleb	Phillips	Texas Tech University	caleb.phillips@ttu.edu
FF0148	Lee	Poeppelman	US Air Force Research Laboratory	Lee.Poeppelman@wpafb.af.mil
FF0035	Sandra	Porter	Digital World Biology & Austin Community College	sandra@digitalworldbiology.com, sandy@geospiza.com
FF0070	Julie	Poulain	Genoscope (French National Sequencing Center)	Poulain@genoscope.cns.fr
FF0121	Amy Jo	Powell	Sandia National Laboratories	ajpowel@sandia.gov
FF0124	Abhishek	Pratap	Joint Genome Institute	apratap@lbl.gov
FF0017	Gary	Qiao	Defense Threat Reduction Agency (DTRA)	Guilin.Qiao@dtra.mil
FF0181	Phil	Rachwal	Defense Science and Technology Laboratory (DSTL)	parachwal@dstl.gov.uk
FF0102	Diana	Radune	J. Craig Venter Institute (JCVI)	Dradune@jcvi.org
FF0162	Thiru	Ramaraj	National Center for Genome Resources (NCGR)	tr@ncgr.org
FF0015	Teri	Rambo Mueller	Roche Applied Science	teri.mueller@roche.com
FF0210	Joanna	Redfern	University of New Mexico	jredfern@unm.edu
FF0189	Krista	Reitenga	Los Alamos National Laboratory (LANL)	reitenga@lanl.gov
FF0153	Scott	Remine	Defense Threat Reduction Agency (DTRA)	scott.remine@dtra.mil
FF0187	Gary	Resnick	Los Alamos National Laboratory (LANL)	resnick@lanl.gov
FF0005	Ernie	Retzel	National Center for Genome Resources (NCGR)	efr@ncgr.org
FF0054	Michael	Rey	Novozymes, Inc.	MWR@novozymes.com
FF0010	Michael	Rhodes	Life Technologies	Michael.Rhodes@lifetech.com
FF0066	David	Roche	Genoscope (French National Sequencing Center)	droche@genoscope.cns.fr
FF0184	Margaret (Peggy)	Rogers	OpGen	progers@opgen.com
FF0227	Nicole	Rosenzweig	Edgewood Chemical Biological Center	nicole.rosenzweig@us.army.mil
FF0076	M.J.	Rosovitz	National Biodefense Analysis and Countermeasures Center (NBACC)	rosovitzmj@nbacc.net

2011 Attendees

FF #	By Last Name		Affiliation	email
FF0098	Surya	Saha	Cornell University	suryasaha@gmail.com
FF0130	Joe	Salvatore	CLC Bio, LLC	jsalvatore@clcbio.com
FF0013	Nan	Sauer	Los Alamos National Laboratory (LANL)	nsauer@lanl.gov
FF0159	Julia	Scheerer	Defense Threat Reduction Agency (DTRA)	julia.scheerer_bna@taurigroup.com
FF0145	Faye	Schilkey	National Center for Genome Resources (NCGR)	fds@ncgr.org
FF0149	John	Schlager	US Air Force Research Laboratory	john.schlager@wpafb.af.mil
FF0046	Brian	Schmidt	National Human Genome Research Institute, NIH	schmidtbr@mail.nih.gov
FF0213	Gary	Simpson	Placitas Consulting Group	garyl.simpson@comcast.net
FF0097	Indresh	Singh	J. Craig Venter Institute (JCVI)	Isingh@jcvi.org
FF0132	Martina	Siwek	JPM - Chemical Biological Medical Systems	Martina.Siwek.ctr@us.army.mil
FF0031	Todd	Smith	Geospiza, Inc.	todd@geospiza.com
FF0224	Annette	Sobel	University of Missouri	sobelan@missouri.edu
FF0022	Shanmuga	Sozhamannan	Naval Medical Research Center	shanmuga.sozhamannan@med.navy.mil
FF0101	Ramunas	Stepanauskas	Bigelow Laboratory for Ocean Sciences	rstepanauskas@bigelow.org
FF0029	Keven	Stevens	Integrated DNA Technologies	kstevens@idtdna.com
FF0006	Fiona	Stewart	New England Biolabs	stewart@neb.com
FF0185	Randy	Stratton	OpGen	rstratton@opgen.com
FF0073	Erick	Suh	Kapa Biosystems	erick.suh@kapabiosystems.com
FF0032	Granger	Sutton	J. Craig Venter Institute (JCVI)	GSutton@jcvi.org
FF0093	Brandon	Swan	Bigelow Laboratory for Ocean Sciences	bswan@bigelow.org
FF0144	Cristina	Takacs-Vesbach	University of New Mexico	cvesbach@gmail.com
FF0106	Haibao	Tang	J. Craig Venter Institute (JCVI)	Htang@jcvi.org
FF0179	Roxanne	Tapia	Los Alamos National Laboratory (LANL)	rox@lanl.gov
FF0018	Ken	Taylor	Integrated DNA Technologies	ktaylor@idtdna.com
FF0034	Clotilde	Teiling	Roche Applied Science	Clotilde.Teiling@roche.com
FF0167	Hazuki	Teshima	Los Alamos National Laboratory (LANL)	hazuki@lanl.gov
FF0062	Graham	Threadgill	Beckman Coulter Inc.	gjthreadgill@beckman.com
FF0105	Nicole	Touchet	Caldera Pharmaceuticals, Inc.	touchet@cpsci.com
FF0068	Steve	Turner	Pacific Biosciences	sturner@pacificbiosciences.com
FF0072	George	Vacek	Convey Computer Corporation	gvacek@conveycomputer.com
FF0206	Eric	van der Walt	Kapa Biosystems	eric.vanderwalt@kapabiosystems.com
FF0079	George	VanDegrift	Convey Computer Corporation	gvandegrift@conveycomputer.com
FF0039	Peter	Vander Horn	Life Technologies	Peter.VanderHorn@lifetech.com
FF0183	Trevor	Wagner	OpGen	twagner@opgen.com
FF0027	Thomas	Walk	USDA - ARS	tom.walk@ars.usda.gov
FF0142	Bruce	Walker	Broad Institute	bruce@broadinstitute.org
FF0171	Mike	Walsh	Beckman Coulter Inc.	mlwalsh@beckman.com
FF0024	Ron	Walters	Pacific Northwest National Laboratory (PNL)	ra.walters@pnl.gov
FF0214	Zhong	Wang	DOE Joint Genomics Institute	zhongwang@lbl.gov
FF0016	Ian	Watson	Defense Threat Reduction Agency (DTRA)	Ian.Watson@dtra.mil
FF0202	Cheng-Cang (Charles)	We	Lucigen Corporation	cwu@lucigen.com
FF0163	Maggie	Werner-Washburne	University of New Mexico	maggieww@unm.edu
FF0136	Patti	Wills	Los Alamos National Laboratory (LANL)	wills@lanl.gov
FF0125	Mark	Wolcott	USAMRIID	Mark.wolcott@us.army.mil
FF0155	Jimmy	Woodward	National Center for Genome Resources (NCGR)	jew@ncgr.org
FF0088	Jim	Woynerowski	Covaris Inc.	jwoynerowski@covarisinc.com
FF0081	Kevin	Wuest	EdgeBio	Kwuest@edgebio.com
FF0033	x	x	x	
FF0186	Nianqing (Nick)	Xiao	OpGen	nxiao@opgen.com
FF0023	Malin	Young	Sandia National Laboratories	mmyoung@sandia.gov
FF0140	Sarah	Young	Broad Institute	stowey@broadinstitute.org
FF0168	Karina	Yusim	Los Alamos National Laboratory (LANL)	kyusim@lanl.gov
FF0134	Ahmet	Zeytun	Los Alamos National Laboratory (LANL)	azeytun@lanl.gov

Map of Santa Fe, NM



History of Santa Fe, NM

Thirteen years before Plymouth Colony was settled by the Mayflower Pilgrims, Santa Fe, New Mexico, was established with a small cluster of European type dwellings. It would soon become the seat of power for the Spanish Empire north of the Rio Grande. Santa Fe is the oldest capital city in North America and the oldest European community west of the Mississippi.

While Santa Fe was inhabited on a very small scale in 1607, it was truly settled by the conquistador Don Pedro de Peralta in 1609-1610. Santa Fe is the site of both the oldest public building in America, the Palace of the Governors and the nation's oldest community celebration, the Santa Fe Fiesta, established in 1712 to commemorate the Spanish reconquest of New Mexico in the summer of 1692. Peralta and his men laid out the plan for Santa Fe at the base of the Sangre de Cristo Mountains on the site of the ancient Pueblo Indian ruin of Kaupoge, or "place of shell beads near the water."

The city has been the capital for the Spanish "Kingdom of New Mexico," the Mexican province of Nuevo Mejico, the American territory of New Mexico (which contained what is today Arizona and New Mexico) and since 1912 the state of New Mexico. Santa Fe, in fact, was the first foreign capital over taken by the United States, when in 1846 General Stephen Watts Kearny captured it during the Mexican-American War.

Santa Fe's history may be divided into six periods:

Preconquest and Founding
(circa 1050 to 1607)

Santa Fe's site was originally occupied by a number of Pueblo Indian villages with founding dates from between 1050 to 1150. Most archaeologists agree that these sites were abandoned 200 years before the Spanish arrived. There is little evidence of their remains in Santa Fe today.

The "Kingdom of New Mexico" was first claimed for the Spanish Crown by the conquistador Don Francisco Vasques de Coronado in 1540, 67 years before the founding of Santa Fe. Coronado and his men also discovered the Grand Canyon and the Great Plains on their New Mexico expedition.

Don Juan de Onate became the first Governor-General of New Mexico and established his capital in 1598 at San Juan Pueblo, 25 miles north of Santa Fe. When Onate retired, Don Pedro de Peralta was appointed Governor-General in 1609. One year later, he had moved the capital to present day Santa Fe.

Settlement Revolt & Reconquest
(1607 to 1692)

For a period of 70 years beginning the early 17th century, Spanish soldiers and officials, as well as Franciscan missionaries, sought to subjugate and convert the Pueblo Indians of the region. The indigenous population at the time was close to 100,000 people, who spoke nine basic languages and lived in an estimated 70 multi-storied adobe towns (pueblos), many of which exist today. In 1680, Pueblo Indians revolted against the estimated 2,500 Spanish colonists in New Mexico, killing 400 of them and driving the rest back into Mexico. The conquering Pueblos sacked Santa Fe and burned most of the buildings, except the Palace of the Governors. Pueblo Indians occupied Santa Fe until 1692, when Don Diego de Vargas reconquered the region and entered the capital city after a bloodless siege.

Established Spanish Empire
(1692 to 1821)

Santa Fe grew and prospered as a city. Spanish authorities and missionaries - under pressure from constant raids by nomadic Indians and often bloody wars with the Comanches, Apaches and Navajos-formed an alliance with Pueblo Indians and maintained a successful religious and civil policy of peaceful coexistence. The Spanish policy of closed empire also heavily influenced the lives of most Santa Feans during these years as trade was restricted to Americans, British and French.

The Mexican Period
(1821 to 1846)

When Mexico gained its independence from Spain, Santa Fe became the capital of the province of New Mexico. The Spanish policy of closed empire ended, and American trappers and traders moved into the region. William Becknell opened the l,000-mile-long Santa Fe Trail, leaving from Arrow Rock, Missouri, with 21 men and a pack train of goods. In those days, aggressive Yankeetraders used Santa Fe's Plaza as a stock corral. Americans found Santa Fe and New Mexico not as exotic as they'd thought. One traveler called the region the "Siberia of the Mexican Republic."

For a brief period in 1837, northern New Mexico farmers rebelled against Mexican rule, killed the provincial governor in what has been called the Chimayó Rebellion (named after a village north of Santa Fe) and occupied the capital. The insurrectionists were soon defeated, however, and three years later, Santa Fe was peaceful enough to see the first planting of cottonwood trees around the Plaza.

Territorial Period
(1846 to 1912)

On August 18, 1846, in the early period of the Mexican American War, an American army general, Stephen Watts Kearny, took Santa Fe and raised the American flag over the Plaza. Two years later, Mexico signed the Treaty of Guadalupe Hidalgo, ceding New Mexico and California to the United States.

In 1851, Jean B. Lamy, arrived in Santa Fe. Eighteen years later, he began construction of the

Saint Francis Cathedral. Archbishop Lamy is the model for the leading character in Willa Cather's book, "Death Comes for the Archbishop."

For a few days in March 1863, the Confederate flag of General Henry Sibley flew over Santa Fe, until he was defeated by Union troops. With the arrival of the telegraph in 1868 and the coming of the Atchison, Topeka and the Santa Fe Railroad in 1880, Santa Fe and New Mexico underwent an economic revolution. Corruption in government, however, accompanied the growth, and President Rutherford B. Hayes appointed Lew Wallace as a territorial governor to "clean up New Mexico." Wallace did such a good job that Billy the Kid threatened to come up to Santa Fe and kill him. Thankfully, Billy failed and Wallace went on to finish his novel, "Ben Hur," while territorial Governor.

Statehood
(1912 to present)

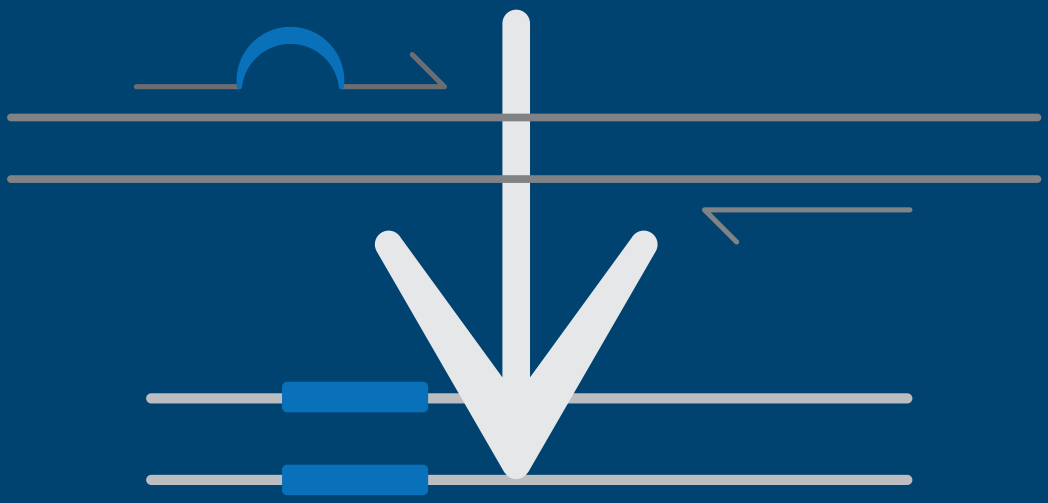
When New Mexico gained statehood in 1912, many people were drawn to Santa Fe's dry climate as a cure for tuberculosis. The Museum of New Mexico had opened in 1909, and by 1917, its Museum of Fine Arts was built. The state museum's emphasis on local history and native culture did much to reinforce Santa Fe's image as an "exotic" city.

Throughout Santa Fe's long and varied history of conquest and frontier violence, the town has also been the region's seat of culture and civilization. Inhabitants have left a legacy of architecture and city planning that today makes Santa Fe the most significant historic city in the American West.

In 1926, the Old Santa Fe Association was established, in the words of its bylaws, "to preserve and maintain the ancient landmarks, historical structures and traditions of Old Santa Fe, to guide its growth and development in such a way as to sacrifice as little as possible of that unique charm born of age, tradition and environment, which are the priceless assets and heritage of Old Santa Fe."

Today, Santa Fe is recognized as one of the most intriguing urban environments in the nation, due largely to the city's preservation of historic buildings and a modern zoning code, passed in 1958, that mandates the city's distinctive Spanish-Pueblo style of architecture, based on the adobe (mud and straw) and wood construction of the past. Also preserved are the traditions of the city's rich cultural heritage which helps make Santa Fe one of the country's most diverse and fascinating places to visit.

Ultramer™ Oligonucleotides



Oligos up to 200 bases
Available in tubes or plates

Applications:

— Cloning
— Mutagenesis
— Gene construction
— shRNA

THE CUSTOM BIOLOGY COMPANY

WWW.IDTDNA.COM

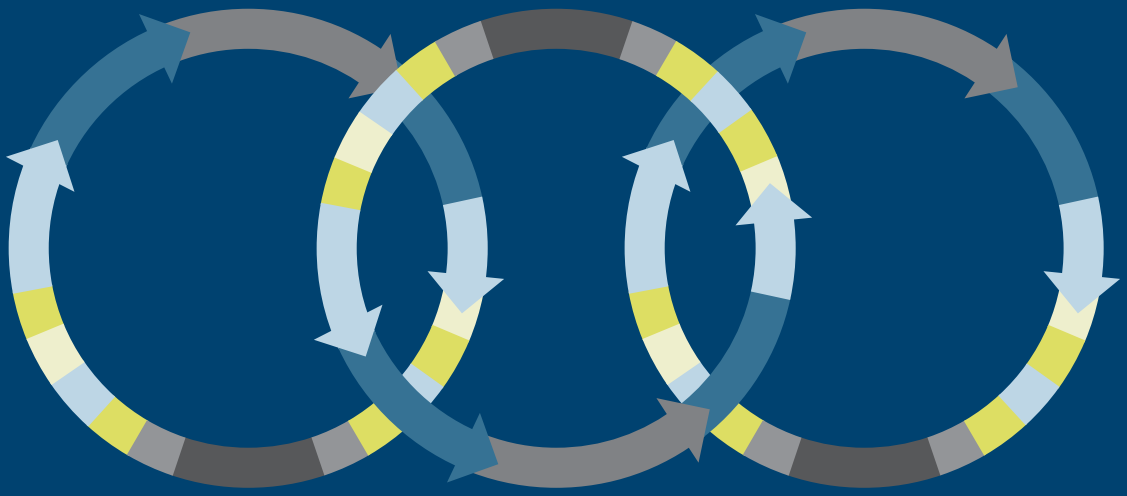
IDT®
INTEGRATED DNA TECHNOLOGIES

Gene Synthesis

miniGenes™ (up to 400 bp)

One flat rate

Shipped in 4-8 business days



Applications:

- **Gene variants and SNPs**
- **Standards for quantitative PCR**
- **Synthetic chromosomes/genomes**
- **Functional genomics**
- **Protein expression**

THE CUSTOM BIOLOGY COMPANY

WWW.IDTDNA.COM

IDT®

INTEGRATED DNA TECHNOLOGIES

http://www.roche-diagnostics.us/
Meet and Greet Party

http://www.lifetechnologies.com
Happy Hour x2

http://www.caliperls.com/
Cooler Bags

http://www.pacificbiosciences.com/
Lunch

http://www.illumina.com/
Lunch

http://www.idtdna.com/Home/Home.aspx
Meeting Guides

http://www.neb.com
Fruit and Juice - Breakfast

NEW ENGLAND BioLabs Inc.

Agilent Technologies

http://www.home.agilent.com
Lunch

http://www.isilon.com/
Break

Covaris®
the sample prep advantage™

http://www.covarisinc.com/
Breaks

http://www.opgen.com/
Break

http://www.clcbio.com/
Break


Roche
Caliper
LifeSciences
life
technologies
illumina
IDT
INTEGRATED DNA
TECHNOLOGIES
Thank You !!!