DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Automated High-Throughput Fosmid Isolation and End-Sequencing Using Agencourt's SprintPrep and Reduced Terminator Cycling Sequencing Reaction Kit

Feng Chen

Technology Development

LBNL-57658

US DOE Joint Genome Institute

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Los Alamos NATIONAL LABORATORY

BERKELEY LAB

University of California Lawrence Livermore National Laboratory

shgc

Formed in 1997 as a MOU between DOE Labs LLNL, LBNL and LANL.

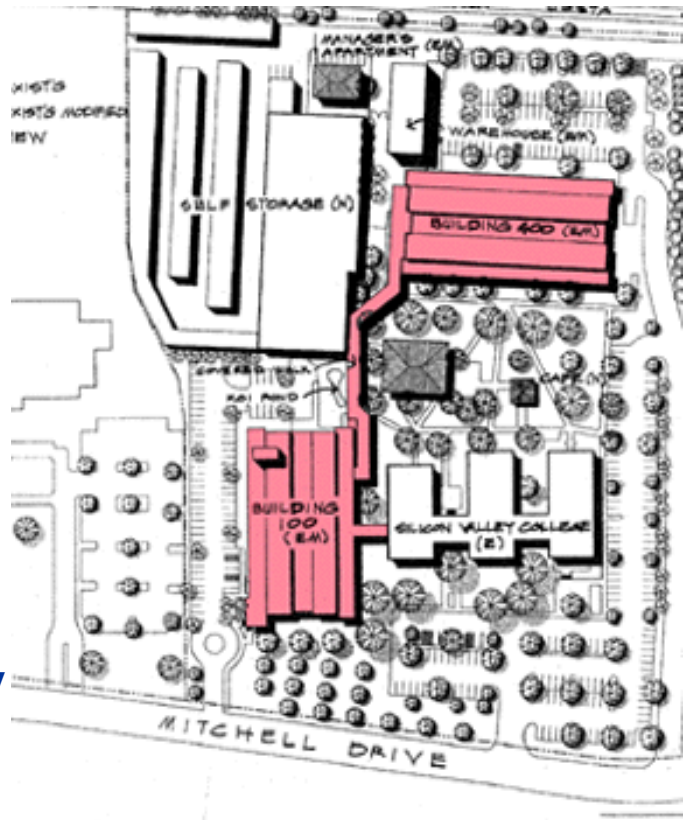
www.jgi.doe.gov

Office of Science
U.S. DEPARTMENT OF ENERGY

~250 FTEs

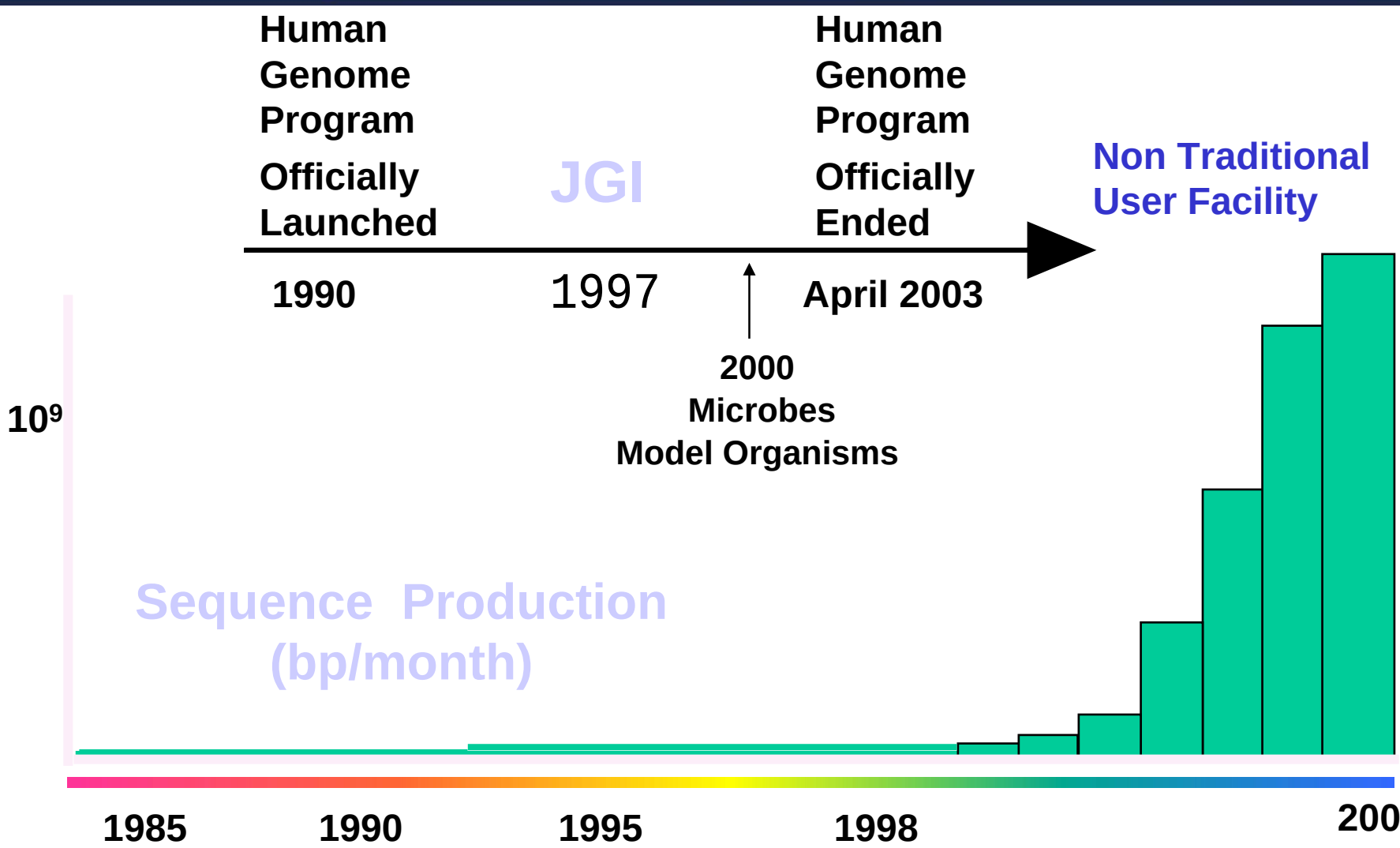
165 FTEs PGF
30 FTEs LANL
50 FTEs SHGC
5 FTEs LLNL
2-3 FTEs ORNL

PGF-Production Genomics Facility
Walnut Creek, CA
2 buildings-60,000 sq. ft.



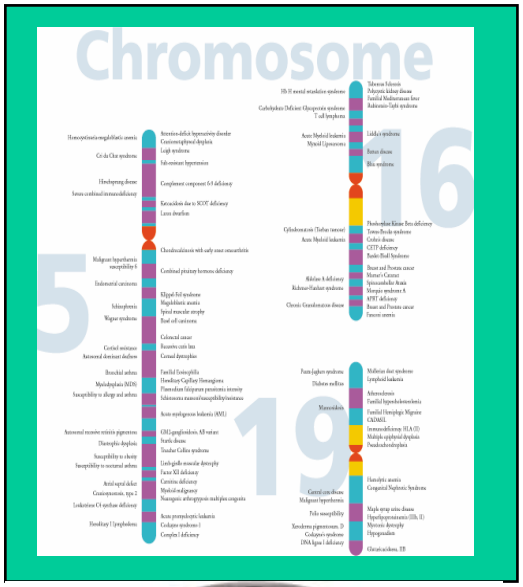
Advancing Science Through DNA Sequence

JGI History and Future



Completion of Three Human Chromosomes at JGI

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE



articles

The sequence and analysis of duplication-rich human chromosome 16

Joel Martin[1], Cliff Han[2], Laurie A. Gordon[1], Astrid Terry[1], Shyam Prabhakar[4], Xinwei She[5], Gary Xie[1,2], Uffe Hellsten[1], Yee Man Chan[6], Michael Altherr[1,2], Olivier Couronne[4], Andrea Aerts[1], Eva Bajorek[6], Stacey Black[6], Heather Blumer[2], Elbert Branscomb[1,3], Nancy C. Brown[2], William J. Bruno[2], Judith M. Buckingham[2], David F. Callen[2], Connie S. Campbell[2], Mary L. Campbell[2], Evelyn W. Campbell[2], Chenier Caoile[6], Jean F. Challacombe[2], Leslie A. Chasteen[2], Olga Chertkov[2], Han C. Chi[2], Mari Christensen[3], Lynn M. Clark[2], Judith D. Cohn[2], Mirian Denys[6], John C. Detter[1], Mark Dickson[6], Mira Dimitrijevic-Bussod[2], Julio Escobar[6], Joseph J. Fawcett[2], Dave Flowers[6], Dea Fotopulos[6], Tijana Glavina[1], Maria Gomez[6], Eidelyn Gonzales[6], David Goodstein[1], Lynne A. Goodwin[2], Deborah L. Grady[2], Igor Grigoriev[1], Matthew Groza[3], Nancy Hammon[1], Trevor Hawkins[1], Lauren Haydu[6], Carl E. Hildebrand[2], Wayne Huang[1], Sanjay Israni[1], Jamie Jett[1], Phillip B. Jewett[2], Kristen Kadner[1], Heather Kimball[1], Arthur Kobayashi[1,3], Marie-Claude Krawczyk[2], Tina Leyba[2], Jonathan L. Longmire[2], Frederick Lopez[6], Yunlan Lou[1], Steve Lowry[1], Thom Ludeman[2], Chitra F. Manohar[3], Graham A. Mark[2], Kimberly L. McMurray[2], Linda J. Meincke[2], Jenna Morgan[1], Robert K. Moyzis[2], Mark O. Mundt[2], A. Christine Munk[2], Richard D. Nandkeshwar[3], Sam Pitluck[1], Martin Pollard[1], Paul Predki[1], Beverly Parson-Quintana[2], Lucia Ramirez[6], Sam Rash[1], James Retterer[6], Darryl O. Ricke[2], Donna L. Robinson[2], Alex Rodriguez[6], Asaf Salamov[1], Elizabeth H. Saunders[2], Duncan Scott[1], Timothy Shough[2], Raymond L. Stallings[2], Malinda Stalvey[2], Robert D. Sutherland[2], Roxanne Tapia[2], Judith G. Tesmer[2], Nina Thayer[1,2], Linda S. Thompson[2], Hope Tice[1], David C. Torney[2], Mary Tran-Gyamfi[1], Ming Tsai[6], Levy E. Ulanovsky[2], Anna Ustaszewska[1], Nu Vo[6], P. Scott White[2], Albert L. Williams[2], Patricia L. Wills[2], Jung-Rung Wu[2], Kevin Wu[6], Joan Yang[6], Pieter DeJong[7], David Bruce[2], Norman A. Doggett[2], Larry Deaven[2], Jeremy Schmutz[6], Jane Grimwood[6], Paul Richardson[1], Daniel S. Rokhsar[1], Evan E. Eichler[5], Paul Gilna[2], Susan M. Lucas[1], Richard M. Myers[6], Edward M. Rubin[1,4] & Len A. Pennacchio[1,4]

[1]*DOE Joint Genome Institute, 2800 Mitchell Avenue, Walnut Creek, California 94598, USA*
[2]*Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[3]*Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA*
[4]*Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, California 94720, USA*
[5]*Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA*
[6]*Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, 975 California Avenue, Palo Alto, California 94304, USA*
[7]*Children's Hospital Oakland, Oakland, California 94609, USA*

Nature 432, 988 - 994 (23 December 2004)

JGI as Non-Traditional User Facility

Users:

- DOE Microbial Program
- Other Governmental Agencies
- Community Sequencing Program (CSP)
  - Will provide the scientific community access to high throughput sequencing at the JGI
  - A wide range of projects will be accepted. Ultimately, the most important factor in determining acceptance is a project's scientific merit
  - The deliverables can range from raw sequence traces to well-annotated assembled genomes

Fosmid End-Sequencing

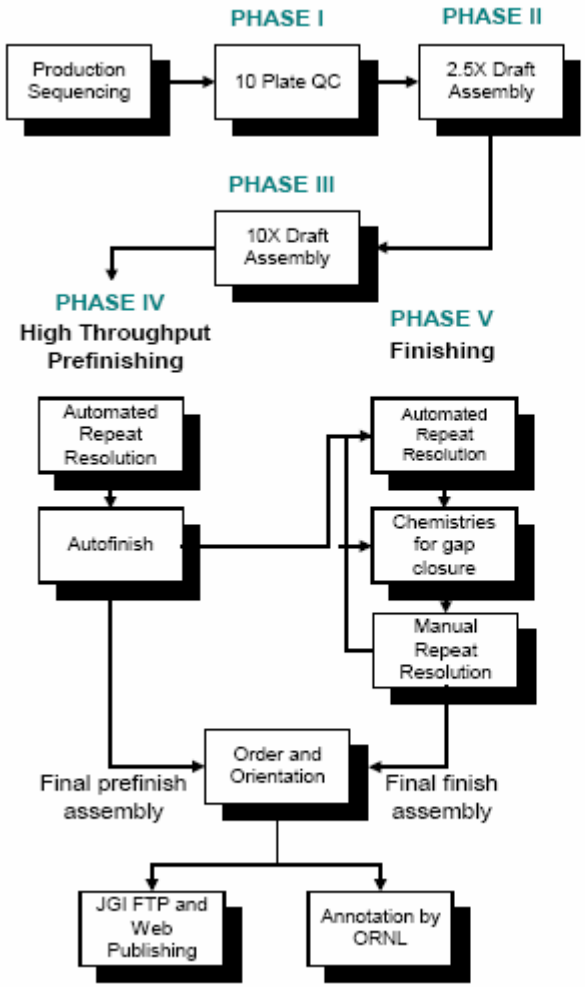
- **Fosmid end-sequencing is critical in whole genome shotgun sequencing**
  - **Building assembly scaffold**
  - **Filling gaps and bridging contigs in finishing process**
- **Obstacles for fosmid end-sequencing**
  - **Cost of sequencing**
  - **Low copy number and low DNA yields**
  - **Labor intensive and difficult-to-automate isolation procedure**

Whole Genome Shotgun Sequencing Strategies

- Shotgun small-insert sequencing
  - 3 kb and 8 kb libraries
  - 10x coverage draft
- Fosmid scaffolding (15x clone coverage)

  ~ 0.5x sequencing coverage
- Assembling
- Prefinishing
- Finishing
- Annotation

WGS Sequencing Strategies Flow Chart



PHASE I - *10 Plate QC*
10 plates are sequenced and QC performed to look for contamination.

PHASE II - *2.5 Draft Assembly*
Draft sequence is peformed to 2.5X coverage. QC is performed to look for contamination.

PHASE III - *10X Draft Assembly*
Draft sequence is performed to 10X coverage. Final draft assembly is done and flagged for Finishing.

PHASE IV - *High Througput Prefinishing*
Semi-automated Prefinishing is accomplished by resolving misassemblies and closing gaps <3kb through Autofinish. Once done, the assembly is order and oriented and the results are sent to ORNL for annotation and posted on the JGI FTP site for public access.

PHASE V - *Finishing*
Assembled contigs from Phase IV are analyzed for gaps and misassemblies. Automated repeat resolution, manual repeat resolution and primer walking are performed in an iterative process to resolve misassembled regions and close remaining gaps. The final assembly is order and oriented and the results are sent to ORNL for annotation and posted on the JGI FTP site for public access.

Fosmid end sequencing critical for finishing

Some Species Sequenced at JGI by WGS

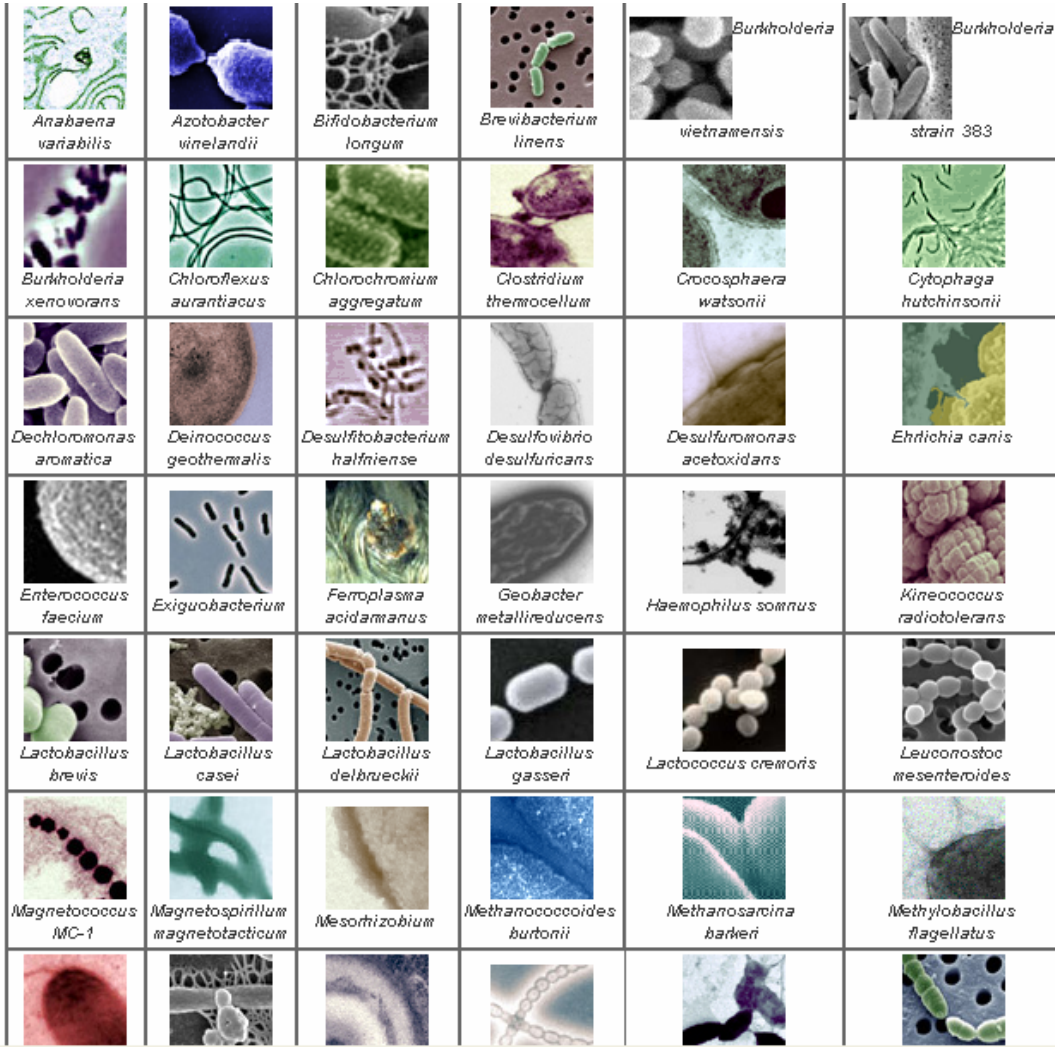
- **Chlamydomonas reinhardtii** (green alga)
- **Emiliania huxleyi** (marine coccolith)
- **Phanerochaete chrysosporium** (white rod fungus)
- **Daphnia pulex** (water flea)
- **Branchiostoma Floridae** (Florida lancelet)
- **Thalassiosira pseudonana** (diatom)
- **AOM** (anaerobic oxidation of methane) **microbial community**
- **Many other microbial genomes**

Diversity of species

Some of these species are G/C rich

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

How many do you recognize?



Fosmid Isolation Methods

- **Filtration based, high yields but labor intensive and time consuming**
  - **Qiagen's REAL prep**
  - **Millipore's Montage BAC_{96}**
  - **Other kits**
- **SPRI magnetic bead based, lower yields but easy to automate and quick**
  - **Agencourt's CosMCPrep**
  - **Agencourt's SprintPrep**
  - **Other SPRI kits**

Fosmid DNA Isolation Procedure Using SprintPrep

Cell culture growth

2-YT in 96-well plate
20 hours 37°C, 85% humidity and 600 rpm

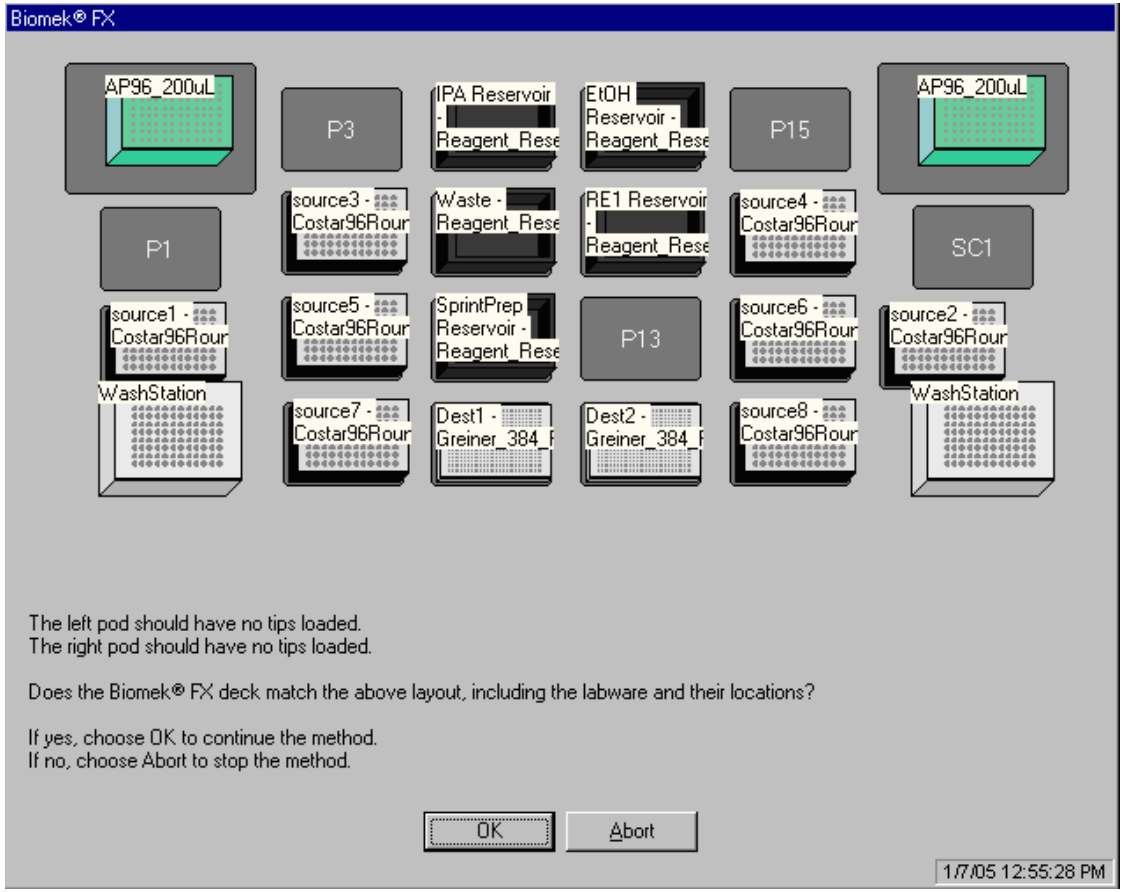
One-step lysis and DNA binding

Add SprintPrep solution
Add isopropanol
Mix 36 times and incubate for 3 min
Incubate on magnet for 5 min
Wash with 70% ethanol for 6 times
Blowing dry at 37°C for 9 min

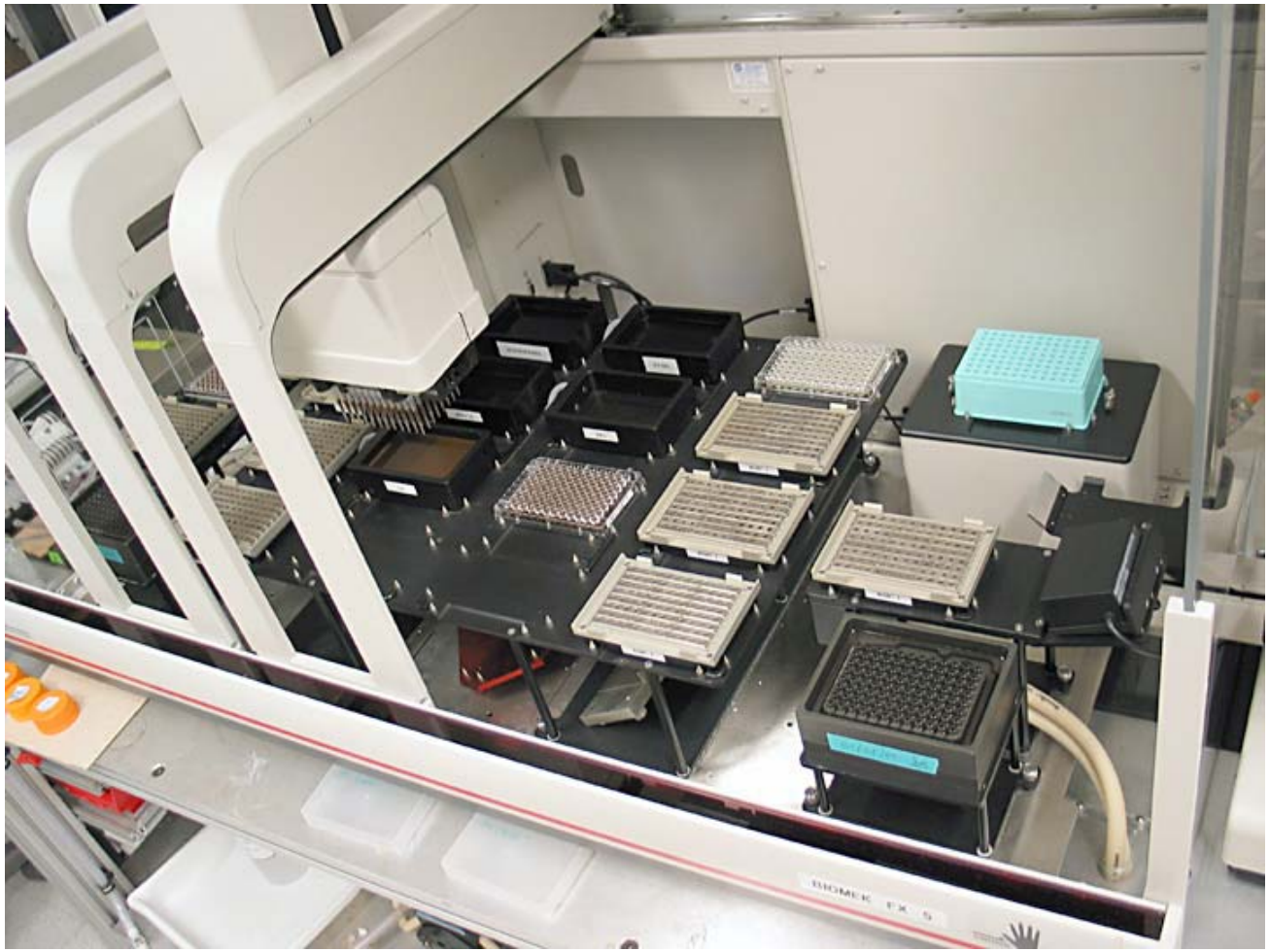
Elution

RE1 with 0.0625% of Triton X-100

Automation on Biomek FX

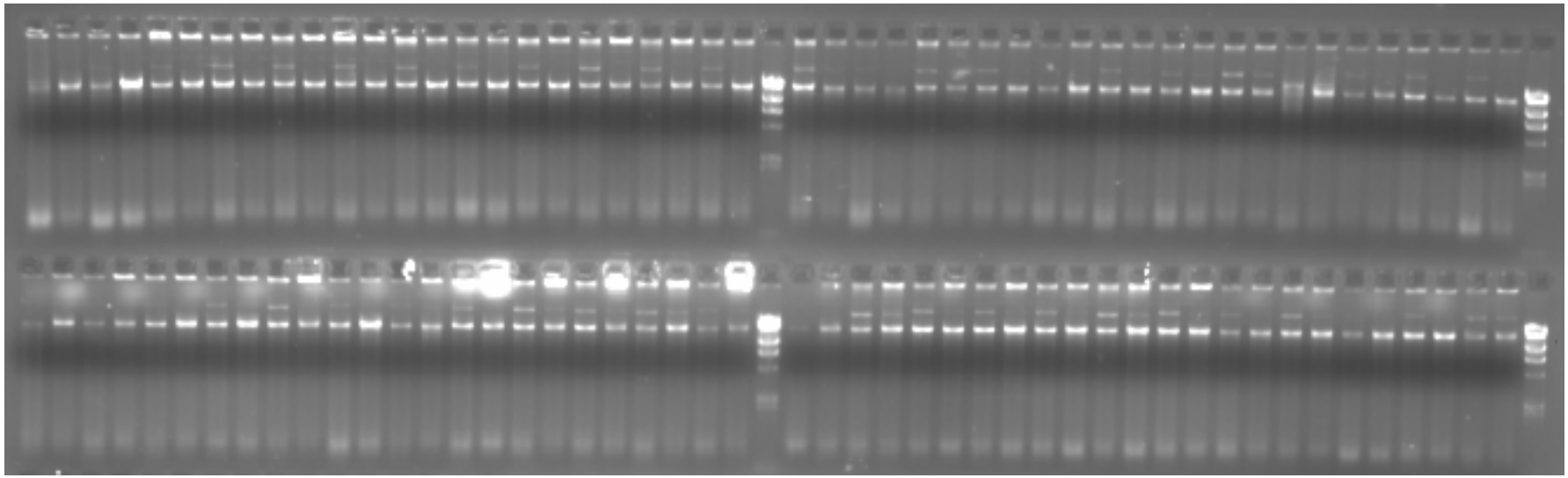


Automation on Biomek FX



Automation on Biomek FX

Prep Optimization

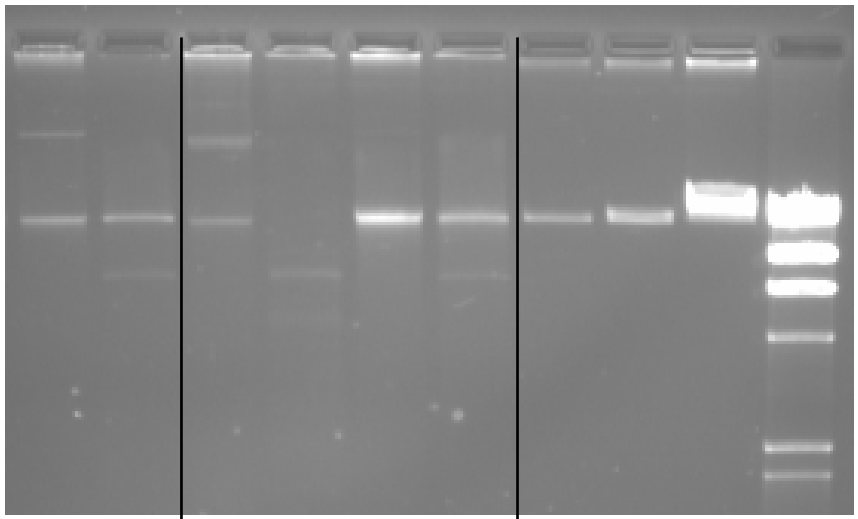


Alternating lanes show DNA from different prep conditions

- Culture volume and condition
- Amount of SprintPrep
- Wash times
- Length of drying

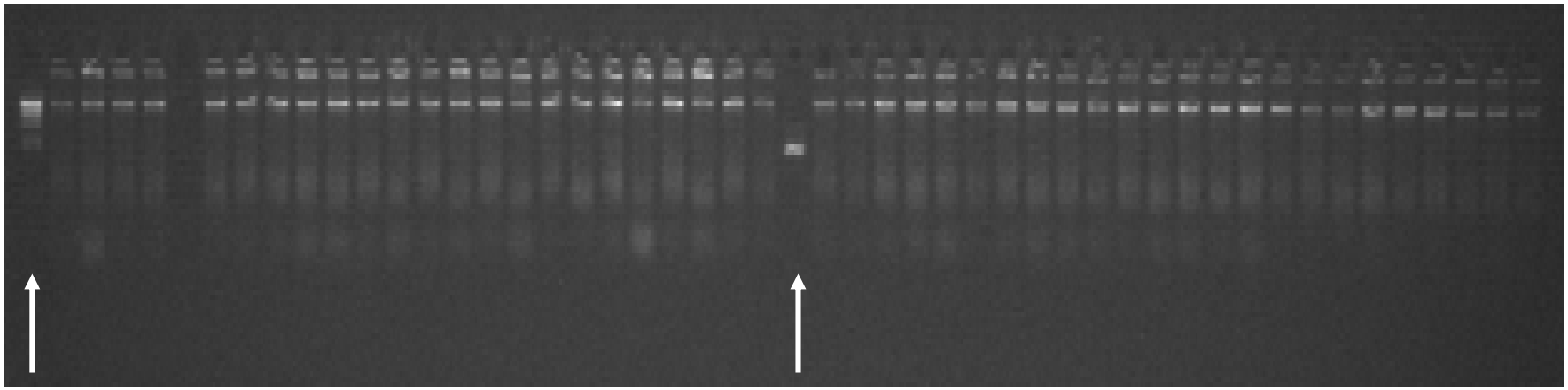
Prep Results

NotI digest



Production Results

QC Agarose gel from 11-29-04 production



Throughput: 8 96-well plates in 1.5 hours

Sequencing Methods

- 1/10 or 1/16 BigDye terminator reaction
  with or without DMSO
- 16% of fosmid DNA from SprintPrep product
- 6 ul total reaction volume
- 99 thermocycles
- Standard magnetic beads clean-up
- ABI 3730xl detection with modified run condition

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

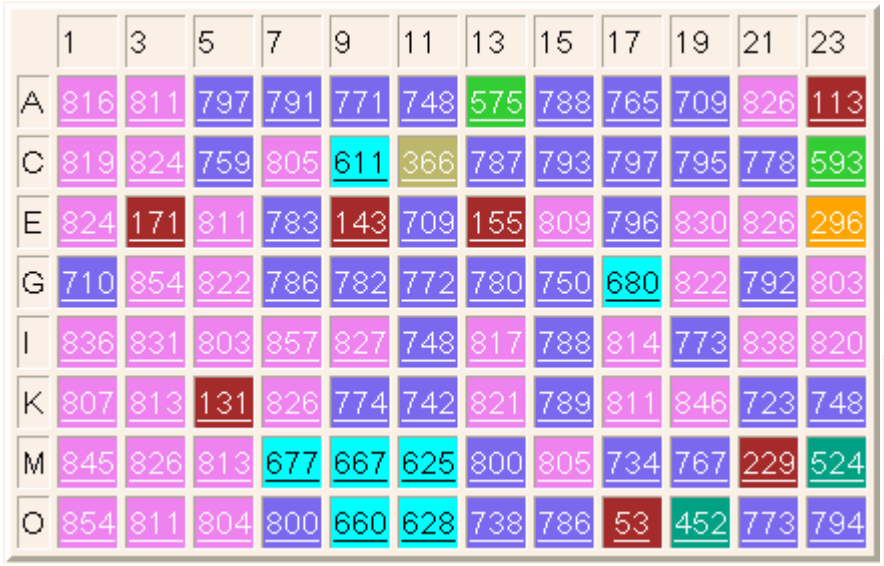
Sequencing Results: Trace View

Read1 (1-140)

Read2 (415-560)

Read3 (660-825)

Sequencing Results: Plate View



	1	3	5	7	9	11	13	15	17	19	21	23
A	816	811	797	791	771	748	575	788	765	709	826	113
C	819	824	759	805	611	366	787	793	797	795	778	593
E	824	171	811	783	143	709	155	809	796	830	826	296
G	710	854	822	786	782	772	780	750	680	822	792	803
I	836	831	803	857	827	748	817	788	814	773	838	820
K	807	813	131	826	774	742	821	789	811	846	723	748
M	845	826	813	677	667	625	800	805	734	767	229	524
O	854	811	804	800	660	628	738	786	53	452	773	794

ASXY0009A

APWS1181A

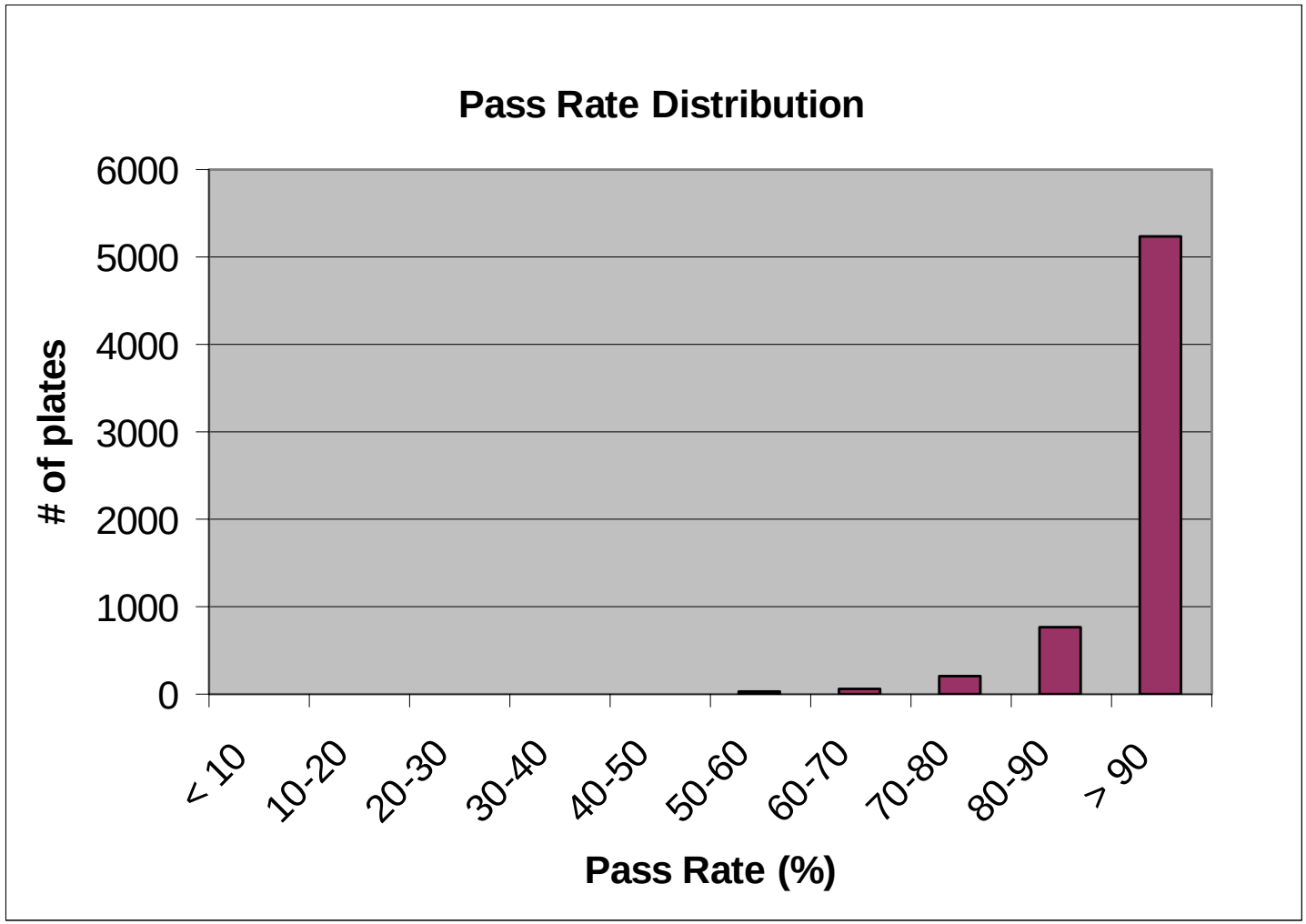
	1	3	5	7	9	11	13	15	17	19	21	23
A	516	787	794	788	793	768	841	800	852	745	283	789
C	782	833	257	811	849	850	800	757	851	829	831	705
E	795	838	720	845	814	839	811	844	833	867	832	844
G	790	826	820	764	790	852	815	695	677	842	812	86
I	706	841	720	840	788	845	832	611	738	831	623	726
K	739	747	814	866	853	814	850	748	815	591	830	794
M	720	690	815	846	818	724	807	831	792	756	794	81
O	725	740	766	800	739	796	738	782	729	729	690	685

0-1 | 2-250 | 251-350 | 351-450 | 451-525 | 526-600 | 601-700 | 701-800 | >800

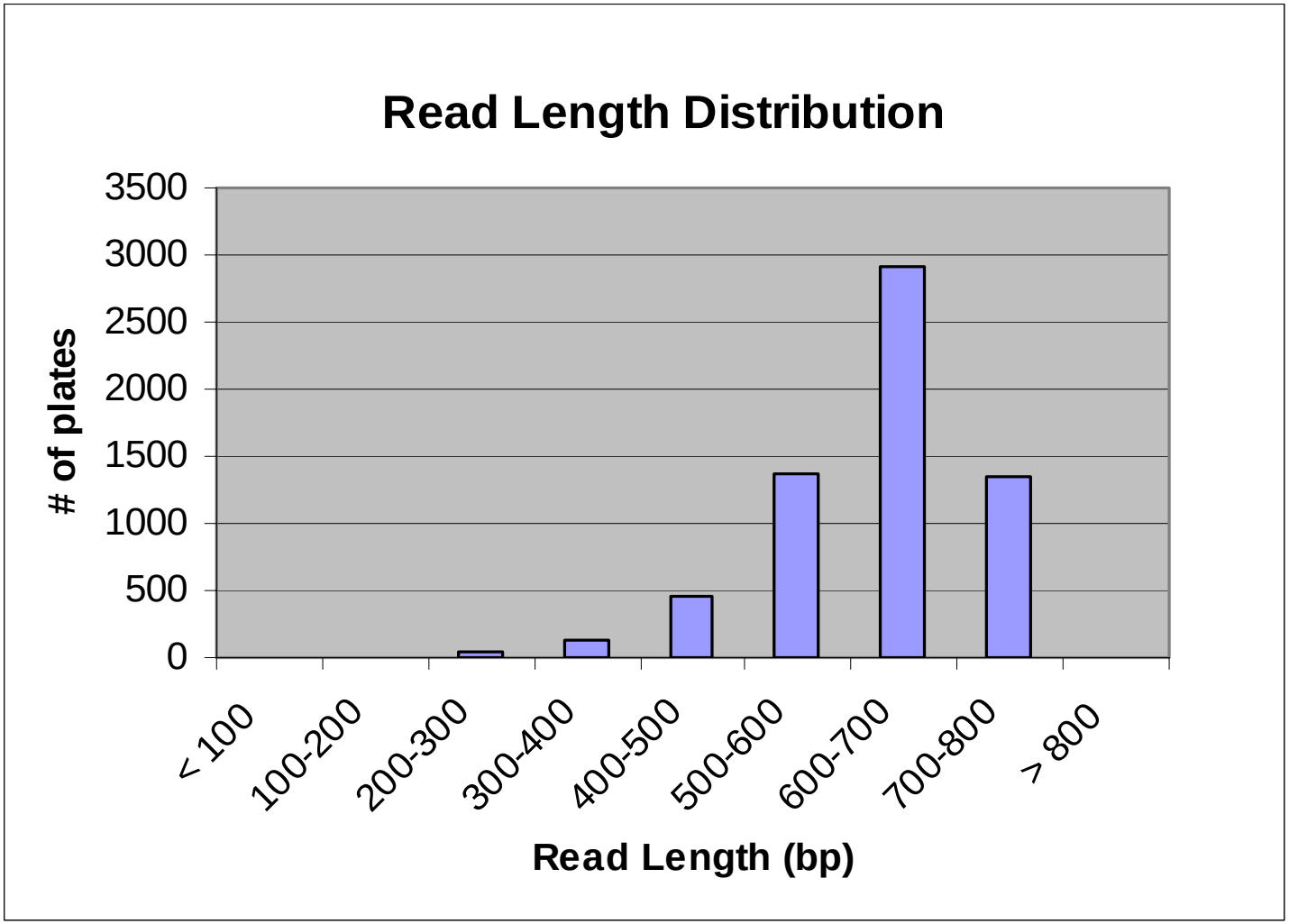
Summary of Sequencing Results
(from last 45 days of 2004 production)

Pass Rate (> 50 bp):	**93.4%**
“Good” Rate (> 450 bp):	**75.4%**
Average Read length (all lanes):	**626 bp (Q20)**
Paired ends (>50 bp):	**90%**

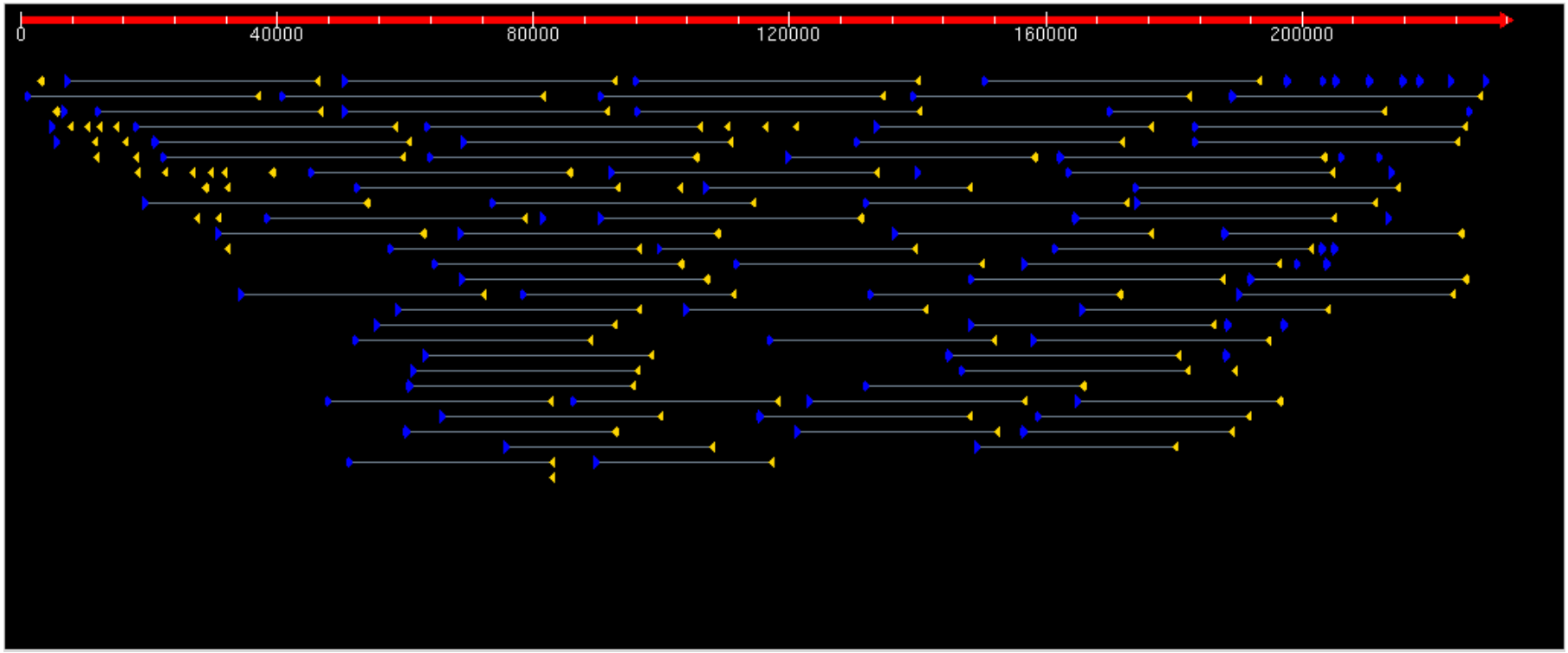
Result Summary



Result Summary

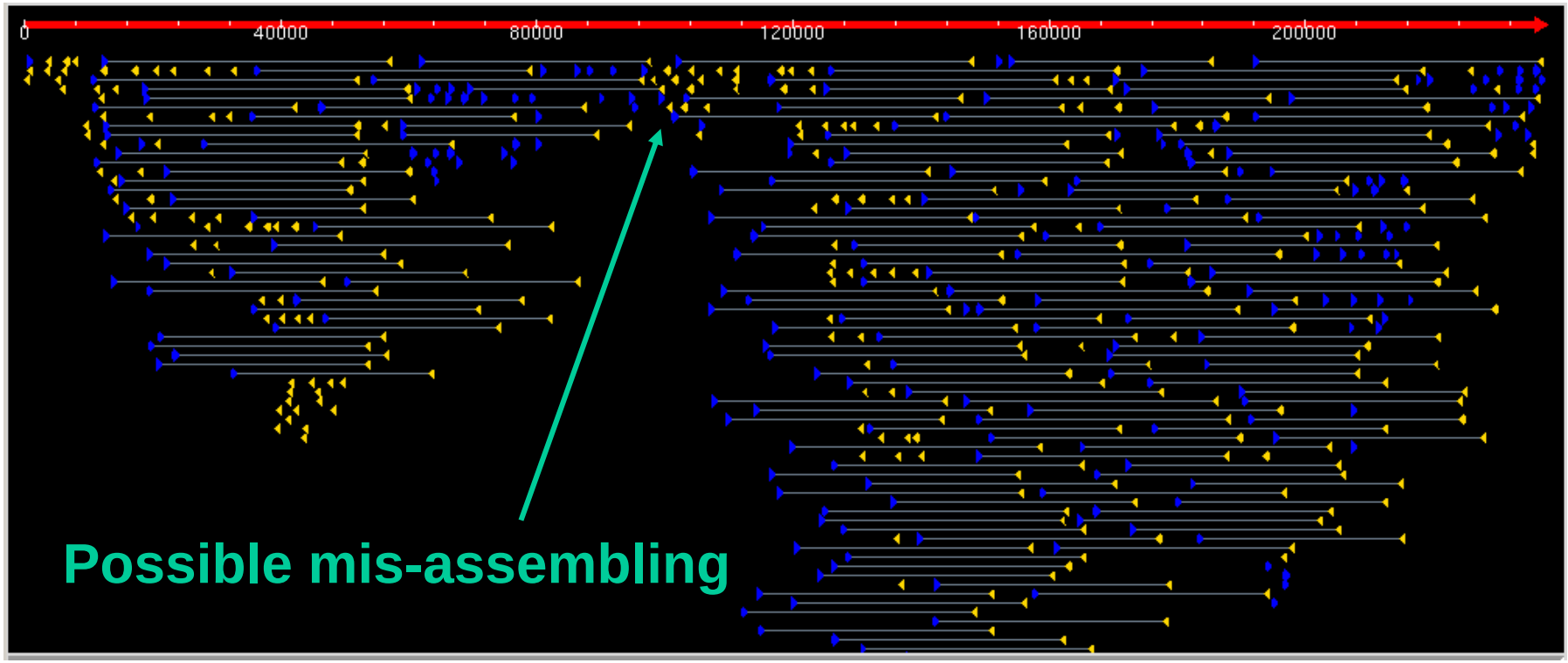


Assembly View with Fosmid Reads Aligned to Contig



Syntrophobacter fumaroxidans

Assembly View with Fosmid Reads Aligned to Contig



Syntrophomonas wolfei

Current Development

- **More automation**
  - Utilizing stackers and relaxing time constraint
- **Higher throughput**
  - From 80 96-well plates to 120
- **384-well format**
  - Reducing culture volume
  - Reducing wash volume
  - Automation

Acknowledgement

- **Genomic Technologies**

  Joe Alessi, Dou-Shuan Yang, Jamie Jett, and Paul Richardson

- **Library Construction**

  Chris Detter

- **Sequencing Production**

  Tijana Glavina, Marty Pollard, and Susan Lucas

- **Microbe Program**

  Alla Lapidus

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36. LBNL-57658

For comments, suggestions and protocol request:

Feng Chen, Ph.D.
Group Leader
Technology Development
US DOE Joint Genome Institute
925 296 5733
fchen@lbl.gov