Joint Genome Institute's High Throughput Production Sequencing Program

Susan Lucas

DOE Genome Contractor & Grantee Workshop

1/31/02

Outline

- **Production Genomics Facility (PGF)**
- **Production changes**
- **Future of PGF Production**

US DOE Joint Genome Institute
Production Genomics Facility (PGF)

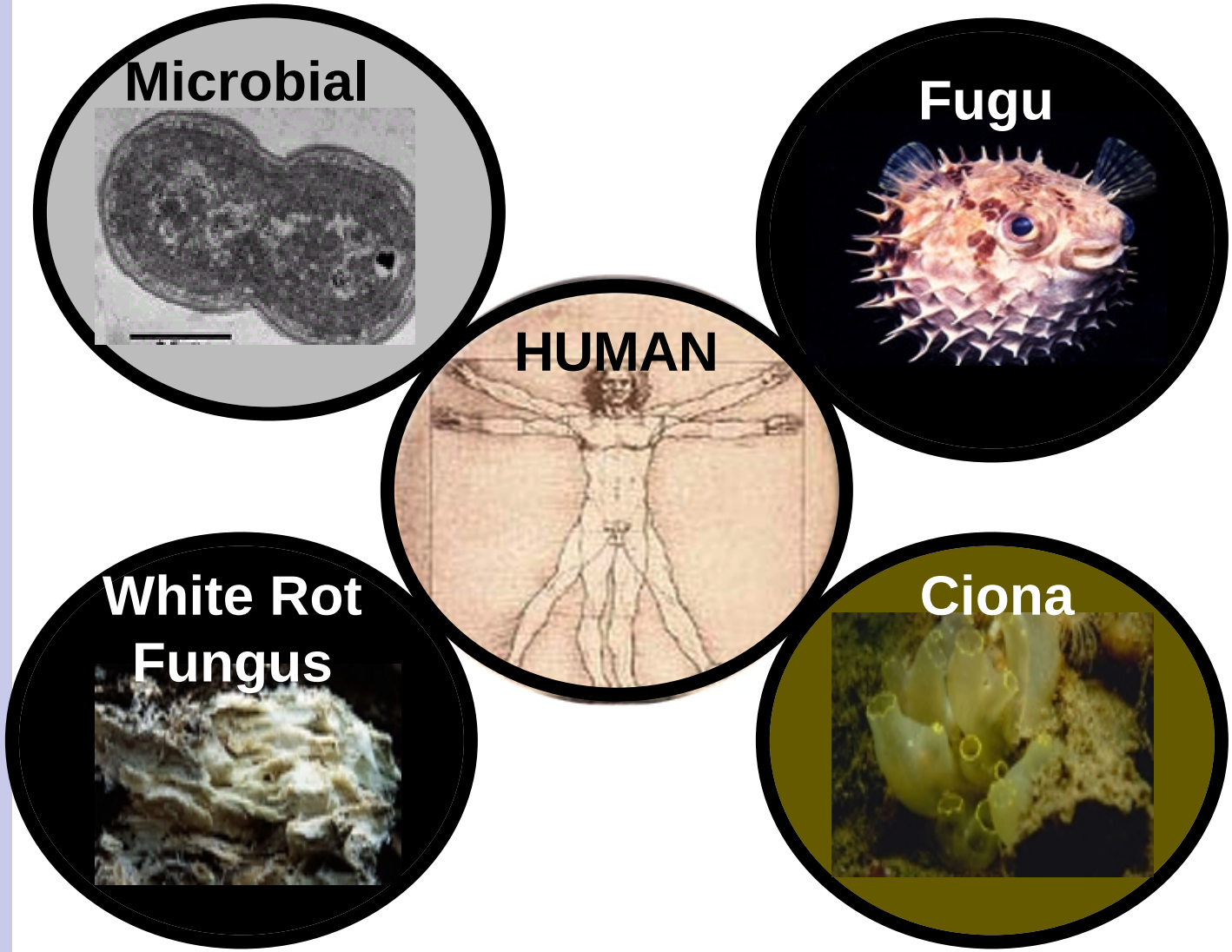
US DOE Joint Genome Institute

Based in Walnut Creek, CA
- ~250 FTEs
- ~$65M annual budget

250 FTE's total
- 150 FTEs PGF
- 60 FTEs LANL
- 25 FTEs SHGC
- 5 FTEs LLNL
- 10 FTEs ORNL

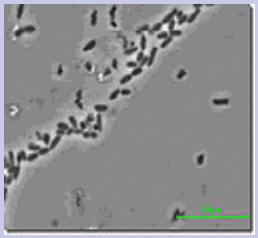
Production Genomics

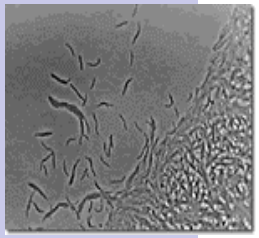
- Fugu draft completed October 2001
- Ciona draft in progress
- JGI mapped & sequenced regions of mouse syntenic to human chromosome 19
- Tools being developed for whole genome comparison

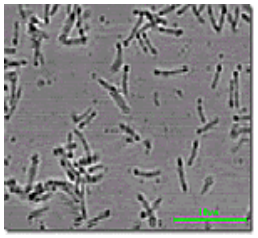


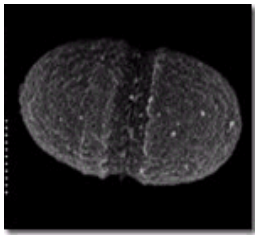
DOE JOINT GENOME INSTITUTE

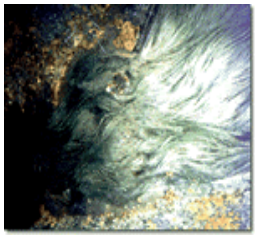
Microbial Genomes
October 2000


Burkholderia cepacia

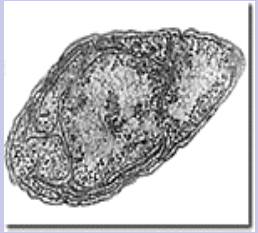

Cytophaga hutchinsonii


Desulfitobacterium halfniense


Enterococcus faecium

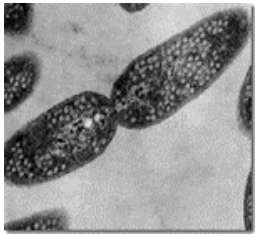

Ferroplasma acidarmanus

Magnetospirillum magnetotacticum

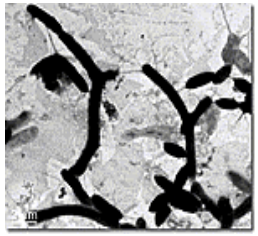

Nitrosomonas europaea

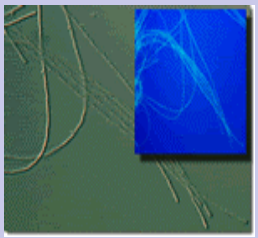
Prochlorococcus marinus

Pseudomonas fluorescens


Rhodobacter sphaeroides

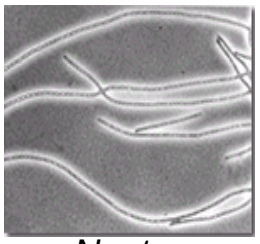
Rhodopseudomonas palustris


Sphingomonas aromaticivorans


Thermomonospora fusca

Trichodesmium erythraeum

Xylella fastidiosa


Nostoc punctiforme

Marine synechococcus

Magnetococcus MC-1

Production Genomics - Finishing

- Chromosomes 5,16,19 will be finished within next 12 months
- Stanford finishing Chromosome 5 &19
- LANL finishing Chromosome 16

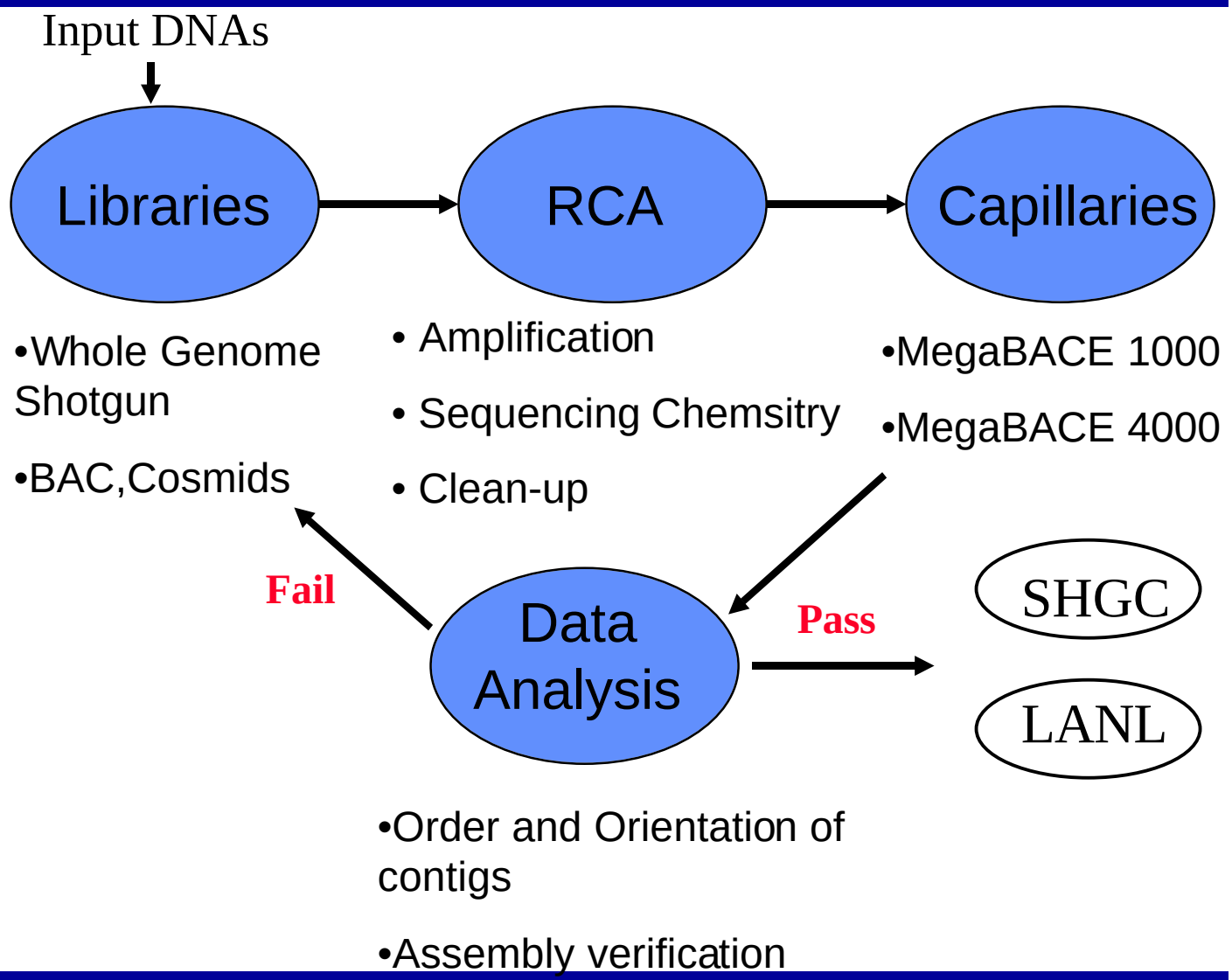
Current Production Pipeline

Production FTE's

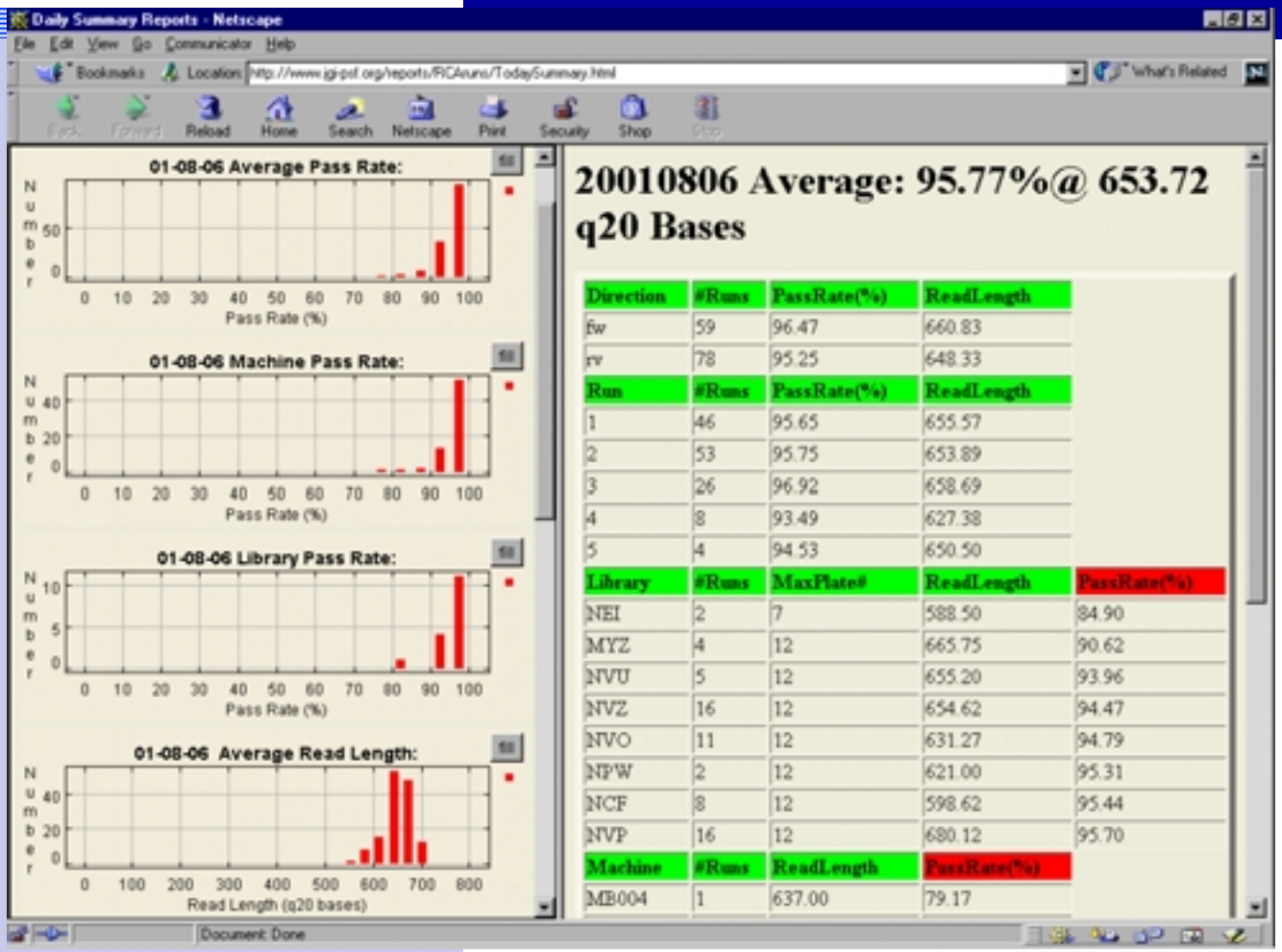
- **Libraries – 10**
- **RCA – 9**
- **Capillaries –17**

Production Support

- **Data Analysis – 4**
- **Production Instrumentation – 6**
- **QA/QC – 1**
- **Production Informatics - 7**



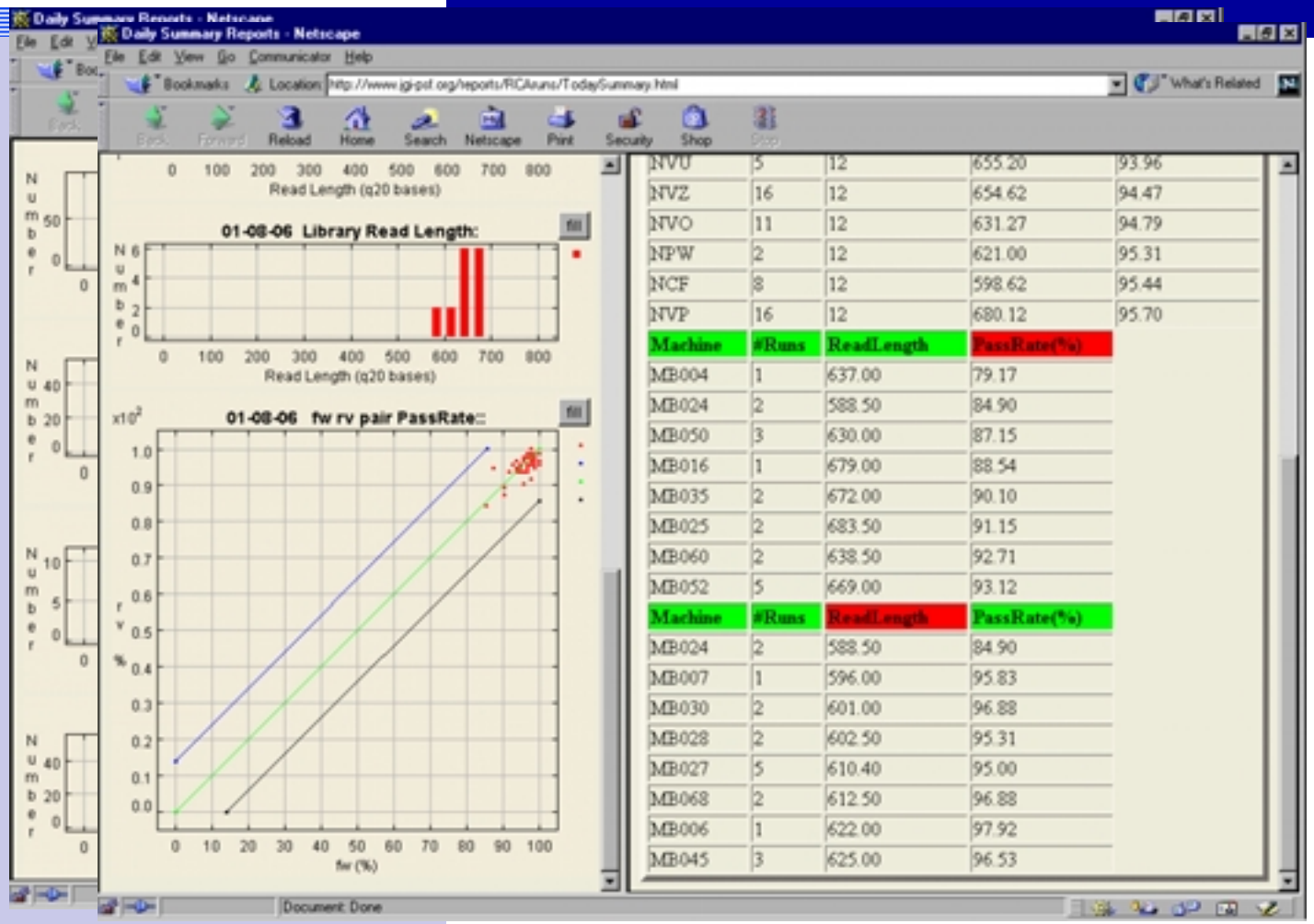
Online tracking of progress – LIMS



20010806 Average: 95.77%@ 653.72 q20 Bases

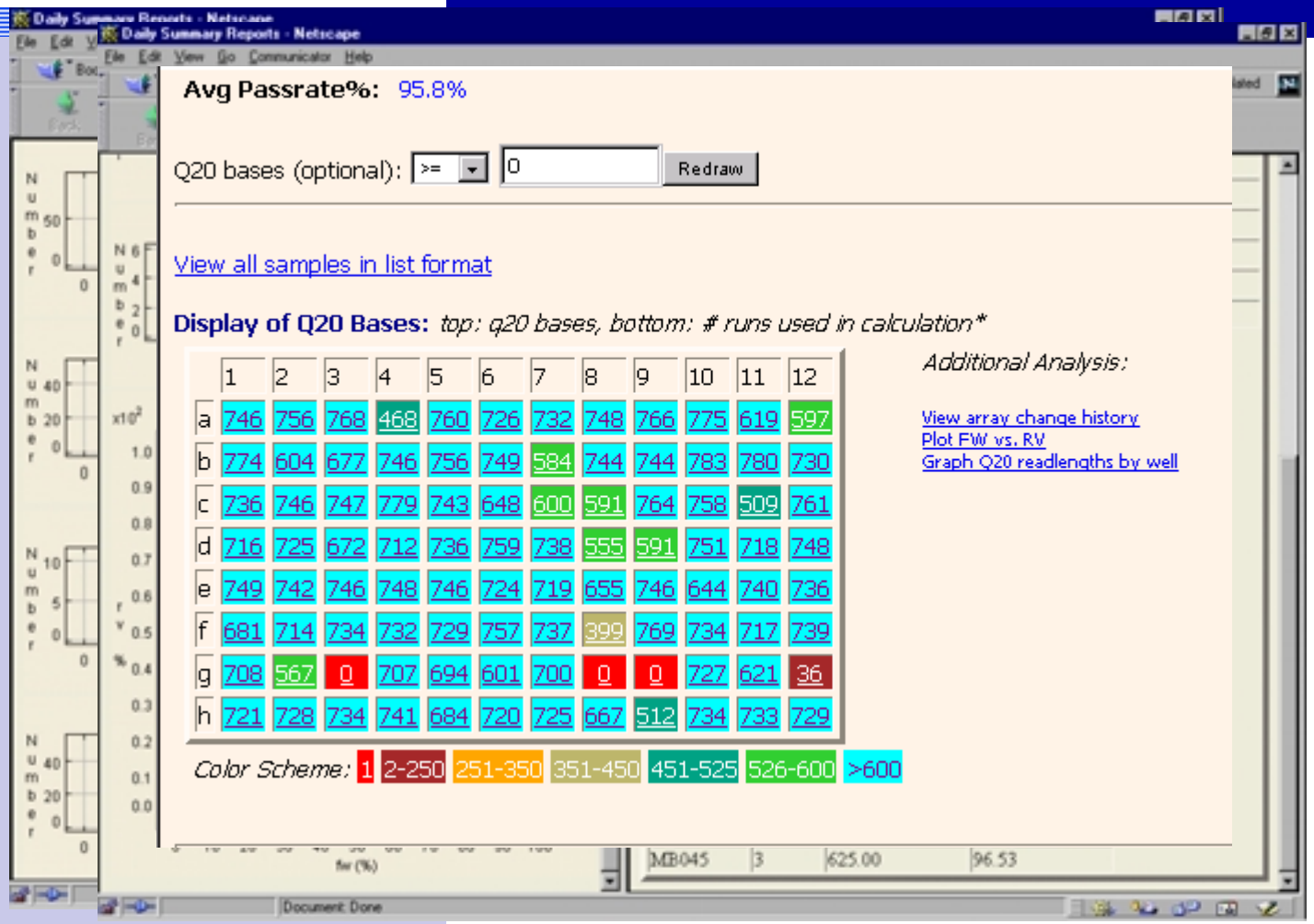
Direction	#Runs	PassRate(%)	ReadLength	
fw	59	96.47	660.83	
rv	78	95.25	648.33	
Run	**#Runs**	**PassRate(%)**	**ReadLength**	
1	46	95.65	655.57	
2	53	95.75	653.89	
3	26	96.92	658.69	
4	8	93.49	627.38	
5	4	94.53	650.50	
Library	**#Runs**	**MaxPlate#**	**ReadLength**	**PassRate(%)**
NEI	2	7	588.50	84.90
MYZ	4	12	665.75	90.62
NVU	5	12	655.20	93.96
NVZ	16	12	654.62	94.47
NVO	11	12	631.27	94.79
NPW	2	12	621.00	95.31
NCF	8	12	598.62	95.44
NVP	16	12	680.12	95.70
Machine	**#Runs**	**ReadLength**	**PassRate(%)**	
MB004	1	637.00	79.17	

Online tracking of progress – LIMS

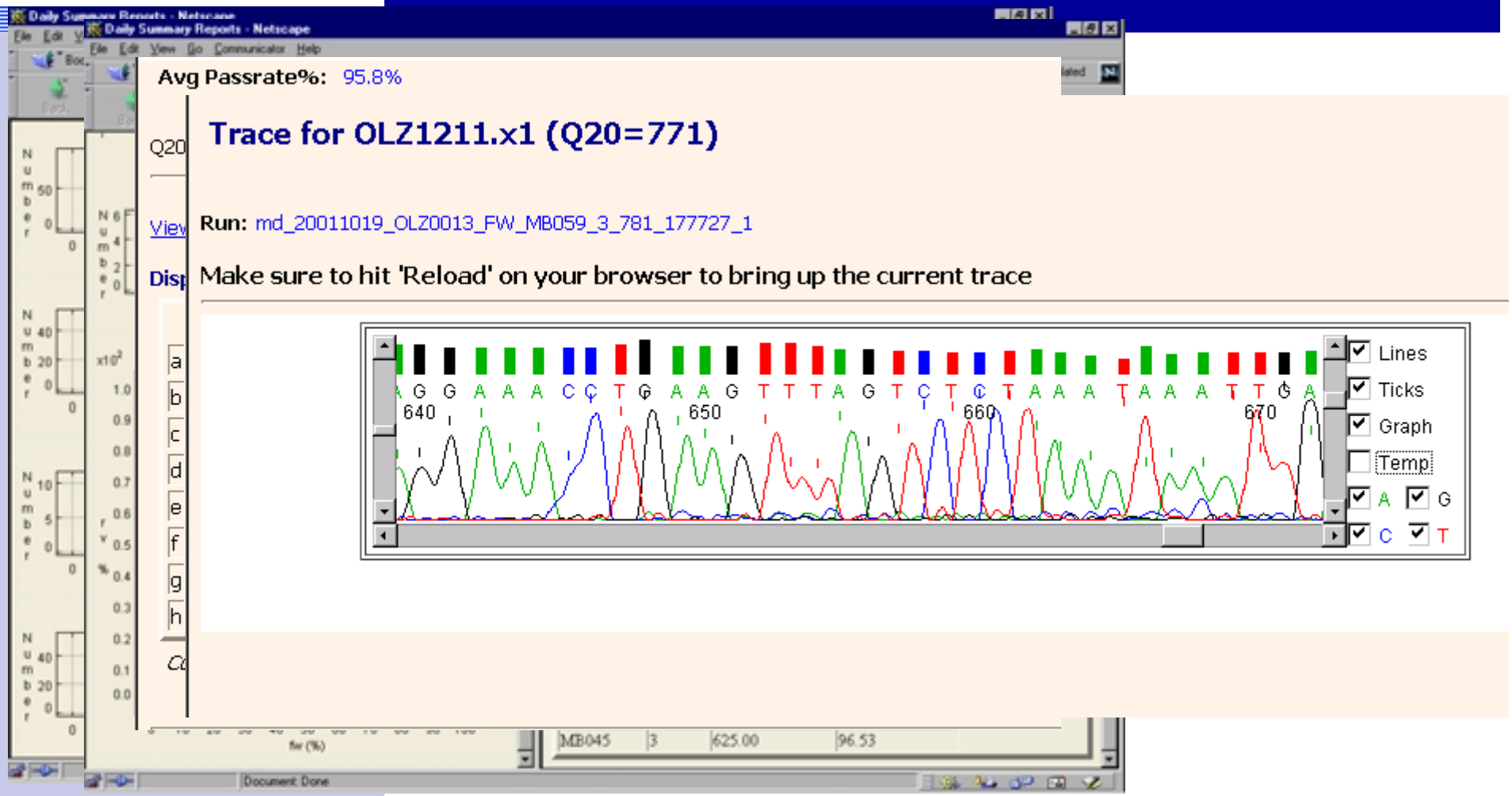


NVU	5	12	655.20	93.96
NVZ	16	12	654.62	94.47
NVO	11	12	631.27	94.79
NPW	2	12	621.00	95.31
NCF	8	12	598.62	95.44
NVP	16	12	680.12	95.70
Machine	#Runs	ReadLength	PassRate(%)	
MB004	1	637.00	79.17	
MB024	2	588.50	84.90	
MB050	3	630.00	87.15	
MB016	1	679.00	88.54	
MB035	2	672.00	90.10	
MB025	2	683.50	91.15	
MB060	2	638.50	92.71	
MB052	5	669.00	93.12	
Machine	#Runs	ReadLength	PassRate(%)	
MB024	2	588.50	84.90	
MB007	1	596.00	95.83	
MB030	2	601.00	96.88	
MB028	2	602.50	95.31	
MB027	5	610.40	95.00	
MB068	2	612.50	96.88	
MB006	1	622.00	97.92	
MB045	3	625.00	96.53	

Online tracking of progress – LIMS



Online tracking of progress – LIMS



Online tracking of progress – LIMS

Online tracking of progress – LIMS

MegaBACE Array QC

Main Page

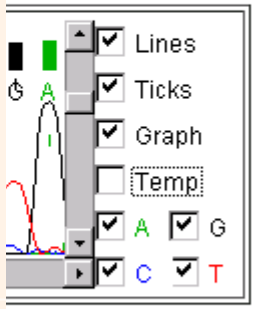
MegaBACE: MB084					Since last array change			
Array	Name	Q20 avg*	Q20>50 avg*	Last changed	Q20 avg+	Q20>50 avg+	%Failed	#Runs
6	ADE649	633	655	07/25/2001	633	655	3	427
2	ADE648	636	659	07/25/2001	636	659	3	427
1	ADE242	646	660	07/25/2001	646	660	2	427
5	ADE261	643	663	07/25/2001	643	663	3	427
3	ADE243	660	680	07/25/2001	660	680	3	427
4	ADE256	672	688	07/25/2001	672	688	2	427

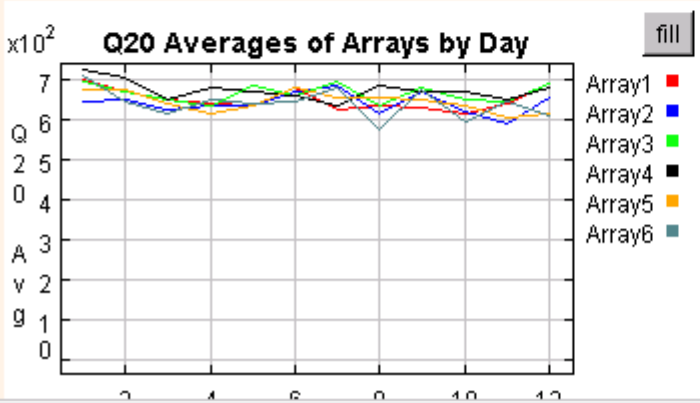
Date Range: 10/06/2001 to 10/19/2001
Keyword: md_
Total Runs: 68
Passrate Criteria: >=65

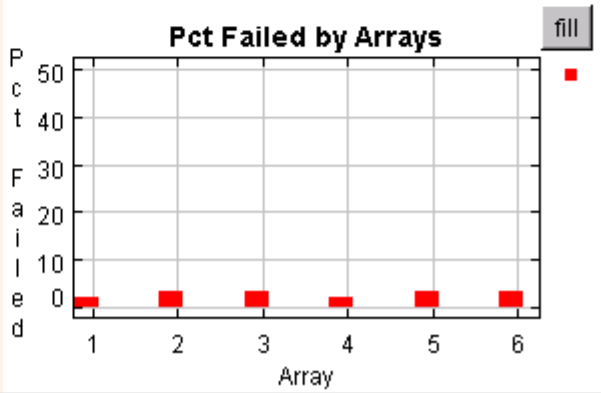
Arrays highlighed in red are based on arrays w/ q20>50 avg <450 since last changed and runs>0 since last changed or arrays w/ failrate% > 30 and runs>0 since last changed.

Note: Array q20 averages are based on runs w/ passrate >= 65 and keyword 'md_'.
#Runs in the 'Since last array change' heading represents all runs within the last array change date.
** denotes qc calculated between the beginning date specified to the ending date specified.*
+ denotes qc calculated between the last array change date or beginning date specified (whichever is later) to the ending date specified.

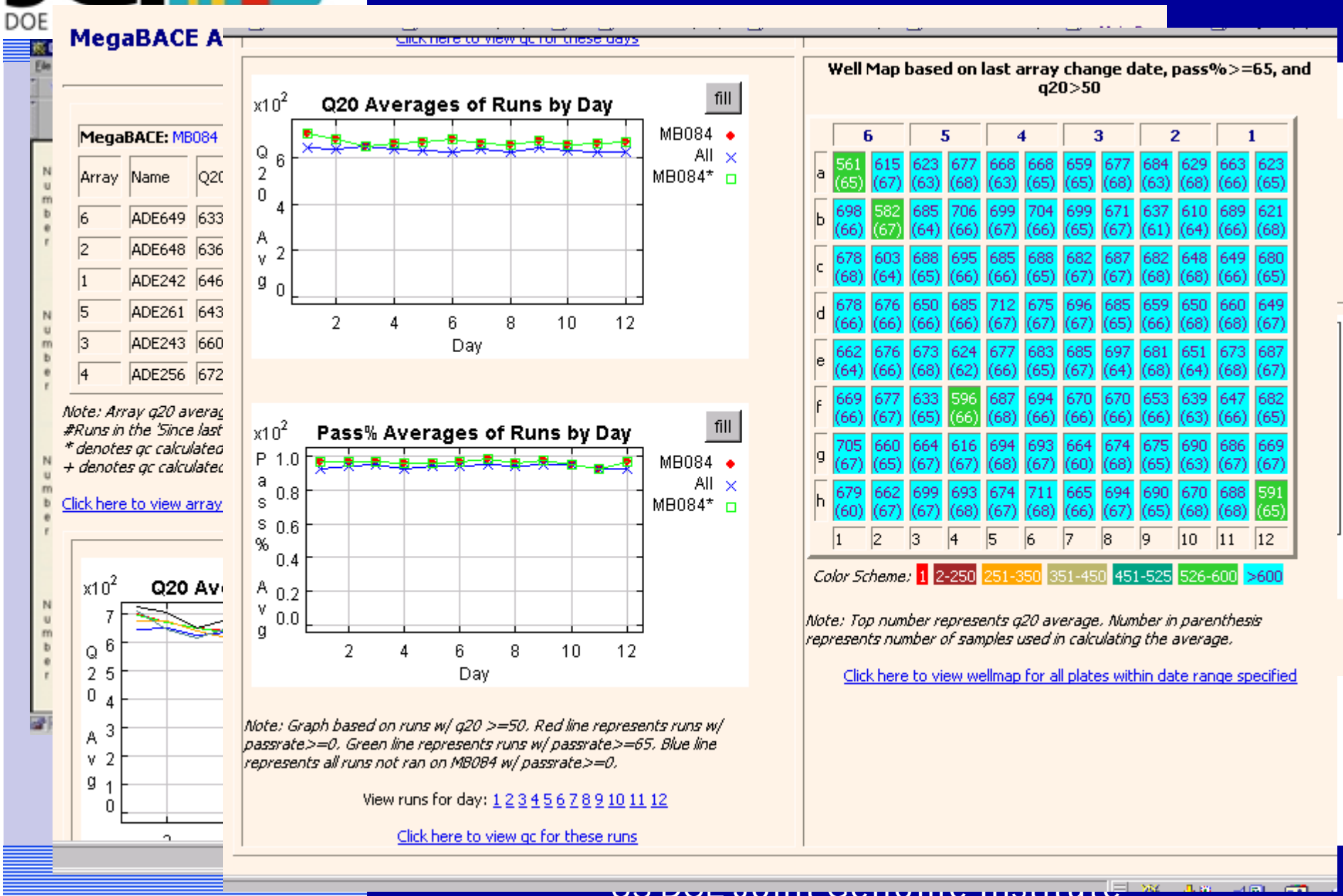
Click here to view array change history







Online tracking of progress – LIMS



Production Changes

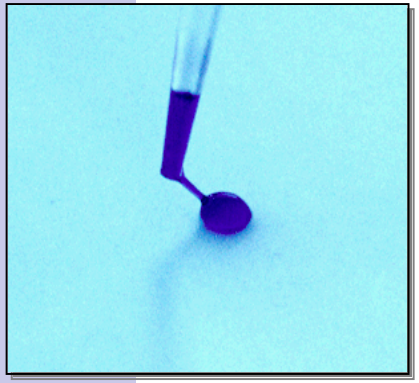
- **Implementation of Rolling Circle Amplification (RCA) - May '01**

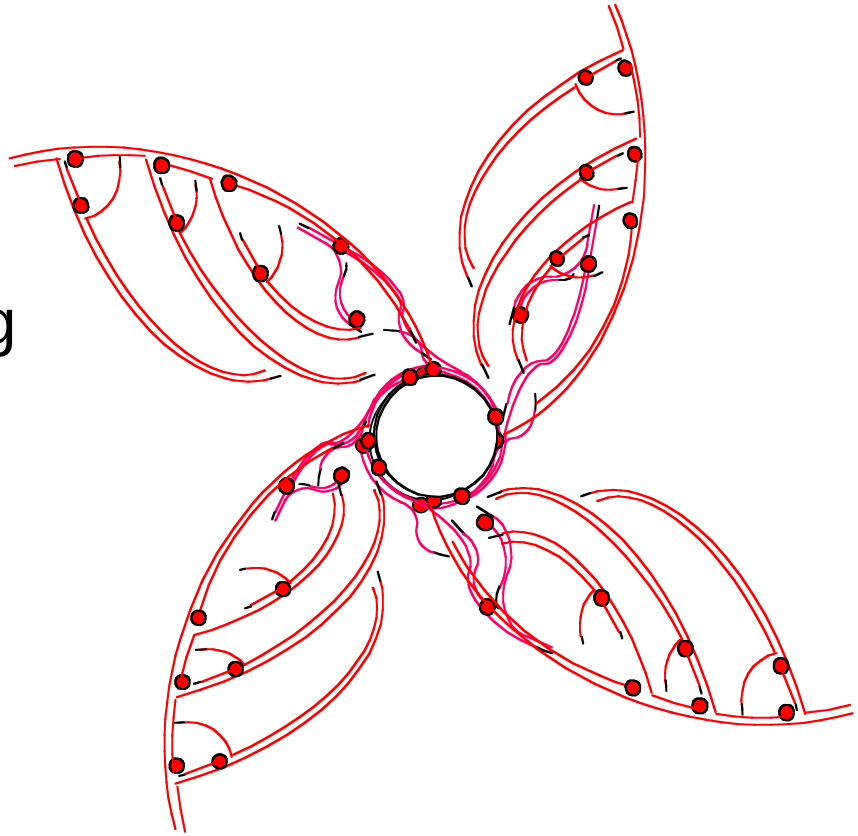
- **MegaBACE 4000 – December '01**

Rolling Circle Amplification
RCA

- **_29 DNA polymerase**
- **TempliPhi**

  -Isothermal reaction
  -Highly processive
  -Random hexamer priming
  -Strand displacement
  -Geometric amplification

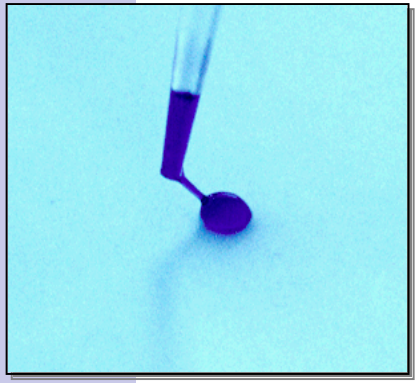


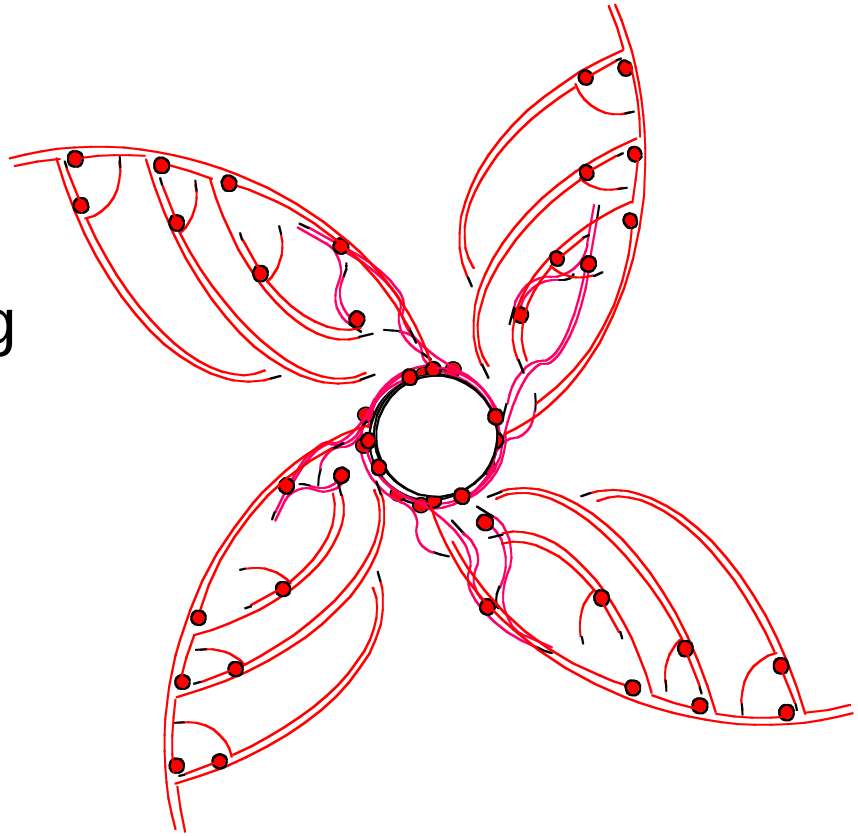


Rolling Circle Amplification
RCA

- **_29 DNA polymerase**
- **TempliPhi**

  -Isothermal reaction
  -Highly processive
  -Random hexamer priming
  -Strand displacement
  -Geometric amplification





Comparison of SPRI vs RCA: Protocols

SPRI

Solid Phase Reversible Immobilization

Inoculate from Glycerol Stocks

18 hour Growth – Higro Incubator

Lysis

Overnight Freeze

90 minute Centrifugation

Transfer Supernatant

Addition of PEG and Magnetic Beads

Incubation on Magnets

6 – 70% ETOH Washes

30 minute Air Dry

Addition of Water

QC - Agarose Gel

VS

RCA

Rolling Circle Amplification

Glycerol Stock Transfer

Heat Lysis

Addition of RCA Mix

18 hour Incubation

Heat Inactivation

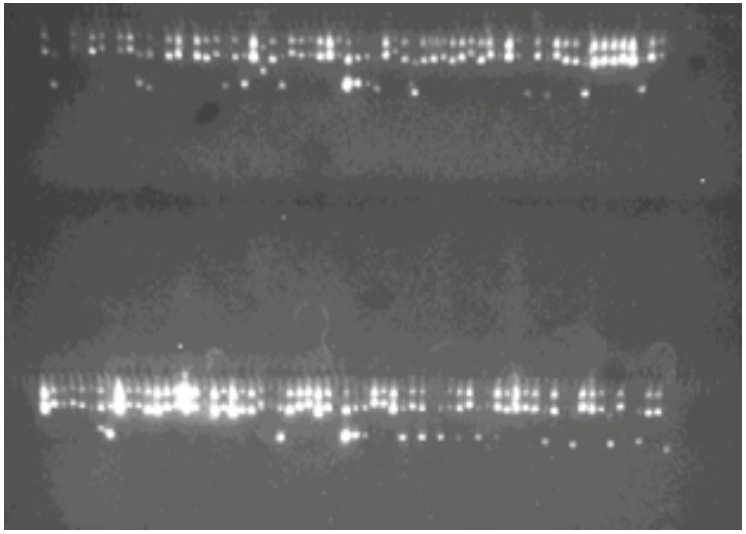
QC - Agarose Gel

In Summary:

- 3 days – 2 days
- Reduction of:
  - Robots
  - Space
  - People
  - Steps

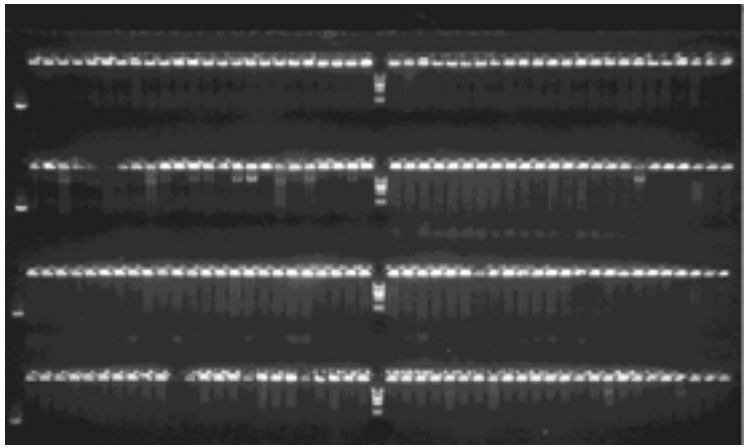
QC Agarose Gels: Amplification vs Purification

SPRI



RCA

200ng



RCA Production Line

- Combine Amplification and Chemistry Groups to 1 lab
- Built a simple automated process with 3 steps:

  1) RCA
  2) Sequencing Chemistry
  3) In House ETOH/Bead Clean-Up

- Control DNA template concentration
- Decrease contamination
- Maintain a 384 well production line

RCA Automation

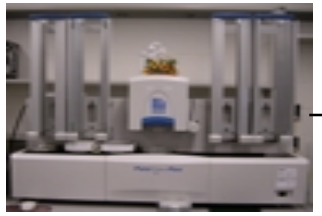
- RCA
- TempliPhi

2. Sequencing Chemistry

- ET Terminators

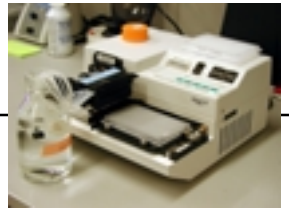
3. BET

- Bead, ETOH, TEG



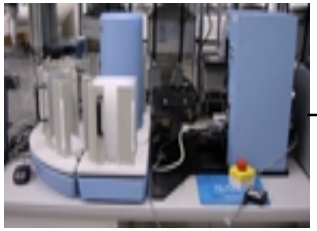
Matrix Platemate

PE Thermocyclers

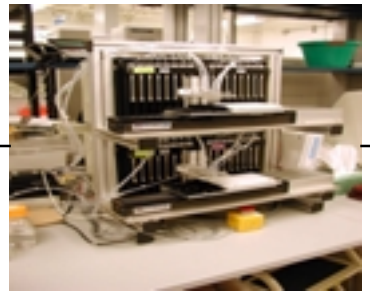


Titertek Multidrop

Incubator



Hydra twister



Cavro Dispense System

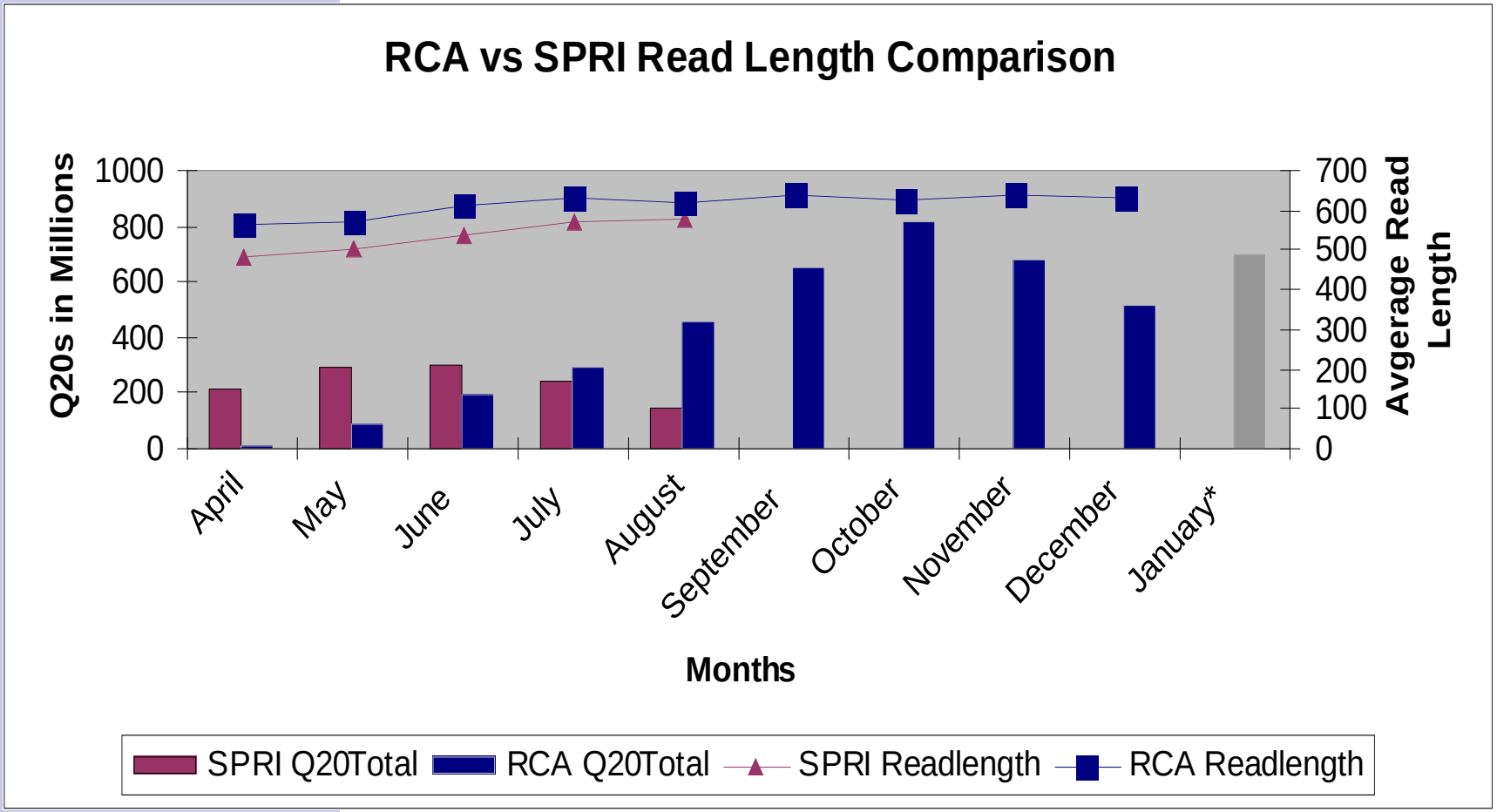
PE Thermocyclers

Biomek FX

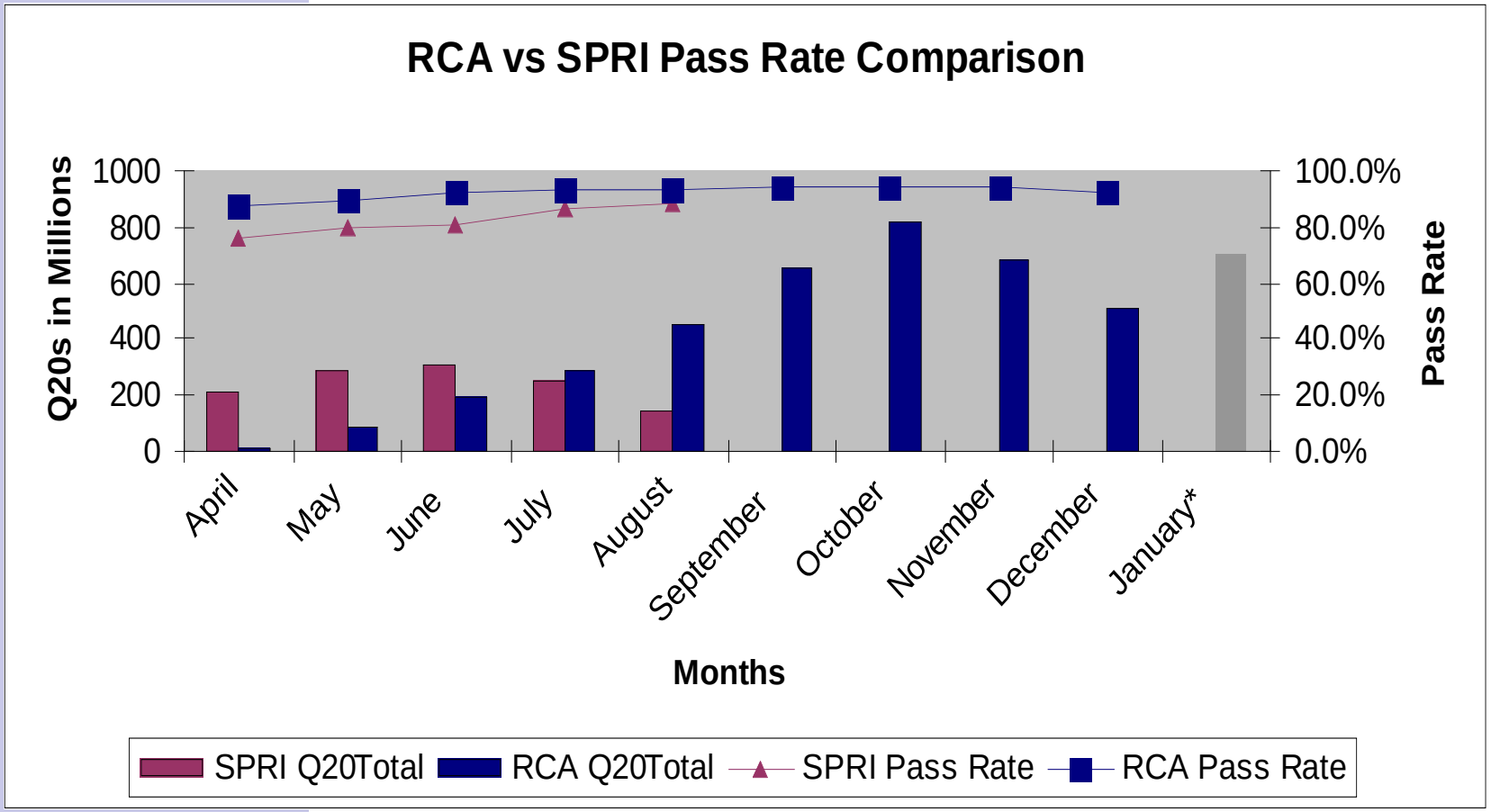
Result of Production Changes

	RCA	SPRI	Change
Days	2-3	5	Increase turnaround
Steps	10	17	7 fewer steps
Space	1,150 sq ft	3500 sq ft	33% space reduction
# Phred 20 bases/day	35M	20M	75% increase
Pass Rate	94%	84%	12% increase
Read Length (q>20)	640	510	130 bases, +25%
FTE's*	9	20	11 FTE reduction

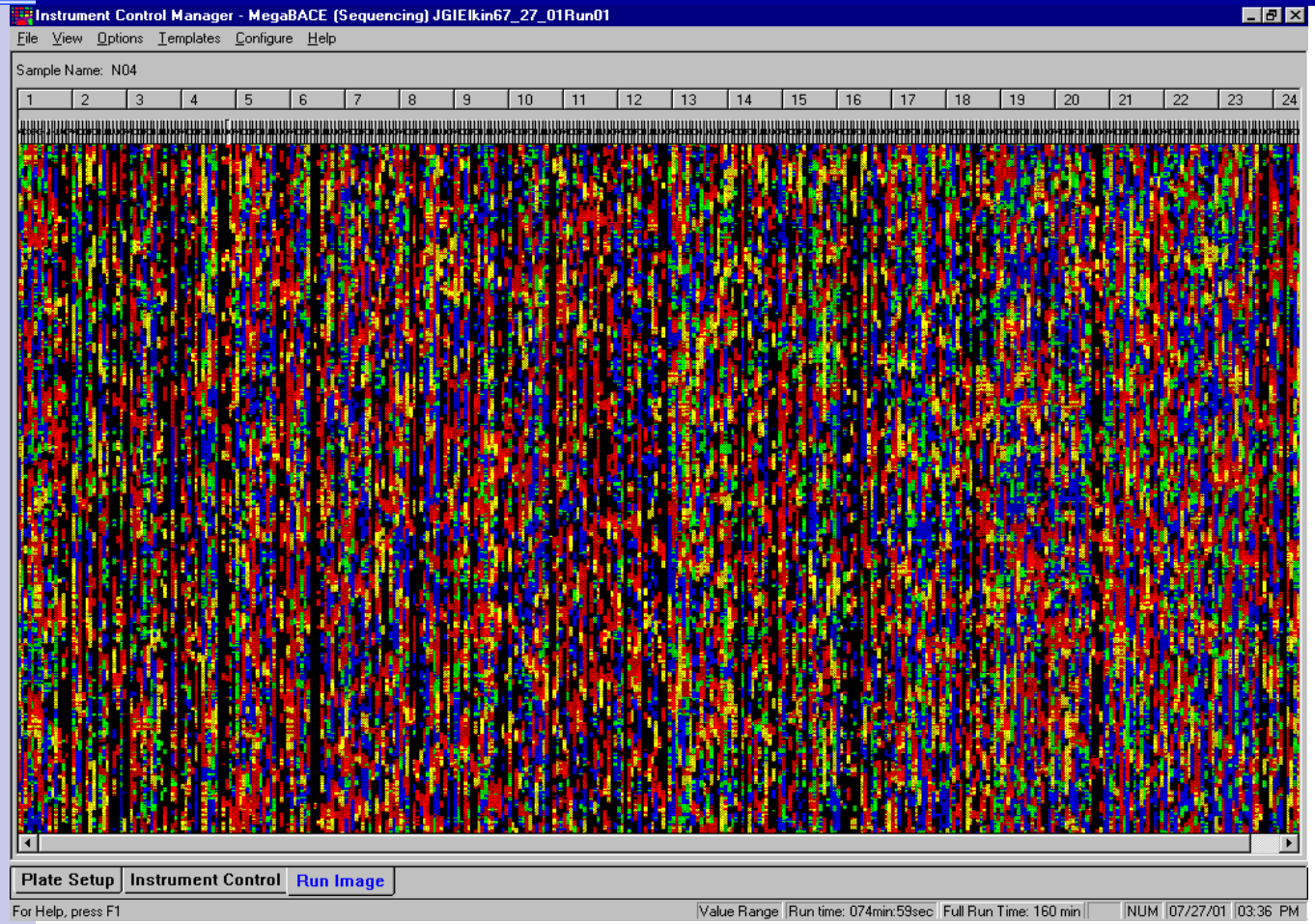
Production RCA Results
Read Length



Production RCA Results
Pass Rate



MegaBACE 4000



MegaBACE 4000

MegaBACE 4000

Throughput Plan:

21 MegaBACE 4000 DNA Sequencers

63 MegaBACE 1000

Sequencers run 5+ days, 24 hours/day

Current month:
700 Mb Q20/month

Goal :
1Gb Q20/month

MB4000 vs MB 1000 Comparison
Read Length & Pass Rate

Averages over January

MegaBACE 1000

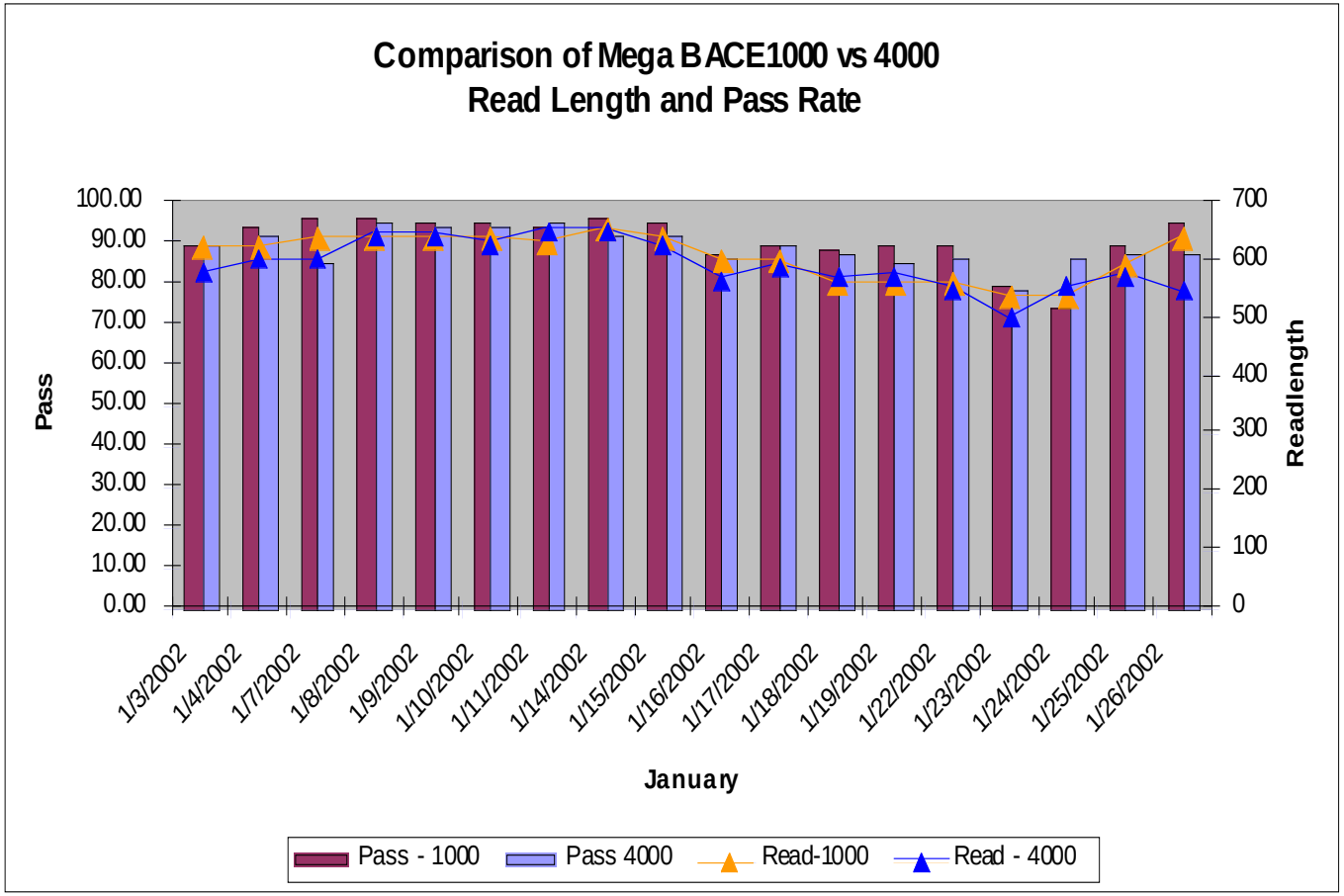
Pass: 90.43%

Read: 603.72

MegaBACE 4000

Pass: 88.54%

Read: 590.17



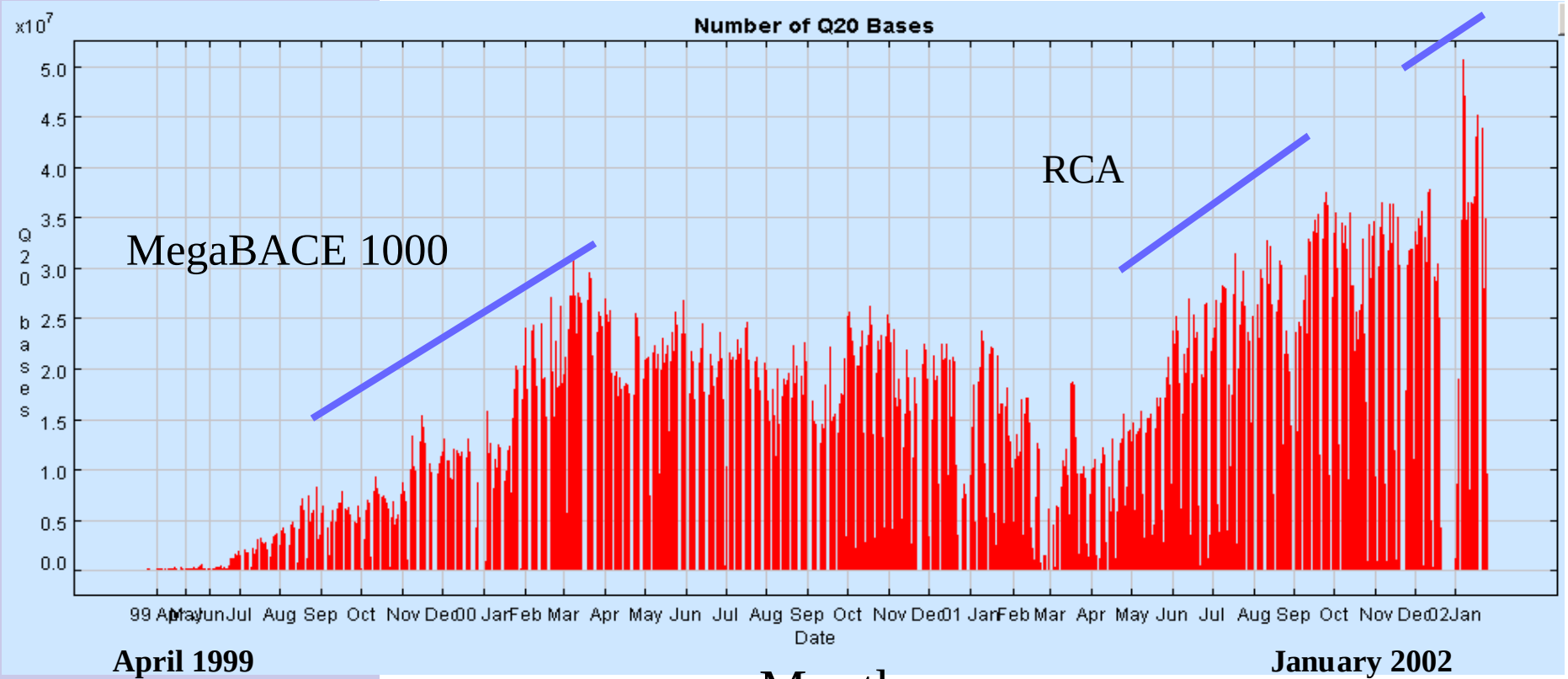
Throughput Goal Summary

- **120 BAC Libraries (120 – 230kb)**
- **250 384 well chemistry plates**
- **1000 96 well or 250 384 well sequencing plates**
- **Average Pass Rate: 90%**
- **Average Readlength: 600**
- **~60 million q20 bases**

Summary of Changes

Daily Q20 bases (in millions)

MegaBACE 4000



Months

Future Of PGF Production

Production:

- Phase out MegaBACE 1000
  - 14 MegaBACE 4000
- 250 384-well plates/day
- Focus on Quality with increased production

Sequencing:

- 1 Gb q20 Sequence/month
- Continue work on Chromosomes 5,16,19
- Microbial Genomes
- Other Large Genomes
  - Xenopus
  - Poplar

Next Microbial Targets

Agrobacterium tumefaciens
Azotobacter vinleandii
Bifidobacterium longum
Clostridium thermocellum
Dechloromonas aromatica
Ehrlichia canis
Ehrlichia chafeensis
Geobacter metallireducens
Haemophilus somnus
Lactobacillus brevis
Lactobacillus casei
Lactobacillus delbrueckii ssp. bulgaricus
Lactobacillus gasseri
Lactococcus lactis ssp. cremoris
Lactococcus lactis ssp. lactis
Leuconostoc mesenteroides
Methanococcoides burtonii
Microbifer degradans
Oenococcus oeni
Pediococcus acidilactici
Pseudomonas syringae pv. syringae
Ralstonia eutropha
Streptococcus suis
Streptococcus thermophilus
Trichodesmium erythraeum
Brevibacterium linens
Pediococcus pentosaceus
Rhodospirillum rubrum
Psychrobacter
Exiguobacterium
Desulfuromonas acetoxidans
Desulfovibrio desulfuricans
Ostreococcus
Thalassiosira pseudonana diatom

Acknowledgements

Tijana Glavina

Nancy Hammon

Lyle Prost

Andrea Aerts

Nathan Bunker

Kristen Kadner

Heather Kimball

Sam Pitluck

Marty Pollard

Sanjay Israni

Production Sequencing Team

Paul Richardson

Chris Detter

Library Development Team

Anne Olsen

Duncan Scott

Joel Martin

Steve Lowry

Jane Grimwood

Jeremy Schmutz

Rick Meyers

Stanford Finishing

Norman Doggett

Mark Mundt

LANL Finishing

Dan Rokhsar

Informatics Team

Jeffrey Boore

Genomic Diveristy Team

Sarah Wenning

Operations Team

Joe Monforte

Lisa Stubbs

Elbert Branscomb

Trevor Hawkins

Marvin Fraiser