

WIKILIMS™

*The Flexible and Cost-Effective Solution for Managing
Your Research Data*



BioTeam at a Glance

- Incorporated in 2002
- Providing high-performance computing, storage and data management service for Life Sciences
- All consultants have both Life Science and IT expertise
- Vendor Agnostics (do not accept vendor commissions)
- Global Client base of > 300
- Partners - Intel, IBM, HP, SGI, Isilon, Illumina, ABI
- Channel Partners - Apple, Univa, Schrodinger

BioTeam HPC, Storage & Data Management Services

- Research Analysis - evaluating your objectives
- Technical Assessment - formulating plans for computing and storage
- Architecture - designing capable, cost-effective solutions
- Implementation - building and integrating
- Testing - validating new and existing systems
- Training - training all users, scientific and technical
- Application development - custom software for research

HPC Informatics Consulting Practice

- Specialty practices & areas of focus:
 - Science-centric IT & Infrastructure Consulting
 - Distributed Resource Management
 - Utility/cloud computing on Amazon EC2

IT & Infrastructure Practice

- *Science-centric* HPC IT consulting & project management including:
 - **Facility**
 - Build-out, technical assessments, relocation/migration projects
 - **System Design**
 - Translate scientific need into IT requirements
 - Turn IT requirements into scalable research IT blueprints
 - **Purchase Assistance**
 - Write RFP documents
 - Evaluate vendor RFP responses & assist with vendor selection
 - Strip inappropriate or unnecessary padded items off of vendor quotes

IT & Infrastructure Practice

- IT & Infrastructure continued ...
 - **Storage Practice**
 - A rapidly growing specialty practice
 - Technical storage audits for life science organizations
 - » Document requirements, estimate growth, identify capability gaps
 - Terabyte to multi-Petabyte storage system design services
 - Integration of terabyte-scale wet lab instruments
 - Confocal microscopy, ultrasound, next-gen sequencing, etc.

Distributed Resource Management

- 10+ years building production clusters & compute farms for Biotech, Pharma, Academic and Government clients
- Deep involvement way beyond “traditional” IT scope:
 - Far more than hardware setup & deployment
 - Installation, deployment & configuration assistance
 - Custom tuning & configuration to match scientific need
 - Scientific application & workflow integration
 - Custom training for end-users, developers & operations staff
- Acknowledged as global experts on Platform LSF and Sun Grid Engine in life science environments
 - Popular community blog <http://gridengine.info> operated by BioTeam
- BioTeam is the only company offering life-science LSF & Grid Engine training
- BioTeam is the only company offering Grid Engine training aimed at end-users

Utility Computing on Amazon EC2

- Since early 2007 **every** active BioTeam consultant has independently used Amazon AWS products to solve real-world customer problems
- After years of BioTeam cynicism regarding “grid” and “utility” computing ...
 - Amazon has finally come through with the proper combination of price, features and capability
 - A fast growing practice area for BioTeam
- Currently working with ISVs and client companies move software and workflows into E22
- BioTeam’s Amazon Cloud milestones:
 - 1st to publicly demonstrate mpiblast operating on EC2
 - 1st to publicly demonstrate self-organizing Grid Engine clusters within EC2
 - Hired by UnivaUD to document Uniclust/EC2 integration
 - Hired by Sun to demonstrate use of EC2 as a “spare pool” for Grid Engine operating under the control of Sun’s Service Domain Management (“SDM”) technology
 - Hired by Pfizer to enable Rosetta ++ Docking Application on the EC2 Cloud.

HPC & Storage Projects to Support Next-Gen Sequencing

- WiCell Research Institute
- MIT Center for Cancer Research
- Helicos BioSciences (WikiLIMS)
- NYU Medical Center (through Sun Microsystems)
- Naval Medical Research Center (WikiLIMS)
- John Hopkins Center for Inherited Disease Research
- MIT Dept Environmental Engineering
- UC Santa Cruz Earth and Planetary Sciences
- Cornell Institute for Biotechnology and Life Sciences Technologies (WikiLIMS)

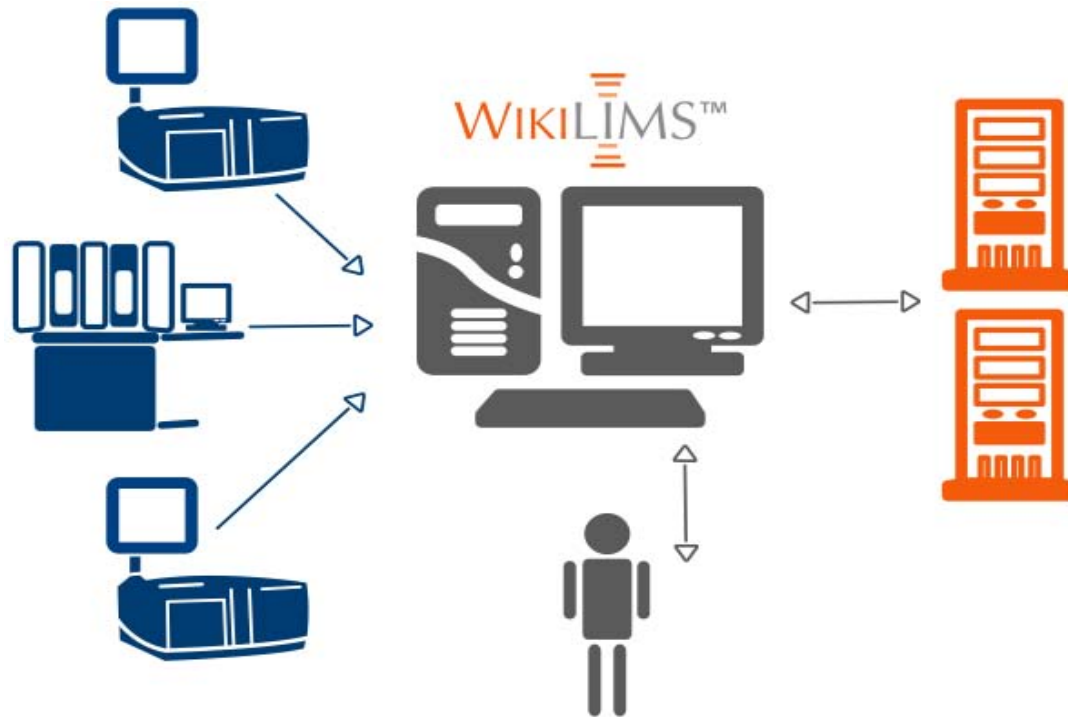
Challenges in Managing Research Data

- High-Throughput Instruments are creating Exponential Data Growth
- New technologies, and changing technologies
- A mix of users: scientific, technical, and informatic
- Multi-platform experimentation (Illumina, 454, Microarrays, Etc.)
- Legacy data locked in out-dated systems and files
- Low volume areas of the lab that are orphaned and have no LIMS
- Data of all types: text, image, video, tabular, relational
- Personnel and conditions change, and closed software isn't maintained

What system can address ALL of these challenges?

The solution is...

- A system that is quickly constructed and easily revised
- A system that can handle many data types
- A modular and transparent system
- A Web-based system



Traditional LIMS vs WikiLIMS

- **Traditional LIMS**

- Complex and unfamiliar interfaces
- Proprietary languages
- Commercial technology platforms
- Complex data schemas
- High cost

- **WikiLIMS**

- Simple user interface
- Use any popular language
- Open-source
- Data *is* the schema
- Low cost

Wikipedia - the world's most used wiki



WIKIPEDIA
The Free Encyclopedia

navigation

- Main page
- Contents
- Featured content
- Current events
- Random article

search

Go Search

interaction

- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Cite this page

languages

Wikipedia is sustained by people like you. Please [donate](#) today.

[article](#) [discussion](#) [view source](#) [history](#)

Wikipedia

From Wikipedia, the free encyclopedia


This article is about the encyclopedia. For the different, similar terms related to Wikipedia, see [Wikipedia \(terminology\)](#). For Wikipedia's non-encyclopedic visitor introduction, see [Wikipedia:About](#).

Wikipedia (pronunciation ^[a]) is a free,^[5] multilingual, open content encyclopedia project operated by the United States-based non-profit Wikimedia Foundation. Its name is a portmanteau of the words *wiki* (a technology for creating collaborative websites) and *encyclopedia*. Launched in 2001 by [Jimmy Wales](#) and [Larry Sanger](#),^[6] it attempts to collect and summarize all human [knowledge](#) in every major language.^[7]

As of April 2008, Wikipedia had over 10 million articles in 253 languages, about a quarter of which are in English.^[2] Wikipedia's articles have been written [collaboratively](#) by [volunteers](#) around the world, and nearly all of its articles can be edited by anyone with access to the Wikipedia website.^[8] Having steadily risen in popularity since its inception,^[1] it is currently the largest and most [popular](#) general reference work on the Internet.^{[9][10][11]}

Critics of Wikipedia target its systemic bias and inconsistencies^[12] and its policy of favoring consensus over credentials in its editorial process.^[13] Wikipedia's reliability and accuracy are also an issue.^[14] Other criticisms are centered on its susceptibility to [vandalism](#) and the addition of spurious or unverified information.^[15] Scholarly work suggests that vandalism is generally short-lived.^{[16][17]}

In addition to being an encyclopedic reference, Wikipedia has received major media attention as an online source of breaking news as it is constantly updated.^{[18][19]} When *Time* magazine recognized "You" as its *Person of the Year* 2006, praising the accelerating success of online collaboration and interaction by millions of users around the world,



Screenshot of Wikipedia's multilingual portal.

URL www.wikipedia.org

Why MediaWiki?

- Stable and supported platform
- Familiar interface
- Very high data capacity using underlying Mysql relational database
- High data connectivity
- Large open source community supporting the code
- Semantic MediaWiki extension for *labeling* of all data
- Highly flexible and programmable
- Versioning and auditing

MediaWiki Physical Requirements

- Sufficient memory, e.g. 4 Gb RAM
- Sufficient space, e.g. 100 Gb disk
- Late model, multi-core processor for maximum performance
- Linux or Mac OS X OS

MediaWiki and its languages

- MediaWiki has its own formal API accessed by HTTP requests
- Perl or Python or Java can all program the Wiki using this API
- PHP *can* be used for deep modifications
- MediaWiki's own template language is used for categorizing and querying
- SQL is not used

MediaWiki Code Base

- Mixture of functional and Object Oriented code
- PHP 5.0
- Complex: 1080 files and ~100K lines of code
- Mysql relational database for all data
- Hundreds of open source extensions: Security, graphics and video, access to other APIs, Semantic MediaWiki, parsing, scheduling and calendars, task management, RSS...

Semantic MediaWiki

- The Semantic Web labels all content to maximize sharing and comprehension
- Semantic MediaWiki is an extension to MediaWiki that allows you apply Semantic Web technology to all your data
- Semantic data is fully inter-related, computable, categorized, and query-able

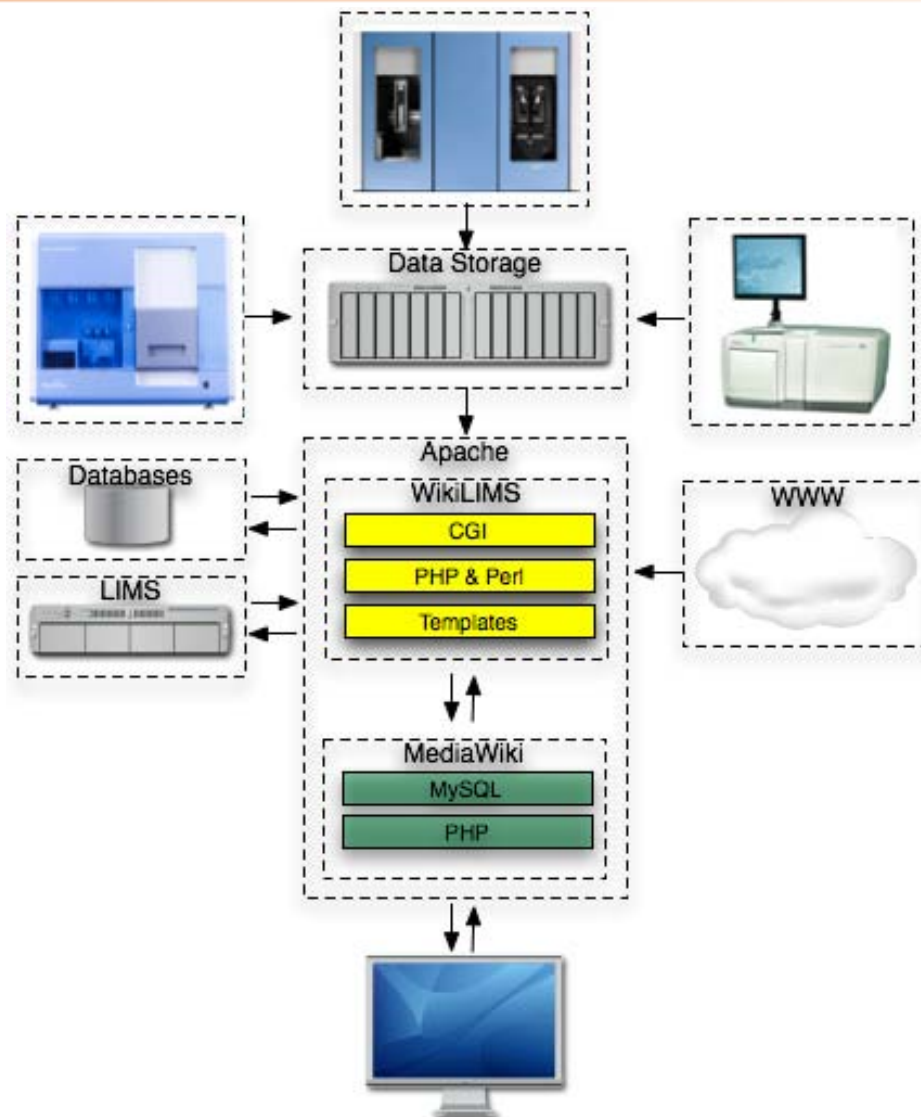
WikiLIMS Developers

- Bill van Etten
- Brian Osborne
- Adam Kraut
- We contribute to BioPerl, DIYA, SNPedia, MediaWiki::Bot, gchart4mw (Google Chart API), Access Control extension, Pywikipedia

WikiLIMS Customers

- Naval Medical Research Center
- National Cancer Institute ABCC
- Cold Spring Harbor Laboratory - Mc Combie Lab
- Cornell Institute for Biotechnology and Life Sciences Technologies
- Emory University - School of Medicine
- University of Connecticut Medical Center
- Helicos BioSciences
- Yeshiva University - Greally Lab Epigenomics
- Pfizer Biotherapeutics and Bioinnovation Center
- Indiana University
- EPA

WikiLIMS Architecture



WikiLIMS Customer Case Studies

- The Navy Biodefense Research Directorate (BDRD)
- John Grealey Lab at the Albert Einstein Medical College
- Brent Graveley Lab at the University of Connecticut
- Core Sequencing Facility at Cornell University
- Dick McCombie Lab at Cold Spring Harbor Laboratory
- Indiana University, Center for Genomics and Bioinformatics
- “multi-national corporation”

Naval Medical Research Center - BDRD















- Situation
 - An expanding collection, greater than 10,000 bacterial strains
 - Need to create rapid sequencing and annotation pipeline
 - Need to launch commands from the Wiki and get results back
 - Need to submit genome sequences to NCBI from the Wiki
 - Need to submit raw data to NCBI Short Read Archive
 - 4 Roche GS instruments in continuous use
 - Affymetrix data
 - Key pages: Strains, Cultures, Projects, Assemblies, Genomes, Runs, Microarrays

BDRD Portal View

page discussion view source history

BDRD Document & Project Management System

(Redirected from Main Page)

 <p>Project portal Desc goes here</p>	 <p>HJF Online Time sheets, admin stuff, etc.</p>	 <p>Server Administration Server administration, etc</p>	 <p>WikiLIMS Lab information management system</p>
 <p>INquiry Portal Bioinformatics portal</p>	 <p>TRAC Version control and development manager</p>	 <p>Do It Yourself Genomics Hosted by sourceforge.net</p>	 <p>BioPerl BioPerl project community</p>
 <p>GBrowse Generic genome browser</p>	 <p>BHS AI Biotechnology HPC Software Applications Institute</p>	 <p>Finch Old BDT Sequencing Database</p>	 <p>NCBI National Center for Biotechnology Information</p>
 <p>FileMaker Database server (admin)</p>	 <p>PhenoDB Phenotype Database</p>	 <p>UMIACS University of Maryland Institute for Advanced Computer Studies</p>	 <p>bioteam Bioteam Tools</p>

A simplified workflow at BDRD

1. Acquire strain, barcode, enter into Wiki (creates **Strain**)
2. Sub-culture **Strain** in the lab, create a **Culture**
3. Organize **Strains** by biological features (creates **Project**)
4. Extract DNA (creates a **Run**)
5. Sequence one or more times (creates **Assembly**)
6. Assemble one or more **Assemblies** from Wiki (creates **Genome**)
7. Annotate one **Genome** or entire **Project** from Wiki
8. Submit **Genomes** to NCBI from Wiki
9. Submit raw data to NCBI Short Read Archive from Wiki

Create a Strain page

Update NS1035 dashboard

Parent: [Species BTH](#)

Siblings: [Strains NS1003, NS1004, and NS1006 ... further results](#)

Children: [Cultures S2452, S6534, and S6546](#)

Grandchildren: [Samples N105, N2052, and N2063 ... further results](#)

Assemblies: [ATCC10792.paired.assembly, P_2008_01_01_00_00_01, and P_2008_01_01_00_00_01_454rig ... further results](#)

Assembly: [P_2008_01_01_00_00_01_454rig](#)

NCBI Project: [29723](#)

locus_tag: [bthur0008](#)

NS1035 updated Tue Jun 24 13:51:50 2008		len	#passed
<input checked="" type="checkbox"/> R_2008_01_01_00_00_01_454FLX_akmala_PMP4xxBTHxxNS1035xxATCC10792	<input checked="" type="checkbox"/> D_2008_01_01_00_00_01_454FLX_akmala_		0

assemble

Create a Culture page

[page](#) | [discussion](#) | [view source](#) | [history](#)

S2452

Culture S2452

date: 2004-11-24

desc: ATCC 10792

location name: BSL3 lab -80 freezer

location code: Box: BA-08, Position: 36

location lab: BDRD-BSL3 lab

Parent: [Strain NS1035](#)

Children: [N105](#), and [N280](#)

Categories: [Culture](#) | [Is a culture](#)

Add a Strain to a Project

Contents [\[hide\]](#)
1 Related topics

Unclassified Project

Description: Incomplete field

Organism: Bacillus

Reference strain: banth0001X

Point of Contact: Incomplete field

[Run Annotation Pipeline \(stable\)](#)

[Run Annotation Pipeline \(dev\)](#)

Related topics

- [Tiling Bacillus genomes](#)
- [DSC2-Clade specific proteins](#)

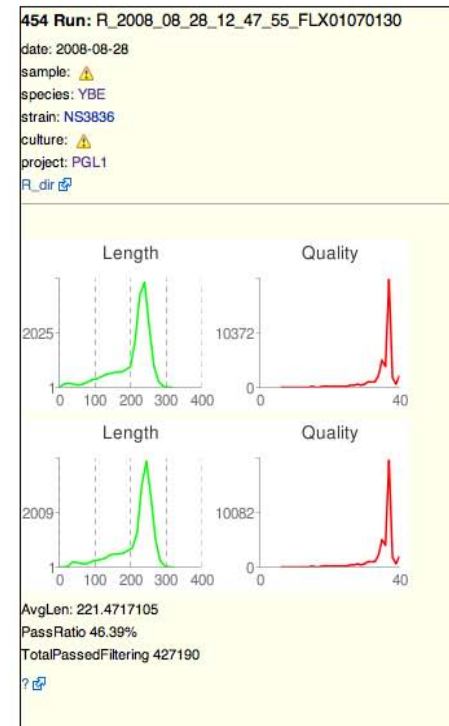
NS1035	ATCC_10792	rev	bthur0008	29723	527031	BTH	Bacillus thuringiensis	1 rdirs	3 pdirs	P_2008_01_01_00_00_01_454rig	bases	reads
NS2805	95/8201	rev	bcere0016	29669	526979	BCE	Bacillus cereus	4 rdirs	5 pdirs	P_2008_02_13_17_56_34_loki	121769577 bases	487766 reads
NS2969	ATCC 10876	rev	bcere0002	29671	526980	BCE	Bacillus cereus	2 rdirs	5 pdirs	P_2008_03_15_14_08_19_loki	255903816 bases	1872088 reads
NS2971	AH_621	rev	bcere0007	29655	526972	BCE	Bacillus cereus	2 rdirs	3 pdirs	P_2008_02_01_13_24_20_loki	113454102 bases	458254 reads
NS2974	ATCC 4342	rev	bcere0010	29665	526977	BCE	Bacillus cereus	4 rdirs	2 pdirs	P_2008_06_25_13_09_24_loki	108850454 bases	406807 reads
NS2996	m1293	rev	bcere0001	29657	526973	BCE	Bacillus cereus	2 rdirs	3 pdirs	P_2008_03_09_15_37_45_loki	260032027 bases	2061922 reads

Sequence and Monitor Quality of the Sequencing Run

R 2008 08 28 12 47 55 FLX01070130

R_2008_08_28_12_47_55_FLX01070130_adminrig_PGL1xxYBExxNS3836xxN2170

*warning: couldn't find a mid for N2170



Assemble Assemblies from the Wiki ...

P_2006_04_14_09_11_15_runAssembly (original 'Old' assembly from 2006 AA)

genus/species: [Yersinia enterocolitica](#)

linklabel: Functional Genomics at NMRC

linkurl: http://www.nmrc.navy.mil/nmrc_db_fg.htm

locale:

name: ATCC_43380

strainname: ATCC 43380

supplier:

[Update NS2457 dashboard](#)

Parent: Species YRO

Siblings: Strains **NS2457**, NS3668, and NS3832 ... further results

Children: Cultures S4182

Grandchildren: Samples N19, and N20

Assemblies: P 2006 04 14 09 11 15 runAssembly, P 2007 10 20 02 23 54 loki, and P 2007 10 23 03 05 33 loki

Assembly: P_2006_04_14_09_11_15_runAssembly

NCBI Project: [29767](#)

locus_tag: yrohd0001

NS2457 updated Thu Sep 11 16:27:21 2008		len	#passed
<input checked="" type="checkbox"/> R_2006_04_13_12_01_37_05240506C_courtneys_yRhodeiRun4x04132006	<input checked="" type="radio"/> D_2006_04_13_12_01_37_05240506C_courtneys_OTFAnalysis	99.227119	250015
<input checked="" type="checkbox"/> R_2006_04_12_15_31_03_05240506C_courtneys_yRhodeiRun3x04122006	<input checked="" type="radio"/> D_2006_04_12_15_31_03_05240506C_courtneys_OTFAnalysis	100.438789	251865
<input checked="" type="checkbox"/> R_2006_04_12_08_25_30_05240506C_courtneys_yRhodeiRun2x04122006	<input checked="" type="radio"/> D_2006_04_12_08_25_30_05240506C_courtneys_OTFAnalysis	99.784096	260758
<input checked="" type="checkbox"/> R_2006_04_11_14_34_57_05240506C_courtneys_yRhodeiRun1x04112006	<input checked="" type="radio"/> D_2006_04_11_14_34_57_05240506C_courtneys_OTFAnalysis	99.150162	228468

[assemble](#)

... and create a Genome page

[page](#)
[discussion](#)
[edit](#)
[history](#)
[delete](#)
[move](#)
[protect](#)
[watch](#)
[refresh](#)

Yrohd0001

***Yersinia rohdei* ATCC_43380**

Sequence	Assembly	Features	Homology	Variation	Links
Length: 4312403 bp	Contigs: 105	Genes: 4039			NS2457
GC Content: 46 %		Average gene length: 907			NCBI Project 29767
Coding content: 84 %		Average intergenic space: 186.86 bp			P directory 2006_04/P_2006_04_14_09_11_15_runAssembly
Topology: linear		Number of overlaps: 552			NCBI Taxon 527004
		Structural RNAs: 57			

Files

[A directory](#)
[genbank](#)
[fsa \(whole genome\)](#)
[fna](#)
[faa](#)
[rfam](#)
[tmscan](#)
[mammer](#)

[Update Yrohd0001](#)
[Launch Annotation](#)
[Submit Yrohd0001 Genome](#)

Launch DIYA annotation software and submit Genomes

page | discussion | view source | history

Yberc0001

Contents [hide]

- 1 *Yersinia bercovieri* ATCC_43970
- 2 Files
- 3 Genbank Submission
- 4 Browser

Yersinia bercovieri ATCC_43970

Sequence	Assembly	Features	Homology	Variation	Links
Length: 4347293 bp	Contigs: 229	Genes: 3947			NS2459
GC Content: 48 %		Average gene length: 903			NCBI Project 16104
Coding content: 82 %		Average intergenic space: 192 bp			P_2006_10_06_10_01_01_runAssembly
Topology: linear		Number of overlaps: 50			NCBI Taxon 634
		Structural RNAs: 67			Project Name: PGL1

Files

[A directory](#)
[genbank](#)
[fsa \(whole genome\)](#)
[fna](#)
[faa](#)
[rfam](#)
[trnscan](#)
[rammer](#)

Genbank Submission

[All annotation files](#)
[Genbank](#)
[ASN.1](#)
[fasta](#)
[AGP](#)
[quality](#)
[tbl](#)
[val](#)

Launch MID Assembly from Wiki

NS5646 (Wed Nov 4 09:51:07 2009)	Select SFF files
R_2009_10_27_17_41_11_FLX01070130_adminrig_NS5643XXNS5646XXNS5649XXN2337XXN2341XXN2345	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR02.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR02.MID5.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR02.MID4.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR02.MID11.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR02.MID1.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID9.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID8.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID7.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID6.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID5.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID4.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID3.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID2.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID12.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID11.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID10.sff
	<input type="checkbox"/> D_2009_10_28_10_24_57_node003_signalProcessing/sff F4RKQXR01.MID1.sff
R_2009_10_02_14_46_34_FLX01070131_adminrig_NS5646xxNS5647xxNS5648xxN2341xxN2340xxN2336	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ02.sff
	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ02.MID6.sff
	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ02.MID5.sff
	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ02.MID4.sff
	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ01.sff
	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ01.MID6.sff
	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ01.MID5.sff
	<input type="checkbox"/> D_2009_10_04_12_27_28_node004_signalProcessing/sff F3G1ZWZ01.MID4.sff

☒ Use -large

Assemble

Genomes viewed in Wiki using GBrowse

http://loki.bdrd:16080/wiki/index.php/Styum0001X

Biological Defense Research Directorate
Organism: *Bacillus*
Database: PMP4
Host: portal.local

Showing 50 kbp from styum0001X, positions 880,556 to 930,555

[Instructions](#) [\[Hide banner\]](#) [\[Bookmark this\]](#) [\[Link to Image\]](#) [\[High-res Image\]](#) [\[Help\]](#) [\[Logout\]](#)

[Search](#)
Choose a sequence...
[Overview](#)
[Region](#)

Region of styum0001X
0M 0.1M 0.2M 0.3M 0.4M 0.5M 0.6M 0.7M 0.8M 0.9M 1M 1.1M 1.2M 1.3M 1.4M 1.5M 1.6M 1.7M 1.8M 1.9M 2M

[Details](#)

Named gene
ORF
CDS

[Clear highlighting](#) [Update Image](#)

[Tracks](#)
[Display Settings](#)
[Add your own tracks](#)

For the source code for this browser, see the [Generic Model Organism Database Project](#). For other questions, send mail to lstein@cshl.org.

\$Id: yeast_chrt1.conf,v 1.9.4.3 2005/07/11 19:43:31 lstein Exp \$

Note: This page uses cookies to save and restore preference information. No information is shared.
Generic genome browser version 1.68

Genomes viewed in Wiki using GBrowse

• http://loli.bdrd/G_labdata/454_data/organized/annotations/2008_04/A_PGL1_2008_04_18_10_27_34/Styum0001X.faa.gp

my

board

AS

Portal

AS

forum

Administration

Search

links here

changes

file

pages

version

next link

Browser

[\[edit\]](#)

Biological Defense Research Directorate
Organism: *Bacillus*
Database: **PMP4**
Host: **portal.local**

Name: styum0001X_9200
Class: locus_tag
Type: gene
Source: Genbank
Position: styum0001X:892679..894445 (+ strand)
Length: 1767

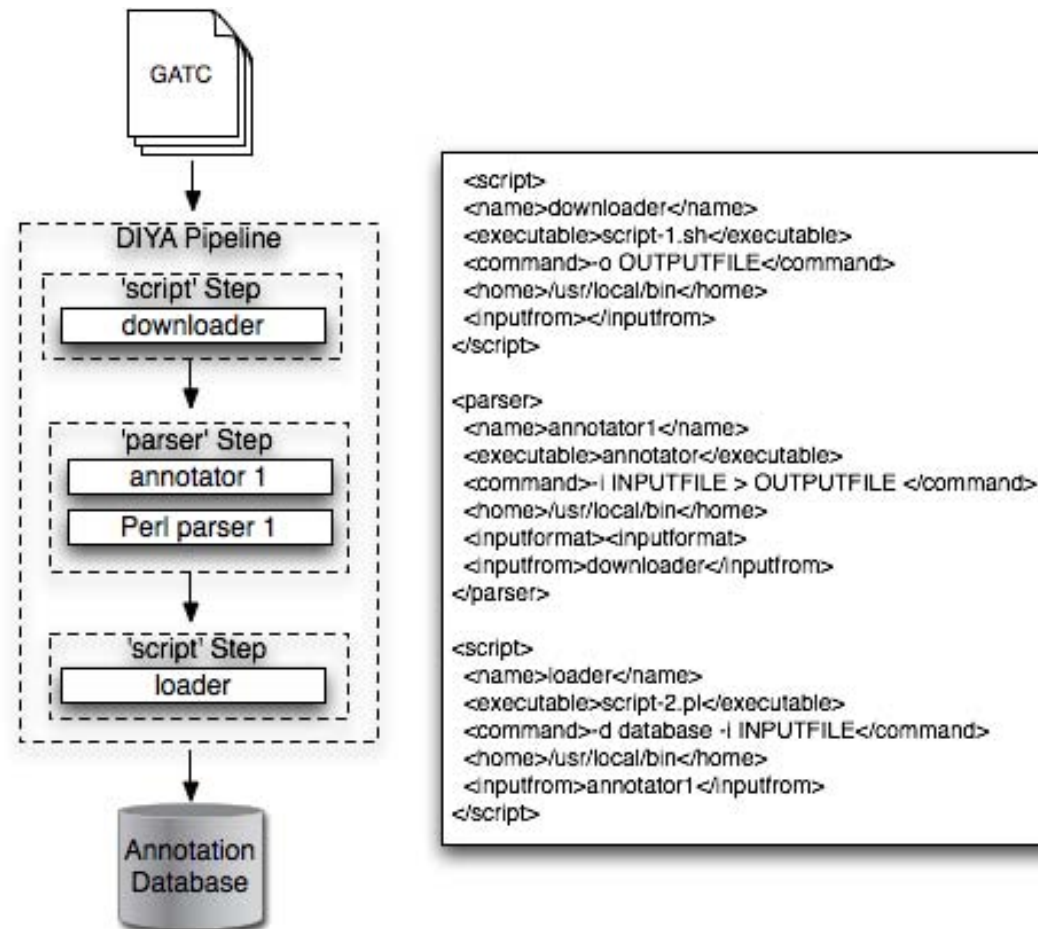
```
>styum0001X_9200 class=locus_tag position=styum0001X:892679..894445 (+ strand)
ATGAAGCTGC CAATCAGAGA ATTGATGCT GTTGTGATTG GTGCGGCTGG CGCAGGTATG CGCGCGGCGC TGCAAAATTC
CCAGAGCGCG CAACACCTGTG CCTGCTCTTC GAAAGTCTTC CGGACCCGTT CCCATAGCGT TTCGCGGCGC GGCAGTATCA
CCGTGCGCGT CGGCAATACC CATGAAGATA ACTGGGATG GCACATGTAC GACACCGTTA AGGGGTGGGA CTACATTGGT
GACCAAGATG CCATCGAATA TATGTGTAG ACCGGACCGG AAGCGATTCT GGAACTGGAG CACATGGGCG TACCCTTCTC
CCGTCTTGAT GATGGACGTA TCTATCAGCG CCCGTCTGCG GCGCAGTCGA AAACTTTGCG CGCGGAGCGC GCGGCACGTA
CCGCGCGCGC GCGTGACCGT ACCGGTCACG CCTTGCTACA CACGCTGTAT CAACAGAATT TGA AAAACCA CACCAAGATT
TTCTCCGAAT GGTATCGGCT GGATCTGGTG AAAAACCAAG ATGCGCGCGT GGTGGGTGT ACCGCACTGT GCATCGAAGC
CGGTGAAGTG GTGTACTTCA AAGCCCGCGC GACGCTGCTG CGGACCGCGC CGCAGGCGCG TATCTACGAG TCCACCACCA
ACGCTCACAT CAACACCGGT GACGCTGTG GTATGCGCT GCGTGCGGT GTCCCGGTGC AGGACATGGA AATGTGGCAG
TTCCACCCGA CGCGCATCGC CGCGCGCGCG GTGCTGGTGA CAGAAAGCTG CCGCGCGGAA GCGGTTTACC TGCTGAACAA
ACACCGCGAG CGCTTTATGG AGCGTTATGC CGCGAACGCC AAGACCTGG CGGCTGTGA CGTGCTGGCG COTTCATCA
TGATCGAAT CGGTGAAGGC CGCGCTGTG ATGGTCCGTG GGGCGCGCAC GCTAAACTGA AACTGGATCA CCTGGGTAAA
GAGGTGCTGG AATCTCGCT GCGGGTATG CTGGAAGCTGT CCCGTACTTT CCGCCACGTT GACCGGTGA AAGAGCGAT
TCCGTTATC CGGACCTGCG ACTACATGAT GGGCGGTATT CCGACTAAAG TGACCGGTCA GCGCTGAGC GTGAACGAGC
AGGCGGAAGA CGTGCTCATT CCGGGCTGT TTGCGGTAGG CGAAATCGCC TGGTATCGG TTCACGGGCG CAACCGTCTG
GCGGTAACCT CACTGCTGGA TCTGTTGTC TTGCGCGCG CCGCGGCGCT GCATTTCAG GAATCCATCG CCGAGCAGCG
CGTCTGCGC GACCGCAGCG AGTCTGACGT TGAAGGTTCT CTGGAGCGCC TGAATCGCT GAACATAAC CGCAATGGCG
AAGATCCGCT GGTATCGCG AAGCGTTGC AGGAGTGTAT CGAGCATAAC TTCTCCGTAT TCCGTGAAGG CGACCGGATG
CGGAAGGCG TTGAACAGTT GAAGTGATC CGCGAGGCTG TGA AAAACCG CCGTCTGAT GACACGTCCA CGGAGTTTAA
TACCGAGCGT GTTGAGTGTG TGGAGCTGGA TAACCTGATG GAACCGGCT ACGCCACTGC CGTATCAGCC AACTTCCGAA
CGGAAGGCG CGCGCGCAT AGCGCTTGC ATTTCCCGGA CGGTGATGAT GCGCACTGCT TGTGTCATAC CCTGTATCAG
CGCMAACCG AATCCATGAC CGCGCGAAGC GTCATATGCG AGCGMAACT CGGTCCGCGC TTCGCGCGGA AGATTGCTAC
TTATTAA
```

For the source code for this browser, see the [Generic Model Organism Database Project](#). For other questions, send mail to lstein@cshl.org.

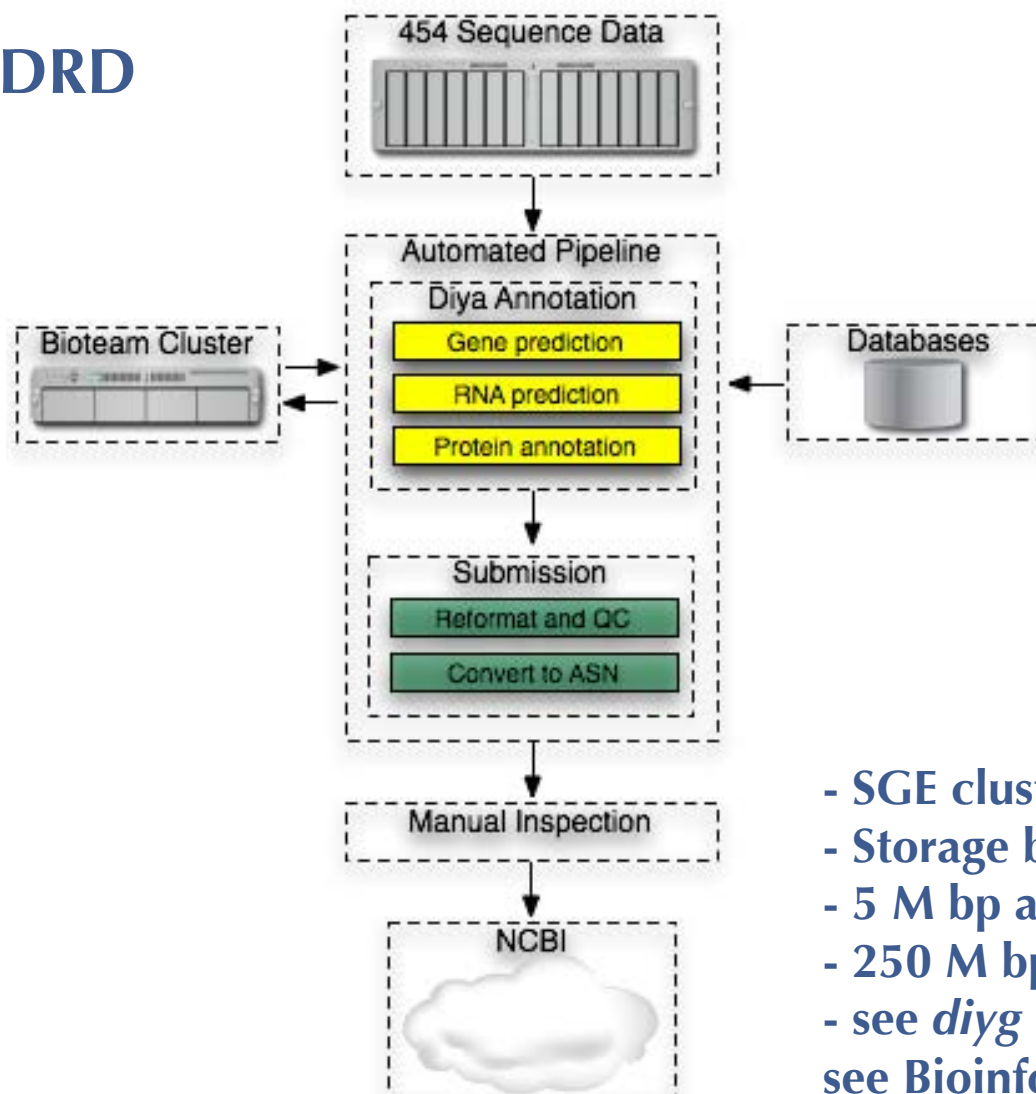
\$Id: yeast_chrl.conf,v 1.9.4.3 2005/07/11 19:43:31 lstein Exp \$

Note: This page uses cookies to save and restore preference information. No information is shared.
Generic genome browser version 1.44

DIYA - open source pipeline software (BioTeam & BDRD)



DIYA at BDRD



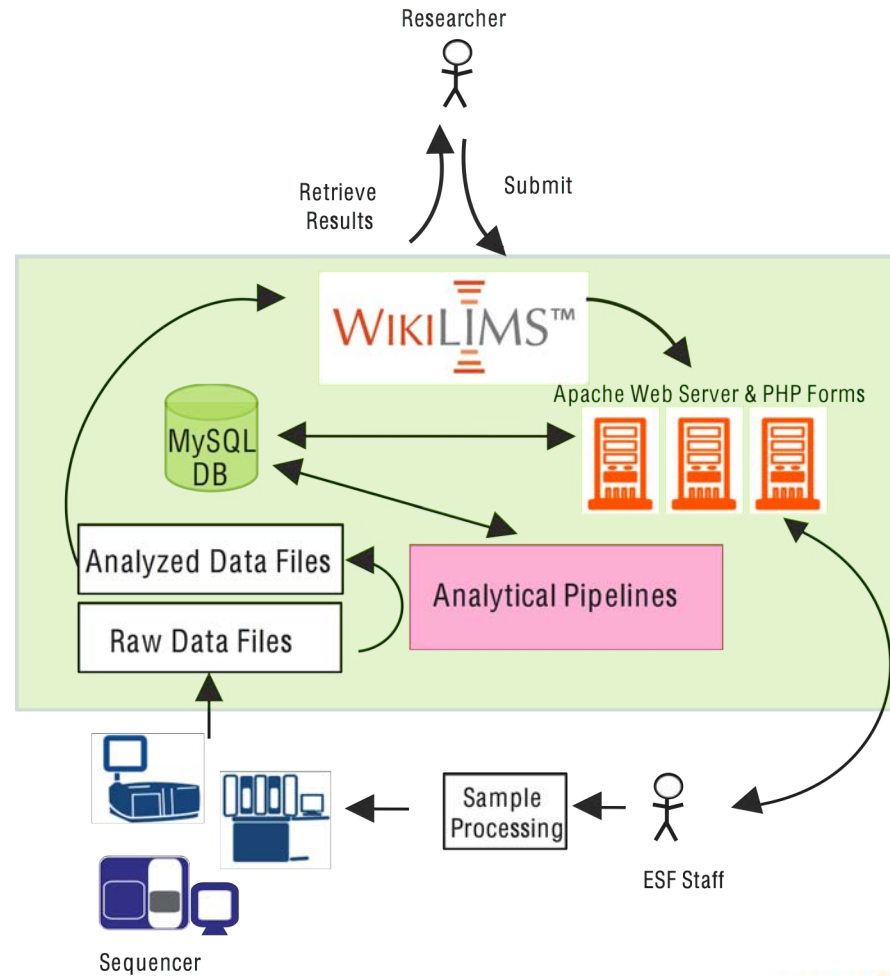
- SGE cluster by BioTeam
- Storage by BioTeam
- 5 M bp annotated, 45 minutes
- 250 M bp submitted in 6 weeks
- see *diyg* at Sourceforge
- see Bioinformatics March 2009

Albert Einstein Medical College Center for Epigenomics

- Situation
 - Core Facilities for Genomics and Epigenomics
 - 1 Roche GS and 1 Illumina GA, NimbleGen microarrays
 - Need to handle sample submissions
 - Need to allow external labs to retrieve their results
 - Need to reserve and schedule technicians and instruments
 - Key pages: Client Request, Samples, Jobs, Notebooks, Analysis, Billing

Albert Einstein Medical College

- *ad hoc* client analysis
- Front-end components
- Customer request UI
- Results and reporting
 - Data Tables
 - Visual Analytics
 - File Management



Managing client requests for sample submission

UCSC Genome Version: [UCSC](#) *

Priority?

Required result date November 2009

Attach files

Sample QC Analysis:

Gel image: ☐ [Upload file](#)

QPCR: ☐ [Upload file](#)

Antibody information:

Antibody name: *

Antibody manufacturer name:

Antibody catalogue number:

Antibody lot number:

Antibody amount used:

Multiple Samples

Samples:

See also [Sample requirements](#)

#	Name	Type	Size (bp) 200-500	Amount (µg) > 10 ng	Conc. (ng/ul) 1-100	A260/280 ≥ 1.8	A260/230 ≥ 1.7	Volume	Buffer
1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
6	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
7	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
8	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
9	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
10	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Sequencing Job Results

- ChIP-Seq, ChIP-chip
- Custom assays
- Analytical jobs
- Custom web reporting
 - Using Mediawiki API
- Launch Custom Apps
 - Gbrowse
 - Jalview

Job description

- **Job Name**
 - CHP-SEQ with anti-GATA1
- **Assay Type**
 - ChIP-Seq
- **Submitted By**
 - Masako Suzuki (Greally Lab)
- **Submitted Date**
 - 07/13/09
- **Completed Date**
 - 09/18/09
- [Click to Show Charts of Job Quality](#)

Sequencing and Alignment Results

Flowcell ID	Sequencing Summary	Sample Name	Lane	Raw Data File	Alignment Result
42DCEAAXX	Click to show	ES_no_cytokines_INPUT	lane_1	Click to download	Show in Genome Browser
		ES_plus_cytokines_INPUT	lane_2	Click to download	Show in Genome Browser
FC42AHHAAXX	Click to show	ES_plus_cytokines_anti-gata1	lane_3	Click to download	Show in Genome Browser
		ES_no_cytokines_anti-gata1	lane_4	Click to download	Show in Genome Browser
		ES_plus_cytokines_anti-gata1	lane_5	Click to download	Show in Genome Browser

Peak Finding Results

Sample Name	Sample Type	Flowcell ID	Lane	Result (1 sample)
ES_no_cytokines_INPUT	Input	42DCEAAXX	lane_1	Click to download <div> Show raw.bed in Genome Browser Show peaks.bed in Genome Browser </div>

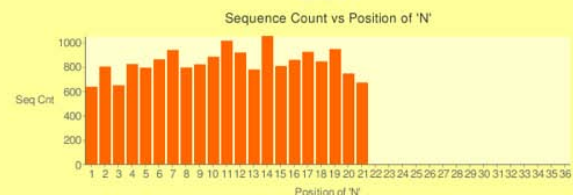
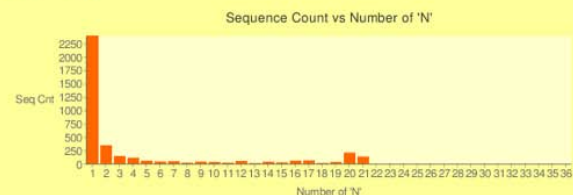
Quality Control Reports

- Google Charts API
- Lane by lane metrics

Sequence Quality Statistics

ES no cytokines INPUT (flowcell: 42DCEAAXX lane 1)

Sequences With and Without 'N'



Sequence Statistics Post Alignment




Run Quality Parameters

These quality metrics are based principally on single read 36bp sequencing of human DNA. The optimal results will vary according to experiment type.

Metric	Result	Uniformity (across tiles)	Notes
Total Yield			Result is 139730 (target is >1Gbases).
Raw Cluster Count			Result is 108688 +/- 9424 (target is >20,000).
% Clusters Passing Filter (PF)			Result is 36.46 +/- 6.93 (target is > 70%). Low % may be indicative of high cluster number (clusters too close together to obtain a clean signal in early cycles)
% Clusters PF that Align Uniquely to Reference			Result is 1.38 +/- 0.12. Optimal value dependent on read-length, genome sequenced and completeness of reference. For 30mers and the human genome, < 80% may be normal.
% Error Rate of Clusters PF			Result is 5.88 +/- 0.51. Should be ~1.5% but in any case, as low as possible.
% Phasing			Result is 0.6900. Should be ~0.5% to no more than 1% but in any case, as low as possible.
% Prephasing			Result is 0.5400. Should both be ~0.5% to no more than 1% but in any case, as low as possible.
First Cycle Intensity			Result is 15 +/- 4. Should be >1000.
20th Cycle Intensity as % of First			Result is 72.12 +/- 14.20. Should be >50%. If too high, suspect relatively low first cycle intensity

WikiLIMS Electronic Lab Notebook (ELN)



navigation

- Main Page
- Community portal
- Current events
- Help

quick links

- Services

search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Print as PDF
- Browse properties

page discussion edit history delete move protect watch refresh

UserPageTest/New Protocols

Genomic DNA extraction protocol

Buffer and reagent:

- Genomic DNA extraction buffer (250ml):
 - 1M Tris.Cl (pH 8.0) 2.5ml
 - 0.5M EDTA (pH 8.0) 50 ml
 - Pancreatic RNase 5 mg
 - 10% SDS 12.5 ml
- Adjust pH to 8.0 and adjust volume to 250ml with ddH₂O
- Saturated phenol (pH 8.0)
- 10M ammonium acetate (NH₄Ac)

Protocol:

1. Weigh 0.5-1g fresh tissue and put in mortar. Add liquid nitrogen to snap freeze tissue and blend tissue to powder.
2. Add 10 ml genomic DNA extraction buffer in 50 ml tube and put tissue powder in.
3. Invert tube to submerge tissue powder and incubate at 37c for 1 hour.
4. Add 50 ul proteinase K (20mg/ml stock), mix gently.
5. Incubate in 50c water bath for 3 hours, shake gently.
6. Let stand in room temperature for 30 min to equilibrate to room temperature.
7. Add 10 ml Phenol, mix gently for 10 min.
8. Centrifuge at 3000 rpm x 15 min.
9. Transfer the viscous aqueous phase to a new tube using a wide-pore glass pipette.
10. Repeat phenol extraction for 2 times or more.
11. Add 2 ml ammonium acetate (10M), mix gently.
12. Add 2 volume of ethanol (in room temperature). Swirl gently and you will see genomic DNA start to form the white mass. Transfer genomic DNA to a new tube by using a "U" shape pipette.
13. Air dry for 5-10min to drive off ethanol and dissolve in ddH₂O or TE buffer.

WIKILIMS™

navigation

- Main Page
- Community portal
- Current events
- Help

quick links

- Services


search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Print as PDF
- Browse properties

UserPageTest/Presentation Images



Description

Gel electrophoresis: 6 "DNA-tracks". In the first row (left), DNA with known fragment sizes was used as a reference. Different bands indicate different fragment sizes (the smaller, the faster it travels, the lower it is in the image); different intensities indicate different concentrations (the brighter, the more DNA). DNA was made visible using ethidium bromide and ultraviolet light.

University of Connecticut

- Situation
 - 1 Roche GS and 1 Illumina GA 2
 - Multiple labs and multiple research projects (and modENCODE)
 - Need to allow data submission and data retrieval from external laboratories
 - Need to track reagent use and work by each user
 - Key pages: Flowcells, Laboratories, Projects, Samples, Reagents, Users, Species

[Main Page](#)

Illumina	Titanium	Total
----------	----------	-------

	Run date	Entry date	Flowcell	Total Kb
090519 HWI-EAS299 0012 4277CAAXX	19 May 2009	27 May 2009 14:16:02	4277CAAXX	3,062,280
090512 HWI-EAS299 0011 4277EAAXX	12 May 2009	19 May 2009 13:02:48	4277EAAXX	5,937,648
090508 HWI-EAS299 0010 427C7AAXX	8 May 2009	14 May 2009 00:41:12	427C7AAXX	4,788,192
090504 HWI-EAS299 0009 427EDAAXX	4 May 2009	13 May 2009 08:56:31	427EDAAXX	4,485,883
090428 HWI-EAS299 0008 4275CAAXX	28 April 2009	13 May 2009 08:57:17	4275CAAXX	4,202,114
090420 HWI-EAS299 0007 313YUAAXX	20 April 2009	13 May 2009 08:57:24	313YUAAXX	3,710,559
090414 HWI-EAS299 0006 313ATAAXX	14 April 2009	13 May 2009 08:57:38	313ATAAXX	4,262,809
090407 HWI-EAS299 0005 3138TAAXX	7 April 2009	13 May 2009 08:57:49	3138TAAXX	2,710,240
090331 HWI-EAS299 0004 313A VAAXX	31 March 2009	13 May 2009 08:56:18	313A VAAXX	1,804,555
090320 HWI-EAS299 0003 315E1AAXX	20 March 2009	13 May 2009 08:56:45	315E1AAXX	2,607,285
090311 HWI-EAS299 0001 30WEDAAXX	11 March 2009	13 May 2009 08:57:07	30WEDAAXX	2,588,722
090303 HWI-EAS299 0002 30VG EAAXX	3 March 2009	9 March 2009 22:24:40	30VG EAAXX	1,558,480

... further results

Total Illumina Kilobases Sequenced: 89630207

Filesystem	512-blocks	Used	Available	Capacity	Mounted on
capipeline01:/data/pipeline	14420876752	9883431024	3804907264	73%	/data/pipeline

navigation

- [Main Page](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

forms

- Add Sample
- Add Flowcell
- Add Species
- Add Laboratory
- Add Project
- Add Machine
- Add Reagent

categories

- Flowcells
- Users
- Samples
- Laboratories
- Projects
- Reagents
- Species

search

Go Search

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Print as PDF](#)
- [Browse properties](#)

- Using this WikiLIMS
- Wiki Editing Basics
- Local Configuration
- WikiLIMS Design
- To Do
- Wikilims Tutorial

Total Number of Illumina Runs



Data current as of June 1, 2009, 09:31. [Refresh page](#)

A Sample has Project and Laboratory data

page

discussion

edit with form

edit

history

delete

move

protect

watch

refresh

S2-DRSC Brr2 RNAi rRNA minus-1

Flowcells with Sample **S2-DRSC Brr2 RNAi rRNA minus-1**: 30B5NAAXX

S2-DRSC Brr2 RNAi rRNA minus-1	
Sample Description	S2-DRSC Brr2 RNAi rRNA minus-1
Sample Type	mRNA-Seq
Library Type	Paired-End
Species	Drosophila melanogaster
User	User:Liyang
Project	ModENCODE
Laboratory	Graveley
Date submitted	2009/02/01
Sample Reagent	

Category: Sample

A Flowcell page, with User view and Flowcell details


30B5NAAXX				
Lane	Sample	User		
1	PhiX	Core		
2	CT-2 MEF-1	Misha		
3	CT-2 CM-2	Misha		
4	Sexual Nonirradiated mRNA PE	Dasaradhi		
5	Asexual Nonirradiated mRNA PE	Dasaradhi		
6	Ago(RNAi) mRNA PE	Dasaradhi		
7	S2-DRSC Brr2 RNAi rRNA minus-1	Liyang		
8	S2-DRSC PS RNAi rRNA minus-1	Liyang		

30B5NAAXX	
Amplification date	2009/02/09
Cluster station 1	N/A
Cluster station 2	N/A
Cluster box 1	N/A
Cluster box 2	N/A
BetaIne	N/A
Lane 1 primer type	Genomic primer 1/2
Lane 1 primer lot	N/A
Lane 1 sample	PhiX
Lane 2 primer type	Genomic primer 1/2
Lane 2 primer lot	N/A
Lane 2 sample	CT-2 MEF-1
Lane 3 primer type	Genomic primer 1/2
Lane 3 primer lot	N/A
Lane 3 sample	CT-2 CM-2
Lane 4 primer type	Genomic primer 1/2
Lane 4 primer lot	N/A
Lane 4 sample	Sexual Nonirradiated mRNA PE
Lane 5 primer type	Genomic primer 1/2
Lane 5 primer lot	N/A








Track work by User

user page	discussion	edit with form	edit	history	delete	move	protect	watch	refresh
---------------------------	----------------------------	--------------------------------	----------------------	-------------------------	------------------------	----------------------	-------------------------	-----------------------	-------------------------

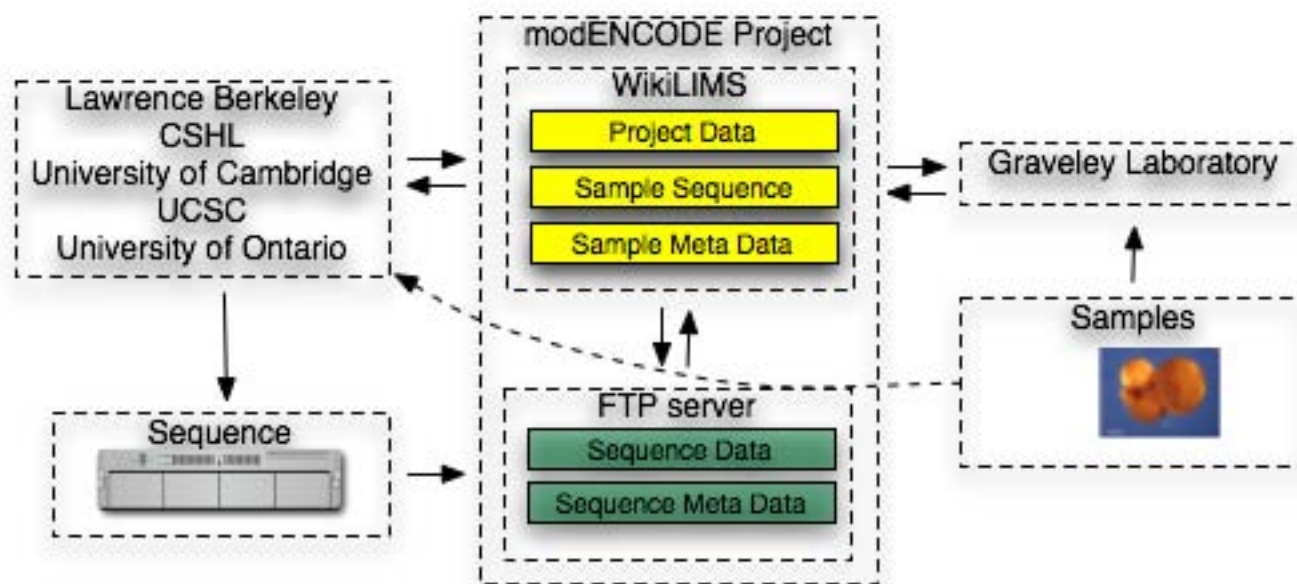
User:Misha

Email Address 

Laboratory Graveley

Samples					Flowcells	
	 Date submitted	 Species	 Sample Has Laboratory	 Sample Has Project		 Amp date
CT-2 CM-1	1 December 2008	Homo sapiens	Graveley	Stem Cell	30CYGAAXX	27 October 2008
CT-2 CM-2	1 February 2009	Homo sapiens	Graveley	Stem Cell	30GGJAAXX	11 November 2008
CT-2 MEF-1	1 February 2009	Homo sapiens	Graveley	Stem Cell	30M66AAXX	9 December 2008
CT-2 MEF-2	1 February 2009	Homo sapiens	Graveley	Stem Cell	30VGEAAXX	12 March 2009
CT-2 TeSR-1	1 February 2009	Homo sapiens	Graveley	Stem Cell	31003AAXX	12 March 2009
CT-2 TeSR-2	1 February 2009	Homo sapiens	Graveley	Stem Cell		
Grabel mES	1 February 2009	Mus musculus	Graveley	Stem Cell		
Grabel mES Sox-1-GFP	1 February 2009	Mus musculus	Graveley	Stem Cell		
H9 CM-1	1 January 2008	Homo sapiens	Graveley	Stem Cell		
H9 CM-2	1 January 2008	Homo sapiens	Graveley	Stem Cell		
H9 ENPd10	1 February 2009	Homo sapiens	Graveley	Stem Cell		
H9 JL-1	1 February 2009	Homo sapiens	Graveley	Stem Cell		
H9 LNPd17RA	1 February 2009	Homo sapiens	Graveley	Stem Cell		
H9 LNPd17c	1 February 2009	Homo sapiens	Graveley	Stem Cell		
H9 MEF miRNA - 1	1 December 2008	Homo sapiens	Graveley	Stem Cell		
H9 MEF miRNA - 2	1 December 2008	Homo sapiens	Graveley	Stem Cell		
H9 MEF-1	1 January 2008	Homo sapiens	Graveley	Stem Cell		
H9 MEF-2	1 January 2008	Homo sapiens	Graveley	Stem Cell		
H9 MEF-CM miRNA - 1	1 December 2008	Homo sapiens	Graveley	Stem Cell		
H9 MNP	1 February 2009	Homo sapiens	Graveley	Stem Cell		
H9 TeSR miRNA - 1	1 December 2008	Homo sapiens	Graveley	Stem Cell		
H9 TeSR miRNA - 2	1 December 2008	Homo sapiens	Graveley	Stem Cell		
H9 TeSR-1	1 January 2008	Homo sapiens	Graveley	Stem Cell		
H9 TeSR-2	1 January 2008	Homo sapiens	Graveley	Stem Cell		

modENCODE Project uses WikiLIMS as project hub



A Project involves internal and external Laboratories

page	discussion	edit with form	edit	history	delete	move	protect	watch	refresh
----------------------	----------------------------	--------------------------------	----------------------	-------------------------	------------------------	----------------------	-------------------------	-----------------------	-------------------------

ModENCODE

Samples for the **ModENCODE** project:

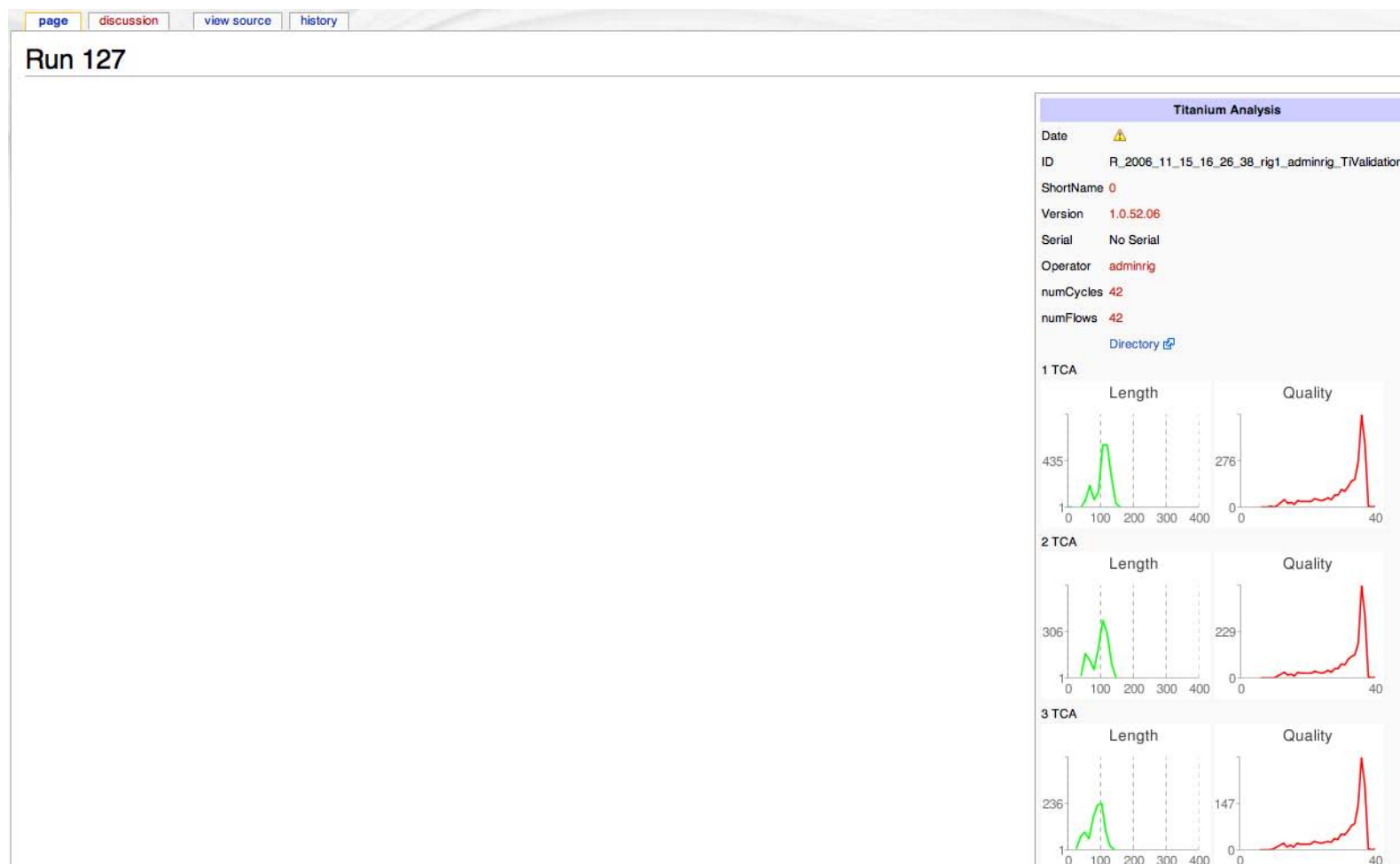
	<input checked="" type="checkbox"/> Sample Has Laboratory	<input checked="" type="checkbox"/> Sample has reagent	<input checked="" type="checkbox"/> Sample Has User
CME W1 Cl.8+-60	Graveley		Liyang
CME W1 Cl.8+-62	Graveley		Liyang
D mel/D sec Hybrid - 450 bp	Graveley		Mcmanus
D. mel adult	Graveley		Mcmanus
D.mel/D.sec mix - 250 bp	Graveley		Mcmanus
D.mel/D.sec mix - 450 bp	Graveley		Mcmanus
Kc167	Graveley		Liyang
Kc167-2	Graveley		Liyang
Kc167-4	Graveley		Liyang
ML-DmBG3-c2-122	Graveley		Liyang
ML-DmBG3-c2-124	Graveley		Liyang
S2-DRSC Brr2 RNAi rRNA minus-1	Graveley		Liyang

ModENCODE	
Project Laboratory	Graveley, Colniker

Cornell University

- Situation
 - 1 Roche GS and 1 Illumina GA 2
 - Need to read customer and sample data from existing LIMS
 - Need to link to existing LIMS
 - Key pages: Samples, Customers, Illumina Runs, Roche Runs, Flowcells

Monitoring quality



Monitoring quality

page

discussion

view source

history

090203 HWI-EAS339 30RF7AAXX

This is a good example of GA2 with Summary.xml

Links:

090203 HWI-EAS339 30RF7AAXX is mentioned by

Illumina Run:

090203_HWI-EAS339_30RF7AAXX

Machine:

HWI-EAS339

Flowcell:

30RF7AAXX

Entry Date:

2009-02-09 11:46:50 AM

Run Date:

2009-02-03

Archived To

false

Tape:

Files:

/data/disk1/dnaservices/090203_HWI-EAS339_30RF7AAXX

30RF7AAXX-IPAR 1.3-Bustard1.3.2 06-02-2009 dnaservices-GERALD 06-02-2009 dnaservices

9 February 2009 17:50:17

30RF7AAXX-IPAR 1.3-Bustard1.3.2 05-02-2009 dnaservices-GERALD 05-02-2009 dnaservices

9 February 2009 17:50:11

Analyses:

30RF7AAXX-IPAR_1.3-Bustard1.3.2_05-02-2009_dnaservices-GERALD_05-02-2009_dnaservices, 30RF7AAXX-IPAR_1.3-Bustard1.3.2_06-02-2009_dnaservices-GERALD_06-02-2009_dnaservices

Get Analyses

090203 HWI-EAS339 30RF7AAXX

Date

2009-02-06 13:05:49

Machine

HWI-EAS339

RunFolder

090203_HWI-EAS339_30RF7AAXX

Yield

1647383112

Raw

75905438

PF

45760642

Lane

Aligned

Passed

Score

Errors

lane 1

0.00%

74.32%

0.00

0.00

lane 2

0.00%

81.79%

0.00

0.00

lane 3

0.00%

78.08%

0.00

0.00

lane 4

94.48%

6.37%

151.22

2.05

lane 5

0.00%

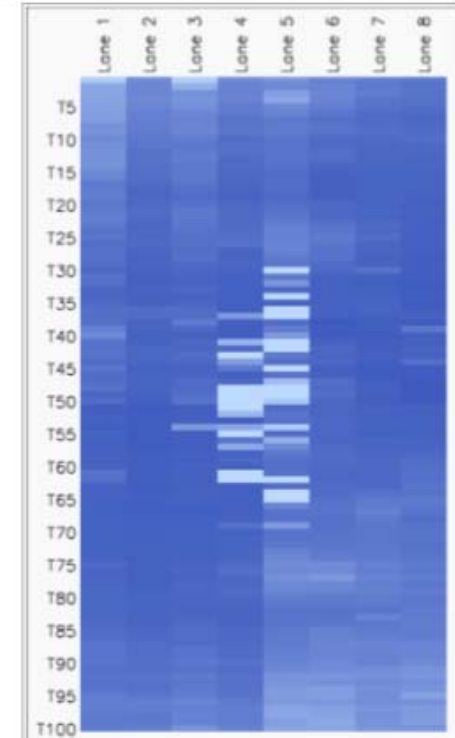
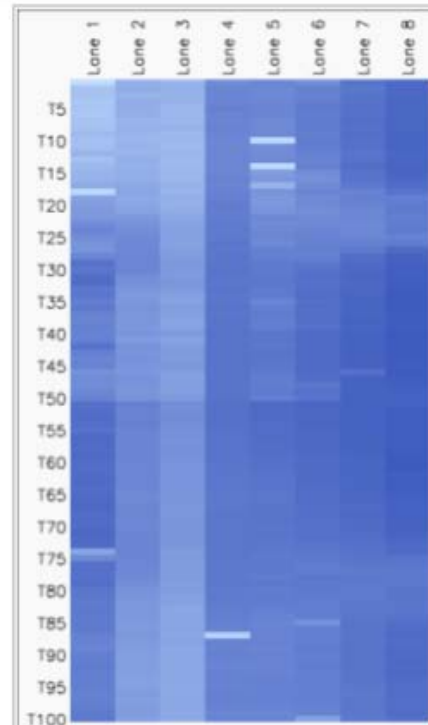
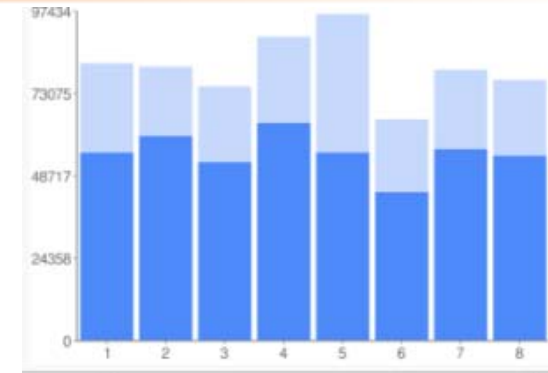
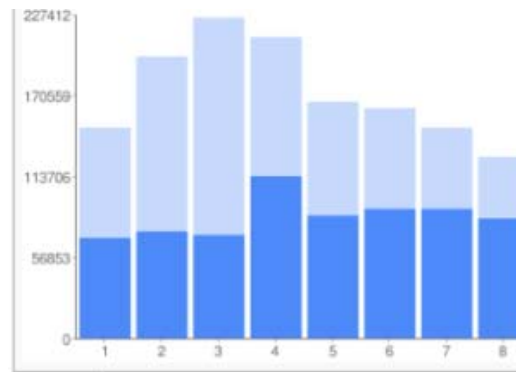
78.54%

0.00

0.00

WIKILIMS™

Monitoring quality



Cold Spring Harbor Laboratory

- Situation
 - 13 Illumina GA sequencers
 - Need to run large number of instruments used by many technicians
 - Need secure environment for clinical samples
 - Key Pages: Illumina Runs, Flowcells, Libraries, PCR Reactions, Genome Amplifications, Machines, Purifications

Create and edit Library pages

[page](#)
[discussion](#)
[edit with form](#)
[edit](#)
[history](#)
[move](#)
[watch](#)

Edit Library: LID2301

Library

Date: : :

Sample_id:

Dna_input:

Contact:

Constructor:

Originator:

Type:

Post-enrichment concentration:

Working dilution:

pM to load

Shearing Method:

Pressure

Duration

Adaptor:

Quantification method:

Size:

Primer:

Monitor ongoing runs

[page](#)
[discussion](#)
[edit](#)
[history](#)
[move](#)
[watch](#)

Recently started runs

For up-to-date values 'edit' then 'save'.

BILLIEHOLIDAY

RD434	BILLIEHOLIDAY	14 March 2009 13:24	6 Days	3d 6h 33m 43s	still going
RD427	BILLIEHOLIDAY	8 March 2009 16:12	2.5 Days	9d 3h 45m 43s	finished

[... further results](#)

BRITNEYSPEARS

RD414	BRITNEYSPEARS	28 February 2009 14:02	2.5 Days	17d 4h 55m 43s	finished
RD407	BRITNEYSPEARS	19 February 2009 17:12	8 days	26d 1h 45m 43s	finished

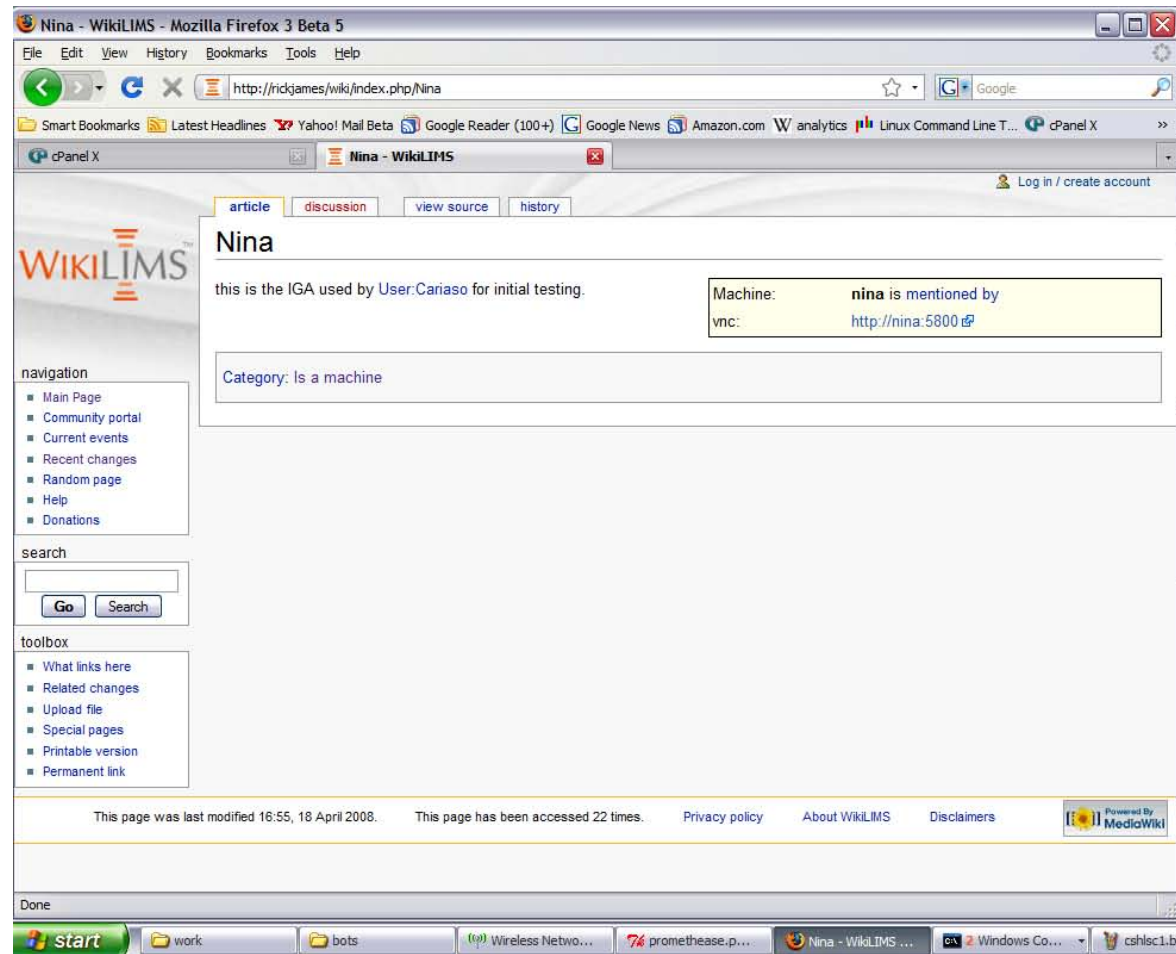
[... further results](#)

FREDDYKRUEGER

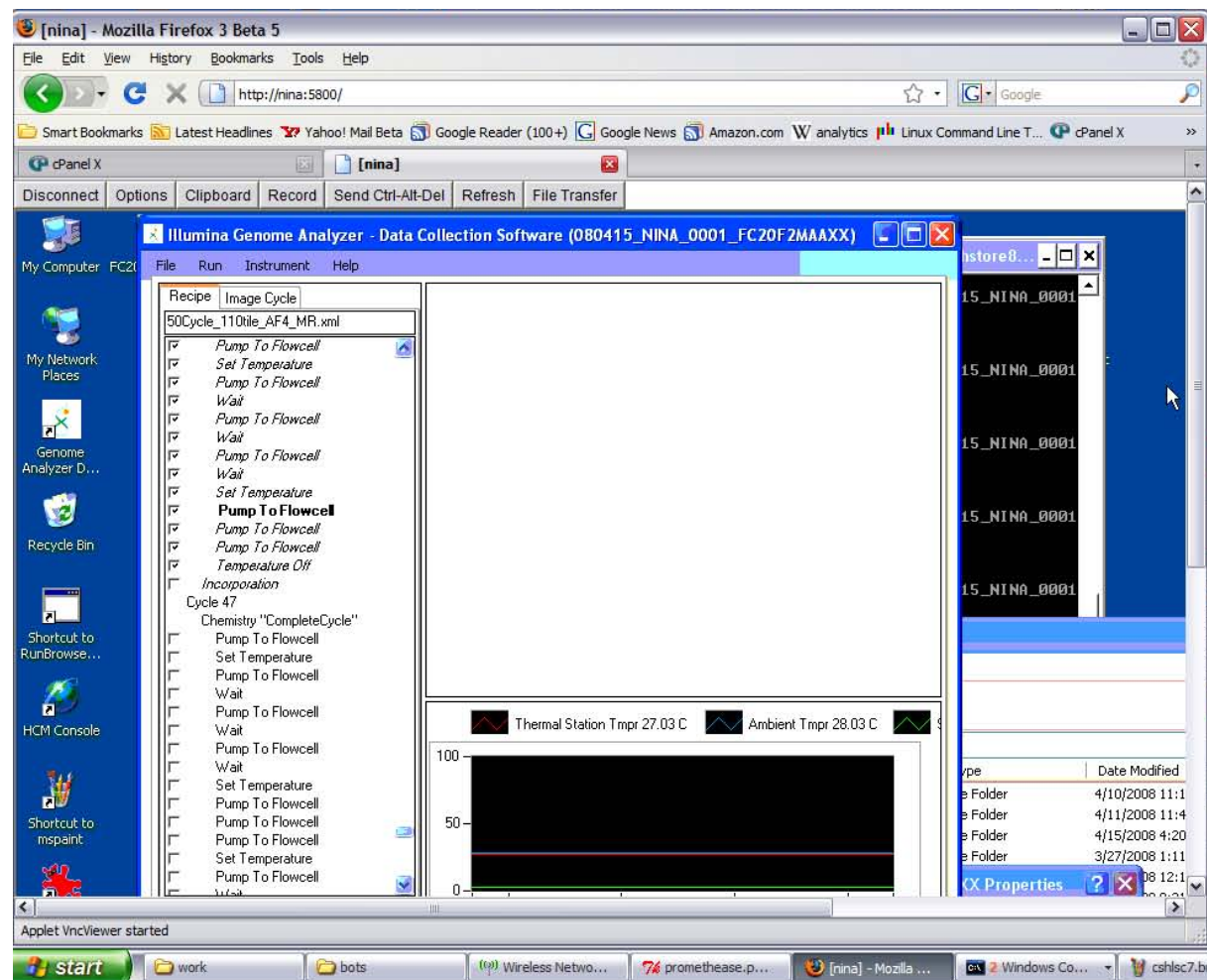
RD433	FREDDYKRUEGER	11 March 2009 10:06	6 Days	6d 9h 51m 43s	finished
RD423	FREDDYKRUEGER	6 March 2009 11:15	5 days	11d 7h 42m 43s	finished

[... further results](#)

Remote Instrument Operation



Remote Instrument Operation



Custom query interface



[special](#)
69.140.171.10 talk for this ip log in / create account

Query	Additional printouts (optional)
[[Category:Is a genotype]] [[Magnitude::>0]]	?Magnitude

Sort by column (optional) [\[Add sorting condition\]](#)

[Find results](#) [Hide query](#) [Querying help](#)

[Previous](#)
[Results 1– 10](#)
[Next](#)
(20 | 50 | 100 | 250 | 500)

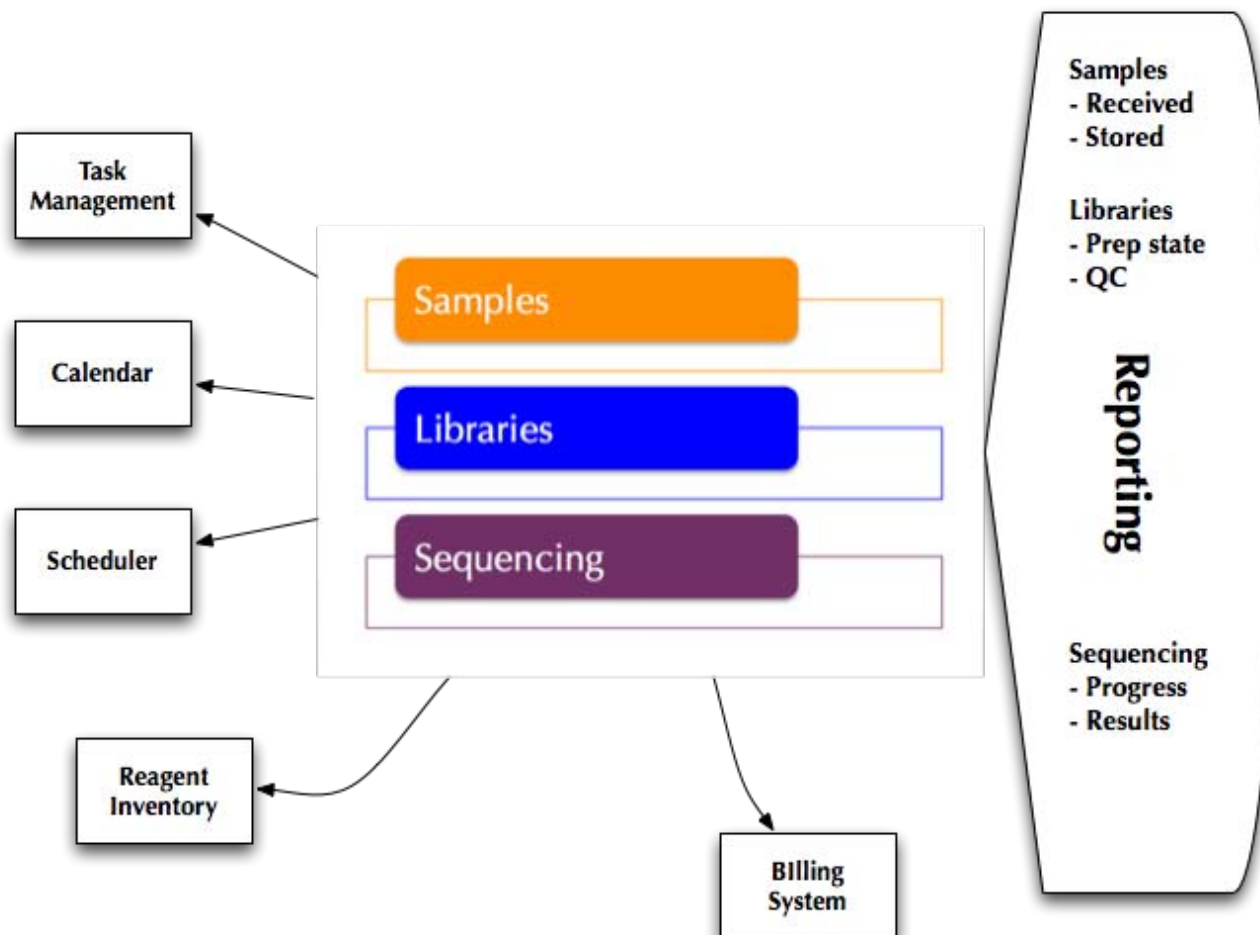
 >>	 << Magnitude
Rs4420638(G;G)	10
Rs6152(A)	3
Rs6265(A;G)	3
Rs671(A;A)	3
Rs671(A;G)	3
Rs7659604(C;T)	2
Rs2903265(A;G)	1
Rs17672135(C;C)	0
Rs1800454(G;G)	0
Rs3788853(G;G)	0

[Previous](#)
[Results 1– 10](#)
[Next](#)
(20 | 50 | 100 | 250 | 500)

Indiana University Center for Genomics and Bioinformatics

- Situation
 - 1 Roche GS and 1 Illumina GA, NimbleGen microarrays
 - Need to track runs, samples, reagents, and group by project
 - Need to track task-level and job-level provenance data
 - Need to send notifications and email alerts
 - Need to carry projects through all the way to billing
 - Key : Projects, Samples, Libraries, Sequencing, Reagents

Management, workflow, and reporting



Managing workflow tasks

Plan	Assign	Confirm	Titration	Sequencing
Status: <input type="text" value="planned"/>				
Final library trace: <input type="text" value="0.05"/>				
Final library quant: <input type="text" value="30.0"/>				
Final library comments: <input type="text" value="Target Library Pool 1"/>				
PTP/Flowcell plan: <input type="text" value="42M2FAAXX"/>				
Run date plan: <input type="text"/>				
Free text: <input type="text"/>				
Summary: <input type="text"/>				

Managing workflow tasks

Plan	Assign	Confirm	Titration	Sequencing
Lab member assigned to titrate library: <input type="text" value="Jaalopez"/>				
Notes on titration: <div><div>Notes for pool 1</div></div>				
Lab member assigned bulk: <input type="text" value="Jaalopez"/>				
Lab member assigned enrichment: <input type="text" value="Jbford"/>				
Lab member assigned run: <input type="text" value="Ahemmeri"/>				
Lab member assigned cluster generation: <input type="text" value="Ahemmeri"/>				
Lab member assigned cluster QA: <input type="text" value="Kmockait"/>				
Free text: <div></div>				

Managing workflow tasks

Plan	Assign	Confirm	Titration	Sequencing
Library receipt confirmed: <input checked="" type="checkbox"/>				
Reagents reserved: <input checked="" type="checkbox"/>				
Proposed schedule for quant and titration: 8 October 2009				
Assignment of bulk confirmed: <input checked="" type="checkbox"/>				
Bulk reagents reserved: <input type="checkbox"/>				
Proposed schedule for bulk: 10 October 2009				
Assignment of enrichment confirmed: <input checked="" type="checkbox"/>				
Enrichment reagents reserved: <input type="checkbox"/>				
Proposed schedule for enrichment: 17 October 2009				
Assignment of run set-up confirmed: <input type="checkbox"/>				
Run reagents reserved: <input type="checkbox"/>				
Proposed schedule for run: <input type="checkbox"/>				
Free text: <div></div>				

Managing workflow tasks

Plan	Assign	Confirm	Titration	Sequencing
------	--------	---------	------------------	------------

Titration completed: ☒

Titration results:

Free text:

Summary:

☐ This is a minor edit ☐ Watch this page

[Cancel](#)

Managing workflow tasks

Plan	Assign	Confirm	Titration	Sequencing
Cycling of bulk confirmed: <input checked="" type="checkbox"/>				
Cycle numbers used for bulk: <input type="text" value="8"/>				
Completion of enrichment confirmed: <input checked="" type="checkbox"/>				
Enrichment results tabulated: <input type="text" value="enrichment_results.xls"/>				
Completion of run set-up confirmed: <input type="checkbox"/>				
Loading of PTP regions or labeling of flowcell lanes confirmed: <input type="checkbox"/>				
Assignment of processing and analysis script confirmed: <input type="checkbox"/>				
Free text: <div><div></div></div>				
Summary: <input type="text"/>				
<input type="checkbox"/> This is a minor edit <input type="checkbox"/> Watch this page				

Sending e-mail notifications

Delete Reply Reply All Forward New Message Note To Do

Q [WikiLIMS] 9 Found

All Mailboxes	Inbox	Entire Message	From	To	Subject	Date Received	Mailbox
			From		Subject		
			www-data		[WikiLIMS] Task updated: Samples/2	August 18, 2009 12:48 PM	All Mail
			www-data		[WikiLIMS] Task updated: Sample 1	October 7, 2009 6:08 PM	All Mail
			www-data		[WikiLIMS] Task updated: Sample 1	October 7, 2009 5:49 PM	All Mail
			www-data		[WikiLIMS] Task updated: Sample 1	October 7, 2009 5:44 PM	All Mail
			www-data		[WikiLIMS] Task updated: Sample 1	October 7, 2009 5:43 PM	All Mail
			www-data		[WikiLIMS] Task updated: Library 2	August 18, 2009 11:25 PM	All Mail
			www-data		[WikiLIMS] New task: Samples/1	August 18, 2009 12:56 PM	All Mail
			www-data		[WikiLIMS] New task: Sample 2	August 18, 2009 1:28 PM	All Mail
			www-data		[WikiLIMS] New task: Library 2	August 18, 2009 3:52 PM	All Mail

From: www-data <www-data@cgb.indiana.edu>
 Subject: [WikiLIMS] New task: Sample 2
 Date: August 18, 2009 1:27:59 PM EDT
 To: Bioteam <kraut@bioteam.net>

Hello Bioteam,

The task "Sample 2" has just been assigned to you http://localhost:8080/wiki/index.php/Sample_2

Here is the task description:
 The database did not find the text of a page that it should have found, named "Sample 2".

This is usually caused by following an outdated diff or history link to a page that has been deleted.

If this is not the case, you may have found a bug in the software.
 Please report this to an [\[\[Special:ListUsers/sysop|administrator\]\]](#), making note of the URL.

Simplified tracking information

← Older edit

Line 5:

```
!Library type=pool
!Lab member assigned to prepare library=Bioteam

}}
```

Line 5:

```
!Library type=pool
!Lab member assigned to prepare library=Bioteam
+ !Sample receipt confirmed=Yes
+ !Sample QA confirmed=No

}}
```

- Email and RSS notifications for every step in workflow
- Wiki Revision Control explains **who** did **what** and **when**
- Lab Managers can revert and undo tasks

Managing tasks by Lab User – Calendar view

User:Bioteam

Email address kraut@bioteam.net

My Samples	My Libraries	My Sequencing
<div>Created on</div> <div>Sample 1 28 August 2009</div>	<div>Created on</div> <div>Library 2 20 August 2009</div>	No Sequencing found

August 2009

◀ Today ▶

August 2009

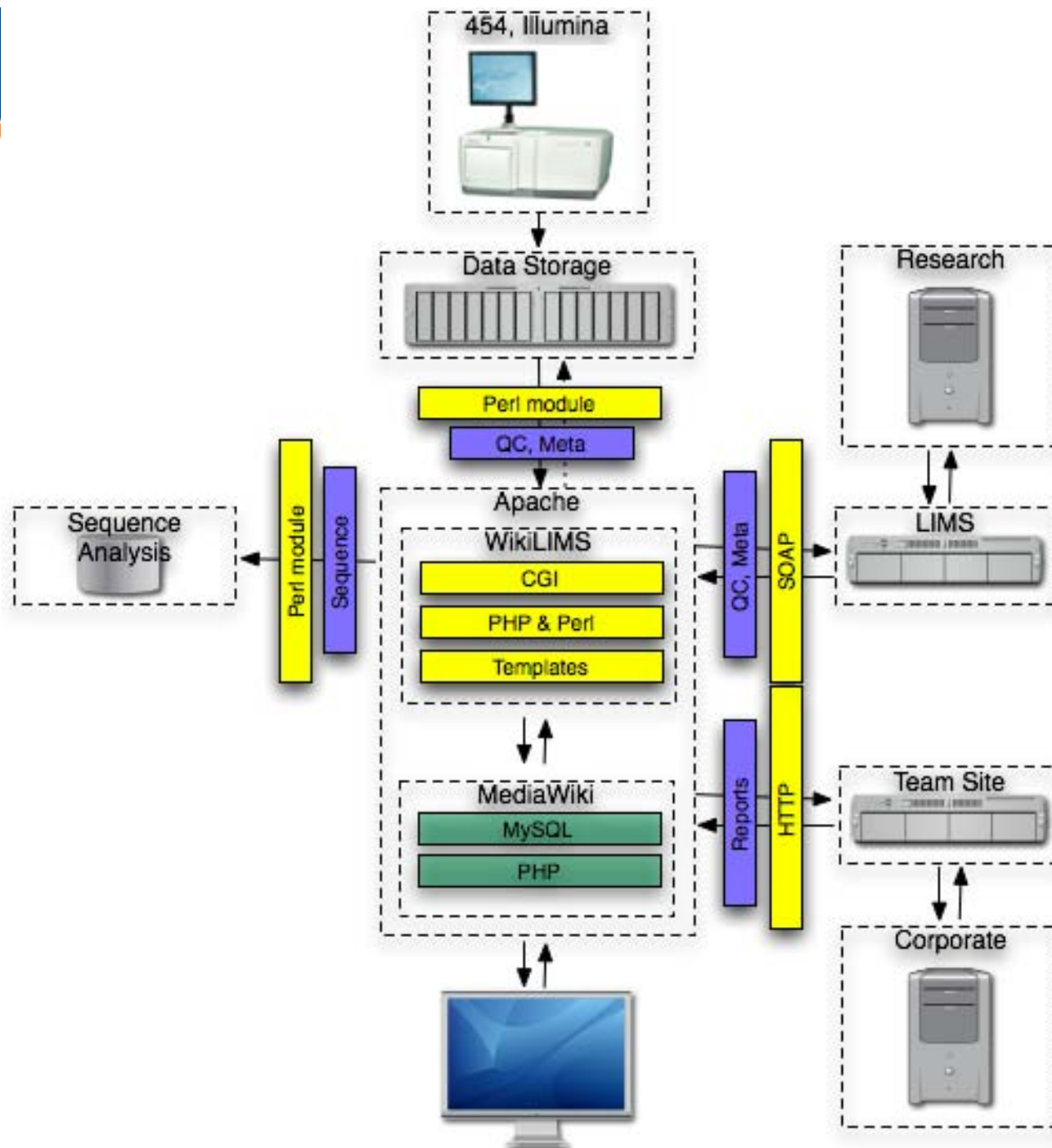
Go to month

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
26	27	28	29	30	31	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	Sample 1	28
30	31	1	2	3	4	5

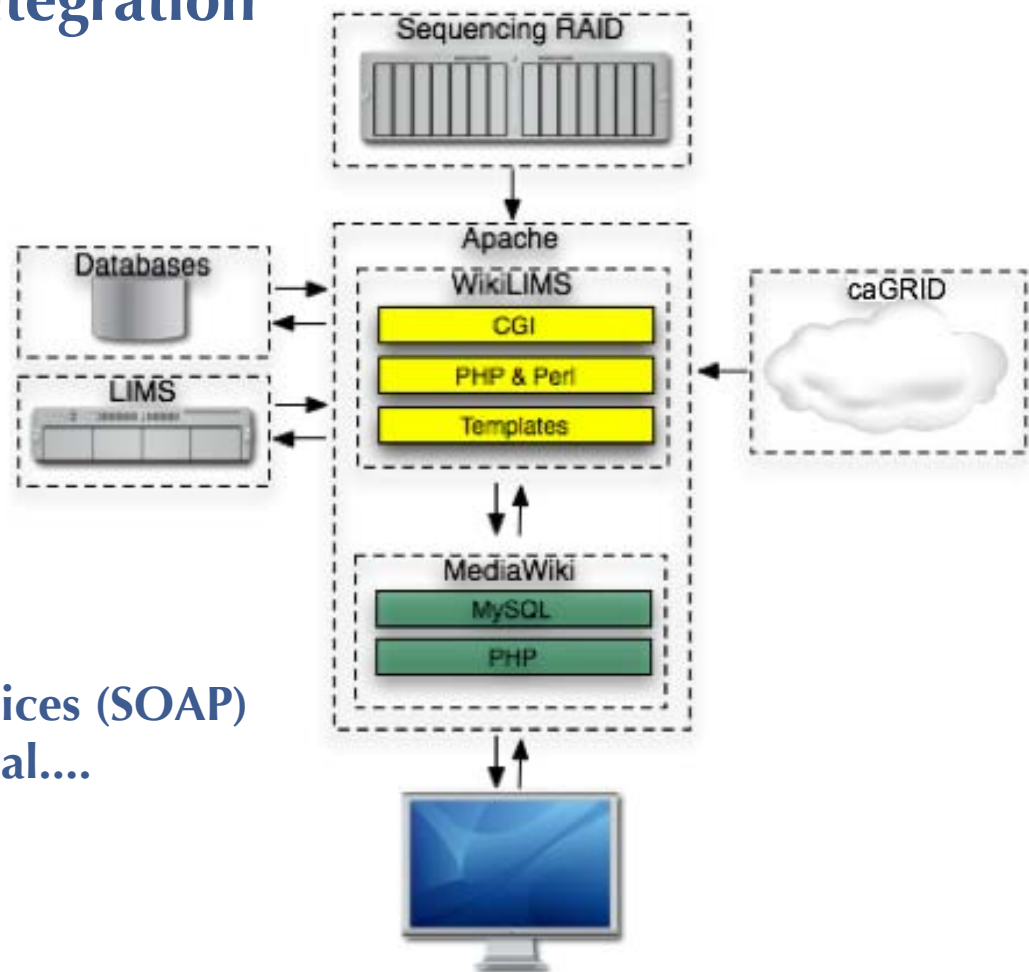
Category: Laboratory Member

“multi-national corporation”

- Situation
 - 3 Roche GS and 2 Illumina GA
 - Existing commercial LIMS system
 - Existing commercial sequence analysis platform
 - Existing collaborative platforms, Web-based
 - Need to make projects visible
 - Need automatic data movement in all directions
 - Key : Projects, Samples, Libraries, Sequencing



Future Directions: caBIG Integration



- WikiLIMS queries caBIG via Web Services (SOAP)
- Gene, protein, microarray, SNP, clinical....

Future Directions: caBIG Integration

[page](#)
[discussion](#)
[edit](#)
[history](#)
[delete](#)
[move](#)
[protect](#)
[watch](#)
[refresh](#)

CAGRID Demo

Proteins from caGRID

	Protein Primary Accession	Checksum	Sequence Length
Q00604	Q00604	D219E8B7F957286A	133
P44444	P44444	B67015EBF8FBA23F	238
P38398	P38398	89C6D83FF56312AF	1,863
P12345	P12345	410321530B95B673	30
A4 HUMAN	P05067	A12EE761403740F5	770
P00107	P00107	CBCDCDEE026A9C64	83

⚠

Retrieve Protein from caBIG

☒ Accession
 ☐ Symbol

Also see [CAGRID Notes](#)


Data current as of March 17, 2009, 15:13. [Clear cache](#)

Future Directions: caBIG Integration

[page](#)
[discussion](#)
[edit with form](#)
[edit](#)
[history](#)
[delete](#)
[move](#)
[protect](#)
[watch](#)
[refresh](#)

P44444

Run BLASTP

Symbol	P44444	Accession
Species		MW
MRPNNRENNQPRQIKITRNYTKHAEGSVLVEFGDTKVLCTATVEDAVPRFLKGQGQGWVTAEYGMLPRSTHSRMQREAAKG		

Find Matches

Category: [CAGRID Protein](#)

Future Directions: caBIG Integration

BRCA1

```
TBLASTN 2.2.10 [Oct-19-2004]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Query= BRCA1_HUMAN
      (1863 letters)

Database: NTs.fa
      9 sequences; 567 total letters

Searching.....done

Sequences producing significant alignments:
```

	Score (bits)	E Value
77209	49	1e-10
1750	49	1e-10
9911	46	7e-10
9082	46	9e-10
305	45	2e-09

```
>77209
      Length = 65

      Score = 48.9 bits (115), Expect = 1e-10
      Identities = 21/21 (100%), Positives = 21/21 (100%)
      Frame = +1

Query: 371 PWITLNSSIQKYMEWFSRDE 391
      PWITLNSSIQKYMEWFSRDE
Sbjct: 1  PWITLNSSIQKYMEWFSRDE 63
```

Run TBLASTN

Symbol	BRCA1
Accession	P38398
Protein	MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQESTRFSQVLEELKI
Species	Homo sapiens
Facts about BRCA1 ⓘ	
CAGRID Gene Protein	MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHI ... HAIGQMCEAPVVTREWVLDVA
CAGRID Gene Symbol	BRCA1 + 🔍


Future Directions: Proteomics

article discussion view source history

GEL1233

This page is all about a 2D gel image. Try moving your mouse over some of the spots. Remember I cooked this one by hand, so not all spots are active. Everything scales, so feel free to edit this page and change, or remove, the size parameter.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce pellentesque odio porta ipsum. Nunc nisl odio, vehicula a, feugiat sit amet, tincidunt sit amet, metus. Sed neque. Donec venenatis vestibulum purus. Duis auctor augue eget metus. Quisque sit amet erat. [Suspendisse](#) a urna. Nunc cursus magna tincidunt arcu. Morbi augue. Suspendisse accumsan odio eu risus. Phasellus fermentum, dui in consequat tempor, ligula magna rutrum ante, in rhoncus tortor metus eu dolor. Sed non dolor et purus vehicula viverra. Suspendisse pede ligula, laoreet et, volutpat eget, dapibus ut, urna.



[Spot 123456](#) is the biggest one. This represents the protein TCF7L2 in the [Pseudomonas Carolinis](#). You may also be interested in [Spot 987654](#) which was later determined to be a contamination error.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce pellentesque odio porta ipsum. Nunc nisl odio, vehicula a, feugiat sit amet, tincidunt sit amet, metus. Sed neque. Donec venenatis vestibulum purus. Duis auctor augue eget metus. Quisque sit amet erat. Suspendisse a urna. Nunc cursus magna tincidunt arcu. Morbi augue. Suspendisse accumsan odio eu risus. Phasellus [fermentum](#), dui in consequat tempor, ligula magna rutrum ante, in rhoncus tortor metus eu dolor. Sed non dolor et purus vehicula viverra. Suspendisse pede ligula, laoreet et, volutpat eget, dapibus ut, urna.

Future Directions: Proteomics

[article](#) [discussion](#) [view form](#) [view source](#) [history](#)

Spot 9999

This spot is sort of interesting.

Category: [Spot](#)

Spot 9999	
Name	Lesotho
Protein	AVPRT1

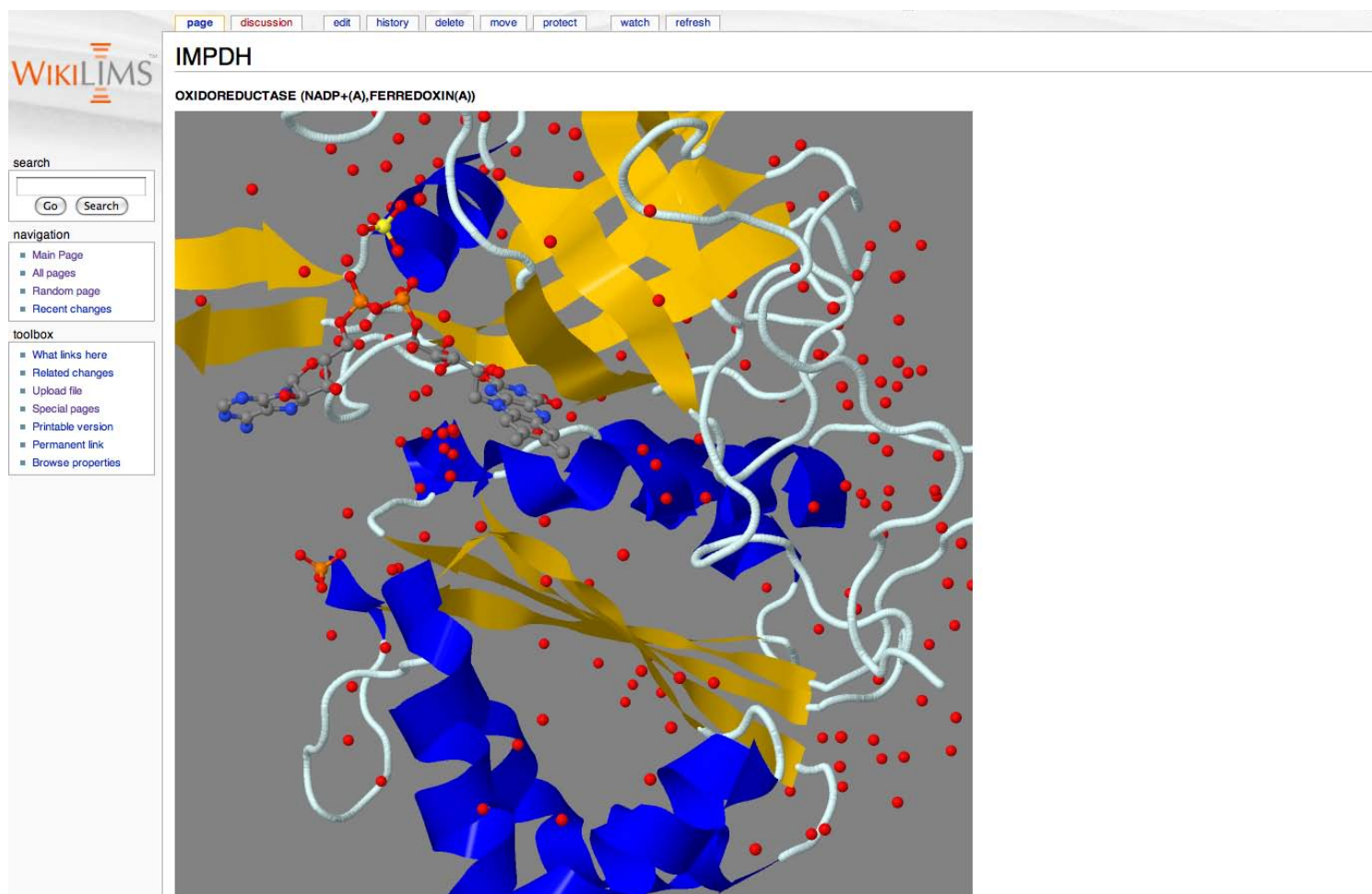
Future Directions: SOAP and RESTful Web Services

The screenshot shows a Wikisysop page for 'AL3A2 HUMAN'. The page has a top navigation bar with links like 'page', 'discussion', 'edit', 'history', 'delete', 'move', 'protect', 'watch', and 'refresh'. On the right, there are user links: 'Wikisysop', 'my talk', 'admin links', 'my preferences', 'my watchlist', 'my contributions', and 'log out'. The main content area is divided into two sections. The left section is titled 'AL3A2 HUMAN' and is mostly empty, with a blue box labeled 'Annotations' pointing to it. The right section is also titled 'AL3A2 HUMAN' and contains a table of properties: Name(s), Domain, Function, Organism, Catalytic activity, Length, and Uniprot. Below these is a 'Sequence' section with a long amino acid sequence. A blue box labeled 'External Data' points to the 'Uniprot' link in the table. At the bottom left, there is a 'Category: Proteins' label.

AL3A2 HUMAN	
Name(s)	Fatty aldehyde dehydrogenase
Domain	Transmembrane
Function	Oxidoreductase
Organism	Homo sapiens (Human)
Catalytic activity	An aldehyde + NAD(+) + H(2)O = an acid + NADH.
Length	485 AA
Uniprot	http://www.uniprot.org/uniprot/P51648
Sequence	MELEVRVRVQAFLSGRSRPLRFRLQQLEALRRMVQEREKDLTAIAADLCKSEFNVYSQE VITVLGEIDFMLENLPEWVTAKPVKKNVLTMLDEAYIQPQPLGVVLIIGAWNYPFVLTIQ PLIGIAAGNAVIKPSSELSSENTAKILAKLLPQYLDQDLYIVINGGVEETELLKQRFDH IFYTGNTAVGKIVMEAAKHLTPVTLELGGKSPCYIDKDCDLIVCRITWGKYMNCGQT CIAPDYILCEASLQNIIVWKIKETVKEFYGENIKESPDYERIINLRHFKRILSLLEGQKI AFGGTDEATRYIAPTDLTDVDPKTKVMQEEIFGPILPVPVKNVDEAINFINEREKPLA LYVFSHNHKLKRMIDETSSGGVTGNDVIMHFTLNSFPFGGVSSGMGAYHGKHSFDTFS HQRPCLLSLKREGANKLRYPPNSQSKVDWGKFFLLKRFNKEKLGILLLTFLGIVAAVLV KAEYY ⚠

Category: Proteins

Future Directions: 3D Structure Viewing (Jmol)



Future Directions: MALDI-TOF Data (R, Gnuplot)

