



**RDECOM**



# Automated Analytical Platform for Rapid Identification and Characterization of Unknown Genomic Samples



EDGEWOOD CHEMICAL BIOLOGICAL CENTER

**TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.**

C. Nicole Rosenzweig, Ph.D.  
ECBC Genomics Integrated Product Team  
Senior Bioinformatics Scientist, OptiMetrics, Inc.  
18 November 2009

Unclassified Approved for Public Release - Distribution **WARFIGHTER FOCUSED.**



To provide reliable sequence data and analysis for rapid, automated pathogen identification

- Sequencing technologies available for decades
- Analysis methods are available for many commonly used procedures
- Computational hardware is cheap



NCBI **genomic BLAST**

BLAST Microbial Fungi Plants Insects Nematodes Environmental Protozoa Eukaryota Help

**BLAST with microbial genomes (1102 bacterial 63 archaeal 193 eukaryotic genomes tree)**

P - indicates the ability to search against protein sequences,   - completed genomic sequence,   - Whole Genome Shotgun,   - add/remove from selection, See [\[Help\]](#) and [\[Article\]](#) for details.

**Attention!**  
Because genome sequencing of DDBJ/EMBL/GenBank, the users of these sequences

**EnsemblBacteria**  
Home > Mycobacterium > Mycobacterium tuberculosis H37Rv  
Location: Chromosome:2,071,036-2,071,914

Location-based displays  
 Whole genome  
 Chromosome summary  
 Region overview  
 Region in detail  
 Comparative Genomics  
 Genomic alignments (1)  
 Synteny (17)  
 Genetic Variation  
 Resequencing (0)  
 Linkage Data  
 Markers  
 Other genome browsers

• Configure this page  
 • Manage your data  
 • Export data  
 • Bookmark this page

Enter your query sequence  
 Select type of query and database  
 Query: DNA Database:  
 You may change BLAST of Expect: 10 Filter:

Chromosome: 2,071,036-2,071,914

Chromosome

Export image

Region overview Region in detail [help](#) Genomic alignments

Contigs  
 EnsemblBacteria ...

Contig	Start	End	Gene
Rv1784	Rv1797	PPE31	Rv1817
Rv1785	Rv1798	PPE32	PPE_PGRS33
Rv1786	ppT	Rv1804c	Rv1809c
PPE25	PPE28	PPE33	Rv1828
PE18	PPE29	Rv1810	seA2
PPE26	ppe30	Rv1812c	spA2
PPE27	PE_PGRS32	Rv1829	glcB
PE19	Rv1806c	Rv1824	PE_PGRS34
esxM	PE20	erg3	Rv1825
esxN	mgcT	gcvH	Rv1843c
Rv1794	Rv1813c	Rv1829	ureA
Rv1795	Rv1815	Rv1830	ureB
mycP5	Rv1816	Rv1831	ureC
		gcvP	ureD
			ureF
			ureG
			ureH

Ensembl Mycobacterium\_tuberculosis\_H37Rv version 55.2a (EB1) Chr

- High-throughput shotgun sequencing requires substantial laboratory expertise and the ability to track laboratory data for efficient reporting
- Many analysis methods are not immediately interoperable
- Capability gaps do exist in this specialized analytical application

1. Track laboratory electronically
2. Provide infrastructure which allows for adequate scalability and automated tracking of analytical methods and parameters used
3. Automate the execution of a standardized analytical pipeline while providing a mechanism for data-driven research

- Capture sample metadata
- Track all reagents, hardware, and most consumables
- Automate calculations
- Experimentally-driven data collection
- Integrated with Roche 454 Sequencers
- Integrated with analysis pipeline



# Laboratory Information Management System (LIMS)



Velox Data Management

Application Manager | Actions | Report Manager | Workflows | Calendar | Audit Log | Account | Product Guide | Logoff

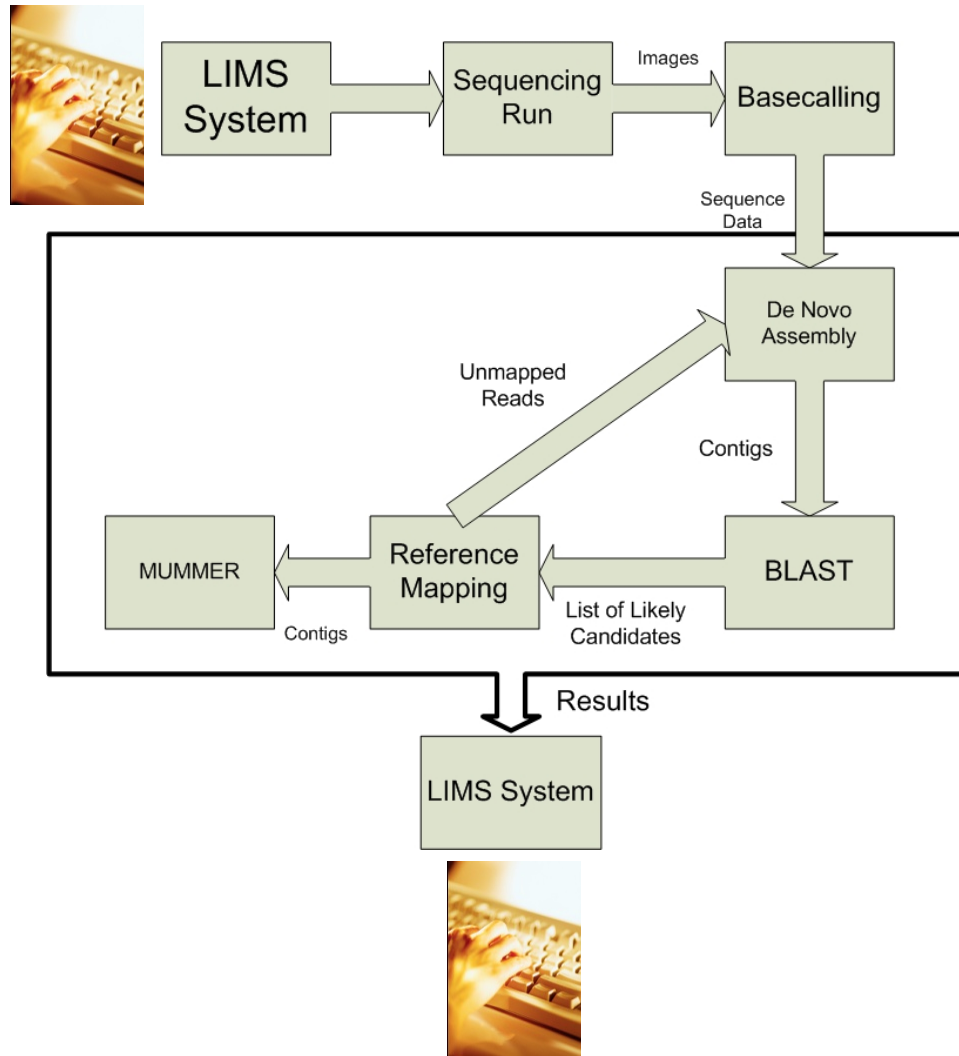
Data | Search | Workflow

De Novo Assembly Results

ID	Median Depth Of Coverage	Total Number Of Reads	Total Number Of Bases	Number of Aligned Reads	Percent Aligned Reads	Number Of Large Contigs	Average Contig Length
1 S000001 : 12-17-2008 : 0	99	1228399	428180393	1220729	99.38	34	12
2 S000001 : 12-17-2008 : 1	60	686448	227069951	680725	99.17	1393	
3 S000001 : 12-17-2008 : 2	60	685812	226853464	680075	99.16	1396	
4 S000002 : 03-19-2009 : 0	83	487990	154850182	483238	99.03	92	1
5 S000002 : 03-19-2009 : 1	76	4717	1257077	3842	81.45	10	
6 S000002 : 03-19-2009 : 2	44	657	66720	89	13.55	1	
7 S000003 : 03-19-2009 : 0	86	518917	162729572	514078	99.07	98	1
8 S000003 : 03-19-2009 : 1	54	1003	125627	229	22.83	1	
9 S000003 : 03-19-2009 : 2	2	564	42560	2	0.35	0	
10 S000004 : 03-19-2009 : 0	81	464371	150878524	460531	99.17	70	2
11 S000004 : 03-19-2009 : 1	0	991	72226	0	0.00	0	
12 S000005 : 03-06-2009 : 0	36	542318	170469663	537407	99.09	256	1
13 S000005 : 03-06-2009 : 1	33	8392	2461165	7822	93.21	9	
14 S000005 : 03-06-2009 : 2	175	5056	1445443	4615	91.28	3	
15 S000006 : 03-06-2009 : 0	32	520246	161201353	516437	99.27	235	1
16 S000006 : 03-06-2009 : 1	39	39512	12117771	38774	98.13	21	
17 S000006 : 03-06-2009 : 2	33	19899	6034450	19286	96.92	13	1
18 S000007 : 03-19-2009 : 0	28	443379	139849073	439396	99.10	254	1
19 S000007 : 03-19-2009 : 1	46	38775	11932070	37655	97.11	15	1
20 S000007 : 03-19-2009 : 2	74	16228	4903164	15219	93.78	8	
21 S000008 : 04-08-2009 : 0	66	521312	133188077	493048	94.58	112	1
22 S000008 : 04-08-2009 : 1	65	1220	130041	248	20.33	1	
23 S000008 : 04-08-2009 : 2	0	729	52688	0	0.00	0	
24 S000009 : 03-26-2009 : 0	32	390119	137532789	387349	99.29	36	11
25 S000009 : 03-26-2009 : 1	23	219004	73409531	214968	98.16	1701	
26 S000009 : 03-26-2009 : 2	23	218760	73328726	214709	98.15	1699	
27 S000010 : 03-26-2009 : 0	31	406721	132539183	404029	99.34	30	13

Results

User: bioinformatics Release 3.1.0 - Build 204



## Identification of nearest neighbors

Number Samples Sequenced	Sample Identity	Predicted Nearest Neighbor
1	<i>Burkholderia pseudomallei</i>	<i>Burkholderia pseudomallei</i>
5	<i>Francisella tularensis</i>	<i>Francisella tularensis</i>
9	<i>Salmonella typhimurium</i>	<i>Salmonella typhimurium</i>
18	<i>Yersenia pestis</i>	<i>Yersenia pestis</i>
13	<i>Bacillus globigii</i>	<i>Bacillus subtilis subsp. subtilis</i>

## Identification of components of samples

Sample Sequenced	Iteration	Predicted Nearest Neighbor
Yersinia pestis Antigua	0	Yersinia pestis Antiqua, complete genome
Yersinia pestis Antigua	1	Yersinia pestis Antiqua plasmid pCD, complete sequence
Yersinia pestis Antigua	2	Yersinia pestis Antiqua plasmid pMT, complete sequence
Yersinia pestis Kim	0	Yersinia pestis KIM, complete genome
Yersinia pestis Kim	1	Yersinia pestis KIM plasmid pCD1, complete sequence
Yersinia pestis Kim	2	Yersinia pestis KIM plasmid pMT-1, complete sequence
Salmonella typhimurium LT2 Derivative	0	Salmonella typhimurium LT2, complete genome
Salmonella typhimurium LT2 Derivative	1	Salmonella typhimurium LT2 plasmid pSLT, complete sequence
Yersenia Pestis Angola	0	Yersinia pestis Angola, complete genome
Yersenia Pestis Angola	1	Yersinia pestis Angola plasmid pMT-pPCP, complete sequence
Yersenia Pestis Angola	2	Yersinia pestis Angola plasmid new_pCD, complete sequence

But what does automation provide at this granularity of analysis

Sample Sequenced	Iteration	Predicted Nearest Neighbor
Y. Pestis Sample	0	Yersinia pestis Nepal516, complete genome
Y. Pestis Sample	1	Bacillus anthracis str. Sterne, complete genome

- Incorporate Institute for Genome Sciences Annotation Engine
- Improved characterization of samples with no near neighbors in the reference database
- Automate Genbank submissions

- ECBC began supporting the TMTI Pathogen Characterization effort in October 2008
- Infrastructure in place to support the sequencing and rapid identification of pathogens
- Initial analytical capability provides baseline analysis while providing means to support research-driven analysis



## ECBC Genomics IPT

Evan W. Skowronski  
Henry S. Gibbons  
Lauren A. McNew

## OptiMetrics, Inc.

Alvin Liem  
Ed Fochler  
Steve Ehrman  
Greg Donarum

## External Collaborators

### LLNL

Tom Slezak

### UDelaware/PIR/Georgetown

Cathy Wu  
Hongzhan Huang  
Darren Natale  
Raja Muzumder

Funding provided by DTRA/TMTI (ECBC)