Information Appliances:

Targeted Support for High-throughput Laboratory Devices .

BRIITE

Strategic planning for IT support of grant-funded research, II

September 22, 2004 - September 24, 2004
Seattle , WA

Daniel E. Geraghty, Ph.D.

Fred Hutchinson Cancer Research Center

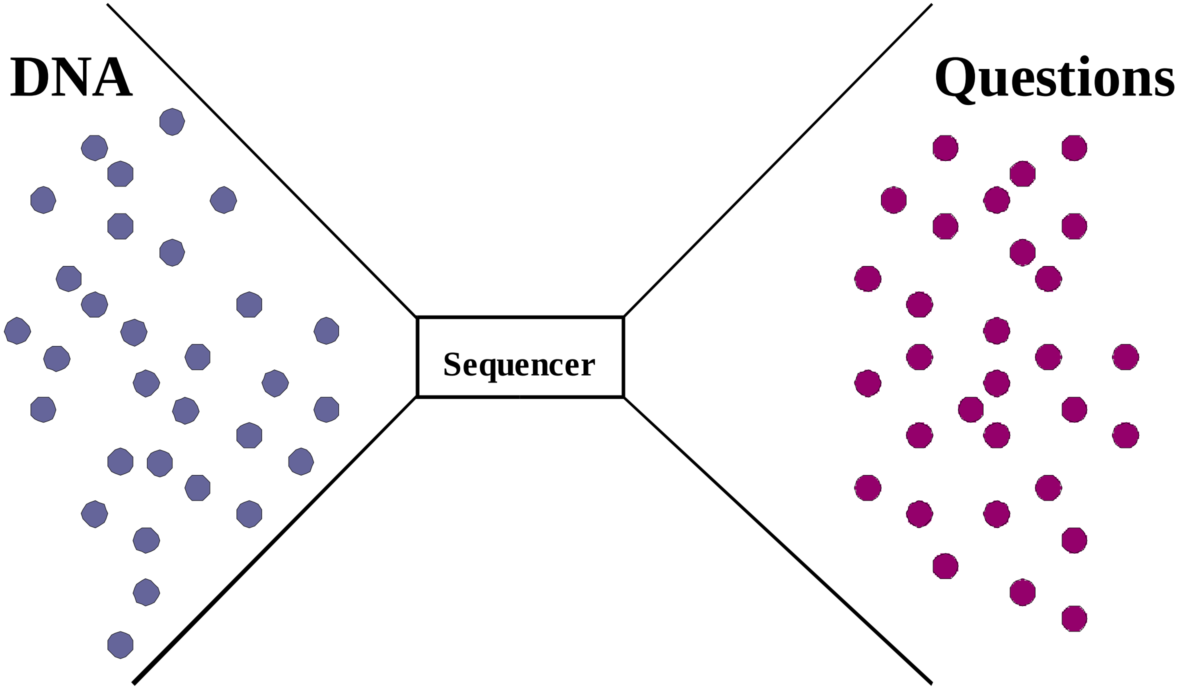
Information Appliances: Targeted Support for High-throughput Laboratory Devices.

1. Outline the scientific questions and experimental methods being applied in our lab.

2. Explain the major experimental limitations of these methods.

3. Describe our local solution for DNA sequencers and its general applicability to a variety of instruments, other sites, and to data sharing.

Information Appliances: Targeted Support for High-throughput Laboratory Devices.



DNA

Sequencer

Questions

Information Appliances: Targeted Support for High-throughput Laboratory Devices.

- Tracking laboratory throughput
- Organization of original data and meta data (machine, reagents, quality, etc.)
- Cost tracking, quality checks
- Data sharing

Creating a modular and extensible framework for future applications (STR, Taqman, HTR, etc.)

Genetic diversity and genomics of the immune response--future impact

- A patient’s precise genetic blueprint can be applied to:
  - *Improving bone marrow and organ transplant outcome through better matching.*
  - *Tailoring medicines to a patient’s genetic makeup.*
  - *Directing cancer treatment based on a tumors genetic makeup.*
  - *Predicting a person’s immune response to infection or vaccination.*
  - *Anticipating disease, including cancer susceptibility, autoimmunity, heart disease.*

- **Proactive vs. Reactive** medical treatment using the predictive ability of genetics to anticipate disease and counsel health.

Genetic diversity and genomics of the immune response.

Current studies:

1) Genetic discovery encompassing a panel of immune response loci.

2) Sequence analysis of uncharted genomic regions.

3) Application of genetic information to correlate genotypes with interesting immunological phenotypes.

Genetic diversity and genomics of the immune response.

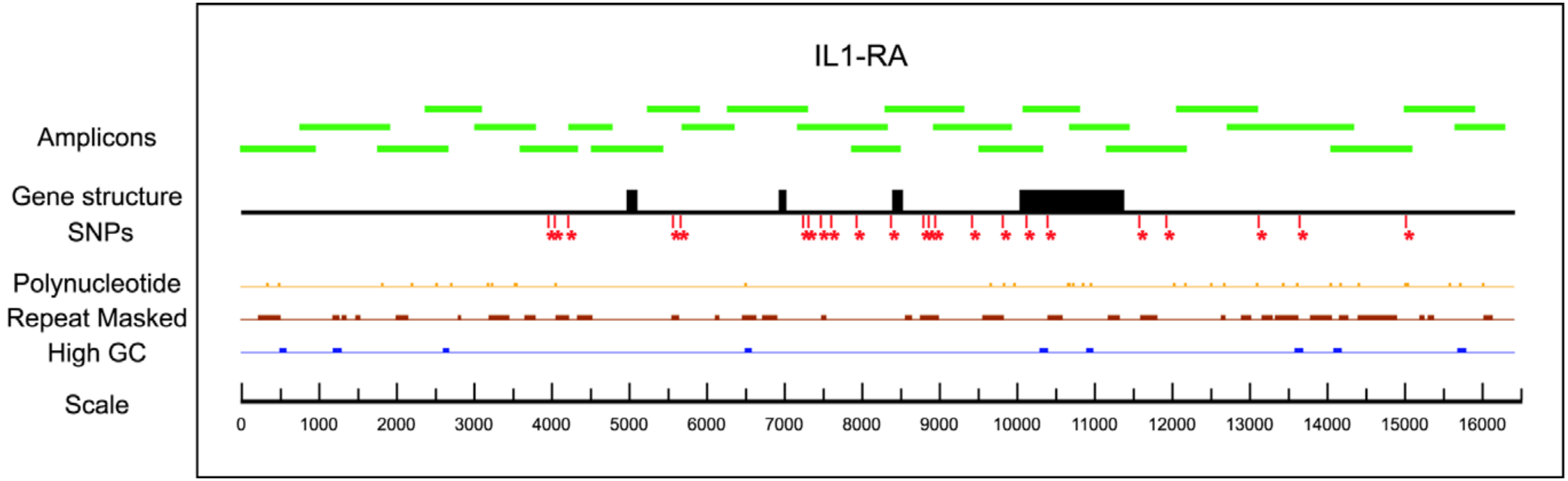
Current studies

1) Genetic discovery encompassing a panel of immune response loci.

   Sequence-based discovery of new genetic data from panels of DNAs. Includes resequencing genes with established sequences.

2) Sequence analysis of new genomic regions.

3) Application of genetic information to correlate genotypes with interesting immunological phenotypes.



1. Obtain sequence data from external databases
2. Analyze for the presence of certain data types
3. Use this information to identify primers
4. Generate sequence data from multiple individuals

Immune Response Gene Diversity Project: data accounting

-40 sequence traces for each of 40 individuals from 103 genes

164,800 chromatogram trace files from a total of 103 distinct subprojects.

Data was combined with microsatellite data from collaborating labs.

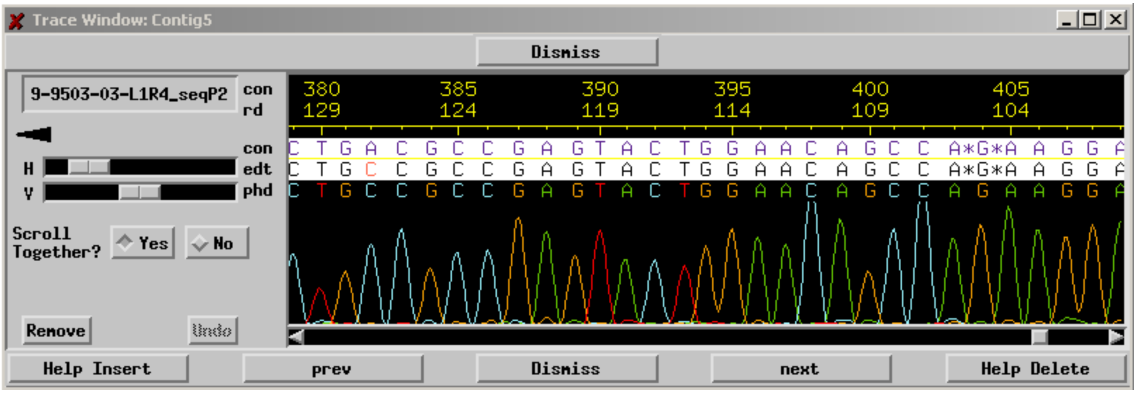
Data was used in derivative projects and submitted to public databases.

Group V Vertical collaboration user group	Petersdorf laboratory, FHCRC	Sequenced and Taqman based genotyping and correlations with marrow transplant outcome.
	Hansen, IHWG, FHCRC	Primer directed sequencing for genotyping in support of research activities. Taqman genotyping in support of disease mapping
	Geraghty laboratory, FHCRC	Primer directed sequencing, shotgun sequencing, STR, Taqman

Immune Response Gene Diversity Project: data accounting

-40 sequence traces for each of 40 individuals from 103 genes

164,800 chromatogram trace files from a total of 103 distinct subprojects.


Trace Window: Contig5
Dismiss
9-9503-03-L1R4_seqP2
con
rd
380 385 390 395 400 405
129 124 119 114 109 104
con C T G A C G C C G A G T A C T G G A A C A G C C A*G*A A G G A
edt C T G C C G C C G A G T A C T G G A A C A G C C A*G*A A G G A
phd C T G C C G C C G A G T A C T G G A A C A G C C A G A A G G A
H
V
Scroll Together?
Yes
No
Remove
Undo
Help Insert
prev
Dismiss
next
Help Delete

Genetic diversity and genomics of the immune response.

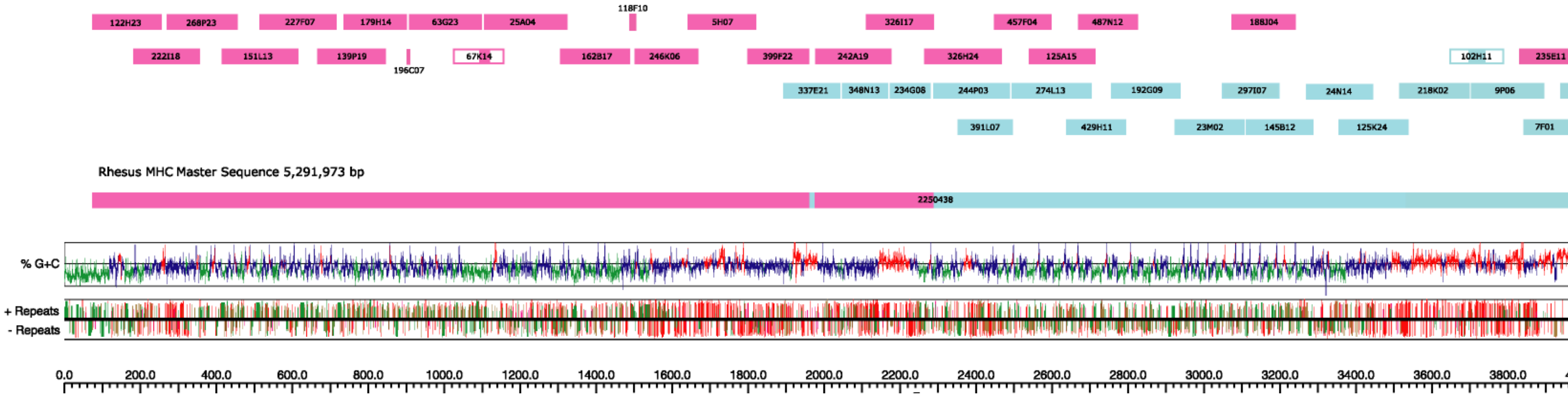
Current studies

1) Genetic discovery encompassing a panel of immune response loci.

2) Sequence analysis of new genomic regions.

   Includes sequencing of novel loci, utilizes shotgun sequencing.

3) Application of genetic information to correlate genotypes with interesting immunological phenotypes.



The Rhesus Macaque Major Histocompatibility Complex (MHC) Sequence analysis and comparison

The rhesus macaque MHC: data accounting

59 BACs each ~180,000 bp

- 10x redundancy shotgun sequencing, one BAC requires ~2,500 traces.

~150,000 chromatograms generated for this project

Subdivided into 59 subprojects of ~2,500 chromatograms per BAC

Work was carried out over 6 months and between two laboratories, one here at the FHCRC and a second across Lake Union at the Institute for Systems Biology.

Group H Horizontal collaboration user group	Lee Rowen, the Institute for Systems Biology	Collaborative projects in basic genomic studies. Shotgun sequencing.
	Geraghty laboratory, FHCRC	Primer directed sequencing, shotgun sequencing, STR.

Genetic diversity and genomics of the immune response.

Current studies

1) Genetic discovery encompassing a panel of immune response loci.

2) Sequence analysis of new genomic regions.

3) Application of genetic information to correlate genotypes with interesting immunological phenotypes.

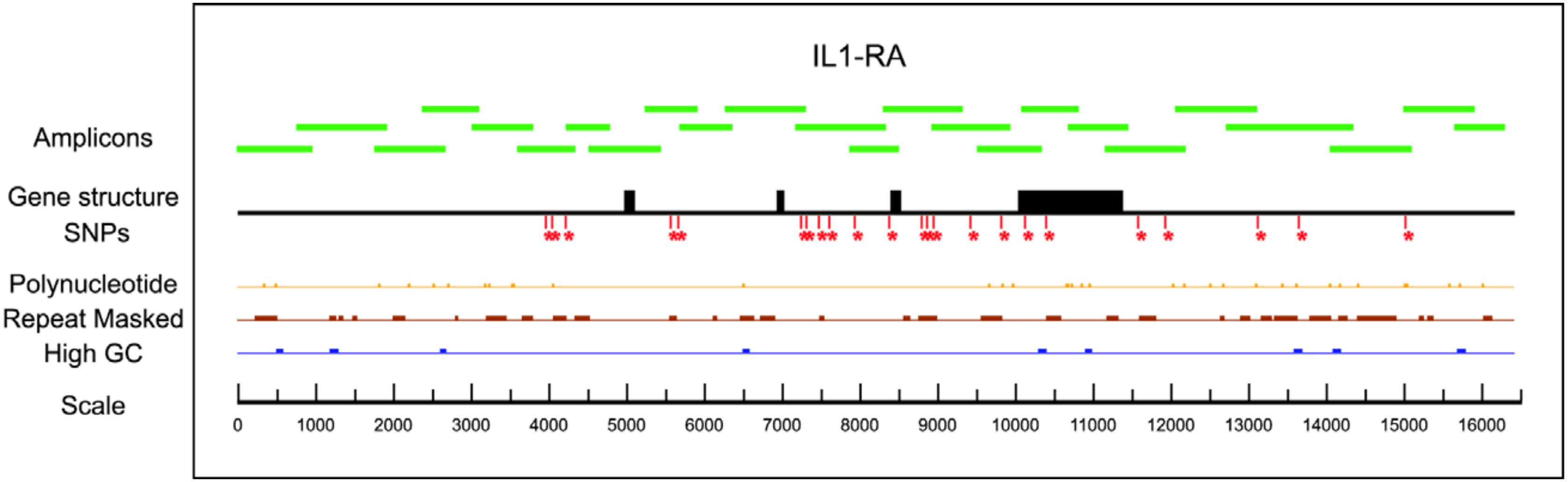
Sequencing used as a genotyping method. Involves examining several to hundreds of genotypes within 1,000s of individuals DNAs.

Relationship between Host Genomic Polymorphisms and Immune Reconstitution during Antiretroviral Therapy in AACTG Protocol A5001

PROPOSING INVESTIGATOR(S) AND INSTITUTION(S)
Janet Andersen, Sc.D.., SDAC
Constance Benson, M.D., University of Colorado
Richard D'Aquila, M.D., Vanderbilt University
Carl Dieffenbach, M.D., NIAID
Daniel E. Geraghty, Ph.D., Fred Hutchinson Cancer Research Center, Seattle
David W. Haas, M.D., Vanderbilt University
*Alan Landay, Ph.D., Rush Medical College (*Protocol Chair)
Derya Unutmaz, Ph.D., Vanderbilt University

STUDY RATIONALE

The central hypothesis of this study is that **polymorphisms in host genes that regulate T cell proliferation, survival, and/or programmed cell death** are associated with interindividual **variability in CD4^{+} T cell increases** during sustained control of plasma viremia by potent antiretroviral therapy.



1. Pick targeted amplicons
2. Generate sequence data from ~1,000 individuals
3. Interpret derived data
4. Export to collaborators for statistical analysis

ACTG data challenges

32 separate loci examined

1,000 individual DNAs

64,000 chromatogram traces generated over 6 weeks.

Data to be interpreted and delivered to several collaborators for a variety of data analyses.

Group E Export server collaboration user group	Watkins laboratory, University of Wisconsin	In support of basic research into vaccine efficacy using a primate model. Primer-based resequencing. Clinical typing lab – sequenced based mamu typing.
	Lakshmi K. Gaur , UW Primate center.	Sequencing and genotyping of class I and II in primates as a discovery resource.
	Geraghty laboratory, FHCRC	Primer directed sequencing, shotgun sequencing, STR.

Information Appliances: Targeted Support for High-throughput Laboratory Devices.

1. Outline the scientific questions and experimental methods being applied in our lab.

2. Explain the major experimental limitations of these methods.

3. Describe our local solution and its general applicability to other instruments, other sites and to data sharing.

Genetic diversity and genomics of the immune response.

The major experimental limitations.

1) Genetic discovery encompassing a panel of immune response loci.

   103 subprojects, 160,000 trace files, data analysis, submission of data to public databases.

2) Sequence analysis of uncharted genomic regions.

   59 subprojects, 150,000 trace files, 2 collaborating labs.

3) Application of genetic information to correlate genotypes with interesting immunological phenotypes.

   64,000 trace files, heterozygous data interpretation, >3 collaborating labs.

Information Appliances: Targeted Support for High-throughput Laboratory Devices.

The major experimental limitations

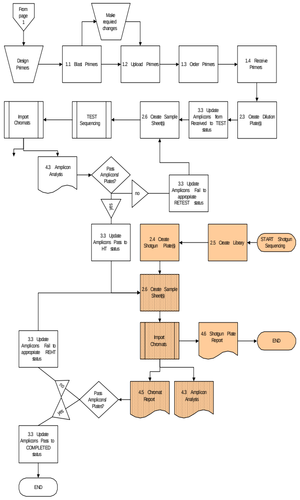
- Tracking laboratory throughput
- Organization of original data and meta data (machine, reagents, quality, etc.)
- Cost tracking, quality checking
- Data sharing

Creating a modular and extensible framework for future applications (STR, Taqman, HTR, etc.)

The major experimental limitations.

- Tracking laboratory throughput
- Organization of original data and meta data (machine, reagents, quality, etc.)
- Cost tracking, quality checks

- **Laboratory organization and data flow.**
  - Solid informatics infrastructure essential for data retrieval (i.e., a lab notebook).
  - Efficient data tracking improves data quality and lowers costs.



Targeted Sequencing

- Primer Design
- Primer Ordering / Storage

- Sample Sheet Creation
- ABI Sequencer Data Feed
- Sequencing Quality Reports
- Parameter Tracking Reporting

Shotgun Sequencing

- Create Library
- Manage Shotgun Plates

- Sample Sheet Creation
- ABI Sequencer Data Feed
- Sequencing Quality Reports
- Library Contamination Report

The major experimental limitations.

- Data sharing

- **Collaboration: the need to easily share data in a secure manner.**
  - Pooling genetic data from common samples (1).
  - Labs at different localities collaborating on a project requiring real time data sharing (2).
  - Pooling resources and expertise among labs in order to answer new questions (3).
  - Publishing and delivering data to public databases (1, 2, 3)

Creating a modular and extensible framework for future applications (STR, Taqman, HTR, etc.)

- **Data interpretation and analysis.**
  - Raw data made available for analysis as tools for analysis evolve.
  - New data types, new instruments.

Information Appliances: Targeted Support for High-throughput Laboratory Devices.

1. Outline the scientific questions and experimental methods being applied in our lab.

2. Explain the major experimental limitations of these methods.

3. Describe our local solution and its general applicability to other instruments, other sites, and to data sharing.

Although information appliances (IAs)* differ from each other as necessary to accomplish their individual tasks, nearly all IAs:

- are designed to support a specific activity, such as music, photography, or writing.
- combine powerful software applications with the ease of use of household appliances.
- are controlled by simple, intuitive user interfaces that require minimal training to use.
- can be used “out of the box”, without requiring complex configuration or set-up activities.
- connect to digital networks for the purpose of gathering or distributing information.
- manage data in standard formats and can share information easily with other similar systems.

***Norman, D. 1998. The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution. MIT Press. Cambridge, Mass.**

Small sequencing and genotyping laboratories need IT solutions to help them deal with their sequencing and genotype data. These labs need data management systems that:

- are designed specifically to support sequencing and genotyping projects.
- combine powerful software applications with the ease of use of household appliances.
- are controlled by simple, intuitive user interfaces that require minimal training to use.
- can be used “out of the box”, without requiring complex configuration or set-up activities.
- connect to digital networks for the purpose of gathering or distributing genetics information.
- manage data in standard formats and can share information easily with other similar systems.

Shared needs of the local (small) genetics data generator.

- Tracking laboratory throughput
- Organization of original data and meta data (machine, reagents, quality, etc.)
- Data sharing
- Cost tracking

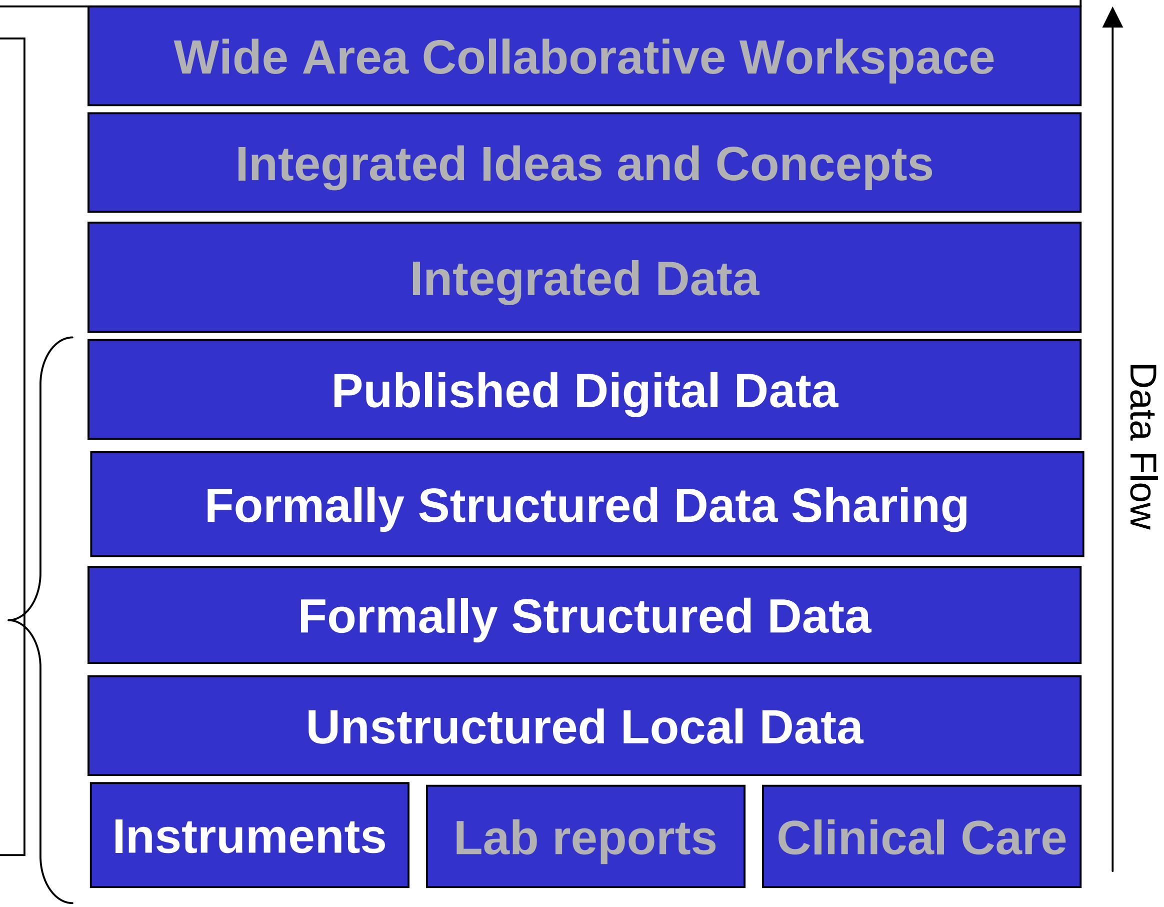
Creating a modular and extensible framework for future applications (STR, Taqman, HTR, etc.)

To address these needs we have built a *Ge*netics *M*anagement *S*oftware suite (*GeMS*).

Geraghty Lab: GeMS Approach

- Wide area data integration is seen as stack of activities

- Focus on bringing full power of high throughput DNA sequencing instruments into hands of small (R01-funded) laboratory



Wide Area Collaborative Workspace

Integrated Ideas and Concepts

Integrated Data

Published Digital Data

Formally Structured Data Sharing

Formally Structured Data

Unstructured Local Data

Instruments

Lab reports

Clinical Care

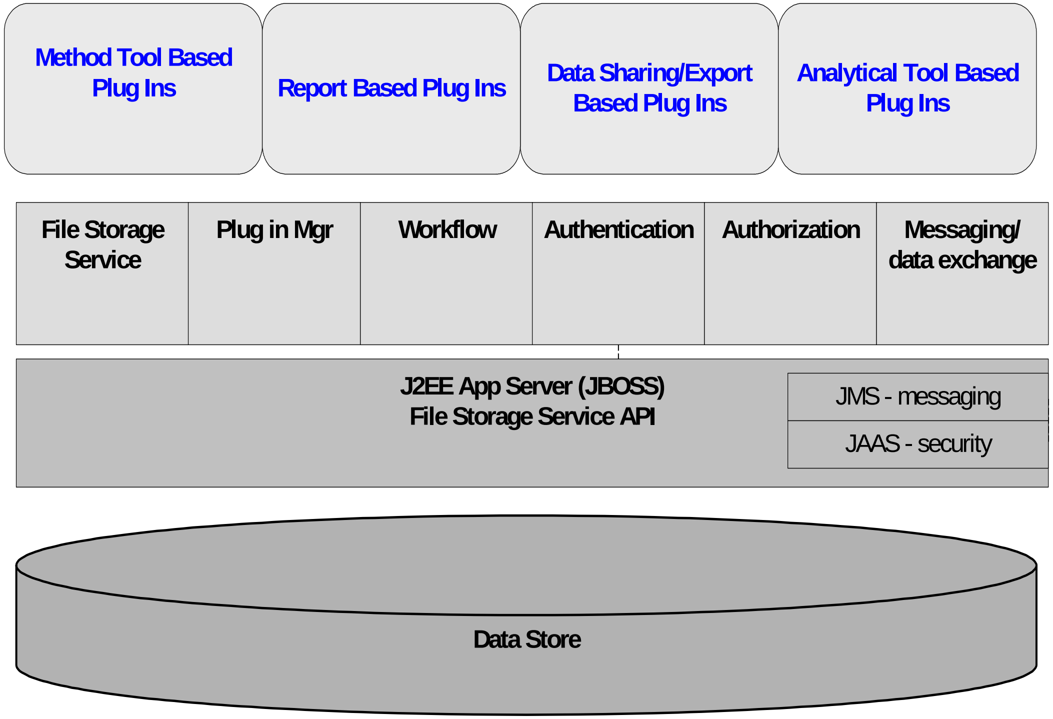
Data Flow

GeMS Architecture

GEMS/IA uses a modular N-tier approach, making it easier to implement and giving it the flexibility necessary

- The data store is accessed through a service API.
- Core services are made available using a J2EE framework. These services are used by the plugins to carry out their functions.
- Plugins represent the functional components that use the core services.

At the base, we have developed a Linux-based "turn-key server" to provide an easy to administer foundation. The GeMS-IA core consists of a PostgreSQL database, a J2EE/JBoss application server



- ***Estimates of ~ 5,000 small laboratory efforts in need of software support for sequencers in the U.S..***

Commercially available software in support of sequence analysis.

Applied Biosystem's SQL-GT and Sequence Collector: $250,000 for the whole set-up.

Scierra Laboratory workflow system: $145,000 plus 18% per year maintenance, plus unknown customization fees.

Geospiza finch server: Costs for various packages from $60,000 plus $24,000 per year to over $100,000 plus 28% of cost/year.

Quotes from survey respondents:

“After we obtain the raw sequence data, it is sent on to our users.”

“Traces are databased haphazardly by individuals.”

“As far as I know, there are no low-cost commercial sequence managers available.”

“We have also 'rolled our own' software here.”

“Unfortunately, there isn’t much out there to the best of my knowledge.”

“…download into Microsoft Access”

“There is an in-lab, home constructed, FileMaker Pro database of text files.”

GeMS-IA Technical Implementation
Open Source Components

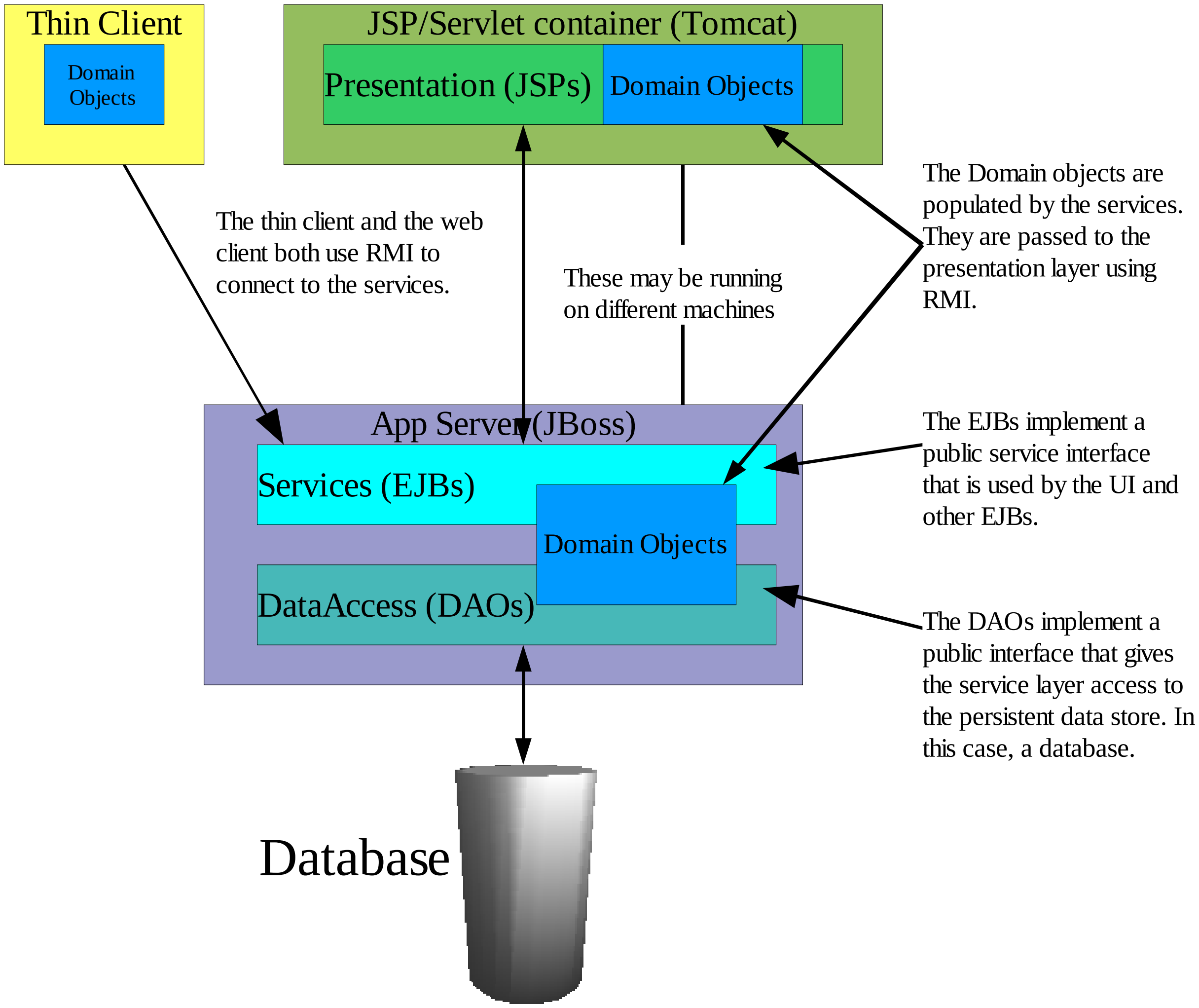
To maximize its cost-effectiveness and extensibility, GeMS/IA has been designed and implemented using open source systems and tools. Specifically,

Operating system:	Linux
System Admin Support:	WebMin
Database:	Postgres
Web server:	Tomcat
J2EE Server:	JBoss
Client Development:	Java

Currently GeMS-IA has 850 classes, and about 140,000 lines of code.

The database has 98 tables.

J2EE Application Layers



The thin client and the web client both use RMI to connect to the services.

These may be running on different machines

The Domain objects are populated by the services. They are passed to the presentation layer using RMI.

The EJBs implement a public service interface that is used by the UI and other EJBs.

The DAOs implement a public interface that gives the service layer access to the persistent data store. In this case, a database.

Current GeMS Application J2EE Application Layers Software Architecture.

GeMS core components

Authentication

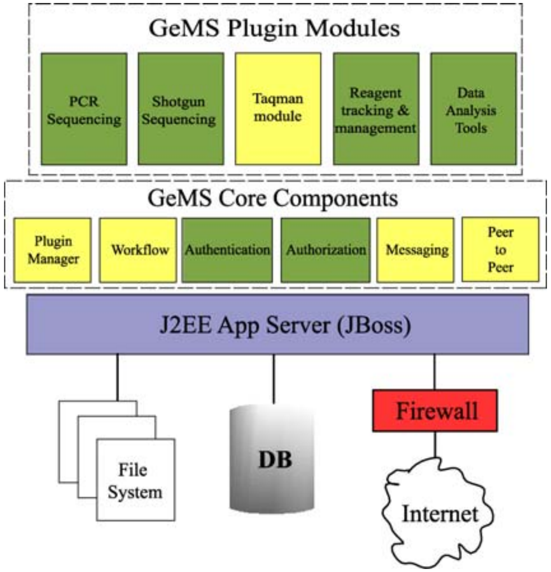
Authentication will be implemented using the J2EE Pluggable Authentication Module (PAM) mechanism.

Authorization

The security requirements of this project require much more flexibility than the standard user/group security model. The requirements specify that access control apply to individual data elements

GeMS-IA Messaging

The messaging component will allow users of the GeMS system to communicate easily and effectively. Users will be able to send and receive messages via email, secure file transfer, adding a message or URL to a web page, and by instant messaging. Recipients may be specified as an individual user or group of users.



GeMS core components

GeMS-IA Work Flow

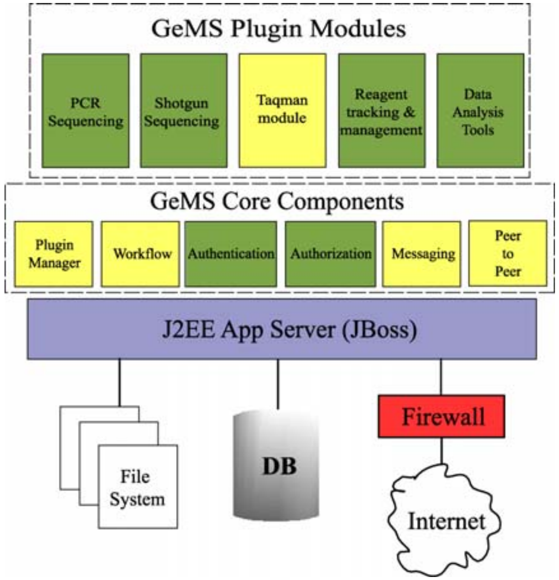
The work flow component will allow users of the GeMS system to collect a series of different tasks into a “work flow.” This will free up the user to perform other work since they will not have to monitor the system as each individual task is completed.

Plug-in management

Support for different protocols and analysis tools will be provided in pluggable modules. These modules are basically J2EE EARs (Enterprise Archives) that build upon the services provided by the platform.

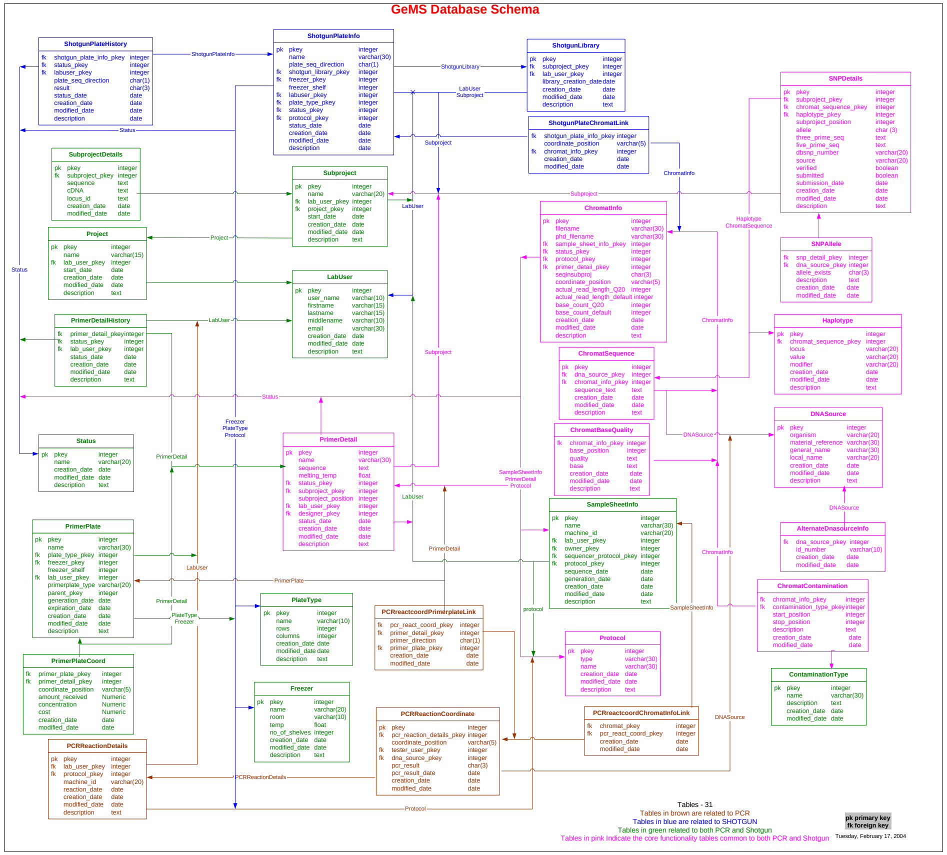
GeMS-IA Peer to peer

The peer to peer component will allow users of the GeMS-IA system to invoke services on other GeMS-IA instances that are set up as collaborators.



GeMS Data Schema

- Schema currently relates all key variables in automated high throughput DNA sequencing to the output files for data analysis, sharing and comparison including
  - DNA Source information
  - SNP Identification
  - Primers
  - amplicons
  - Haplotypes
  - Sequencers
  - Technicians
  - PCR Thermocyclers



GeMS Database Schema

	ShotgunPlateHistory	
fk	shotgun_plate_info_pkey	integer
fk	status_pkey	integer
fk	labuser_pkey	integer
	plate_seq_direction	char(1)
	result	char(3)
	status_date	date
	creation_date	date
	modified_date	date
	description	date

	ShotgunPlateInfo	
pk	pkey	integer
	name	varchar(30)
	plate_seq_direction	char(1)
fk	shotgun_library_pkey	integer
fk	freezer_pkey	integer
	freezer_shelf	integer
fk	labuser_pkey	integer
fk	plate_type_pkey	integer
fk	status_pkey	integer
fk	protocol_pkey	integer
	status_date	date
	creation_date	date
	modified_date	date
	description	date

	ShotgunLibrary	
pk	pkey	integer
fk	subproject_pkey	integer
fk	lab_user_pkey	integer
	library_creation_date	date
	creation_date	date
	modified_date	date
	description	text

	ShotgunPlateChromatLink	
fk	shotgun_plate_info_pkey	integer
	coordinate_position	varchar(5)
fk	chromat_info_pkey	integer
	creation_date	date
	modified_date	date

	SNPDetails	
pk	pkey	integer
fk	subproject_pkey	integer
fk	chromat_sequence_pkey	integer
fk	haplotype_pkey	integer
	subproject_position	integer
	allele	char (3)
	three_prime_seq	text
	five_prime_seq	text
	dbsnp_number	varchar(20)
	source	varchar(20)
	verified	boolean
	submitted	boolean
	submission_date	date
	creation_date	date
	modified_date	date
	description	text

	SubprojectDetails	
pk	pkey	integer
fk	subproject_pkey	integer
	sequence	text
	cDNA	text
	locus_id	text
	creation_date	date
	modified_date	date

	Subproject	
pk	pkey	integer
	name	varchar(20)
fk	lab_user_pkey	integer
fk	project_pkey	integer
	start_date	date
	creation_date	date
	modified_date	date
	description	text

	Project	
pk	pkey	integer
	name	varchar(15)
fk	lab_user_pkey	integer
	start_date	date
	creation_date	date
	modified_date	date
	description	text

	LabUser	
pk	pkey	integer
	user_name	varchar(10)
	firstname	varchar(15)
	lastname	varchar(15)
	middlename	varchar(10)
	email	varchar(30)
	creation_date	date
	modified_date	date
	description	text

	PrimerDetailHistory	
fk	primer_detail_pkey	integer
fk	status_pkey	integer
fk	lab_user_pkey	integer
	status_date	date
	creation_date	date
	modified_date	date
	description	text

	ChromatInfo	
pk	pkey	integer
	filename	varchar(30)
	phd_filename	varchar(30)
fk	sample_sheet_info_pkey	integer
fk	status_pkey	integer
fk	protocol_pkey	integer
fk	primer_detail_pkey	integer
	seqinsubproj	char(3)
	coordinate_position	varchar(5)
	actual_read_length_Q20	integer
	actual_read_length_default	integer
	base_count_Q20	integer
	base_count_default	integer
	creation_date	date
	modified_date	date
	description	text

	SNPAllele	
fk	snp_detail_pkey	integer
fk	dna_source_pkey	integer
	allele_exists	char(3)
	description	text
	creation_date	date
	modified_date	date

	Haplotype	
pk	pkey	integer
fk	chromat_sequence_pkey	integer
	locus	varchar(20)
	value	varchar(20)
	modifier	varchar(20)
	creation_date	date
	modified_date	date
	description	text

	ChromatSequence	
pk	pkey	integer
fk	dna_source_pkey	integer
fk	chromat_info_pkey	integer
	sequence_text	text
	creation_date	date
	modified_date	date
	description	text

	ChromatBaseQuality	
fk	chromat_info_pkey	integer
	base_position	integer
	quality	text
	base	text
	creation_date	date
	modified_date	date
	description	text

	DNASource	
pk	pkey	integer
	organism	varchar(20)
	material_reference	varchar(30)
	general_name	varchar(30)
	local_name	varchar(20)
	creation_date	date
	modified_date	date
	description	text

	Status	
pk	pkey	integer
	name	varchar(20)
	creation_date	date
	modified_date	date
	description	text

	PrimerDetail	
pk	pkey	integer
	name	varchar(30)
	sequence	text
	melting_temp	float
fk	status_pkey	integer
fk	subproject_pkey	integer
	subproject_position	integer
fk	lab_user_pkey	integer
fk	designer_pkey	integer
	status_date	date
	creation_date	date
	modified_date	date
	description	text

	SampleSheetInfo	
pk	pkey	integer
	name	varchar(30)
	machine_id	varchar(20)
fk	lab_user_pkey	integer
fk	owner_pkey	integer
fk	sequencer_protocol_pkey	integer
fk	protocol_pkey	integer
	sequence_date	date
	generation_date	date
	creation_date	date
	modified_date	date
	description	text

	AlternateDnasourceInfo	
fk	dna_source_pkey	integer
	id_number	varchar(10)
	creation_date	date
	modified_date	date

	PrimerPlate	
pk	pkey	integer
	name	varchar(30)
fk	plate_type_pkey	integer
fk	freezer_pkey	integer
	freezer_shelf	integer
fk	lab_user_pkey	integer
	primerplate_type	varchar(20)
	parent_pkey	integer
	generation_date	date
	expiration_date	date
	creation_date	date
	modified_date	date
	description	text

	PlateType	
pk	pkey	integer
	name	varchar(10)
	rows	integer
	columns	integer
	creation_date	date
	modified_date	date
	description	text

	PCRreactcoordPrimerplateLink	
fk	pcr_react_coord_pkey	integer
fk	primer_detail_pkey	integer
	primer_direction	char(1)
fk	primer_plate_pkey	integer
	creation_date	date
	modified_date	date

	ChromatContamination	
fk	chromat_info_pkey	integer
fk	contamination_type_pkey	integer
	start_position	integer
	stop_position	integer
	description	text
	creation_date	date
	modified_date	date

	Protocol	
pk	pkey	integer
	type	varchar(30)
	name	varchar(30)
	creation_date	date
	modified_date	date
	description	text

	PrimerPlateCoord	
fk	primer_plate_pkey	integer
fk	primer_detail_pkey	integer
	coordinate_position	varchar(5)
	amount_received	Numeric
	concentration	Numeric
	cost	Numeric
	creation_date	date
	modified_date	date

	Freezer	
pk	pkey	integer
	name	varchar(20)
	room	varchar(10)
	temp	float
	no_of_shelves	integer
	creation_date	date
	modified_date	date
	description	text

	PCRReactionCoordinate	
pk	pkey	integer
fk	pcr_reaction_details_pkey	integer
	coordinate_position	varchar(5)
fk	tester_user_pkey	integer
fk	dna_source_pkey	integer
	pcr_result	char(3)
	pcr_result_date	date
	creation_date	date
	modified_date	date

	PCRreactcoordChromatInfoLink	
fk	chromat_pkey	integer
fk	pcr_react_coord_pkey	integer
	creation_date	date
	modified_date	date

	ContaminationType	
pk	pkey	integer
	name	varchar(30)
	description	text
	creation_date	date
	modified_date	date

	PCRReactionDetails	
pk	pkey	integer
fk	lab_user_pkey	integer
fk	protocol_pkey	integer
	machine_id	varchar(20)
	reaction_date	date
	creation_date	date
	modified_date	date
	description	text

Tables - 31

Tables in brown are related to PCR

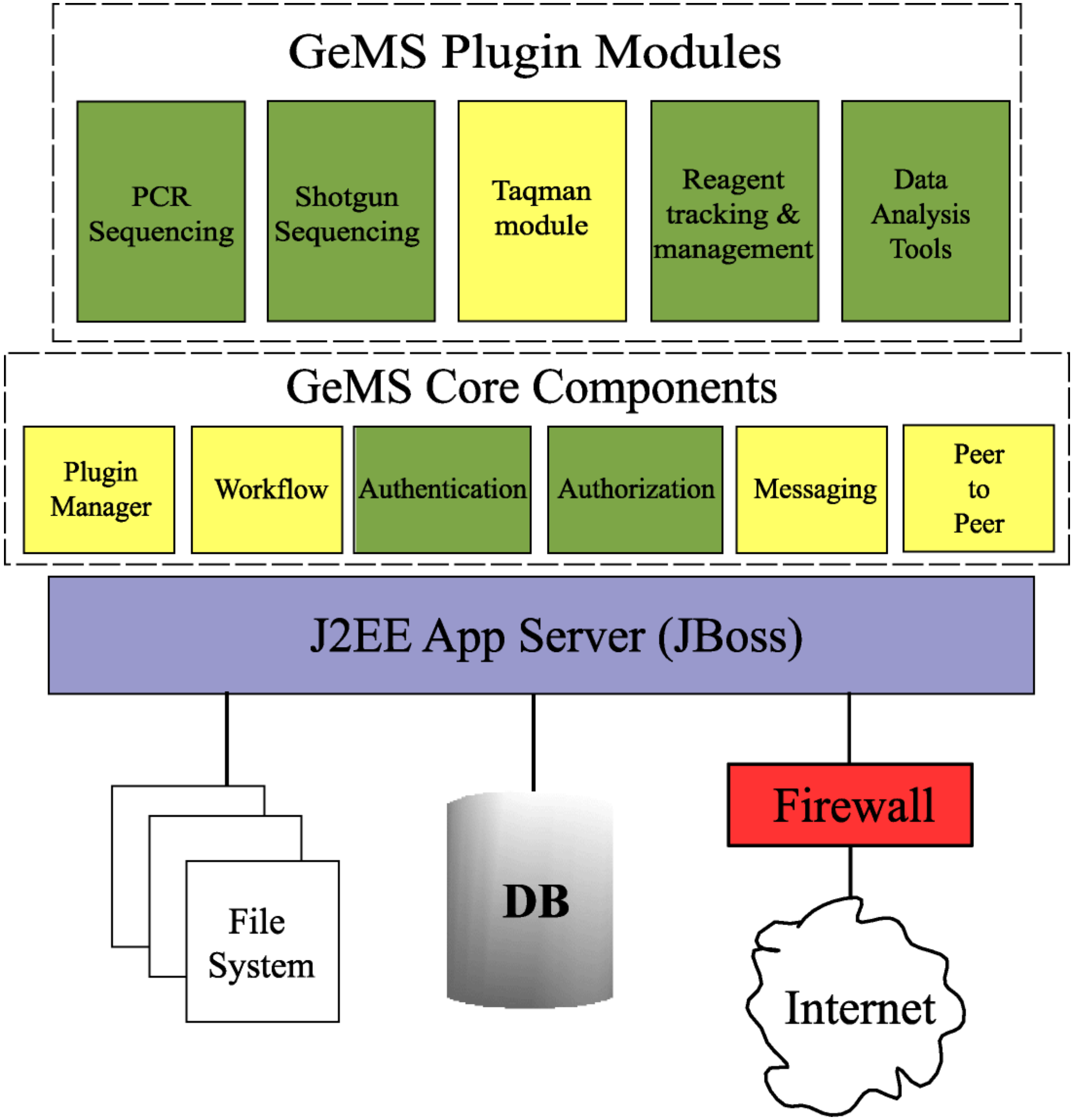
Tables in blue are related to SHOTGUN

Tables in green related to both PCR and Shotgun

Tables in pink Indicate the core functionality tables common to both PCR and Shotgun

pk primary key
fk foreign key

Tuesday, February 17, 2004

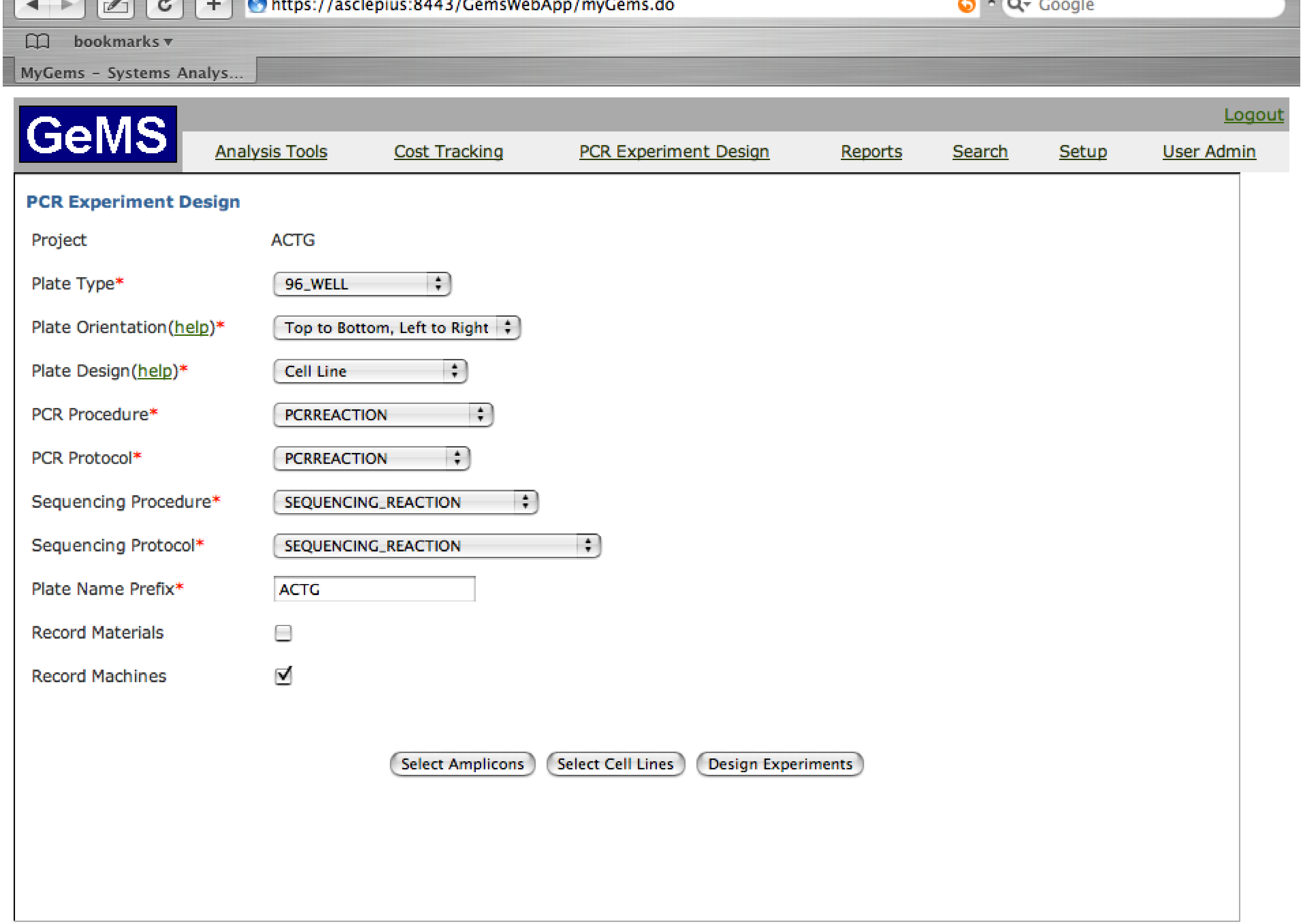

GeMS Plugin Modules
PCR Sequencing
Shotgun Sequencing
Taqman module
Reagent tracking & management
Data Analysis Tools
GeMS Core Components
Plugin Manager
Workflow
Authentication
Authorization
Messaging
Peer to Peer
J2EE App Server (JBoss)
File System
DB
Firewall
Internet

GeMS Plug-in modules

1) PCR sequencing and 2) Shotgun sequencing.
In the current GeMS, the PCR and Shotgun modules are packaged together in the GemsSequencing Module.
Chromatogram quality reporting
Sequence assembly reporting
Retrieval of chromats

3) GeMS sequence analysis tools
The programming is now broken down into four modular functions, with three directly used for primer design for PCR sequencing.
Assemble/View Chromats:
RepeatMasker utility:
Primer3 utility:
Blast Primer utility:

4) Cost accounting and reagent tracking.
The primary function of cost tracking is to determine the cost of running a particular protocol. Compares expected costs with inventory reports.



PCR Experiment Design

Project ACTG

Plate Type* 96_WELL

Plate Orientation(help)* Top to Bottom, Left to Right

Plate Design(help)* Cell Line

PCR Procedure* PCRREACTION

PCR Protocol* PCRREACTION

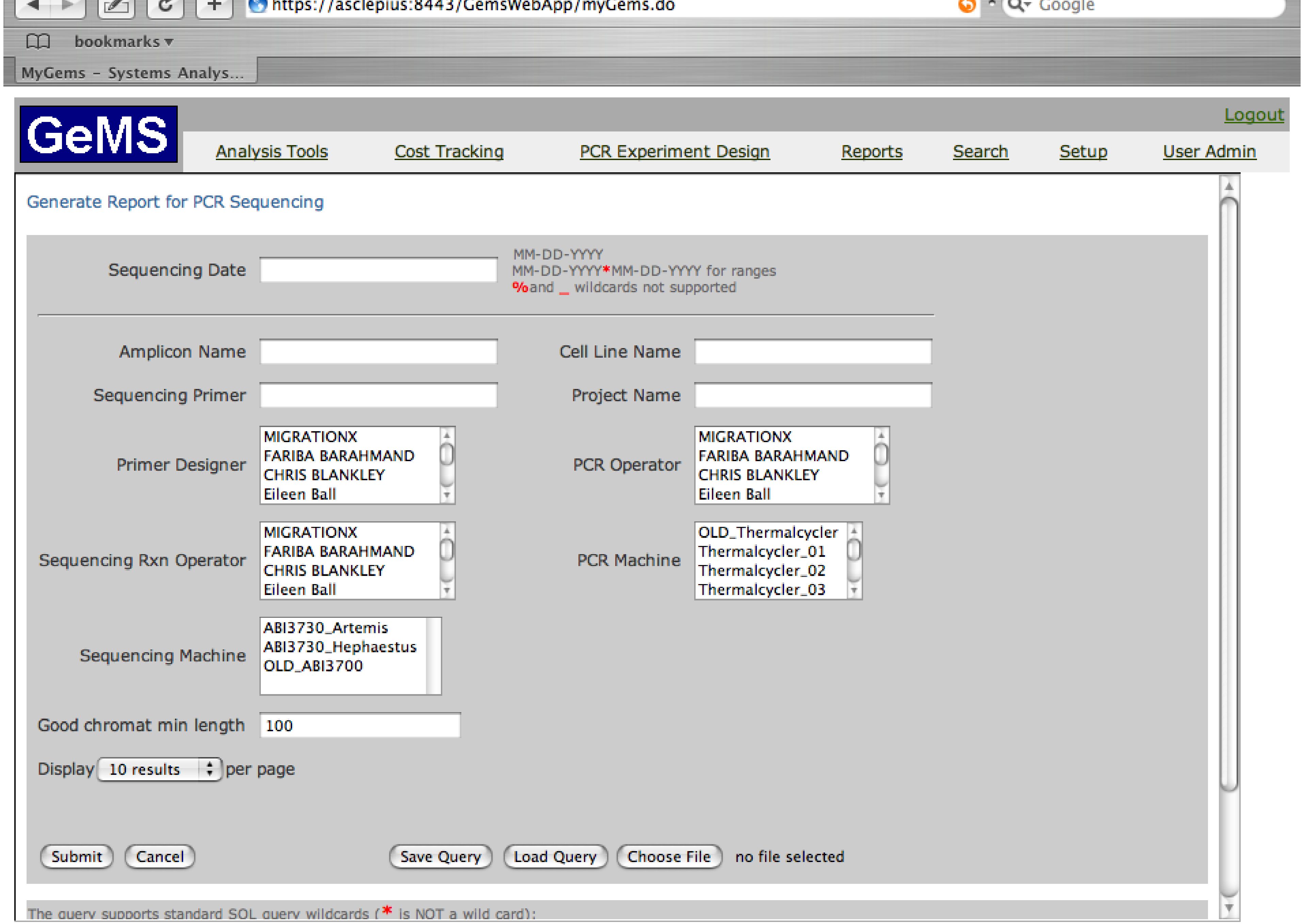
Sequencing Procedure* SEQUENCING_REACTION

Sequencing Protocol* SEQUENCING_REACTION

Plate Name Prefix* ACTG

Record Materials ☐

Record Machines ☑



Generate Report for PCR Sequencing

Sequencing Date

MM-DD-YYYY
MM-DD-YYYY*MM-DD-YYYY for ranges
%and _ wildcards not supported

Amplicon Name

Cell Line Name

Sequencing Primer

Project Name

Primer Designer: MIGRATIONX, FARIBA BARAHMAND, CHRIS BLANKLEY, Eileen Ball

PCR Operator: MIGRATIONX, FARIBA BARAHMAND, CHRIS BLANKLEY, Eileen Ball

Sequencing Rxn Operator: MIGRATIONX, FARIBA BARAHMAND, CHRIS BLANKLEY, Eileen Ball

PCR Machine: OLD_Thermalcycler, Thermalcycler_01, Thermalcycler_02, Thermalcycler_03

Sequencing Machine: ABI3730_Artemis, ABI3730_Hephaestus, OLD_ABI3700

Good chromat min length: 100

Display 10 results per page

no file selected

The query supports standard SQL query wildcards (* is NOT a wild card):

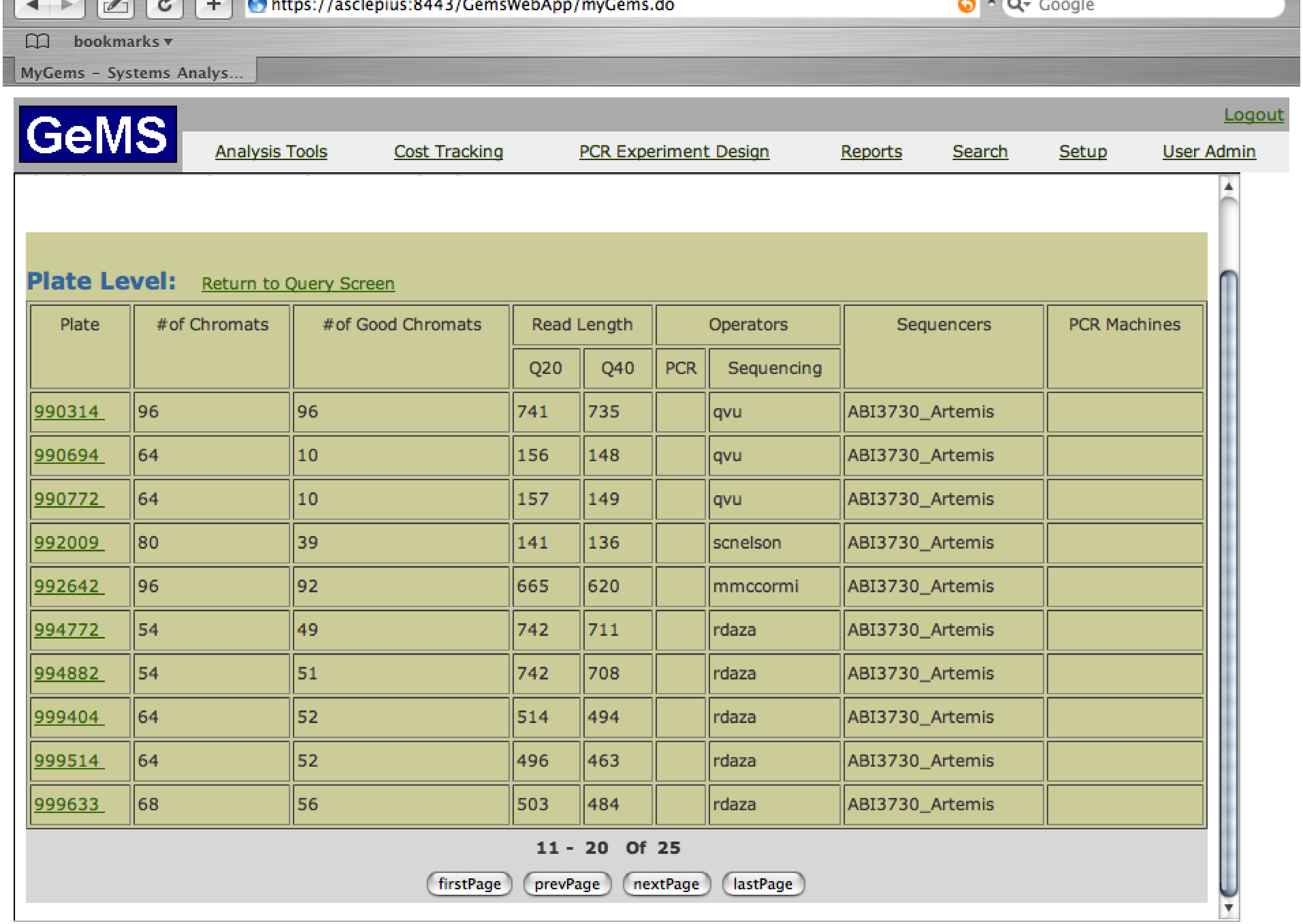


Plate Level: Return to Query Screen

Plate	#of Chromats	#of Good Chromats	Read Length		Operators		Sequencers	PCR Machines
			Q20	Q40	PCR	Sequencing		
990314	96	96	741	735		qvu	ABI3730_Artemis	
990694	64	10	156	148		qvu	ABI3730_Artemis	
990772	64	10	157	149		qvu	ABI3730_Artemis	
992009	80	39	141	136		scnelson	ABI3730_Artemis	
992642	96	92	665	620		mmccormi	ABI3730_Artemis	
994772	54	49	742	711		rdaza	ABI3730_Artemis	
994882	54	51	742	708		rdaza	ABI3730_Artemis	
999404	64	52	514	494		rdaza	ABI3730_Artemis	
999514	64	52	496	463		rdaza	ABI3730_Artemis	
999633	68	56	503	484		rdaza	ABI3730_Artemis	

11 - 20 Of 25

Immediate gains from the implementation of GeMS in the Geraghty lab.

Parameter	**Improvement**
Homologies Mapping	time reduced four fold (estimated 20 hours/year).
Primer Quality	eliminated design errors (start/end pos.) from 5% of all primers to none. reduced strand errors from 1% of all primers to none (combined estimated 100 hours/year including laboratory time saved).
Primer Ordering	automation saved one hour per plate (40 hours/year).
Sample Sheet Creation	automation saved 5 minutes per plate (200 hours/year).
PCR/Seq plate map creation	shows user which cell lines, primer(s), go in each well. Reduces user errors and save time setting up experiments (estimated 400 hours/year including laboratory time saved).
Chromatogram Quality Reports	saved 30 minutes per quality output summary (estimated 100 hours/year) eliminated naming errors – saved variable time depending on number and complexity of naming errors (estimated total 200 hours/year including laboratory time saved).
Data Organization	Able to easily group together chromats based on a list of criteria. (e.g. group all chromats from one cell line, or all chromats from one amplicon, etc.) Saved variable time and reagent cost checking quality criteria depending on the size of the project. (estimated $10,000 reagent costs and 100 hours/year laboratory time saved).
Managing assemblies	GeMS saves chromats in a centralized place and can dynamically create assemblies in any combination desired. This avoids data duplications and saves both space and analysis time (estimated 200 hours/year).
Cost Tracking	Improved from general estimation to detailed tracking that related work effort and reagent cost to a specific protocol being run over time (saved 20% reagent costs or $30,000 annualized).

Total 1360 hrs + $40,000 reagent costs/yr

Immediate gains from widespread implementation of GeMS.

•***Estimates of ~ 5,000 small laboratory efforts in need of software support for sequencers in the U.S..***

We are very much at the beginning of an emerging science that, for some time to come, will require the flexibility and innovation that only a multitude of investigator initiated research efforts can provide.

In this arena not only must small efforts individually use the technology effectively and efficiently, but they must also be in a position to take advantage of the collective value of data exchange and sharing.

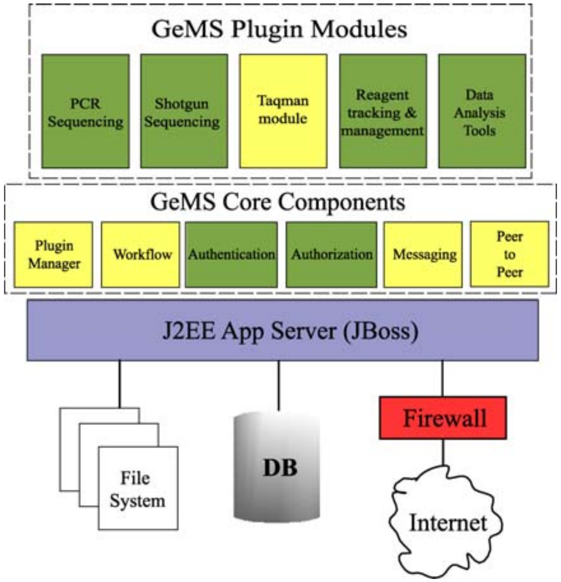
Data Analysis

An example of a plug-in module for GeMS:

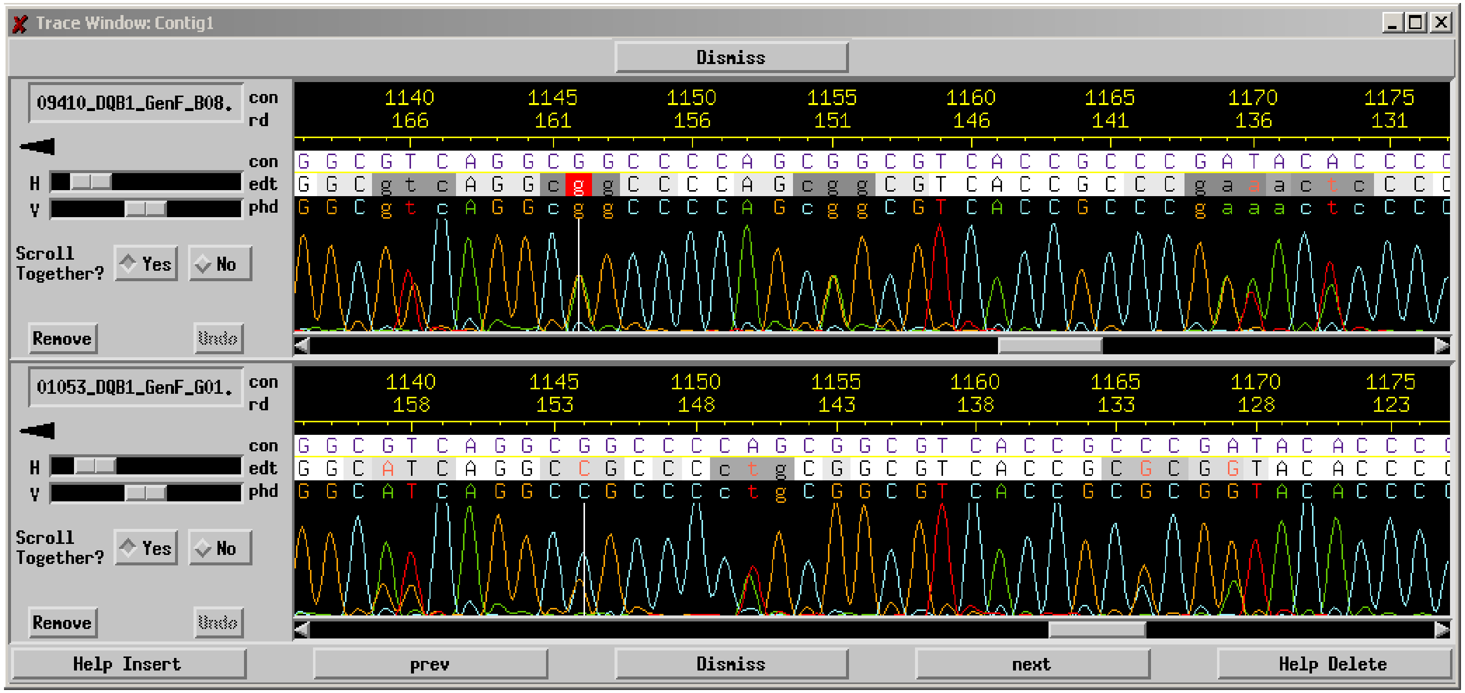
There is a substantial need for high throughput and *inexpensive* HLA typing.

The need arises from marrow transplants but extends now to a multitude of clinical networks and ongoing clinical trials.

Many methods for HLA typing are used, the gold standard being sequence-based typing.



The main limitation for sequence based typing is:



```
09410_DQB1_GenF_B08.
con   1140        1145        1150        1155        1160        1165        1170        1175
rd     166         161         156         151         146         141         136         131
con  G G C G T C A G G C G G C C C C A G C G G C G T C A C C G C C C G A T A C A C C C C
edt  G G C g t c A G G c g g C C C C A G c g g C G T C A C C G C C C g a a a c t c C C C
phd  G G C g t c A G G c g g C C C C A G c g g C G T C A C C G C C C g a a a c t c C C C

01053_DQB1_GenF_G01.
con   1140        1145        1150        1155        1160        1165        1170        1175
rd     158         153         148         143         138         133         128         123
con  G G C G T C A G G C G G C C C C A G C G G C G T C A C C G C C C G A T A C A C C C C
edt  G G C A T C A G G C C G C C C c t g C G G C G T C A C C G C G C G G T A C A C C C C
phd  G G C A T C A G G C C G C C C c t g C G G C G T C A C C G C G C G G T A C A C C C C
```

2 cell lines, 9 polymorphic positions,
1146 is trimorphic

Heterozygous Trace Resolution software (HTR)

- Interprets heterozygous DNA sequence data directly from the chromatogram without manual interpretation.
- Written in Java
- Current implementation does not have user interface.
- Undergoing upgrades to improve accuracy and to deliver data quality metrics.

ATCG

Program	Amps	CLs	Snp Conf	Snp Called	genotype errors	%genotype errors	FP Calls	FN Calls
HTR script	71	20	199	204	53	1.3	44	4
Polyphred 80	71	20	199	298	759	12.7	406	105
SoftGenetics 20_0	69	20	194	201	448	11.1	35	308

MHC

Program	Amps	CLs	Snp Conf	Snp Called	genotype errors	%genotype errors	FP Calls	FN Calls
HTR script	26	31	389	396	176	1.4	92	49
Polyphred 80	26	31	389	500	645	4.2	304	119
SoftGenetics 20_0	26	31	389	459	774	5.4	99	441

Low and medium complexity heterozygous data

DQB1 generic

	right	wrong	no call	right	wrong	no call
HTR script	182	4	6	95%	2%	3%
polyp 80	145	21	26	76%	11%	14%
softg_20_20	93	64	35	48%	33%	18%

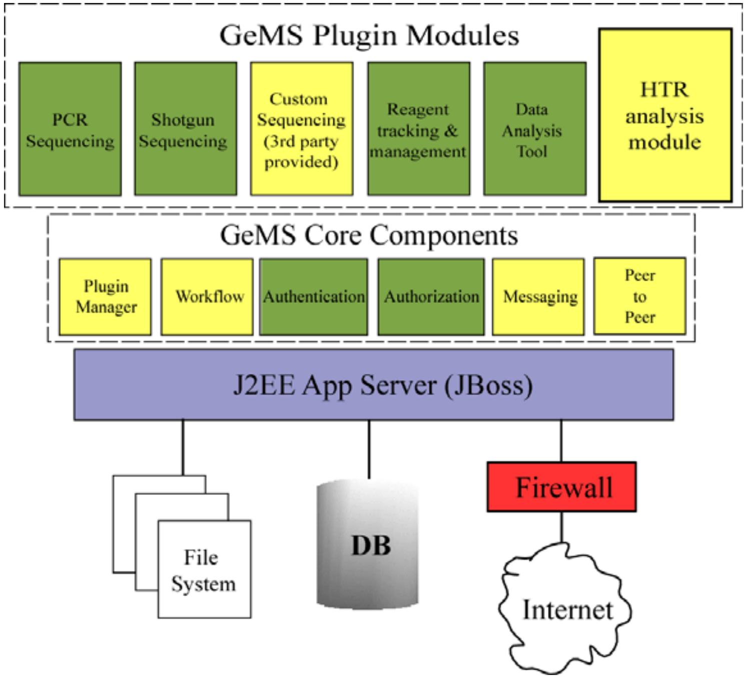
HTR script

	right	wrong	no call
DQB1	211	0	2
DPB1	57	0	0

High complexity heterozygous data HLA class II

Heterozygous Trace Resolution software:

An advanced plug-in for GeMS.



HTR software works best when provided with knowledge of

1) The method that produced the sequence data.
2) The machine on which the data was produced.
3) When the data was produced.
4) Details about primers used, amplicon sequence, etc.

All of these data are already part of the GeMS database, affording the use of considerable existing programming.

A major goal of GeMS is to improve researchers’ ability to share appliance-generated high-throughput biological data.

Using the Information Appliance model we can address a basic data sharing problem.

Collaboration among labs using different custom solutions can be difficult, as different data models require the pair wise development of custom data exchange systems.

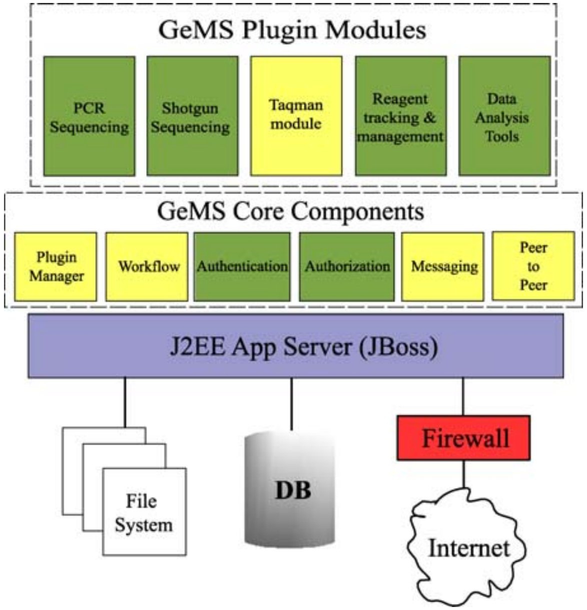
This is required in order to define and translate common data elements (CDEs), and can be the major expense of such an effort.

Simplifying data sharing is central to the information appliance model, because it implies user *adoption* of CDEs that automatically goes along with use of the information appliance.

We propose this solution is best suited to high throughput devices common to basic and clinical research.

Peer to Peer implementation:

- A clearly defined API that encapsulates all security, translation, and transport protocols so that application developers only need to understand the GeMS API that they want to remotely execute.
- This model allows for both data and computational sharing.
- Uses existing GeMS authentication and authorization security model.
- All communication is exchanged as compressed XML over HTTPS/SSL.



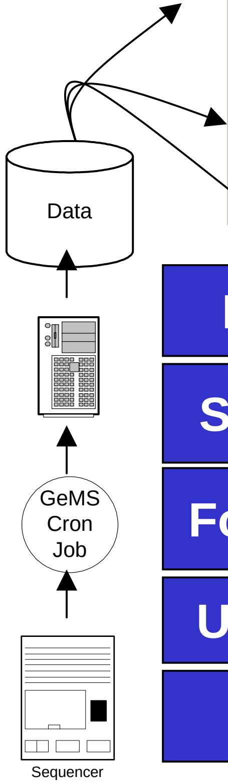
Group	**Lab**	**Interest**
Group V Vertical collaboration user group	Petersdorf laboratory, FHCRC	Sequenced and Taqman based genotyping and correlations with marrow transplant outcome.
	Hansen, IHWG, FHCRC	Primer directed sequencing for genotyping in support of research activities. Taqman genotyping in support of disease mapping
	Geraghty laboratory, FHCRC	Primer directed sequencing, shotgun sequencing, STR, Taqman
Group H Horizontal collaboration user group	Lee Rowen, the Institute for Systems Biology	Collaborative projects in basic genomic studies. Shotgun sequencing.
	Geraghty laboratory, FHCRC	Primer directed sequencing, shotgun sequencing, STR.
Group E Export server collaboration user group	Watkins laboratory, University of Wisconsin	In support of basic research into vaccine efficacy using a primate model. Clinical typing lab – sequenced based mamu typing.
	Lakshmi K. Gaur , UW Primate center.	Sequencing and genotyping of class I and II in primates as a discovery resource.
	Geraghty laboratory, FHCRC	Primer directed sequencing, shotgun sequencing, STR.

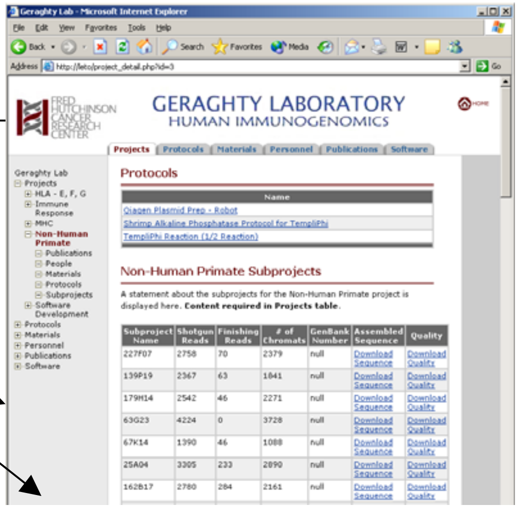
Future plans for GeMS-IA

1. To adapt the GeMS server engine towards wider applicability. Working with collaborators to test and improve GeMS.
2. To develop a GeMS operating client tool with the capability of allowing different labs to share data between and among themselves (peer to peer *plus*).
3. To use the tools developed in aims 1 and 2 as a framework for the establishment of software and database structures that will constitute an ‘export server’ capable of platform-independent data exchange and interoperability.

From Data Generation to Data Publication

- Nightly Data pick up by system
- Unstructured and unrelated data sent to GeMS server for processing
- Data related to associated parameters
- Subset of data made available to the Geraghty website




Geraghty Lab - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
http://leto/project_detail.php?id=3
Go
FRED HUTCHINSON CANCER RESEARCH CENTER
HOME

GERAGHTY LABORATORY
HUMAN IMMUNOGENOMICS

Projects | Protocols | Materials | Personnel | Publications | Software

Geraghty Lab
- Projects
  - HLA - E, F, G
  - Immune Response
  - MHC
  - **Non-Human Primate**
    - Publications
    - People
    - Materials
    - Protocols
    - Subprojects
  - Software Development
- Protocols
- Materials
- Personnel
- Publications
- Software

Protocols

Name
Qiagen Plasmid Prep - Robot
Shrimp Alkaline Phosphatase Protocol for TempliPhi
TempliPhi Reaction (1/2 Reaction)

Non-Human Primate Subprojects

A statement about the subprojects for the Non-Human Primate project is displayed here. **Content required in Projects table.**

Subproject Name	Shotgun Reads	Finishing Reads	# of Chromats	GenBank Number	Assembled Sequence	Quality
227F07	2758	70	2379	null	Download Sequence	Download Quality
139P19	2367	63	1041	null	Download Sequence	Download Quality
179H14	2542	46	2271	null	Download Sequence	Download Quality
63G23	4224	0	3728	null	Download Sequence	Download Quality
67K14	1390	46	1088	null	Download Sequence	Download Quality
25A04	3305	233	2890	null	Download Sequence	Download Quality
162B17	2780	284	2161	null	Download Sequence	Download Quality

Published Digital Data

Structured Data Sharing

Formally Structured Data

Unstructured Local Data

Instruments

Data Flow

more Future plans for GeMS-IA

1. To build a new module for an additional genetics data generating instrument (Taqman).

2. To create and maintain the ability to connect distributed installations supporting the two distinct types of genetics instruments (sequencers and Taqman).

    Many of the problems associated with data sharing between labs *simply disappear* if the labs employ common informatics systems and common data models.

3. To create and maintain an adaptation of the existing EDRN Research Network Exchange (ERNE using OODT) that will assist Import/export functions for distributed GeMS installations with other widely available databases containing genetics data.

Genetic diversity and genomics of the immune response.

Immune response genes
Quyen Vu
Skylar Nelson

GeMS software development
Lee Davis*
Mike McCormick*
Simon Fortelny*
Bethany Richards
Ruihan Wang*
Zhihong Zhang

HTR software package
Ruihan Wang*
Wade Smith*

Robert J. Robbins, Ph.D.
(VP for Administrative IT, FHCRC)

Mark Thornquist,Ph.D.
(Public Health Sciences Division, FHCRC), EDRN related initiative

Thomas Geraghty*, (COO, Immunogenomics Inc.) Off-site testing and requirements gathering