NCIBI Education and Outreach

NCIBI Graduate Students

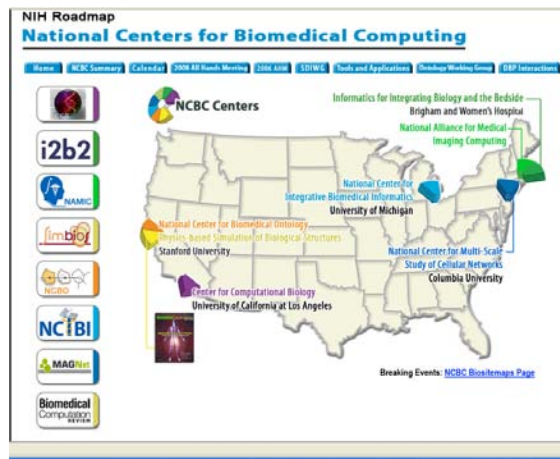| Student	Mentor	Project
Jing Zhang	H.V. Jagadish	Provenance of data in MiMI
Anna Shevardian	H.V. Jagadish	Interaction graphs to visualize and study protein interaction networks
Abhik Shah	Peter Woolf	Integrating existing knowledge and experimental data using Bayesian Networks
Sira Sarntivija	Brian Athey	Ontology Mapping; Cell Line Ontology
Arzucan Ozgur	Drago Radev	NLP Methods for information extraction
Nerit Glazer	Barbara Mirel	Evaluation and Assessment of User Heuristics
Gang Su	Fang Meng	Visualization and Data integration for protein-protein and metabolite interactions

NCBC Collaboration

- Big "P" in program

- 3 working groups across NCBCs
  - Ontologies
  - Yellow pages → **BIO**SITEMAPS
  - DBP Impact   http://www.ncbcs.org/biositemaps/

- www.ncbcs.org

Shared Events Calendar

NCIBI 2008-2009 Pilot Projects

- Open call for pilot projects, but with preference for comorbidity of substance abuse and bipolar disorder

- Selected collaborative project looking at Bipolar Disorder with co-morbid substance abuse among African American women

- Three subprojects:
  - *Genes for Bipolar Disorder among African Americans*, Sharon Lewis, Langston University
  - *Estrogen and Substance Abuse*, Tonya Gerald, North Carolina Central University
  - *Text mining*, Raphael Isokpehi, Wellington Ayensu, Jackson State University

NCIBI/RCMI Workshop on Translational Bioinformatics
July 29-30, 2009

- Provide RCMI investigators with training in the use of NCIBI tools and resources for translational research
- Build research relationships with the goal of expanding informatics resources for the study of health disparities
- Introductory hands-on tutorial with the NCIBI tools, followed by sessions tentatively focused on RTRN translational research clusters that overlap with projects in cancer, diabetes and nephropathy, and bipolar disorder
- Estimated number of attendees ~20
- Identify projects for coming year to expand outreach for NCIBI

Other Current and Future Collaborations

- Building Bridges Post-doc
  - Rich McEachin – Bridge with Laura Beirut Lab at WUSTL
  - Zuofeng Li (Hong Yu, UW Milwaukee) –

- I2B2/NCIBI – Tool integration with HIVE
- GenePattern (Broad Institute)
- Institute for Systems Biology (Seattle)

- Kay Robbins (UT San Antonio)
- Youping Deng (U. Southern MS), and Jack Yang (Harvard)

Tools and Technologies Seminars

Tools & Technology
Seminar Series

- Regular weekly seminars presenting novel development, algorithms and technologies.

- Broadcast over Adobe Connect for all to attend

- Video archived and available for review

http://portal.ncibi.org/gateway/tt.html
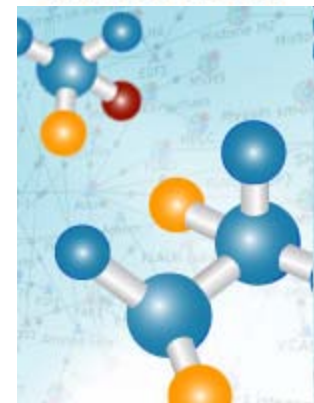
Meetings and Demonstrations

- ISMB 2008 – Toronto – Technology Track demonstration

- AMIA Translational Bioinformatics Summit

- Jackson State University - Demonstration

- Yesterday's NCIBI Workshop !

- Upcoming Tutorials with UM HSL

- Upcoming ISMB meeting (July 2009)

Dissemination of Software Tools

MiMI Web Statistics

Geographic Distribution of top 90% of visits
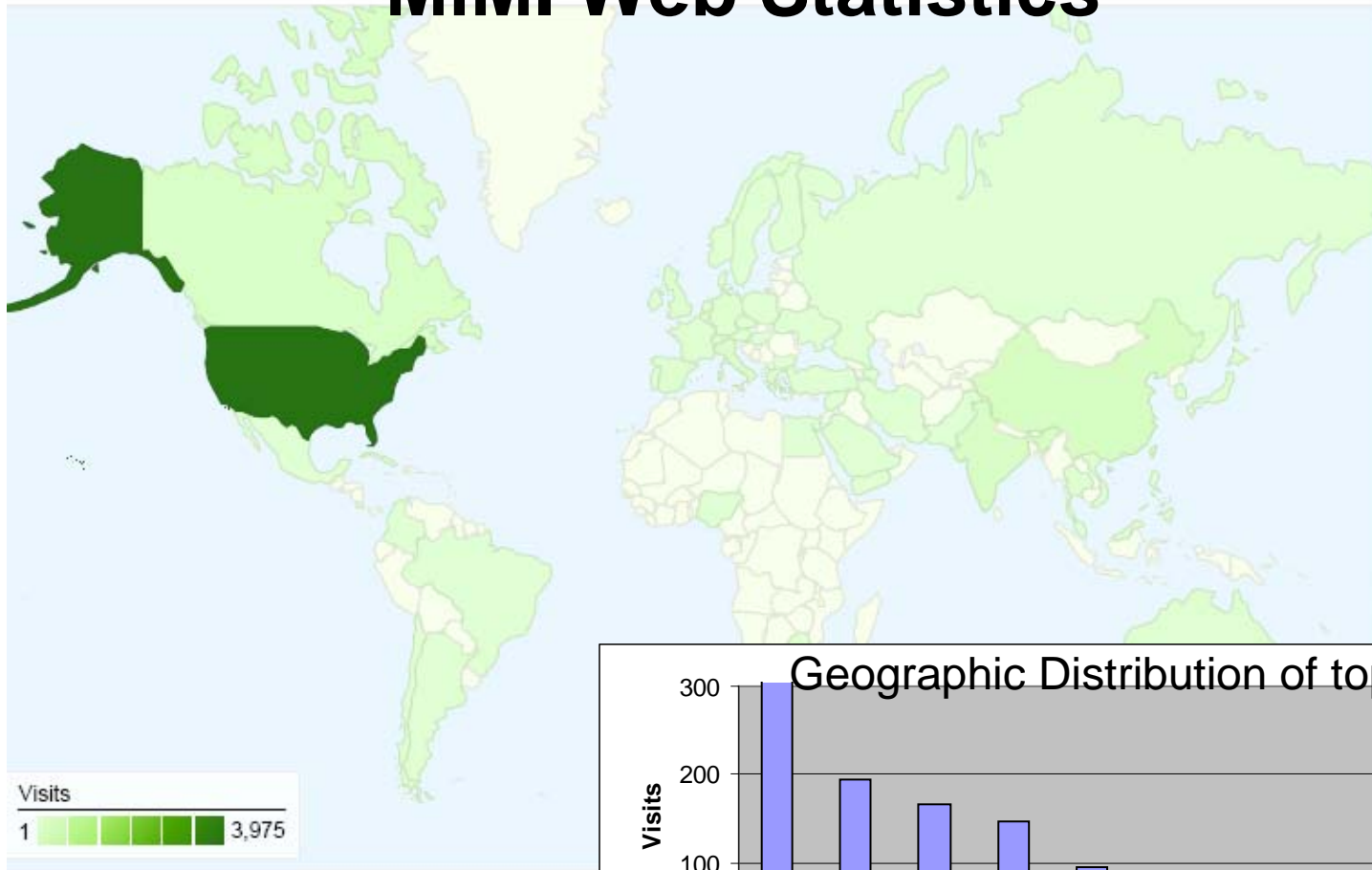
5320 total visits from 63 Countries
41% new visits
59% returning
~ 5.5 pages per visit
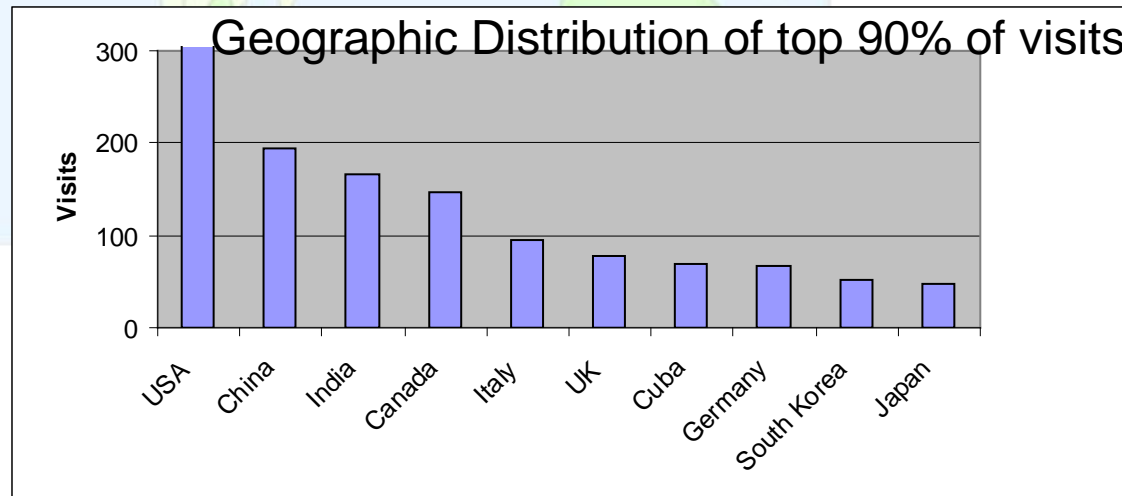
Tool Usage Statistics

	total hits	unique	returning	% unique	% returning
MiMI	6238	2521	3717	40	60
Gene2Mesh	2591	1468	1123	57	43
Metab2Mesh	4501	4080	421	91	9
MiSearch	1572	872	700	55	45
Biosearch2D	376	230	146	61	39

% Total tool hits

- biosearch2D
- Gene2Mesh
- Metab2Mesh
- MiMI
- MiSearch

% Returning tool hits

- biosearch2D
- Gene2Mesh
- Metab2Mesh
- MiMI
- MiSearch
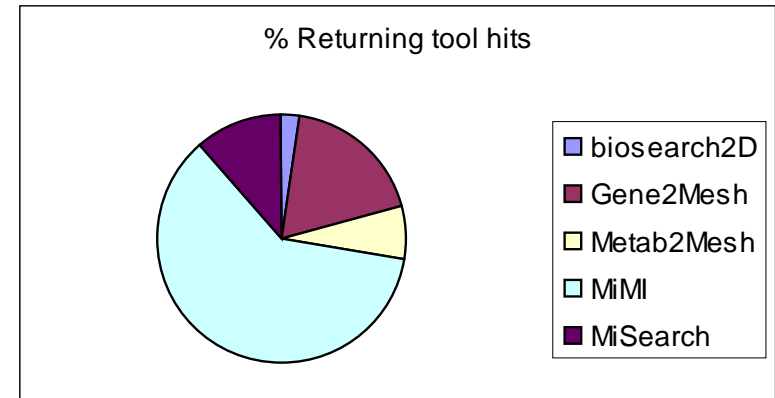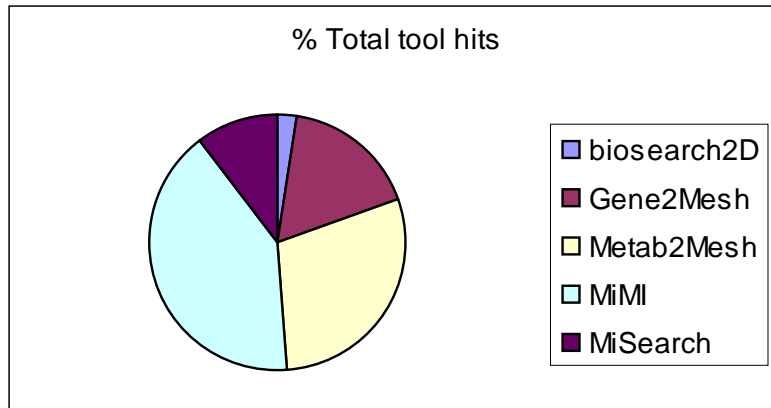
Cytoscape Plugin Downloads

520　　　　828　　　　890　　　458　　　　　　　　　　　　1017

| 2.0 | 2.1 | 2.2 | 3.0 | 3.01 |

8/31/07　　　10/8/07　　　　4/14/08　　　7/1/08　 8/12/08

present

MiMI Plugin

159

0.8

3/15/09

MetScape Plugin

NCIBI

NATIONAL INSTITUTES OF HEALTH
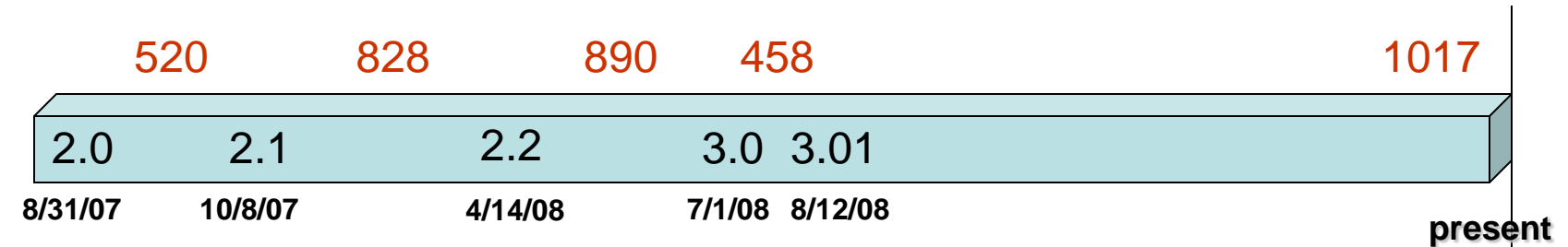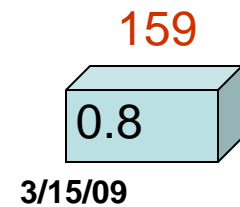
MiMI Web and Plugin Compared

- Two main routes of entry to get at MiMI data
    - Plug-in via Cytoscape
    - MiMI Website

	Total Visits			Unique Visits			Returning Visits per IP address	
	2008	2009	fold change	2008	2009	fold change	2008	2009
MiMI Web	3990	6238	2	370	2521	7	11	2
MiMI Plugin	4144	8515	2	435	742	2	10	11

NCIBI Core 4

- Role of Core 4 is twofold
  - Provide reliable, consistent and high quality IT infrastructure for all NCIBI projects and collaborations
  - Explore and implement cutting edge computational capacity and technology for research

- Computer Infrastructure
  - Compute intense (RAM and CPU) servers
  - Virtual Servers
  - Storage
  - Relational Databases
  - Backup / Recovery

Network Overview

Green Court

Palmer Commons

MACC

Site-to-Site VPN

Public Network

Lyons Sequencing Core

Private Network

Virtual Servers

Storage

Virtual Servers

Compute Servers

Storage

Compute Servers

Database Servers

Storage

Storage

Storage

LTO4 Backup

Offsite Backup

Combined Storage: 120TB RAW
Compute Servers: 3
Database Servers: 7
Virtual Servers: 15

NCiBI

Fourth Annual National Center for Integrative Biomedical Informatics Meeting 2009

NATIONAL INSTITUTES OF HEALTH

Compute Servers

- Server / Compute cycles
  - Servers with large RAM provide the bulk of CPU resources for Memory intensive NLP projects
  - Compute servers for databases
    - Standard server configuration: dual quad-core + 64 GB RAM per server (we started with 8, then 32)

- DC-NCIBI Cluster
  - 152 Opteron cores on 32 computing nodes
  - 40T Unified cluster storage
  - 12T Oracle Grid with 12 Opteron cores and 64 G
  - Time machine backup of critical files
  - Green computing: computing nodes turned off when not used

- GPU servers
  - Using Graphics Processing Units for algorithms amenable to Vector calculation
  - Enormous speed up from parallel pipe architecture of the GPU.

NCIBI

Virtual Servers and Storage

- ## Virtual Servers
  - Servers to support less intense needs; Web servers
  - Reduces energy / cooling costs by sharing hardware needs
  - Easily reconfigured and adapted to changing needs without additional purchase costs

- ## Storage
  - Current storage is built on 'fast' and 'slow' disks
  - Expansion of storage is a regular and essential part of NCIBI
  - Current capacity is ~ 120 TB (raw) (we started with 24)
  - Configured for 2 site mirroring of critical data (first level backup)
  - Tape storage for offline backup of selected data

Databases and Backup

- Relational Databases
  - Databases are essential core function for NCIBI
  - SQLServer and Oracle are primary RDBMS
    - Prototype, Test and Production
  - Tuning and security assessment and performance testing

- Backup Recovery Planning
  - Servers setup for snapshot recovery (immediate retrieval of deleted files)
  - RAID 5, 6, 10 disk arrays (protects against HD failure)
  - Battery backup, monitoring for status, temperature ** (power loss, AC loss control)
  - Offsite Mirror of critical data (protects against single site disaster (fire, sprinkler, etc)
  - Tape Backup for selected data files (long term recovery / restoration)

Data Currency

- Data Warehouse scripts automatically download over 20 biology/bioinformatics data sources totaling over 1.5 terabytes.
  - available to all of the compute servers and therefore all of the NCIBI researchers.

- Nightly Download of PubMed
  - Automatic replication of the nightly downloaded and weekly parsed PubMed data.

Thank you !

- Jerome Kinlaw
- LihShwu Ke
- Vasile Negrea
- Bruce Taylor
- Kurt Seyfried
- Aaron Bookvich
- Zach Wright
- Jim Rees
- Alex Terzian