

ApiDB: User Studies and Impact on Development

Eileen Kraemer

Steve Fischer

UGA

U. Penn

eileen@cs.uga.edu

stevef@pcbi.upenn.edu

ApiDB Bioinformatics Resource Center

ApiDB.org

- ApiDB: **A Bioinformatics Resource Center** for Biodefense and Emerging/Re-emerging Infectious Diseases.
 - ApiDB.org -- umbrella /integrated site for apicomplexan parasites
 - CryptoDB.org --
 - *Cryptosporidium*; causes cryptosporidiosis
 - PlasmoDB.org
 - *Plasmodium* species; causes malaria
 - ToxoDB.org
 - *Toxoplasma*; causes toxoplasmosis
-

Outline

- About the workshop
 - Surveys
 - Interviews
 - Video capture of lab exercise sessions
 - Card-sorting analysis of query categories
 - Tour
 - Conclusions
-

About the workshop

- June 26th - 29th, 2006 at UGA
 - approx. 30 participants from global audience
 - selected on the basis of their need to become an expert user of one or more of the ApiDB database resources (CryptoDB, ToxoDB, PlasmoDB).
 - introduction to apicomplexan database resources that are part of the ApIDB.org Bioinformatics Resource Center (PlasmoDB, ToxoDB, CryptoDB and ApiDots).
 - “tours” of databases and tools, hands-on exercise sessions
 - **User surveys, studies, exercises**
-

Pre-workshop survey

- Online, anonymous survey at start of first session.
 - What do users know? What kinds of adaptations do we need to make?
 - Participants presented with a list of 147 terms related to biological databases and the analysis of biological data.
 - For each term, the participants asked to rank their familiarity as:
 - not at all familiar
 - heard of it
 - slightly familiar
 - very familiar
-

Pre-workshop survey – Results

- For only 21 of the 147 terms (14.3%) were 50% or more of the participants able to claim that they were “very familiar”.
 - Terms:
 - AA sequence
 - Annotated
 - BLAST
 - cDNA
 - Chromosome
 - Coil
 - Contig
 - Daltons,
 - GC number
 - Gametocyte
 - GenBank
 - gene
 - Helix
 - Locus
 - Mitochondrion
 - NCBI
 - Oligo
 - Promoter
 - sequence similarity
 - Strand
 - translation
-

Pre-workshop survey – Results

- For 31 of the 147 terms (21%) 50% or more of the participants responded that they were “not at all familiar”. These terms were:
 - **Affymetrix Genotyped SNP probes**
 - **annotation density**
 - **ApiDoTS alignments**
 - **Boolean intersect**
 - **Boolean join**
 - **Boolean subtract**
 - **CryptoCyc Metabolic Pathway**
 - **Eimeria Gene Models**
 - **e-PCR data**
 - **expression timing**
 - **FullPhat**
 - **Genes by Volatility (Mutability)**
 - **GenPept protein**
 - **GLEAN gene**
 - **GO component**
 - **GO process**
 - **MR4 reagents**
-

Unfamiliar terms, cont'd

- ❑ ncRNA
 - ❑ Optical maps
 - ❑ Profile
 - ❑ OrthoMCL
 - ❑ PATS
 - ❑ PlasMit
 - ❑ PlasmoA
 - ❑ ProDom
 - ❑ Pubcrawler
 - ❑ Refseq
 - ❑ *T. gondii* UniGene EST alignments
 - ❑ TigrScan Gene
 - ❑ TwinScan Gene Models
 - ❑ unigene
-

Pre-workshop survey – Impact

- Gave developers a clearer idea of the types of expertise possessed by users
 - Revisions to site/help text
 - Dropped “Boolean”, annotation density terms
 - Allowed workshop presenters to adjust details/level of aspects of workshop presentations
-



My Query History

Release 5.2

Home Queries & Tools

My Query History
(4 queries)

Data Sources Downloads

About PlasmoDB Support

Tutorials

Upcoming features...

Gene query history

ID	Query	Size
4 <input type="button" value="Rename"/>	GO Term	1142 orthologs view download revise delete
3 <input type="button" value="Rename"/>	Expression Timing (P.f)	340 orthologs view download revise delete
2 <input type="button" value="Rename"/>	Expression Timing (P.f)	1 orthologs view download revise delete
1 <input type="button" value="Rename"/>	Snp Characteristics	1608 orthologs view download revise delete

Removed "Boolean" terminology

Email:

Password:

[forgot?](#)

Combine results: [eg: 1 or ((4 and 3) not 2)]

Be careful: This will delete all your queries on PlasmoDB.

Understanding AND, OR and NOT:

- 1 and 2** Genes that 1 and 2 have in common. You can also use "1 intersect 2".
- 1 or 2** Genes present in 1 or 2, or both. You can also use "1 union 2".
- 1 not 2** Genes in 1 but not in 2. You can also use "1 minus 2".

* If you want to delete a query, you must first delete all other boolean queries that uses this one as a component.



Tutorials

Release 5.2

Home

Queries &
ToolsMy Query History
(no queries)Data
Sources

Downloads

About
PlasmoDB

Support

Tutorials

Upcoming features...

Email: Password:

forgot?

Login

Register or Subscribe

Did you know ?

Did you know [you can download a file containing the mapping of former Gene IDs to new Gene IDs?](#)

[Read all the 'FAQs'](#)

Featured data source

[Transcriptome: TIGR Gene Indices for P. berghei](#)

[See all data sources](#)



Introduction [view](#)

Duration: 2:31 minutes Size: 12.7 mb
Introduction to the PlasmoDB functionality.

Queries and Tools [view](#)

Duration: 3:07 minutes Size: 25.3 mb
Introduction to the PlasmoDB queries and tools.

Query History [view](#)

Duration: 2:27 minutes Size: 5.6 mb
Using the query history to combine queries.

Candidate Vaccine Targets Query [view](#)

Duration: 3:03 minutes Size: 24.8 mb
Query to identify genes that are potential vaccine targets. Query includes secreted proteins, phylogenetic profile and expression profile queries.

Candidate Vaccine Targets Query With SNPs [view](#)

Duration: 5:21 minutes Size: 41.3 mb
Query to identify genes that are potential vaccine targets. Query includes secreted proteins, phylogenetic profile, SNPs and expression profile queries.

Candidate Drug Targets Query [view](#)

Duration: 2:56 minutes Size: 16.8 mb
Query to identify genes that are candidate drug targets. Query includes expression profile, proteomics evidence and GO function queries.

Candidate Drug Targets Query With SNPs [view](#)

Duration: 5:21 minutes Size: 37 mb

Individual interviews

- Conducted with six of the participants
 - **Have you used these databases before?**
 - **Problems? Explain.**
 - **Feature requests?**
 - **Elicit "usage scenarios"**
 - **Anything else we should know?**
-

Individual Interviews - Results

- Output formatting:
 - Don't want all the columns that are returned
 - ✓ column configuration in progress
 - More descriptions / better explanations
 - ✓ Tutorial movie clips
 - Consistency with other sources:
 - data not synchronized with GeneDB and GenBank, leads to problems with analysis.
 - ?
-

Individual Interviews - Results

- Naming problems:
 - Gene synonymy; mapping of gene names & products to common names
 - Frequent changes, occasional broken links frustrating.
 - Want more explanation of how orthologs are determined and clearer descriptions of relationships between gene and orthologs
-

Individual Interviews: Desired Features

- Ability to save queries
 - ✓ In progress ..
 - Support for local and high-throughput analysis: Web Services, XML format data format, etc.
 - Some elements available now/in progress
 - simpler-to-parse format coming; XML possibly on the horizon
 - Ability to load a list of gene-ids, and then be able to just click “next” or choose from a list, right on the gene page.
-

Individual Interviews: Desired Features

- map view, as at NCBI.
 - A codon usage table, as at kazusa.jp
 - Link added
 - Better support for finding small pieces and identifying the location of hits.
-

Video capture of lab exercise sessions

- Video/audio capture during hands-on exercise sessions (with participant consent).
 - Participants asked to work in pairs; person at keyboard used a microphone headset.
 - “SnapZPro” was used for video/audio capture.
 - (Camptazia Studio / Wink)
 - Video segments were both burned to CD and placed on a secure web site, reviewed, text annotations added.
-

Video capture of lab exercise sessions -- Results

- Highlighted user difficulties in locating the appropriate queries to answer questions
 - Lack of familiarity with terminology used
 - Navigation problems
 - “Hidden” information
-

-
- A nice-looking interface
-

PlasmoDB

The Plasmodium Genome Resource

Release 5.2



PlasmoDB.org hosts genomic and proteomic data (and more) for different species of the parasitic eukaryote Plasmodium, the cause of Malaria. It brings together data provided by numerous laboratories worldwide (see the [Data Sources](#) page), and adds its own data analysis. Publications relying on PlasmoDB should please [acknowledge](#) the database developers and the scientists who have made their data available. PlasmoDB is part of an NIH/NIAID funded [Bioinformatics Resource Center](#) to provide [Apicomplexan Database Resources](#).

Features not yet available in PlasmoDB 5.2 may still be accessed via [PlasmoDB 4.4](#), and the results of PlasmoDB 4.4 queries may be exported to PlasmoDB 5.2 (see [PlasmoDB 4.4 Query History](#)).

As part of our ongoing education efforts, we are proud to announce the second of four annual [apicomplexan database workshops](#). The registration deadline is January 19, 2007.

- Queries & Tools
- History (series)
- Downloads
- Support
- als
- atures...
- Login
- subscribe
- now?
- ou can & and es?
- Qs'
- ia source

Related Sites:

- [ApiDB](#)
- [ToxoDB](#)
- [CryptoDB](#)
- [OrthoMCL-DB](#)
- [Recent Plasmodium publications](#)

News

- [05 December 2006 ApiDB Workshop, June 3-6, 2007](#)
- [17 November 2006 PlasmoDB 5.2 is released](#)
- [16 August 2006 PlasmoDB 5.1 is released](#)
- [25 May 2006 PlasmoDB 5.0 is fully released \(PlasmoDB 4.4 still available\)](#)
- [All PlasmoDB News](#)

[PlasmoDB Events](#) | [Release Notes](#)

Quick Tools [[Genome browser](#) | [BLAST](#) | [Sequence Retrieval](#) | [PlasmoCyc](#)]

Search Genes

PlasmoDB Gene ID

Gene ID: [▶go](#)

Keyword (search product name, notes, etc.)

Text term: [▶go](#)

Gene Type

Gene type: [▶go](#)

Search Genomic Sequences

Genomic Sequence ID

Sequence ID: [▶go](#)

All available queries:

Search genes

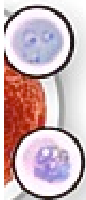
[▶](#) [▶](#) [▶go](#)

-
- But can be difficult for novice to locate desired query ...



PlasmoDB

The Plasmodium Genome Resource



PlasmoDB.org hosts genomic and proteomic data (and more) for different species of the parasitic eukaryote *Plasmodium* by numerous laboratories worldwide (see the [Data Sources](#) page), and adds its own data analysis. Public developers and the scientists who have made their data available. PlasmoDB is part of an NIH/NIAD funded Database Resources.

Features not yet available in PlasmoDB 5.2 may still be accessed via [PlasmoDB 4.4](#), and the results of [PlasmoDB 4.4 Query History](#).

As part of our ongoing education efforts, we are proud to announce the second of four registration deadline is January 19, 2007.

Queries & Tools

History (es)

Downloads

Support

Features...

Login

Subscribe

How?

Can and?

es!

source

Related Sites:

- [ApiDB](#)
- [ToxoDB](#)
- [CryptoDB](#)
- [OrthoMCL-DB](#)
- [Recent Plasmodium publications](#)

News

- **05 December 2006** [ApiDB Workshop, June 3-6, 2007](#)
- **17 November 2006** [PlasmoDB 5.2 is released](#)
- **16 August 2006** [PlasmoDB 5.1 is released](#)
- **25 May 2006** [PlasmoDB 5.0 is fully released \(PlasmoDB 5.0\)](#)
- [All PlasmoDB News](#)

[PlasmoDB Events](#) | [Release Notes](#)

Quick Tools

[Genome browser](#) | [BLAST](#) | [Sequence](#)

Search Genes

PlasmoDB Gene ID

Keyword (search product name, notes, etc.)

Gene Type

Search Genomic Sequences

Genomic Sequence ID

All available queries:

Search genes

Gene Type

Exon Count

Predicted Proteins =====

Isoelectric Point

Molecular Weight

PDB 3D Structures

Predicted 3D Structure (P.f.)

Low Complexity Proteins (P.f.)

Protein Secondary Structure

Similarity/Pattern =====

BLAST

Protein Motif Pattern

Interpro Domain

Population Biology =====

Snp Characteristics

Putative Function =====

EC Number (P.f.)

GO Term

Metabolic Pathway (P.f.)

Y2H Protein Interaction (P.f.)

Predicted Functional Interaction (P.f.)

Transcript Expression =====

Expression Timing (P.f.)

Expression Profile Similarity (P.f.)

Expression Percentile (P.f.)

Gametocyte Expression (P.f.)

Expression Percentile (P.b.)

EST evidence

Protein Expression =====

Protein Expression (P.f.)

Protein Expression (P.b.)

Cellular Location =====

Subcellular Localization

Transmembrane Domain Count

Predicted Signal Peptide

Exported Protein

Evolution =====

Homology Phylogenetic Profile (P.f.)

Orthology Phylogenetic Profile

Orthology ID



All Categories

Chromosomal Location

>go!

>go!

>go!

>go!

>go!

Video capture of lab exercise sessions -- Impact

- Restructuring of query menus -> query grid
 - Additional tutorials in form of movie clips
 - Rethinking of some help text
-



PlasmoDB Queries and Tools

Release 5.2

Home **Queries & Tools**

My Query History
(no queries)

Data Sources Downloads

About PlasmoDB Support

Tutorials

Upcoming features...

Email:

Password:

[forgot?](#)

Did you know?

Did you know [you can get Plasmodium clones and reagents?](#)

[Read all the 'FAQs'](#)

Featured data source

[: P.berghei protein expression](#)

[See all data sources](#)



Queries

Query Availability: **A** = ApiDB **C** = CryptoDB **P** = PlasmoDB **T** = ToxoDB

Identify Genes Based On:

- | Genomic Position | Gene Attributes | Population Biology |
|--|--|--|
| <ul style="list-style-type: none"> Chromosomal Location A C P T Proximity to Centromeres P Proximity to Telomeres P Non-nuclear Genomes P T | <ul style="list-style-type: none"> Type (e.g. rRNA, tRNA) A C P T Exon/Intron Structure A C P T | <ul style="list-style-type: none"> SNPs P T Microsatellites |
| Transcript Expression | Similarity/Pattern | Protein Expression |
| <ul style="list-style-type: none"> EST Evidence C P T SAGE Tag Evidence Microarray Evidence P | <ul style="list-style-type: none"> Protein Motif A C P T Interpro/Pfam Domain A C P T BLAST similarity A C P T | <ul style="list-style-type: none"> Mass Spec. Evidence C P |
| Predicted Proteins | Putative Function | Cellular Location |
| <ul style="list-style-type: none"> Molecular Weight A C P T Isoelectric Point P T Secondary Structure P Crystal Structure P Predicted 3D Structure P | <ul style="list-style-type: none"> GO Term A C P T EC Number A C P T Metabolic Pathway C P T Y2H Interaction P Predicted Interaction P | <ul style="list-style-type: none"> Signal Peptide A C P T Transmembrane Domain A C P T Organellar Compartment P Exported to Host P |
| Evolution | Other Attributes | |
| <ul style="list-style-type: none"> Orthologs/Paralogs A P Orthology Profile A C P T Homology Profile P | <ul style="list-style-type: none"> Keyword A C P T ID A C P T Species A C P Available Reagents P | |

Identify Genomic Sequences Based On:



ApiDB.org Queries and Tools

**My
Queries:**
0

Queries

Query Availability: A = ApiDB C = CryptoDB P = PlasmoDB T = ToxoDB

Identify Genes Based On:

Genomic Position

- Chromosomal Location A C P T
- Proximity to Centromeres P
- Proximity to Telomeres P
- Non-nuclear Genomes P T

Gene Attributes

- Type (e.g. rRNA, tRNA) A C P T
- Exon/Intron Structure A C P T

Population Biology

- SNPs P T
- Microsatellites

Transcript Expression

- EST Evidence C P T
- SAGE Tag Evidence
- Microarray Evidence P

Similarity/Pattern

- Protein Motif A C P T
- Interpro/Pfam Domain A C P T
- BLAST similarity A C P T

Protein Expression

- Mass Spec. Evidence C P

Predicted Proteins

- Molecular Weight A C P T
- Isoelectric Point P T
- Secondary Structure P
- Crystal Structure P
- Predicted 3D Structure P

Putative Function

- GO Term A C P T
- EC Number A C P T
- Metabolic Pathway C P T
- Y2H Interaction P
- Predicted Interaction P

Cellular Location

- Signal Peptide A C P T
- Transmembrane Domain A C P T
- Organellar Compartment P
- Exported to Host P

Evolution

- Orthologs/Paralogs A P
- Orthology Profile A C P T
- Homology Profile P

Other Attributes

- Keyword A C P T
- ID A C P T
- Species A C P

Card-sorting analysis of query categories

- Goals:

- Determine the groups of the queries that users view as belonging together
 - Determine appropriate names for the groups that emerge.
 - Produce better menu structure
-

How The Exercise Was Conducted

- Users given 39 cards with titles and descriptions of queries available on PlasmODB 4.4
- Sample card

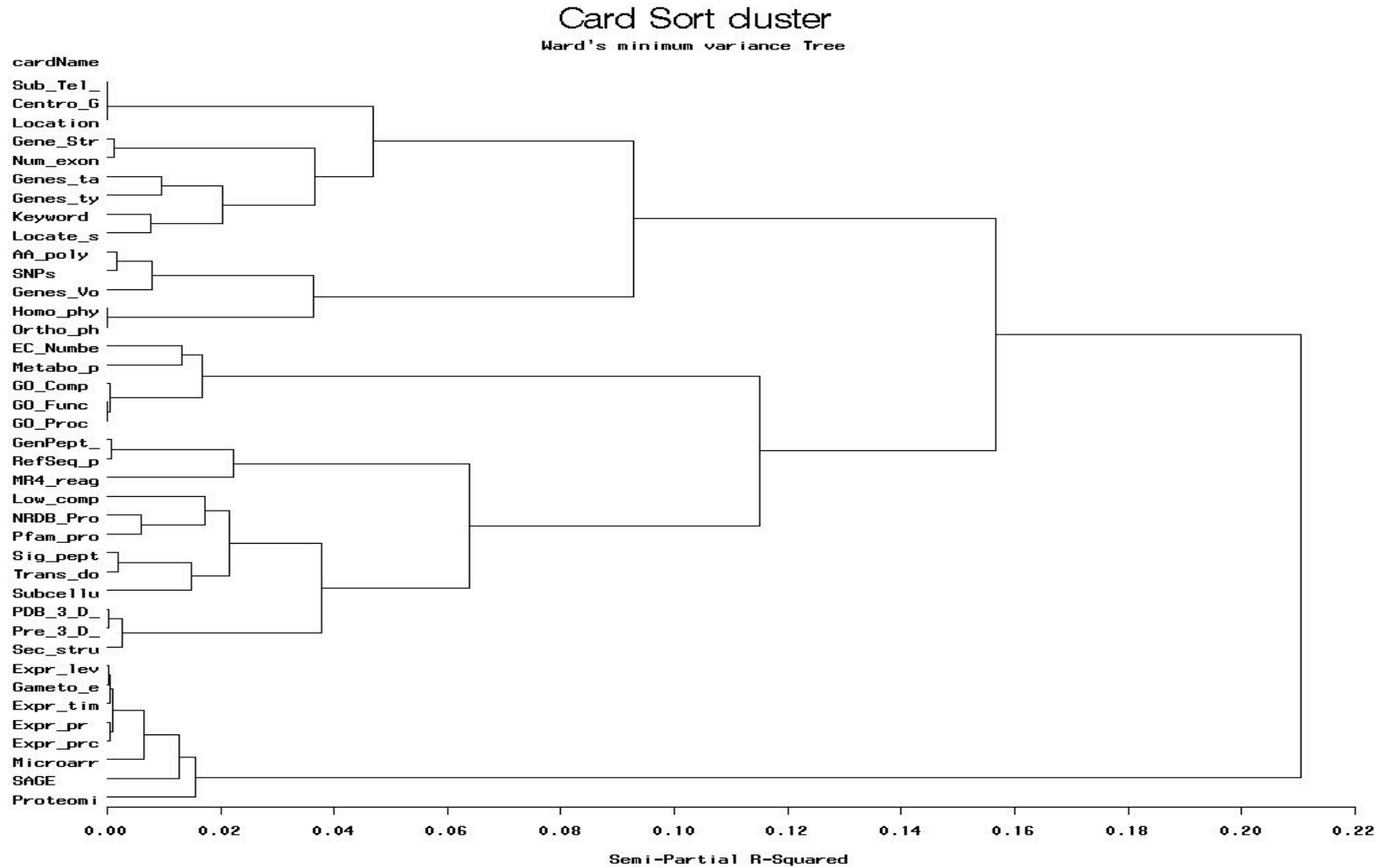
Signal peptide

Search for genes whose protein products contain predicted signal peptides, as predicted by SignalP 3.0.

Card Sorting Analysis

- Users asked to place cards into meaningful groups, clip them together, and label the groups.
 - Hierarchies allowed -- rubber band together and label clipped stacks
 - Cluster analysis and other analyses
 - 7 different methods; similar results across methods
-

Sample Cluster Output - Tree



Site Tour

- with Steve Fischer



Usability Guidelines

- A good place to start:
 - [AskTog: First Principles of User Interaction Design](#)
 - Important Lessons:
 - Preference != performance
 - Clutter <-> visibility tradeoffs
 - Interaction behavior more important than appearance
-

Conclusions

- User studies can answer questions and guide development
 - Survey
 - provides info about user preferences (not performance!)
 - Easy to administer/analyze
 - Video capture
 - Provides info about user performance
 - Tedious to analyze
 - Biggest impact:
 - Card-sorting exercise
 - Individual interviews
 - Both are easy to administer; straightforward to analyze
-