InterMine & MODs

Richard Smith
richard@flymine.org

www.intermine.org
InterMine

- data integration
- query-optimised data warehouse
- works with any data model
- flexible web interface
- web services
Data Integration

Existing data sources:
- UniProt XML
- e!
- Chado
- OMIM disease
- OBO ontology
- PSI XML
- Inparanoid

Custom Data Sources:
- FASTA
- GFF3
- XML

Configurable data integration

Java and Perl APIs

InterMine data warehouse
Web Interface

- **Template queries** - pre-defined queries
- **QueryBuilder** - custom queries, advanced
- **Report pages** - configurable, templates
- **Lists** - upload, use in queries, analysis widgets
- **Export**
- **Search**
New Features

- report page update
- region search
- better galaxy integration
- expanded web services
- new results tables (soon)
InterMOD

Yeast

Zebrafish

InterMine

Worm

Mouse

Rat

FlyMine

3 years funding, thanks NHGRI!
InterMOD

Julie Sullivan
Alex Kalderimis
Progress

RGD
SGD
ZFIN
MGI
WormBase
FlyMine
YeastMine is Awesome!
Demo
Lists
custom queries
region search
enrichment widgets
similarity search (soon)
more data
Results tables

Widgets

Displayers

Lists
custom queries
region search
enrichment widgets
similarity search (soon)
more data
Template query embed links....
The plan

• Delete existing webapp
  • Java/JSP - old and slow to develop
• All functionality as web services
• Key components as JavaScript
• Lightweight new webapp
PathQuery q = new PathQuery(model);

q.setView("Protein.primaryIdentifier,
           Protein.genes.primaryIdentifier");

q.addConstraint("Protein.proteinDomains.name",
                Constraints.eq("Homeobox"));
q.addConstraint("Protein.organism.shortName",
                Constraints.eq("D. melanogaster"));
Result formats

- JSON (objects)
- TSV
- BED
- CSV
- JSON (rows)
- JSONP
- FASTA
- GFF3
• MODs maintain data
• regular releases (weekly)
• similar data models
• powerful web services
mash ups
1. Users can query across model organisms
2. InterMine API demos for developers
Demo
Caching

1. Query results are cached in the InterMine server

2. Each Mine has a release number:
   yeastmine.org/service/release
   if the number is the same no data has changed.

Easy to add local caches and check version numbers.
Why is this useful

- Data integration is hard
  - resolution of ids, old ids, update schedules
- Problems with stale data
- Small databases can present more data without local storage
Conclusion

- MODs maintain data
- InterMine provides query software
  - focus on performance
- Bioinformaticians build webapps!
  - easy to attach to multiple MODs
InterMine Team

Biologists
Rachel Lyne, Adrian Carr, Mike Lyne

Developers
Sergio Contrino, Julie Sullivan, Alex Kalderimis, Fengyuan Hu, Daniela Butano, Radek Stepan, Richard Smith

Sys Admin
Dan Tomlinson

PI
Gos Micklem

MODs

RGD
Andrew Vallejos, Elizabeth Worthey, (Simon Twigger)

SGD
Kalpana Karra, Ben Hitz, Gail Binkley, Mike Cherry

ZFIN
Sierra Taylor, Judy Blake, Nathan Dunn, Monte Westerfield

MGI
(Michael O’Keefe), Joel Richardson

WormBase
Todd Harris, Lincoln Stein

www.intermine.org
richard@flymine.org