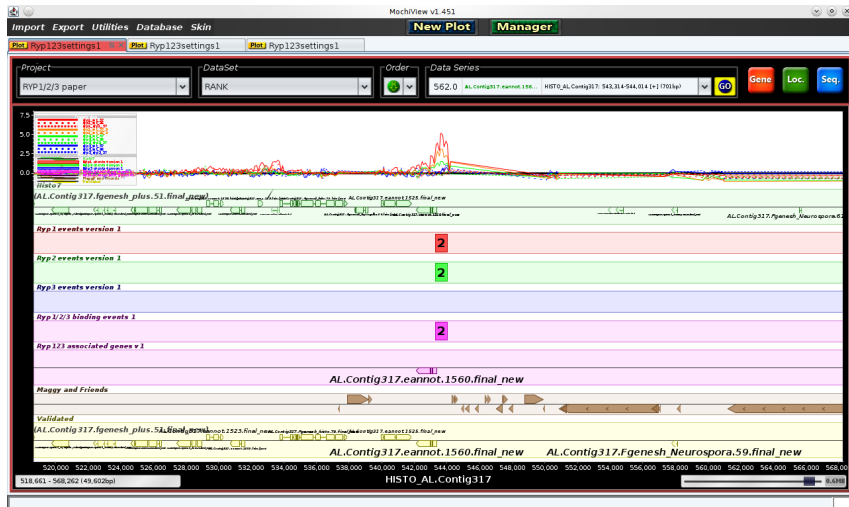


Genome Browsing and Course Summary

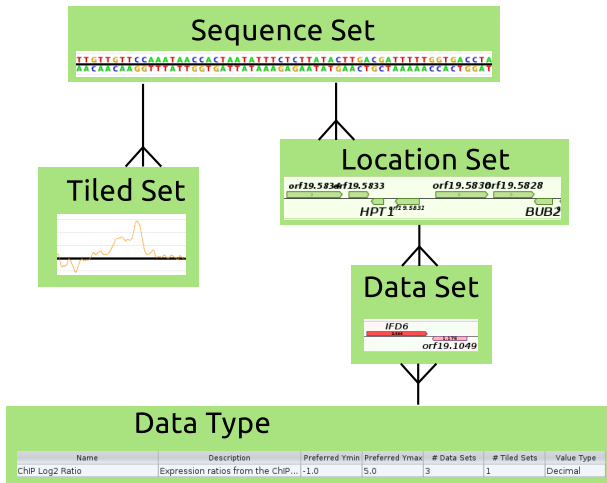
Mark Voorhies

4/9/2012

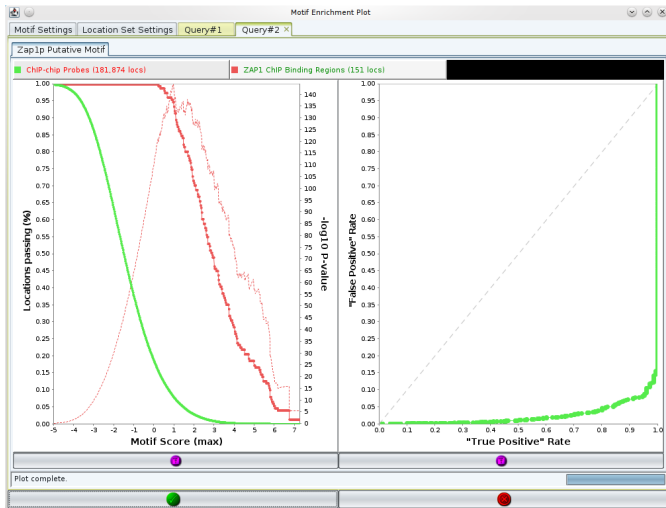
MochiView



MochiView Schema



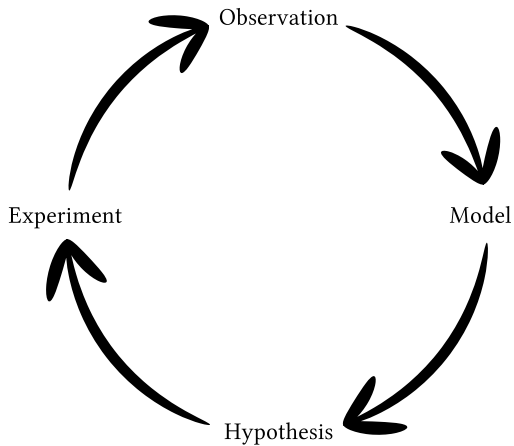
Choosing thresholds with ROC plots



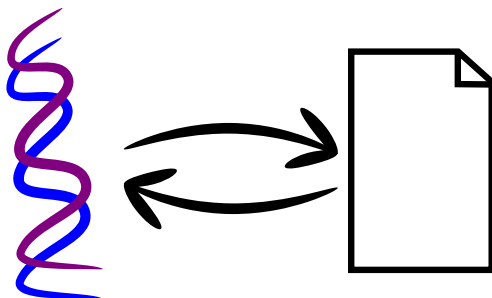
RTFM



- For tools:
 - Read the manual
 - Read the paper
- Good general references:
 - The O'Reilly BLAST book
 - Durbin, Eddy, Krogh, and Mitcheson (HMMs)
 - Numerical Recipes
 - Branden & Tooze

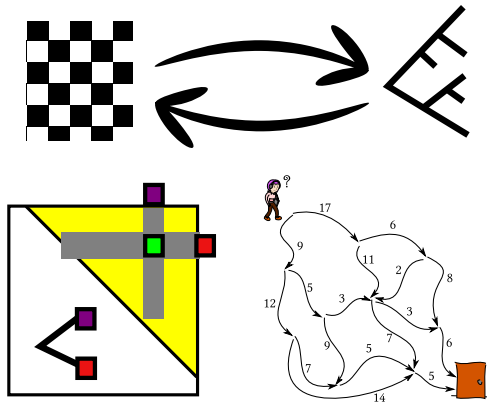


Every object should have an isomorphism to a file



- Export, audit, edit, and import *independent* of a given program.
- Standard file formats for portability.
- Don't be afraid to look inside and hack on *your* data files.

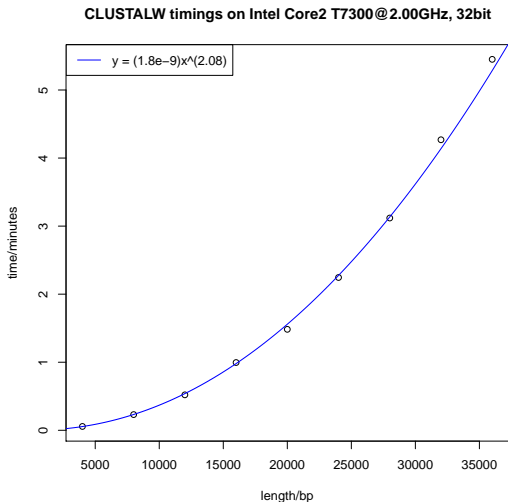
A few techniques can solve many problems



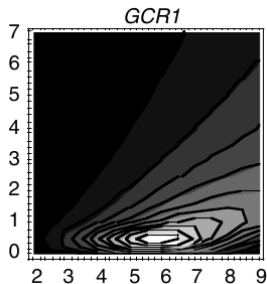
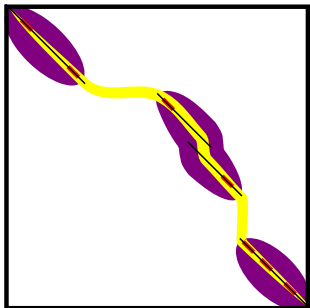
Iteration, clustering, dynamic programming, ...



Run times are predictable and measurable

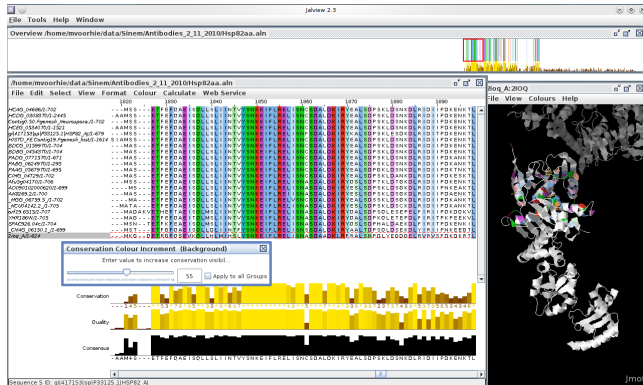


Heuristics and stochastic sampling for hard problems



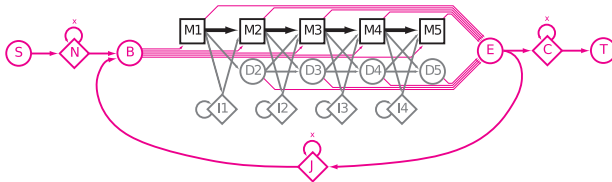
BLAST, HMMer3, MrBayes, ...

Evolution is a rich source of information



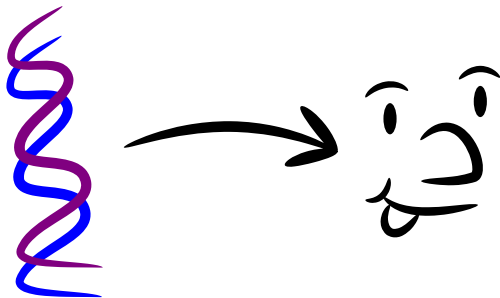
- Infer homology from sequence similarity
- More sequences provide more information

HMMs capture position and gap information

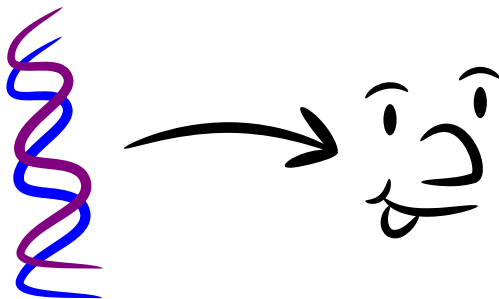


(Image from Sean Eddy, PLoS Comp. Biol. 4:e1000069)

Phenotype is more diverse than Genotype

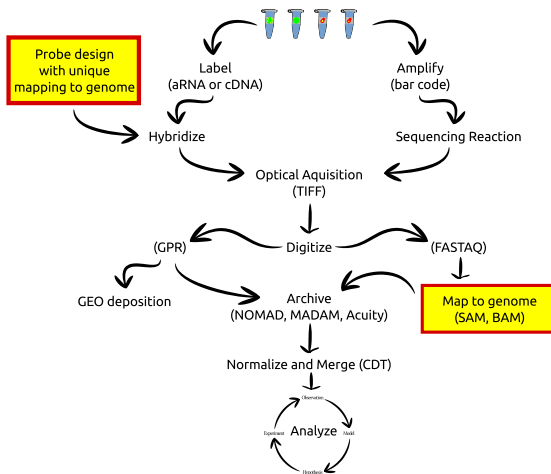


Phenotype is more diverse than Genotype

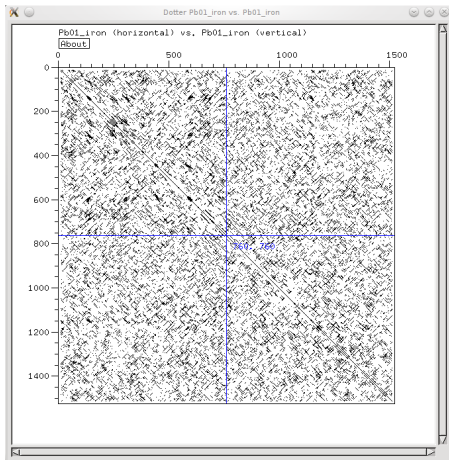


- Make sure you know what you are measuring
- Nucleic acid sequences are especially easy to address
- Many phenotypes can be analyzed by common numerical methods

Sequencing methods sidestep design-time decisions



Start from an unbiased view



Tools should support aggregation and annotation

Java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt

File Settings Analysis Export Window Help

Dendrogram

View Status
Row: 115 (Yof)
Column: 49 (spt)
Value: 1.34

Usage Hints
Mouse over to get info

GeneTreeAnno: /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt

Sporulation

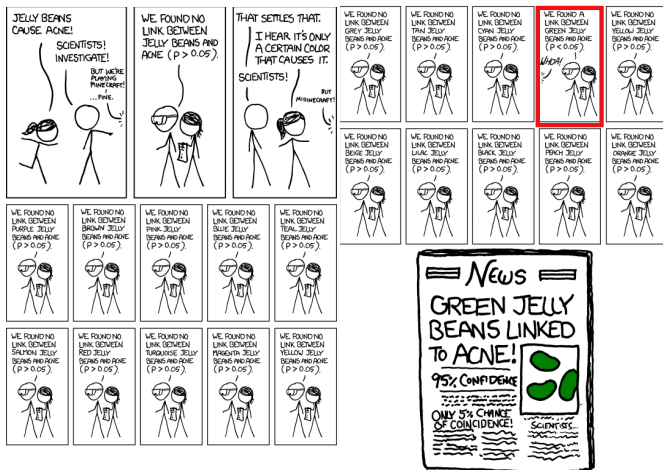
Name: Sporulation Annotation: Genes upregulated in sporulation

NODEID	LEFT	RIGHT	CORRELAT...	NAME	ANNOTATI...
NODE184...	NODE184...	NODE152...	0.627369	Sporulation	Genes up...
NODE184...	NODE184...	GENE56X	0.627369		
NODE184...	NODE184...	NODE178...	0.627369		
NODE184...	NODE150...	GENE177...	0.627287		

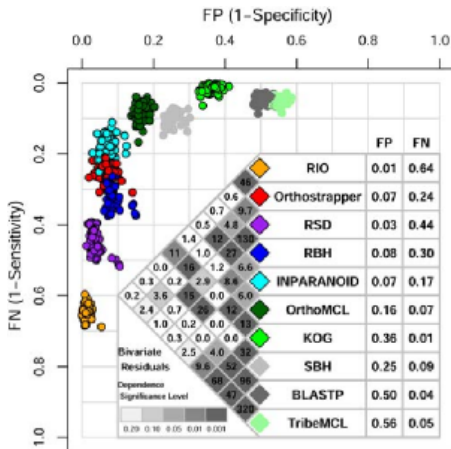
Dock Close

YFR029C CDC14 MITOSIS PROTEIN PHOSF
 YML085W ORC1 DNA REPLICATION ORIGIN F
 YIL139C REV7 DNA REPAIR DNA POLYMEF
 YNL318C NONE TRANSPORT HEXOSE PERM
 YFR023W PES4 DNA REPLICATION UNKNOWN;
 YHR015W MIP6 MRNA EXPORT, PUTATIVE RNA
 YDR263C DIN7 DNA REPAIR (PUTATIVE) DNA
 YLR045C STU2 CYTOSKELETON SPINDLE
 YOR033C DHS1 DNA REPAIR EXONUCLEASE
 YIL156W BNR1 CYTOSKELETON ACTIN FI
 YKL042W SPC42 CYTOSKELETON SPINDLE
 YNL225C CNM67 CYTOSKELETON SPINDLE
 YCR092C CDC10 CYTOKINESIS GTP BINDING
 YLR210W CLB4 CELL CYCLE G2/M CYCLIN
 YLR314C CDC3 CYTOKINESIS SEPTIN
 YBR045C GIP1 GLUCOSE REPRESSION (PUTA
 YDL156W CLB3 CELL CYCLE G2/M CYCLIN
 YDR118W APC4 CELL CYCLE ANAPHASE-PF
 YDR253C MET32 METHIONINE METABOLISM TRP
 YMR196W CIK1 CYTOSKELETON SPINDLE
 YDR113E POS1 CELL CYCLE ANAPHASE-TA

Account for Multiple Hypothesis Testing



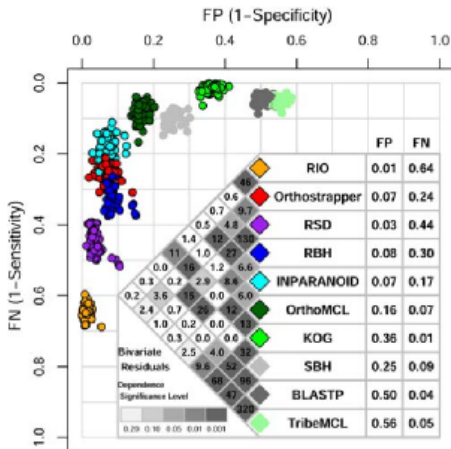
Choose appropriate statistical tools



- Solid theoretical basis
- Reasonable power
- Estimates significance *and* effect size

PLoS ONE 2:e383

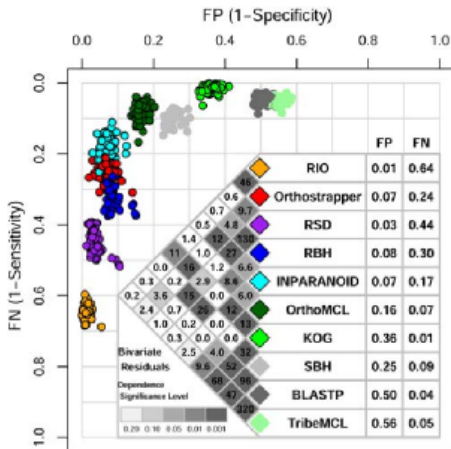
Choose appropriate statistical tools



- Solid theoretical basis
- Reasonable power
- Estimates significance *and* effect size
- Addresses the question that you are asking

PLoS ONE 2:e383

Choose appropriate statistical tools



- Solid theoretical basis
- Reasonable power
- Estimates significance *and* effect size
- Addresses the question that you are asking
- When in doubt, find positive and negative controls

PLoS ONE 2:e383

Science is a Conversation

- Follow computational methods as they evolve (Web of Science, PubMed RSS...)

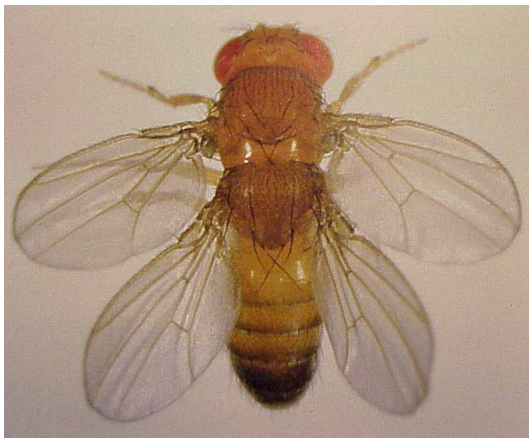
Science is a Conversation

- Follow computational methods as they evolve (Web of Science, PubMed RSS...)
- As a reviewer, insist on availability of source code

Science is a Conversation

- Follow computational methods as they evolve (Web of Science, PubMed RSS...)
- As a reviewer, insist on availability of source code
- Learn to program and to work in a UNIX environment
 - Mark Lutz: Learning Python
 - Linux: Ubuntu or Knoppix Live CDs

We understand systems by breaking them



Source: Peter A. Lawrence via <http://www.bio.davidson.edu/courses/molbio/ubx/ubx.html>