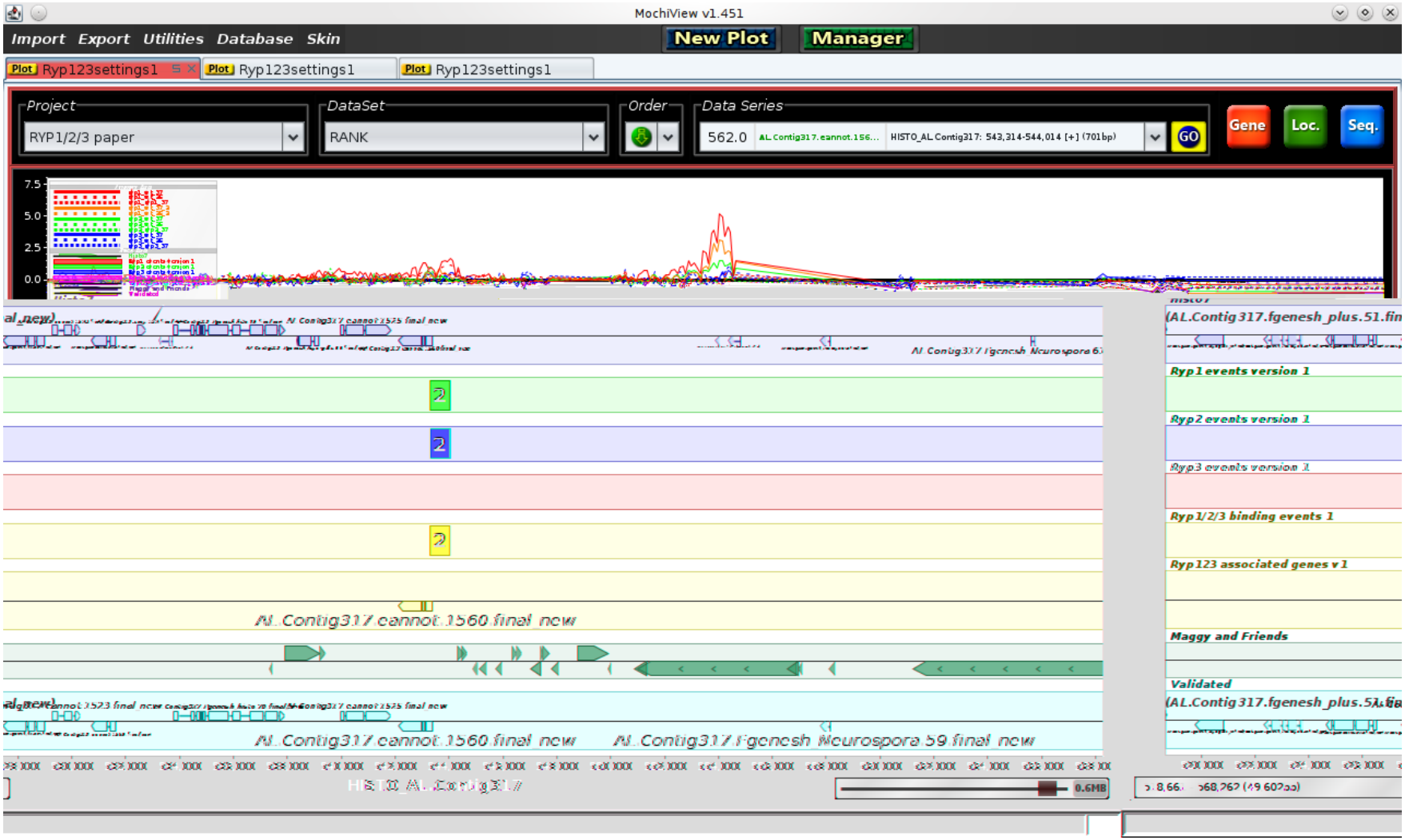
Genome Browsing and Course Summary

Mark Voorhies

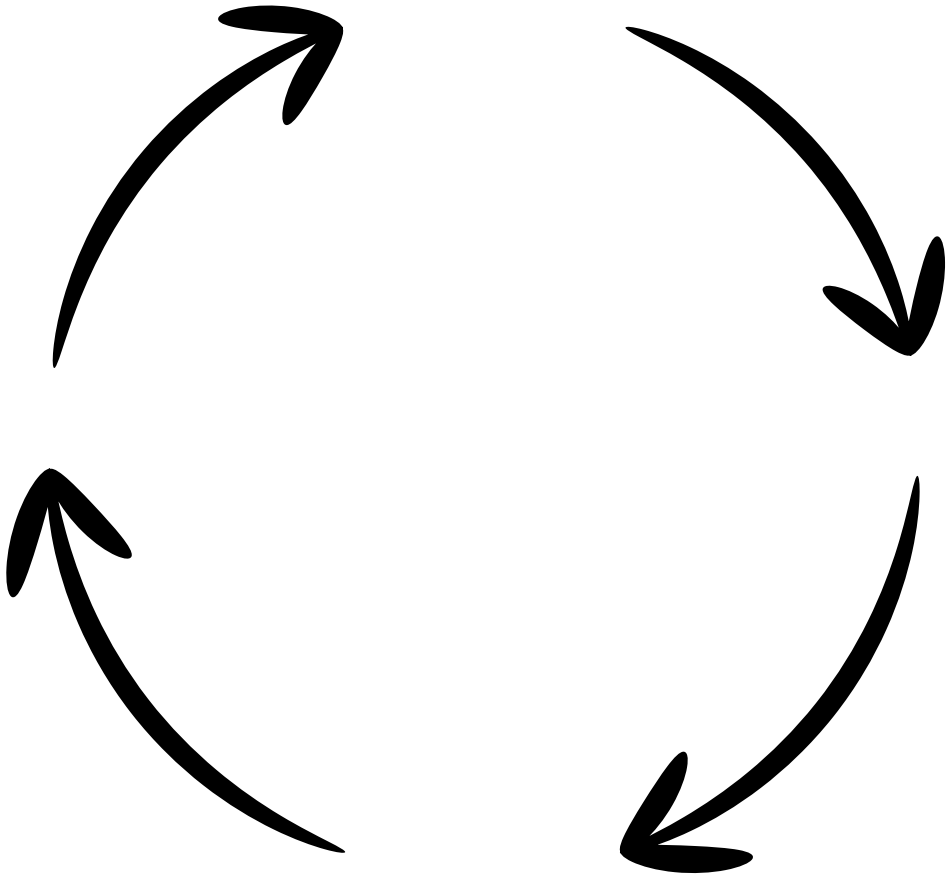
4/8/2011

MochiView

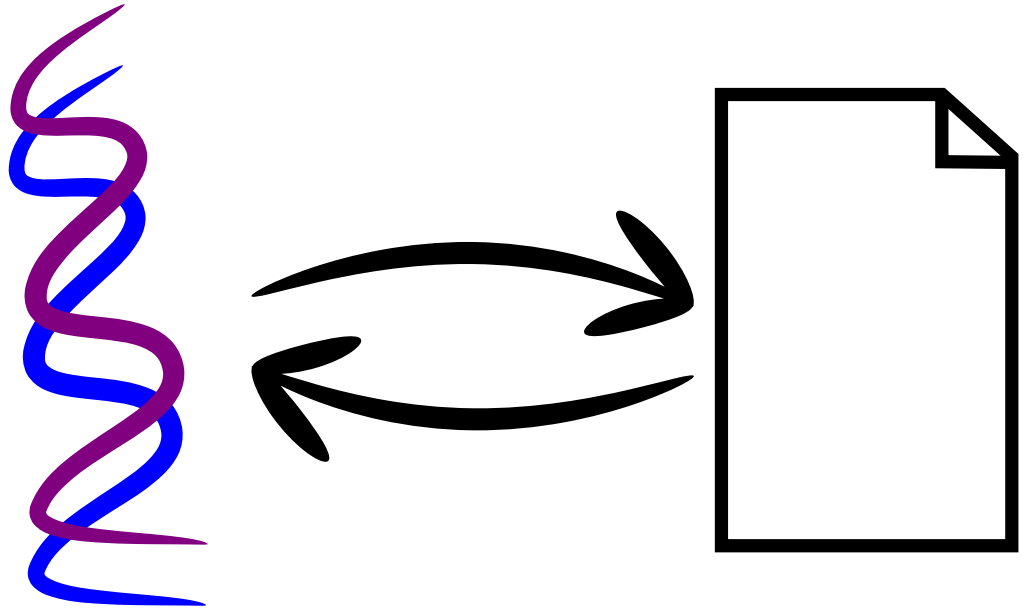

MochiView v1.451
Import Export Utilities Database Skin
New Plot
Manager
Plot Ryp123settings1
Plot Ryp123settings1
Plot Ryp123settings1
Project
RYP1/2/3 paper
DataSet
RANK
Order
Data Series
562.0
GO
Gene
Loc.
Seq.
7.5
5.0
2.5
0.0
Al_Contig317_cannot_1560_final_new
Al_Contig317_fgenesh_Neurospora_59_final_new
(AL.Contig317.fgenesh_plus.51.fin
Ryp1 events version 1
Ryp2 events version 1
Ryp3 events version 1
Ryp1/2/3 binding events 1
Ryp123 associated genes v1
Maggy and Friends
Validated
(AL.Contig317.fgenesh_plus.5
0.6MB

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

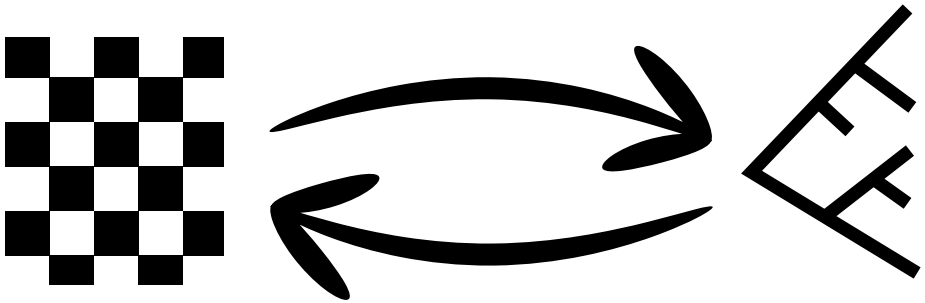


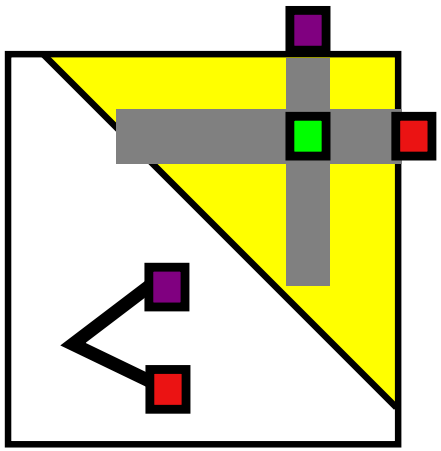
Every object should have an isomorphism to a le

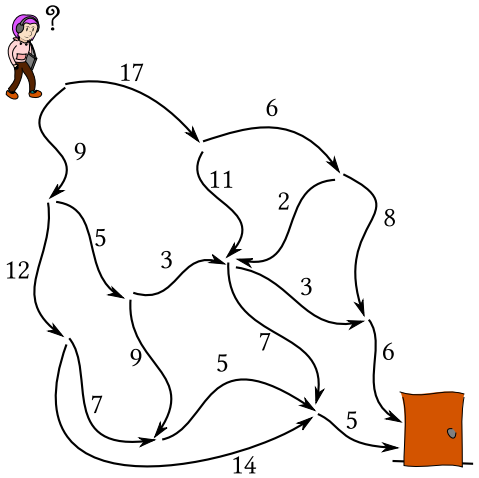


- Export, audit, edit, and import *independent* of a given program.
- Standard file formats for portability.
- Don't be afraid to look inside and hack on *your* data files.

A few techniques can solve many problems

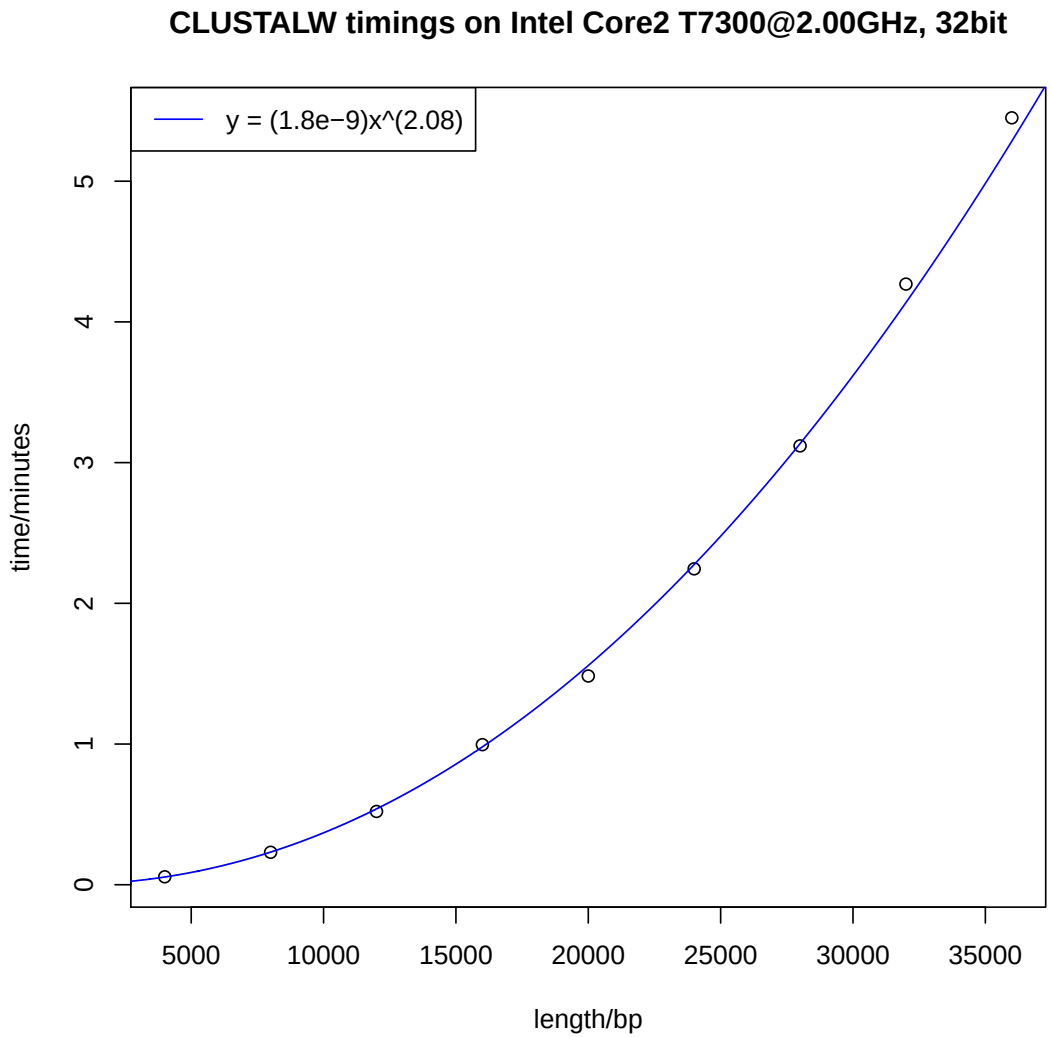




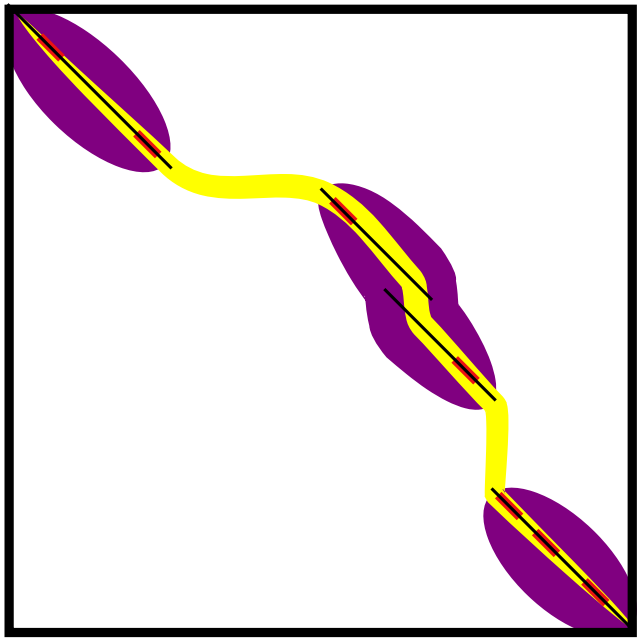


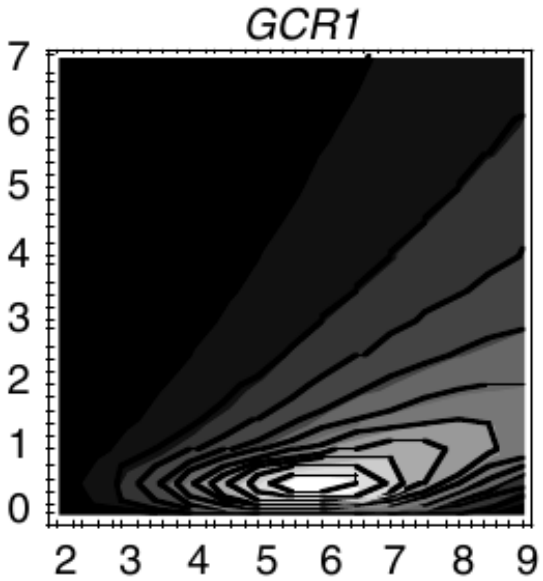
Iteration, clustering, dynamic programming, ...

Run times are predictable and measurable



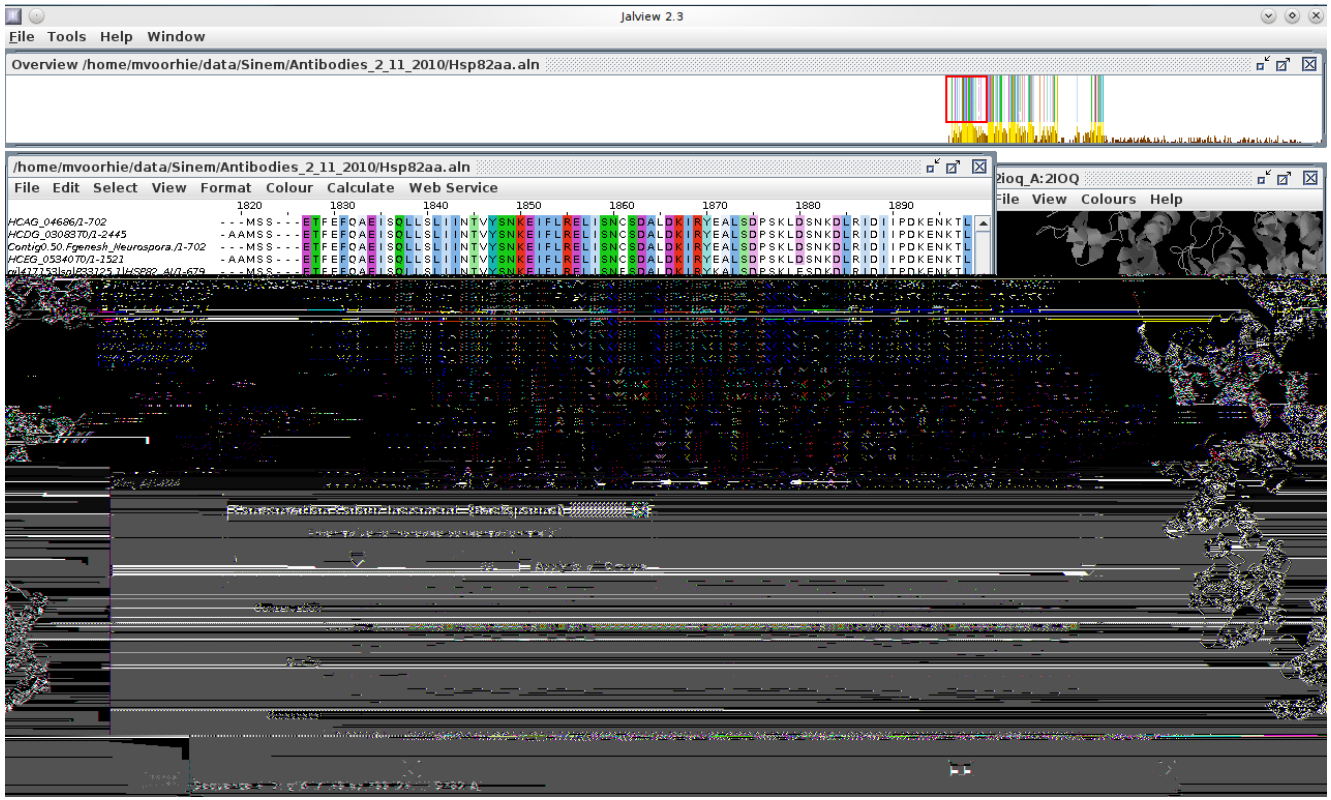
Heuristics and stochastic sampling for hard problems





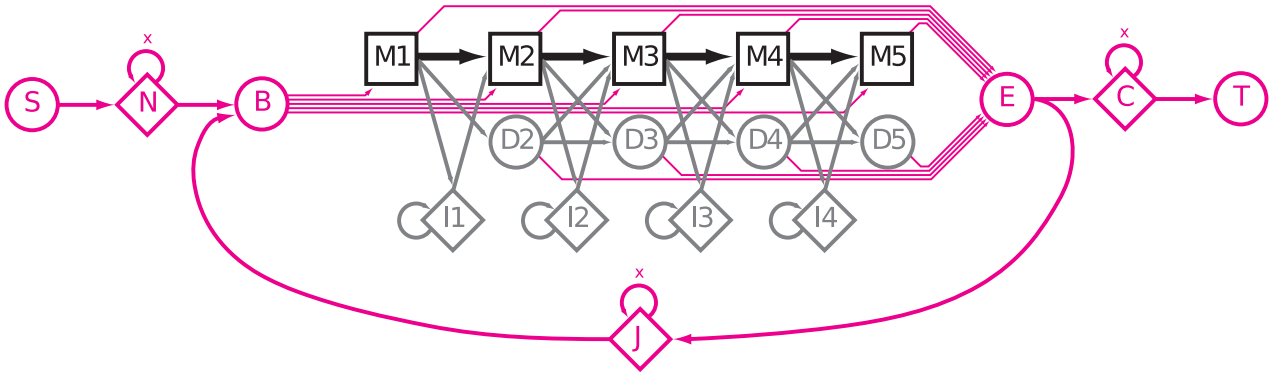
BLAST, HMMer3, BAGEL, MrBayes, ...

Evolution is a rich source of information



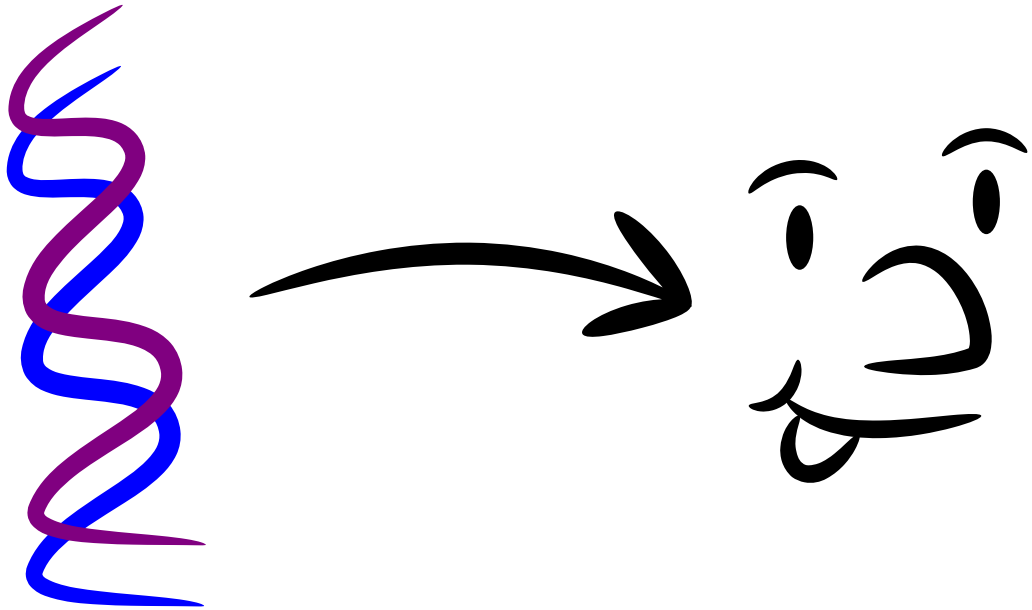
- Infer homology from sequence similarity
- More sequences provide more information

HMMs capture position and gap information

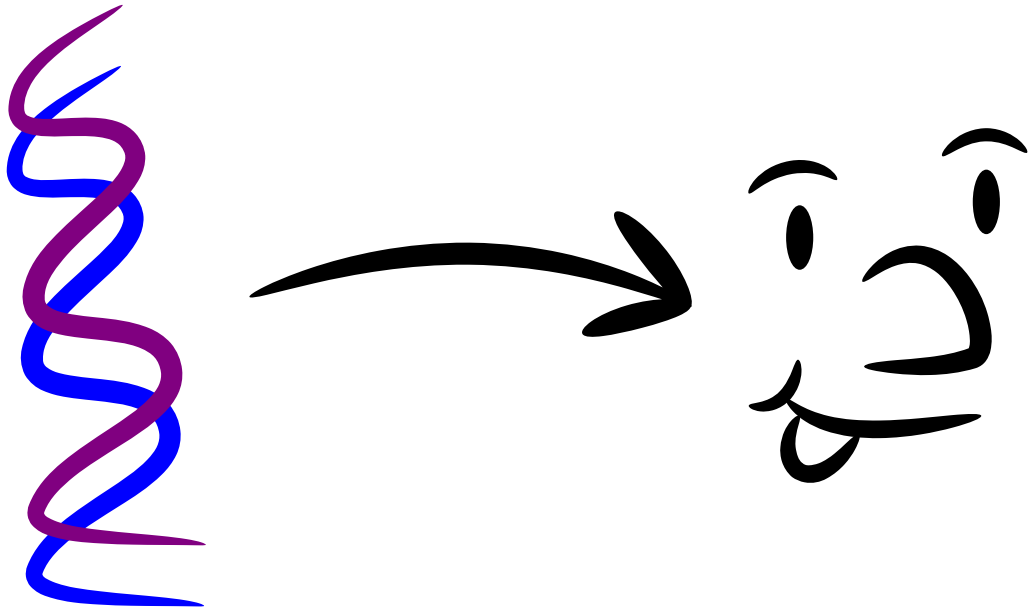


(Image from Sean Eddy, PLoS Comp. Biol. 4:e1000069)

Phenotype is more diverse than Genotype

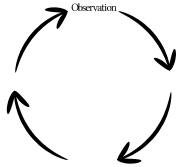


Phenotype is more diverse than Genotype

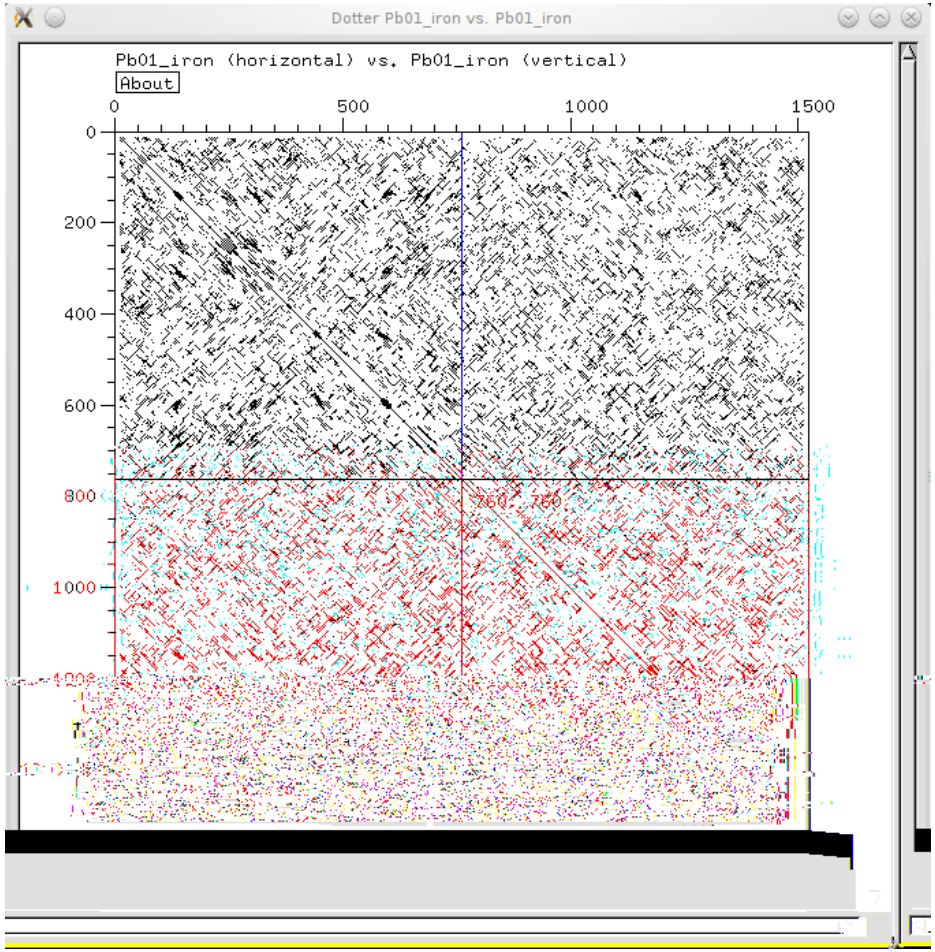


- Make sure you know what you are measuring
- Nucleic acid sequences are especially easy to address
- Many phenotypes can be analyzed by common numerical methods

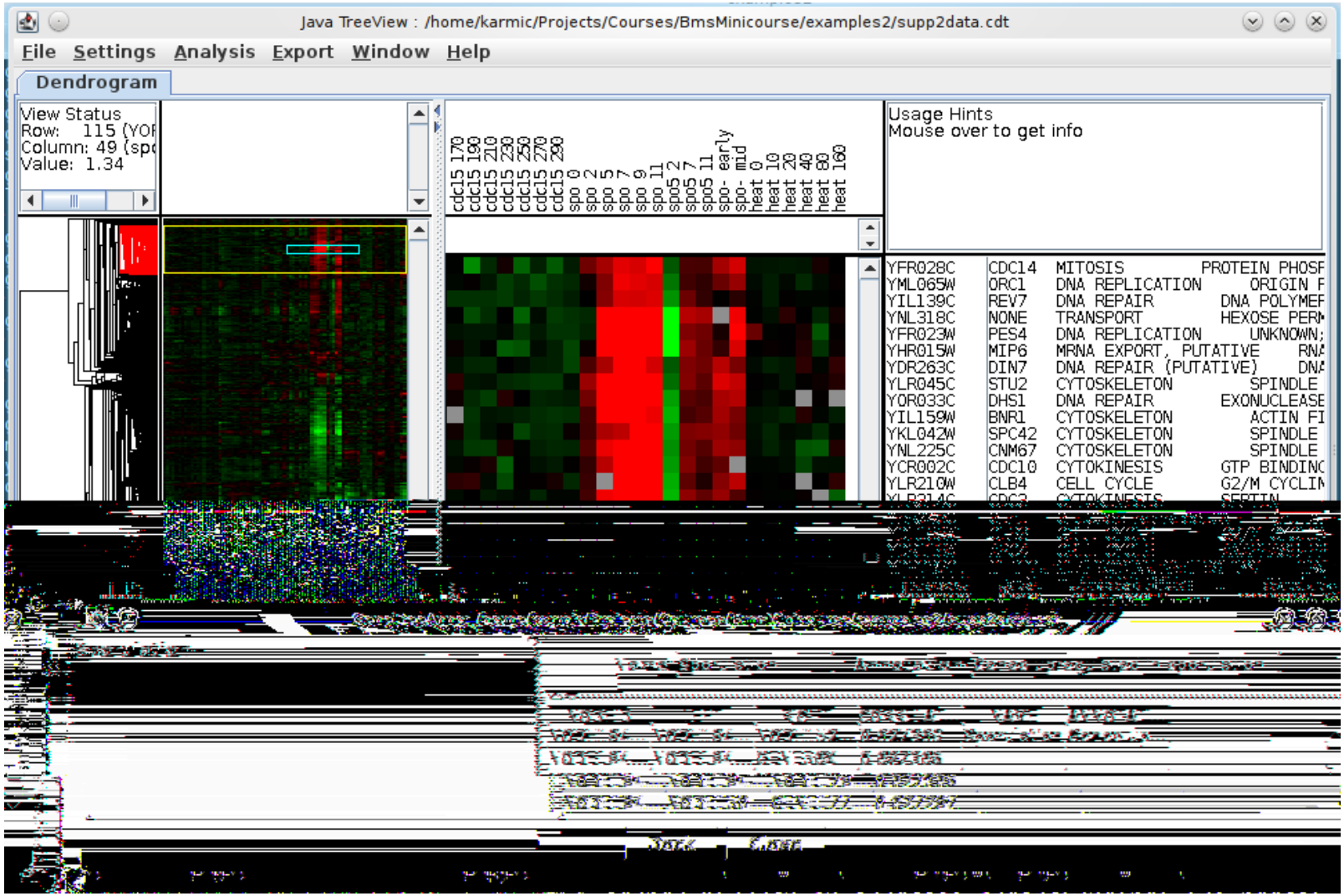
Sequencing methods sidestep design-time decisions



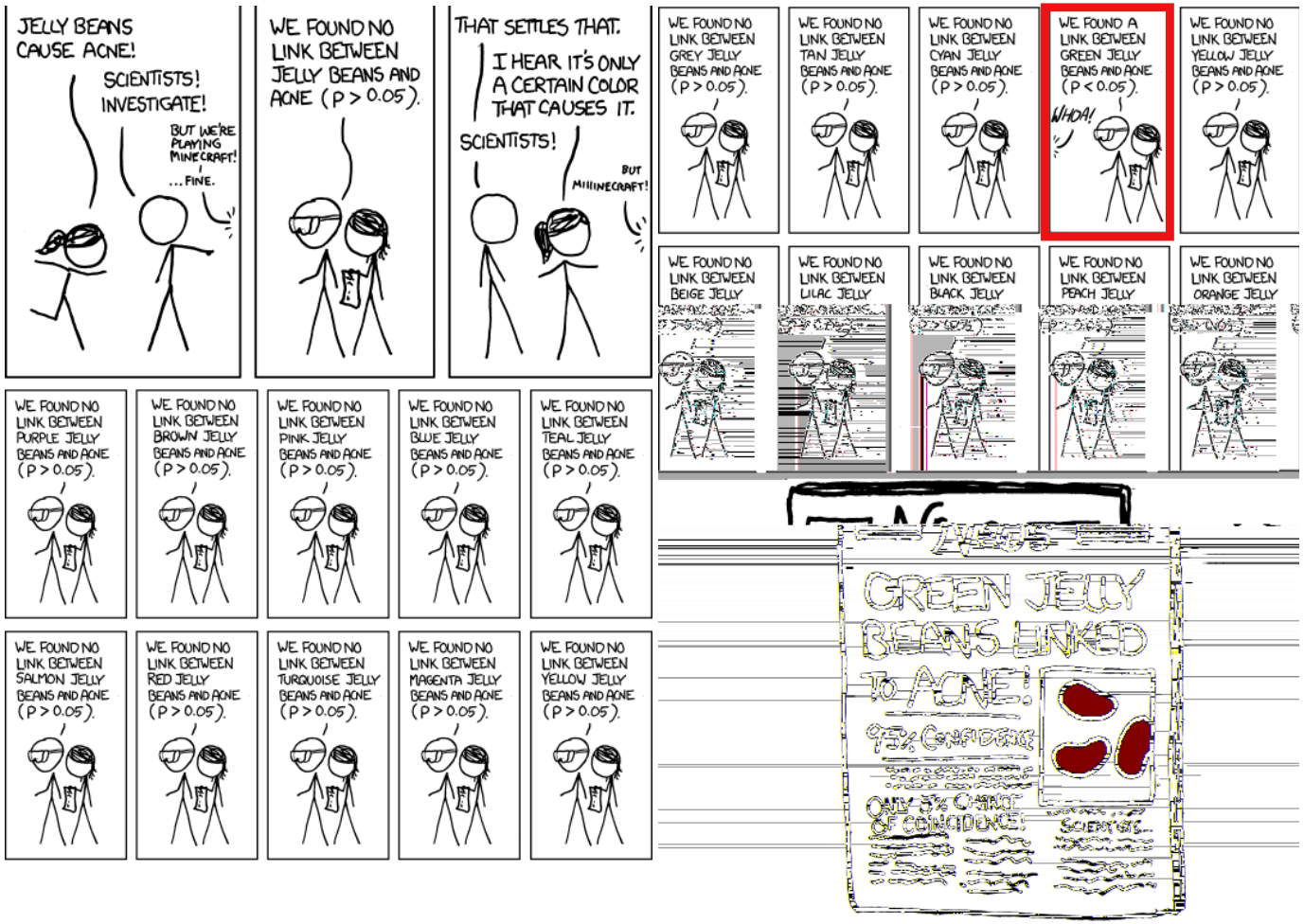
Start from an unbiased view



Tools should support aggregation and annotation

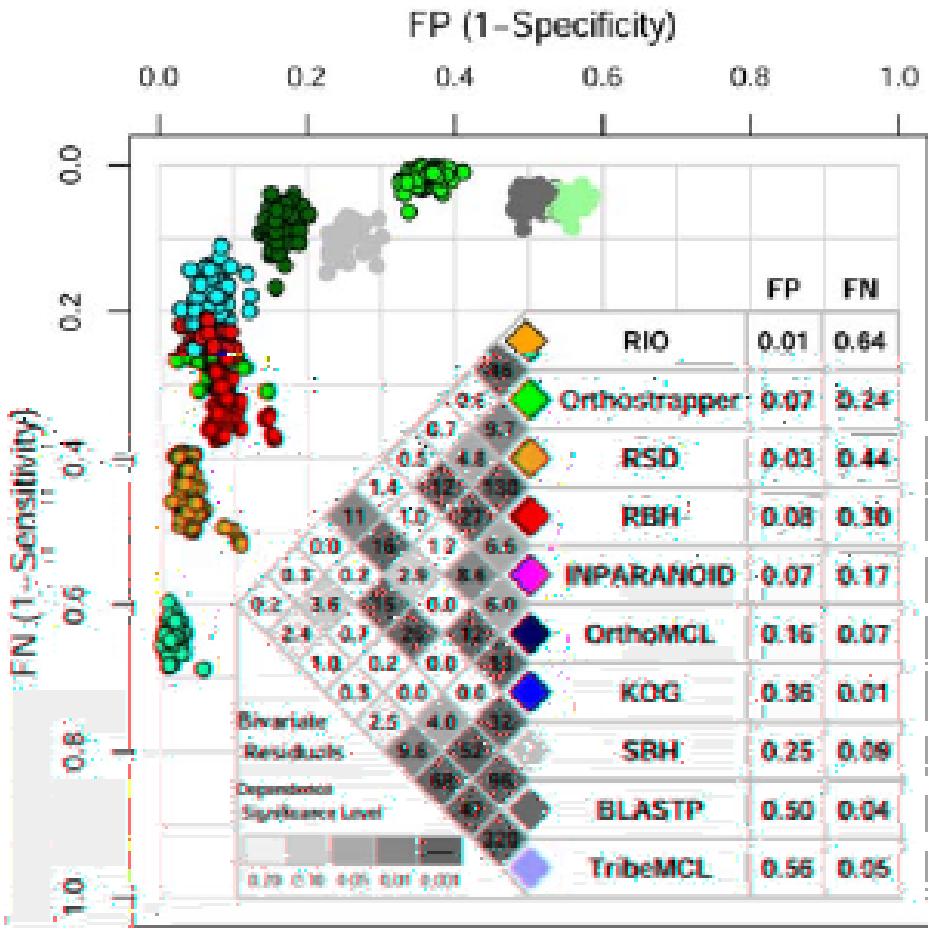


Account for Multiple Hypothesis Testing



http://xkcd.com/882/

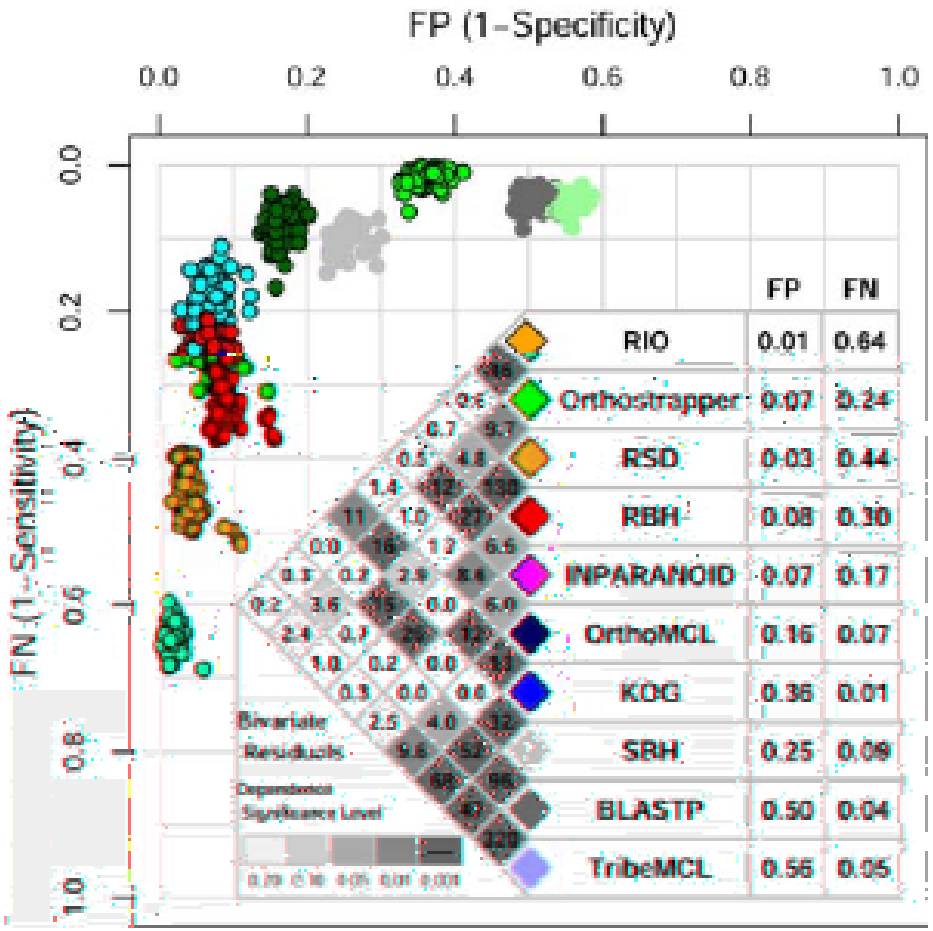
Choose appropriate statistical tools



PLoS ONE 2:e383

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

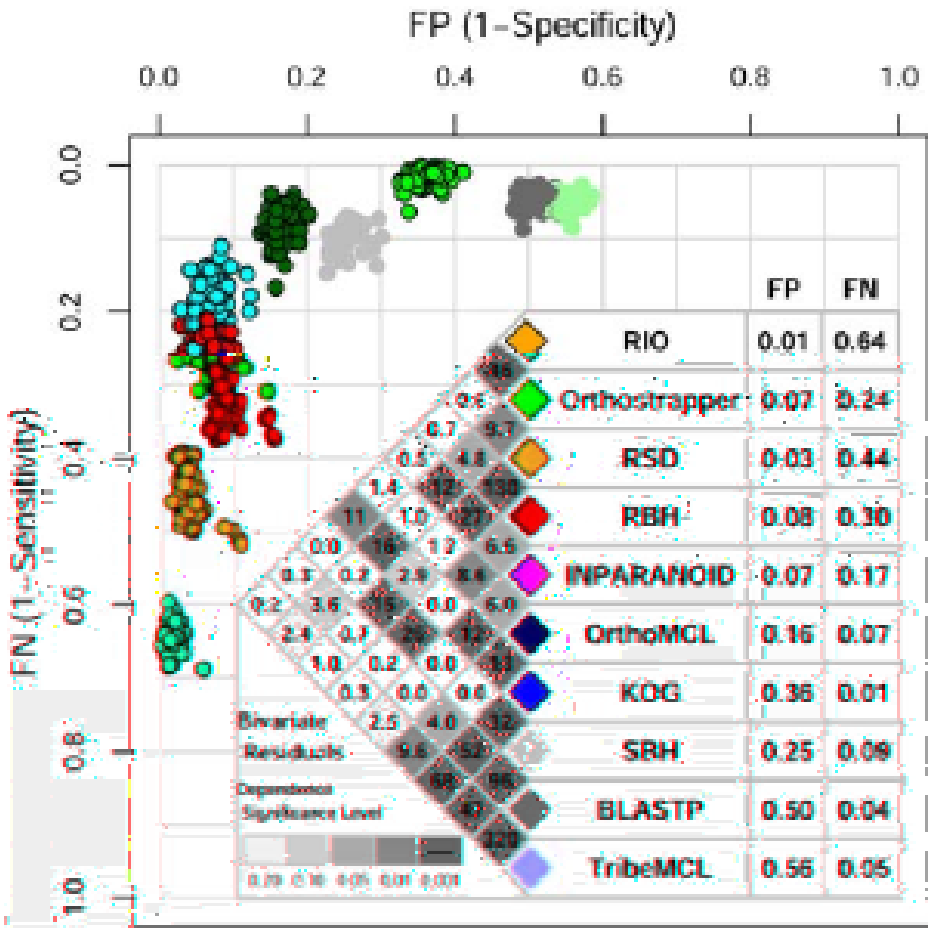
Choose appropriate statistical tools



PLoS ONE 2:e383

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

Choose appropriate statistical tools



PLoS ONE 2:e383

- 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

Science is a Conversation

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

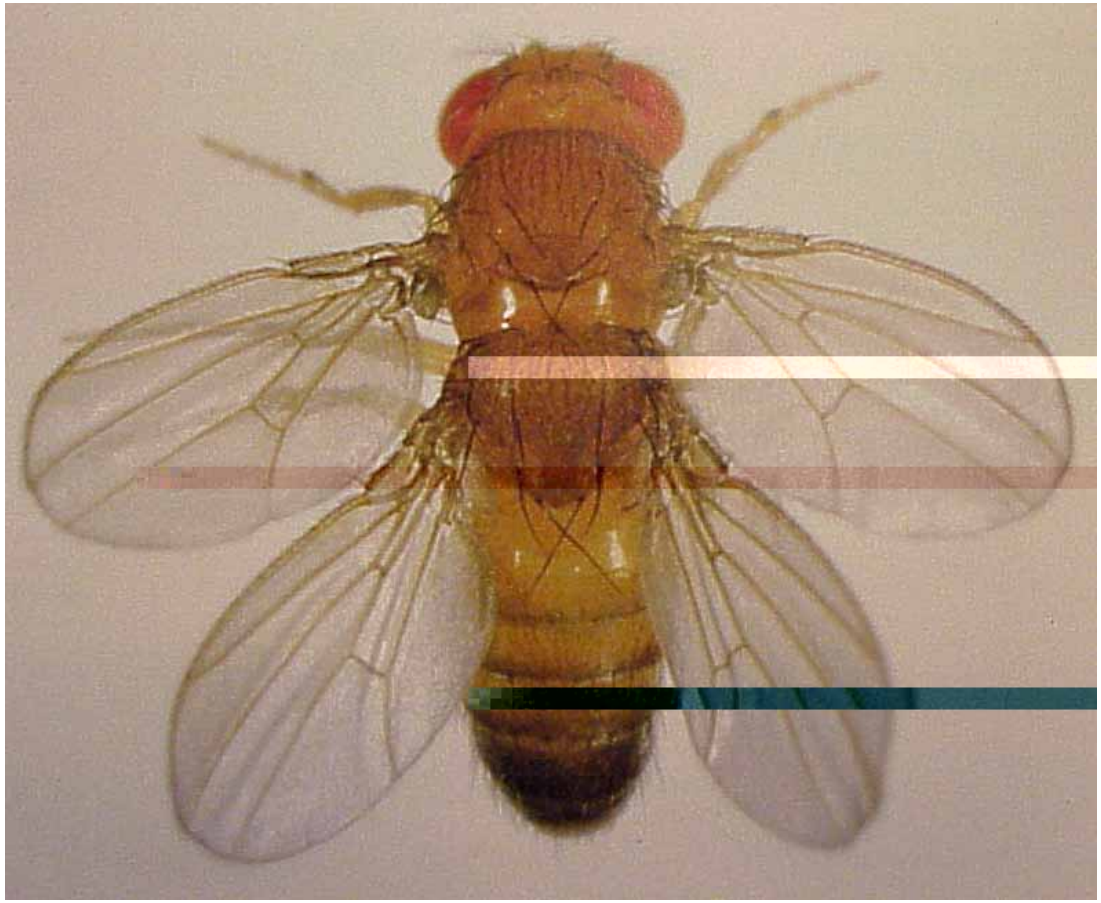
Science is a Conversation

- 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