

Lies, damn lies, and genomics

you, your data, your perceptions and
reality

Christopher West Wheat



Goal of this lecture

- Present a critical view of ecological genomics
- Make you uncomfortable by sharing my nightmares
- Encourage you to critically assess findings and your expectations in light of publication biases

Disclaimer

I'm a positive person

I love my job and the work we all do

I'm just sharing scrumptious food for thought

What if

How would that
affect your
expectations
and work?

50% of your
favorite studies
had conclusions
that were just
wrong?

If the biomedical science has the most money and oversight, then

Their findings should be robust:

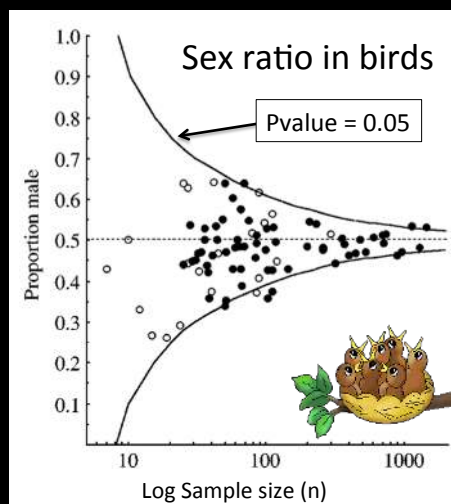
- Repeatable effect sizes
- The same across different labs
- The same across years

Publication replication failures

- **Biomedical studies**
 - Of 49 most cited clinical studies, 45 showed intervention was effective
 - Most were randomized control studies (robust design)
- **Mouse cocaine effect study, replicated in three cities**
 - Highly standardized study

Ioannidis 2005 JAMA; Lehrer 2010

Assessing reality using funnel plots



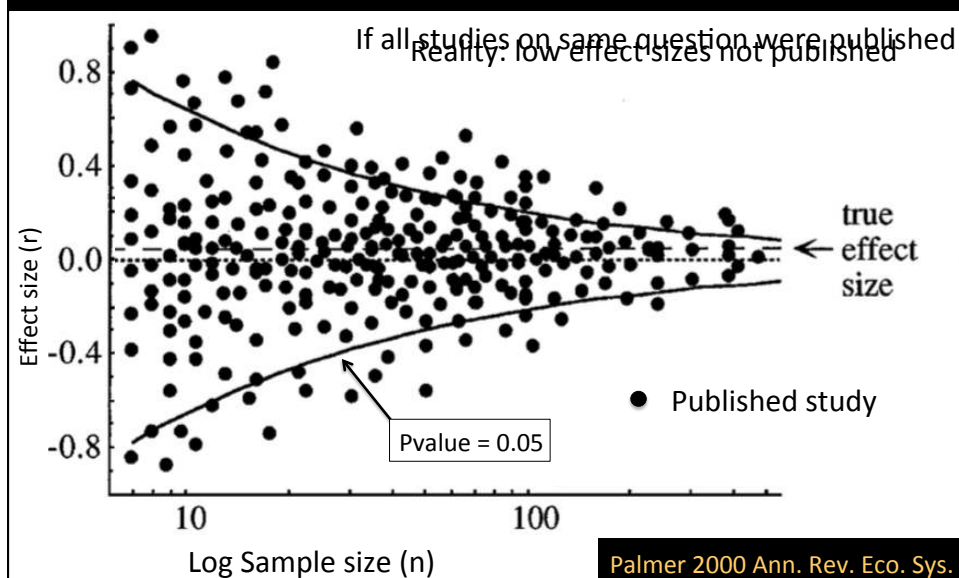
Small sample sizes affect measurement accuracy

Each dot = a study and has error

Study estimates are randomly distributed about the real value

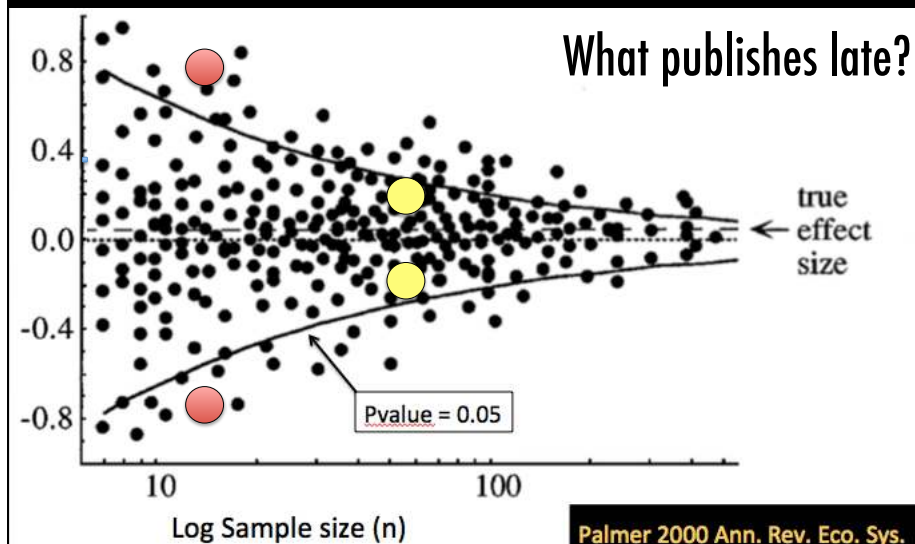
Your study is just a random estimate of some idealized value

Publication bias increases effect size



What if there is no replication?

What is most likely to publish first & where?



Why Most Published Research Findings Are False

A research finding is less likely to be true when:

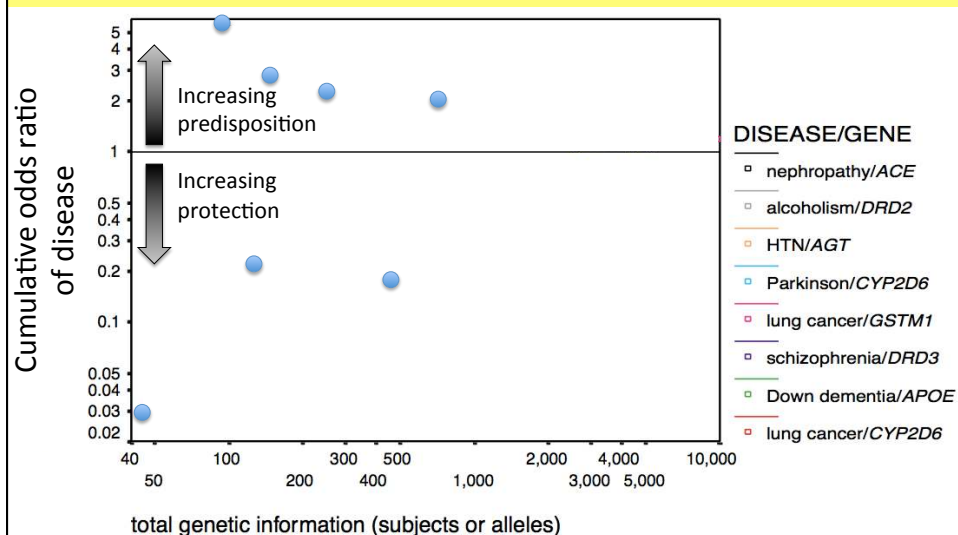
- ✓ the studies conducted in a field have a small sample size
- ✓ when effect sizes are small
- ✓ when there are many tested relationships using tests without *a priori* selection
- ✓ where there is greater flexibility in designs, definitions, outcomes, and analytical modes
- ✓ when there is greater financial and other interest and prejudice
- ✓ when more teams are involved in a scientific field, all chasing after statistical significance by using different tests

Ioannidis 2005 Plos Med.

But surely, this doesn't
apply to genomics

Or does it?

8 topics first reported with $P < 0.05$



Ioannidis, J. P., E. E. Ntzani, T. A. Trikalinos, and D. G. Contopoulos-Ioannidis. 2001. Replication validity of genetic association studies. *Nat Genet* 29:306–309.

There are lies, damn lies,
and

But wait, is that fair?

Are these really lies?

Where does this bias come from?

- Population heterogeneity
 - Space and time
- Publication bias
 - Large & significant effects publish fast and with high impact
 - Small & non-significant effects publish slow with low impact

Where does this bias come from?



YOU!!

And me All of us

Its arises from humans doing science

The way we think

The way our institutions work

Apophenia

A universal human tendency to seek patterns in random information and view this as important



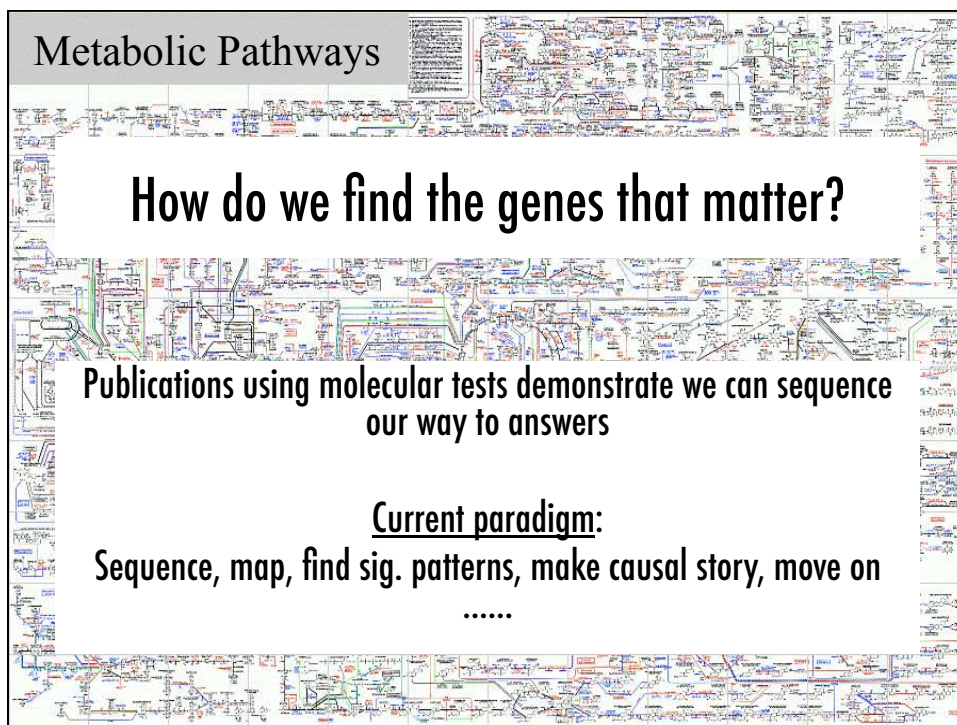
Story telling of Type 1 errors

Celebration of the false positives

Outline

- Are there biases understanding the genomic architecture of adaptations?
- What is the power of molecular tests of selection?
- What does the dissection of some classic comparative genomics study reveal?

Metabolic Pathways

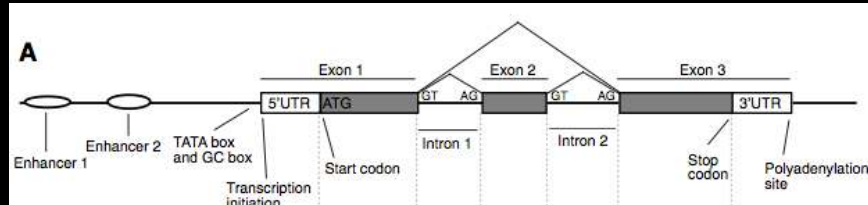


How do we find the genes that matter?

Publications using molecular tests demonstrate we can sequence our way to answers

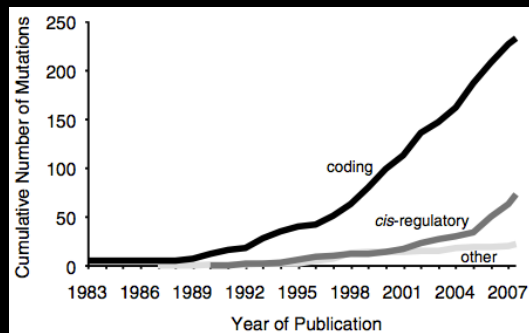
Current paradigm:
Sequence, map, find sig. patterns, make causal story, move on
.....

What is the architecture of a causal variant?



How predictable are adaptations?

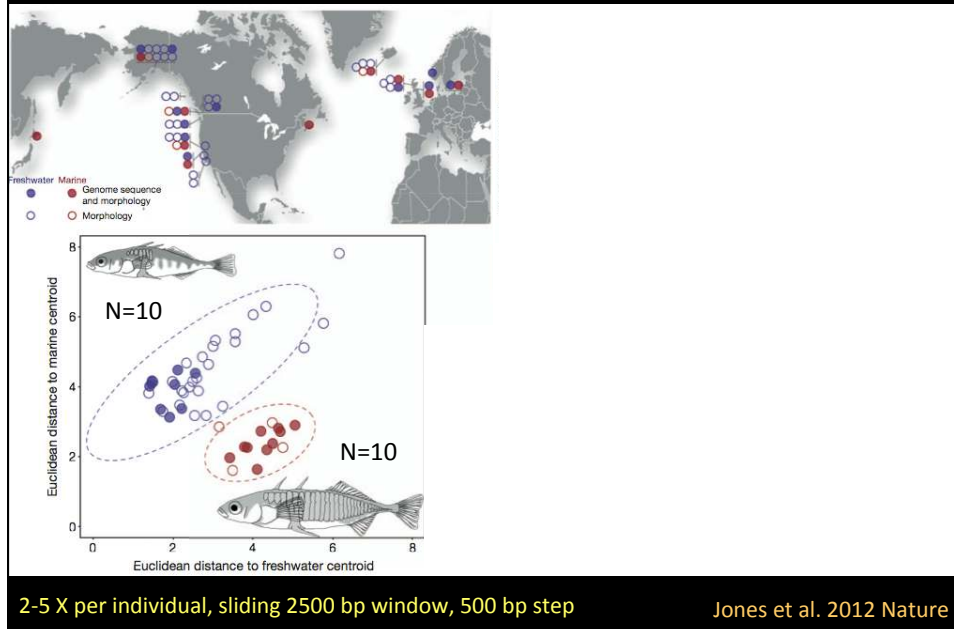
	Plants	Animals
Coding ¹	71	163
<i>Cis</i> -regulatory	26	48
Other ²	16	7
Total	113	218
Null ³	67	32



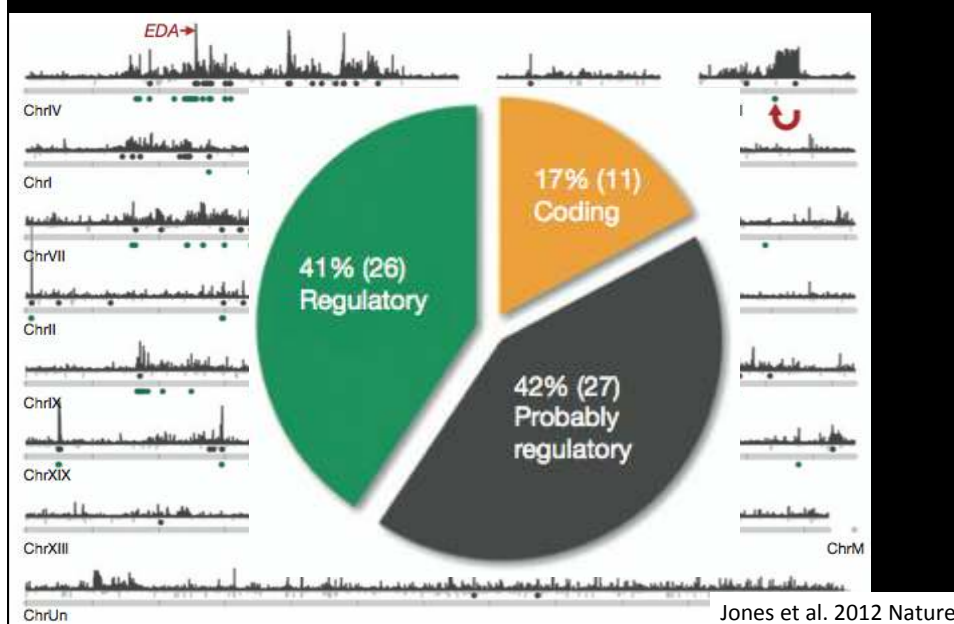
	Morphology	Physiology	Behavior
Coding ³	62	170	2
<i>Cis</i> -regulatory	43	29	2
Other ⁴	3	20	0
Total	108	219	4
Null ⁵	41	58	0

Stern & Orgogozo 2008 Evolution

Individual genome sequencing: powerful insights



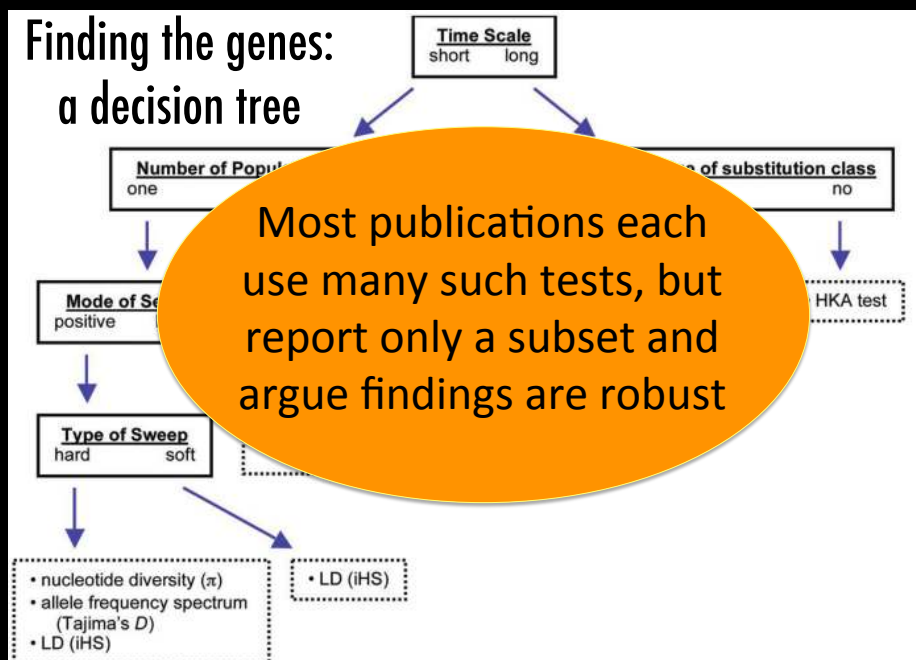
Which regions are more important? Coding or expression?



How do we identify the genes that matter?

- Molecular tests of selection are popular, but ...
 - What are their assumptions and power?
- What are these tests detecting?
 - What is a footprint of selection?
 - How are they formed?
 - How large are they?
 - How long do they last?

Finding the genes: a decision tree



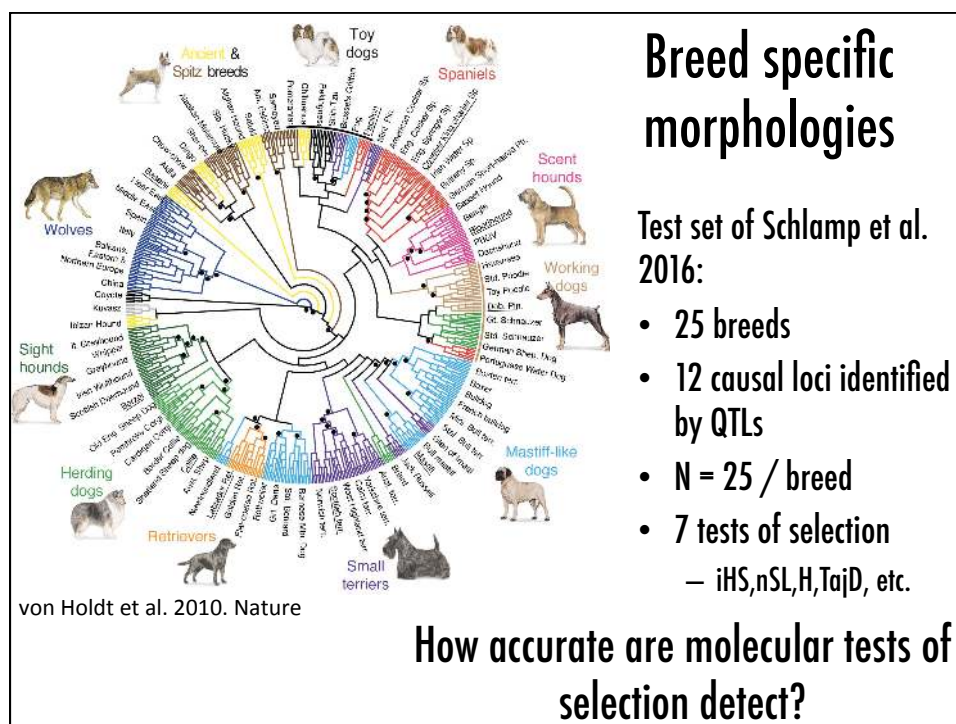
Hohenlohe et al. 2010 Int. J. Plant Science

What power do we have to detect evolution by natural selection?

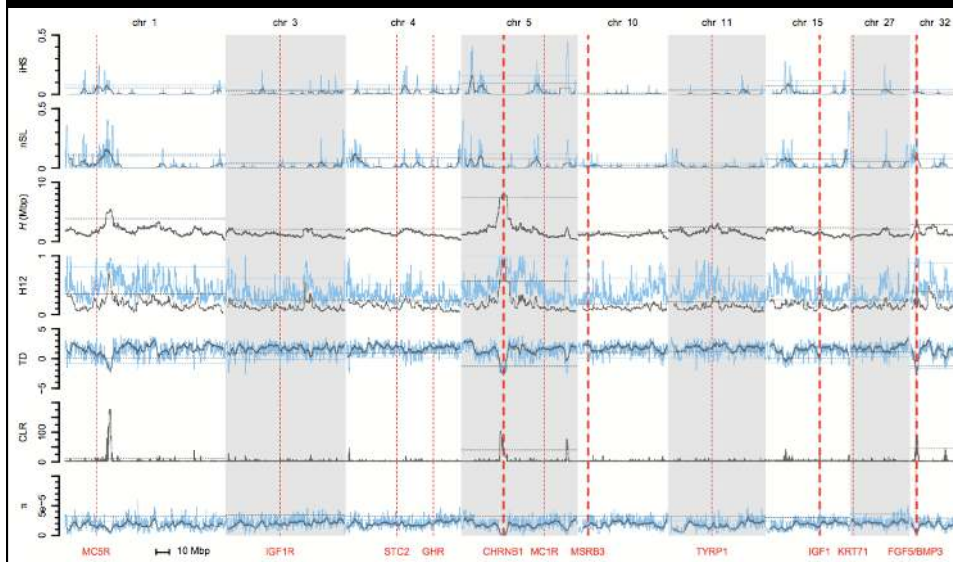
What is statistical power?

Power is the probability that the test will reject the null hypothesis when the alternative hypothesis is TRUE

Using a t-test, you would want power $> 90\%$ at reasonable sample size, right?



French Bulldog sample: low power (high type II error)



Schlamp et al. 2016. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Molecular Ecology* 25:342–356.

Why don't these tests have much power?

Biological reality
vs.
theoretical population genetics?

Directional selection: an example of the expectations of hard selection

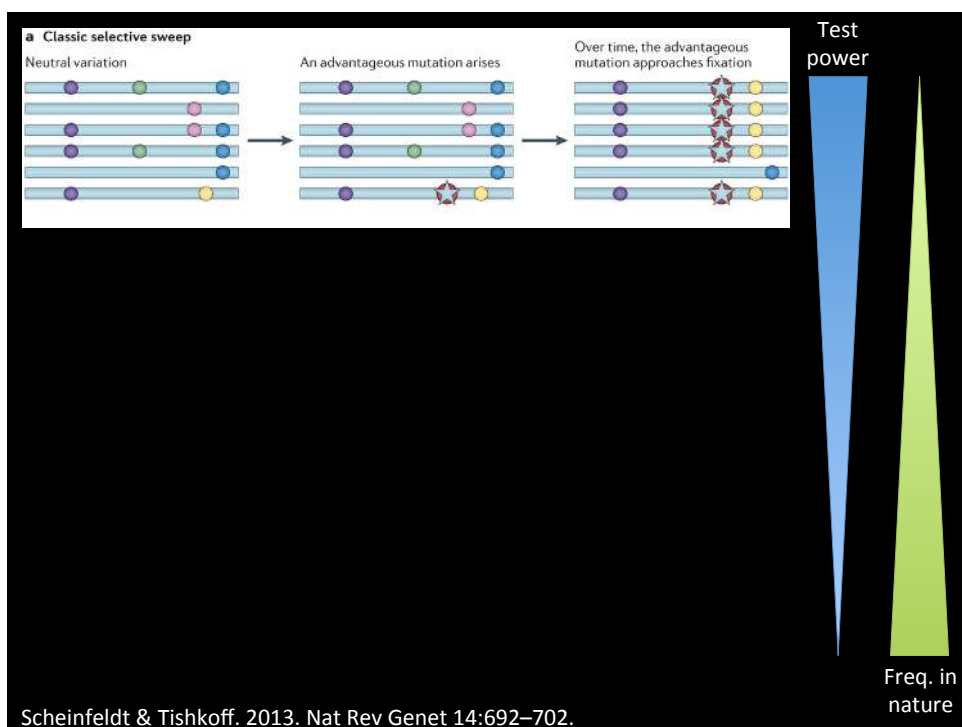
```

ATGGTAGGTCATATTGATCAGGGTGAATGTGCTAGAACATA
ATGCTAGATCAAAGTGATCATGGTGAATGTGCTAGAACATA
ATGGTAGATCAAATTGATCATGGTGCATGTGCTAGATCATA
ATGCTAGATCATATTGATGATGGTGAATGTGCTAGATCATA
ATGCTAGATCATATTGATCATGGTGAATGTGCTAGAACATA
ATGCTAGGTCATATTGATCATGCTGAAAGTGGTAGATCATA
  
```

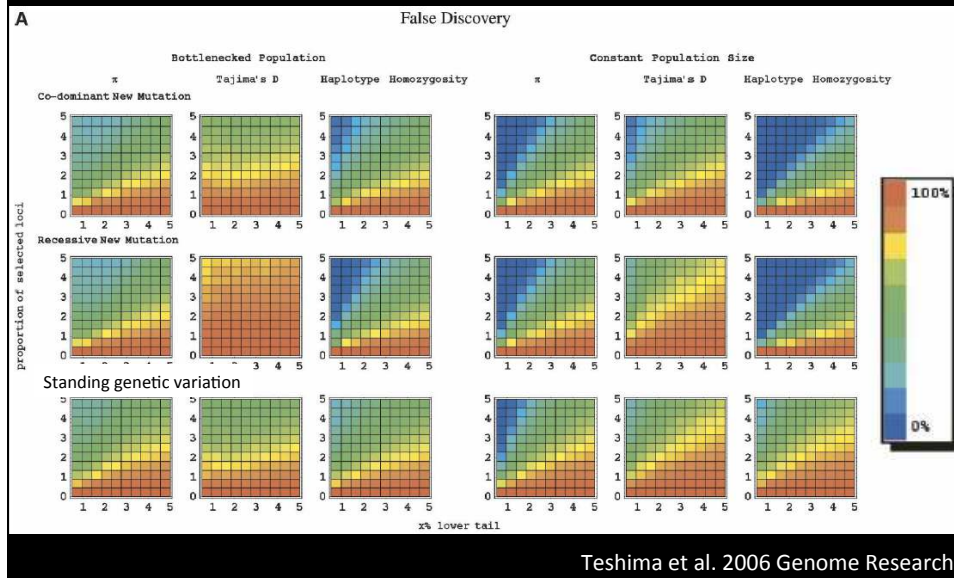
Population genomics has been dominated by developing methods to detect hard sweeps for past two decades

- But a proper 'null model' continues to be elusive, resulting in a high false positive rate since their inception

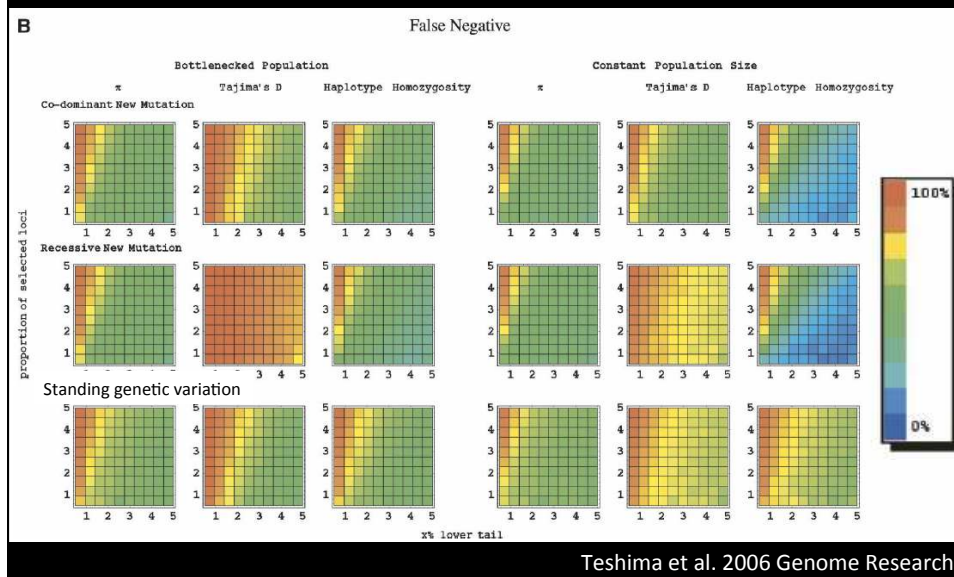
Storz 2005 Mol. Ecology



Estimate of error rates using , Tajima's D, and haplotype homozygosity under the models for a human population



Estimate of error rates using , Tajima's D, and haplotype homozygosity under the models for a human population



Simulation conclusions

- Simulations suggest
 - empirical approaches will identify several interesting candidates
 - But will also miss many—in some cases, most—loci of interest
- Power is lower when
 - directional selection involves a recessive rather than a co-dominant allele
 - when it acts on a previously neutral rather than a new allele
 - Demographic changes rather than constant population size

Genomic scans yield an unrepresentative subset of loci that contribute to adaptations

Molecular tests ...

BASED ON 20 YEARS OF PUBLICATIONS

- Are still chasing an elusive null model
 - Each performs better than previous ones under a specific set of conditions, all have poor null model
- But ... under realistic biological conditions, they all
 - Have very low power (high type II error rates)
 - Have high false positive rates

How common are hard sweeps in nature?

- “we argue that soft sweeps might be the dominant mode of adaptation in many species”

Messer and Petrov 2013 TREE

Certainly not everyone agrees



- This is an important read, critical of
 - assumptions underlying soft sweep (selection on standing variation)
 - the low power of molecular tests to detect hard & soft sweeps

How likely does natural selection use standing variation in your species?

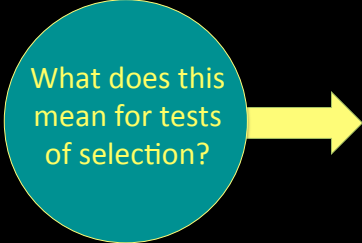
Thought experiment:

What fraction of species respond to selection in the lab?

Why?

If populations have variation, how likely is selection to use it?

What's likelihood of selection on standing variation in wild?



What does this mean for tests of selection?

We have not been studying the dominant form of selection in the wild & cannot reliably detect it

Age and type of selection matters

- Novel mutation, large effect, hard sweep that goes to fixation
 - Probability of detection 20 – 90%, depending on demography, etc.
- Old mutation and / or polygenetic that does not sweep to fixation
 - Probability of detection close to 0
- Finding the causal mechanism
 - Coding > expression (but allele specific expression can be lightening rod for expression)
 - SNPs > more complex mutations (indel, TE, CNV)
 - Ongoing gene flow & grouping by phenotype across replicate populations helps a lot
- Demographic effects
 - Nearly all species have experienced a major demographic change in the past 10,000 generations
 - Demographic change significantly reduces power and increases false positive rates.
- What is the relative frequency of these?
 - What will be the architecture of your phenotype?
 - What does your method have the highest power to detect?

Get ready, here come the 1000ⁿ genomes

- Roughly 20 arthropods sequenced to date
— plans to sequence many more
- Many other large genomes being sequenced

An unprecedented opportunity for large scale errors?

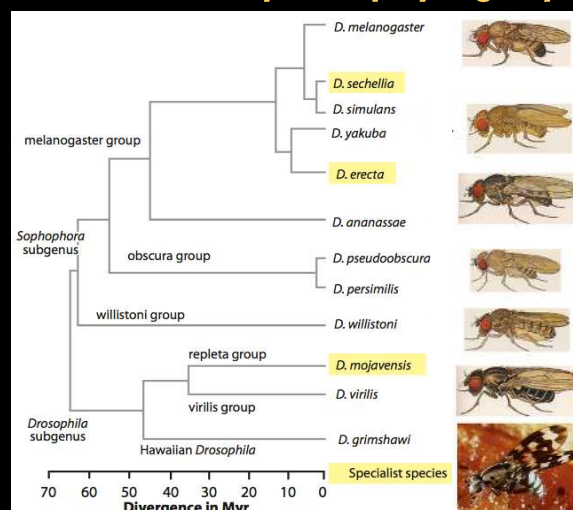
Studying: evolutionary relationships

- Genome evolution
- Functional insights into genes and genomic features (e.g. regulation and inheritance)



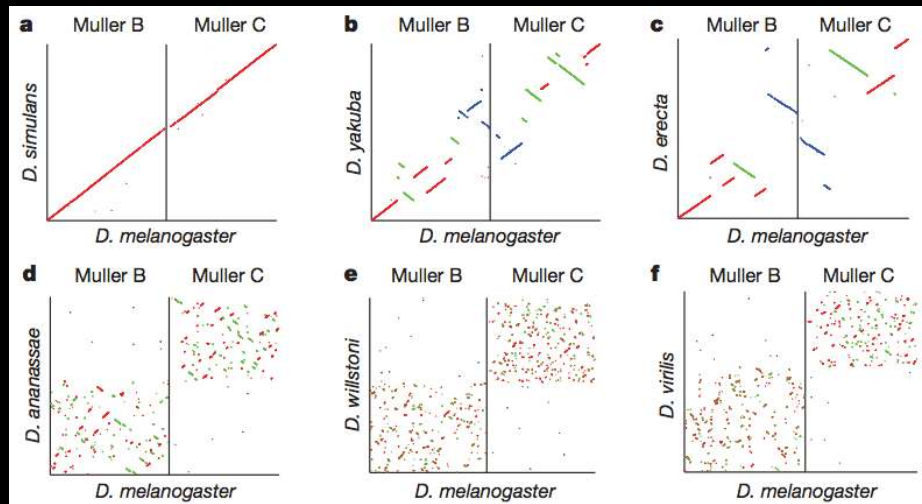


Classic study: Evolution of genes and genomes on the *Drosophila* phylogeny



Drosophila 12 Genomes Consortium 2007 Nature

Tempo and mode of chromosome evolution

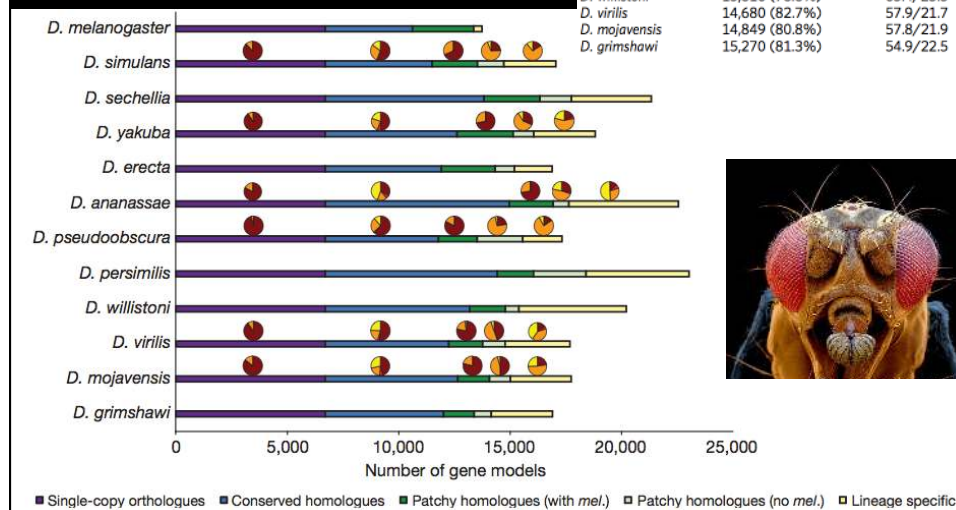


- > 20 My, chromosomal order completely reshuffled in Diptera

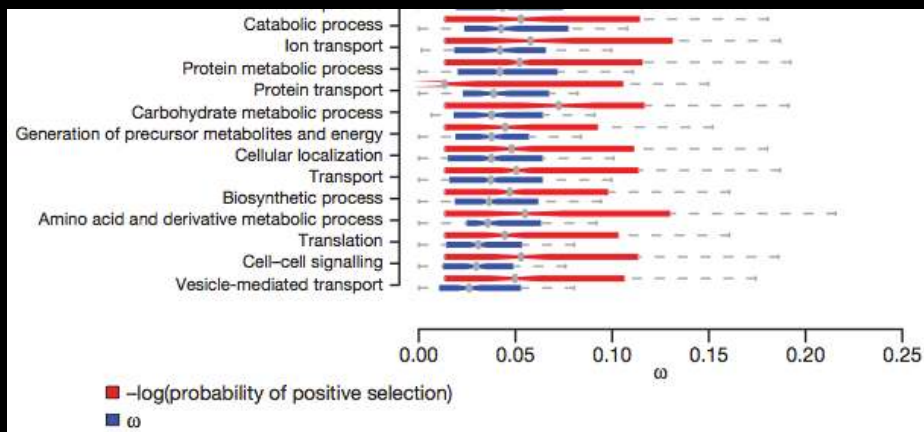
Drosophila 12 Genomes Consortium 2007 Nature

Genome evolution

Drosophila 12 Genomes Consortium 2007 Nature



Selection dynamics across functional categories

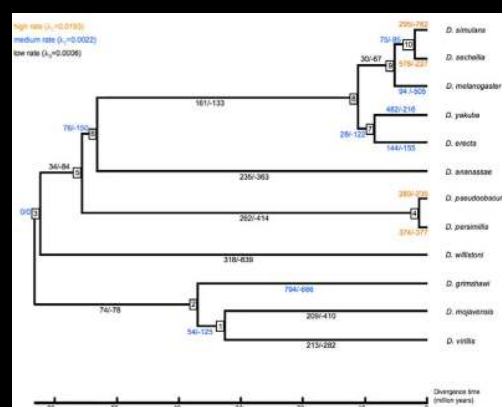


- 33.1% of single-copy orthologues have experienced positive selection on at least a subset of codons.

Drosophila 12 Genomes Consortium 2007 Nature

Gene Family Evolution across 12 Drosophila Genomes

- One fixed gene gain/ loss across the genome every 60,000 yr
- 17 genes are estimated to be duplicated and fixed in a genome every million years



Drosophila 12 Genomes Consortium 2007 Nature
Hahn et al. 2007 Plos Genetics

Comparative Genomics : a house of cards?

- Data scale is too large to thoroughly assess errors ...
 - Perhaps the findings are just wrong
- All conclusions, at some stage, rest upon
 - Simple bioinformatics
 - Assumptions that get incorporated into seemingly unbiased methods



Lets exploring two pillars of these studies, their error and repercussions

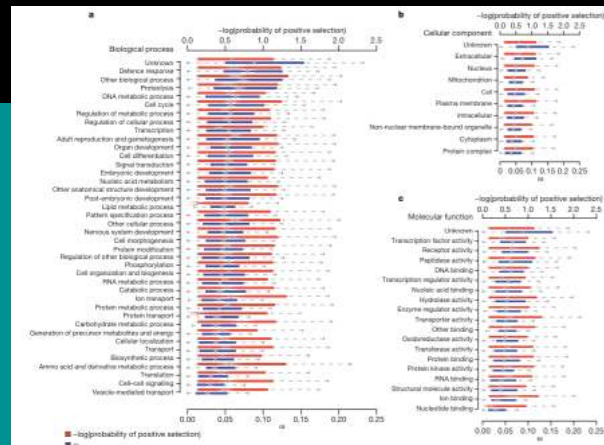
- Gene alignments in detecting positive selection
- Calibrations in temporal analysis

Published studies allow ...

Follow up studies to reveal limitations

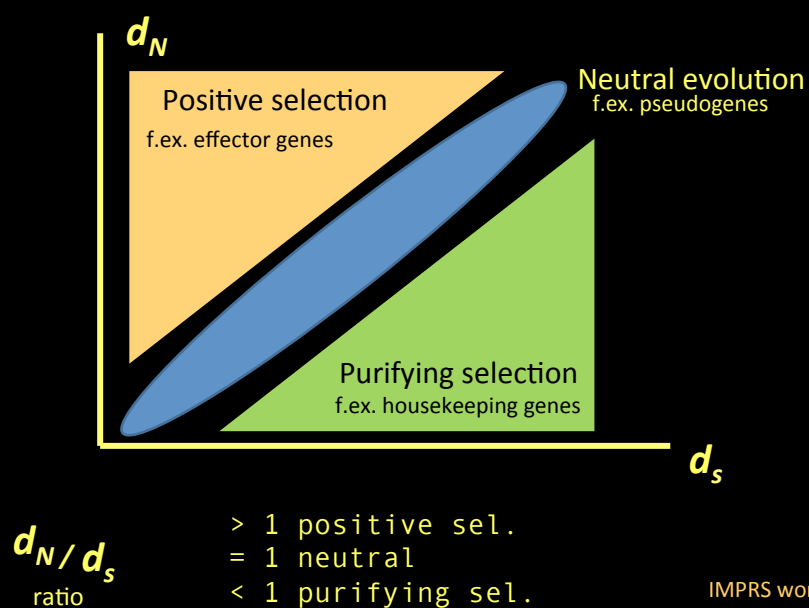
But, must have enough details to be
repeatable

Genome-wide selection dynamics:



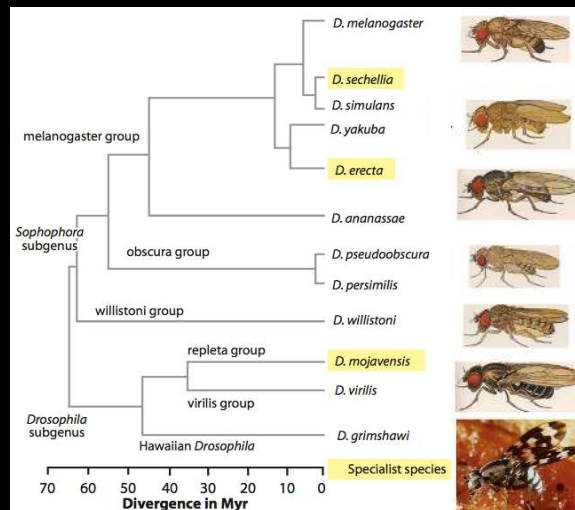
How robust are these conclusions?

Codon based tests of selection



IMPRS workshop,
Comparative Genomics

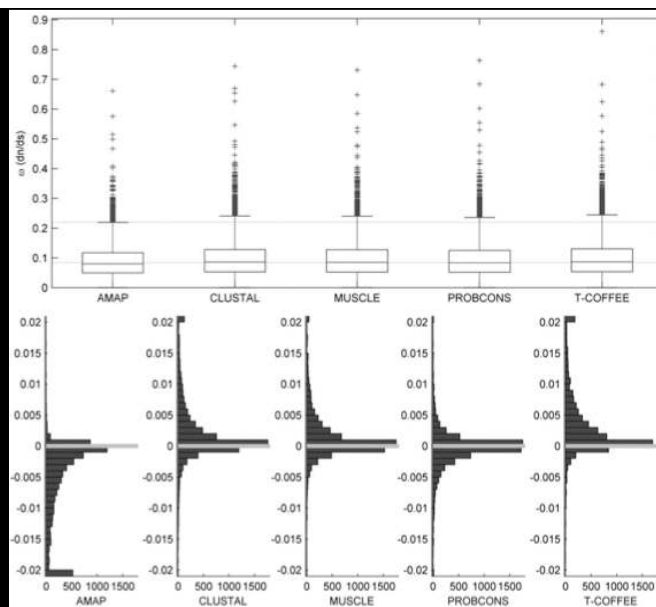
Evolution of genes and genomes on the *Drosophila* phylogeny



Drosophila 12 Genomes Consortium 2007 Nature

dN/dS estimates by aligner

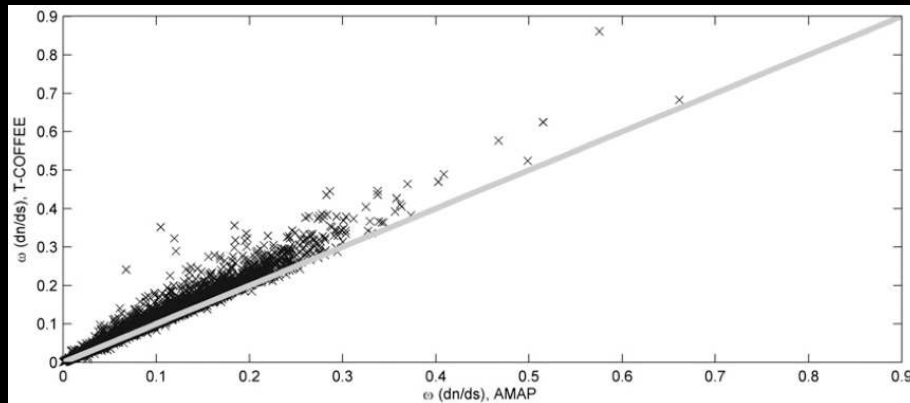
- 6690 orthologs
- 5 alignment methods
- Alignment methods affect dN/dS estimates



Markova-Raina & Petrov 2011 Genome Biology

Comparing results across methods is responsible
bioinformatics!!!!

Since we can't look at our data, we need approaches that
allow 1st principal assessments



Markova-Raina & Petrov 2011 Genome Biology

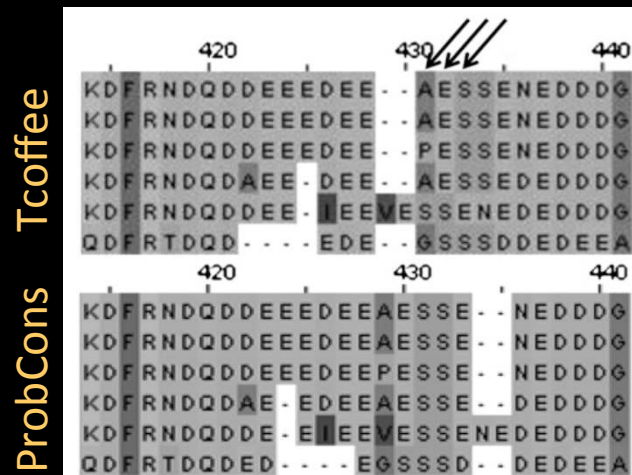
Aligner has a larger effect than
biological signal

Aligner	12 genomes, M7/8		12 genomes, M1a/2a		12 genomes, M7/8, with removed gaps		Melanogaster group, M7/8	
	95% (a)	99% (b)	95% (c)	99% (d)	95% (e)	99% (f)	95% (g)	99% (h)
AMAP	817	213	256	110	558	104	973	257
MUSCLE	1043	306	379	192	764	155	1134	366
ProbCons	1013	281	346	180	801	182	1128	371
T-Coffee	1290	479	612	353	824	173	1248 (909)	463 (218)
ClustalW	902	261	244	117	666	112	1269	453
Total in 5	1902	673	799	441	1562	384	1737 (1723)	652 (620)
PRANK	468	49	49	16	258	42	581	70

Markova-Raina & Petrov 2011 Genome Biology

Alignment results highlight importance of alignment score!

- Tcoffee finds 3 selected sites indicated by arrows
- ProbCons identifies region with low alignment score, not used



Markova-Raina & Petrov 2011 Genome Biology

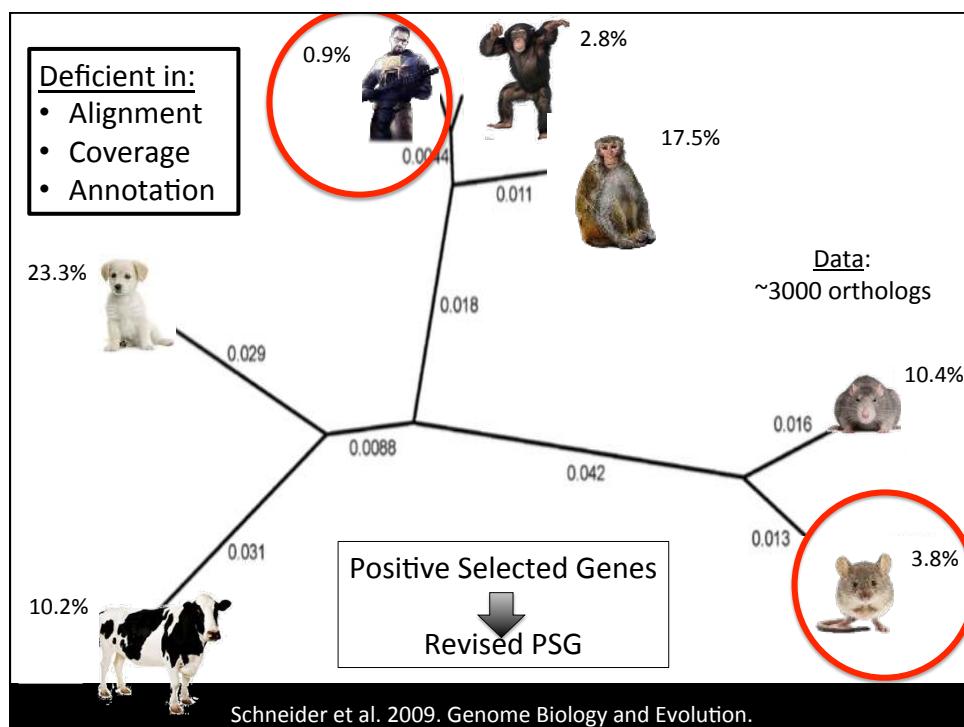
What about recent genomes?

Surely they are better?

and mammals ... they have good genomes

and alignment problems rarely happen

... right?



Post-genomics challenge

"What we can measure is by definition uninteresting and what we are interested in is by definition unmeasurable"

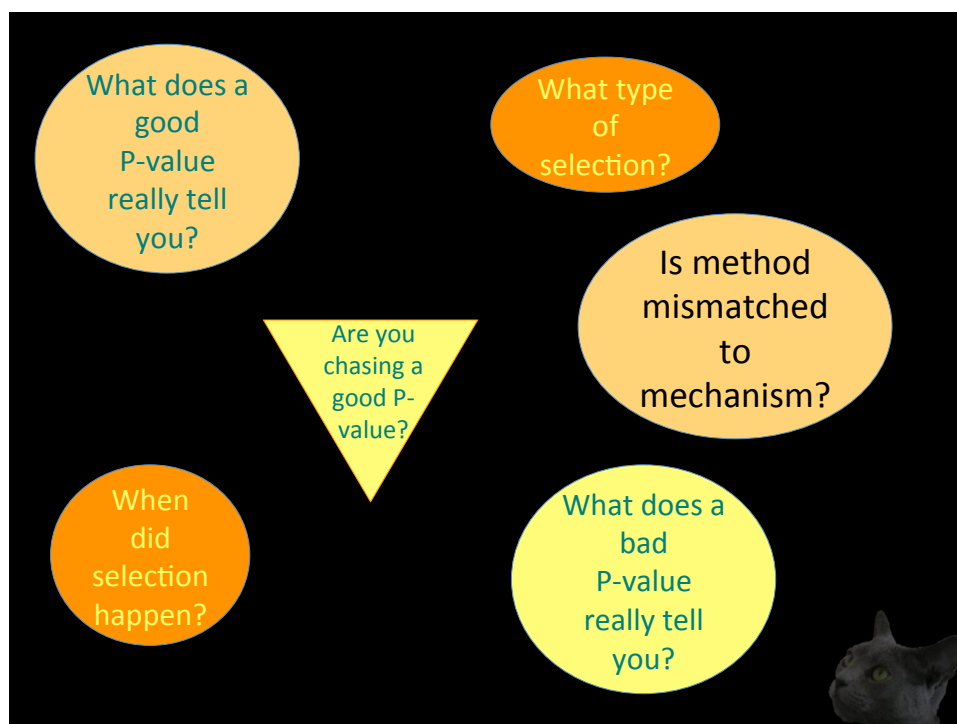
- Lewontin 1974

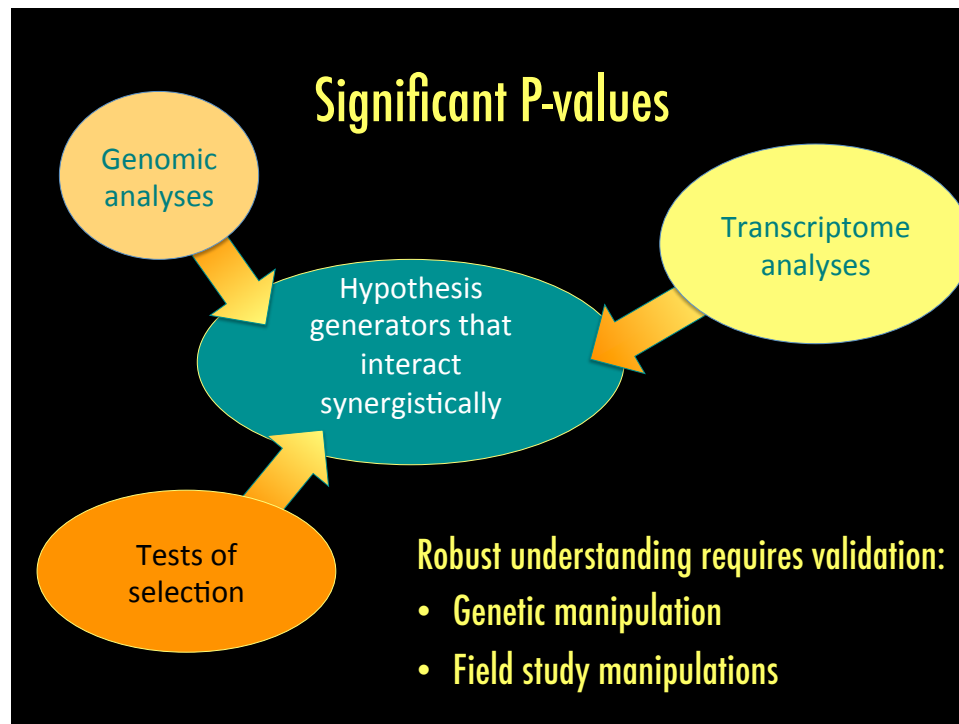
"What we understand of the genome is by definition uninteresting and what we are interested in is by definition very damn difficult to sequence and assemble and annotate and analyze at genomic scale"

- Wheat 2015

For example:

- indels & inversions
- gene family dynamics
- evolutionary dynamics





Goal of this lecture

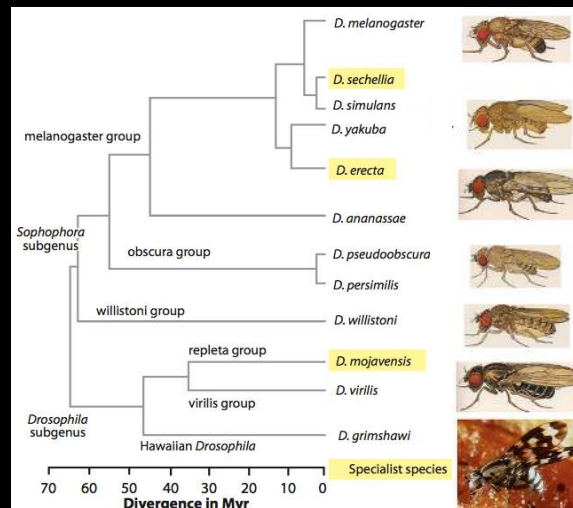
- Present a non-typical view of ecological genomics
 - So you have a more complete view of the field
- Make you uncomfortable
 - Provide a context for understanding your results
- Encourage you to rethink the reality presented by publication biases
 - Overcoming this bias is a continual challenge



Temporal inference:

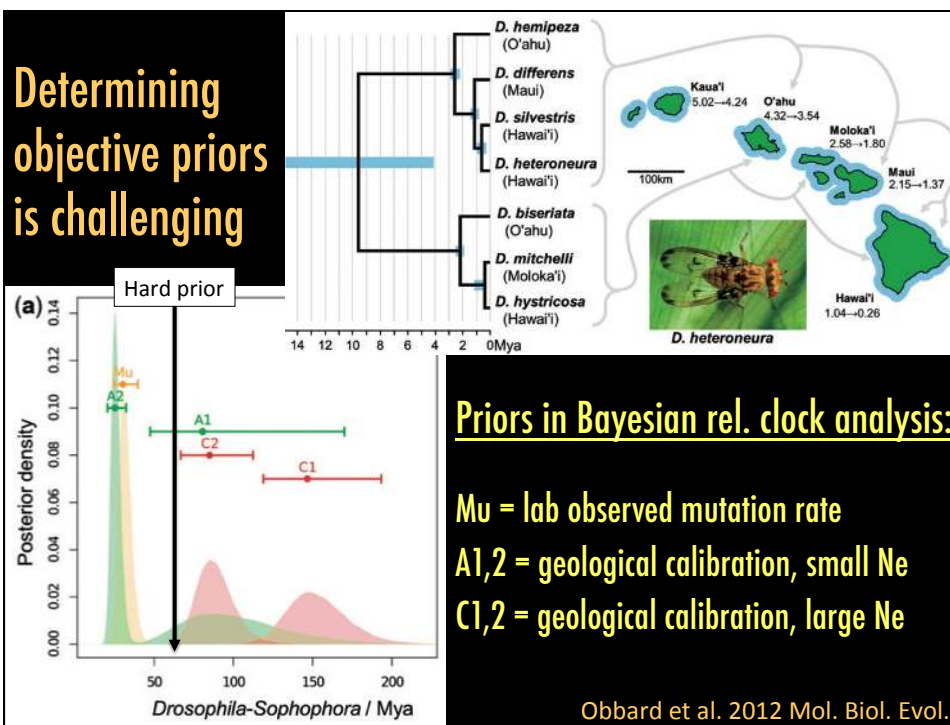
fact or fiction?

Evolution of genes and genomes on the *Drosophila* phylogeny



Drosophila 12 Genomes Consortium 2007 Nature

Determining objective priors is challenging



Priors in Bayesian rel. clock analysis:

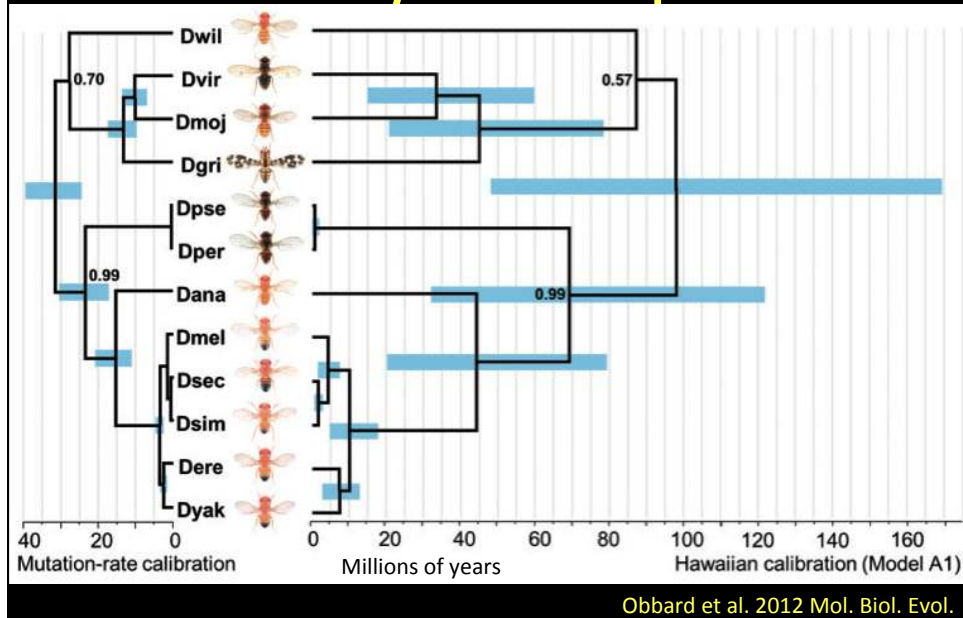
μ = lab observed mutation rate

A1,2 = geological calibration, small N_e

C1,2 = geological calibration, large N_e

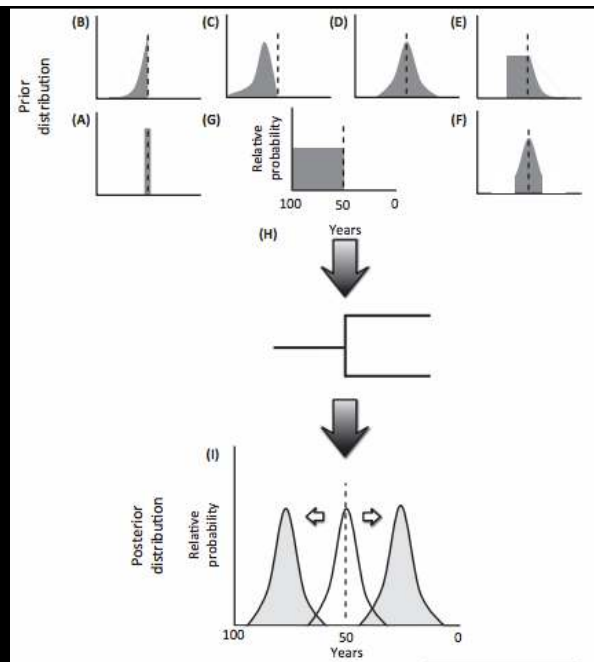
Obbard et al. 2012 Mol. Biol. Evol.

Priors directly influence posteriors



Prior distributions matter

- Integrative science is challenging
- Discuss or collaborate with experts to evaluate your approach.



Wheat and Wahlberg 2013 Trends Ecology & Evolution

How do we gain dating confidence when we are in the dark?

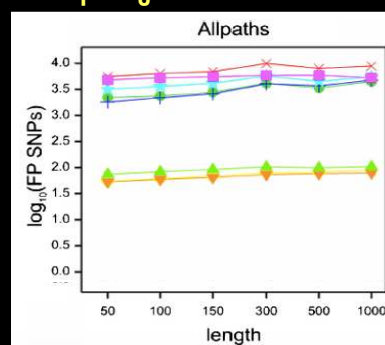
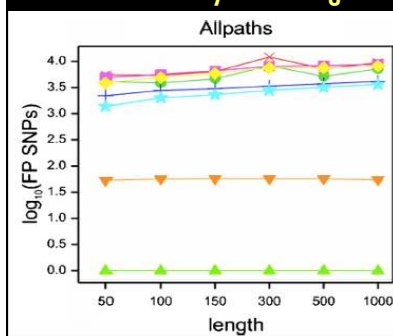
- Fossils and DNA are likely to rarely agree
- How can we assess the temporal signal in the DNA in a robust manner?
 - Reducing prior biases and using lots of DNA data, while modeling likely violations of analysis models



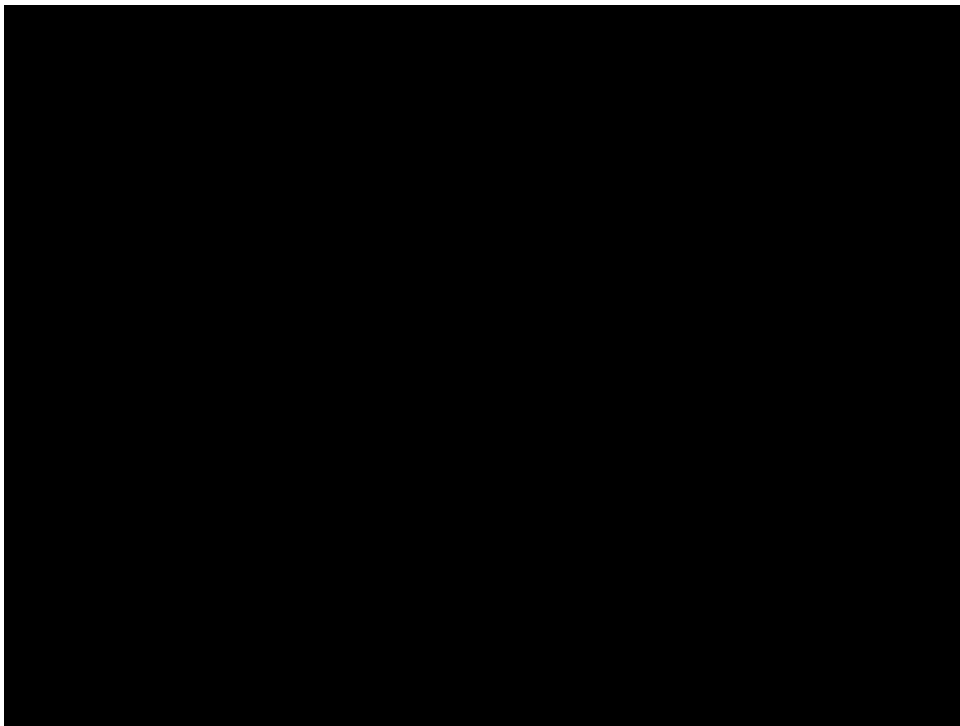
Wheat and Wahlberg 2013
Trends Ecology & Evolution

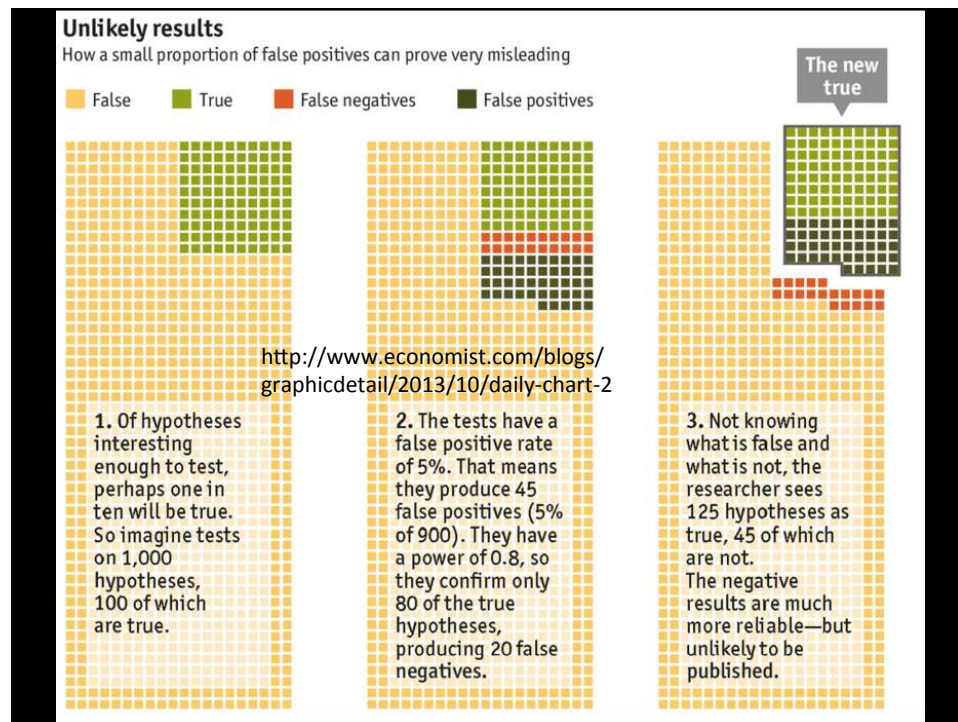
•

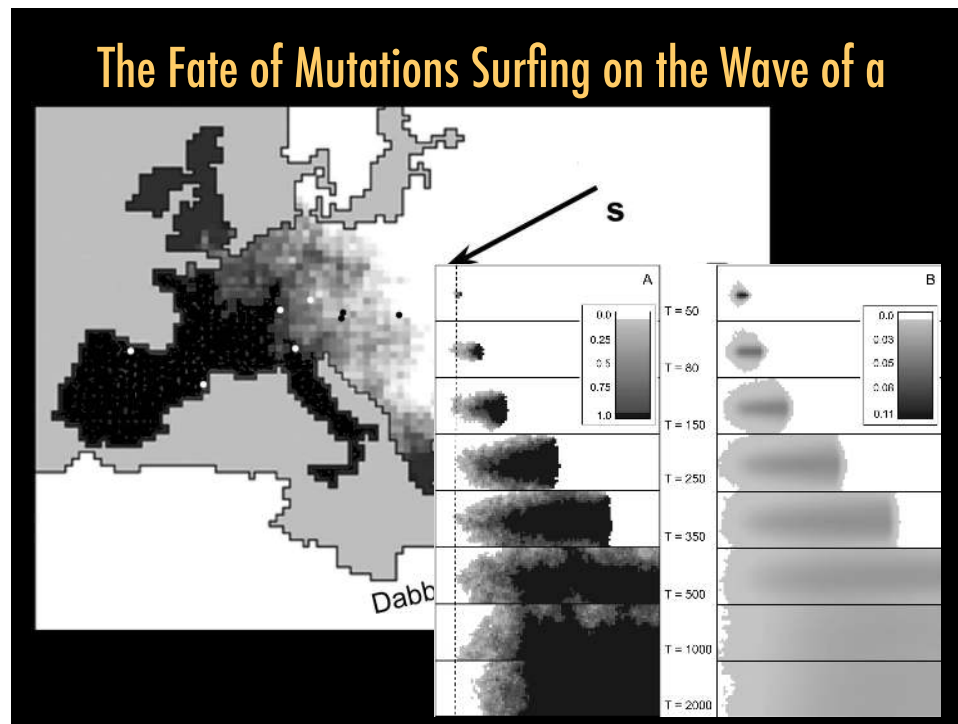
- 1000's of false positive SNPs (FP SNPs) result from misassembly x mapping x calling
- Genome was small (~ 125 Mbp) with few repeats (*Arabidopsis thaliana*)
- FP rates likely much higher with larger, more complex genomes



Ribeiro et al. 2015 BMC
Bioinformatics





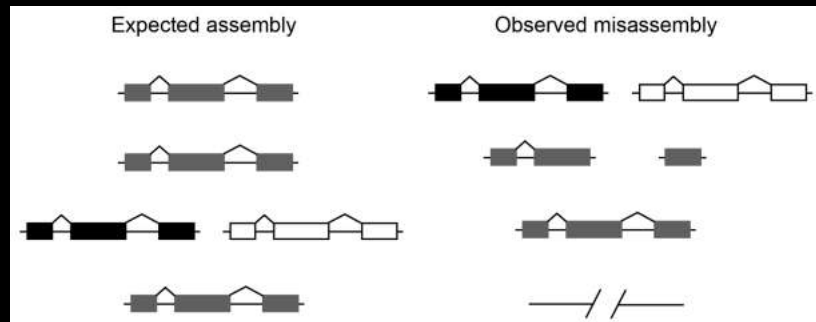


JOURNAL OF NEGATIVE RESULTS

- ECOLOGY & EVOLUTIONARY BIOLOGY -

Now handling genomic data

<http://www.jnr-eeb.org/index.php/jnr>



Microevolution effects

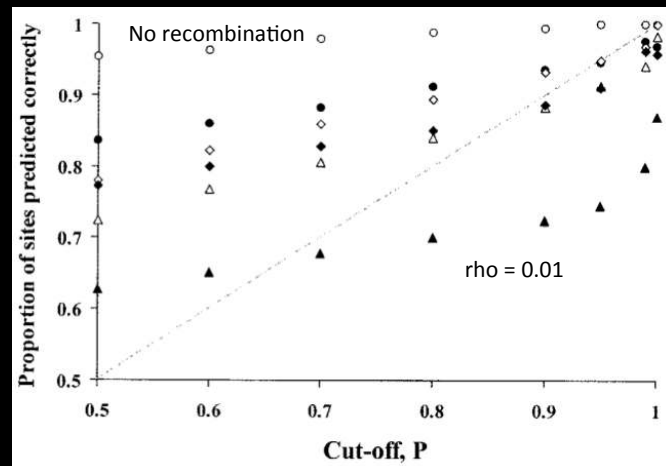
Previous examples were at deep evolutionary time scales

Surely such problems don't exist at the within genera level Right?

Recombination violates dN/dS tests

Codeml
inferred
selection:

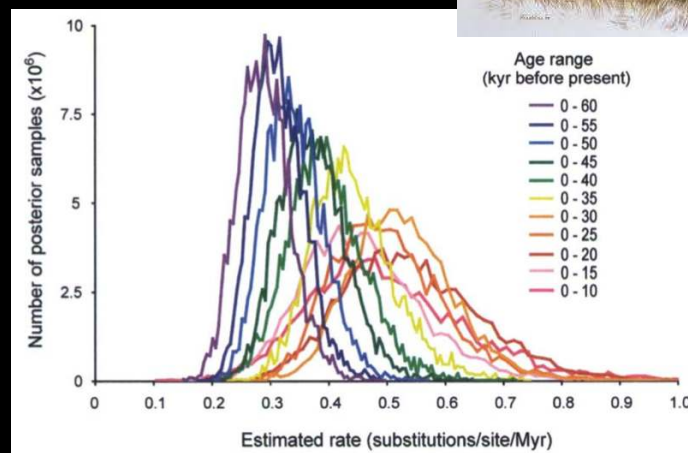
False
positives can
increase to
over 30%



- 13% of sites simulated at $\omega = 2.5$
- Sample size = 30 sequences

Anisimova 2003 Genetics

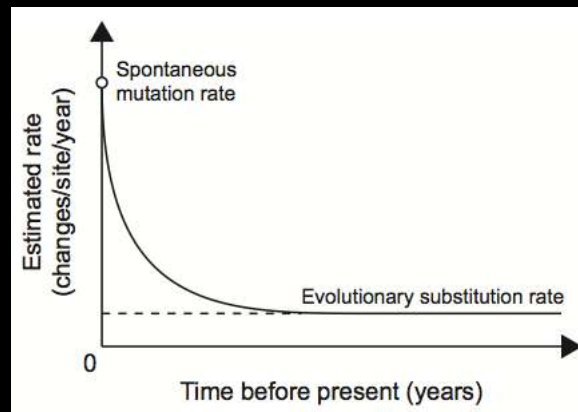
Posterior distribution estimates of substitution rates from mitochondrial control region from Beringian bison



Ho et al. 2007 Systematic Biology

Time dependent rates of molecular evolution

Significant implications for phylogeographic studies that use fixed rates to assess demographic with environmental change



Ho et al. 2011 Molecular Ecology

What power do we have to detect balancing selection?

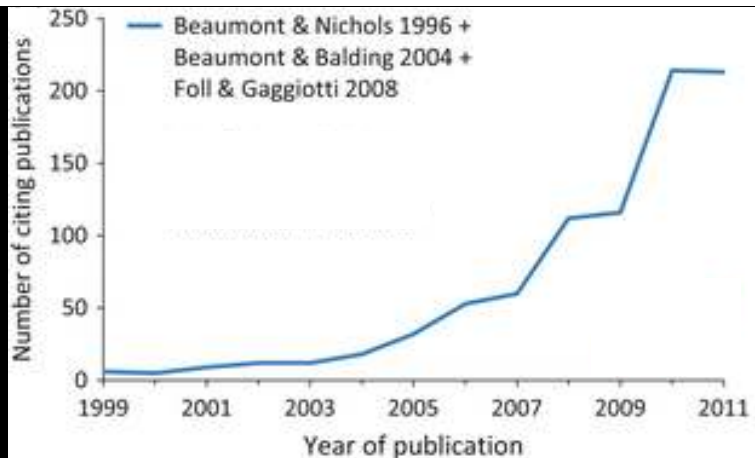
ρ	Width of window (bp)				
	25	50	100	200	1000
1	85.6	90.2	92.8	93.5	83.8
3	80.8	85.3	86.3	83.5	44.7
10	69.0	69.9	64.5	51.0	4.1
30	48.1	42.5	31.0	15.7	0.1
100	20.5	15.6	8.9	2.4	0.0

Tajima's D
% finding selection of 5000 simulations

- For *Drosophila melanogaster*, power = 50% with window size of 200 bp, using 24 diploid individuals.
- For species with larger population size, power likely lower
- Recombination and gene conversion destroy 'footprint' rather quickly

Nordborg and Innan 2003 Genetics

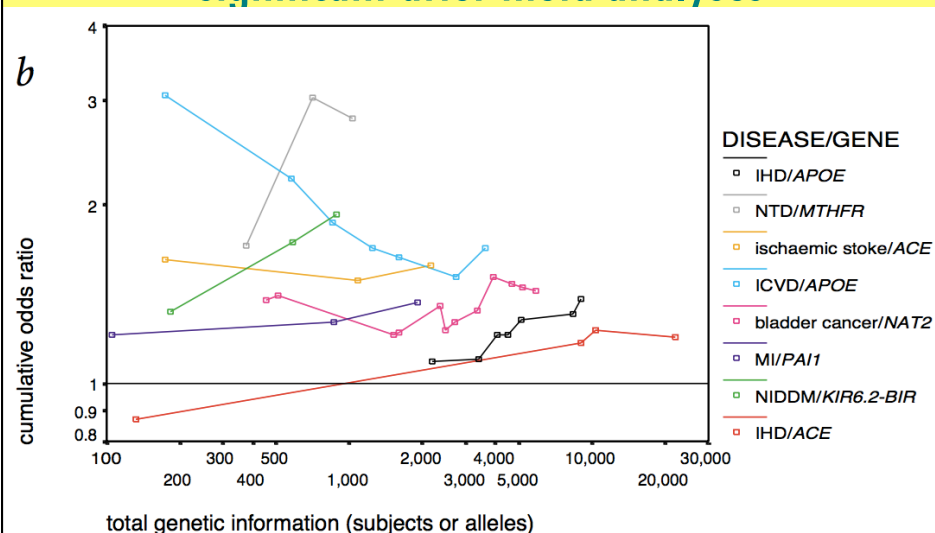
Fst outlier analyses are common



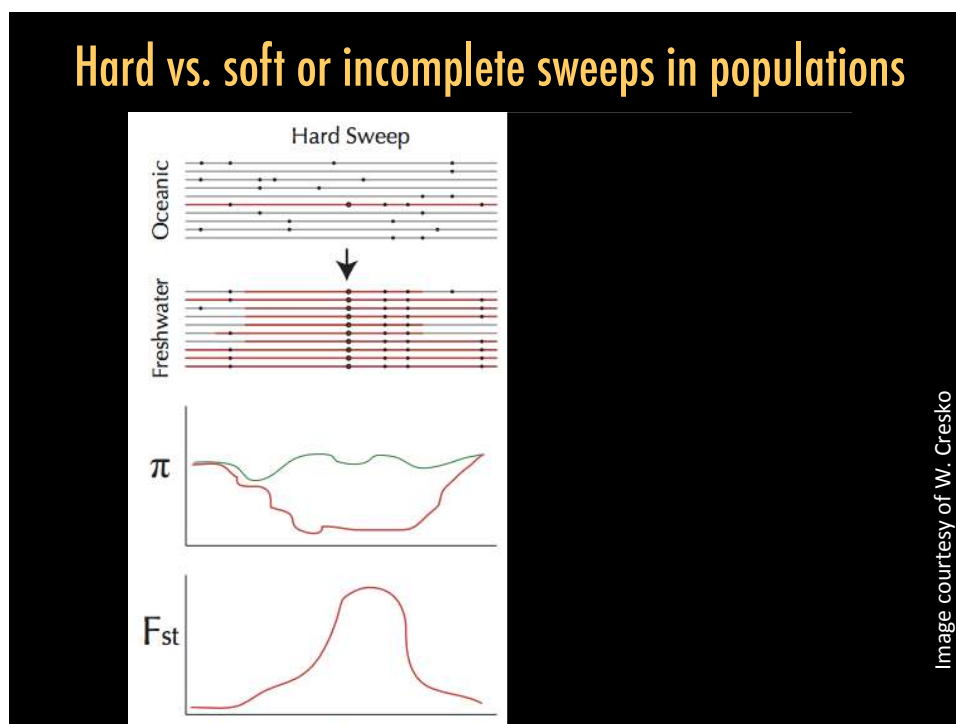
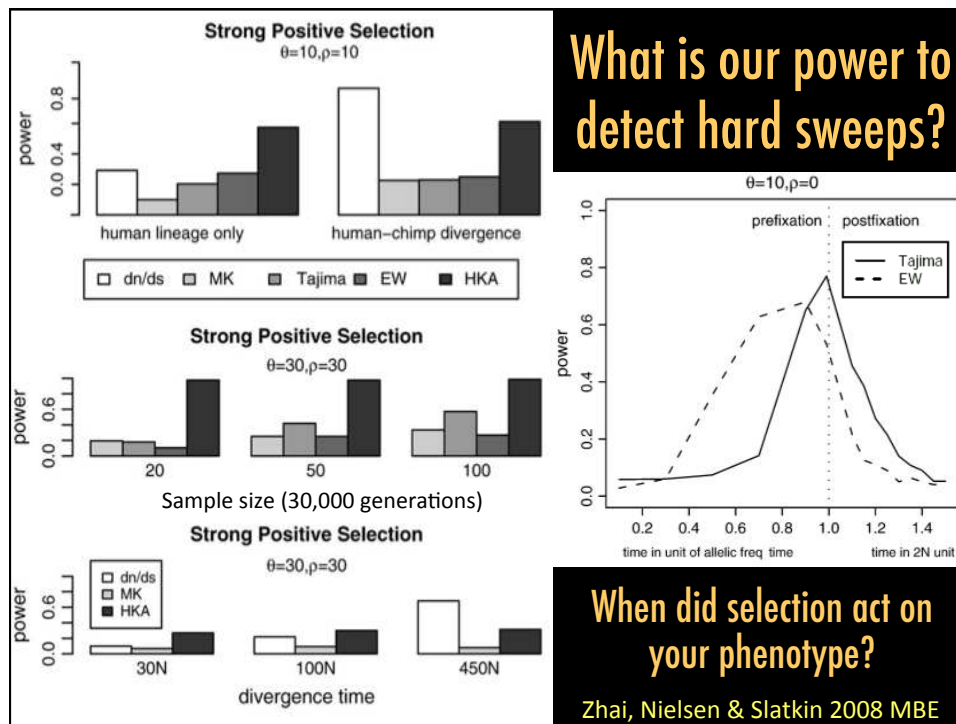
Pervasive selection or is it...? why are *FST* outliers sometimes so frequent?

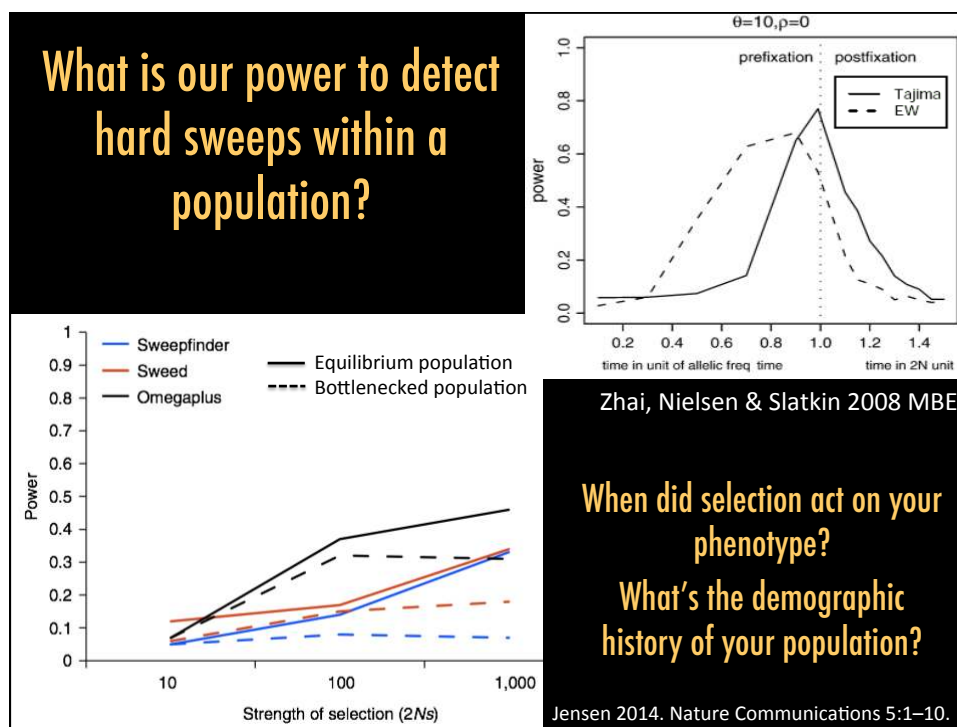
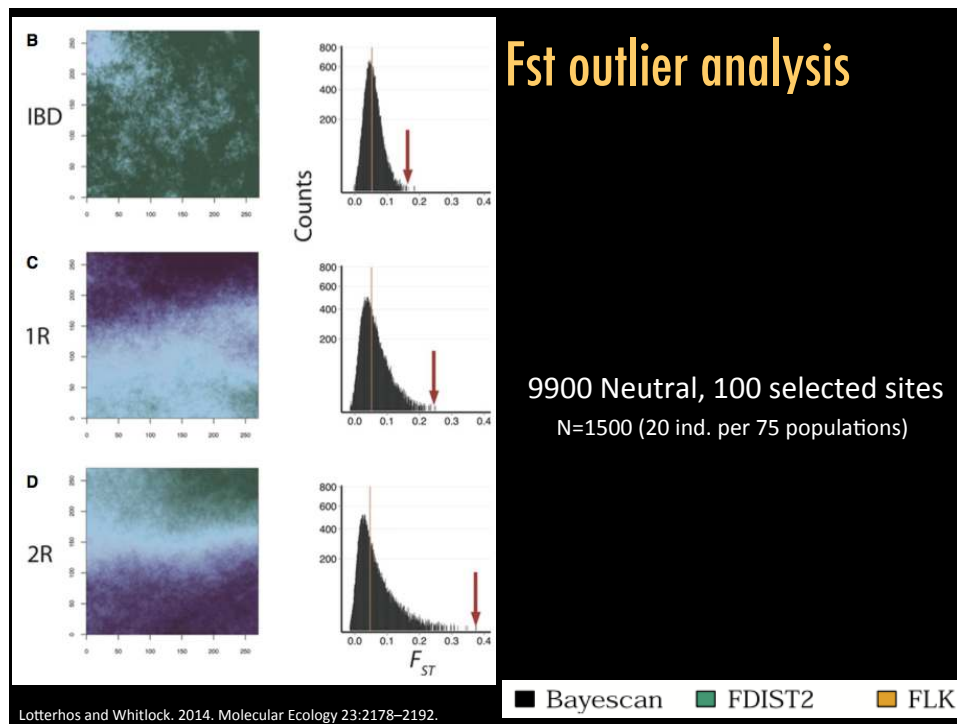
Bierne et al. Molecular Ecology 2013

8 topics where first study $P > 0.05$, but became significant after meta-analyses



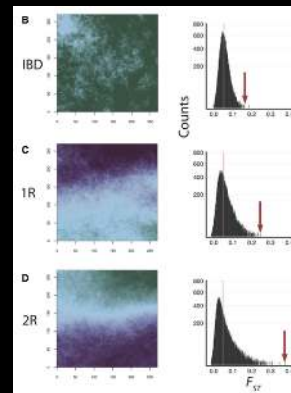
Ioannidis, J. P., E. E. Ntzani, T. A. Trikalinos, and D. G. Contopoulos-Ioannidis. 2001. Replication validity of genetic association studies. Nat Genet 29:306–309.



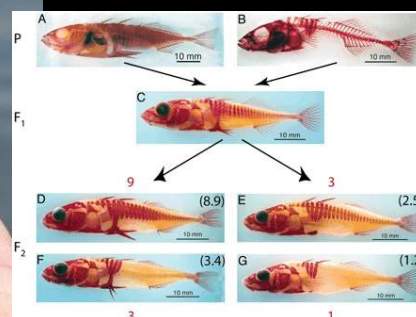
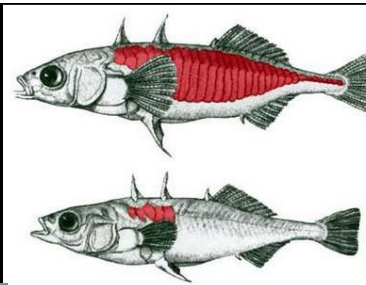
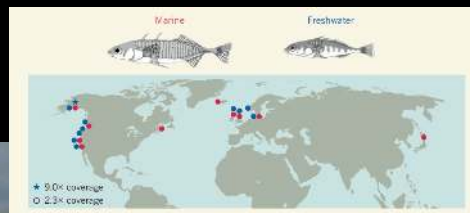


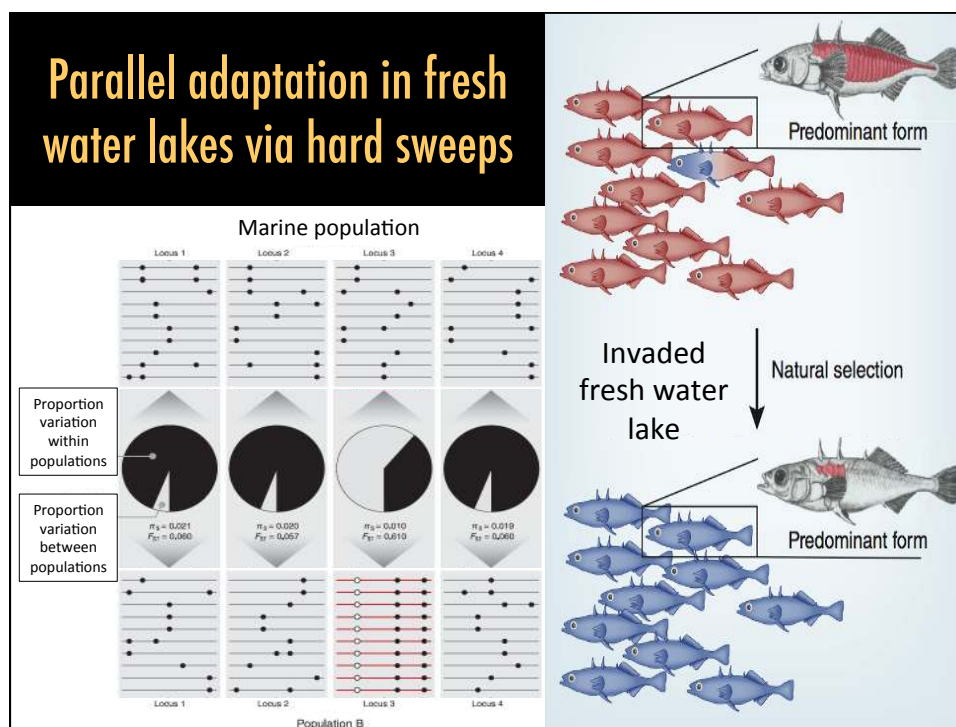
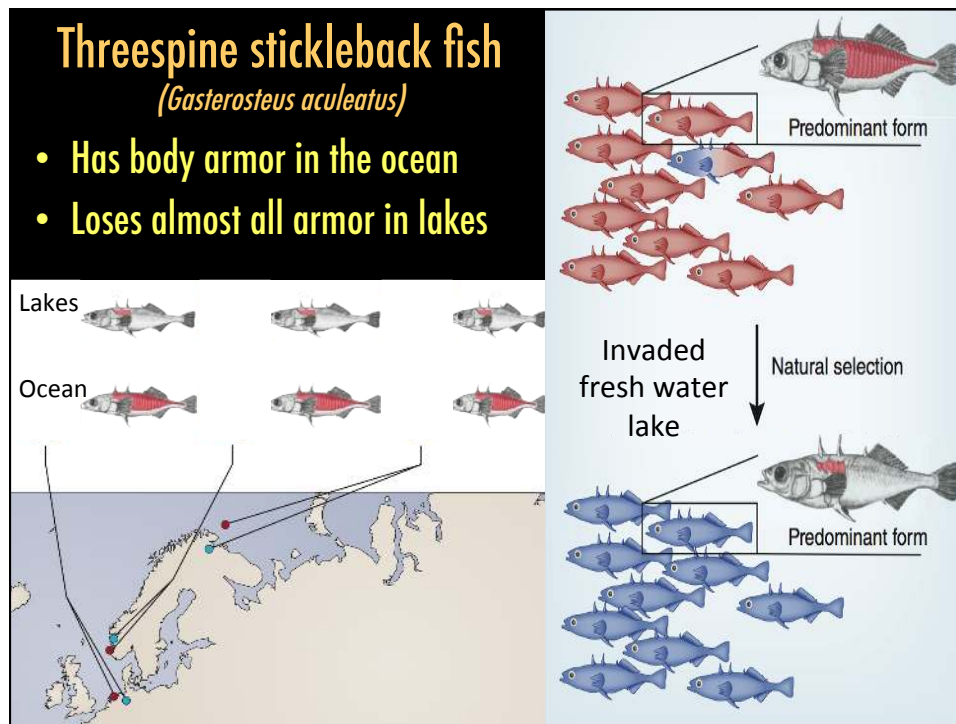
What's a good way to assess molecular tests?

- Computer simulations of evolution
 - Across range of demographic scenarios
- What else?
- Testing them on real data where we know the targets of selection = real world validation
 - Which ones work and when
 - We could then use this to make better tests, right? (very rare)



Hard selection case example: threespine stickleback fish



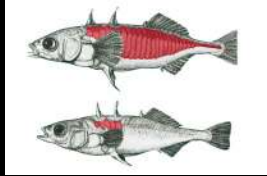


Non – adaptive



disease, aging, height, etc.

Adaptive



salinity, color, resistance, etc.

generally ...

1000's of loci, each of small effect size

One or several loci of large effect

Is this a publication bias?

Will your trait have 1000's of small effect genes, or a few genes of large effect?

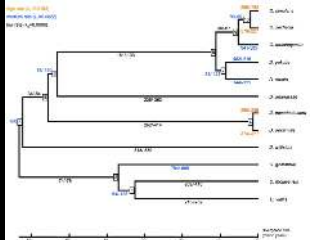
Sear (2010) ... Is bigger always better?

Rockman (2011) ... All that's gold does not glitter

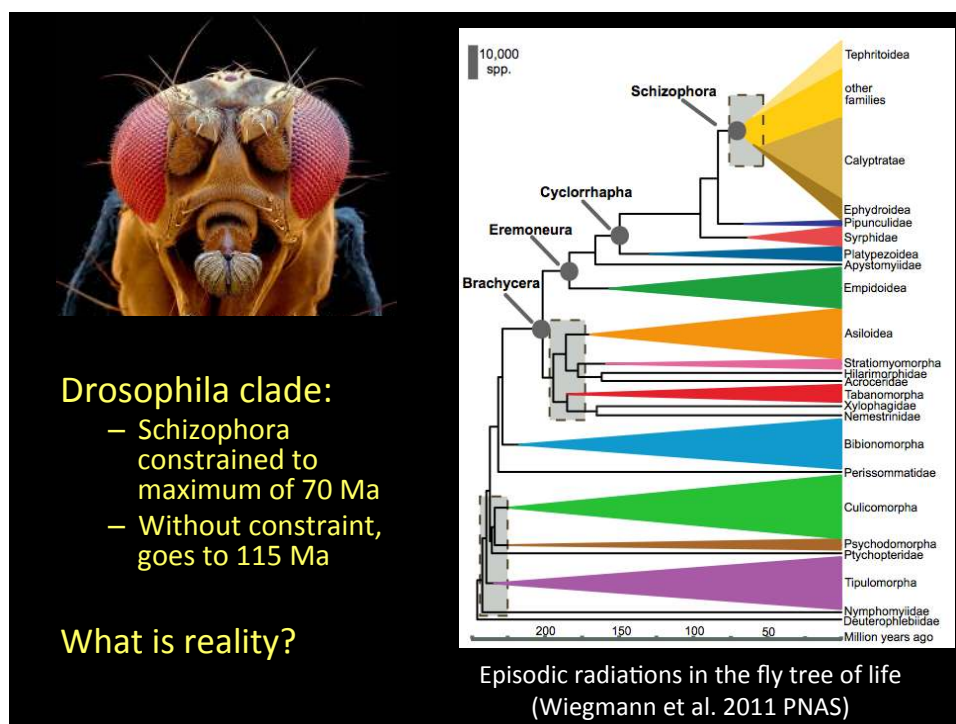
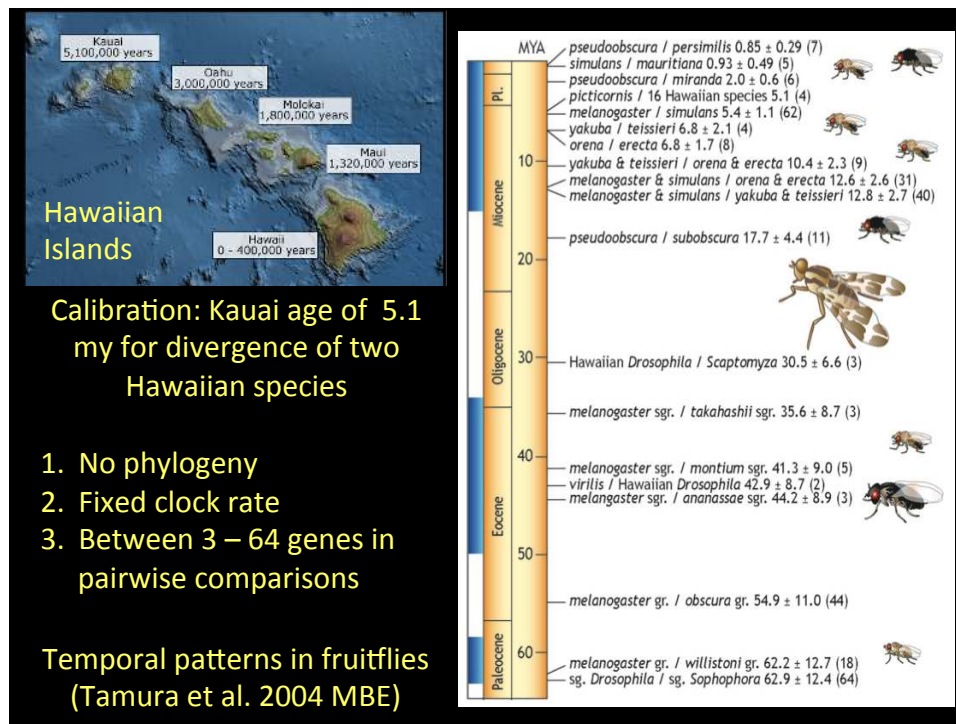


Timing of divergence

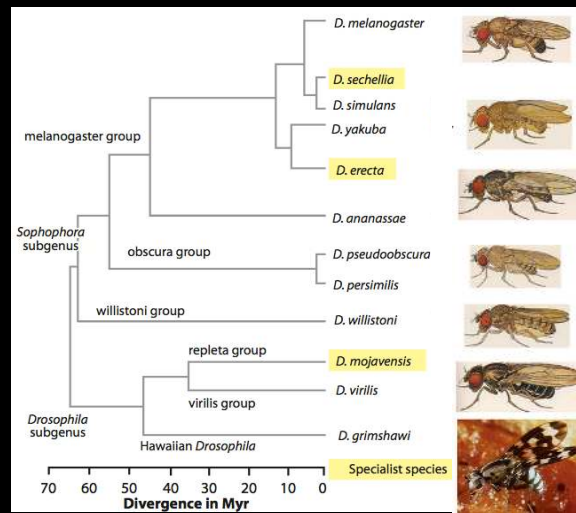
- **Directly affects rate estimates**
- **Deriving unbiased dates from molecular data**
 - Large field of software development
- **Bayesian methods, while potentially informative and unbiased**
 - Can be easily, and are routinely, abused




Wheat and Wahlberg 2013 TREE



Thus, the age of this clade is fiction



Drosophila 12 Genomes Consortium 2007 Nature