Genetics, Microarrays & Evolution: Issues as a Statistician

(sans formula)

Brian S. Yandell
Horticulture, Statistics & Biometry
University of Wisconsin-Madison

who am I (professionally)?

- Professor Brian S. Yandell
- joint appointment across colleges:
  - 50% Horticulture (CALS)
  - 50% Statistics (Letters & Sciences)
  - UW-Madison since 1982
- Biometry Program
- teaching & research

biometry program

MS degree

- coadvise with biologist
- bridge biology & stats
- project & oral report
- consulting experience
- 10 completed, 1 current

  Botany, Dairy Sci (2), Genetics, Hort, Land Resources, Meat & Animal Sci, Wildlife Ecology (2), Zoology

consulting facility

- statistical consulting
  - 5 faculty, 2-4 students
- computing assistance
  - 2 staff + operators
- self-help model
  - guide research ideas
  - build skill sets
- collaboration
  - students, faculty, staff
  - CALS, VetMed, L&S

research & teaching

- statistical genetics
  - QTLs in *Brassica* & mice
  - genetic architecture
    - gene action
    - epistasis
  - microarrays
    - differential expression
    - genetical genomics
- statistical ecology
  - population ethology
  - individual-based simulations
- stats consulting
  - communication skills
  - write, plot, talk
  - bridge stats & biology
- linear models
  - experimental design
  - complicated analysis
  - problems directly from consulting
  - published textbook

what is statistics?

We may at once admit that
any inference from the particular to the general
must be attended with
some degree of uncertainty,
but this is not the same as to admit that
such inference cannot be absolutely rigorous,
for the nature and degree of the uncertainty
may itself be capable of rigorous expression.
— Sir Ronald A. Fisher
(1935 *The Design of Experiments*)
digital.library.adelaide.edu.au/coll/special/fisher

what is biology?

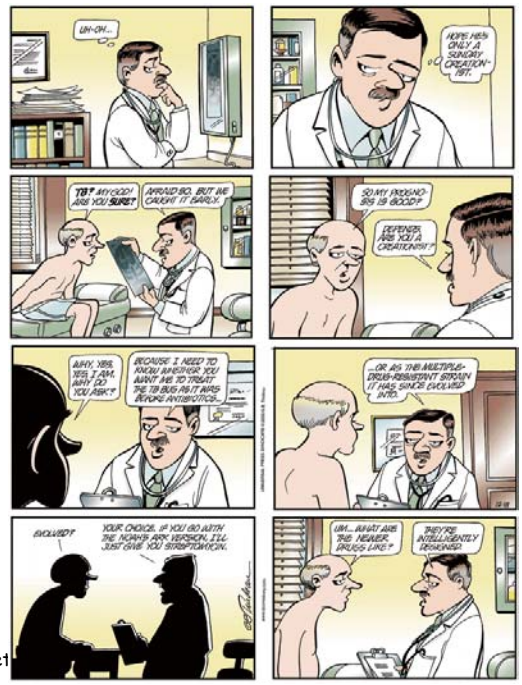
Biology … consists of two rather different fields,
mechanistic (functional) biology and
historical [evolutionary] biology.

Functional biology deals with … cellular processes, including those of the genome. …

[Evolutionary biology] involve[s] the dimension of historical time.

— Ernst Mayr at 100
(*What Makes Biology Unique?* 2004 Cambridge U Press)

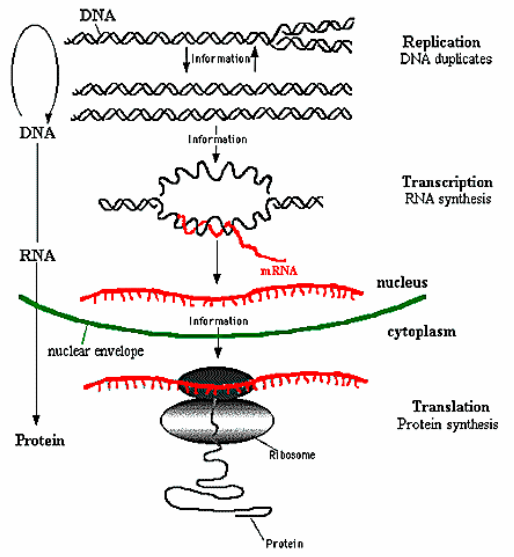
Doonesbury
by Garry Trudeau
18 dec 2005
www.doonesbury.com



functional biology: how does life work?

- broad scientific questions
  - how do plants modify flowering time?
  - how do mice (humans) develop diabetes?
- molecular investigations
  - genetic association: QTL(s)
  - fine mapping: candidate gene(s)
  - biochemical pathways: causal models

DNA
↓
RNA
↓
protein
↓
metab-
olites

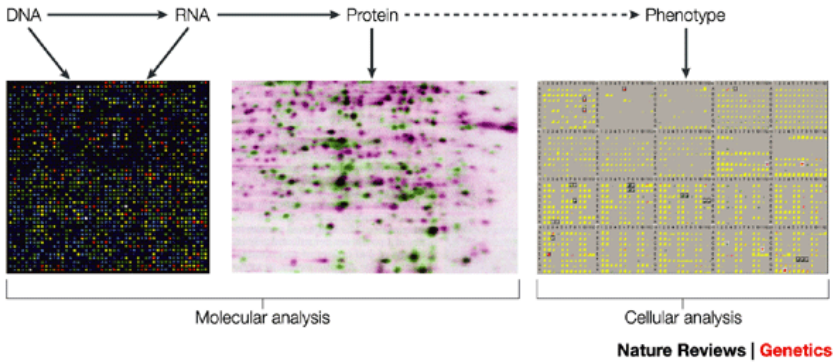


J Watson & F Crick (1953)
R Franklin (1953)

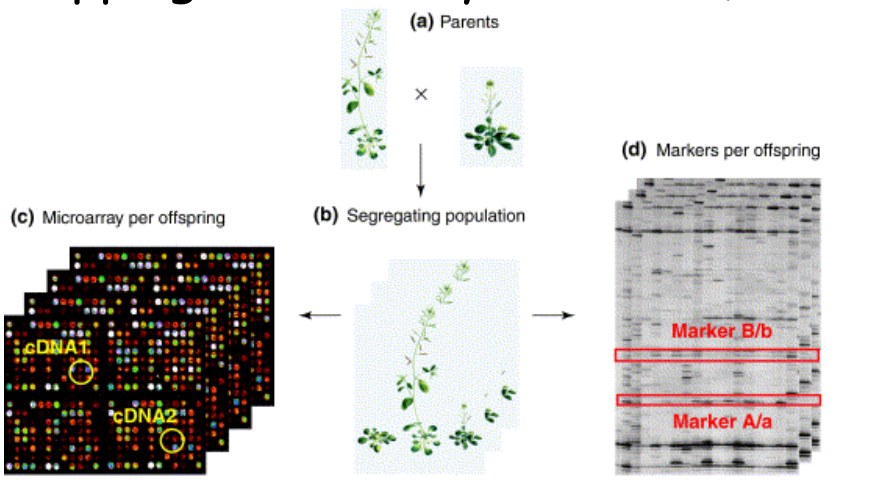
The Central Dogma of Molecular Biology

www.accessexcellence.org/RC/VL/GG/central.html

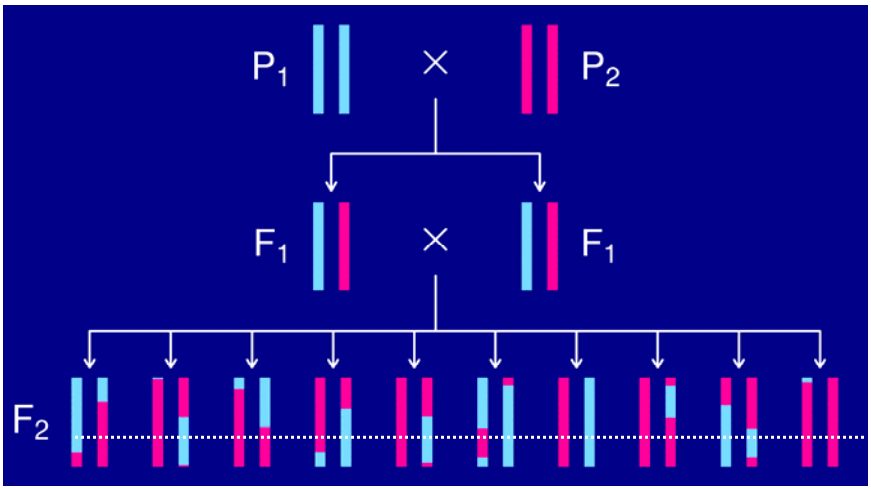
central dogma via microarrays (Bochner 2003)



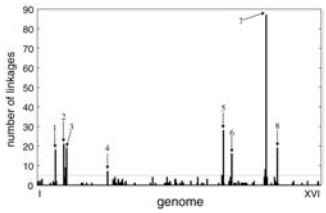
genetical genomics: mapping microarrays (Jansen Nap 2001)



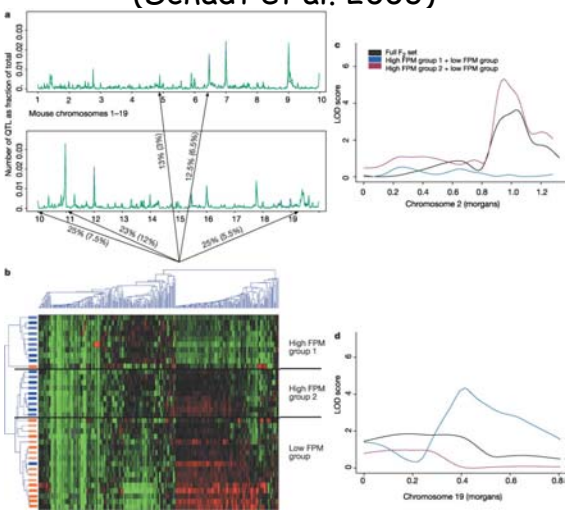
intercross: genetic mosaic of P1,P2
(from KW Broman)



expression
pleiotropy
in yeast genome
(Brem et al. 2002)

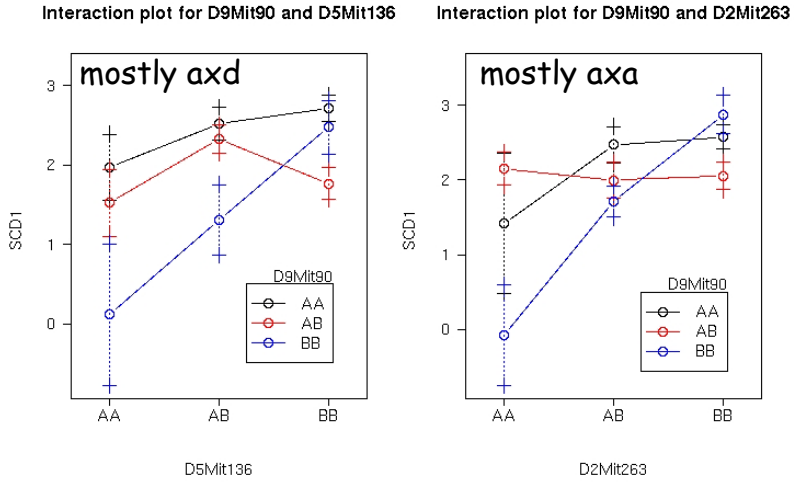


coordinated expression
in mouse genome
(Schadt et al. 2003)



SCD1: epistatic interaction plots

(chr 5x9 p=0.007; chr 2x9 p<0.0001)



genetic architecture for mRNA (SCD1) expression

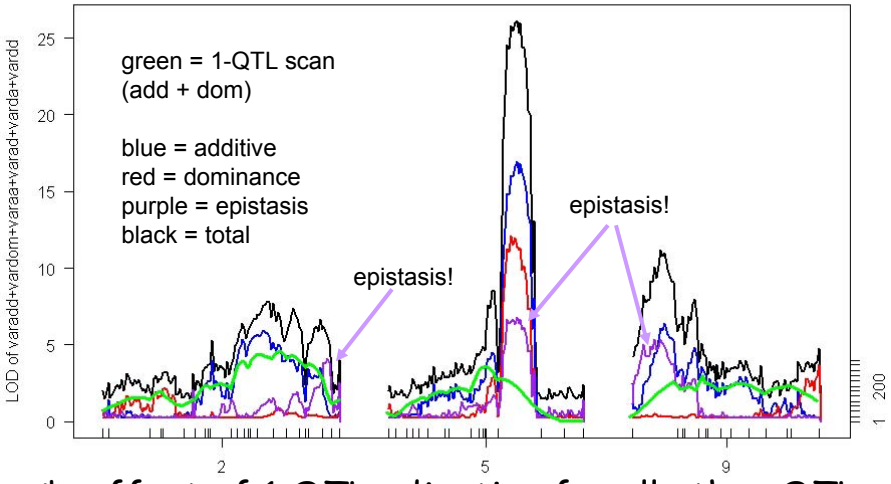
Model:

	Df	MS	F	Pr(F)	
Model	7	69.060	73.10	0.00000	***
Error	99	0.945			
Total	106	70.005			

Single term deletions

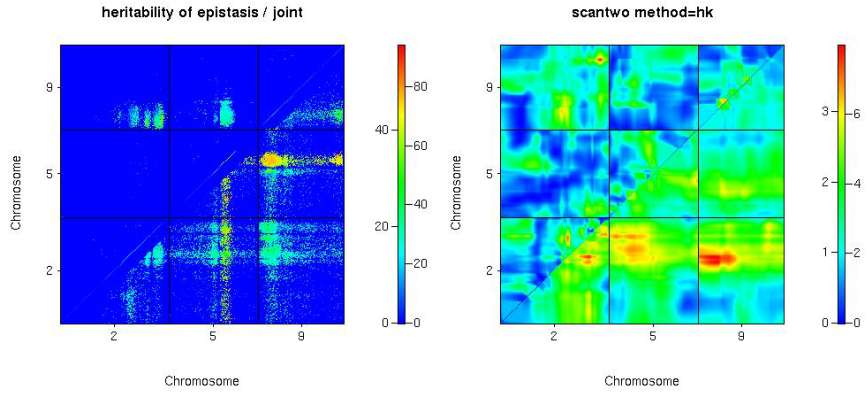
	Df	MS	F	Pr(F)	
Chr2@80	1	9.073	9.60	0.0025	**
Chr2@105	1	0.073	0.08	0.78	
Chr5@67	1	0.218	0.23	0.63	
Chr9@67	1	7.156	7.57	0.0070	**
Chr9@67dom	1	0.106	0.11	0.74	
Chr2@105:Chr9@67	1	15.612	16.52	0.000096	***
Chr5@67:Chr9@67dom	1	7.211	7.63	0.0068	**

SCD1: marginal* LOD by locus



*: effect of 1 QTL adjusting for all other QTL

SCD1: marginal 2-QTL scan
examine 2 QTL in presence of others



marginal 2-QTL scan

2-QTL only scan

how to assess genetic architecture via genetical genomics?

- screen mRNA across segregating panel
  - which show strong evidence of heritability?
- organize into functional groups
  - correlation across panel
  - (Lan, Chen et al. 2006 PLoS Genetics)
- infer genetic architecture by group
  - allow for multiple QTL and epistasis
- validate
  - comparative genomics, KOs, etc.
- infer biochemical pathways

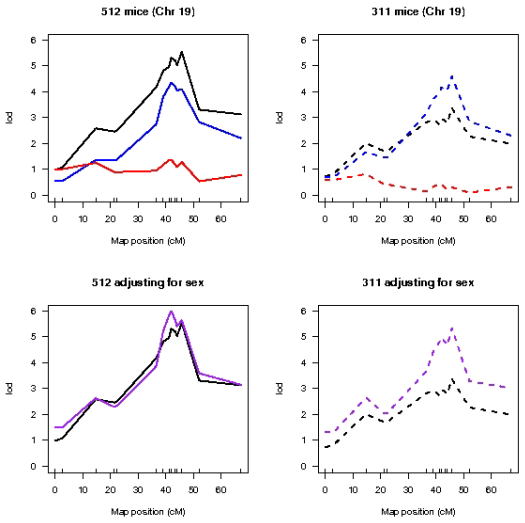
from QTL to candidate gene?

trait: log10(insulin) at 10 weeks in mice

QTL on Chr 19

solid=512 mice (now)
dashed=311 mice (then)

black=all
blue=male
red=female
purple=sex-adjusted



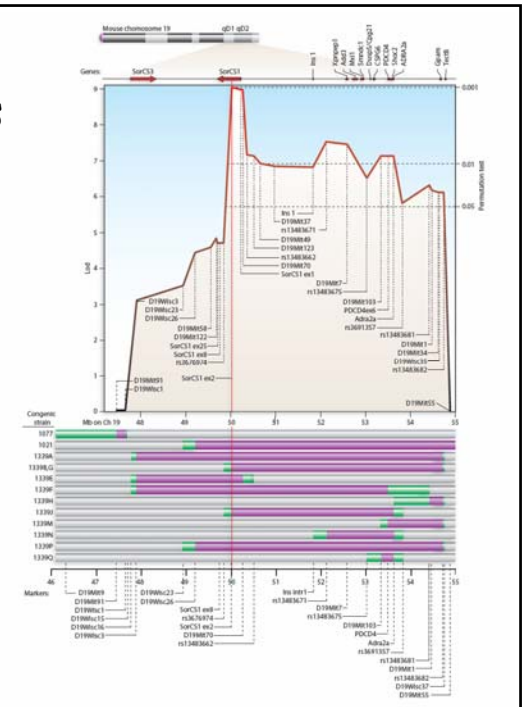
fine mapping with meta-analysis

candidate gene: Sorcs1

meta-analysis across 11 sub-congenic strains

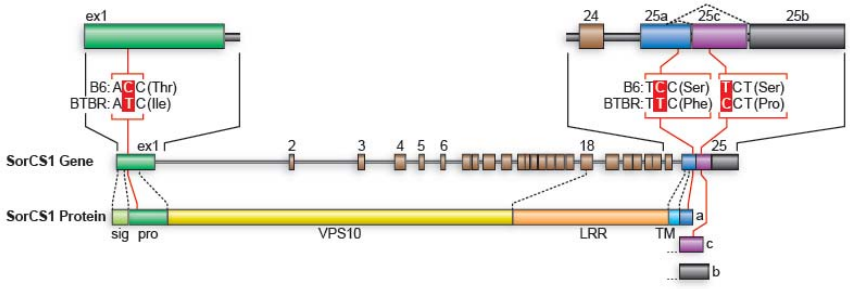
marker regression & within-strain permutations

Clee et al. (2006 in review)

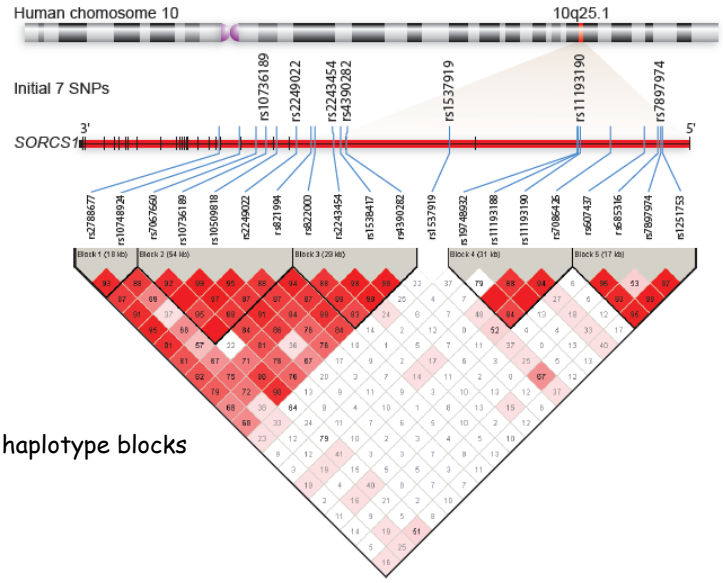


Sorcs1 gene & SNPs

(only gene in region with SNP variation between parent strains)



Sorcs1 study in humans

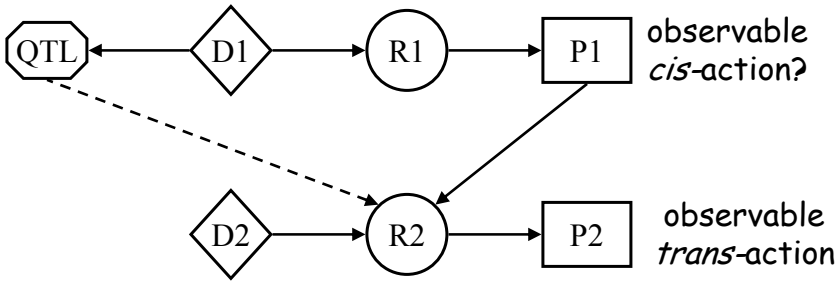


5 haplotype blocks

goal: unravel biochemical pathways

(with Elias Chaibub)

- candidate genes in QTL regions
- lab experiments on pathway components
- graphical models via genetical genomics



evolutionary biology: why did life end up this way?

- natural selection
  - why do certain species persist/perish?
  - why do some characteristics emerge/persist/perish?
- modeling individuals in a population
  - historical narratives
  - time-based simulations
  - event-based simulations

what is natural selection?

Darwin's "variation and selection"
1. random variation
   mutation, meiosis, gamete meeting
2. differential selection
   selection of the best (fittest)
   culling/elimination of the worst

selection on individuals, not populations
   survival selection
   sexual selection (reproductive success)

— Ernst Mayr at 100
(*What Makes Biology Unique?* 2004 Cambridge U Press)

individual-based models in population ethology

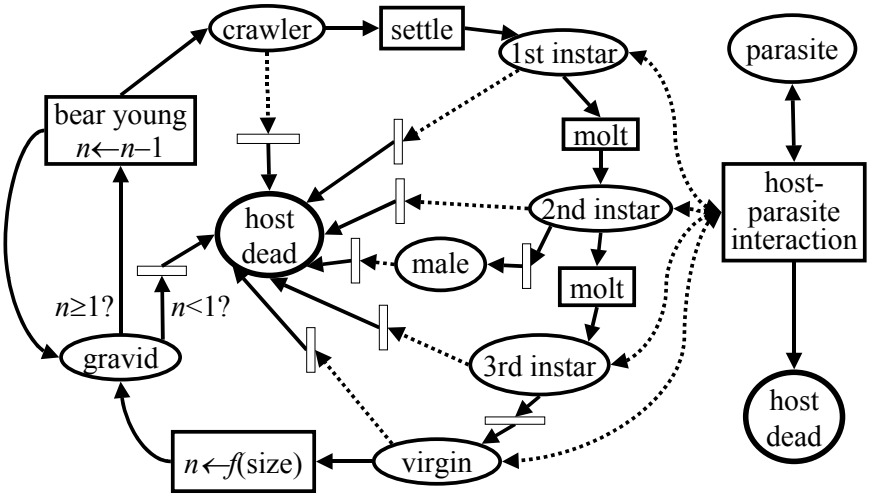
California red scale damage to orange

two parasitoids

dead red scale

www.ipm.ucdavis.edu/PHENOLOGY/models.html

event-driven model: red scale life history



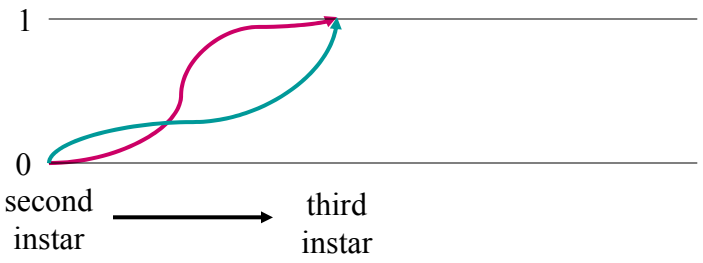
what do we measure?

- trait: aspect of an individual that is constant or slowly/rarely changing over span of study
- event: significant biological change in state that is instantaneous at resolution of study
- types of events
  - birth, death
  - change of "health" or "development" state
    - flower opening, seed setting, dormancy
  - interactions with other individuals
    - predation, parasitism, plant harvesting
    - reproduction, pollination, fruit/seed harvesting

(Bland Ewing et al. 2002 *Ecol Model*)

how to incorporate genetics?

- genotype is a trait of an individual
- differential probabilities for events
  - depending on genotype, event history
  - change event rate or shape



what is our scope of measurement?

- measurements are in context of span & resolution
  - mechanics of measuring
  - focus of key questions in time and space
- span: largest amount of time/space studied, with aspects over longer intervals considered constant or slowly varying
- resolution: smallest increment of time/space contributing useful biological information, with processes over smaller scales assumed to be instantaneous

orange:	resolution (1 day)	span (1 year)
molecular:	resolution (1 picosec)	span (1 msec)

(Bland Ewing et al. 2002 *Ecological Modeling*)

modeling natural selection

- depends on genetics, event history, environment
- differential survival
  - chance of surviving parasitism
  - development rate
  - ability to migrate to new habitat
- sexual selection
  - mate preference based on genotype
  - other gene flow vectors (co-evolution)
  - barriers to reproduction (speciation)

how do we infer general properties from historical narrative simulations?

- cannot predict exact outcomes
  - probability over range of scenarios
  - variability across multiple simulations
- "black box" testing
  - sensitivity to changes in conditions
    - temperature, initial population sizes
    - rates of occurrence of life events
  - comparison with historical records

summary

- who am I?
- what do statisticians do?
- what is biology?
- how have I studied genetics?
- how can microarrays deepen our models?
- how might I study evolution?