



# Genetic architecture of quantitative traits

William G. Hill  
Institute of Evolutionary Biology  
School of Biological Sciences  
University of Edinburgh

QTL – MAS Poznan Mqy 2010

# Architecture of quantitative traits

Some questions:

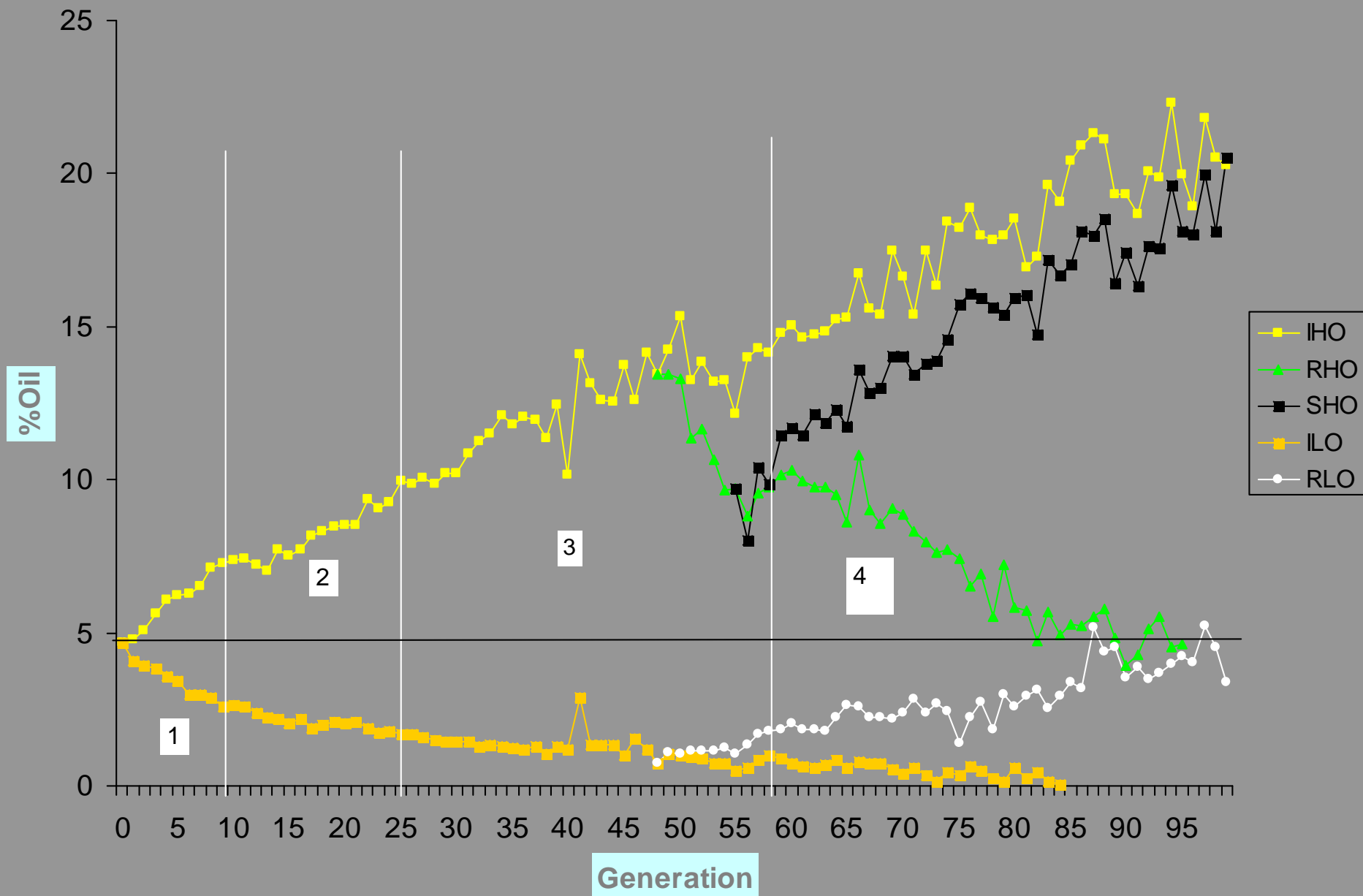
- What is distribution of gene effects and their interactions?
- What is distribution of gene frequencies, and relation to their effects?
- What do the genes do?
- What are the evolutionary forces that lead to these outcomes and to total variation maintained?
- What are implications for e.g. selection programmes?

Information on most of these points is weak or lacking. But data now being acquired.

# Variation in quantitative traits

- Genetic variation is ubiquitous
- Response to selection continues
- How is variation maintained in populations?
  - Mechanisms: e.g. drift, selection
- Explaining the magnitude of standing variation remains a poorly resolved problem
- What do we expect to see in new analyses?
- What have we learnt/might we learn?

# Oil means



# Consequences of and inferences from artificial selection

- Long continued response in most populations
- Range many times initial standard deviation (genetic and phenotypic)
- Some plateaux obtained
  - Waited long enough?
  - Often fitness associated, e.g. major genes
- Implies many loci involved (> 50 direct evidence from Illinois lines), and eventually contribution from mutation (required to explain continued response in Illinois lines with small  $N_e$ )
- Implications of large numbers of loci correspond to e.g. those on distributions of traits and height in humans from GWAS

# Maintenance of variation in quantitative traits

- Origin: mutation
  - Measured as mutational variance,  $V_M$ , new genetic variance due to mutation each generation
  - Estimates of rates, various lab species
  - $V_M/V_E \sim 0.001$  (range 0.0001 to 0.01)
  - Thus for a trait with  $h^2 = 0.5$ ,  $V_G$  maintained is about  $1000 V_M$  (range 100 to 10000)
- Sources of loss
  - Genetic drift (sampling),  $\Delta V_G = 1/(2N_e)$
  - Hence if neutral for fitness  $V_G = 2N_e V_M$   
(Here and subsequently assume additive gene action unless stated otherwise – most variance additive even if effects are not)
  - But we do not see close relationship  $V_G$  to  $N_e$ , so not just mutation – drift balance
  - Selection (but also a reason for maintenance of variation?)

# Types of selection

- Directional selection on trait: extremes favoured
- Stabilising selection on trait: intermediates favoured (traditionally, human birth weight)
- Selection varying in time and/or space
- Selection wholly/mainly unrelated to the trait (pleiotropic effects)
- Heterozygote superiority, and equivalent epistatic models

# Directional selection

- Directional selection on trait
  - Under infinitesimal model (all genes tiny effect), variance maintained is as for neutral model
  - Otherwise, variance maintained depends on proportion of variance due to favourable mutations (and on  $N_e$ , not on  $i$ )
- Directional selection on other traits
  - Depends on proportion increasing trait in index
- Directional selection, but direction varying in time
  - Total change in gene frequency over generations depends on sum of selective values: if non-zero, implies net directional selection. Stability requires long cycles of fluctuation relative to generation interval
- Directional selection, but direction varying in space
  - Alternative niches with different directions of selection: stability requires migration rate to be small relative to selection at locus level
- Heterozygotes favoured – little evidence



# Stabilising selection on trait

- Typically assume non-optimal model, e.g. Gaussian fitness function
  - Strength depends on ratio ‘variance’ of fitness function to variance of trait,  $V_S$
  - Estimates of  $V_S$  from natural populations, traditionally 20, but ??
- Fate of mutation
  - If population is centred around optimum, any mutant which influences the trait is at a disadvantage ( $s \propto a^2 V_S$ )
  - Hence variation lost, that contributed before loss is *independent* of gene effect,  $a^2$ .
  - If selection solely on the target trait, variation maintained is  $V_G = 4\lambda V_S$ , where  $\lambda$  is number of mutant loci for the trait. What is  $\lambda$ ? If consider whole genome involved, it could be large. But...
  - Alternative infinitesimal type models. But...

# Pleiotropy - observations

- Genetic correlations
- *P*-element mutations:
  - indicate large mutational target size of most quantitative traits (Mackay *et al.*). Of mutations screened, approximate proportion of inserts affecting the trait were:
    - 22% abdominal bristle number
    - 23% sternopleural bristle number
    - 41% starvation stress resistance
    - 6% olfactory behavior in response to a single odorant
    - 22% wing shape
    - 37% locomotor startle response
    - 35% aggressive behaviour
  - Many genes: Vast majority of those tagged were novel and not previously annotated to affect adult quantitative traits
  - Most inserts have discernible effect on fitness

# Pleiotropic models

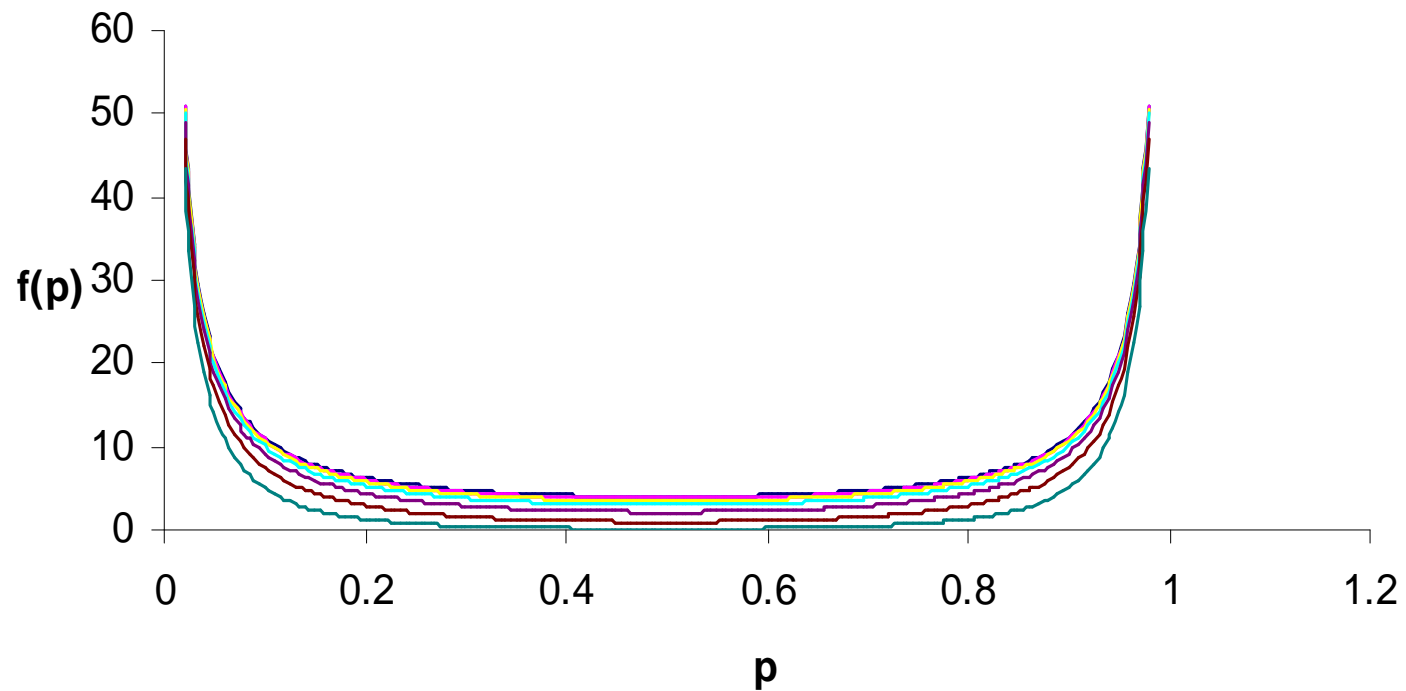
- Selection exclusively (predominantly) on other traits, with trait under consideration having little or no impact on the fate of the mutation
- Examples: (semi-)lethals for viability or fertility
  - Genes reducing fitness found in artificial selection lines of *Drosophila*
- Those of less deleterious effect, but  $s$  values much larger than those for stabilising selection
  - e.g. gene effect  $0.1SD$  on a trait has  $s \sim 10^{-3}$  with  $V_S = 20$ .
- Almost all new mutations appear to be deleterious for fitness; favourable mutants occur – but traverse quickly
- Much evidence of pleiotropy
- If mutants have pleiotropic effect on other traits also under stabilising selection, total selection is greater and variance maintained is less than predicted from stabilising selection alone.
- *Joint models*: stabilising selection on trait and any pleiotropic effect
- In all cases, tendency is to keep mutants at low frequencies

# Frequency distribution and heterozygosity

- Neutral model with rare (non-recurrent) mutation
  - Consider time spent in each frequency class to fixation
  - For a mutant with initial frequency  $p$ ,  $f(p) \propto 1/p$
  - If cannot distinguish mutant & ancestral type
$$f(p) \propto 1/p + 1/(1-p) = 1/[p(1-p)]$$
  - Heterozygosity ( $H$ ) maintained =  $2N_e$  x initial  $H$  (1 if  $N_e = N$ )
  - $H(p) \propto p(1-p)f(p) = 1$ ,
    - i.e. uniform distribution of variance for  $p$
- Deleterious mutant,  $s =$  fitness
  - $f(p) = 2[1 - e^{-4Ns(1-p)}] / [p(1-p)(1 - e^{-4Ns})]$   
strongly concentrated in the tails if  $Ns$  is large
- If consider mutations from both sides
  - $H(p) = [2 - e^{-4Ns(1-p)} - e^{-4Nsp}] / (1 - e^{-4Ns})$

# Gene frequency distribution - neutral/deleterious genes

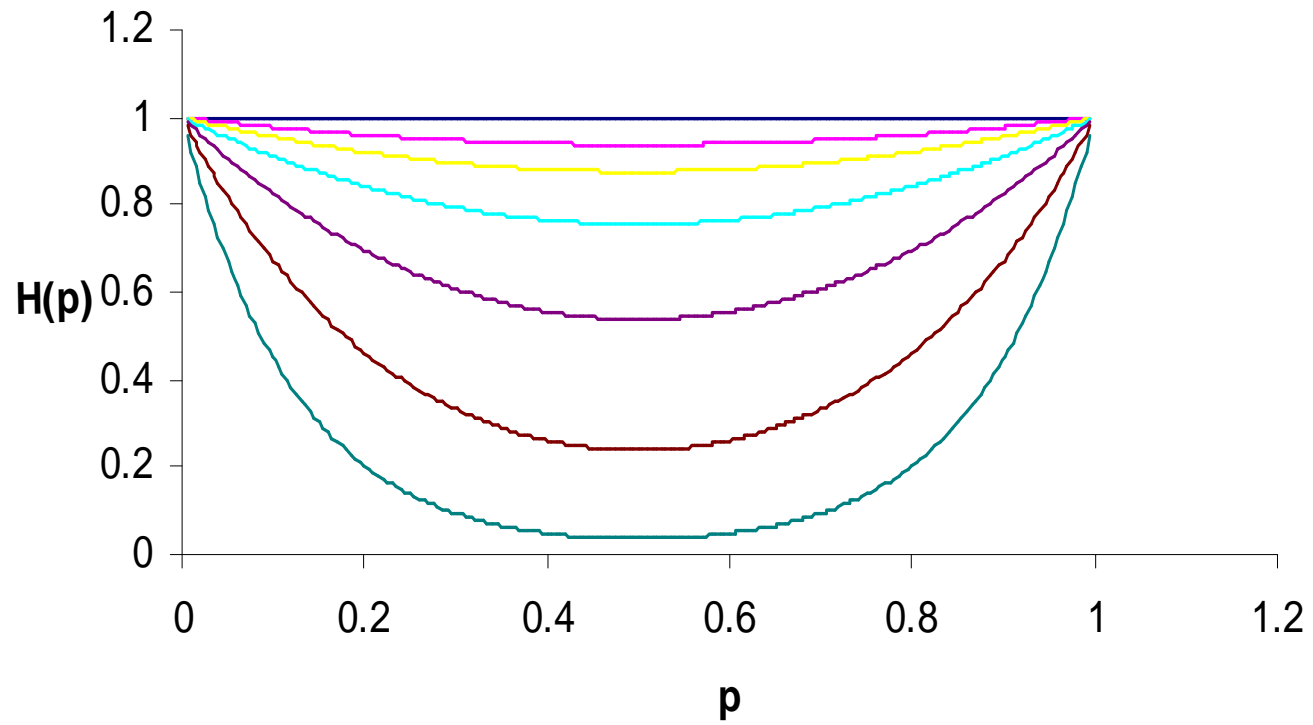
$2N = 1000$



$S = -4Ns$



# Distribution of heterozygosity



$-4Ns$

— 0 — 0.5 — 1 — 2 — 4 — 8 — 16

# Distributions of effects of mutants on fitness and on traits

- Mutation accumulation laboratory experiments:
  - Can only detect mutants of large effects
  - Miss mutants that are very deleterious as homozygotes
- Analysis of frequency distribution of molecular variants fixed in one species and segregating in another (eg human/chimpanzee) with ancestral allele known. Infer distribution of  $N_e s$ 
  - May miss those of most extreme deleterious effects (too rare to see)
  - Cannot elucidate distribution of fitness increasing mutants – go to fixation fast

(Keightley, Eyre-Walker *et al.*)

# Estimates of distributions of mutation effects

- Typically fit (reflected) gamma distribution
  - Shape = 1 is exponential
  - Shape < 1 is more leptokurtic (more near 0 & more very large)
- Fitness (gamma):
  - Estimates from population analyses: leptokurtic, shape  $\sim 0.2$ , mean for humans of  $4Ns \sim -3000$  [less for livestock?]
  - Estimates from mutation accumulation experiments in lab animals: platykurtic, shape  $> 1$ , mean  $s \sim -0.2$
- Traits (reflected gamma)
  - Mutation accumulation experiments: some evidence bimodality, but in any case much less extreme distribution, shape  $> 1$  (? - estimates range widely)



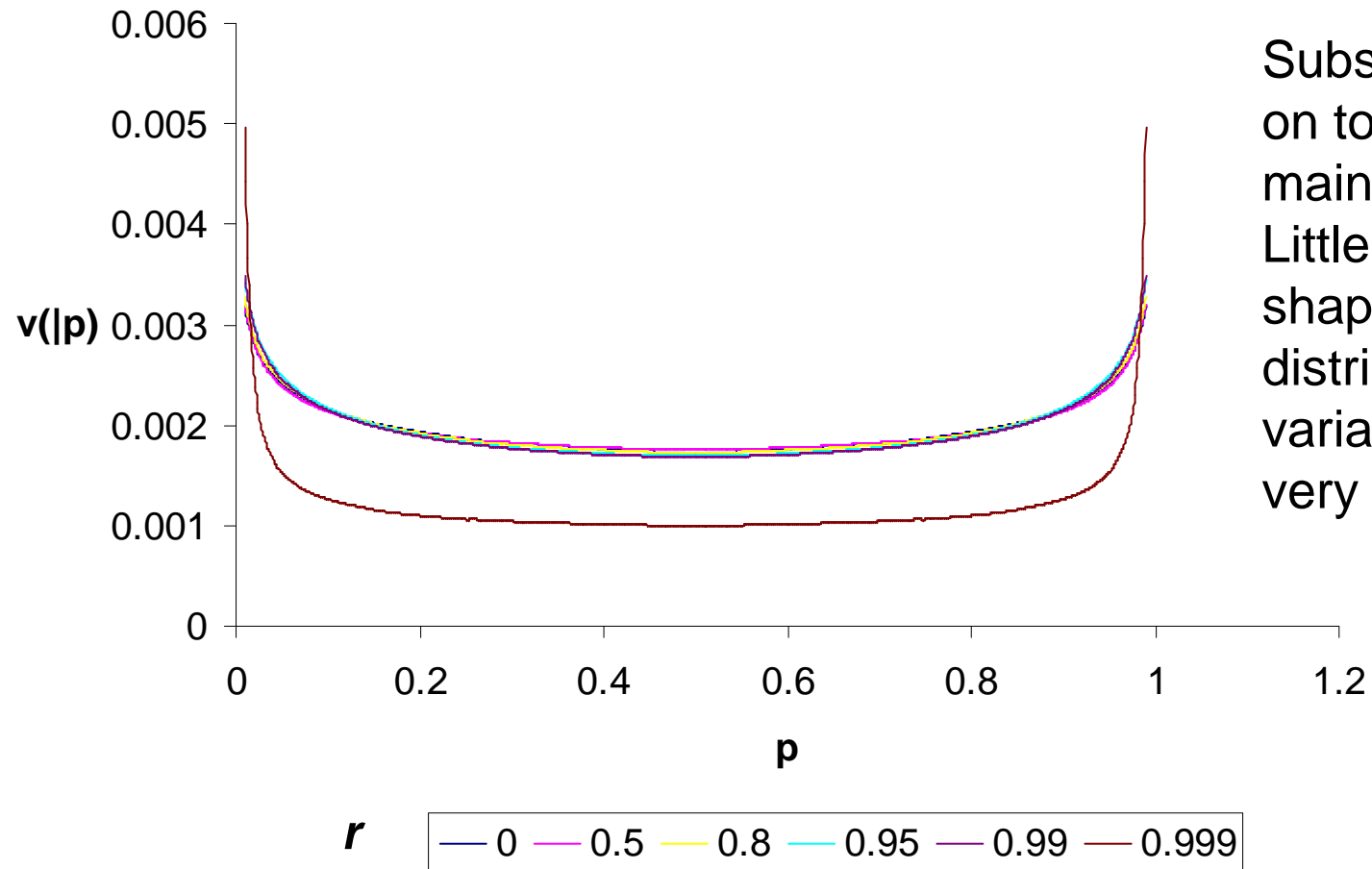
# Combining fitness and trait distributions

- Information on joint distributions mostly lacking.
- Hypothesis: Association of mutation's effect on fitness (deleterious) and on trait (increasing or decreasing)
- Assume correlation  $r$  between fitness ( $s$ ) and absolute value of effect on trait ( $|a|$ ). Neutrality on trait if  $r = 0$ , but heterozygosity is still reduced
- Analysis of Eyre-Walker (PNAS 2010) and some simulation (Hill, primitive)
- Assume fitness ( $s$ ) has gamma distribution, 'independent' part of trait has gamma distribution, and mutant effect on trait has sum

$$|a^*| = sr\sqrt{V(a)/N(s)} + a\sqrt{1 - r^2}$$

## Deleterious mutations – pleiotropic model

### Distribution of variance



Substantial effect on total variance maintained. Little effect on shape of distribution of variance unless  $r$  very high

# Causes of non-ascertainment in GWAS

## Causes

- Small amount of variation contributed to trait  $2p(1 - p)a^2$ 
  - Small effects on trait  $a^2$
  - Low heterozygosity  $2p(1 - p)$
  - Effects and heterozygosity negatively correlated
- Inadequate LD - overall a function of  $2p(1 - p)a^2r^2$ 
  - Low density of markers, and low  $r^2$
  - Low frequency of QTL ( $p$ ) relative to MAF ( $q$ ) of markers
    - If  $D' = 1$  (missing haplotype), max  $r^2$  is  $[p(1 - q)]/[(1 - p)q]$
- Inadequate sample size, confounding etc



# Distributions of variance and response

- If much of variation centred in tails, expect that variance would rise with selection as rare genes become more frequent: more than allowing for loss of those at high frequency
- Analysis shows this to be expected for unlinked loci
- Using *Drosophila* model, Bulmer effect/LD compensates
- In mammals, no experiments from natural population base
- In livestock, no evidence that variation is falling as a result of artificial selection: broilers  $h^2 \sim 0.25$  (still);  $h^2$  has risen in dairy cattle ( $\sim 0.25$  to  $>0.35$ , but also better recording and management)

# Some conclusions

- Problem of explaining magnitude and maintenance of quantitative genetic variation remains
- Most models predict very extreme distribution of frequencies of genes, including trait genes
  - Less extreme if  $N_e$  small
  - But ascertaining genes e.g. in GWAS contributing to variance easier than ascertaining gene number
  - Distribution of effects on trait still to be determined but opportunities from such studies
  - But only for those of reasonably large effect
- Implication: extreme frequencies imply homozygotes at both loci rarely present
  - Little dominance variance or epistatic variance – more extreme than simple neutral case (Hill et al. 2008)

# Acknowledgments

Adam Eyre-Walker

Xu-Sheng Zhang

Peter Keightley

Peter Visscher

Mike Goddard

USS





## Impact on partition of variance

‘U’ is neutral case. Stronger pleiotropic selection expected to increase proportion of variance that is additive rather than dominance or epistasis

Genetic model	Distribution of allele frequencies			
	$p = 1/2$	Uniform	‘U’ ( $N = 100$ ) <sup>b</sup>	‘U’ ( $N = 1000$ )
Dominance without epistasis $d = 1/2a$	0.89	0.91	0.93	0.93
Dominance without epistasis $d = a$	0.67	0.75	0.80	0.80
Dominance without epistasis $d = 0$	0.00	0.33	0.50	0.50
A × A without dominance	0.00	0.67	0.87	0.92
Duplicate factor 2 loci	0.27	0.56	0.71	0.75
Duplicate factor 100 loci	0.00	0.00	0.00	0.00
Complementary 2 loci	0.57	0.67	0.74	0.76

<sup>a</sup>Models defined in Methods section

<sup>b</sup>Population size

doi:10.1371/journal.pgen.1000008.t002

# Impact on dominance variance

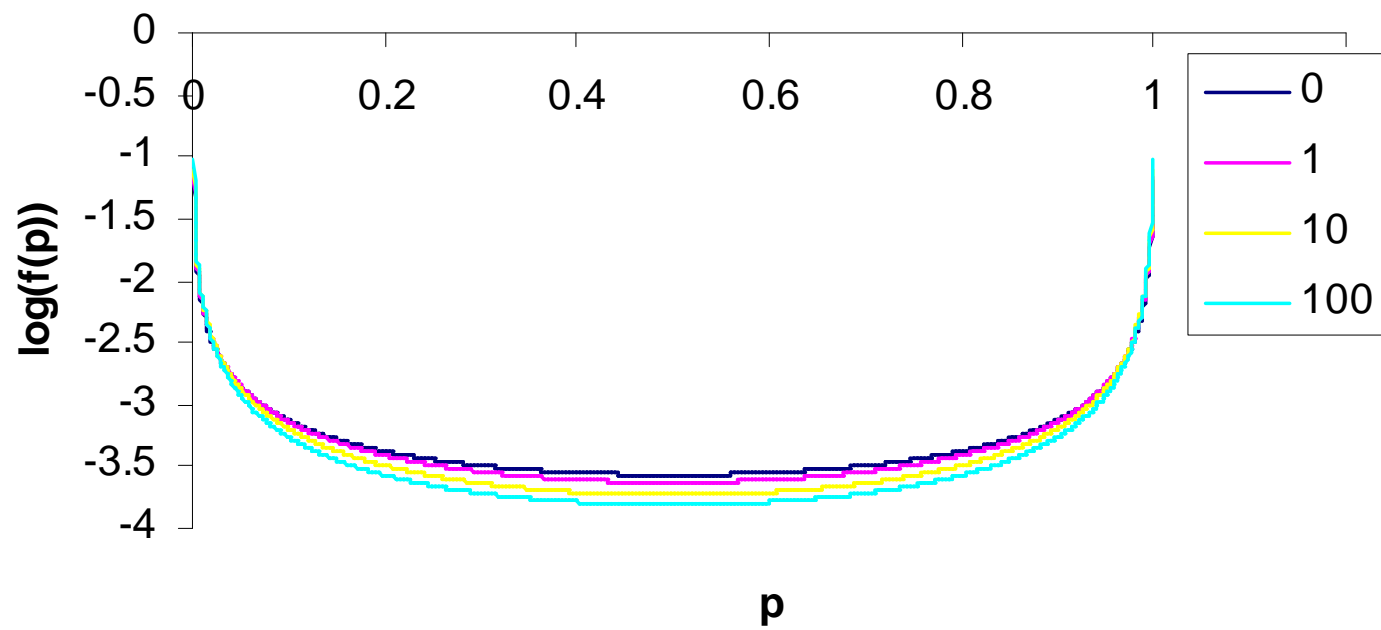
## Stabilising (+ pleiotropic) selection – ‘Drosophila’

Distributions of effects of genes ( $|a|/\sigma_P$ ) on the quantitative trait in terms of numbers of loci segregating and proportion of  $V_A$  contributed. Neutral models ( $N = 250$ ) and natural selection models ( $N = 10000$ ) that maintain heritability  $1/3$  for  $V_M/V_E = 0.1\%$

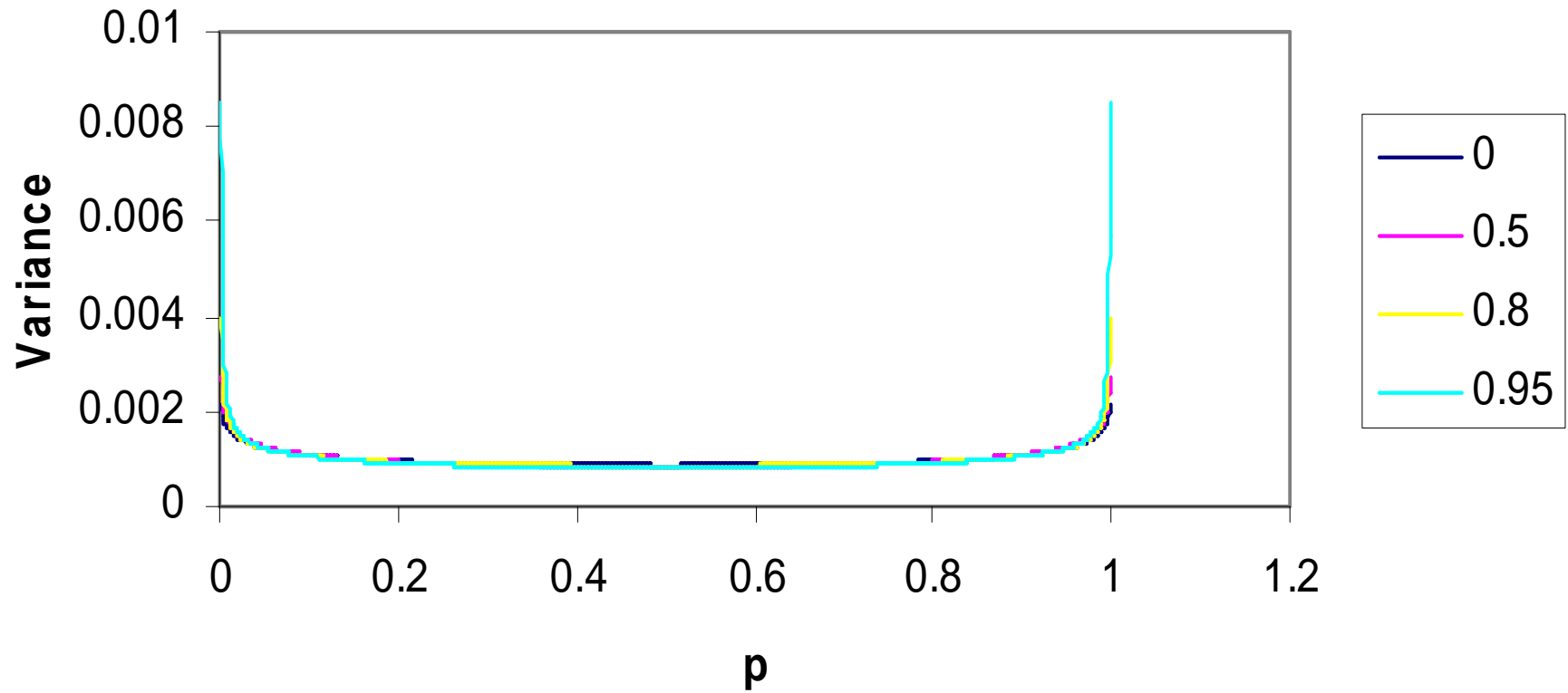
Effect $ a /\sigma_P$	< 0.01	0.01 - 0.1	0.1 - 0.25	0.25 - 0.5	> 0.5
Neutral					
No. loci	644	483	129	108	21
$V_A$ (%)	0.2	14	41	34	10
Natural selection					
No. loci	2758	2038	394	51	3
$V_A$ (%)	0.2	27	50	19	3



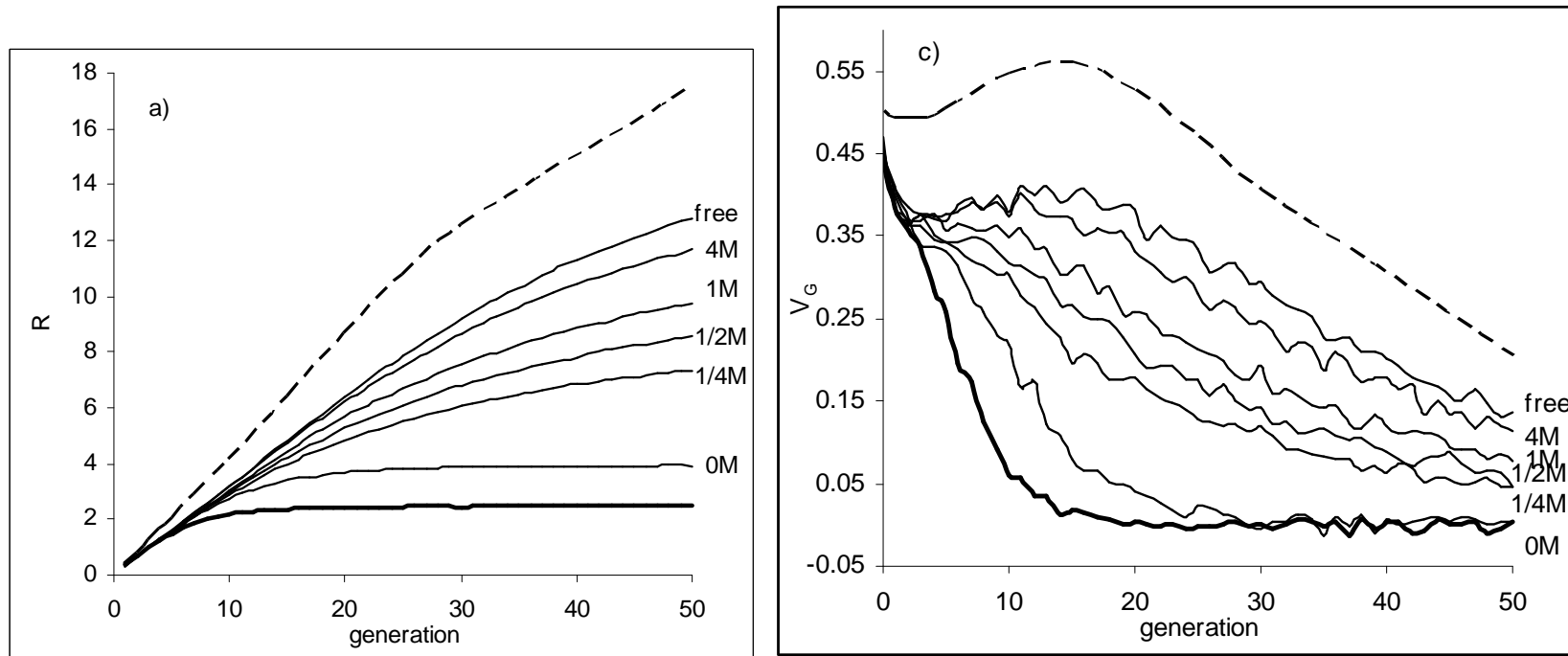
# log(frequency distribution)



### Distribution of variance - pleiotropic $N_s = -100$

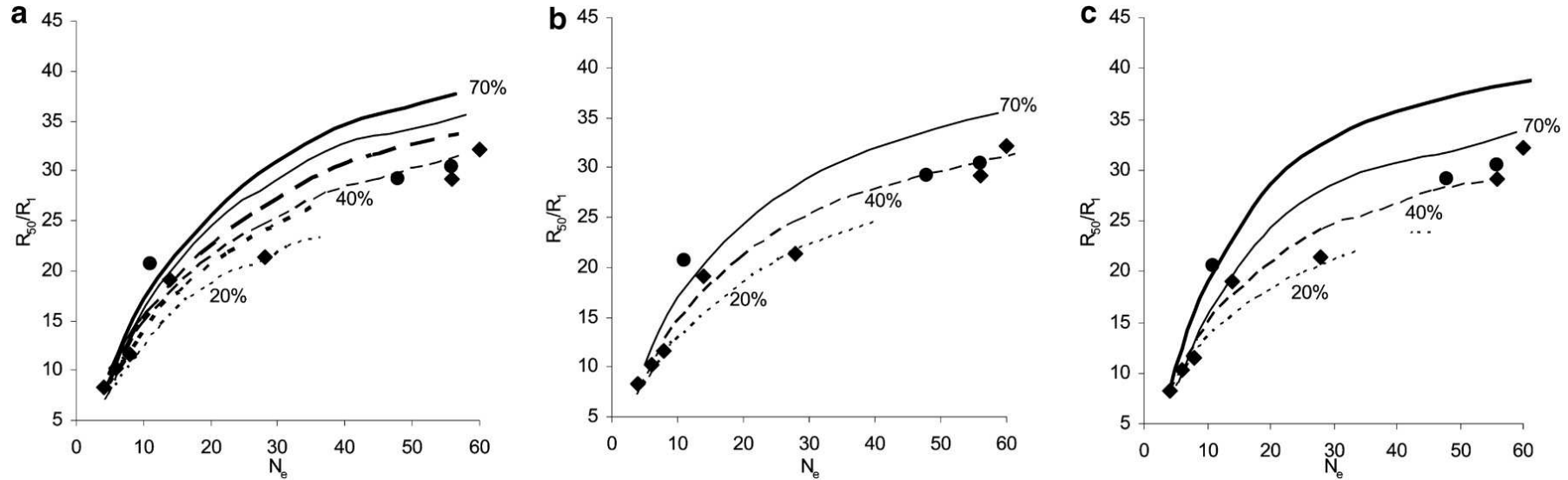


## Expected response and variance under directional selection



Influences of linkage and the Bulmer effects on a) response, and c) the within-line genetic variance. The base population was from a natural population at mutation-selection balance. The natural population size  $M$  was 10 and the equilibrium heritability 0.5. A cage population of size 160 was drawn from this before artificial selection (40% selected) and size  $N=40$ . Two methods: transition matrix for independent individual loci (dash lines ---), and MC simulation for genome of three chromosomes each of length 0,  $\frac{1}{4}$ ,  $\frac{1}{2}$ , 1, 4 and  $\infty$  Morgan (M) (thin lines—), or 1 completely linked chromosome. (Z & H 2005)

## Robustness of selection response to model assumed



Predicted and experimental values of total response after 50 generations, relative to initial response ( $R_{50}/R_1$ ), Theoretical for selection of 20% (dotted lines), 40% (dashed), and 70% (solid) Data points (diamonds,  $i > 1.0$ ; circles,  $i < 1.0$ ) results from *Drosophila*, (a) Artificial selection starting from base populations drawn immediately from a natural population at MSB, with (thick solid lines) and without contributions from new mutations (thin solid lines). (b) As (a) but first kept for 32 generations  $N_e = 160$ , including new mutations. (c) The predictions from the infinitesimal model, including mutation and linkage.. Robertson predictions with new mutation and the Bulmer effect (thick solid line, 40% selected)