Lecture 12 : Genetic variance within and between populations

Genetics & Genomics Winter School

Valentin Hivert

contact: v.hivert@imb.uq.edu.au
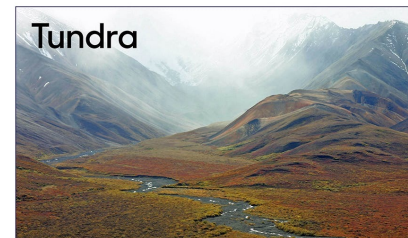
Institute for Molecular Bioscience

Program in Complex Trait Genomics

Introduction

How to explain differences in heritability estimates and loss of prediction accuracy of polygenic score across populations?

Introduction

Environment

National Geographic

Introduction

Genotype-by-Environment interactions

Tropical Rainforest

Temperate Forest

Boreal Forest

Desert

Savanna

Grassland

Marine

Freshwater

Tundra

National Geographic

×

Science

Introduction

Genotype-by-Genotype interactions

Science

Introduction
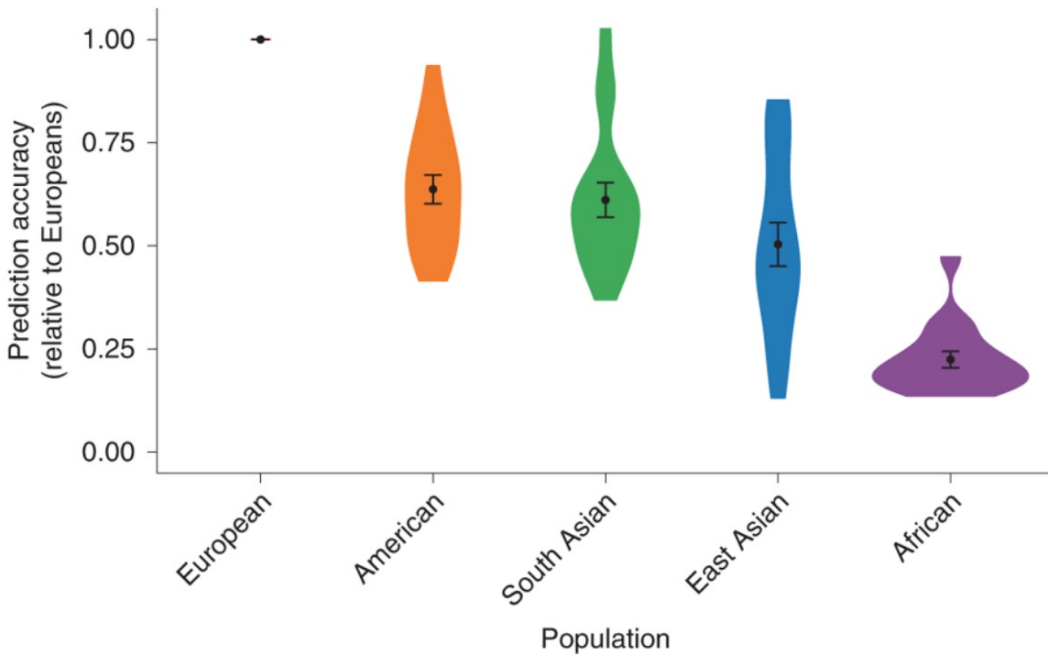
Martin *et al.* (2019)

ARTICLE
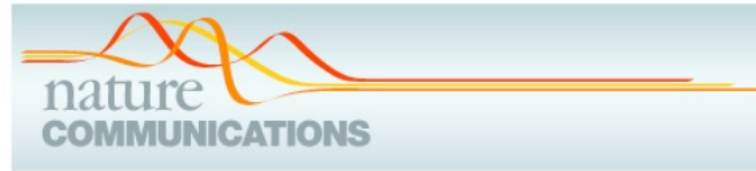
https://doi.org/10.1038/s41467-020-17719-y   OPEN

Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations

Ying Wang [1], Jing Guo [1], Guiyan Ni [1], Jian Yang [1,2], Peter M. Visscher [1] & Loic Yengo [1]

➡ LD and allele frequency differences between populations account for the majority of the loss of prediction accuracy

Introduction

Assume an additive genetic model with constant environment and effect sizes across populations:

➔ Observed genetic variation within and between-population results from the demographic and evolutionary history of populations

Evolutionary forces

- Genetic Drift

- Mutation
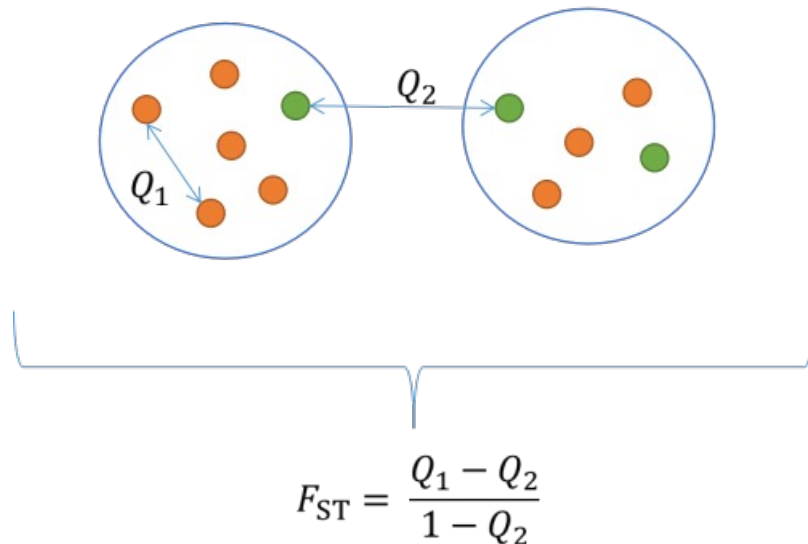
- Migration (gene flow)

- Selection

How do the evolutionary forces act on within and between-population genetic variation?

Population divergence model
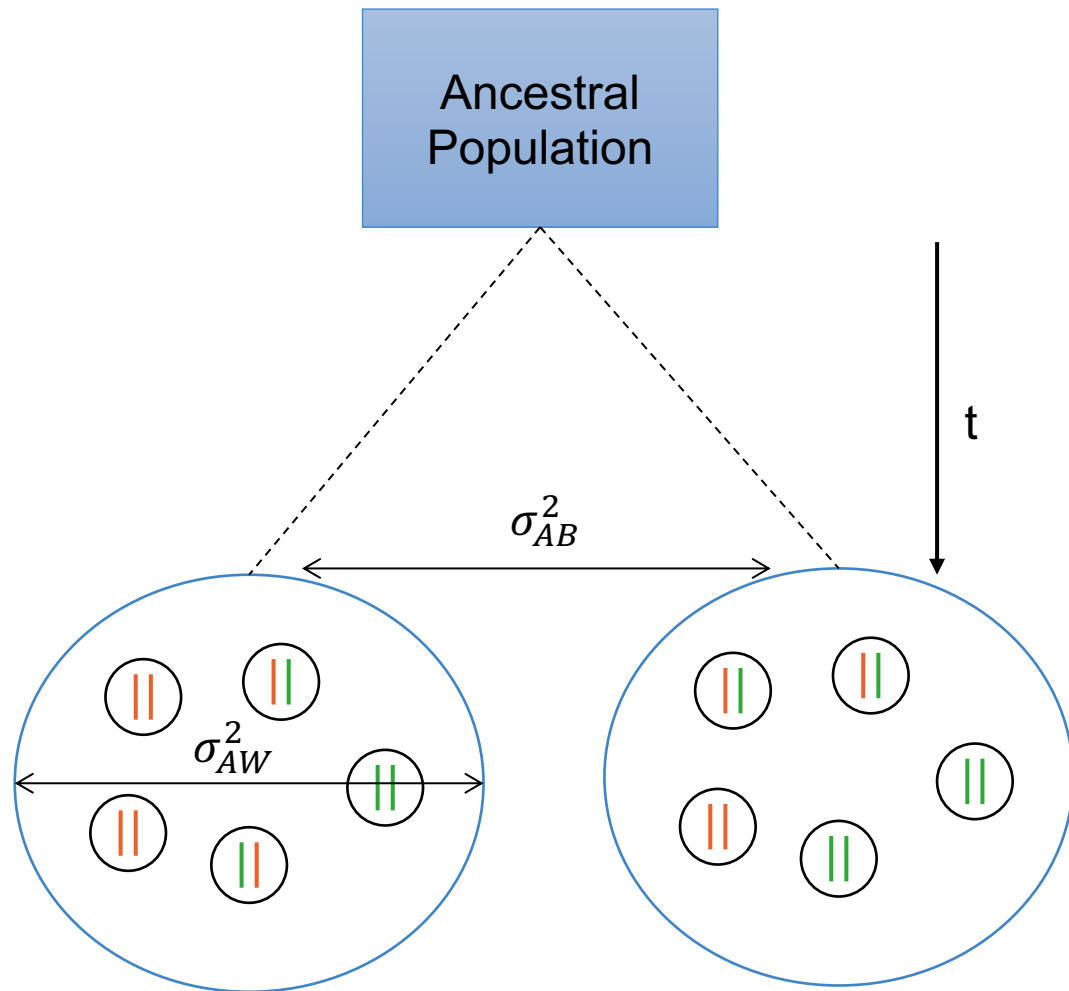
The total additive genetic variance of the meta-population can be partitioned into within (σ^2_{AW}) and between-population (σ^2_{AB}) additive genetic variance.

Wright (1951) fixation index F_{ST}

$$F_{ST} = \frac{Q_1 - Q_2}{1 - Q_2}$$

Q1 the probability of identity in state of a random pair of alleles sampled within a subpopulation
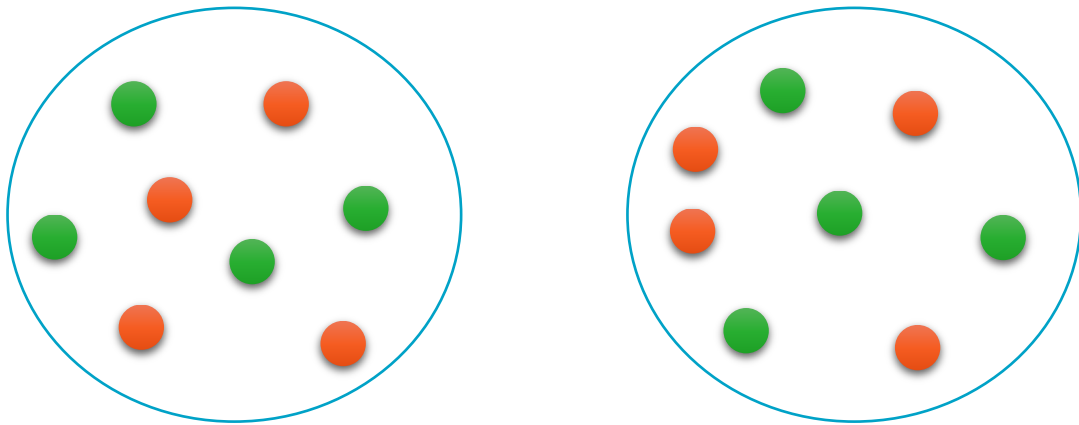
Q2 the probability of identity in state of a random pair of alleles sampled between subpopulations

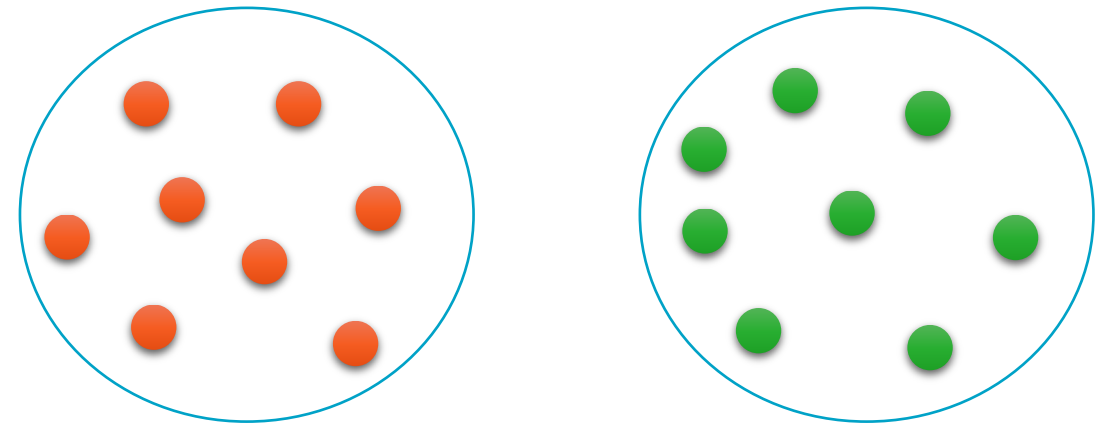F_{ST} can be estimated from marker data:

Weir and Cockerham (1984) ➜ estimator of F_{ST} in an anova framework

Wright (1951) fixation index F_{ST}

$$F_{ST} = 0$$

$$F_{ST} = 1$$

Evolutionary forces

- Genetic Drift

- Mutation

- Migration (gene flow)

- Selection

Genetic Drift

Assume a discrete and panmictic population of effective population size Ne diploid individuals.

Random sampling of $2Ne$ gametes to generate the next generation → Random change of allele frequencies

One biallelic locus with alleles A and B, p the allele frequency of allele A at generation t:

- $\mathbb{E}\left(p_{A_{t+1}}\right) = \text{p}$ but $\text{Var}\left(p_{A_{t+1}}\right) = \frac{p(1-p)}{2Ne}$
- $P(fixation) = p$
- $P(loss) = 1 - p$
- $\mathbb{E}(time\ to\ fixation) \approx 4N\text{e}$ generations
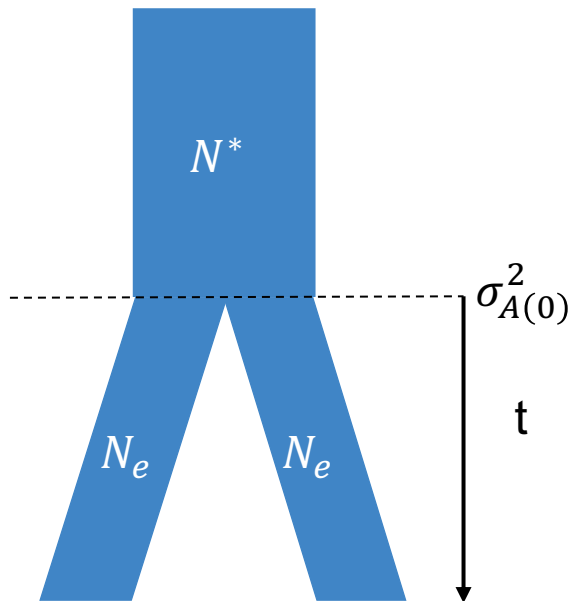
Genetic Drift

Genetic drift reduce heterozygosity over time:

$$H_t = H_0(1 - \frac{1}{2Ne})^t$$

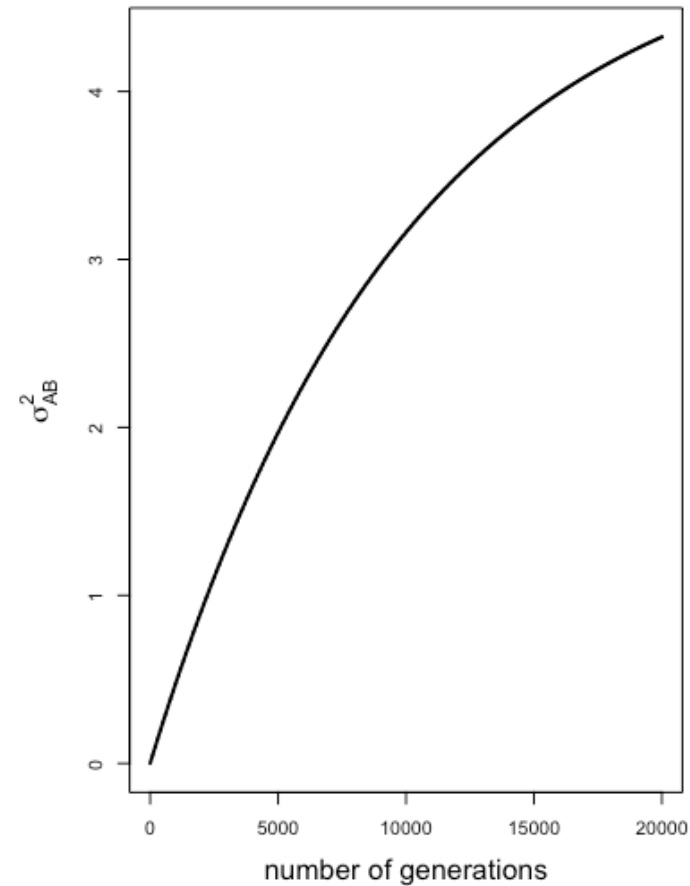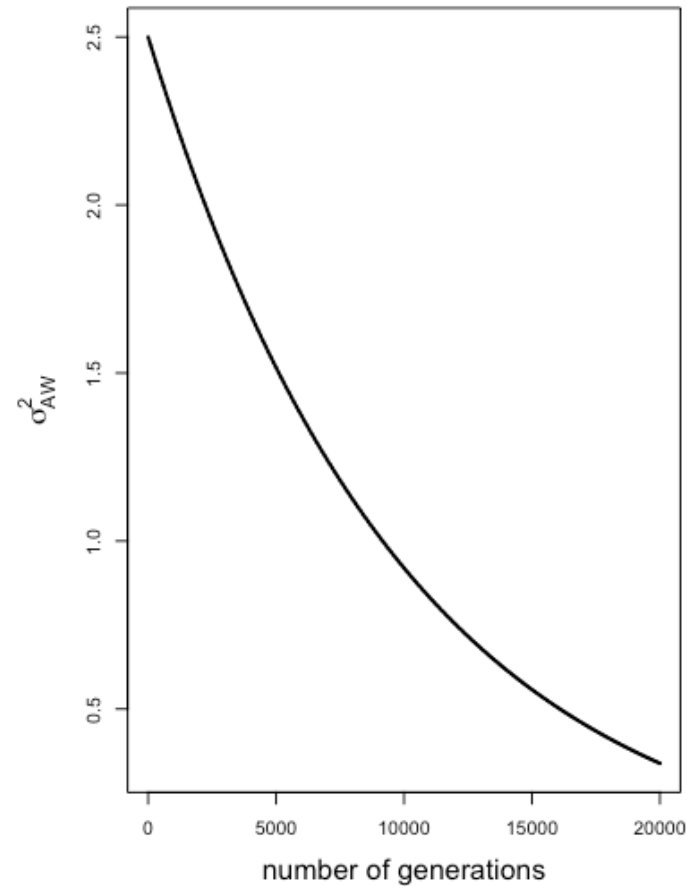Recall that $\sigma^2_{A(0)} = 2p(1-p)\beta^2 = H_0\beta^2$, therefore:

$$\sigma^2_{AW(t)} = \sigma^2_{A(0)}(1 - \frac{1}{2Ne})^t = \sigma^2_{A(0)}(1 - F_{ST}) \approx \sigma^2_{A(0)}e^{\frac{-t}{2Ne}}$$

$$\sigma^2_{AB(t)} = \left(\frac{1}{N_{f0}} + 2f_t\right)\sigma^2_{A(0)}$$

N^*

N_e N_e

$\sigma^2_{A(0)}$

t

With N_{f0} the number of founders per line, t the number of generations since divergence and $f_t = 1 - \left(1 - \frac{1}{2N_e}\right)^t = F_{ST}$. $\sigma^2_{A(0)}$ is the additive genetic variance in the base population at time t=0 (time of population split).

Genetic drift
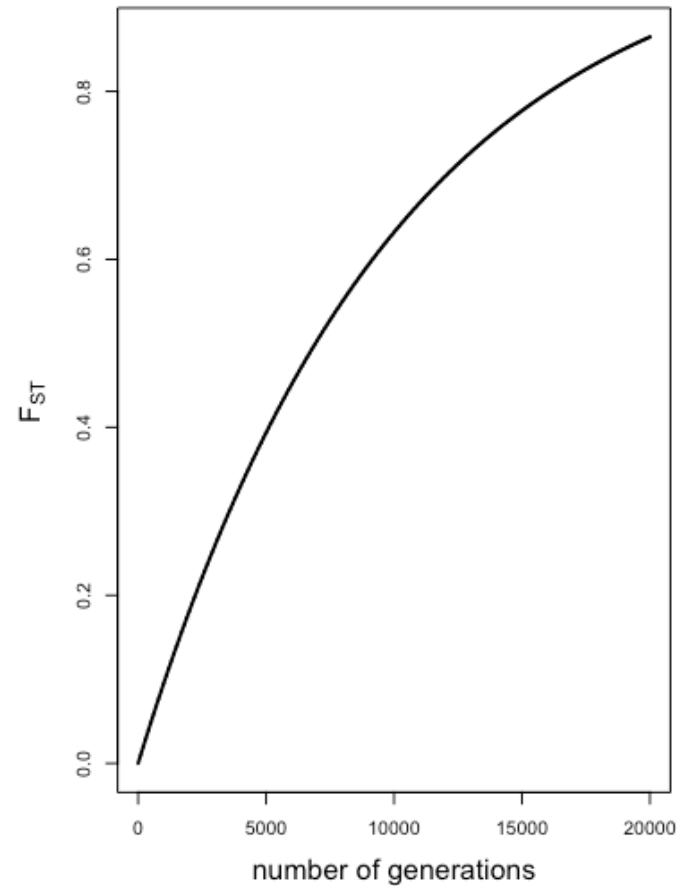
N_e = 5000 diploids
$\sigma^2_{A(0)}$ = 2.5

—— Genetic drift only

Evolutionary forces

- Genetic Drift ➔ removes within-population additive genetic variance over time and increase the between-population additive genetic variance

- Mutation

- Migration (gene flow)

- Selection

Evolutionary forces

- Genetic Drift ➔ removes within-population additive genetic variance over time and increase the between-population additive genetic variance

- Mutation

- Migration (gene flow)

- Selection

Mutations to the rescue of genetic variation

- Mutations introduce new genetic variation in the populations

  per base mutation rate

- The mutational variance $\sigma_m^2 = 2N_e L \mu \sigma_\beta^2$ is the amount of additive genetic variance introduced by new mutations at every generation

- $\dfrac{\sigma_m^2}{\sigma_e^2}$ is called the mutational heritability

  - Estimates range from 0.01 to 1e-4

Table 12.1 Estimates of the mutational heritability for a variety of organisms and characters.

Species	Character	h_m^2	Reference
Drosophila melanogaster	Abdominal bristle number	0.0035	See text
	Sternopleural bristle number	0.0043	See text
	Enzyme activities	0.0022	Clark et al. 1995b
			Harada 1995
	Ethanol resistance	0.0009	Weber and Diggins 1990
	Body weight	0.0047	Clark et al. 1995b
	Wing dimensions	0.0020	Santiago et al. 1992
	Viability	0.0003	Mukai 1964
			Mukai et al. 1972
			Cardellino and Mukai 1975
			Ohnishi 1977
Tribolium castaneum	Pupal weight	0.0091	Goodwill and Enfield 1971
Daphnia pulex	Life-history traits	0.0017	Lynch 1985
Mouse	Lengths of limb bones	0.0234	Bailey 1959
	Mandible measures	0.0231	Festing 1973
	Skull measures	0.0052	Carpenter et al. 1957
			Deol et al. 1957
			Hoi-Sen 1972
	6-week weight	0.0034	Caballero et al. 1995
Arabidopsis thaliana	Life-history traits	0.0039	Schultz et al. (in prep.)
Maize	Plant size	0.0112	Russell et al. 1963
	Reproductive traits	0.0073	Russell et al. 1963
Rice	Plant size	0.0030	Oka et al. 1958
	Reproductive traits	0.0028	Sakai and Suzuki 1964
Barley	Life-history traits	0.0002	Cox et al. 1987

From Lynch and Walsh (1998)

Mutations to the rescue of genetic variation
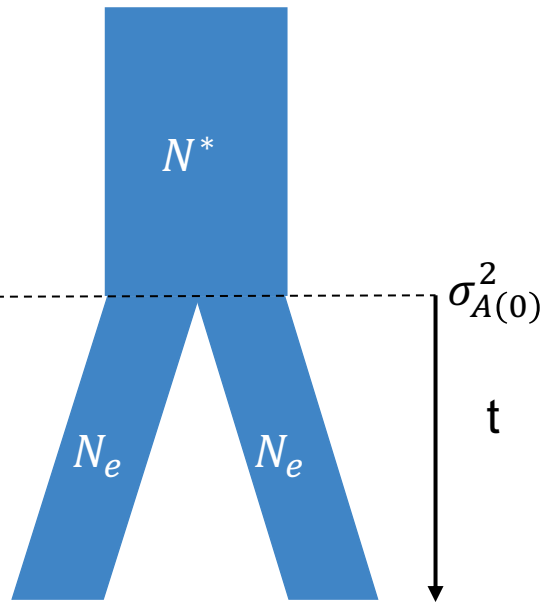
N^*

$\sigma^2_{A(0)}$

N_e N_e

t
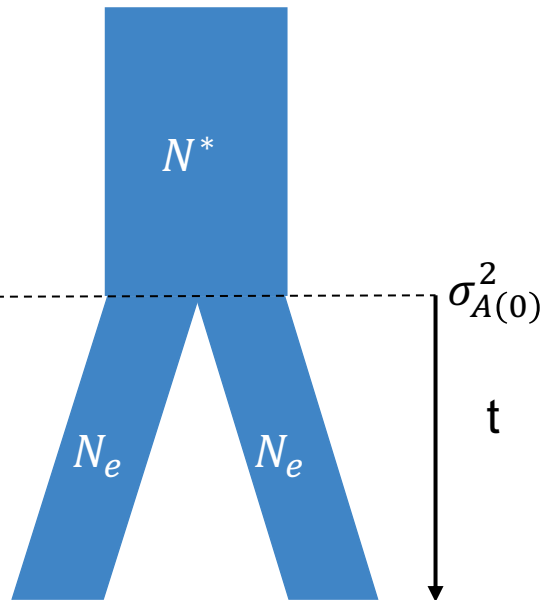
$$F_{\text{ST(t)}} = \frac{\theta_{w(t)} - \theta_{B(t)}}{1 - \theta_{B(t)}}$$

$$\theta_{w(t)} \approx \frac{1}{1 + 4Ne\mu} + \left(\theta_{w(0)} - \frac{1}{1 + 4Ne\mu}\right)\left(1 - 2\mu - \frac{1}{2Ne}\right) t$$

$\theta_{w(0)}$ is the IBD value before split: $\frac{1}{1+4N^*\mu}$ with N^* the effective population size of the ancestral population and Ne the effective population size of each derived population.

$$\theta_{B(t)} \approx \theta_{w(0)}(1 - \mu)^{2t}$$

Mutations to the rescue of genetic variation

$$\sigma^2_{AW(t)} = 2Ne\sigma^2_m + \left(\sigma^2_{A(0)} - 2Ne\sigma^2_m\right) \times e^{\frac{-t}{2Ne}}$$

$$= \sigma^2_{A(0)} \times e^{\frac{-t}{2Ne}} + 2Ne\sigma^2_m(1\text{-}\ e^{\frac{-t}{2Ne}})$$
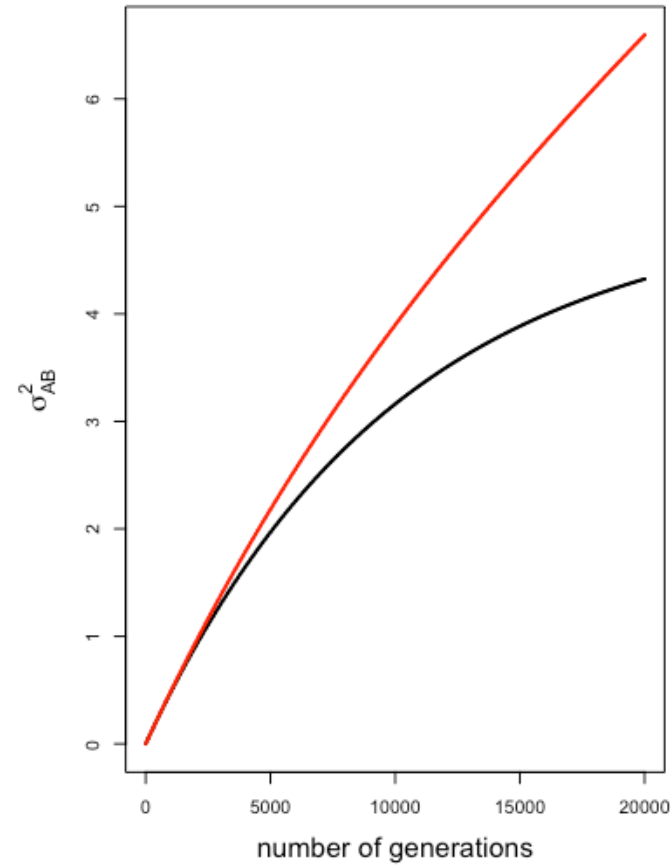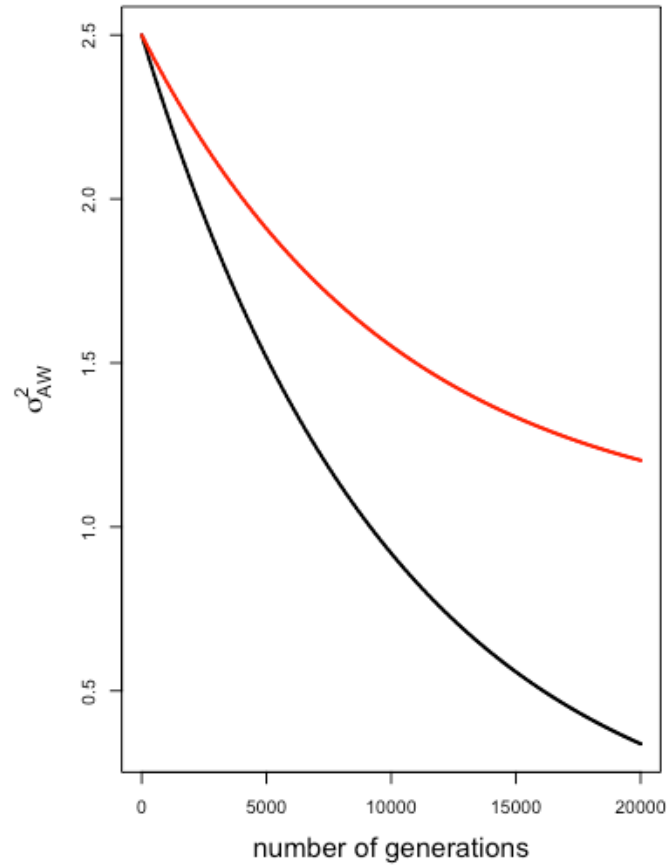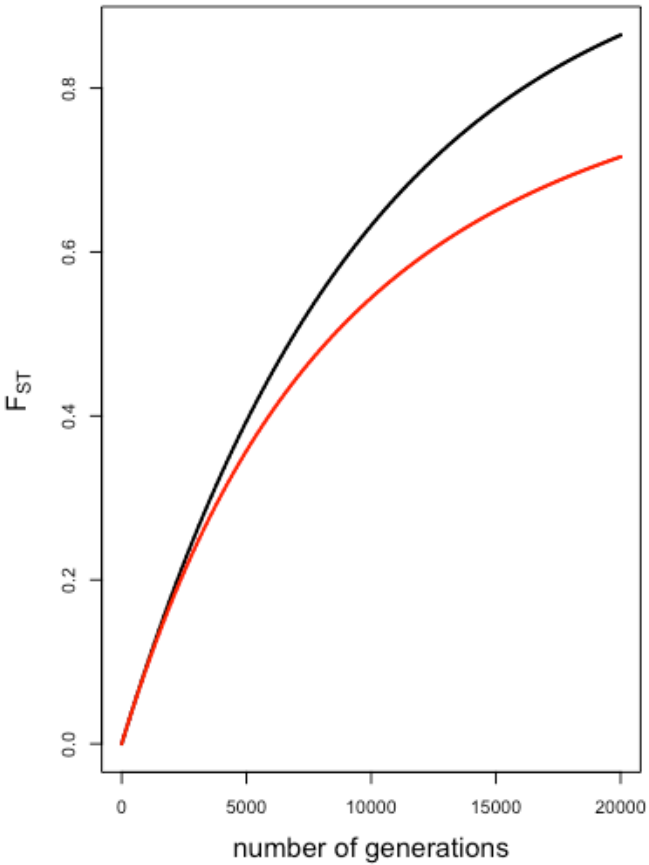
mutation-drift equilibrium (when $4Ne\mu \ll 1$) $\rightarrow$ $\sigma^2_{AW(t)} = 2Ne\sigma^2_m$

$$\sigma^2_{AB(t)} = 2t\sigma^2_m + 2\left(\sigma^2_{A(0)} - 2Ne\sigma^2_m\right) \times (1 - e^{\frac{-t}{2Ne}})$$

$$\sigma^2_{AB} \approx 2t\sigma^2_m \text{ for very large t}$$

N^*

$\sigma^2_{A(0)}$

N_e N_e

t

If the ancestral population is in mutation-drift equilibrium, then $\sigma^2_{A(0)} = 2N^*\sigma^2_m$ when $4N^*\mu \ll 1$.

Mutations to the rescue of genetic variation

N_e = 5000 diploids
$\mu = 10^{-8}$
$\sigma^2_{A(0)} = 2.5$
$\sigma^2_m = 1e - 4$

— Genetic drift only
— Genetic drift + mutation

Evolutionary forces

- Genetic Drift ➔ removes within-population additive genetic variance over time and increase the between-population additive genetic variance

- Mutation ➔ decreases the effect of drift from the base population variation and increase the between-population additive genetic variance

- Migration (gene flow)

- Selection

Evolutionary forces

- Genetic Drift ➔ removes within-population additive genetic variance over time and increase the between-population additive genetic variance

- Mutation ➔ decreases the effect of drift from the base population variation and increase the between-population additive genetic variance

- Migration (gene flow)

- Selection

Gene flow

Gene flow homogenize the allele frequencies within populations
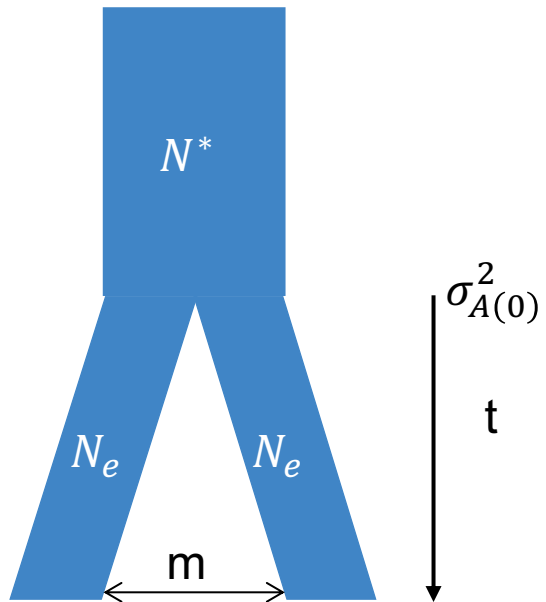
Cockerham and Weir (1987)

N^*

$\sigma^2_{A(0)}$

t

N_e N_e

m

$$\theta_{2,t+1} = \rho[a(1 - \gamma + \gamma\theta_{2,t}) + (1 - a)\theta_{3,t}]$$

$$\theta_{3,t+1} = \rho[b(1 - \gamma + \gamma\theta_{2,t}) + (1 - b)\theta_{3,t}]$$

$$\theta_{2,t+1} - \theta_{3,t+1} = \rho d[\gamma(\theta_{2,t} - \theta_{3,t}) + (1 - \gamma)(1 - \theta_{3,t})].$$

$$F_{ST} = \frac{\theta_2 - \theta_3}{1 - \theta_3}$$

mutation-drift-migration equilibrium ➔ $F_{ST} \approx \dfrac{1}{1 + 4Ne(\mu + m\frac{n_d}{n_d - 1})}$

Migration rate between demes

$$\rho = (1 - \mu)^2$$

$$\gamma = 1 - \frac{1}{2N}$$

$$a = (1 - m)^2 + \frac{m^2}{n_d - 1}$$

$$b = \frac{1 - a}{n_d - 1}$$

$$(a - b) = d = (1 - m\alpha)^2$$

$$\alpha = \frac{n_d}{n_d - 1}$$

Number of deme

Gene flow

From Lande (1972), we have the following effects of gene flow on the within and between-population additive genetic variance:
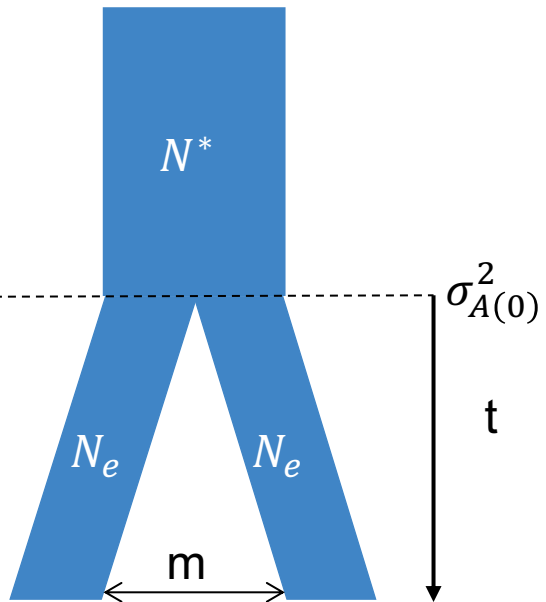
$$\Delta\sigma^2_{AW} \approx m\,\sigma^2_{AB} - \frac{1}{2Ne}\sigma^2_{AW} + \sigma^2_m$$

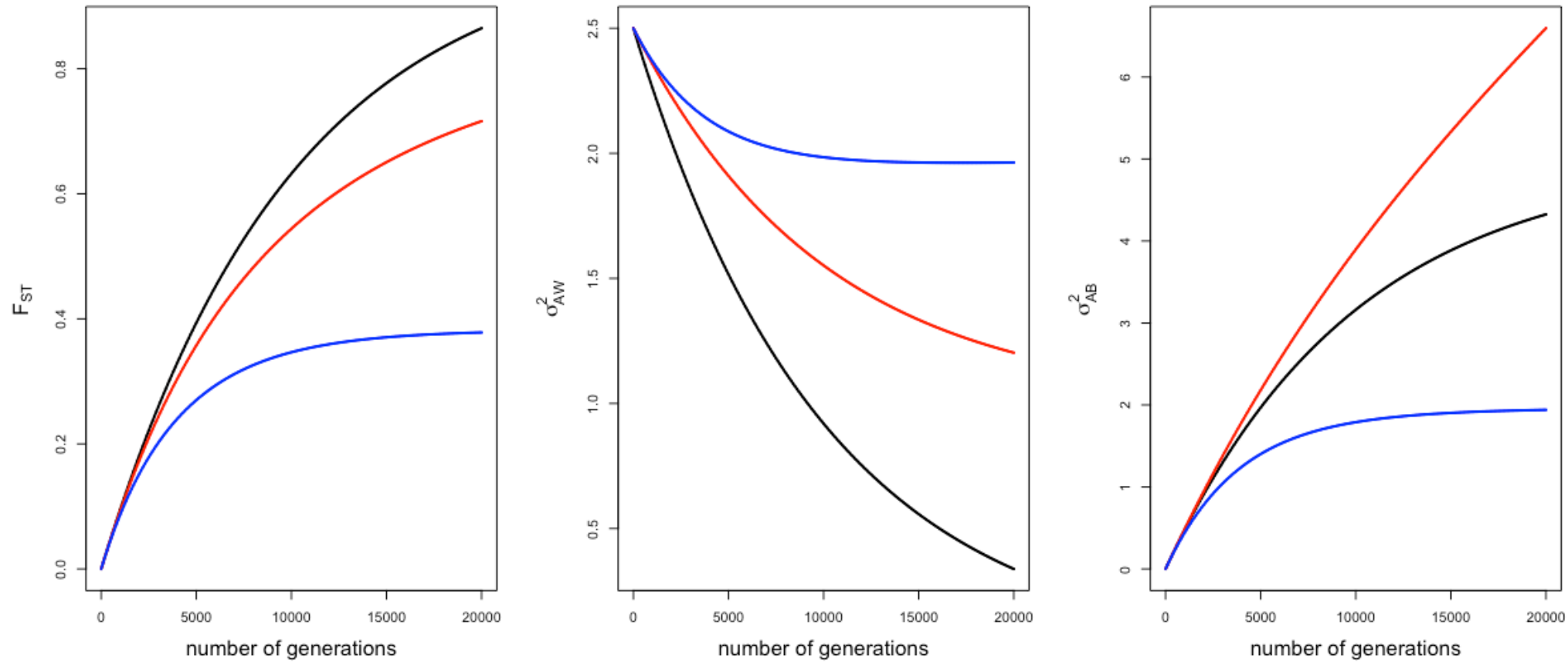$$\Delta\sigma^2_{AB} \approx -\frac{2mn_d}{n_d - 1}\sigma^2_{AB} + \frac{1}{Ne}\sigma^2_{AW}$$

At mutation-drift-migration equilibrium:

$\sigma^2_{AW} = 2n_d Ne\sigma^2_m$ ➜ n_d times the expected within-population additive genetic variance in a single panmictic population with genetic drift and mutation

$$\sigma^2_{AB} = \frac{n_d - 1}{m}\,\sigma^2_m$$

N^*

N_e N_e

$\sigma^2_{A(0)}$

t

m

Gene flow

N_e = 5000 diploids
$\mu = 10^{-8}$
$\sigma^2_{A(0)} = 2.5$
$\sigma^2_m = 1e - 4$
$m = 5e - 5$

— Genetic drift only
— Genetic drift + mutation
— Genetic drift + mutation + migration

Evolutionary forces

- Genetic Drift ➔ removes within-population additive genetic variance over time and increase the between-population additive genetic variance

- Mutation ➔ decreases the effect of drift from the base population variation and increase the between-population genetic variance

- Migration (gene flow) ➔ homogenize the within and decrease between-population additive genetic variance

- Selection

$$Q_{ST} = \frac{\sigma^2_{AB}}{\sigma^2_{AB} + 2\sigma^2_{AW}}$$ (Spitze, 1993), the trait-based analog of F_{ST}

Q_{ST} vs F_{ST} (estimated from putatively neutral markers) comparisons

Table 12.2 Interpretation of Q_{ST} versus F_{ST} comparisons.

Observation	Interpretation
$Q_{ST} = F_{ST}$	Consistent with divergence expected under drift. Does not rule out selection, but does not support it either.

$$Q_{ST} = \frac{\sigma_{AB}^2}{\sigma_{AB}^2 + 2\sigma_{AW}^2}$$ (Spitze, 1993), the trait-based analog of F_{ST}.

Q_{ST} vs F_{ST} (estimated from putatively neutral markers) comparisons

Table 12.2 Interpretation of Q_{ST} versus F_{ST} comparisons.

Observation	Interpretation
$Q_{ST} > F_{ST}$	**Divergent selection**: spatial variation in trait values in excess of neutral expectation.
$Q_{ST} = F_{ST}$	Consistent with divergence expected under drift. Does not rule out selection, but does not support it either.
$Q_{ST} < F_{ST}$	**Convergent selection**: spatial variation in trait values less than neutral expectation. Similar trait values are favored over populations.

Evolutionary forces

- Genetic Drift ➜ removes within-population additive genetic variance over time and increase the between-population additive genetic variance

- Mutation ➜ decreases the effect of drift from the base population variation and increase the between-population additive genetic variance

- Migration (gene flow) ➜ homogenize the within and decrease between-population additive genetic variance

- Selection

Selection

https://www.apsnet.org/
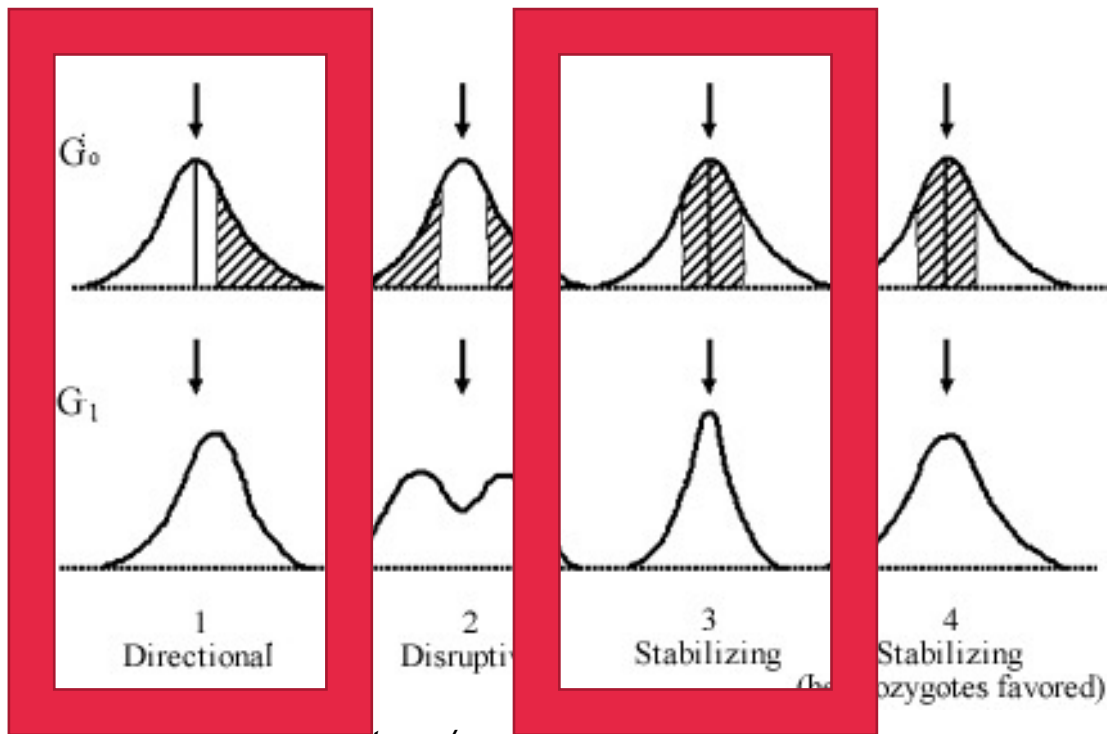
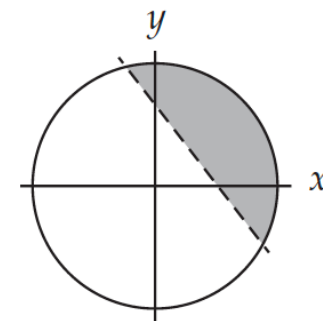Bulmer (1971)

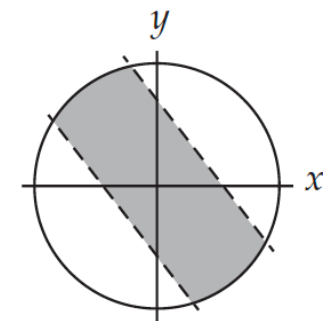$$\sigma_A^2 = \sigma_a^2 + d$$

Genic variance   covariance between allelic effects at different loci

Directional and stabilizing selection: $d < 0$

Walsh and Lynch (2018)
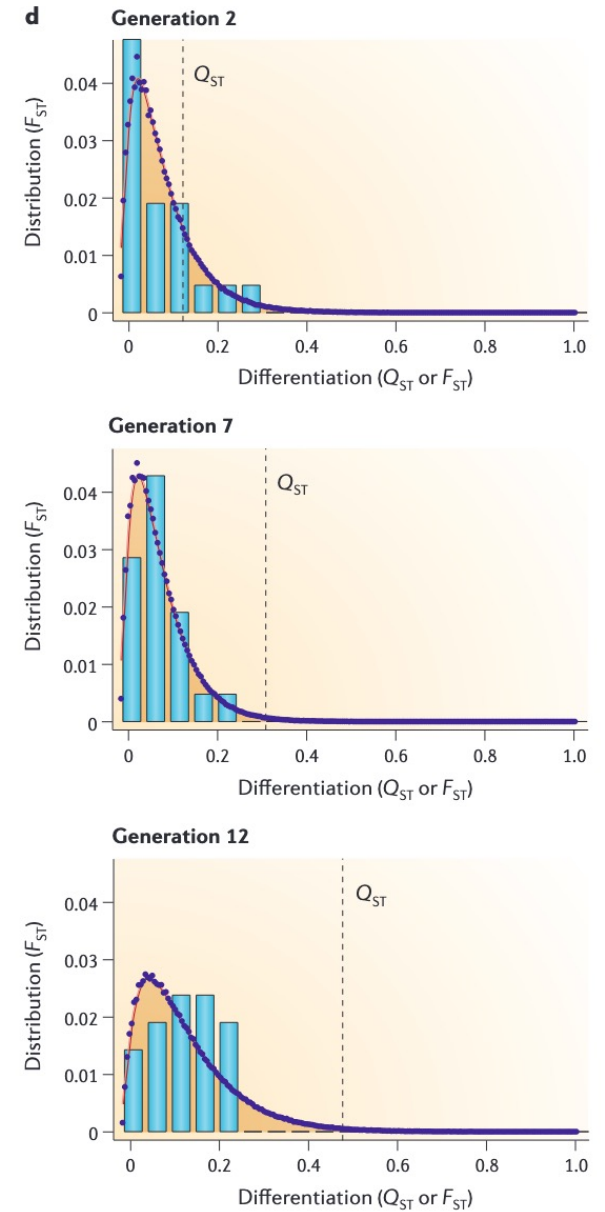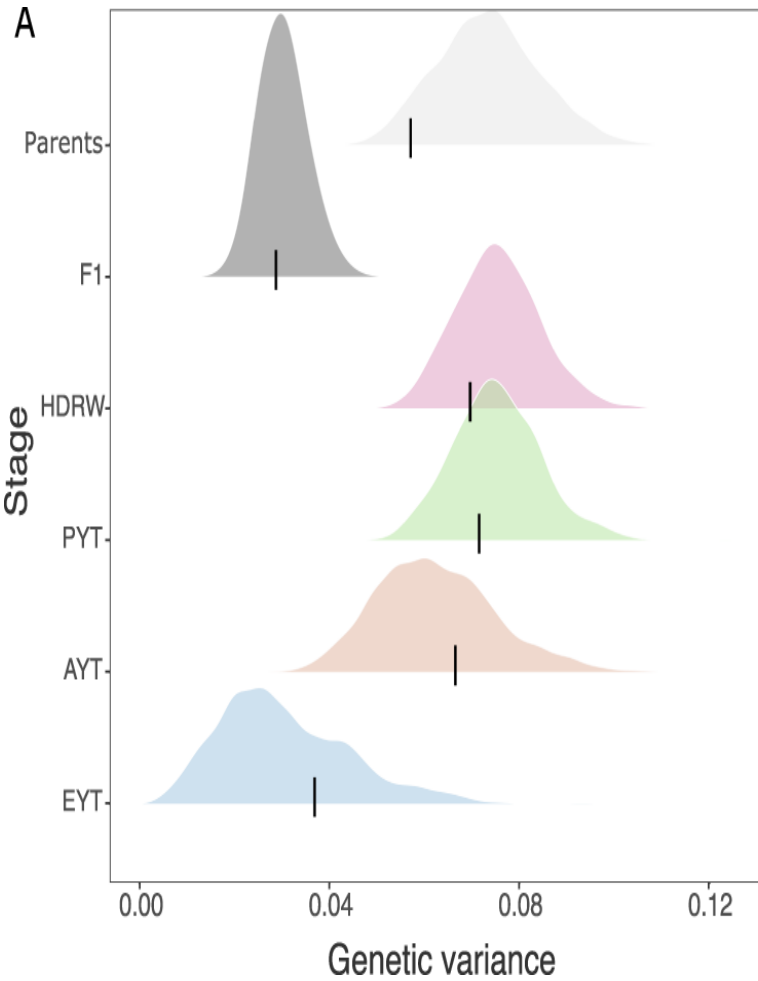
Selection
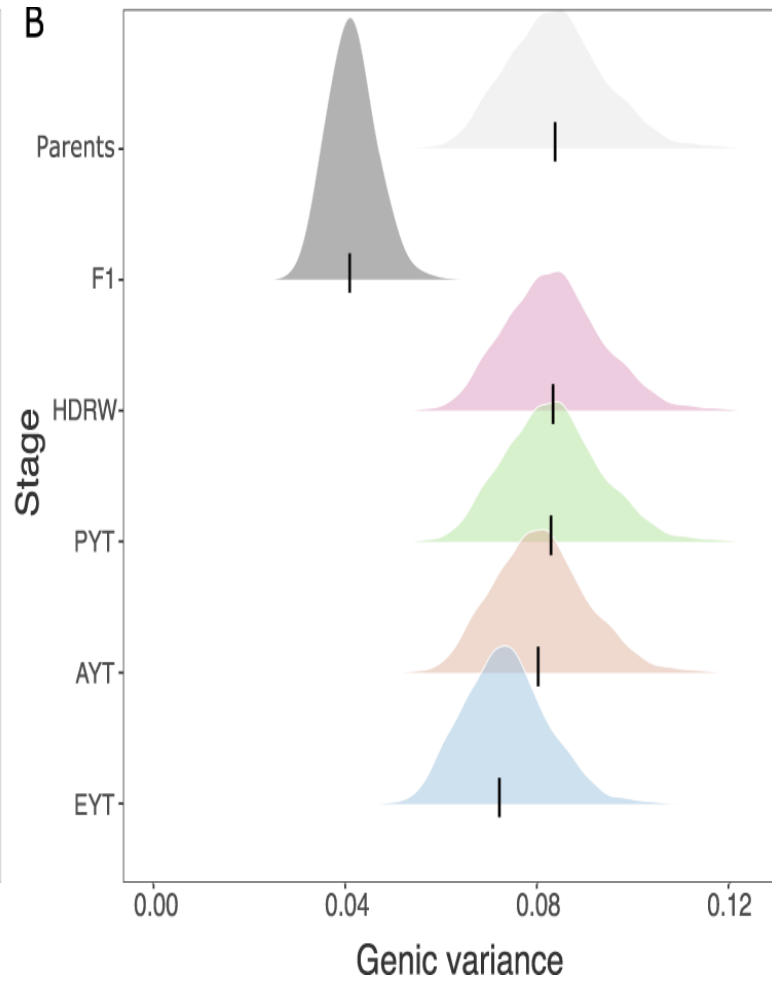
De C. Lara *et al.* (2022)

Leinonen *et al.* (2013)

Selection

Directional selection:

Decrease the within-population additive genetic variance and increase the between-population additive genetic variance

➔ $Q_{ST} > F_{ST}$

Stabilizing selection:

Decrease the within and between-population additive genetic variance

➔ $Q_{ST} < F_{ST}$

Conclusion

- Genetic Drift ➔ removes within-population additive genetic variance over time and increase the between-population additive genetic variance

- Mutation ➔ decreases the effect of drift from the base population variation and increase the between-population genetic variance

- Migration (gene flow)➔ homogenize the within and decrease between-population additive genetic variance

- Selection ➔ removes within-population additive genetic variation and can increase or decrease the between-population additive genetic variance

Questions?

In humans:

- Empirical data are consistent with causal variants being largely shared across populations

  ➔ Small mutational heritability ?

  ➔ Pervasive stabilizing selection ?

nature COMMUNICATIONS

ARTICLE
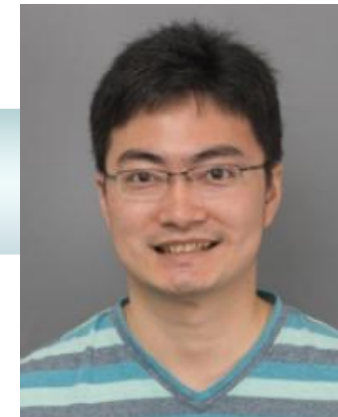
https://doi.org/10.1038/s41467-021-21446-3   OPEN

Widespread signatures of natural selection across human complex traits and functional genomic categories

Jian Zeng[1✉], Angli Xue[1], Longda Jiang[1], Luke R. Lloyd-Jones [1], Yang Wu [1], Huanwei Wang [1], Zhili Zheng[1], Loic Yengo [1], Kathryn E. Kemper[1], Michael E. Goddard[2,3], Naomi R. Wray [1,4], Peter M. Visscher [1] & Jian Yang [1,5,6✉]

Acknowledgments

All the colleagues from the Program in
Complex Trait Genomics group