

Breeding and quantitative genetic insight on the genetic of populations

AX 100217

AGRY611

Outline

1. Introduction

1. Hardy-Weinberg equilibrium
2. Fisher-Wright process
3. Beavis effect

2. Changes in allele frequency

1. Mutation
2. Migration
3. Selection
4. Genetic drift
5. Hill-Robertson effect

3. Population management

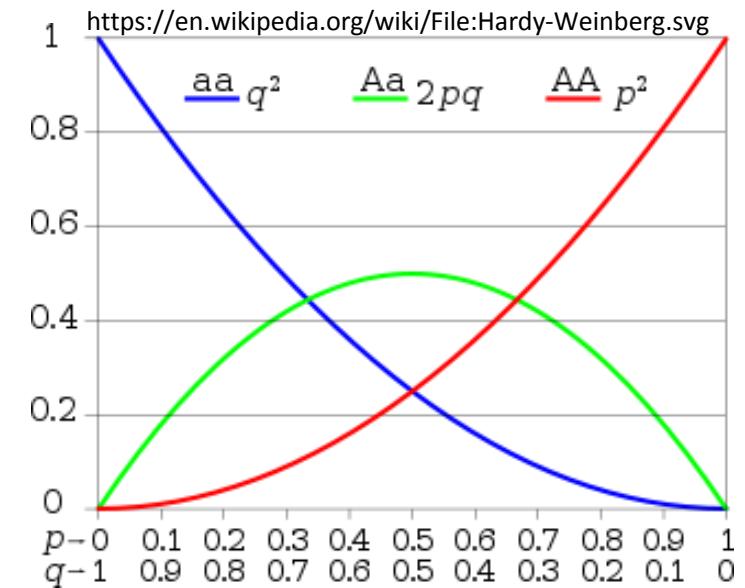
1. Effective population size
2. Short and long term gains
3. Inbreeding and heterosis
4. Pedigree and relatedness

1. Introduction

1.1 Hardy-Weinberg equilibrium

Hardy-Weinberg equilibrium

- **STATEMENT:** Allele and genotype frequencies are constant across generations. It is used to define a hypothetical “idealized population”. HWE can always be achieved with one generation of random mating.
- **IN STATISTICS:**
 - Alleles follow a **binomial distribution**
 - Mean: $E[x] = 2p$
 - Variance: $D[x] = 2p(1 - p)$
- **CONDITIONS:**
 - Random mating
 - No evolutionary pressure (mutation, selection, migration)
 - No overlapping populations
- **FOR BREEDING:** Observed within (but not across) F2 populations and open pollination populations. Used to create genetic maps and to estimate segregation distortion. HWE is often assumed to simplify calculations.



Hardy-Weinberg equilibrium

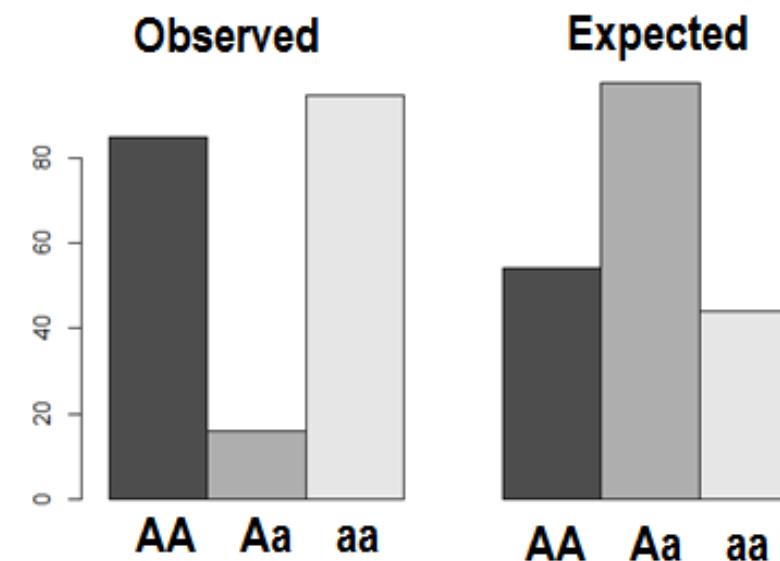
- **Notation (single-locus)**
 - Probability of observing each state of an allele
 - $\text{Pr}(A) = p$
 - $\text{Pr}(a) = q = (1 - p)$
 - Probability of observing each state of a locus (genotypic frequencies)
 - $\text{Pr}(AA) = p^2$
 - $\text{Pr}(Aa) = 2p(1 - p)$
 - $\text{Pr}(aa) = (1 - p)^2$
- **Single-locus expression:**
 - Sum of probabilities always add to 1
 - Allele: $p + (1 - p) = 1$
 - Locus: $p^2 + 2p(1 - p) + (1 - p)^2 = 1$
- **Two-loci case:**
 - Allele: $\text{Pr}(AB), \text{Pr}(Ab), \text{Pr}(aB), \text{Pr}(ab)$
 - Loci: $\text{Pr}(AABB), \text{Pr}(AABb), \text{Pr}(AAAb), \dots$
 - Takes many generations for neighbor loci HWE
- **FOR BREEDING:** Allelic and genotypic frequencies from multiple loci across the genome are used to detect which genomic changes are occurring due to selection, drift, migration, bottlenecks.

Hardy-Weinberg equilibrium

- **Checking HWE:** Do my genotypic frequencies (the proportions of AA, Aa, aa) match with what is expected from the allele frequencies (p and q)?

```
> x
 [1] 2 0 2 0 1 0 2 2 0 0 1 2 0 2 2 0 0 0 1 2 1 0 2 2 0 0 2
[28] 0 0 2 0 0 0 2 2 0 0 0 0 0 0 2 1 1 2 0 2 2 2 0 2 0 2 2
[55] 2 2 2 2 2 1 2 2 2 2 0 2 0 2 2 2 2 0 2 2 2 2 2 2 2 2 2
[82] 0 0 0 0 0 0 2 2 0 2 0 2 2 2 0 0 1 0 0 2 2 2 0 0 2 0
[109] 0 2 0 0 1 2 0 0 0 2 2 0 0 2 1 2 2 2 2 2 0 0 0 1 0 2 2
[136] 2 1 2 2 2 1 1 1 0 2 0 0 0 0 2 0 0 2 0 2 2 2 0 0 1 2 0
[163] 0 0 0 2 0 2 0 2 0 0 0 2 2 2 2 0 0 2 0 2 2 2 2 0 2 2 0
[190] 0 2 2 0 0 2 0
> p = mean(x)/2
> n = length(x)
> Expectation = c(AA=p^2,Aa=2*p*(1-p),aa=(1-p)^2) * n
> round(Expectation)
AA Aa aa
54 98 44
> table(x)
x
 0  1  2
85 16 95
> chisq.test( table(x), p = Expectation )
Chi-squared test for given probabilities

data: table(x)
X-squared = 144.62, df = 2, p-value < 2.2e-16
```

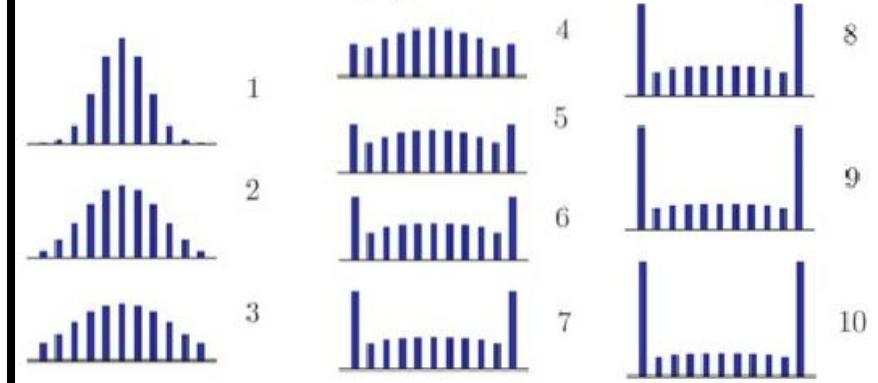


1.2 Fisher-Wright process

Fisher-Wright process

- **CONDITIONS:**
 - Finite populations
 - Non-overlapping populations
 - Diploid behavior
 - Ongoing selection
 - Allows for mutation
- **CONSEQUENCE:** More realistic model. Not commonly observed in nature, that is an acceptable scenario of plant breeding. **FW process describes how alleles are driven to fixation in finite populations.** It allows to trace the proportion of alleles in any generation (forward and backward).
- **WATCH OUT FOR:** FW process is an stationary Markov Chain, the rate of fixation is estimated from a small time frame and extrapolated. However, the relative contribution of any given locus or allele to complex traits is believed to change overtime.
- **FOR BREEDING:** Regression-type models for genomic prediction must be recalibrated every so often to capture the dynamic changes in the FW model: changes in population composition, allele effect, etc.

Figure 2.3: Genetic drift in the Wright–Fisher model with $N = 5$ and $p = \frac{1}{2}$. The abscissae show the population number of A genes.

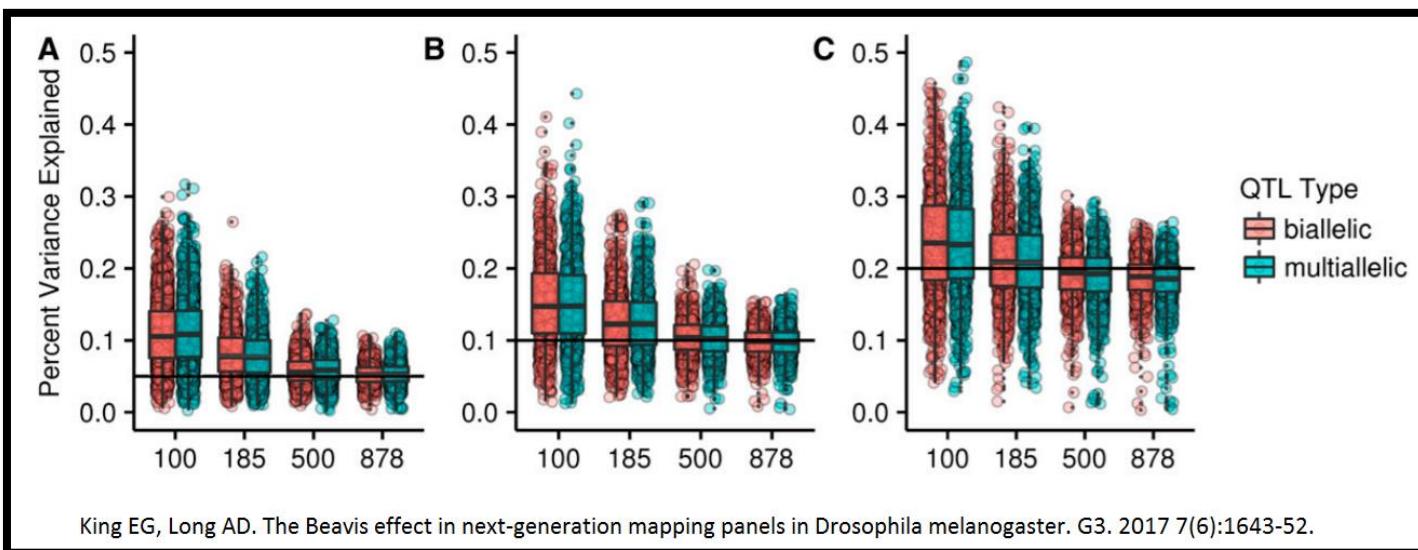


Christiansen, F. B. (2014). Theories of population variation in genes and genomes. Princeton University Press.

1.3 Beavis effect

Beavis effect

- **DEFINITION:** A matter of statistical power in genetics. Small population sizes ($N=100$) greatly overestimate the estimate effect of a QTL, whereas a reasonable population ($N=1000$) provides a good approximation.
- **Source:** Beavis WD (1998) QTL analyses: power, precision, and accuracy. Molecular dissection of complex traits, 1998, 145-162.
- **FOR BREEDING:** Population size has a major effect on genetic studies in applied breeding: QTL identification for MAS and genome-wide prediction.



Xu, S. (2003). Theoretical basis of the Beavis effect. Genetics, 165(4), 2259-2268.

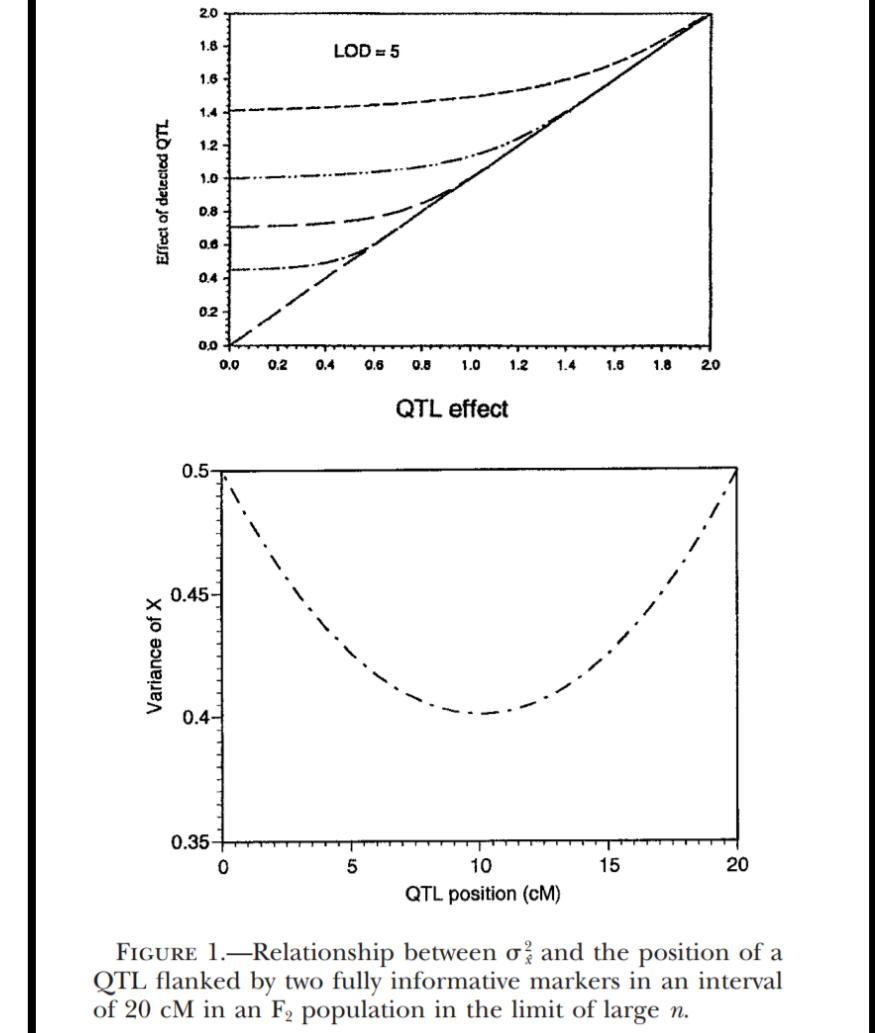


TABLE 4
**Comparisons of predicted and observed (estimated) biases in estimated QTL effects and variances from
 Beavis F₂ simulation experiments**

Simulated conditions ^a	Variance explained			Additive effect			Average estimated location
	Simulated	Observed	Predicted ^b	Simulated	Observed	Predicted ^c	
10-30-100	3.00	16.76	16.0537	2.45	4.96	5.6410	11.3
10-30-500	3.00	4.33	4.1890	2.45	2.89	2.8617	10.53
10-30-1000	3.00	3.02	3.1846	2.45	2.56	2.4868	10.8
10-63-100	6.25	12.65	16.5984	3.55	4.68	5.7328	10.51
10-63-500	6.25	7.08	6.5581	3.55	3.73	3.5829	10.96
10-63-1000	6.25	6.34	6.3566	3.55	3.60	3.5500	11.04
10-95-100	9.50	18.68	17.3883	4.36	5.85	5.8466	10.58
10-95-500	9.50	10.1	9.7082	4.36	4.49	4.3607	11.08
10-95-1000	9.50	9.67	9.6028	4.36	4.44	4.3600	11.19
40-30-100	0.75	15.78	15.6270	1.22	4.40	5.5436	10.83
40-30-500	0.75	3.17	3.3332	1.22	2.35	2.5671	10.17
40-30-1000	0.75	1.46	1.7961	1.22	1.85	1.8790	10.17
40-63-100	1.56	16.31	15.7983	1.77	4.71	5.5999	10.45
40-63-500	1.56	3.54	3.5783	1.77	2.59	2.6582	10.13
40-63-1000	1.56	1.96	2.1435	1.77	2.09	2.0494	10.37
40-95-100	2.40	16.55	15.9694	2.18	5.02	5.6236	10.45
40-95-500	2.40	3.97	3.9190	2.18	2.79	2.7641	10.12
40-95-1000	2.40	2.58	2.6970	2.18	2.36	2.2784	10.29

^a Numerical values denote the number of QTL-heritability-number of progeny.

^b Using Equation 17.

^c Using Equation 8.

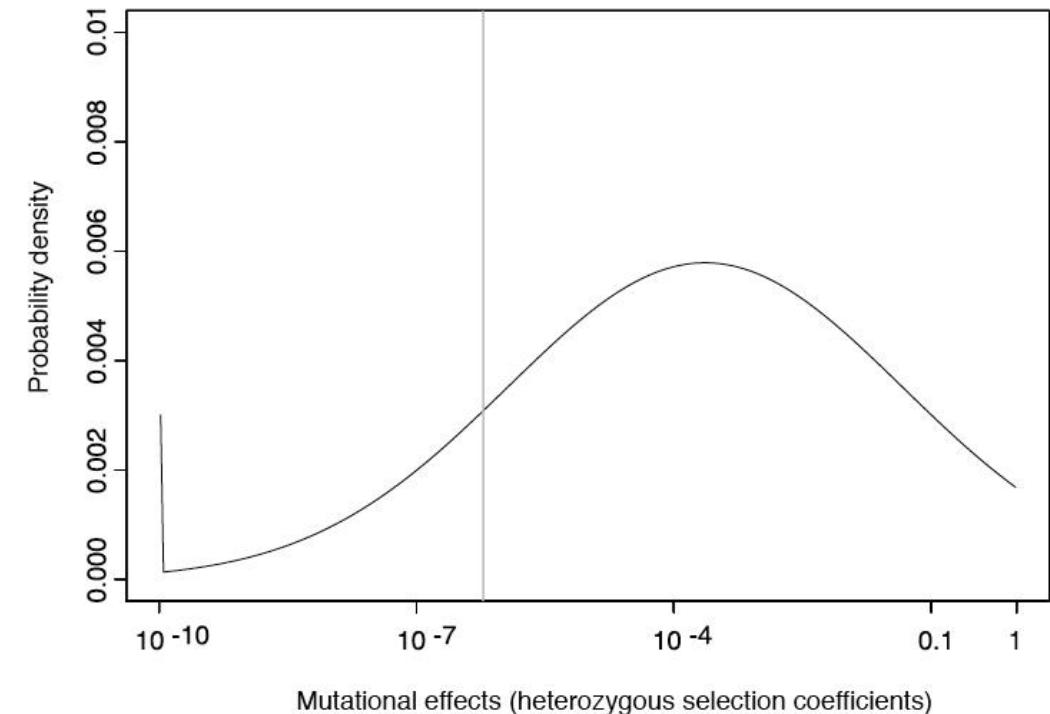
2. Changes in allele frequency

2.1 Mutation

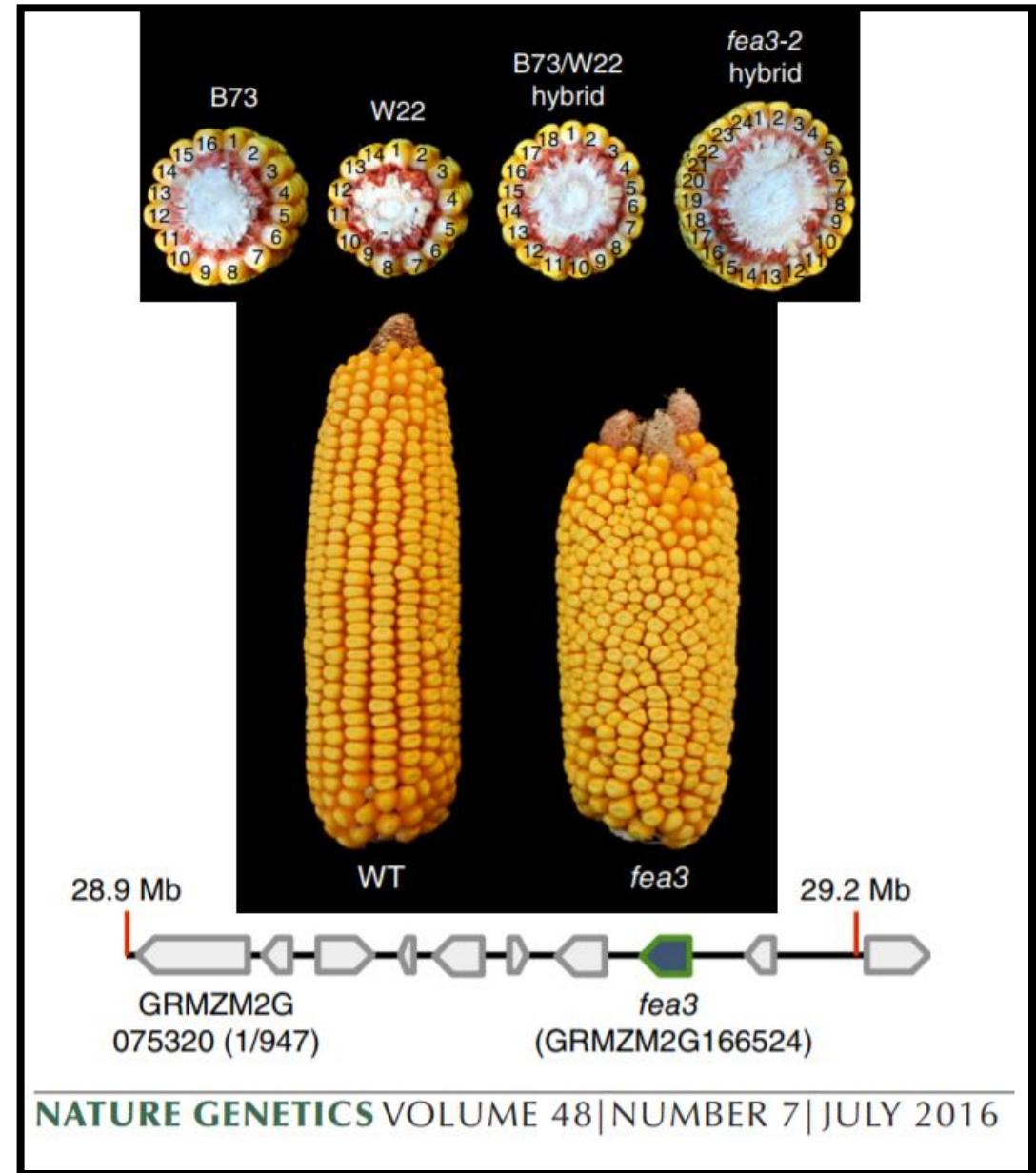
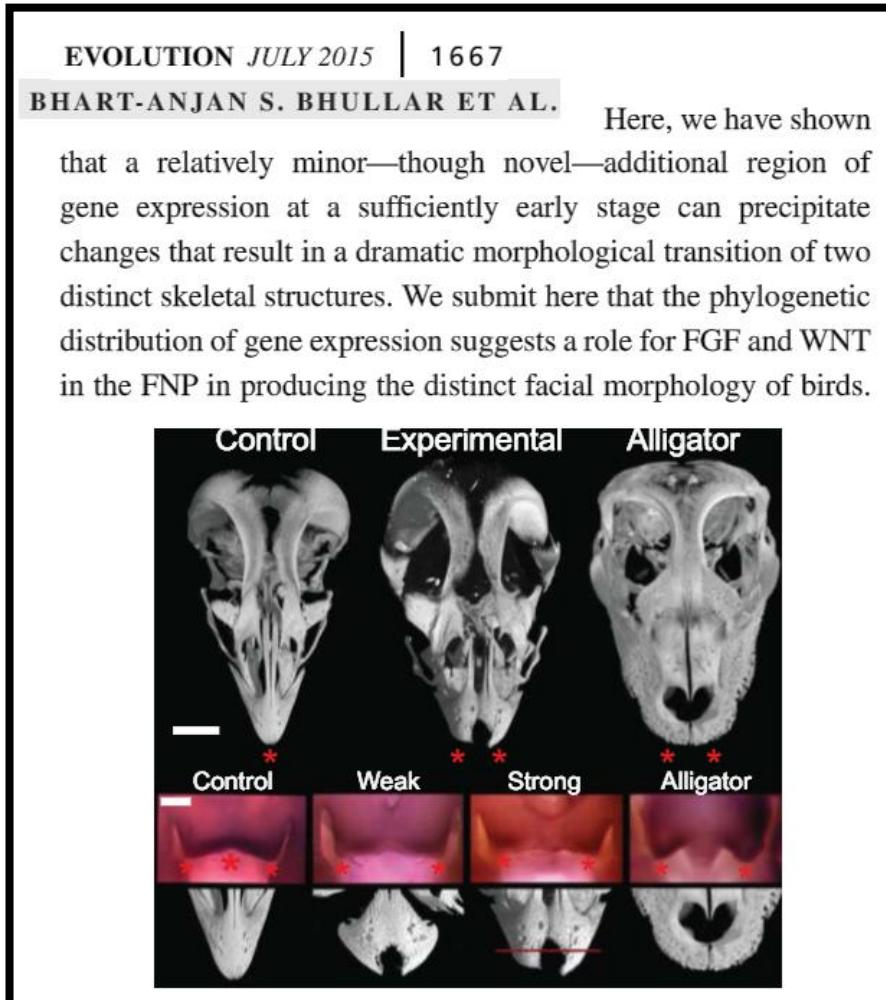
Mutation

- **RELEVANCE**: Mutation is the only natural source of new variation in all species. Important for phylogenetics to trace speciation.
- **RANDOMNESS**: Mutation are more likely to occur in some parts of the genome, such as *euchromatin* (less condensed DNA).
- **CHANCES**: Mutagenic agents increases the probability of mutation from 10^{-7} to 10^{-4}
- **FOR BREEDING**: Mutational breeding is deprecated in crops, but it is still big deal in some clonally propagated species. The updated version is called '***gene editing***' based on CRISPR-Cas9 (site specific changes as opposed to random modification).
- **WATCH OUT FOR**: Most mutation have deleterious effects.

Loewe, L. (2008) Genetic mutation. *Nature Education* 1(1):113



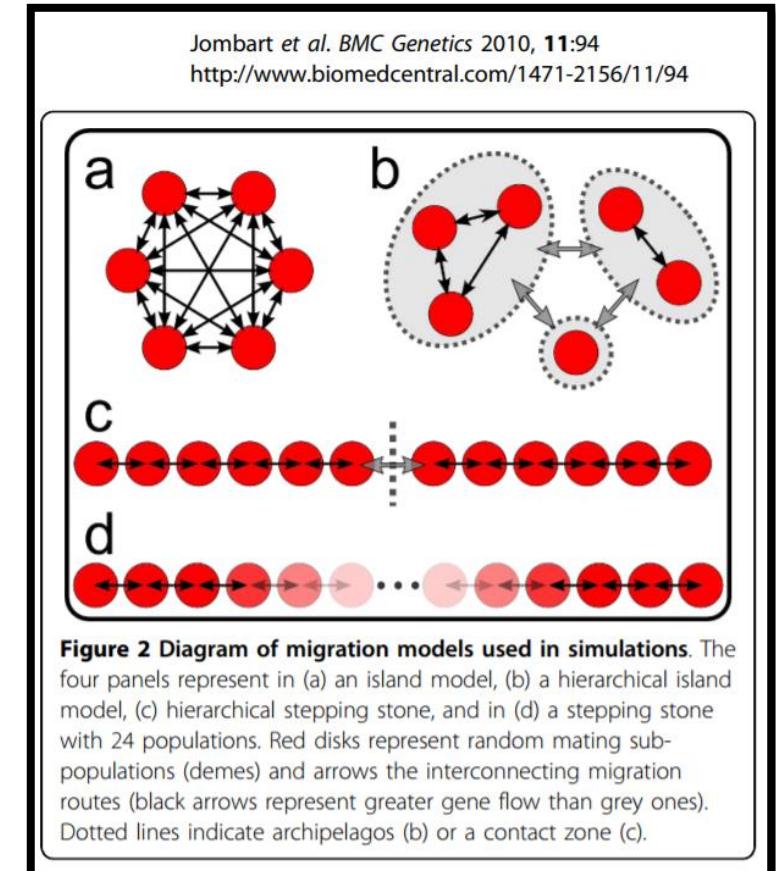
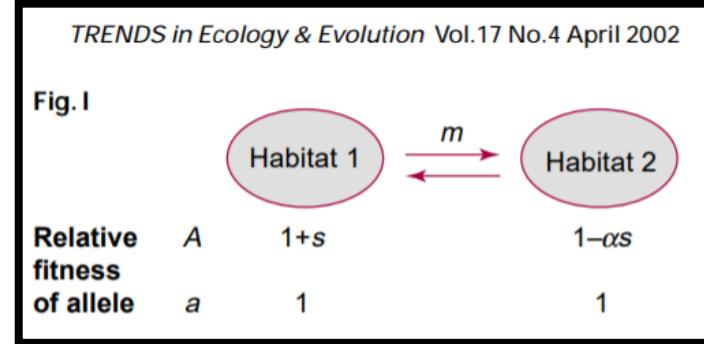
Examples



2.2 Migration

Migration and gene flow

- **RELEVANCE:** Exogenous sources of alleles from other populations
- **FOR NATURE:** Outbreeding
- **FOR BREEDING:** Occurs intentionally or unintentionally
 - (1) exchange of genetic material among breeders;
 - (2) incorporation of new material;
 - (3) seed or pollen contamination.
- **WATCH OUT FOR (1):** Contaminations in small population with high genetic value: **crossing blocks** and **seed production**.
- **WATCH OUT FOR (2):** Gene flow when managing genetically modified material.



More on gene flow: Implication on breeding

Gene flow by pollen: implications for plant conservation genetics

Norman C. Ellstrand **OIKOS 63: 77–86. Copenhagen 1992**

Table 1. Traits of rare species at risk for gene flow-mediated hazards.

Breeding system	Obligate outcrossing (Dioecy, self-incompatibility, etc.) OR If self-fertile, high outcrossing rate
For outbreeding depression:	
Population structure Differentiation	Multiple populations with at least two within mating distance (generally < 10 km) Strong between populations
For problems of interspecific hybridization:	
Proximity of congener Compatibility of congener Magnitude of gene flow source	Sympatric or parapatric (generally < 10 km) Compatible enough to readily affect fertilization (seed set is not required; see text) Congener population numerically greater than vulnerable population (generally at least twice as many individuals) OR Congener population reproductively more vigorous than vulnerable population (in terms of pollen production or pollen export)

Word of caution for Crossing blocks

Gene flow by pollen in small populations

One factor often cited as a source of gene flow variation is population size. Generally, the rate of gene flow by pollen is expected to increase as population size decreases. Two reasons are offered for this expectation: 1) As population size increases, the number of targets for a fixed amount of "pollen rain" increases and, consequently, the average rate of fertilization by that pollen should decrease (Handel 1983). 2) For zoophilous species, optimally foraging pollinators will spend more time within large populations than small populations, affecting proportionately more intrapopulation matings (Levin 1981).

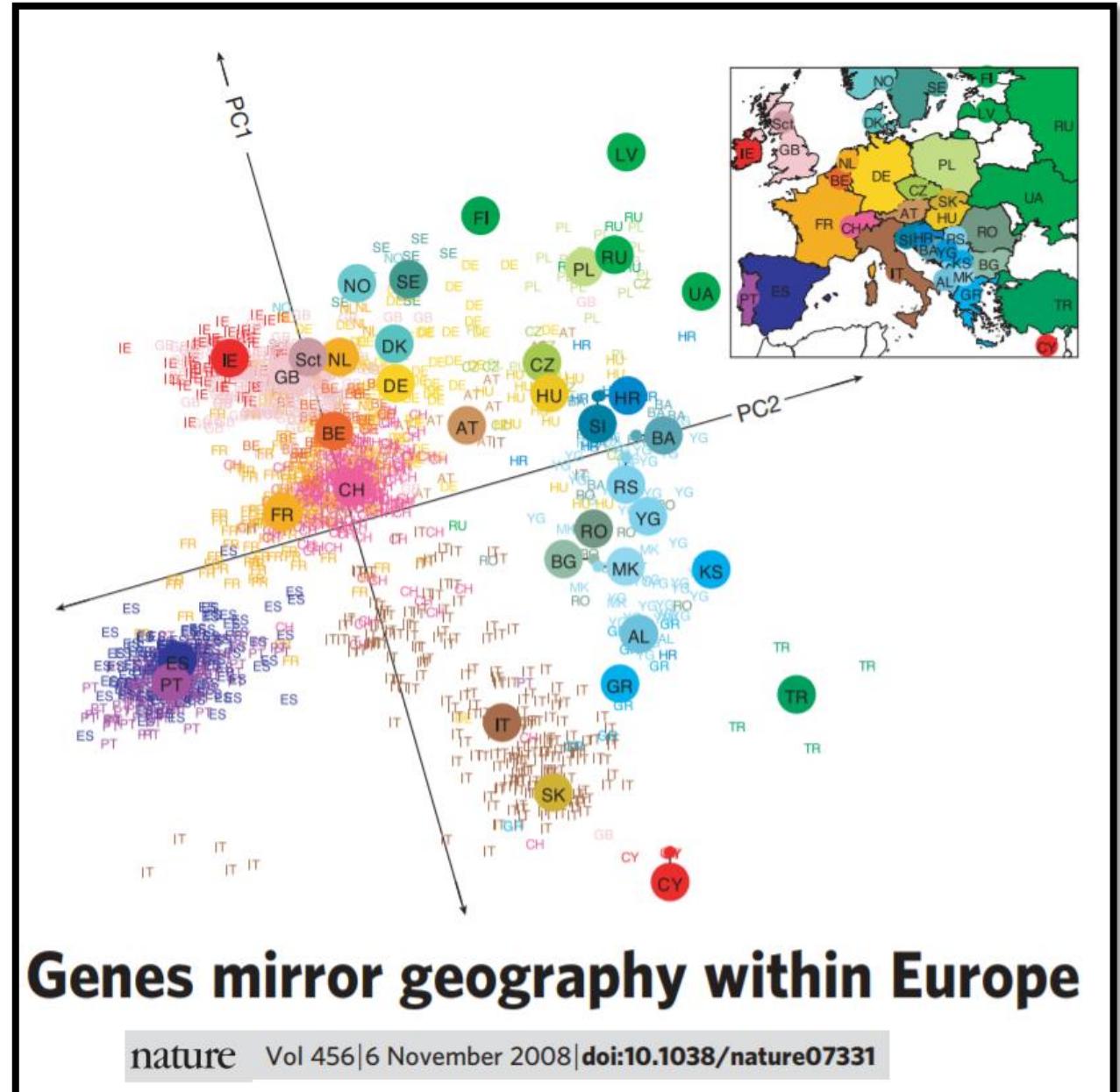
OIKOS 63: 77–86. Copenhagen 1992

IS GENE FLOW THE MOST IMPORTANT EVOLUTIONARY FORCE IN PLANTS?

American Journal of Botany 101(5): 737–753. 2014.

Applied consequences—The range of human concerns relating to the genetic composition of plant populations is staggering. Natural gene flow and human-mediated gene flow are a component of many of these concerns. As discussed, breeders and seed-producers are concerned with maintaining the genetic purity of seed at a certain level and consequently preventing genetic contamination by accidental pollen- and seed-based gene flow. With regard to plant conservation, significant gene flow changes can alter the genetic composition of the populations of endangered species for better or worse (Ellstrand and Elam, 1993).

Illustration of signature of demography and gene flow using PCs



2.3 Selection

General types of selection

- **Directional selection:** One direction increases the fitness: either higher and lower values, but not both.
- **Stabilizing selection:** An equilibrium exist and the population is better off under an intermediate phenotypic value (or because extremes are lethal).
- **Disruptive selection:** Either extreme is a better fit than an intermediate form.
- **FOR BREEDING:** yield undergoes *positive directional selection* (more is better), while composition and maturity often undergo *stabilizing selection* (keep germplasm within target maturity).
- **WATCH OUT FOR:** Most traits of interest are genetically correlated, the *best phenotype* is provided by combination of traits: breeders must manage genetic tradeoffs (**often through selection indexes**).

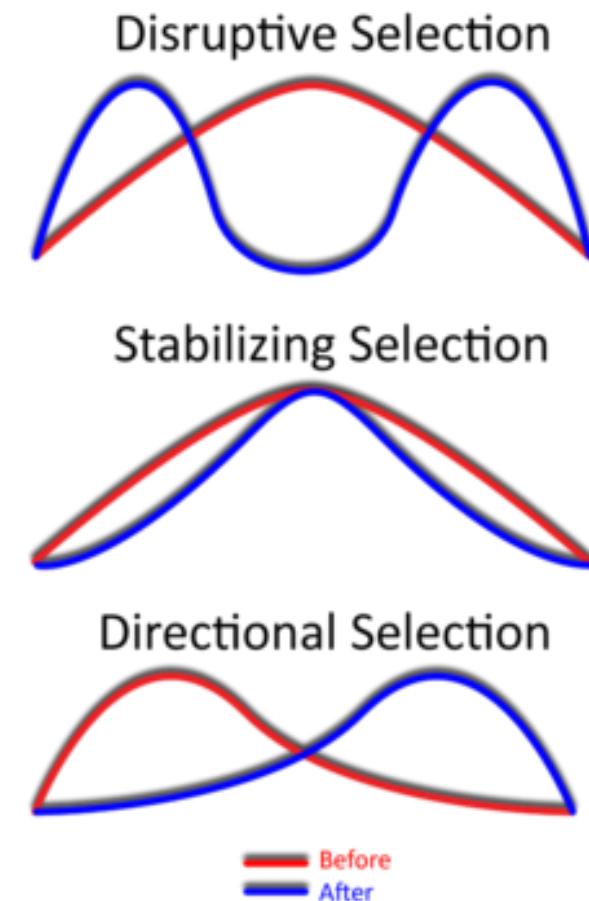
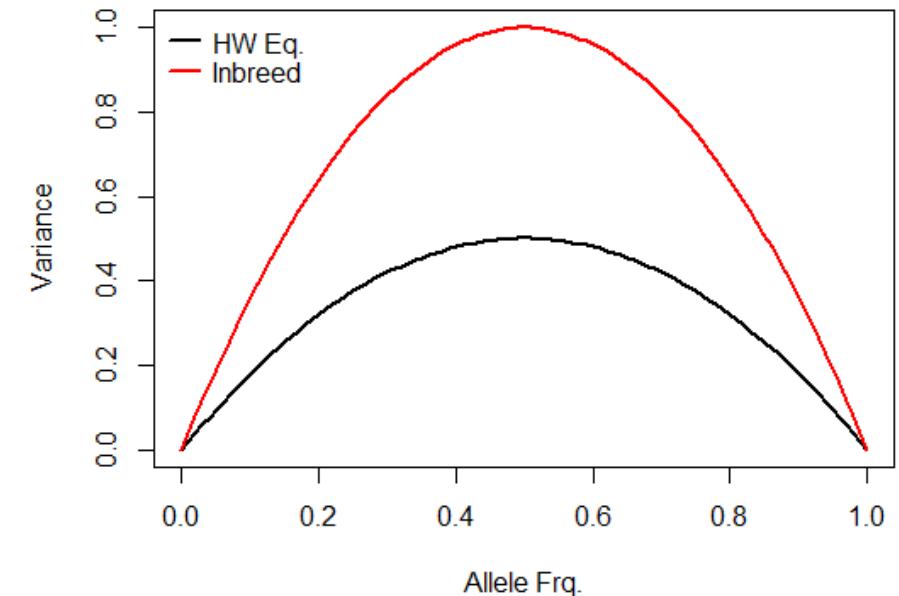


Image source: en.wikipedia.org/wiki/Directional_selection

Selection and allele frequency

- **RISE AND FATE OF ALLELES:**
 - Maximum variability point occurs at $p=0.5$
 - Overtime, high fitness alleles trend to go to **fixation** whereas low fitness allele trend to go to **extinction**
 - Heritability is proportional to the frequency of alleles that affect the trait (eyes color has $h^2=0$ in a population with no variation for the trait)
- **CONSEQUENCE:** Phenotype of the population shifts towards the most fit balance.
- **FOR BREEDING:** Allele come from introducing new lines (migration rather than mutation). Breeders' job is to sort out which allele are valuable for the program.
- **WATCH OUT FOR:** Pushing selection *too hard* to avoid compromising long term genetic gains



Selective sweep and diversity

- **BOTTLENECK:** Kohn et al. report natural selection causing hitchhiking over a substantial genomic region around fitness related genes.
- **CONSEQUENCE:** Loss of genetic diversity in the genomic regions surrounding these major fitness genes.
- **FOR BREEDING:** “QTL drag” is caused by MAS or selection on traits controlled by major QTLs.
- **WATCH OUT FOR:** Pushing selection *too hard* to avoid compromising long term genetic gains.

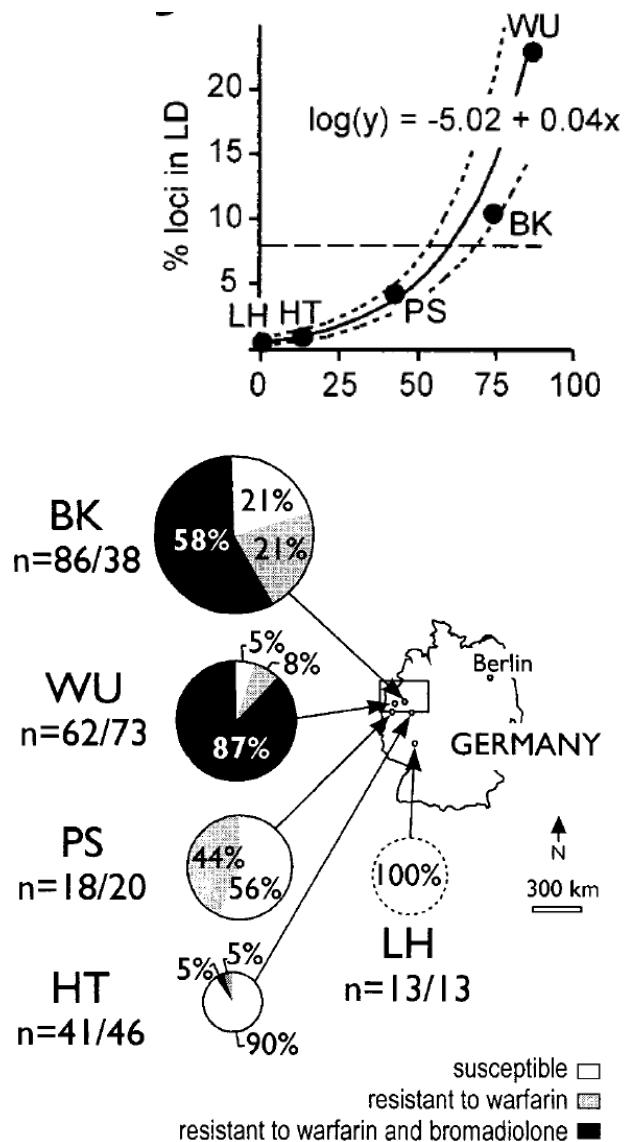
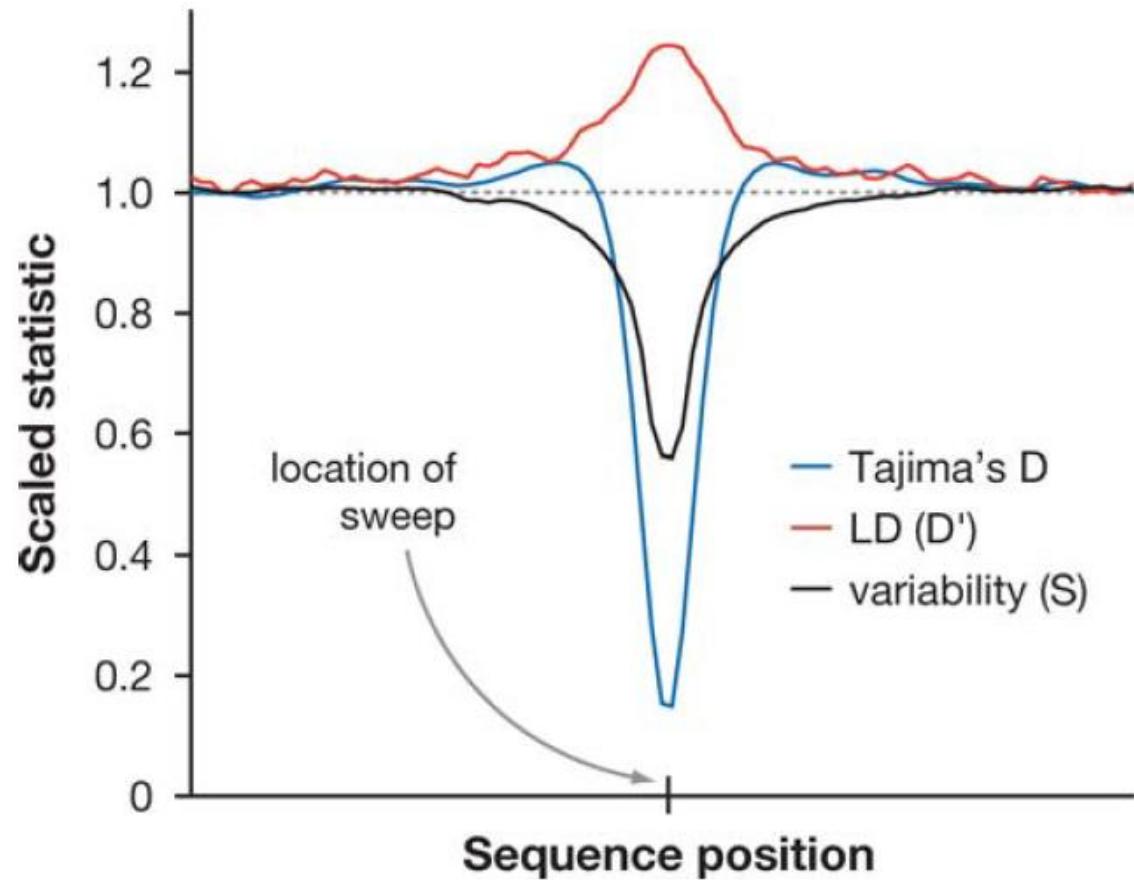


Fig. 1. Origin of rats, their resistance level, and relative population size. Collection sites and the nearest town (in parentheses) are as follows: BK (Olfen); WU (Stadtlohn); PS (Dorsten); HT (Drensteinfurt); all are located within the previously described resistance area that includes the city of Münster (box; ref. 33). Sampling site LH (Ludwigshafen) was located outside the resistance area.

Illustration of bottleneck or selective sweep



Molecular Signatures of
Natural Selection

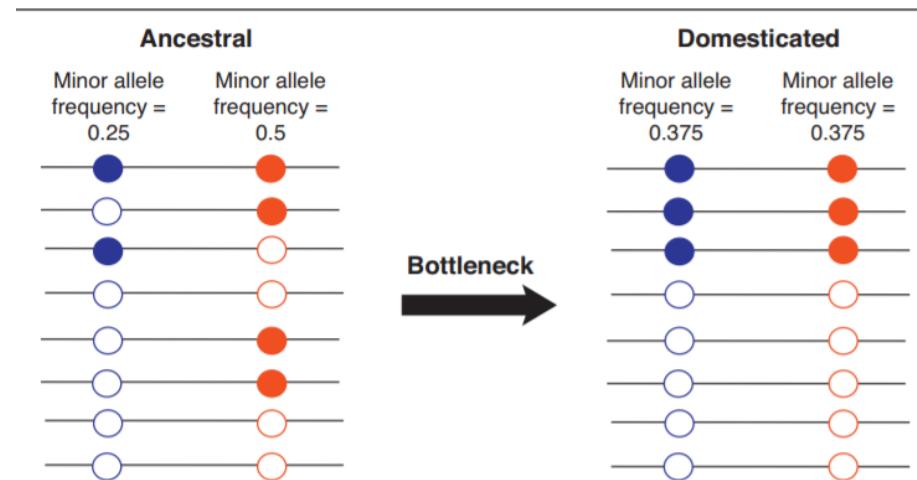
Rasmus Nielsen

Center for Bioinformatics and Department of Evolutionary Biology, University of Copenhagen, 2100 Copenhagen Ø, Denmark; email: rasmus@binf.ku.dk

2.4 Genetic Drift

Sampling effect

- **RANDOMNESS:** The allele frequency of most polymorphism with little or no effect on fitness fluctuates at random, often leading alleles to fixation or extinction.
- **FOUNDER EFFECT:** The magnitude of genetic drift is aggravated under sampling bottlenecks – when a new population is derived from a non-representative subset of the original population.
- **CONSEQUENCE:** Diversity losses and random changes in the genomic composition. Effects of drift in the genome can be easily confounded with selection.
- **FOR BREEDING:** Most breeders are not concerned about drift. Much of the selection performed in early generations causes drift. Having the same family being selected in separate experiments mitigates drift.
- **WATCH OUT FOR:** Unlike practical breeding *per se*, geneticist that attempt to infer signatures of selection should pay close attention to drift (**keyword: replication**).



$$d_{ns} \approx d_s$$

Table 1 Statistics of SNPs in whole genome and genic regions of wild and cultivated soybean accessions

Whole genome	Number of SNPs	Non-synonymous SNPs	Synonymous SNPs	Nonsyn/Syn
Wild soybean	5,924,662	106,716	78,701	1.36
Cultivated soybean	4,127,942	77,291	55,883	1.38

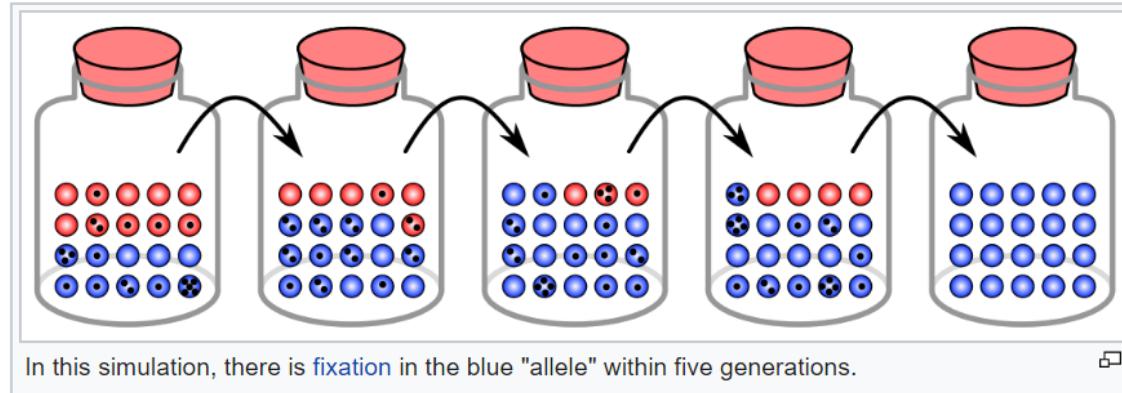
VOLUME 42 | NUMBER 12 | DECEMBER 2010 NATURE GENETICS

Competition between random genetic drift and natural selection play a central role in evolution: Whereas nonbeneficial mutations often prevail in small populations by chance, mutations that sweep through large populations typically confer a selective advantage.

Hallatschek et al. 19926–19930 | PNAS | December 11, 2007 | vol. 104 | no. 50

Genetic drift

From Wikipedia, the free encyclopedia

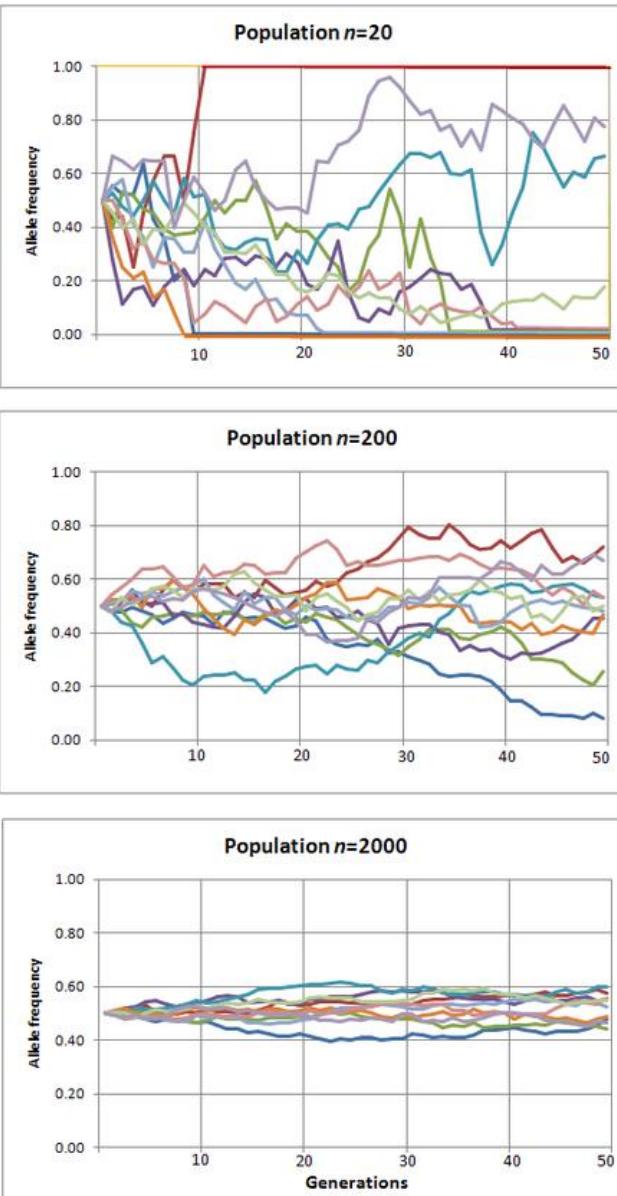
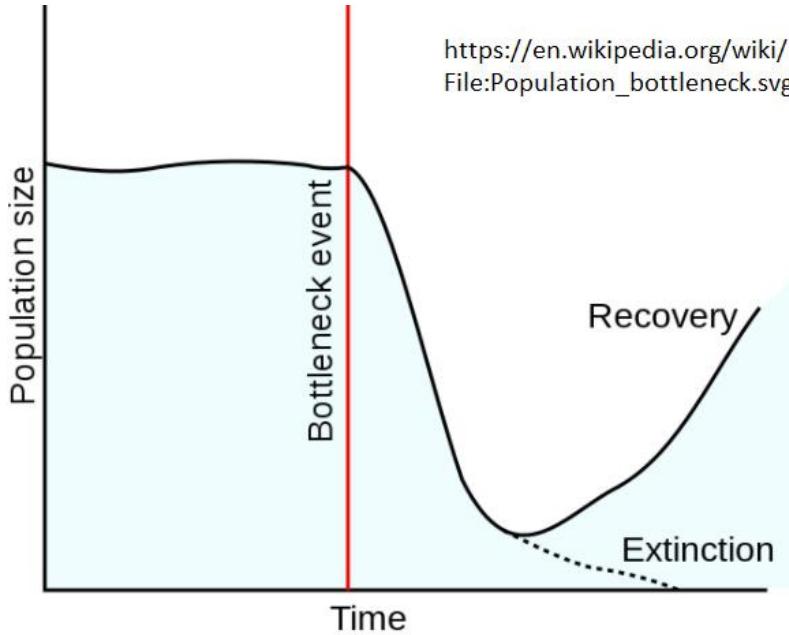


Trends in Genetics March 2011, Vol. 27, No. 3

Mutation–drift equilibrium and effective population size

Genetic variation is caused by random mutations that occur at a characteristic rate (about 10^{-8} /bp/generation for SNPs [78]). Most mutations arising at functional sites are strongly deleterious and are quickly eliminated by purifying selection, whereas strongly advantageous mutations are rare and are quickly fixed in the population. As a consequence, the majority of genetic variation segregating within populations is either weakly selected or neutral. Furthermore, most mutations are lost to drift when they are young, producing an allele frequency spectrum shifted toward lower frequencies.

All populations are finite and thus experience genetic drift.



https://en.wikipedia.org/wiki/File:Random_genetic_drift_chart.png

Effects on Genetic Diversity

Genetic drift changes the distribution of genetic variation in two ways: (i) the decrease of variation within populations (loss of heterozygosity and eventual fixation of alleles), and (ii) the increase of differentiation among populations. Every finite population experiences genetic drift, but the effects become more pronounced as population size decreases (31, 38). Wright (120) predicted that drift will substantially alter the organization of genetic variation of populations when $1/4N_e$ is much greater than the mutation rate (μ) and the selection coefficient (s) where N_e is the effective population size.

Effective population size is the number of individuals in an ideal population that would have the same genetic response to random processes as a real population of size N (23, 120). This concept is important because most population genetic theory deals with ideal populations. To best apply the predictions of population genetics, estimates of effective population sizes in nature are necessary. The effective population size is often depressed below the census size by factors such as deviations from one-to-one sex ratios, overlapping generations, variation in progeny production, and fluctuations in population size (37, 63, 100). While effective population sizes in nature are often difficult to measure, the ratio N_e/N is often expected to fall between 0.25 and 1.0 (Nunney & Campbell, in preparation).

Populations with continually small effective population sizes will be especially susceptible to the loss and reorganization of variation by genetic drift. However, any population that undergoes occasional fluctuations to small population size may also suffer from loss of variation by chance.

Ellstrand, N. C., & Elam, D. R. (1993). Population genetic consequences of small population size: implications for plant conservation. *Annual review of Ecology and Systematics*, 24(1), 217-242.

2.5 Hill-Robertson effect

Hill-Robertson effect

- **CONCEPT:** Genetic drift and mutation tend to slow down the process of evolution by selection. Linkage/LD between sites under selection harm the overall genetic progress. Introduced by Hill and Robertson (1966, *Genetical Research* 8:269–294).
- **CONSEQUENCE:** Drift and mutation (and epistasis) reduce the effectiveness of selection in finite population. It results in more drift and lower N_e than expected. Recombination becomes key to overcome HR effect.
- **FOR BREEDING:** Producing crosses with large offspring reduce chances of getting stuck with suboptimal haplotypes.
Transgressive segregation often means to overcome HR effect.
- **WATCH OUT FOR:** The heritability and number of genes involved: More complex/polygenic the trait is, the harder is to obtain true transgressive individuals. Larger offspring mitigate HR by mutation, drift, hitchhiking and undesirable epistasis.

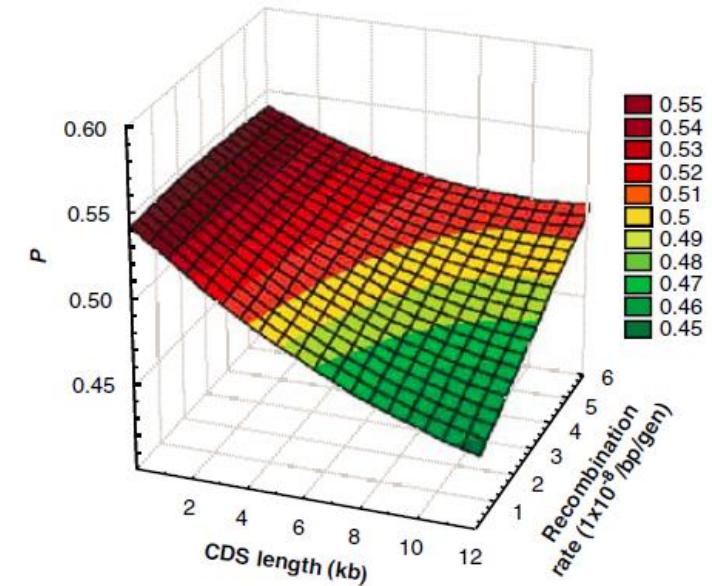


Figure 8 Observed relationship between the length of coding sequences (CDS), recombination rate (rate of crossing-over) and the frequency of preferred codons (P)

Cameron et al. *Heredity* (2008) 100, 19-31

More genes controlling the trait (L) and smaller the population (ρ), less effective the selection (P) will be

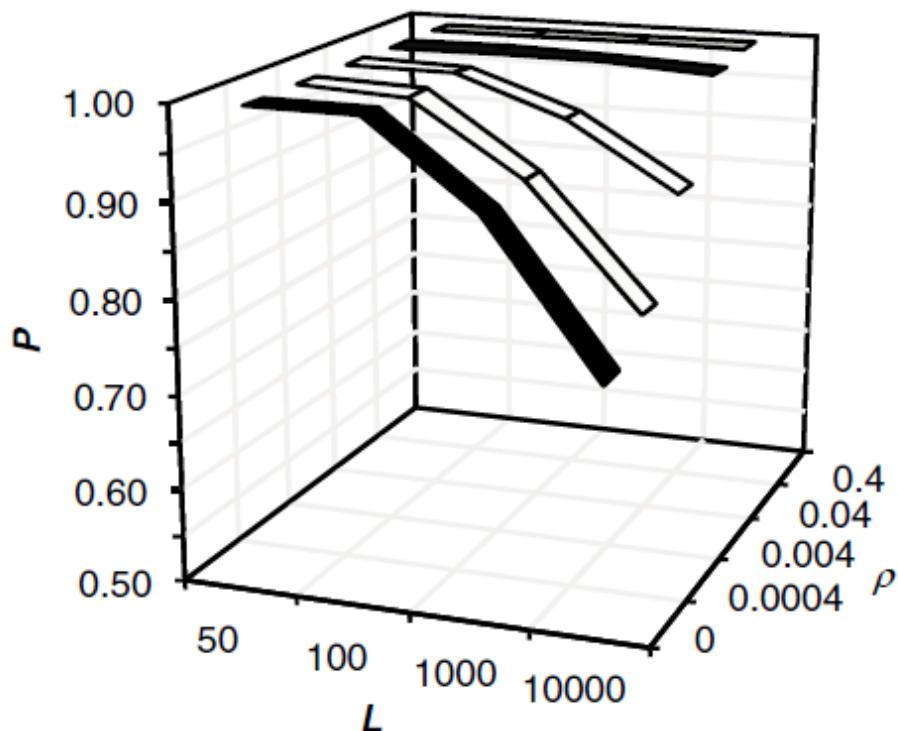


Figure 4 Simulation results showing the influence of variable number of sites under selection (L) and variable-scaled recombination rates ($\rho = N_e c$) on the frequency of sites with the preferred variant (P).

Cameron et al. Heredity (2008) 100, 19-31

3. Population management

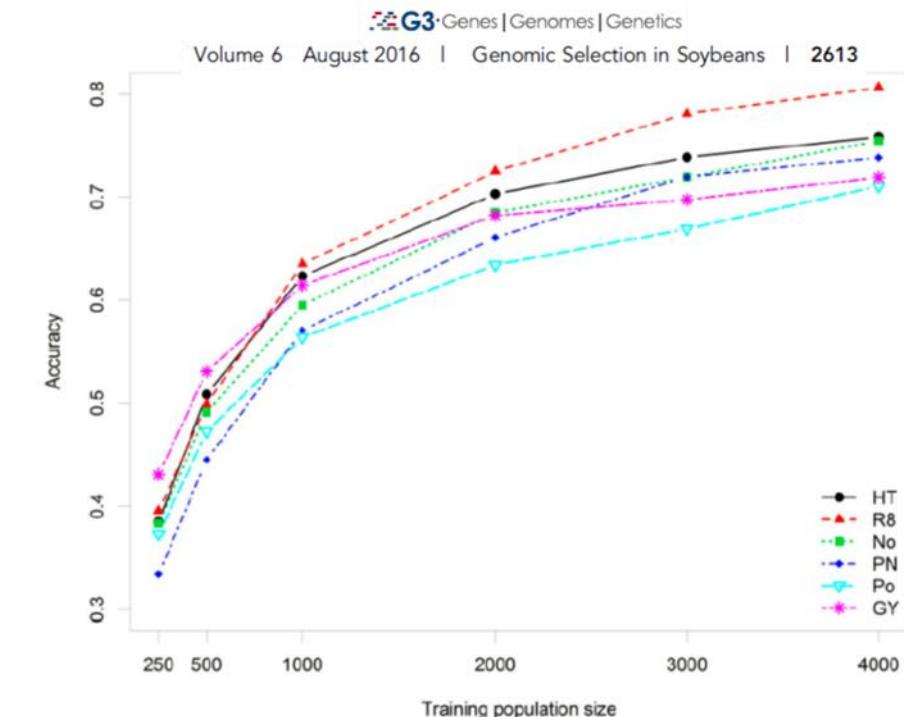
3.1 Effective population size

Effective population size (N_e)

- **DEFINITION:** There exist several definitions for N_e , depending on the context or research area. N_e can be thought as a canonical representation of a given population that comprises all the genetic variability in a minimal number of individuals under random mating.
- **Some factors that influence N_e :**
 - Both mutation and recombination increase N_e
 - Incorporating unrelated individuals increases N_e
 - Number of lines used as parents for the next generation
- **FOR BREEDING:** Recurrent estimation of N_e is a practical way to monitor the level of inbreeding and genetic variability of breeders' germplasm.
- **WATCH OUT FOR:** Controlling the number of crosses and individuals per cross, as well as the relatedness among the parental lines.

Genetics, Vol. 189, 633–644 October 2011

$$E(r^2) = \text{Var}(r) \approx \frac{1}{3N_e},$$



3.2 Effective Population Size

When considering the effective size of a population, we always have to remember that effective size is *not* a property of the population. It is a number that helps us describe some of the effects of random genetic drift in the population, in particular the convergence to genetic homogeneity, in that the factor λ describing this convergence is specified as

$$\lambda = 1 - \frac{1}{2N_e}.$$

The effective population size is a reference to the gamete pool model that will give us an intuitive feel for various properties of the population.

Christiansen, F. B. (2014). Theories of population variation in genes and genomes. Princeton University Press.

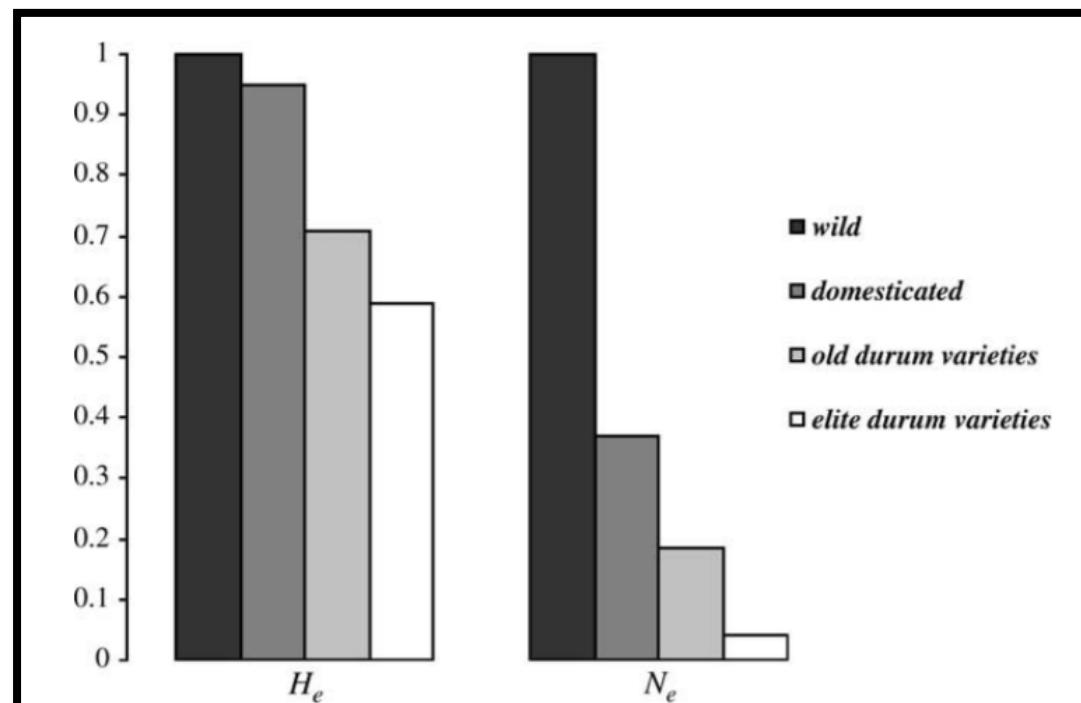


FIGURE 3.—Reductions of heterozygosity (H_e) and effective population size (N_e) from the wild form, *T. t. dicoccoides*, to the elite varieties of durum wheat. Values are shown relative to the value in the wild form that is retained in the subsequent population.

A.-C. Thuillet *et al.*
Genetics 169: 1589–1599 (March 2005)

3.2 Short and long term genetic gains

Short and long term gains

- **CHALLENGE:** Dealing with the side effects of directional selection in a breeding population.
- **FOR BREEDING:** Mostly about the tradeoff imposed of selection intensity.
- **WATCH OUT FOR:**
 - **Sort-term losses:** Control number of crosses and offspring size; target crosses with better odds of transgressive segregation.
 - **Long-term losses:** Controlled selection intensity; avoid inbreeding with within-family selection; maintain or expand effective population size (incorporating new germplasm, new QTLs).

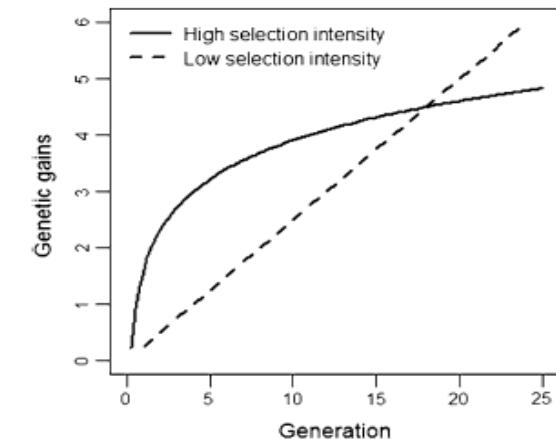
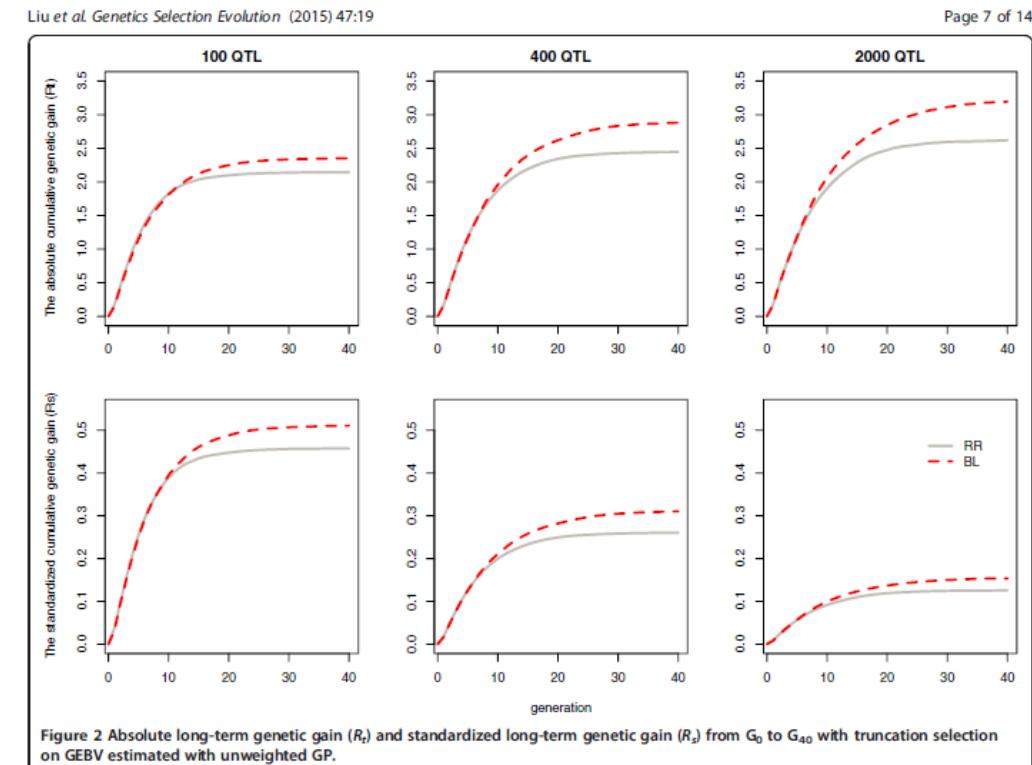
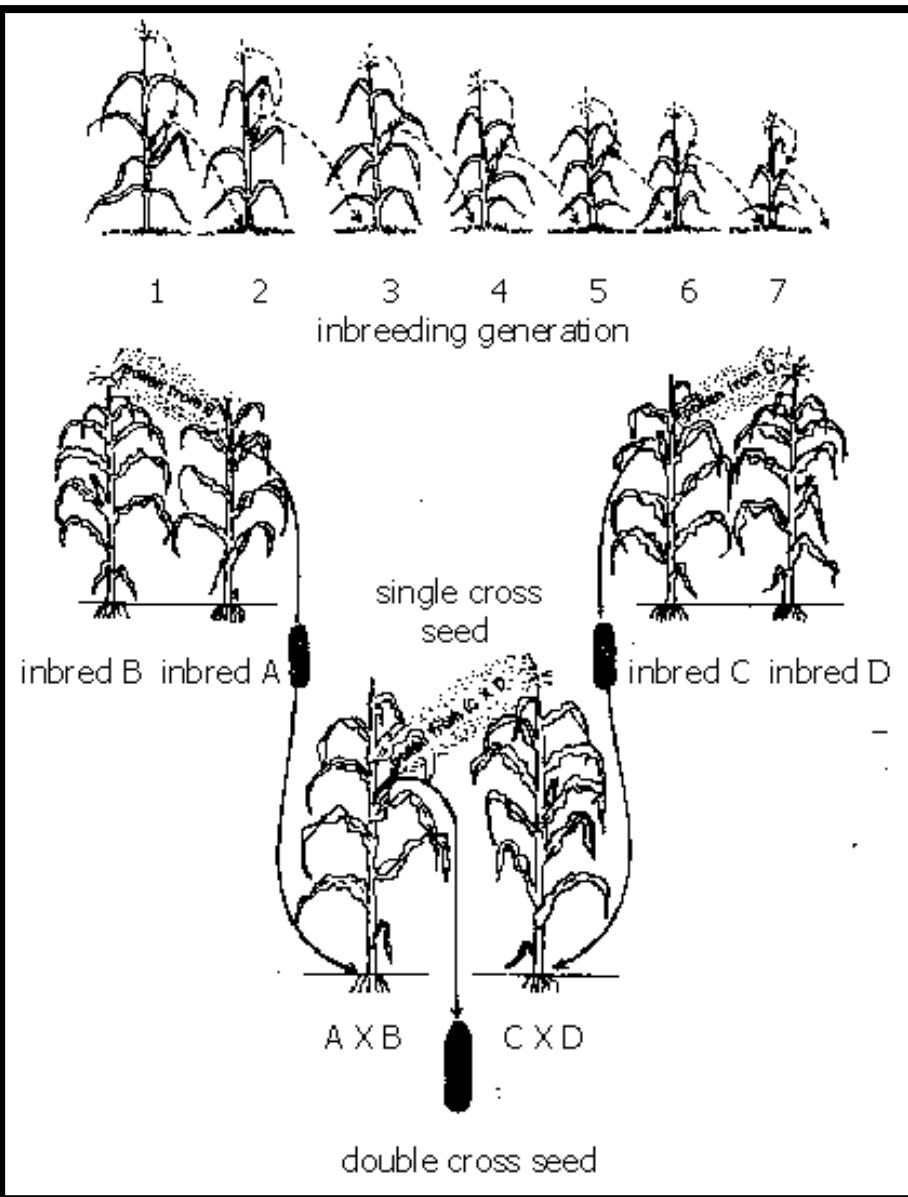


Fig. 4 Illustration of the consequences of high and low selection intensity on genetic gains over generations of selection for a given quantitative trait



3.3 Inbreeding and heterosis

A brief preview



Inbreeding

- Inbreeding is the result of selfing (or double-haploids)
 - In maize, they inbred crosses within the heterotic group
 - Inbred lines are just used as parents

Heterosis: Hybrid from single cross

- Heterosis or hybrid vigor is the result of outbreeding
 - The **maximum level of heterosis** is observed by crossing inbred lines from **two different heterotic groups**

Decay of heterosis: Double-cross hybrid

- Cross between two hybrids: Parents will be related at some extent
- **Does not present the same level of heterosis** as single-cross hybrids

Inbreeding and heterosis

- **Inbreeding depression**: In many plant species (cassava, corn, tomatoes) and all animal species, driving a population to a restrict genetic background has serious implications of vigor and fitness – due to the *accumulation of deleterious alleles in homozygous state*.
- **Heterosis (hybrid vigor)**: The inverse scenario of inbreeding depression. Improvement in fitness (or yield) attributed to dominance due to outbreeding: *Hybrids F1 over perform the parental lines*.
- **FOR BREEDING**:
 - Heterosis is the foundation of commercial breeding for some crops (maize, canola, sorghum, and also rice)
 - Limitation – Not all crop have a reproductive biology that allows for feasible hybrid production
 - Breeding programs develop the parental liner by their general (GCA) and specific combining ability (SCA)
- **WATCH OUT FOR**:
 - Watch out for how one define the parental genetic pools (referred to as **heterotic groups**)
 - Heterosis may not be big thing for all crops
 - **Heterosis is not a heritable property**

Maize is best known example

Lu et al (2009). Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. TAG, 120(1), 93-115.

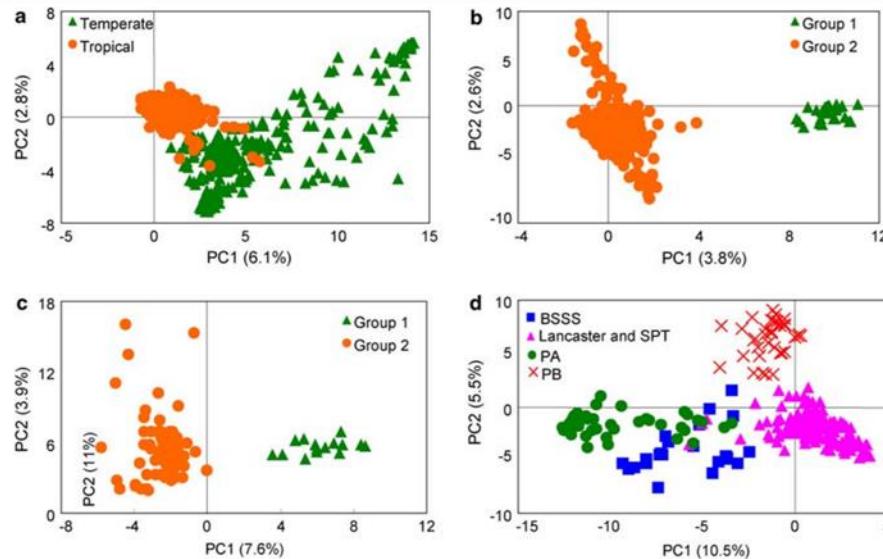


Fig. 4 Principal component analysis for the entire set of maize lines (temperate and tropical/subtropical lines) (a), CIMMYT inbred lines (b), Brazilian inbred lines (c), and Chinese inbred lines (d)

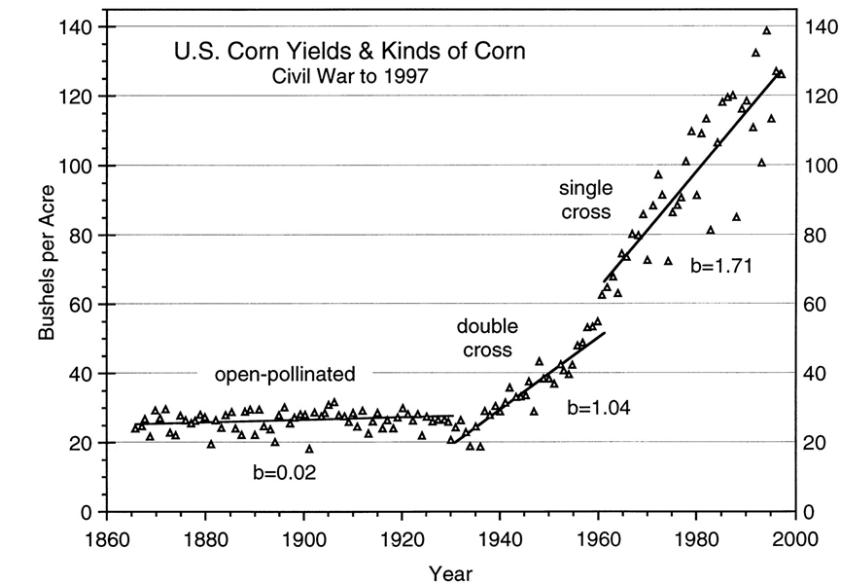
Heterotic groups

Heterosis

Springer, N. M., & Stupar, R. M. (2007). Allelic variation and heterosis in maize: how do two halves make more than a whole?. Genome research, 17(3), 264-275.



Crow, J. F. (1998). 90 years ago: the beginning of hybrid maize. Genetics, 148(3), 923-928.



Impact in agriculture

Heterosis is not a heritable property

Theor Appl Genet (2015) 128:1647–1667

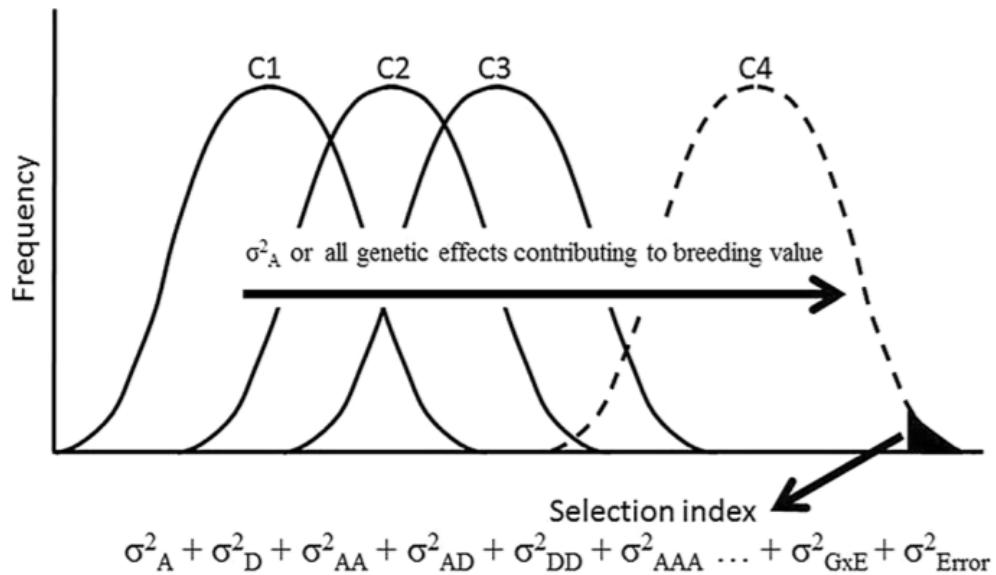
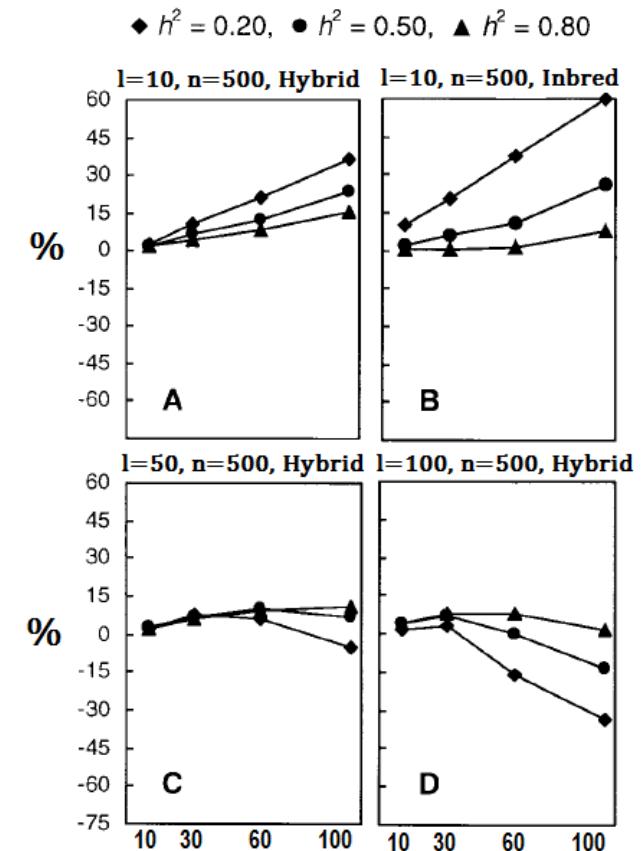


Fig. 1 Illustration of a truncated phenotypic recurrent selection scheme such as the one currently used in cassava breeding. C1, C2, C3 and C4 are the successive cycles of selection. Shifts in allelic frequencies gradually occur in the different versions of the population represented by the successive cycles. This genetic progress is achieved mostly exploiting additive genetic effects. The selection of a successful clone, however, is affected by all genetic effects as well as experimental errors and the ever confounding effect of genotype-by-environment interaction



Bernardo, R. (2001). What if we knew all the genes for a quantitative trait in hybrid crops? Crop Science, 41(1), 1-4.

$l = \# \text{ of alleles}$

$n = \# \text{ of individuals}$

Instead of breeding the hybrids, programs that take advantage of heterosis breed parents with good combining ability – capable of generating better hybrids when crossed with another group of parents

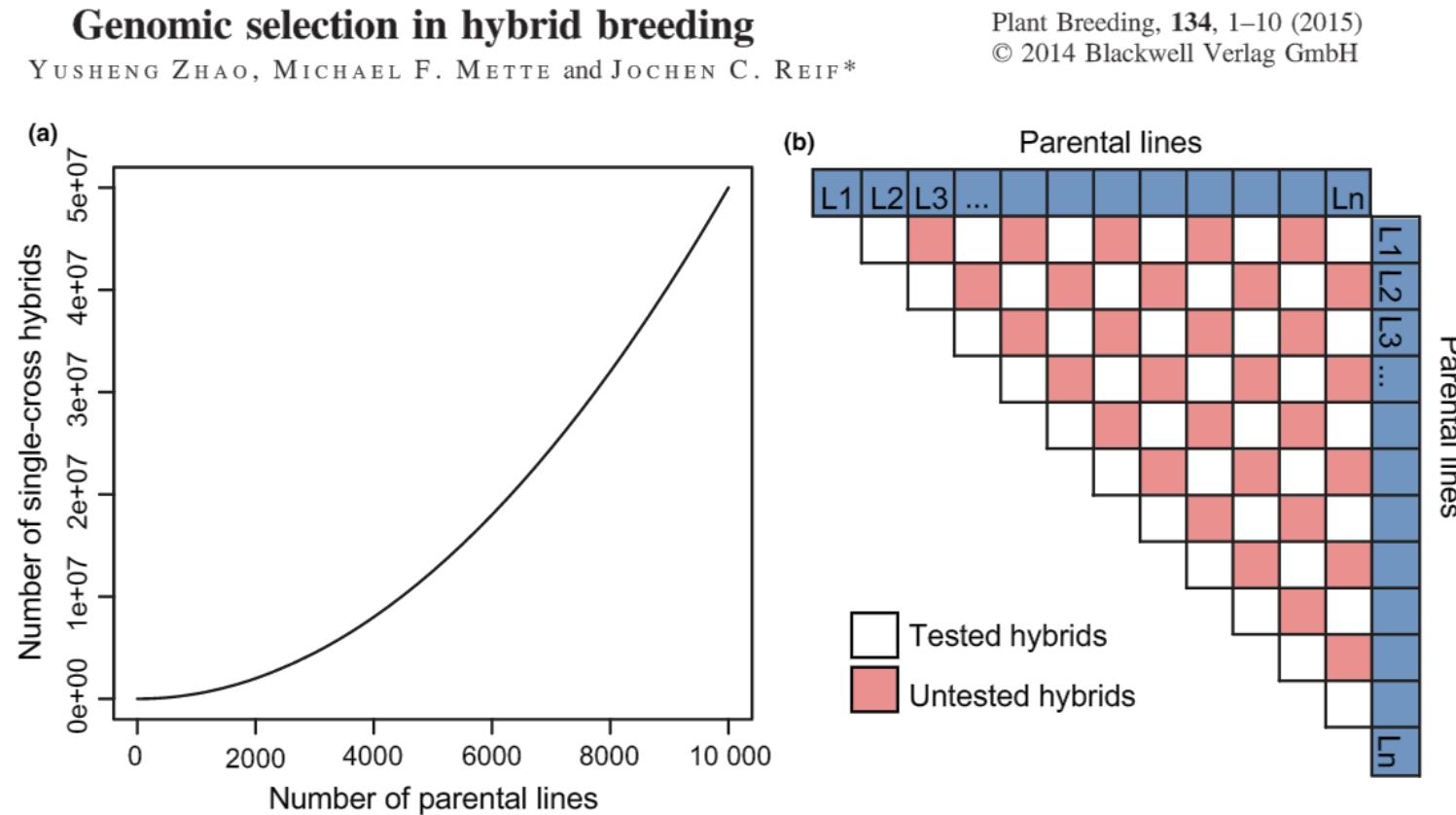
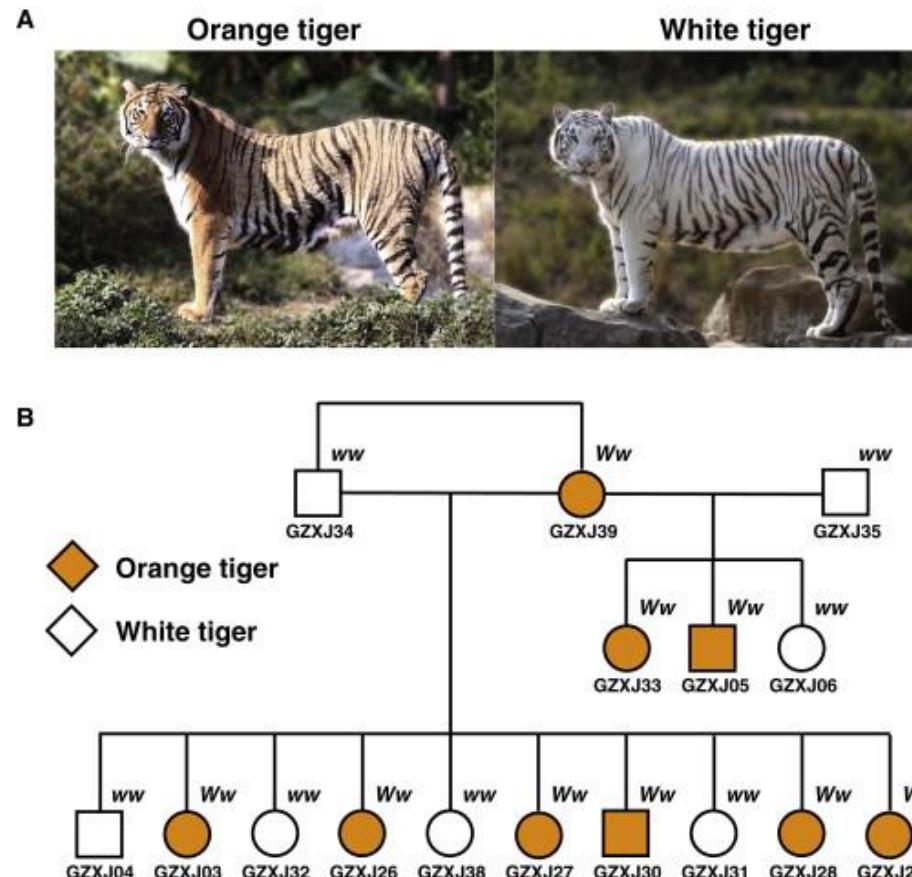


Fig. 2: Application of genomic selection in hybrid breeding. (a) Number of potential single-cross hybrids in dependency on the number of parental lines. (b) Possible genomic selection scheme to identify heterotic groups.

Inbreeding depression – big deal in animals

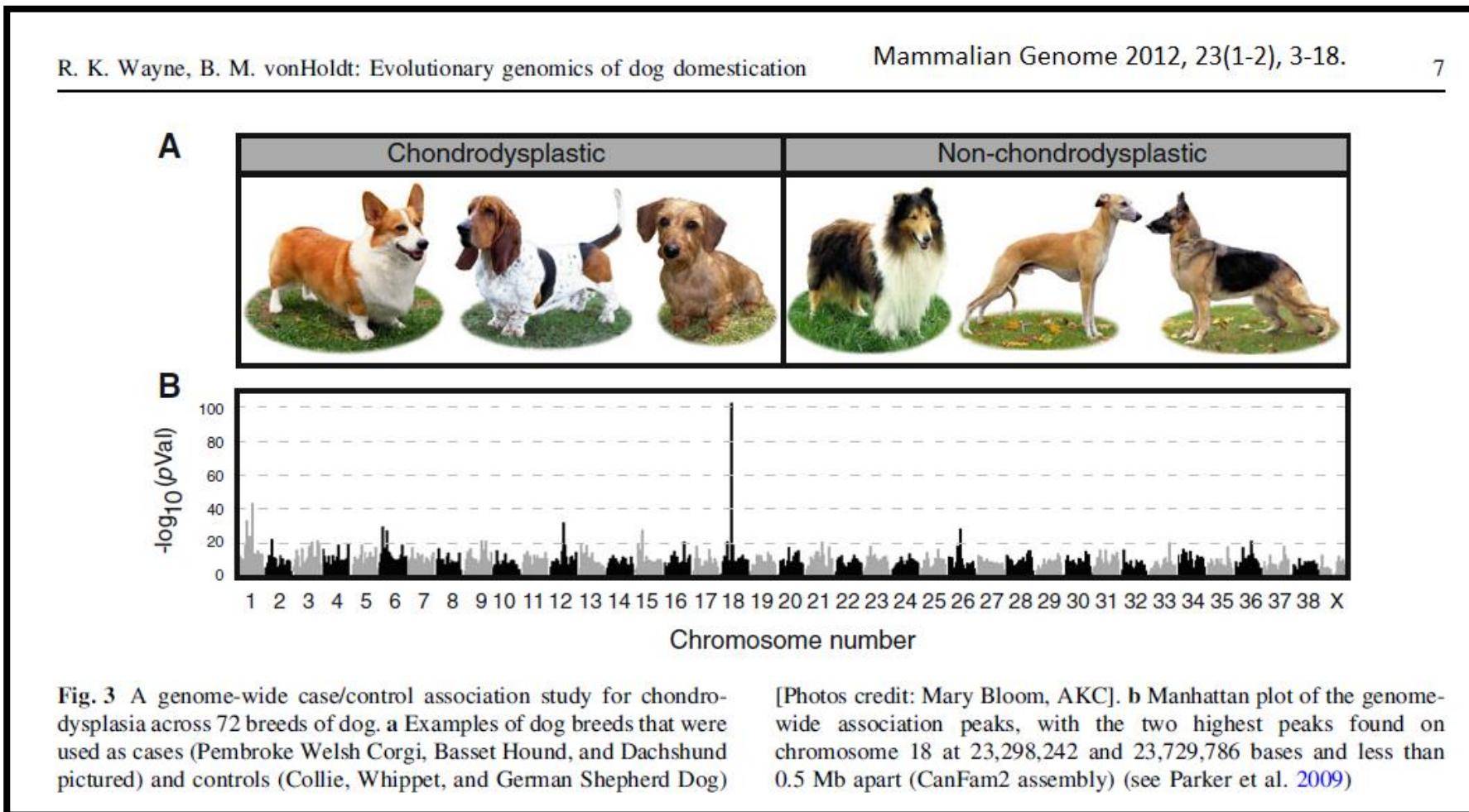


Xu et al (2013). The genetic basis of white tigers. Current biology, 23(11), 1031-1035.

A single amino acid change in transporter SLC45A2 causes the white tiger phenotype

- **Fixation index** – the ratio between the variance of a subpopulation and the variance of the total population

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{p(1-p)}$$
- FST is 1 when an allele is fixed in different directions between two populations



Inbreeding and structure

It has probably occurred to the reader that the coefficient of inbreeding may mean very different things in different cases. (1) There may be division of the population into completely isolated small strains, within each of which there is random mating. The inbreeding coefficient of individuals relative to the total is here due merely to the relationship of all members of the same strain and disappears at once with random mating among strains. (2) There may be frequent mating of close relatives but no permanent separation of strains. Here again random mating at once reduces the inbreeding coefficient to zero. (3) The sires used may be rather limited in number and derived from even more limited numbers of grandsires and great-grandssires. In this case there may be little apparent close inbreeding at any time, contrary to (2), and no division into strains, contrary to (1), but the value of F , relative to the foundation stock, keeps rising and cannot be much reduced by random mating. There are other possibilities. Clearly we need something more than a single value of F to give an adequate description of structure.

Wright, S. (1949). The genetical structure of populations. *Annals of Human Genetics*, 15(1), 323-354.

(LITERATURE MASTER PIECE !!)

3.4 Pedigree and relatedness

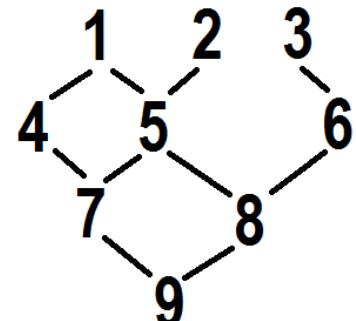
Pedigree and relatedness

- **CONCEPT**: The “*coefficient of relationship*” was introduced by Samwell Wright (1922) as the expected consanguinity between individuals based on their shared ancestry.
- **FOR BREEDING**:
 - **CROSSING**: Pedigree information is used for crossing decisions to avoid inbreeding or driving the germplasm towards a specific subset of families.
 - **ADVANCEMENTS**: Pedigree is used in mixed models to define the **additive** relationship among individuals. BLUPs generated with pedigree information are referred to as “**breeding values**”, which corresponds to the values that are heritable. Breeding values are believed to be better proxies for the selection of complex traits by taking into account the **information from relatives**, hence leading to a better distinction between the genetic effect of interest and environmental effects (or non-additive genetics).
- **WATCH OUT FOR**: Complexity of the pedigree (number of generations you want to trace back the material under evaluation).

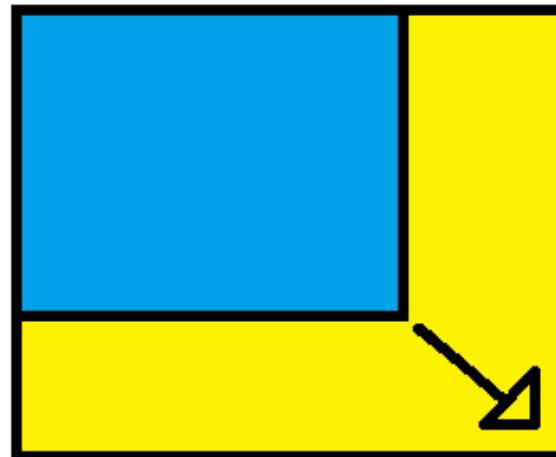
Pedigree-based relationship matrix

The pedigree **must** be written in time order, starting from individuals of unknown parentage

ID	Dam	Sire
1	0	0
2	0	0
3	0	0
4	0	1
5	1	2
6	3	0
7	4	5
8	5	6
9	7	8



The pedigree relationship matrix expands one individual at time, which adds one column and row



Each cell (a_{ij}) in of A represents the relationship between a pair of individuals, estimated as the average between the maternal (d_i) and paternal (s_j) relatedness between individuals

$$a_{ij} = \frac{d_i + s_j}{2}$$

Pedigree-based relationship matrix

Function

```
# Pedigree matrix
Amat=function(ped){
n = nrow(ped)
A=diag(0,n)
for(i in 1:n){
for(j in 1:n){
if(i>j){A[i,j]=A[j,i]}else{
d = ped[j,2]
s = ped[j,3]
if(d==0){A1d=0}else{A1d=A[1,d]}
if(s==0){A1s=0}else{A1s=A[1,s]}
if(d==0|s==0){Asd=0}else{Asd=A[d,s]}
if(i==j){A1j=1+0.5*Asd}
if(i!=j){A1j=0.5*(A1d+A1s)}
A[i,j]=A1j}}}
return(A)}
```

Input

	id	dam	sire
[1,]	1	0	0
[2,]	2	1	0
[3,]	3	1	0
[4,]	4	1	0
[5,]	5	1	0
[6,]	6	2	3
[7,]	7	0	3
[8,]	8	4	5
[9,]	9	6	7
[10,]	10	8	3
[11,]	11	9	10

Output

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]
[1,]	1.000	0.500	0.500	0.500	0.500	0.500	0.250	0.500	0.375	0.500	0.438
[2,]	0.500	1.000	0.250	0.250	0.250	0.625	0.125	0.250	0.375	0.250	0.312
[3,]	0.500	0.250	1.000	0.250	0.250	0.625	0.500	0.250	0.562	0.625	0.594
[4,]	0.500	0.250	0.250	1.000	0.250	0.250	0.125	0.625	0.188	0.438	0.312
[5,]	0.500	0.250	0.250	0.250	1.000	0.250	0.125	0.625	0.188	0.438	0.312
[6,]	0.500	0.625	0.625	0.250	0.250	1.125	0.312	0.250	0.719	0.438	0.578
[7,]	0.250	0.125	0.500	0.125	0.125	0.312	1.000	0.125	0.656	0.312	0.484
[8,]	0.500	0.250	0.250	0.625	0.625	0.250	0.125	1.125	0.188	0.688	0.438
[9,]	0.375	0.375	0.562	0.188	0.188	0.719	0.656	0.188	1.156	0.375	0.766
[10,]	0.500	0.250	0.625	0.438	0.438	0.438	0.312	0.688	0.375	1.125	0.750
[11,]	0.438	0.312	0.594	0.312	0.312	0.578	0.484	0.438	0.766	0.750	1.188

Genomic relatedness: genetic distance

Individual inbreeding coefficient (F) and pairwise relatedness (r) are fundamental parameters in population genetics and have important applications in diverse fields such as human medicine, forensics, plant and animal breeding, conservation and evolutionary biology. Traditionally, both parameters are calculated from pedigrees, but are now increasingly estimated from genetic marker data.

Wang, J. (2016). Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient?. *Theoretical population biology*, 107, 4-13.

- **DEFINITION:** Genetic distance is a measure of the genetic divergence between species or between populations within a species
- **USE:** Characterize the structure of a population and to verify relatedness among individuals.
 - **NOTICE:** Genetic distances may vary according to the dataset (ie. number and quality of the SNPs).
- **COMMON METRICS GENETIC DISTANCES:**
 - Nei Distance (Linear): $d_{ND} = -\ln \left(\frac{x'y}{\sqrt{x'x * y'y}} \right)$
 - Provesti Distance (Absolute) = $d_{PD} = \frac{\sum_j |x_j - y_j|}{2m}$
 - Modified Rogers' Distance (Euclidean): $d_{MRD} = \frac{\sum_j (x_j - y_j)^2}{2\sqrt{m}}$

$$\left. \begin{array}{l} \text{AA-AA or aa-aa} = 0 \\ \text{Aa-AA or Aa-aa} = 0.5 \\ \text{AA-aa} = 1 \end{array} \right\}$$

Genetic distance is quite intuitive

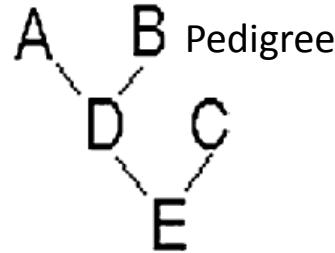
Provesti
distance

Single-SNP averages. We use S_{Bj} to denote the genotype of B at the j^{th} diallelic SNP, coded as 0, 1 or 2. Analogous with the definition of coancestry, a natural way to score the similarity of two individuals at each SNP is as the probability of a match between alleles drawn at random from each of them. In that case matching homozygotes (0,0) or (2,2) score 1; discordant homozygotes (0,2) score 0; while (0,1), (1,1) and (1,2) all score 0.5. Averaged over m SNPs, this gives an allele-sharing coefficient^{29,38} as follows, where X_B is a (row) vector with j^{th} entry $S_{Bj} - 1$.

Speed, D., & Balding, D. J. (2015). Relatedness in the post-genomic era: is it still useful?. Nature Reviews Genetics, 16(1), 33-44.

Breeding values from pedigree

Field map (line and its yield)		
A = 27	Missing	E = 21
E = 27	C = 20	B = 27
B = 21	A = 25	Missing



Example from Xavier et al. 2016 TAG, 129(10), 1933-1949

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{Z}'\mathbf{X} \\ \mathbf{X}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$(\lambda = \sigma_e^2 / \sigma_a^2)$

Mixed model equation (Henderson Model 3)

INPUT

$$\mathbf{y} = \begin{bmatrix} \text{Yield} \\ 25 \\ 27 \\ 27 \\ 21 \\ 20 \\ 21 \\ 27 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \text{Intercept} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} \\ \mathbf{A} & 1 & 0 & 0 & 0.5 & 0.25 \\ \mathbf{B} & 0 & 1 & 0 & 0.5 & 0.25 \\ \mathbf{C} & 0 & 0 & 1 & 0 & 0.5 \\ \mathbf{D} & 0.5 & 0.5 & 0 & 1 & 0.5 \\ \mathbf{E} & 0.25 & 0.25 & 0.5 & 0.5 & 1 \end{bmatrix}$$

OUTPUT

$$\mathbf{b} = [23.812] \quad \mathbf{u} = \begin{bmatrix} 1.191 \\ 0.172 \\ -1.291 \\ 0.799 \\ -0.060 \end{bmatrix} \quad \sigma_a^2 = 4.004 \quad \sigma_e^2 = 6.987$$

Using genomic information instead of pedigree to build the relationship matrix

- There are many different methodologies, the resulting variance components may differ a little but the rank of breeding is expected to be the same. For example:

$$a. \quad G = \frac{(M-P)(M-P)'}{2 \sum_j p_j(1-p_j)}$$

$$b. \quad G = \frac{MM'}{k}$$

$$c. \quad G = 1 + \frac{MM'}{m}$$

$$d. \quad G = 1 - \frac{\sum_j |m_{xj} - m_{yj}|}{2m}$$

Linear genomic relationship matrices for VCA

Zeng ZB, Wang T, Zou W (2005) Modeling quantitative trait loci and interpretation of models. Genetics 169(3):1711-25.

The G2A-model represents the Cockerham's model in a multiple regression context as

$$g = \mu + h_a \alpha + h_d d$$

where $p = p_A$, $q = (1 - p_A)$, h_a and h_d are defined as

$$\mathbf{Z} = \begin{cases} (2 - 2p) & \text{for genotypes } \begin{cases} AA \\ Aa \\ aa \end{cases} \\ (1 - 2p) & \\ -2p & \end{cases}$$

$$\mathbf{W} = \begin{cases} -2q^2 & \text{for genotypes } \begin{cases} AA \\ Aa \\ aa \end{cases} \\ 2pq & \\ -2p^2 & \end{cases}$$

Additive

$$\mathbf{z} = \text{code}\{\mathbf{0}, \mathbf{1}, \mathbf{2}\} - \text{mean}(\mathbf{z})$$

Dominance

$$\mathbf{w} = \text{code}\{\mathbf{0}, \mathbf{1}, \mathbf{0}\} - \text{mean}(\mathbf{w})$$

Xu S (2013) Mapping quantitative trait loci by controlling polygenic background effect. Genetics 195(4): 1209-1222.

Table 1 Formulas used to calculate marker generated kinship matrices

Type of effect	Original kinship matrix	Normalization factor ^a	Kinship matrix
Additive (a)	$K_a^* = \sum_{k=1}^m Z_k Z_k^\top$	$c_a = \text{mean}[\text{diag}(K_a^*)]$	$K_a = (1/c_a)K_a^*$
Dominance (d)	$K_d^* = \sum_{k=1}^m W_k W_k^\top$	$c_d = \text{mean}[\text{diag}(K_d^*)]$	$K_d = (1/c_d)K_d^*$
Additive \times additive (aa)	$K_{aa}^* = \sum_{k=1}^{m-1} \sum_{k'=k+1}^m (Z_k \# Z_{k'})(Z_k \# Z_{k'})^\top$	$c_{aa} = \text{mean}[\text{diag}(K_{aa}^*)]$	$K_{aa} = (1/c_{aa})K_{aa}^*$
Dominance \times dominance (dd)	$K_{dd}^* = \sum_{k=1}^{m-1} \sum_{k'=k+1}^m (W_k \# W_{k'})(W_k \# W_{k'})^\top$	$c_{dd} = \text{mean}[\text{diag}(K_{dd}^*)]$	$K_{dd} = (1/c_{dd})K_{dd}^*$
Additive \times dominance (ad)	$K_{ad}^* = \sum_{k=1}^{m-1} \sum_{k'=k+1}^m (Z_k \# W_{k'})(Z_k \# W_{k'})^\top$	$c_{ad} = \text{mean}[\text{diag}(K_{ad}^*)]$	$K_{ad} = (1/c_{ad})K_{ad}^*$
Dominance \times additive (da)	$K_{da}^* = \sum_{k=1}^{m-1} \sum_{k'=k+1}^m (W_k \# Z_{k'})(W_k \# Z_{k'})^\top$	$c_{da} = \text{mean}[\text{diag}(K_{da}^*)]$	$K_{da} = (1/c_{da})K_{da}^*$

^a Each marker-generated kinship matrix is normalized so that the diagonal elements are roughly equal to one. This is achieved by dividing the original kinship matrix by the mean value of all the diagonal elements of the original matrix. Using the normalized kinship matrix will bring the estimated genetic variance in the same scale as the residual error variance.

- PS. You can simply use the function G2A_Kernels(Genotypes) from NAM

Example: G Matrices (from Xu 2013)

- Matrix Z codes additive (-101) and the matrix W codes dominance (010)

> Z	ch1p7	ch1p330	ch2p129	ch2p399	ch3p227	ch4p32	ch4p315	ch5p128	ch5p382	ch6p124	ch7p34	ch7p312	ch8p111	ch8p418	ch9p217	ch10p1	ch10p259
id18	1	-1	0	0	-1	1	1	0	0	1	1	0	0	0	0	0	0
id56	1	1	0	0	1	0	0	0	0	1	0	0	0	-1	0	-1	-1
id102	1	1	-1	1	0	1	1	-1	0	-1	0	1	1	1	0	0	0
id159	-1	-1	0	0	0	1	0	0	1	1	-1	-1	0	0	0	1	1
id164	0	0	-1	0	1	-1	-1	1	1	1	0	0	-1	-1	0	0	0
id177	-1	-1	-1	0	-1	0	0	1	1	0	1	1	-1	0	-1	0	0
id232	-1	-1	1	1	1	0	0	-1	1	0	1	1	1	1	0	0	-1
id251	-1	0	0	0	1	1	1	0	0	0	-1	0	0	0	0	0	-1
id374	1	1	0	0	0	1	1	-1	-1	1	-1	-1	0	0	0	0	0
id395	1	0	1	1	1	1	1	0	-1	0	-1	0	0	0	0	0	0
> W	ch1p7	ch1p330	ch2p129	ch2p399	ch3p227	ch4p32	ch4p315	ch5p128	ch5p382	ch6p124	ch7p34	ch7p312	ch8p111	ch8p418	ch9p217	ch10p1	ch10p259
id18	0	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	1
id56	0	0	1	1	0	1	1	1	1	0	1	1	1	0	1	0	0
id102	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	1	1
id159	0	0	1	1	1	0	1	1	0	0	0	0	1	1	1	0	0
id164	1	1	0	1	0	0	0	0	0	0	1	1	0	0	1	1	1
id177	0	0	0	1	0	1	1	0	0	1	0	0	0	1	0	1	1
id232	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1	1	0
id251	0	1	1	1	0	0	0	1	1	1	0	1	1	1	1	1	0
id374	0	0	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1
id395	0	1	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1

Example: G Matrices (from Xu 2013)

- Below, we are estimating the **additive** and **dominance** relationships

- $\mathbf{A} = \frac{\mathbf{ZZ}'}{k_a}$, where $k_a = \text{mean}(\text{diag}(\mathbf{ZZ}'))$
- $\mathbf{D} = \frac{\mathbf{WW}'}{k_d}$, where $k_d = \text{mean}(\text{diag}(\mathbf{WW}'))$

```
> ZZ = tcrossprod(z); A = ZZ/mean(diag(ZZ))
> WW = tcrossprod(w); D = ZZ/mean(diag(ZZ))
> A
      id18     id56     id102    id159     id164     id177     id232    id251     id374     id395
id18  0.7865169  0.0000000  0.1123596  0.1123596 -0.2247191  0.2247191  0.0000000 -0.1123596  0.2247191  0.1123596
id56  0.0000000  0.7865169  0.0000000 -0.3370787  0.3370787 -0.3370787 -0.1123596  0.1123596  0.3370787  0.1123596
id102 0.1123596  0.0000000  1.2359551 -0.3370787 -0.5617978 -0.2247191  0.2247191  0.1123596  0.3370787  0.2247191
id159 0.1123596 -0.3370787 -0.3370787  1.0112360  0.1123596  0.1123596  0.0000000  0.2247191  0.1123596  0.0000000
id164 -0.2247191  0.3370787 -0.5617978  0.1123596  1.0112360  0.3370787 -0.2247191 -0.1123596 -0.3370787 -0.2247191
id177 0.2247191 -0.3370787 -0.2247191  0.1123596  0.3370787  1.1235955  0.1123596 -0.1123596 -0.6741573 -0.3370787
id232 0.0000000 -0.1123596  0.2247191  0.0000000 -0.2247191  0.1123596  1.3483146  0.2247191 -0.4494382  0.0000000
id251 -0.1123596  0.1123596  0.1123596  0.2247191 -0.1123596 -0.1123596  0.2247191  0.6741573  0.2247191  0.2247191
id374 0.2247191  0.3370787  0.3370787  0.1123596 -0.3370787 -0.6741573 -0.4494382  0.2247191  1.0112360  0.2247191
id395 0.1123596  0.1123596  0.2247191  0.0000000 -0.2247191 -0.3370787  0.0000000  0.2247191  0.2247191  1.0112360
> D
      id18     id56     id102    id159     id164     id177     id232    id251     id374     id395
id18  0.7865169  0.0000000  0.1123596  0.1123596 -0.2247191  0.2247191  0.0000000 -0.1123596  0.2247191  0.1123596
id56  0.0000000  0.7865169  0.0000000 -0.3370787  0.3370787 -0.3370787 -0.1123596  0.1123596  0.3370787  0.1123596
id102 0.1123596  0.0000000  1.2359551 -0.3370787 -0.5617978 -0.2247191  0.2247191  0.1123596  0.3370787  0.2247191
id159 0.1123596 -0.3370787 -0.3370787  1.0112360  0.1123596  0.1123596  0.0000000  0.2247191  0.1123596  0.0000000
id164 -0.2247191  0.3370787 -0.5617978  0.1123596  1.0112360  0.3370787 -0.2247191 -0.1123596 -0.3370787 -0.2247191
id177 0.2247191 -0.3370787 -0.2247191  0.1123596  0.3370787  1.1235955  0.1123596 -0.1123596 -0.6741573 -0.3370787
id232 0.0000000 -0.1123596  0.2247191  0.0000000 -0.2247191  0.1123596  1.3483146  0.2247191 -0.4494382  0.0000000
id251 -0.1123596  0.1123596  0.1123596  0.2247191 -0.1123596 -0.1123596  0.2247191  0.6741573  0.2247191  0.2247191
id374 0.2247191  0.3370787  0.3370787  0.1123596 -0.3370787 -0.6741573 -0.4494382  0.2247191  1.0112360  0.2247191
id395 0.1123596  0.1123596  0.2247191  0.0000000 -0.2247191 -0.3370787  0.0000000  0.2247191  0.2247191  1.0112360
```

Example: G Matrices (from Xu 2013)

- Below, we are estimating the epistatic relationships
 - $A \times A = \frac{ZZ' \circ ZZ'}{k_{aa}}$, where $k_{aa} = \text{mean}(\text{diag}(ZZ' \circ ZZ'))$
 - $A \times D = \frac{ZZ' \circ WW'}{k_{ad}}$, where $k_{ad} = \text{mean}(\text{diag}(ZZ' \circ WW'))$
 - $D \times D = \frac{WW' \circ WW'}{k_{dd}}$, where $k_{dd} = \text{mean}(\text{diag}(WW' \circ WW'))$

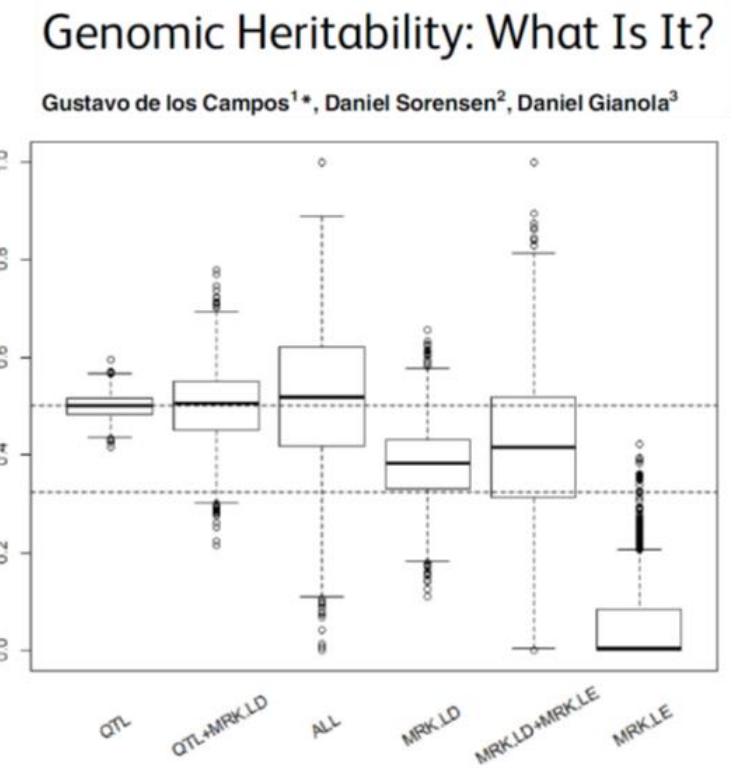
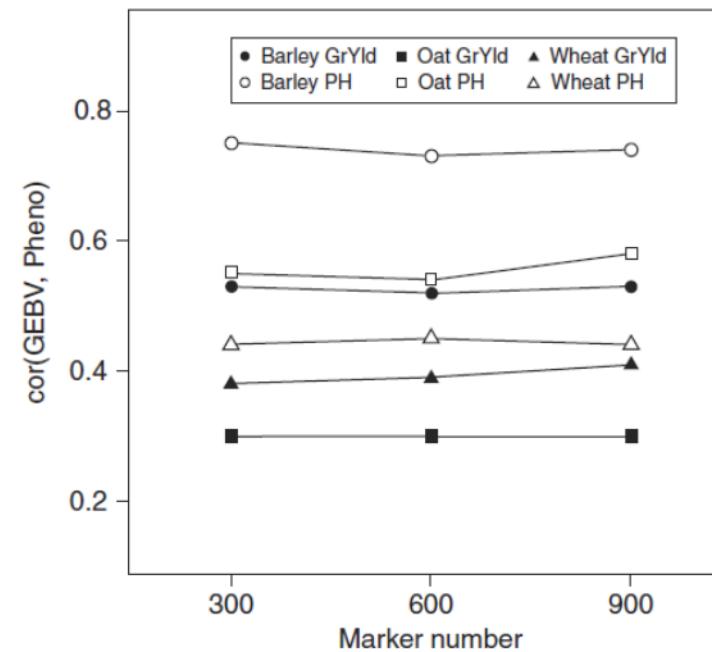
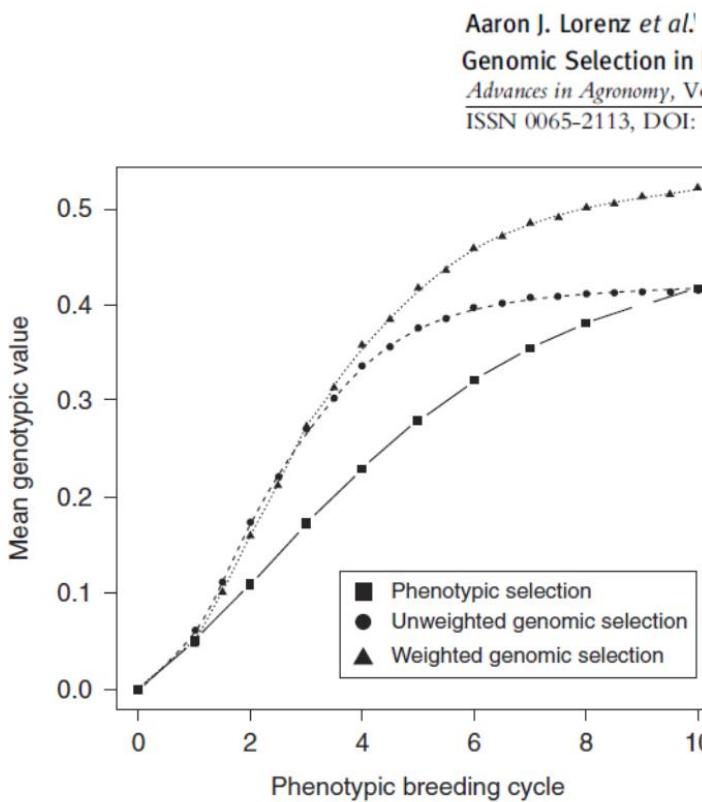
```
> AA = ZZ*ZZ; AA = AA/mean(diag(AA))
> AD = ZZ*WW; AD = AD/mean(diag(AD))
> DD = WW*WW; DD = DD/mean(diag(DD))
> AA
      id18     id56     id102    id159     id164     id177     id232     id251     id374     id395
id18  0.59538275 0.00000000 0.01215067 0.01215067 0.04860267 0.04860267 0.00000000 0.01215067 0.04860267 0.01215067
id56  0.00000000 0.59538275 0.00000000 0.10935601 0.10935601 0.10935601 0.01215067 0.01215067 0.10935601 0.01215067
id102 0.01215067 0.00000000 1.47023086 0.10935601 0.30376671 0.04860267 0.04860267 0.01215067 0.10935601 0.04860267
id159 0.01215067 0.10935601 0.10935601 0.98420413 0.01215067 0.01215067 0.00000000 0.04860267 0.01215067 0.00000000
id164 0.04860267 0.10935601 0.30376671 0.01215067 0.98420413 0.10935601 0.04860267 0.01215067 0.10935601 0.04860267
id177 0.04860267 0.10935601 0.04860267 0.01215067 0.10935601 1.21506683 0.01215067 0.01215067 0.43742406 0.10935601
id232 0.00000000 0.01215067 0.04860267 0.00000000 0.04860267 0.01215067 1.74969623 0.04860267 0.19441069 0.00000000
id251 0.01215067 0.01215067 0.01215067 0.04860267 0.01215067 0.01215067 0.04860267 0.43742406 0.04860267 0.04860267
id374 0.04860267 0.10935601 0.10935601 0.01215067 0.10935601 0.43742406 0.19441069 0.04860267 0.98420413 0.04860267
id395 0.01215067 0.01215067 0.04860267 0.00000000 0.04860267 0.10935601 0.00000000 0.04860267 0.04860267 0.98420413
```

Follow up & remarks

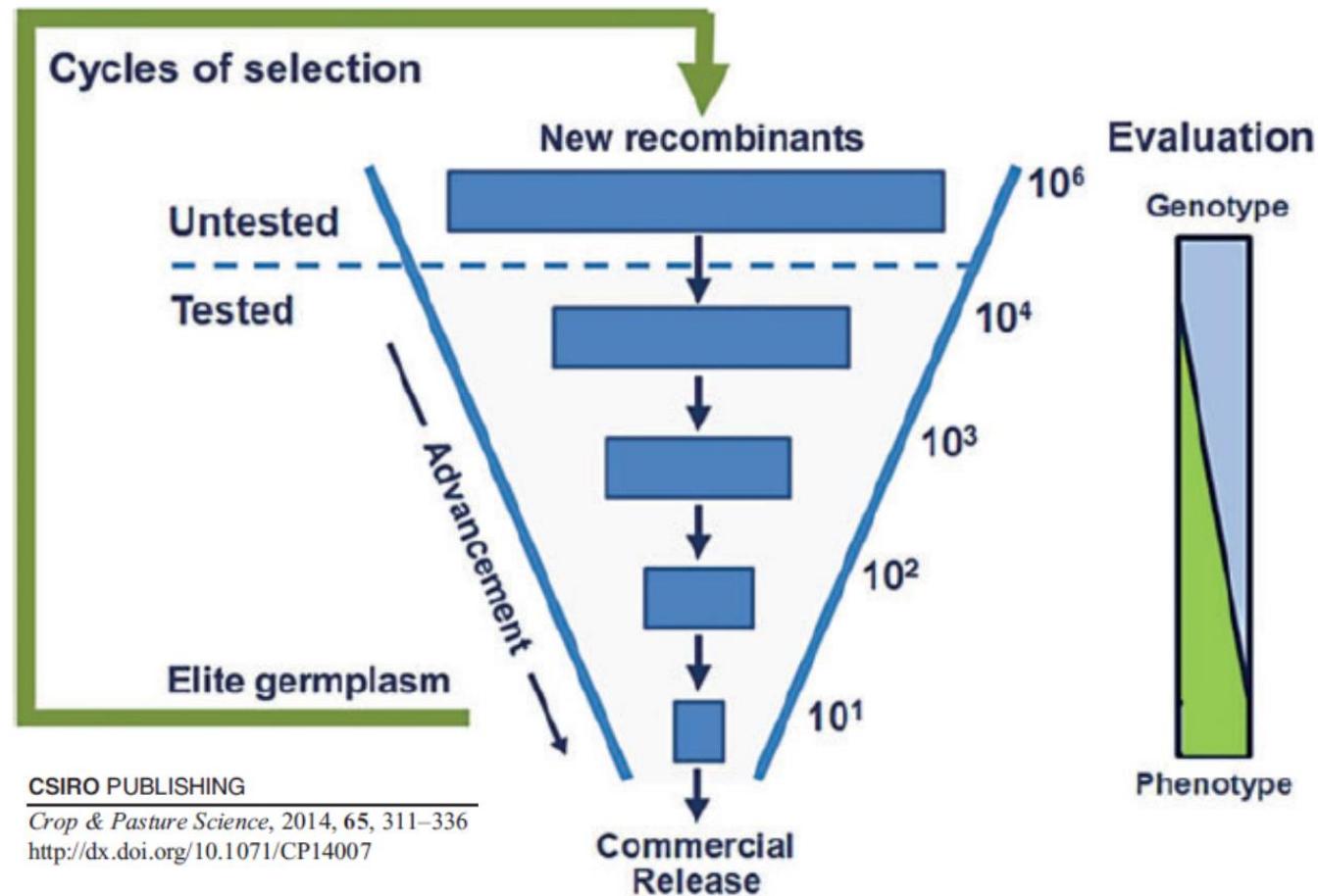
Follow up - recommended literature

- Sewall Wright (1922) Coefficients of inbreeding and relationship
- Douglas S Falconer (1964) Introduction to quantitative genetics
- Rasmus Nielsen (2005) Molecular signatures of natural selection
- Warren J Ewens (2006) Mathematical Population Genetics
- Vlastimil Kirivan (2008) Evolutionary games and population dynamics

Successful selection with or without genomic tools is obtained with the proper combination of **strategy**-**data**-**method**



As the selection stage advances within a breeding cycle, the contribution from more and better phenotypic information overweight genomic information



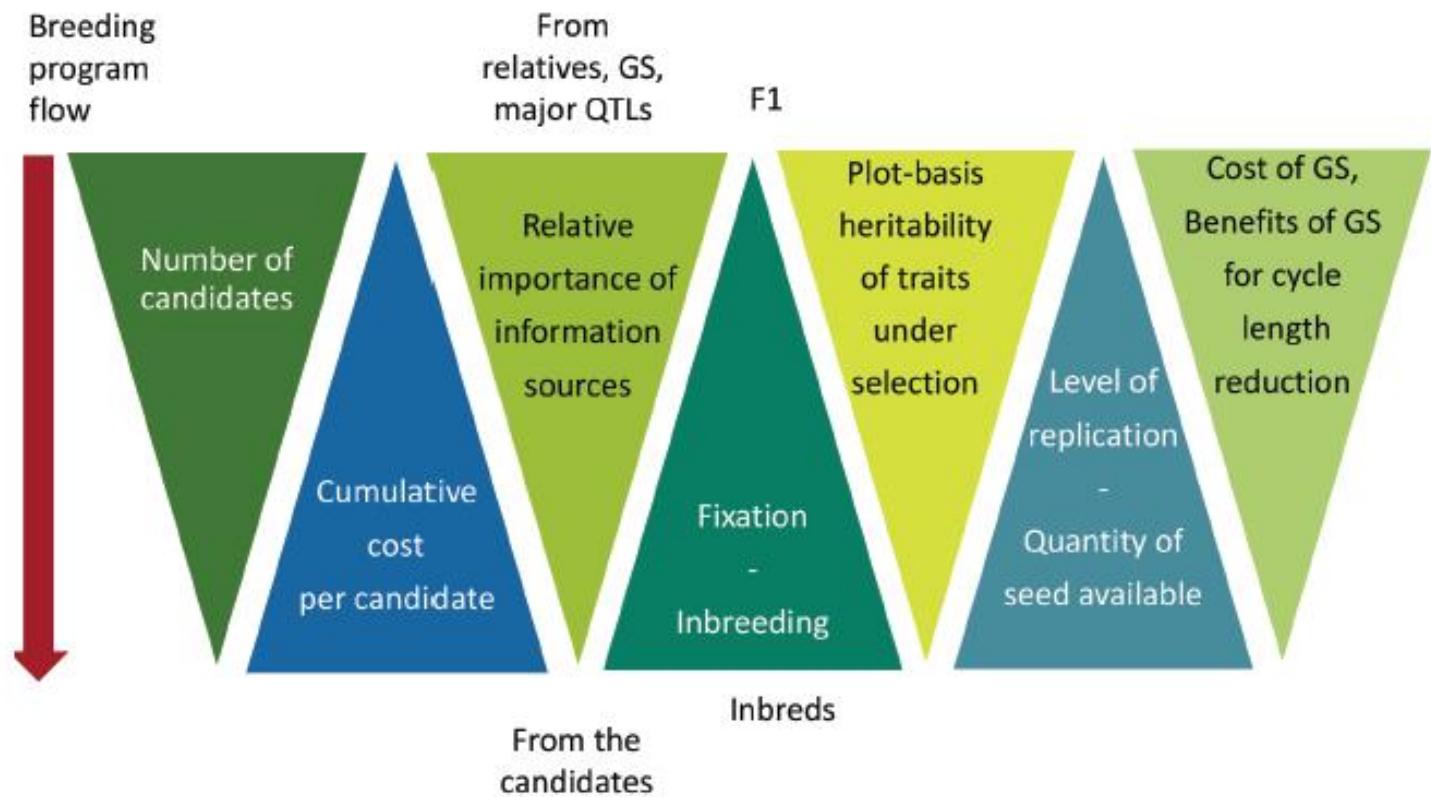


Figure 1. Key parameters and changes during a breeding cycle, to consider in implementing genomic selection (GS). The triangles indicate increase or decrease of the quantity considered. QTL, quantitative trait loci.

Heslot, N., Jannink, J. L., & Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Science*, 55(1), 1-12.

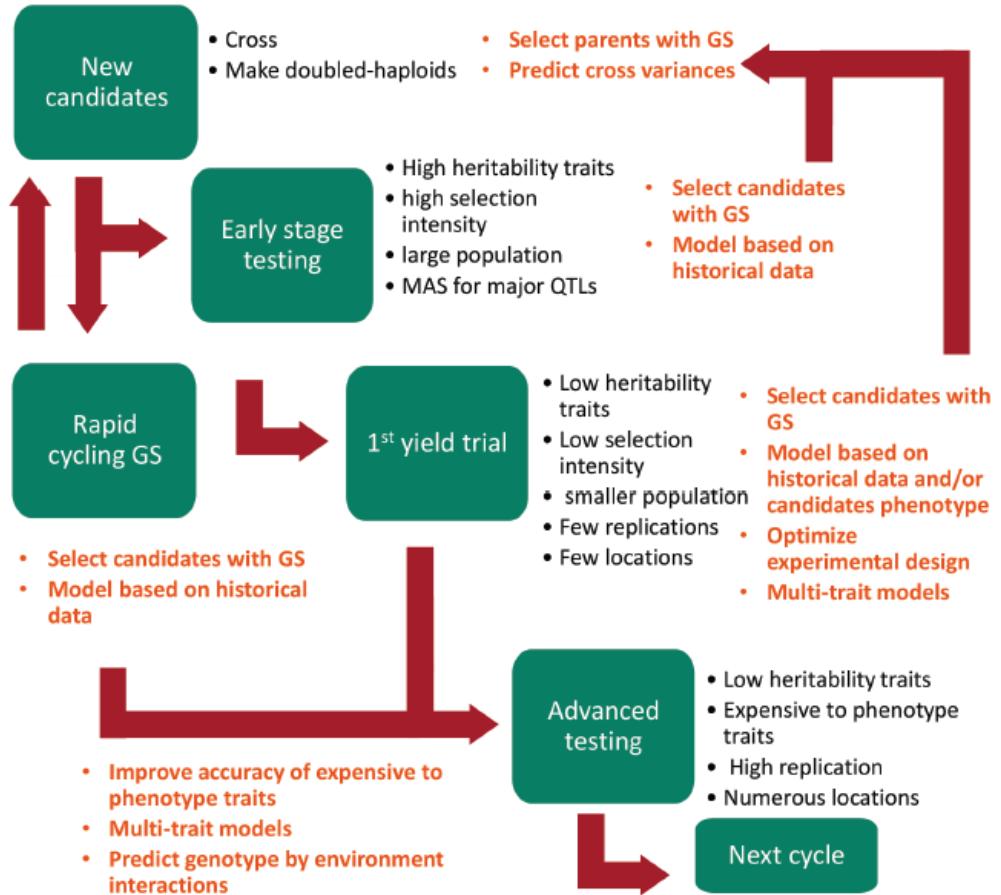
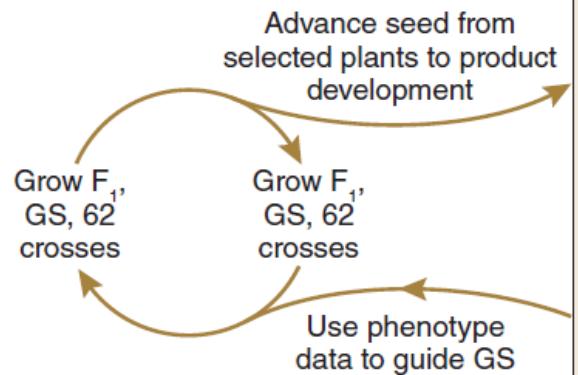


Figure 2. Simple scheme of a breeding cycle with what genomic selection (GS) could bring for each stage (orange). Arrows indicate the flow of germplasm. Upward arrows correspond to early re-crossing. For the sake of simplicity, the scheme uses doubled-haploids. MAS, marker-assisted selection; QTL, quantitative trait loci.

Heslot, N., Jannink, J. L., & Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Science*, 55(1), 1-12.

Population improvement



Product development			
Year	Generation	Number of plants	Action
1	F_1	124 half-sib families	Increase in greenhouse
2	F_2	1,000 plants per family	Bulk 50 plants per family
3	F_3	1,000 plants per family	Bulk 50 plants per family
4	F_4	1,000 plants per family	Derive new lines from 50 plants per family
5	$F_{4:5}$	6,200 headrows	Advance 1,000 lines
6	PYT, $F_{4:6}$	1,000 lines	Yield trial, genotype
7	AYT, $F_{4:7}$	100 lines	Yield trial
8	EYT, $F_{4:8}$	10 lines	Yield trial
9	EYT, $F_{4:9}$	10 lines	Yield trial
10	$F_{4:10}$	1 line	Release variety

Figure 3 A variant of the two-part breeding program design for plant breeding. GS, genomic selection; PYT, primary yield trials; AYT, advanced yield trials; EYT, elite yield trials.

Hickey, J. M., Chiurugwi, T., Mackay, I., & Powell, W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature genetics*, 49(9), 1297.

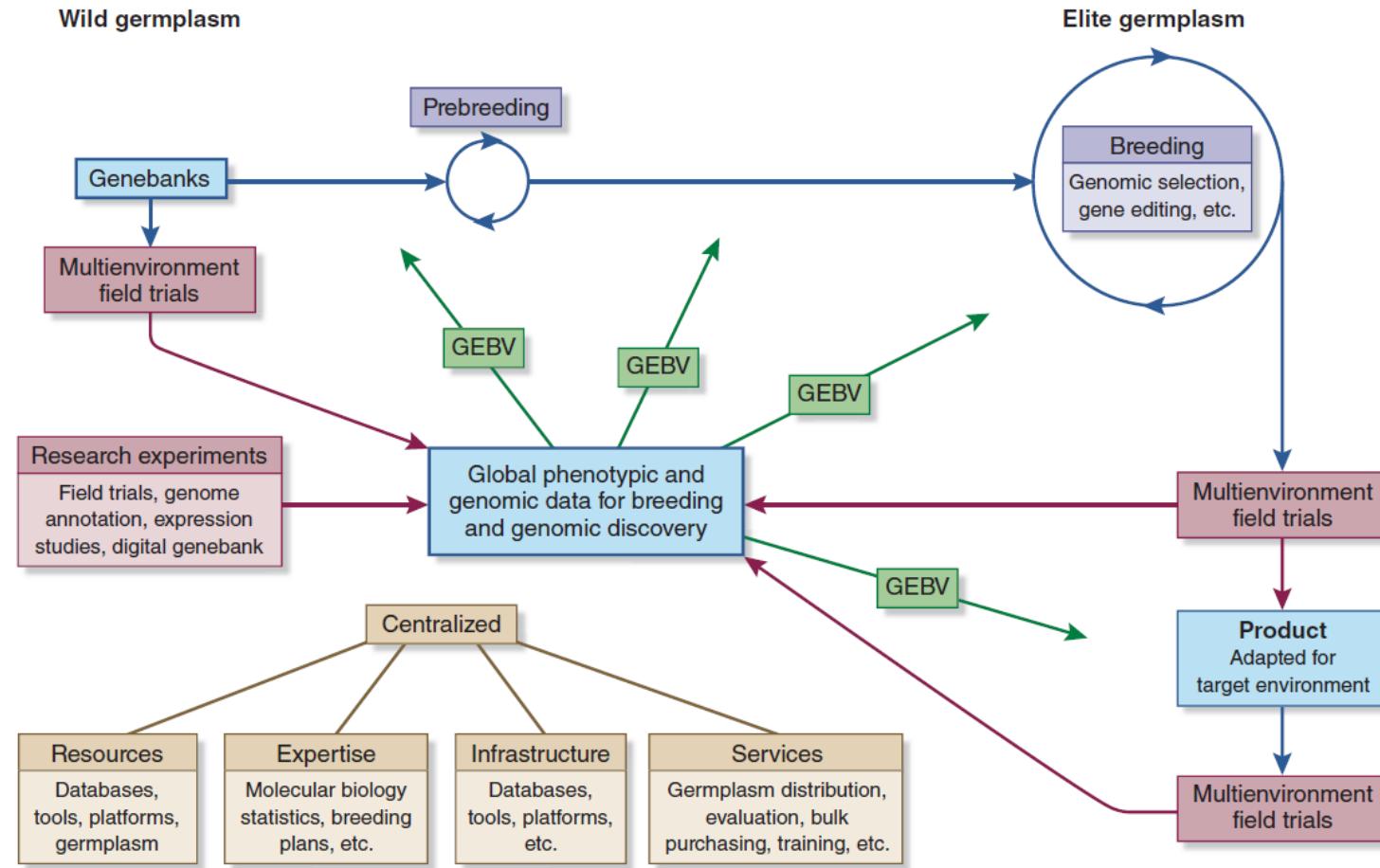


Figure 4 Framework for combining approaches. Capturing new opportunities to accelerate the pace of genetic gain based on efficient and targeted access to genetic diversity, coordinated phenotyping across environments, cost-effective sequencing, genomic prediction and genome editing.

Hickey, J. M., Chiurugwi, T., Mackay, I., & Powell, W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature genetics*, 49(9), 1297.