Introduction to Plant Breeding

A Parochial view

Origins of crops
Scientific approaches 1850... present
Plant & animal breeding compared
Achievements & questions

Matthew 7:18-7:20 A good tree cannot bring forth evil fruit, neither can a corrupt tree bring forth good fruit. Every tree that bringeth not forth good fruit is hewn down, and cast into the fire. Wherefore by their fruits ye shall know them.

The Scientific approach to plant breeding

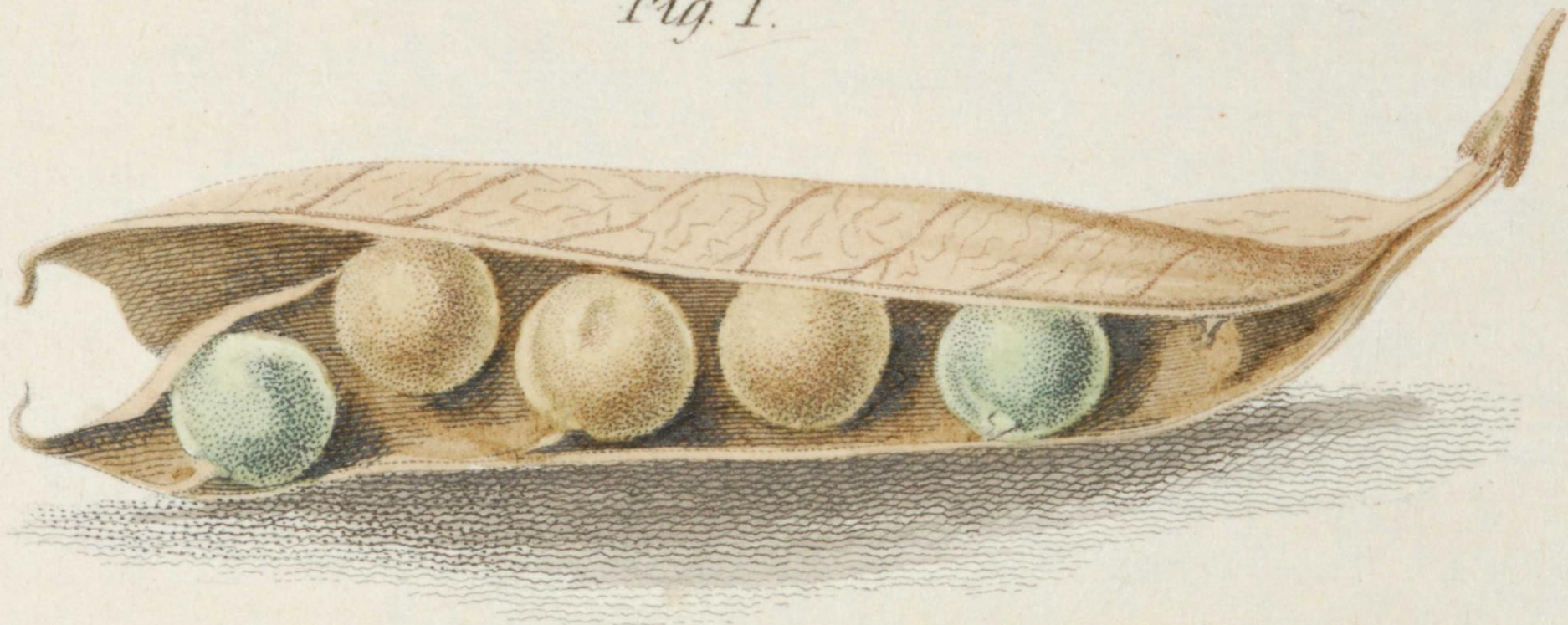Two strands:

1. Mendelian:
   Incorporate information from genes into selection decisions
   championed by plant breeders

2. Biometric:
   Incorporate information from relatives into selection decisions
   championed by animal breeders

   Prospects: we now have the technology to combine the two.

Fig. 1.

John Goss (1824) On Variation in the Colour of Peas, occasioned by Cross Impregnation
Horticultural Transactions (Series 1) Vol:5, p. 234-237 + 1 fig

Some milestones in Mendelian genetics & breeding

1823:   Knight: Dominance, recessiveness, and segregation observed in peas

1900:   Rediscovery and verification of Mendel's principles

1903:   Biffen: resistance to stripe rust of wheat is Mendelian recessive.

1908:   Nilsson-Ehle: seed colour in wheat is due to 3 Mendelian factors.

1923:   Sax: linkage between quantitative and qualitative traits in beans.

1956:   Flor: gene for gene hypothesis for host-parasite resistance

1965-70   Borlaug: Green Revolution (India & Pakistan) based on dwarfing genes.

1983:   Beckmann & Soller : RFLPs for genome wide QTL detection and breeding

2001:   Meuwissen *et al* : Genomic selection proposed

Wheat
Genetic history: plant breeding.

Dwarfing genes reduced the weight of straw, changing the distribution of resources and resulting in:

Which increased susceptibility to disease. But plants were protected by newly developed:

Dwarfing genes allow increases:

•Nitrogen fertiliser levels.

•Higher grain yields.
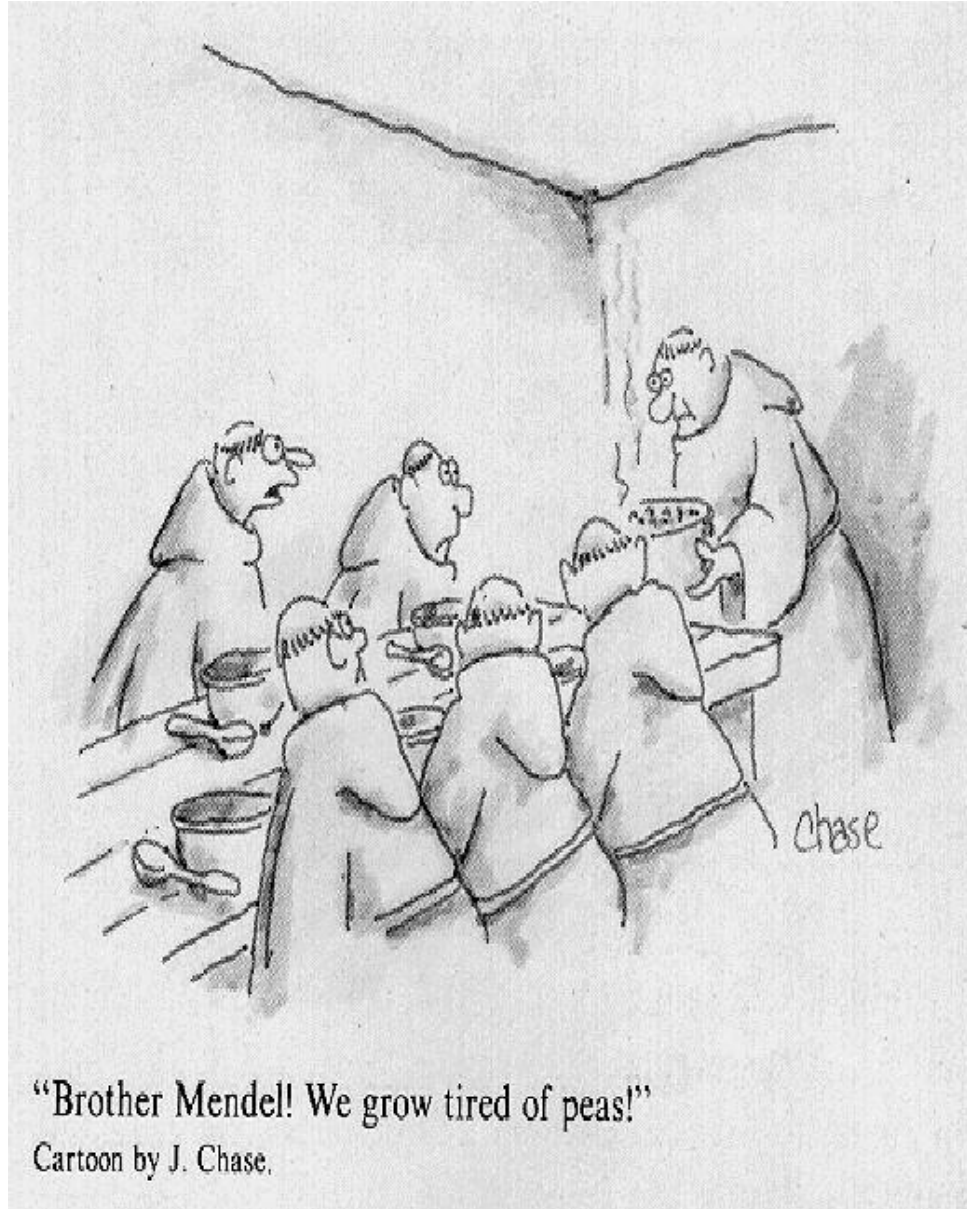
In addition, pleiotropic effects of the dwarfing gene include more grains per ear.

•Fungicide

Quantitative methods in plant breeding –
March '09

Information from genes.

"Brother Mendel! We grow tired of peas!"
Cartoon by J. Chase.

Some milestones in biometrical genetics & breeding

1840-50 de Vilmorin: progeny test in wheat, oat, and sugar-beet breeding.

1889 Galton: publishes Natural Inheritance, a statistical statement of the relative influence of parents

1921 Wright: relationships between relatives

1936 Smith: selection index

1947 Lush : Family merit & individual merit as a basis for selection

1953 Henderson: origins of BLUP

1971 Patterson & Thompson REML

2001 Meuwissen *et al* : Genomic selection proposed

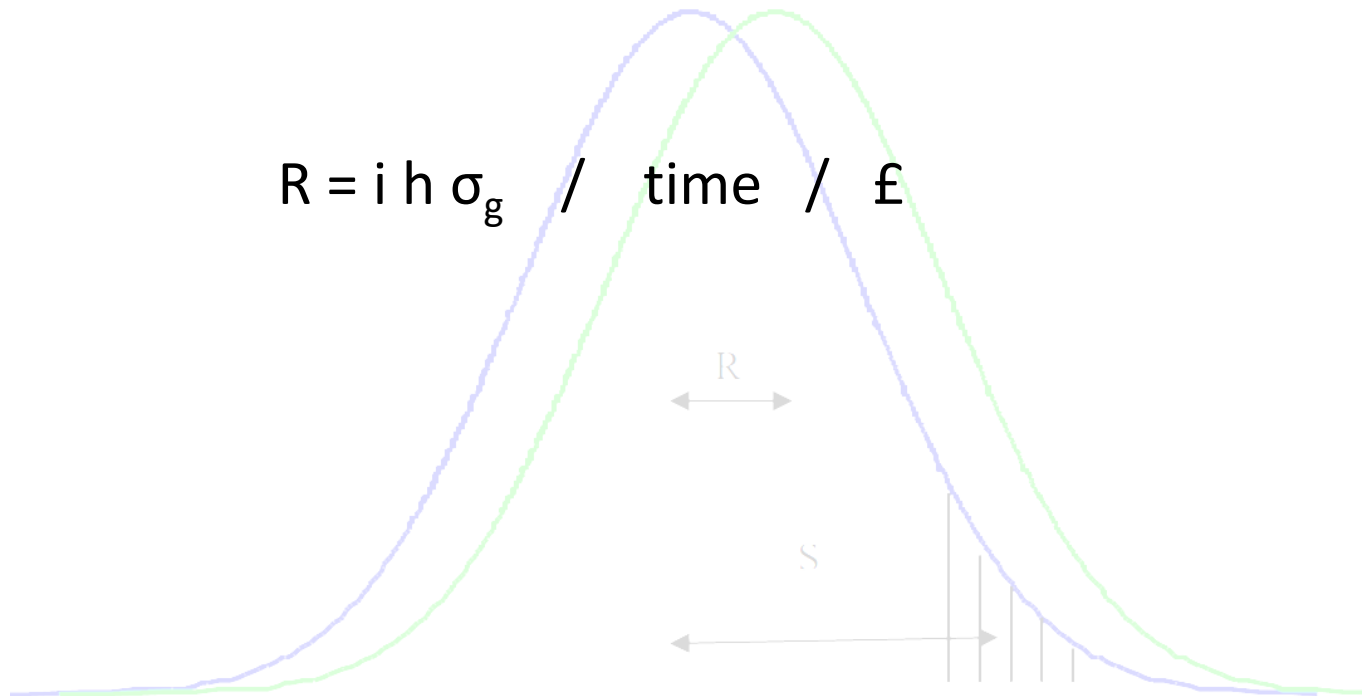Both approaches are linked by the breeders' equation $R = h^2S$.
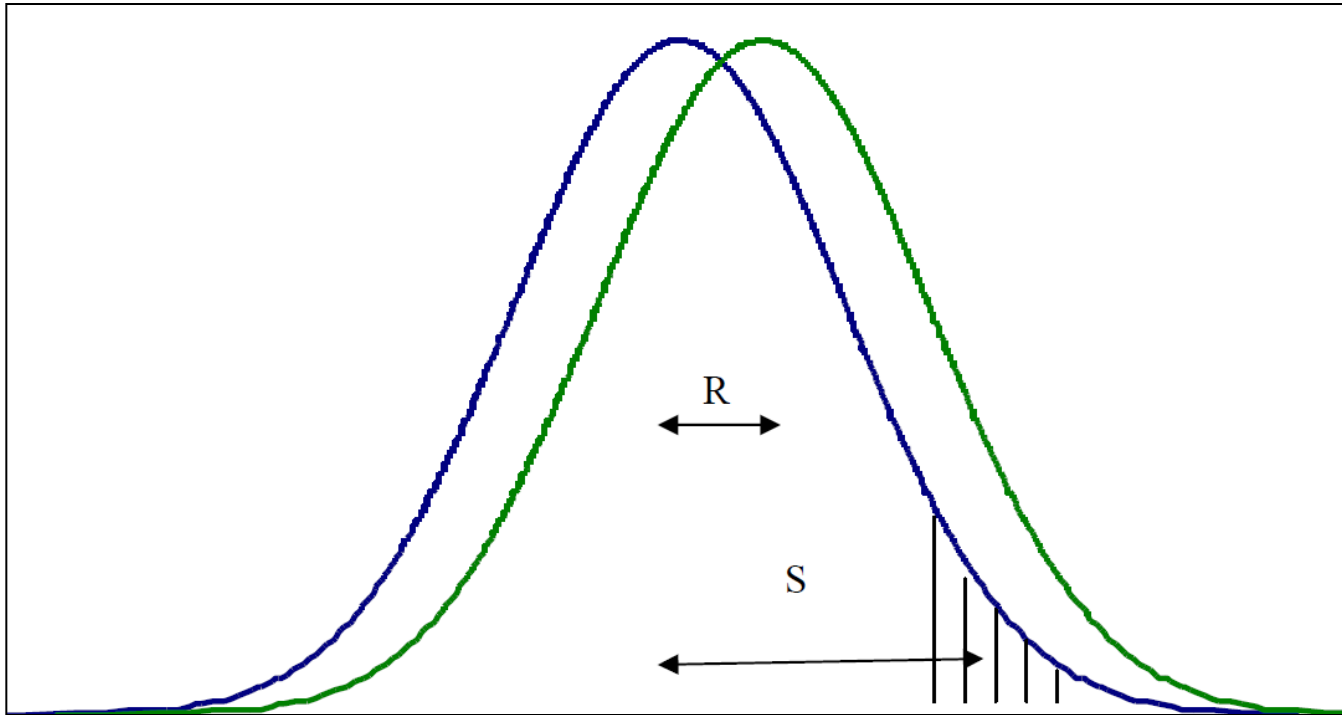
Everything in plant (and animal) breeding can be judged by its effect on "the breeders' equation."

The breeders' equation $R = h^2 S$.

standardized as:

$R = i \, h \, \sigma_g \quad / \quad \text{time} \quad / \quad £$

R

S

Some arbitrary dates in plants breeding methods

1840-50   de Vilmorin   progeny testing

1909   Nilsson-Ehle   scientific wheat breeding:
                                         pedigree breeding, bulk breeding

1878-81   Beal   corn hybrids yield more

1909   Shull: use of F1 hybrids between inbreds in corn breeding

1924   Blakeslee & Belling   report doubled haploids

1939:   Golden   single seed descent

1936   ?   haploids and polyploids

Some features of plant breeding methods

Replicate genotypes:	clones
	inbred lines
	DH lines
	F1 hybrids
Heritabilities	vary through replication
Inbreeding is quick	self: S1,S2..Sn, doubled haploids
Mating systems:	selfing, outcrossing
	gms, cms, S alleles, …
Polyploids:	haploids, allopolyploids, autopolyploids
Use of ancestral species	eg synthetic wheat
GxE	generally larger than in animals
Half sibs	have a common female parent

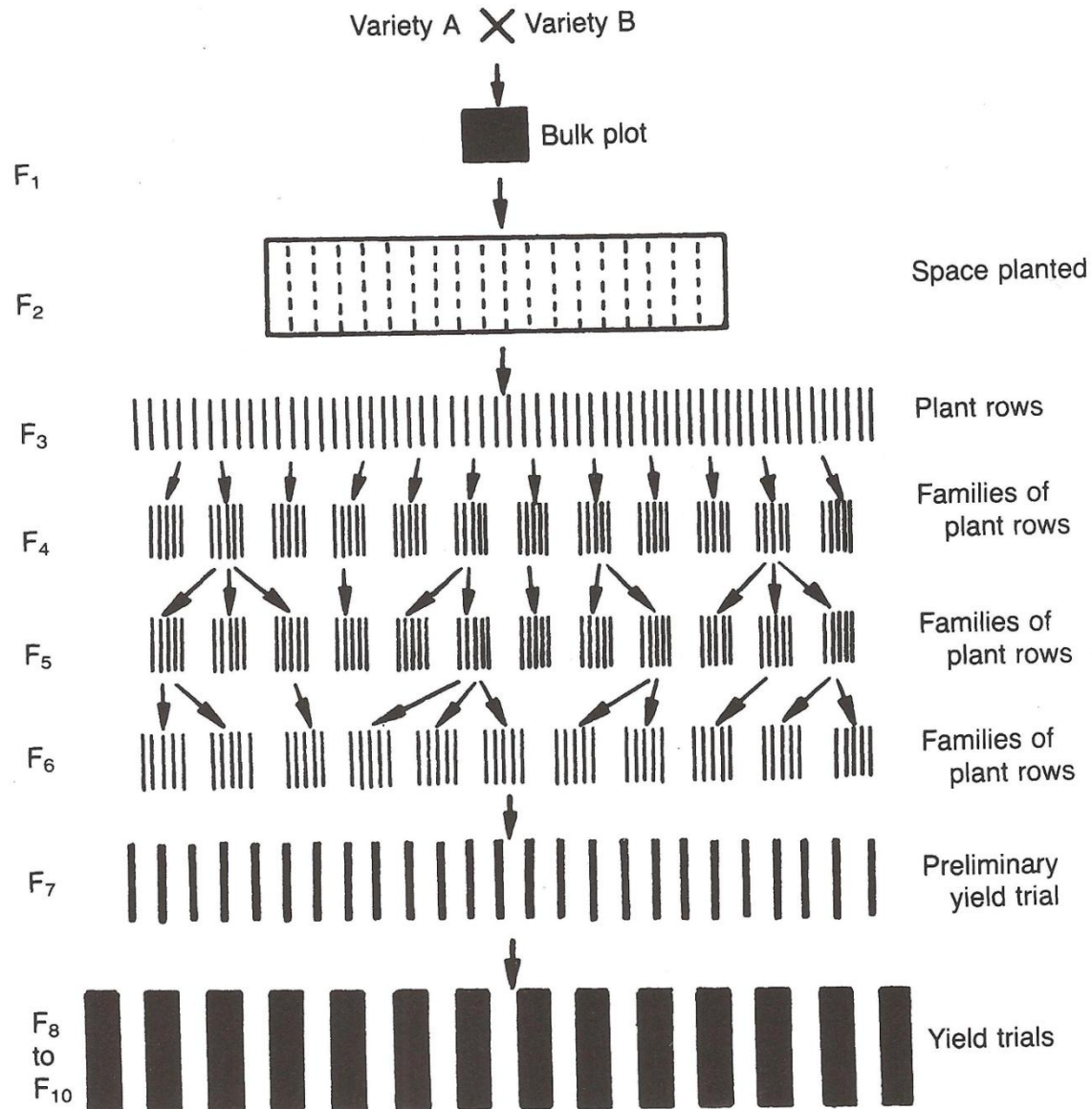# Methods for selection within crosses

Pedigree breeding

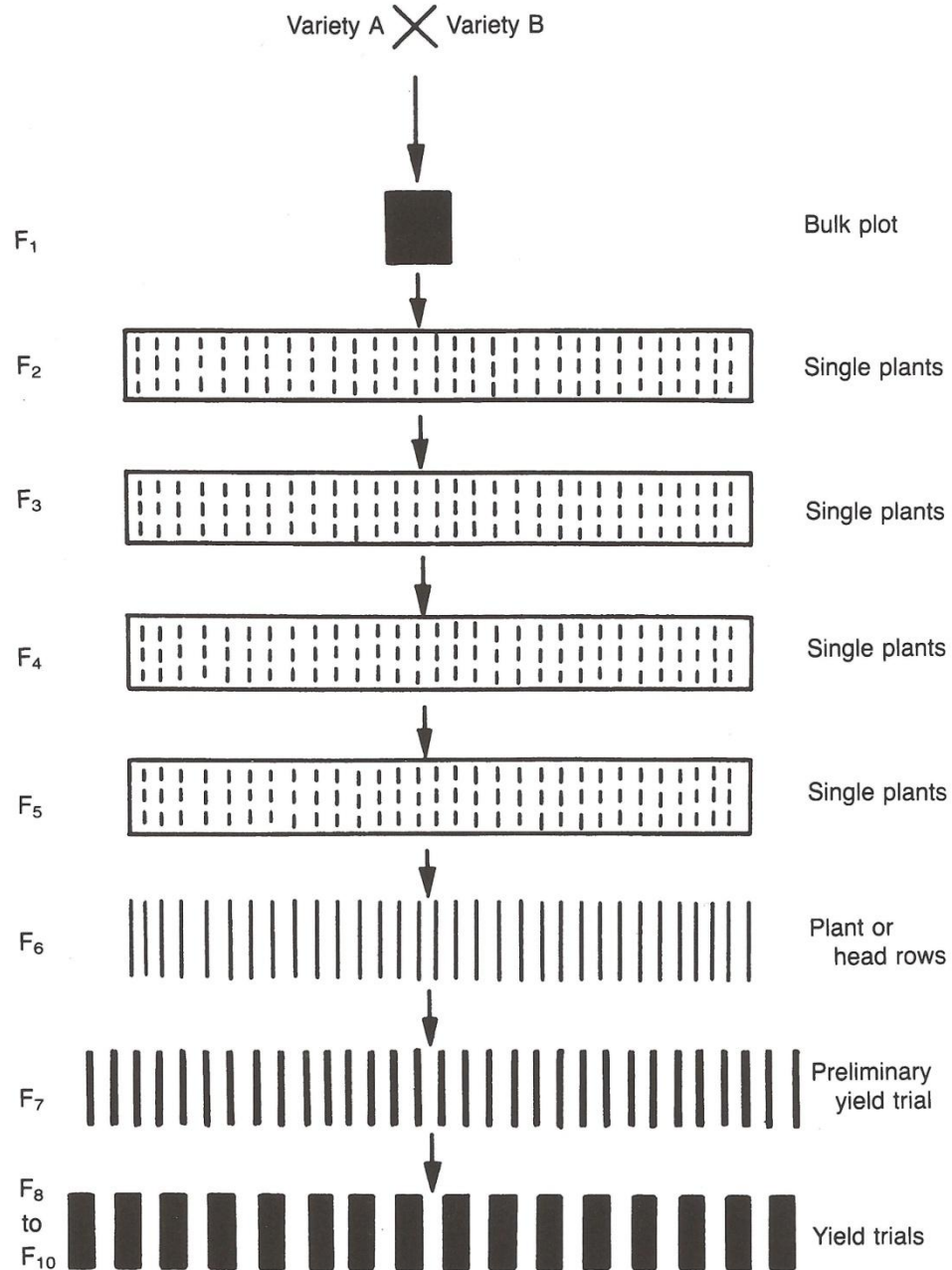Single seed descent

Doubled haploids

Bulk breeding

Pedigree method

Single Seed Descent

Single Seed Descent

Goulden   (1939)
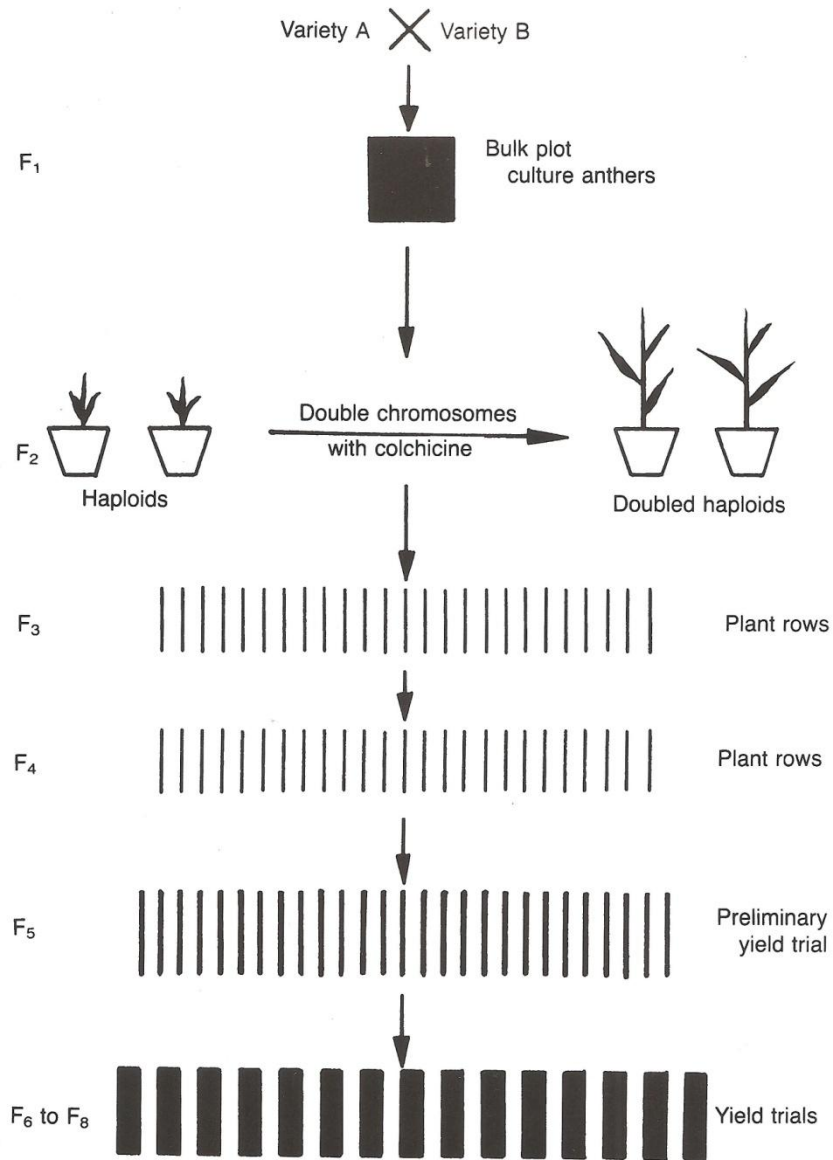Knott & Kumar (1975)   wheat

Pedigree breeding: inbreeding & selection concur

SSD: separate inbreeding from selection (faster)

Proposed and developed for breeding.

Use in trait mapping is more recent.

Doubled Haploids

Doubled Haploids

"The practical importance of haploids and polyploids in plant breeding is being quickly recognised and it seems possible that their artificial production will be simply a matter of technique in the near future." *Imperial Bureau of Plant Genetics, 1936*
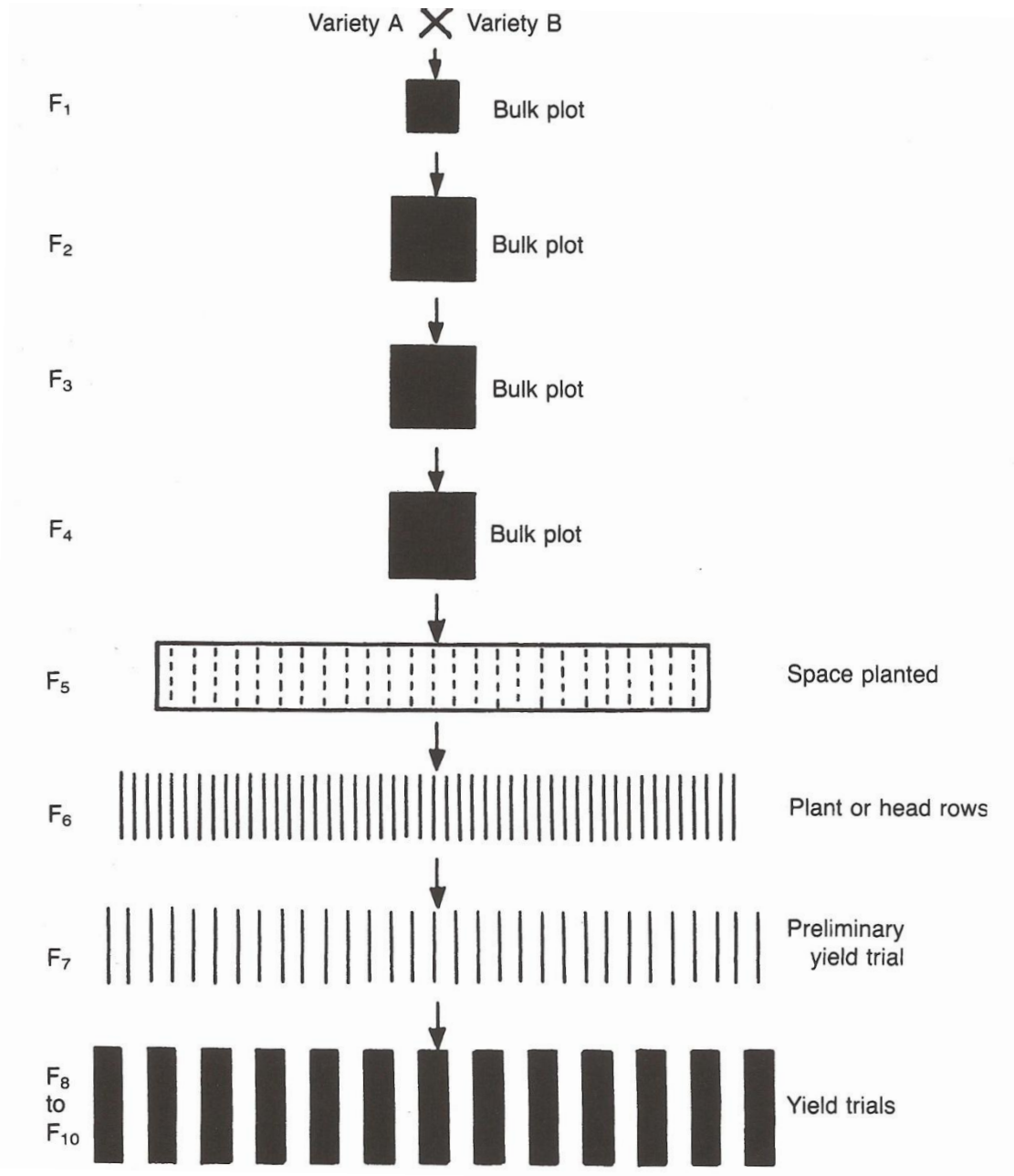
Faster than SSD

Expensive
Low efficiency in some crops
Less recombination

Bulk Breeding

Variety A ✕ Variety B

F_1 — Bulk plot

F_2 — Bulk plot

F_3 — Bulk plot

F_4 — Bulk plot

F_5 — Space planted

F_6 — Plant or head rows

F_7 — Preliminary yield trial

F_8 to F_{10} — Yield trials

Bulk breeding

As slow as pedigree breeding

Encourage selection in the bulk (natural & artificial)

F2s contribute unequally to inbred lines

Long history (Allard, Harlan)

Not much used in commercial plant breeding.

Regularly rediscovered by academics. And funded!

Hybrid breeding

General combining ability

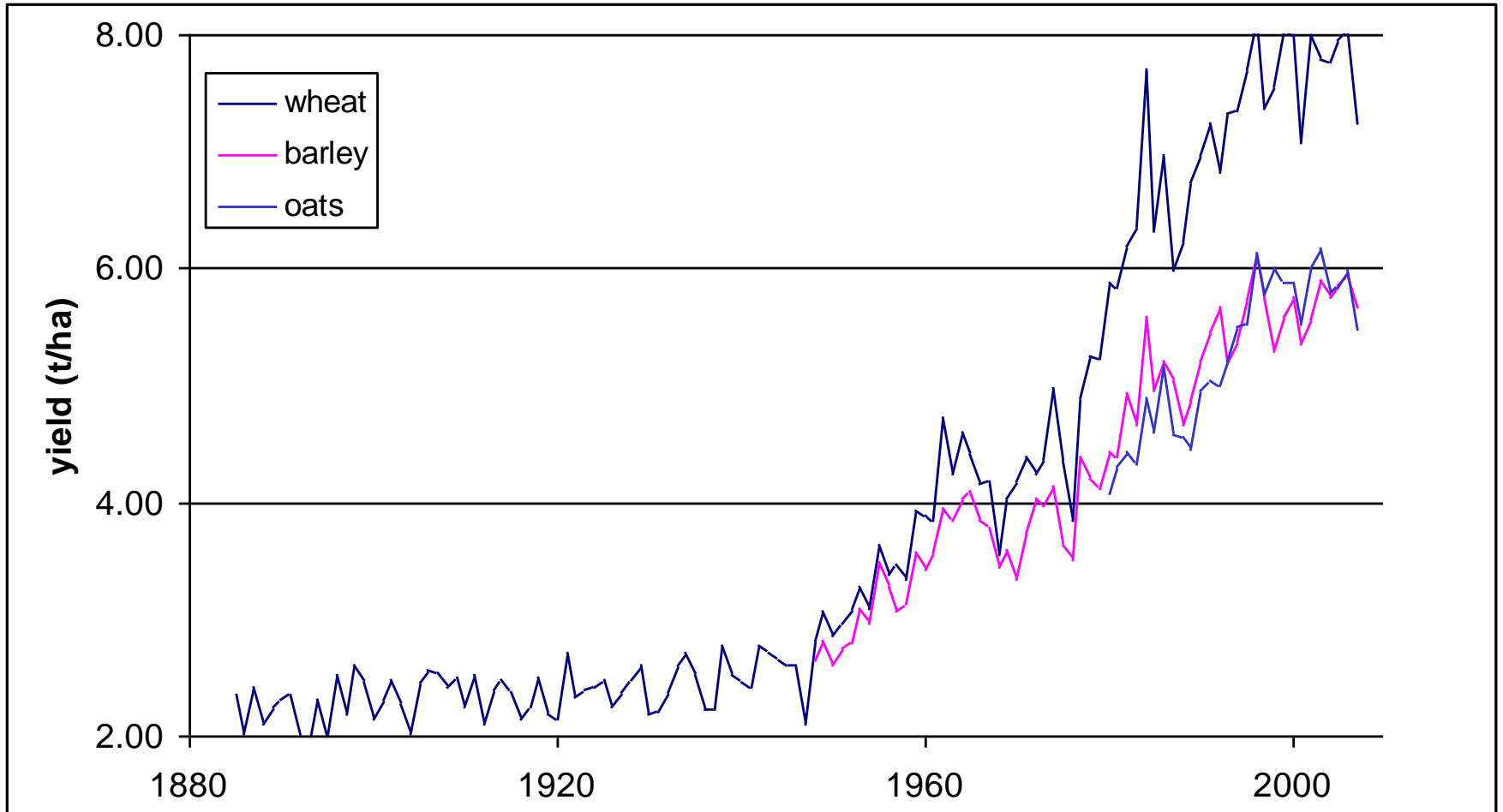Specific combining ability

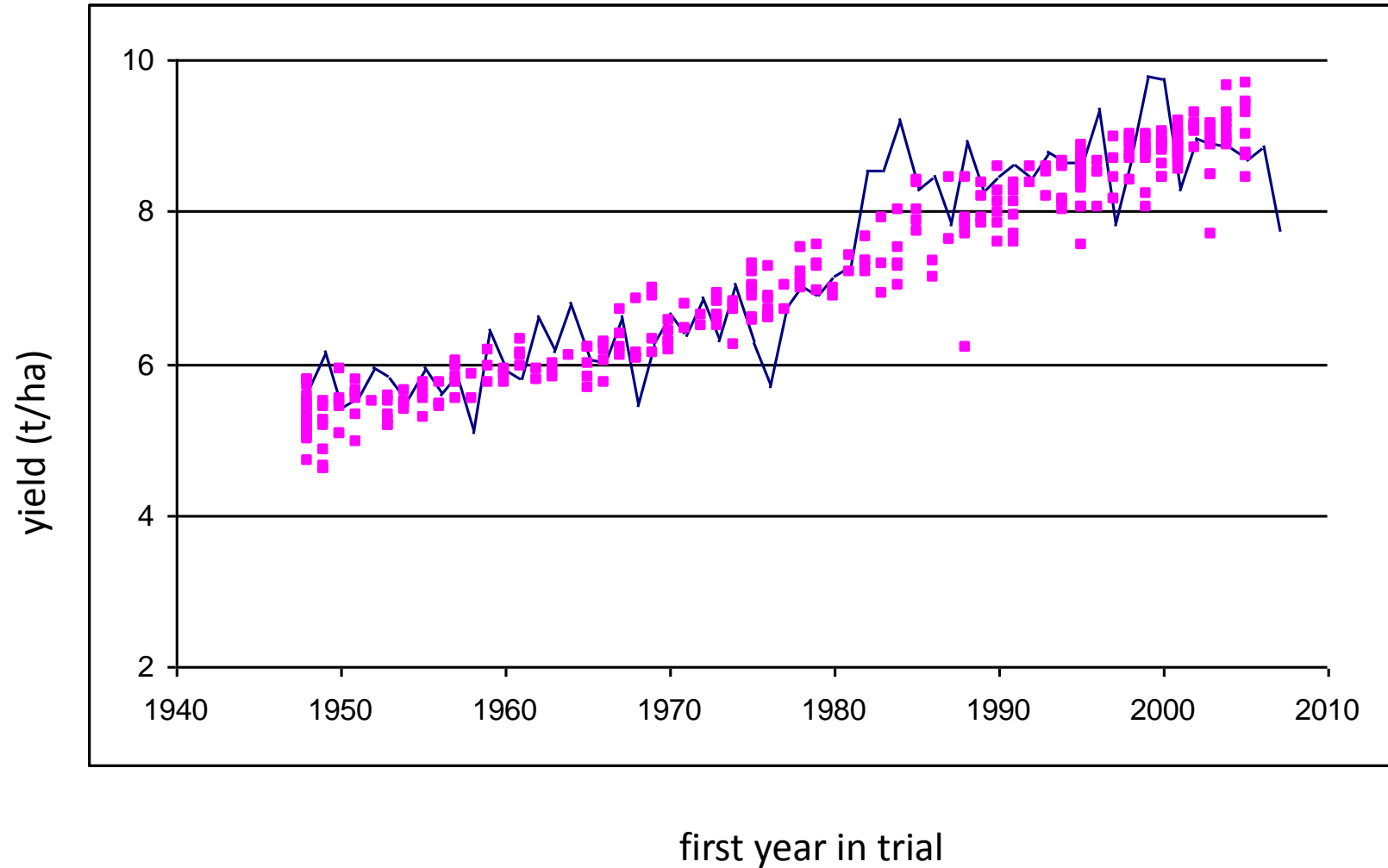Circulant partial diallels

Heterotic groups

Reciprocal recurrent selection

More money

Cereal yields in the UK

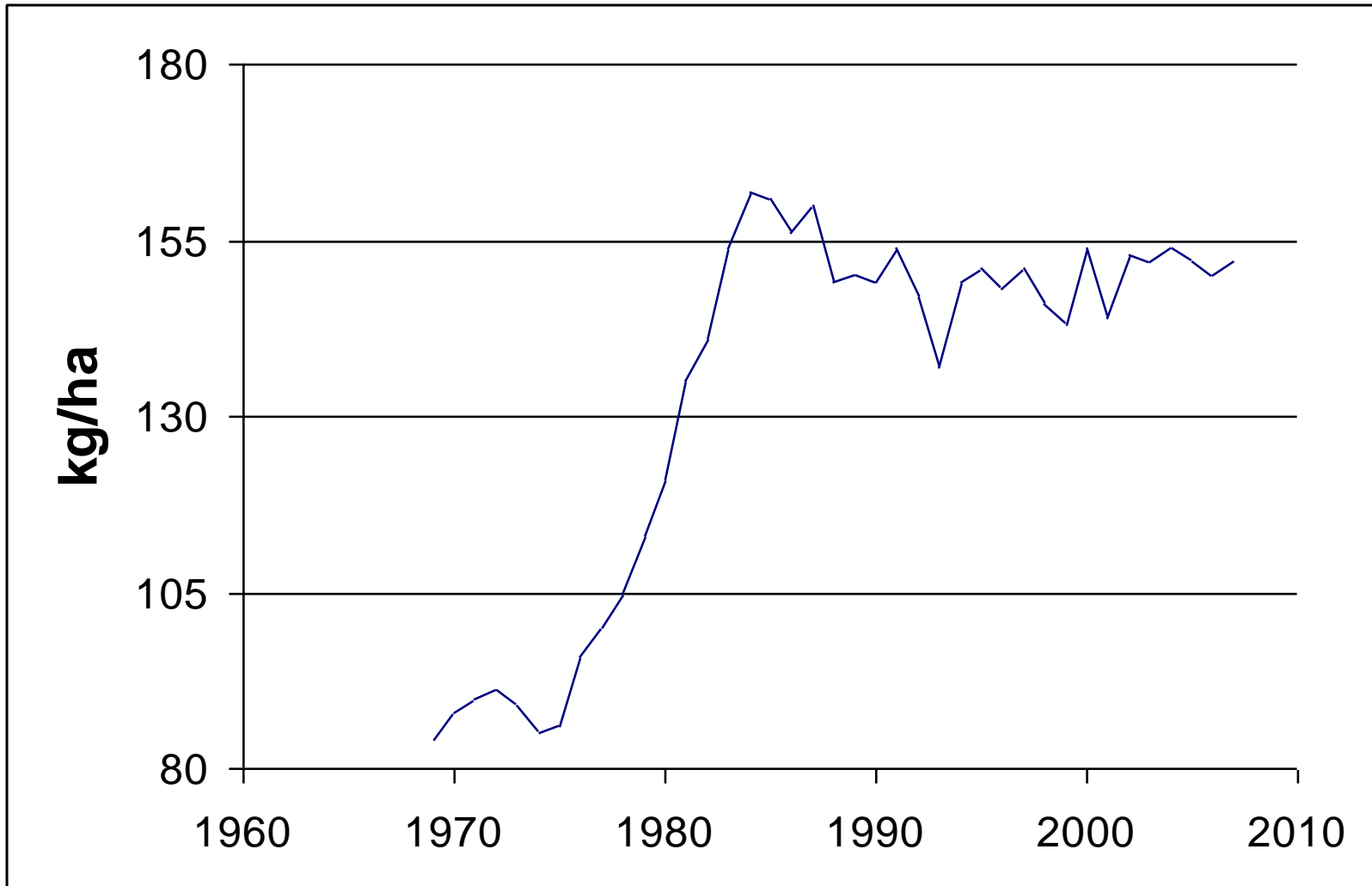winter wheat genetic and environmental trends

yield (t/ha)

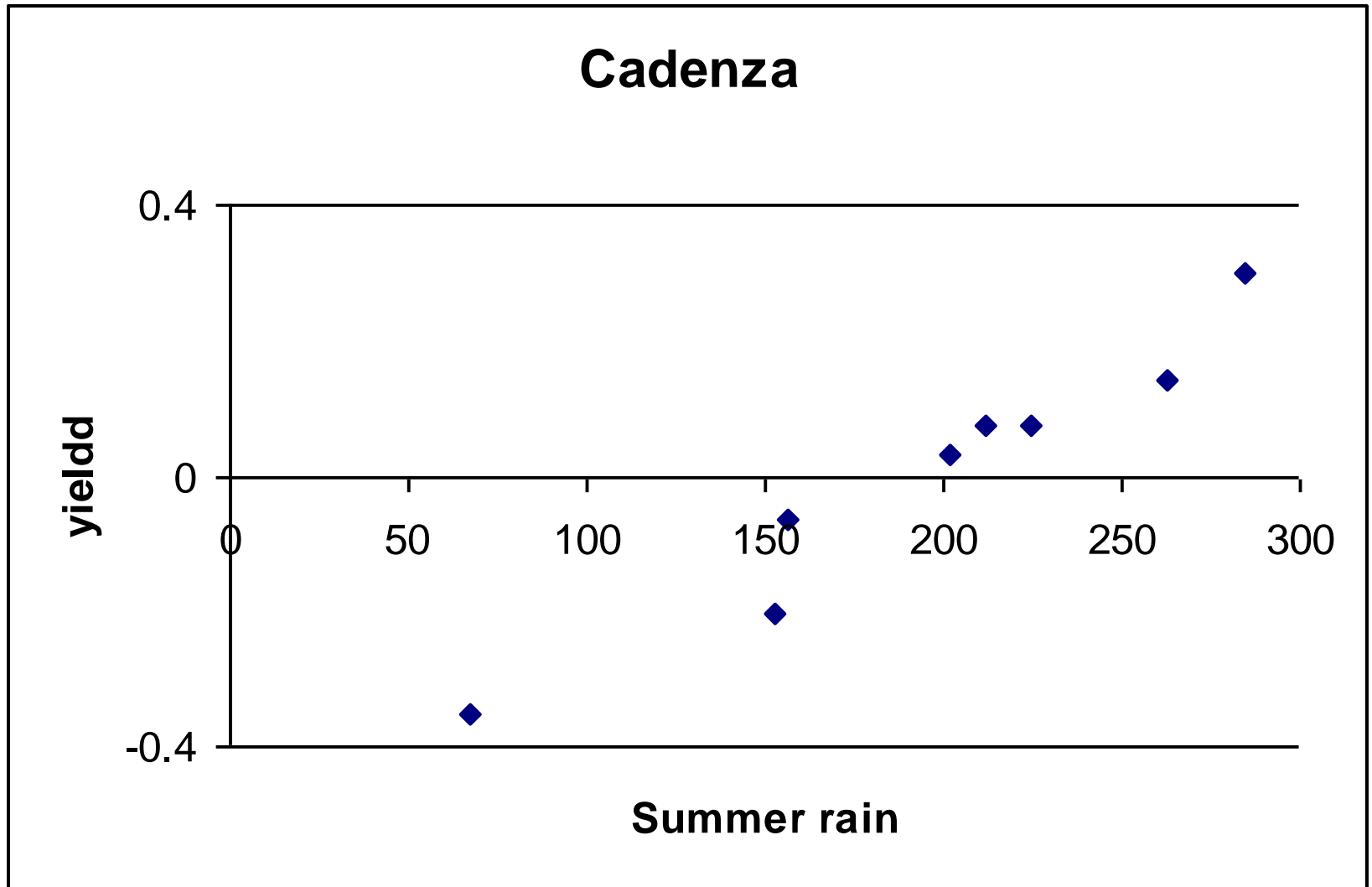first year in trial

Linear trends in yield (t/ha)

1982-2007 NL/RL trials

	varieties	years
winter wheat	0.074	0.010
spring barley	0.060	-0.006
winter barley	0.071	0.010
maize	0.109	0.108
sugar beet	0.105	0.112
oilseed rape	0.064	-0.019

N use for tillage crops: England & Wales

Screen for sensitivity to climatic stress?

Some challenges & questions; a personal view

Have yields stopped rising?

Should we care about GxE?

What proportion of quantitative variation has originated by mutation since domestication: should we sample wild and old germplasm for yield QTL?

Do we get enough recombination?

Why are yield and quality negatively correlated?

Are the days of breeding to exploit natural variation numbered by GM?

What is the best design of a breeding programme to exploit GS?

Monday pm

- Population genetics and linkage disequilibrium

Population Genetics

Books

Felsenstein

http://evolution.genetics.washington.edu/

Weir   Genetic Data Analysis 2nd ed.

http://statgen.ncsu.edu/powermarker/

GH Hardy 1877-1947

"There is no permanent place in the world for ugly mathematics."

"I am reluctant to intrude in a discussion concerning matters of which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists."

Hardy-Weinberg Equilibrium 1908

A sufficient condition for no evolution to occur within a Mendelian population is that mutation, selection, and chance effects are all absent and that mating is at random.

The hereditary mechanism, of itself, does not change allele frequencies. The constancy of genotype frequencies then follows from the presence of random mating.

Population Genetics

The Hardy-Weinberg Law

Nothing changes except for:

mutation
selection
sampling variation (drift)
migration
non-random mating

Population Genetics

The Hardy-Weinberg Law

genotype	AA	Aa	aa
frequency	X	2Y	Z
alleles	all A	½ A, ½ a	all a

Frequency of A gamete $\quad$ $X + \tfrac{1}{2} 2Y$ = p say

Frequency of a gamete $\quad$ $Y + \tfrac{1}{2} 2Y$ = 1-p = q say

with p + q = 1

	female gamete (freq)	
	A (p)	a (q)
male gamete (freq)		
A $\quad$ (p)	AA (p^2)	Aa (pq)
a $\quad$ (q)	Aa (pq)	aa (q^2)

$\rightarrow$

AA	Aa	aa
p^2	2pq	q^2

Frequency A: $\qquad$ $p^2 + \tfrac{1}{2} 2pq = p(p+q) = p$

Polyploids

$$(p_1 A_1 + p_2 A_2 + p_3 A_3 \ldots p_n A_n)^p$$

Eg   Bufo pseudoraddei baturae

Population Genetics

Non-random mating.

	AA	Aa	aa
	$p^2 + pqf$	$2pq(1-f)$	$q^2 + pqf$

Selfing series

generation	AA	Aa	aa
0	p^2	$2pq$	q^2
1	$p^2 + pq/2$	pq	$q^2 + pq/2$
2	$p^2 + pq3/4$	$pq/2$	$q^2 + pq3/4$
3	$p^2 + pq5/8$	$pq/4$	$q^2 + pq5/8$
∞	$p^2 + pq = p$	0	$q^2 + pq = q$

Population Genetics

Mixed selfing and random mating

	AA	Aa	aa
observed	$p^2 + pqf$	$2pq(1-f)$	$q^2 + pqf$

Just as before,   but

$$f = s / (2-s)$$
where s is the proportion of seed set by selfing or

$$f = (1-t) / (1+t)$$
where t is the proportion of seed set by random mating

Population Genetics

Wahlund effect

Subdivided populations have reduced heterozygosity:

Frequency in population 1　　　　=　p_1　= p+x

Frequency in population 2　　　　=　p_2　= p-x

Average heterozygosity　　　　=　　　$(2p_1q_1 + 2p_2q_2) / 2$

　　　　　　　　　　　　　　　=　　　$(p+x)(1-p-x) + (p-x)(1-p+x)$

　　　　　　　　　　　　　　　=　　　$2pq - 2x^2$

Cross pops– observe excess of hets:

　　　　　　　$(p+x)(1-[p-x]) + (1-p-x)(p-x)$
=　　　$2pq + 2x^2$

Explanation　　　for heterotic pools

　　　　　　　and composite varieties

Linkage Disequilibrium

Random mating between individuals generates equilibrium genotype frequencies at a single locus.

(Hardy-Weinberg equilibrium)

Random assortment of chromosomes in meiosis generates equilibrium frequencies between loci.

(Linkage equilibrium)

At equilibrium:

	loc B	r (B)	s (b)
Loc A			
p (A)	pr AB		ps Ab
q (a)	qr aB		qs ab

Rearranging:

AB	Ab	aB	ab
pr	ps	qr	qs

Same in the next generation

With arbitrary frequencies

Loc A	loc B	B	b
A		w	x
a		y	z

Compare observed and expected with χ^2

	AB	Ab	aB	ab
Observed	w	x	y	z
Expected	pr	ps	qr	qs
O – E	+D	-D	-D	+D

D = observed frequency minus expected frequency

AB	Ab	aB	ab
pr	ps	qr	qs
+D	-D	+D	-D

$$D = p(AB) - p(A).p(B)$$

or $\quad -D = p(aB) - p(a).p(B)$

etc.

Some properties of the D

Max value is 0.25, when $p(A)=p(B)=0.5$

At other allele freqs. max. value can be small eg

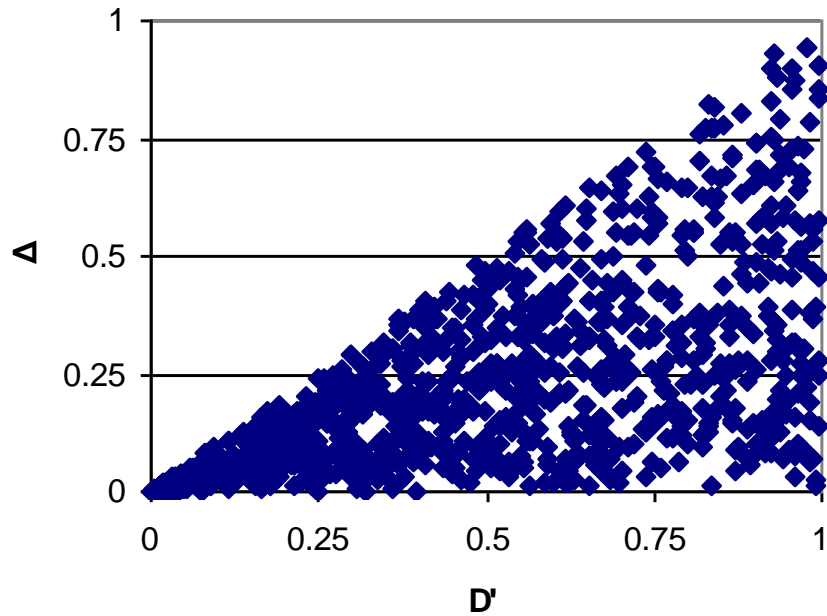$p(A)=p(B)=0.9$ $D_{max} = 0.09$

To make interpretation easier, define:
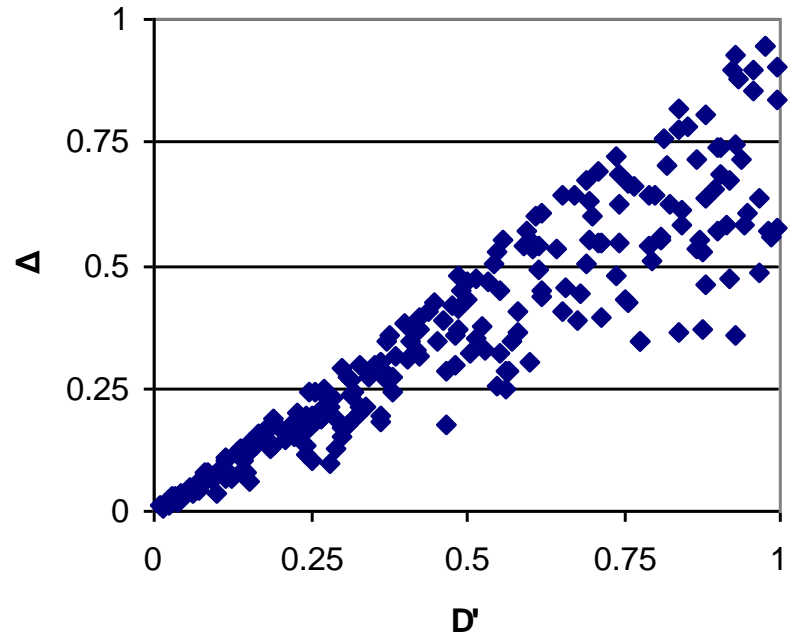
$D' = D / D_{max}$ range 0-1

or

$$\Delta = \frac{D}{\sqrt{p(A)p(a)p(B)p(b)}}$$ range 0-1

Comparison of LD measures

1000 random SNPs

Rare allele >0.25

$\Delta \rightarrow$ 1: allele freqs match, two haplotypes

D' $\rightarrow$ 1: allele freqs don't matter, three haplotypes

LD measures for multiple alleles

Calculate D' or r^2 for each pair of alleles in turn.

Take the average, weighted by the expected frequency ($p_1 p_2$)

Estimates tend to be biased upwards in small samples. The bias can be quite large.

Correct by permutation testing.

The decay of Linkage Disequilibrium

$$D_1 = (1-\theta)\,D_0$$

$$D_t = (1-\theta)^t\,D_0$$

# gens	unlinked	5cM	0.5cM	50k
0	1	1	1	1
1	0.50	0.95	1	1
10	0	0.60	0.95	1
100	0	0.01	0.61	0.95
1000	0	0	0.01	0.61
10000	0	0	0	0.01

Proof

To decay, LD needs recombination. Recombination need double heterozygotes

AB/ab occur at a frequency $2(pr + D)(qs + D)$
Ab/aB ------------ditto-------- $2(ps - D)(qr - D)$

Arbitrarily select gamete type AB to follow over 1 generation:

$P_{(AB)} = 2(pr + D)(qs + D) (1 - \theta) / 2 + 2(ps - D)(qr - D) \theta / 2$

             (non recs from AB/ab)   (recs from Ab/aB)
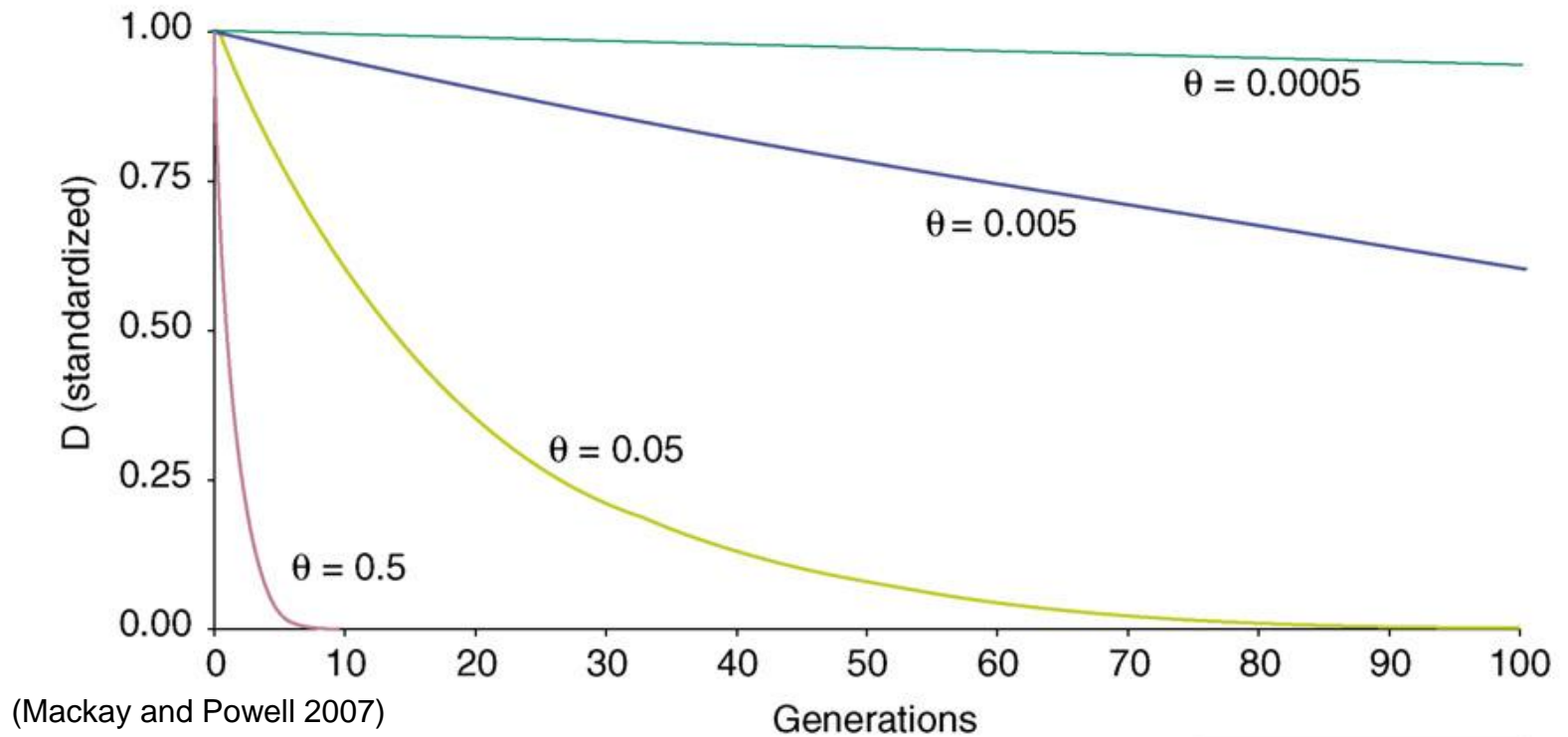
Ignore terms not involving θ to get change in $P_{(AB)}$

$= \quad [- (pr + D)(qs + D) + (ps - D)(qr - D)]\theta \qquad = \qquad -\theta D$

New value of D is therefore $D - \theta D = D(1 - \theta)$
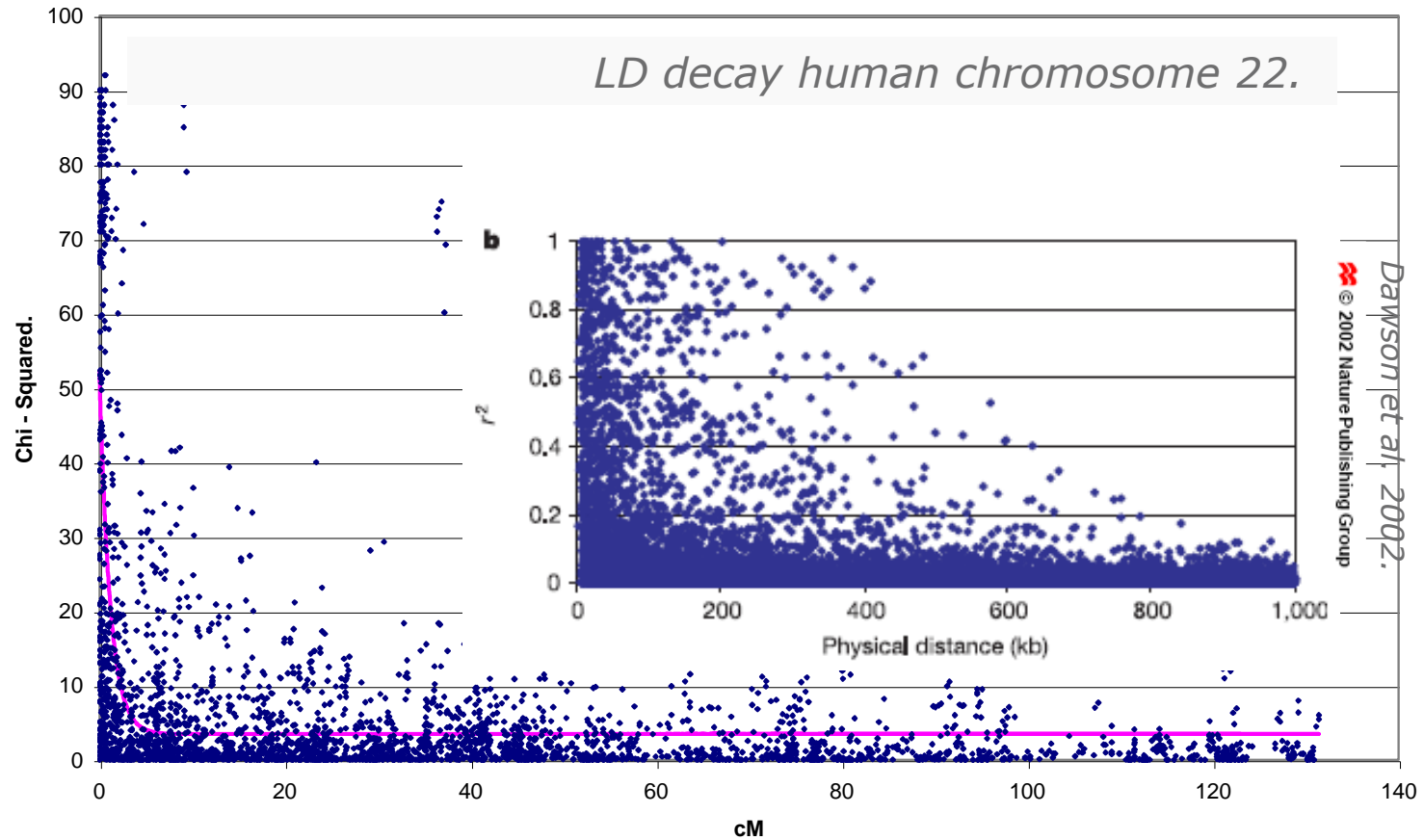
Over t generations: $D_t = D_0 (1 - \theta)^t$

LD decays with time and recombination fraction
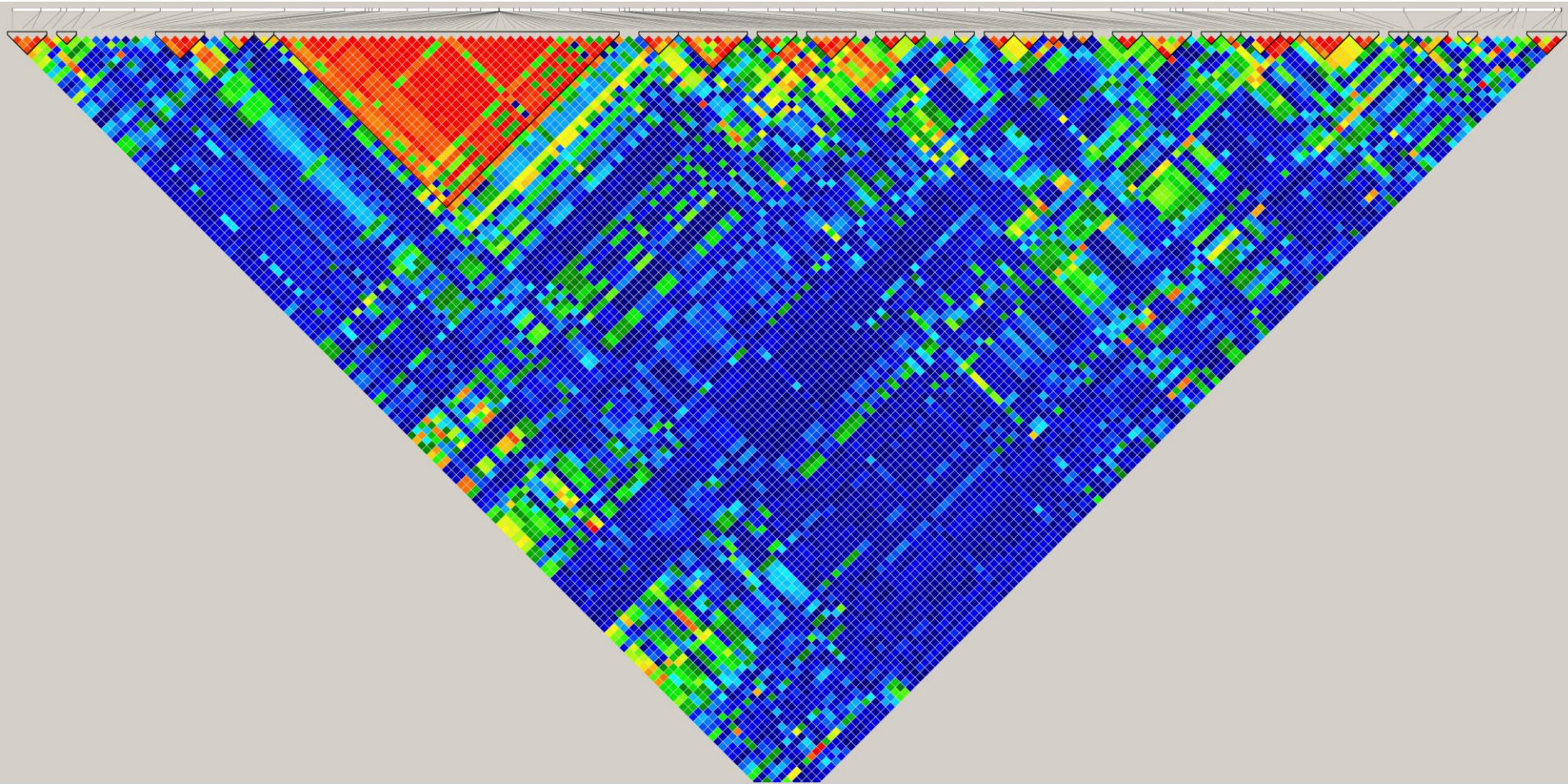

(Mackay and Powell 2007)

TRENDS in Plant Science

Decline in LD with genetic distance

Decline of between marker association over genetic distsnce. UK wheat all genomes.

LD decay human chromosome 22.

LD in Barley varieties

Chromosome 2, Barley, AGUEB SNP data

The Causes of Linkage Disequilibrium

Mutation

Sampling   drift, founder effect

Migration

Selection

Mutation

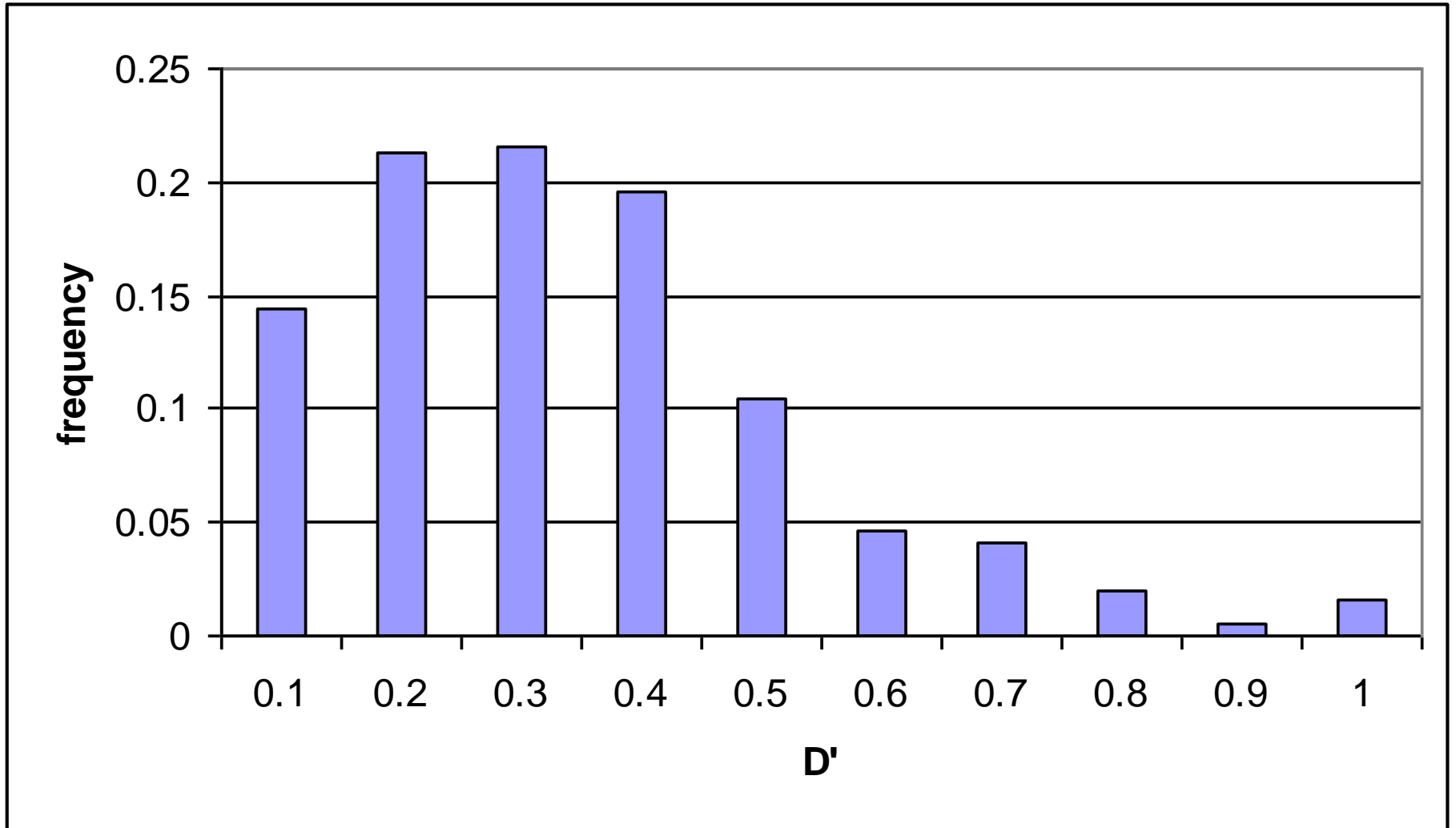| Gen.	Allele freq	D'	Δ
0	1/2N	1	0
x	?	?	?

Although mutation generates LD, this is not very interesting. It is the fate following mutation which is important.

Drift

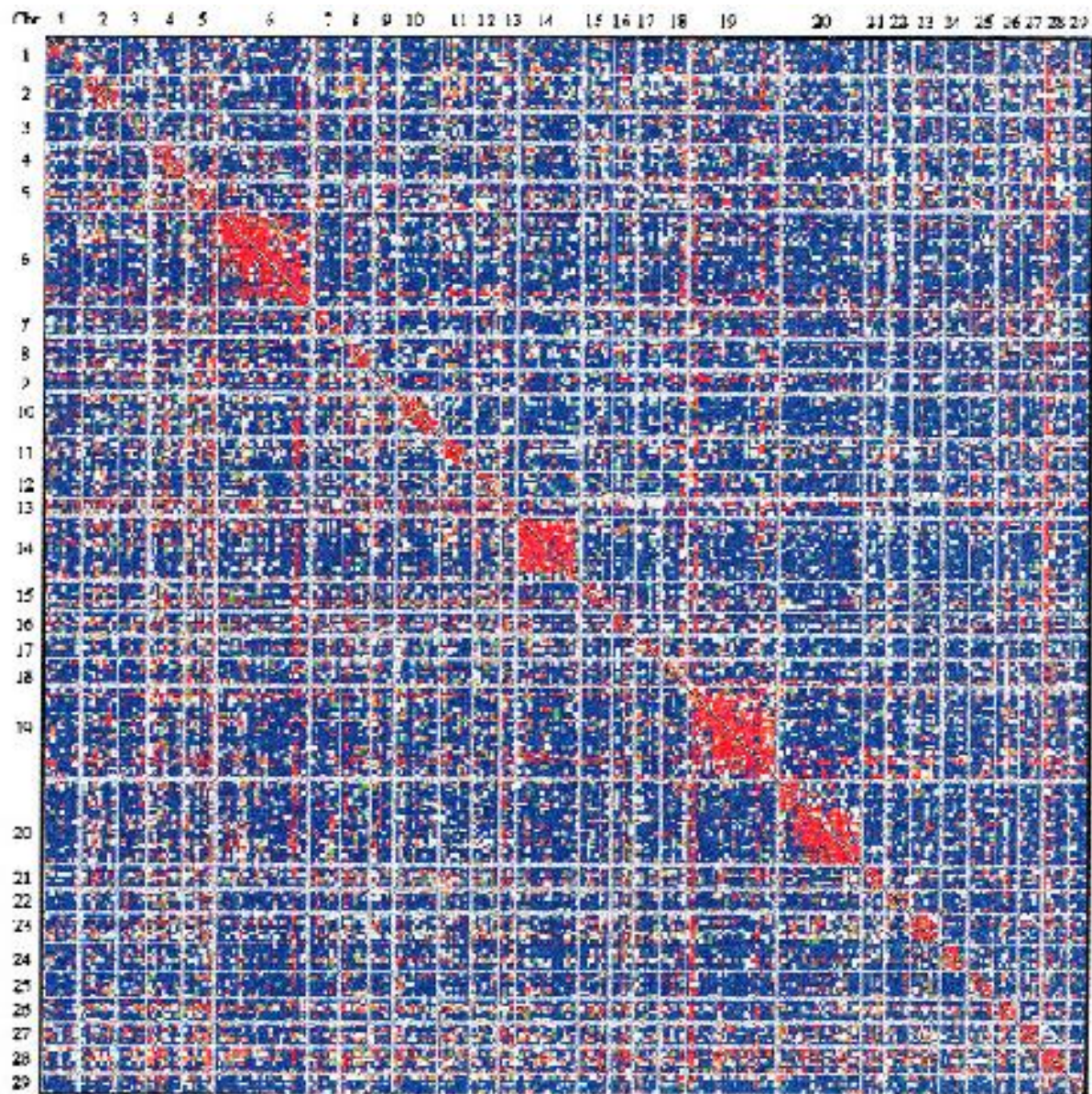$$\varepsilon(\Delta^2) = \frac{1}{1 + 4N_e\theta}$$

On average, as population size and recombination increase, LD falls

Distribution of LD in founder population size 10

Migration

Pop 1 (no LD)   Pop 2 (no LD)

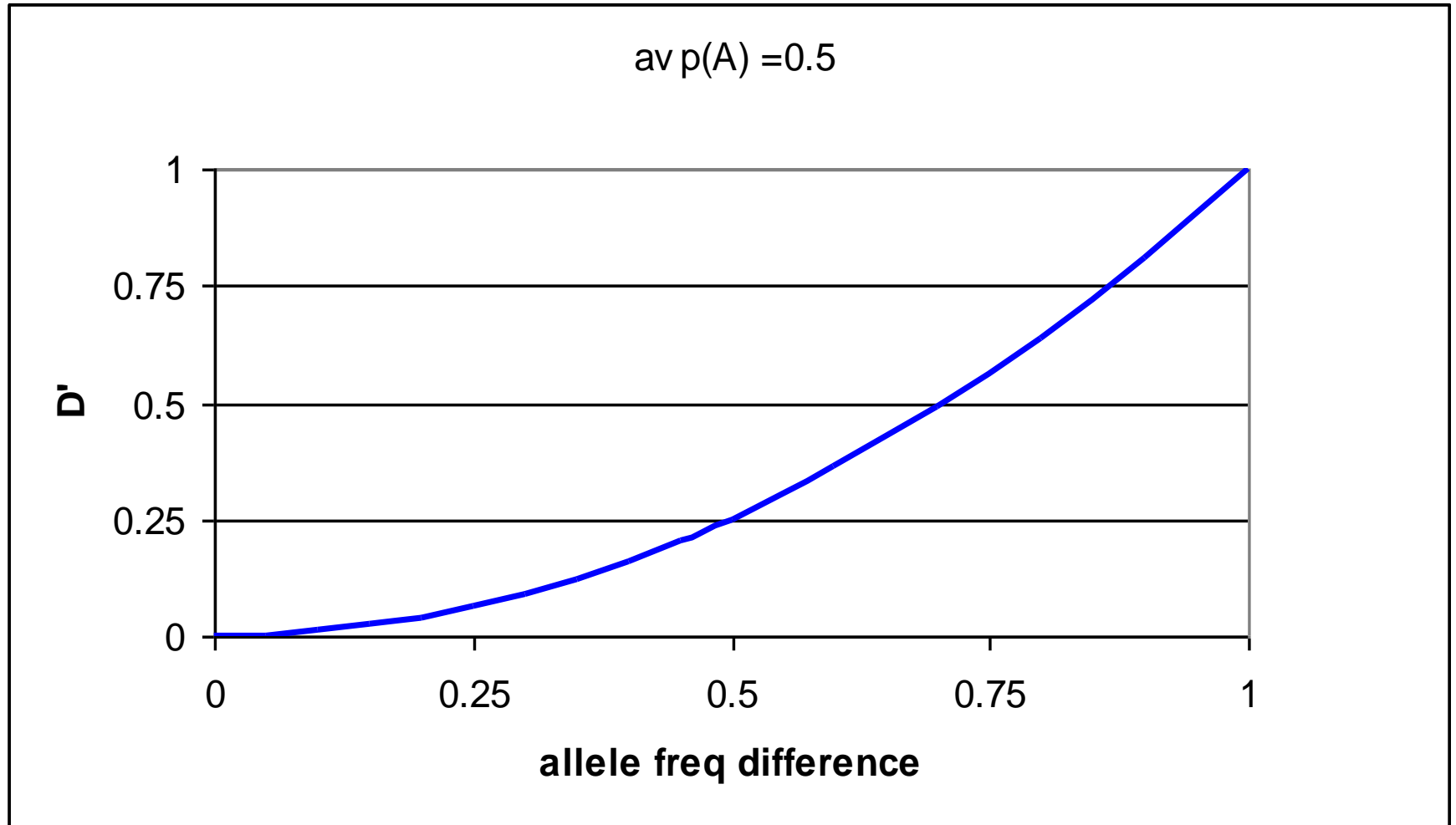p_1r_1 (AB) p_2r_2 (AB)

1:1 mix

What is the freq. of AB

Observe $\frac{1}{2} (p_1r_1 + p_2r_2)$

Expect $\frac{1}{4} (p_1 + p_2)(r_1 + r_2)$

$D = \frac{1}{4} (p_1 - p_2)(r_1 - r_2)$

Zero if $p_1 = p_2$ or $r_1 = r_2$

Migration – population admixture

av p(A) =0.5

D'

allele freq difference

Hitch-hiking

Allele frequencies change at a locus as a result of selection.

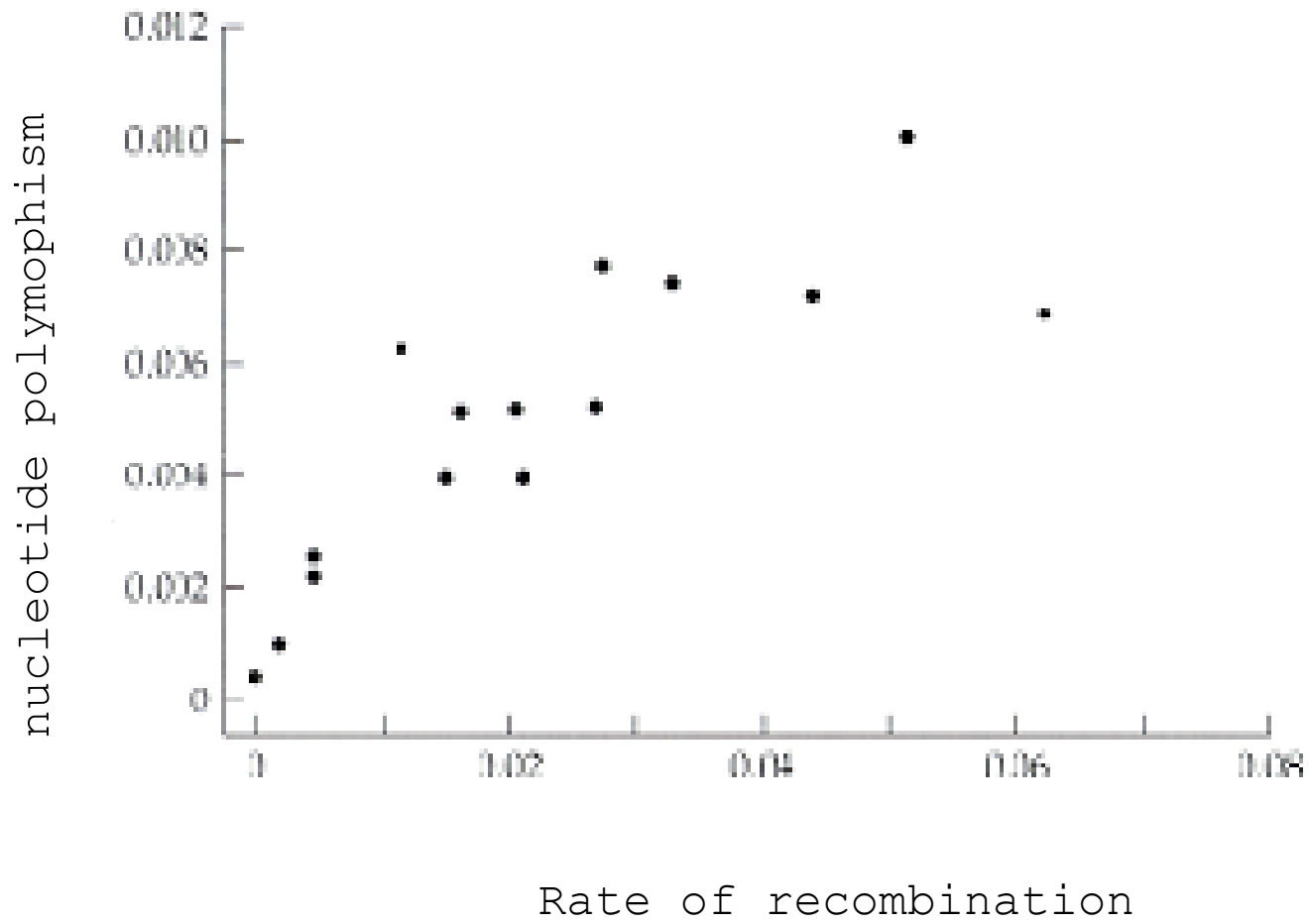As a result, closely linked polymorphisms change in frequency too.

Hitch-hiking generates LD over the whole linked region.

Is important in regions of low recombination.

These are the gene-rich regions – more opportunities for selection.

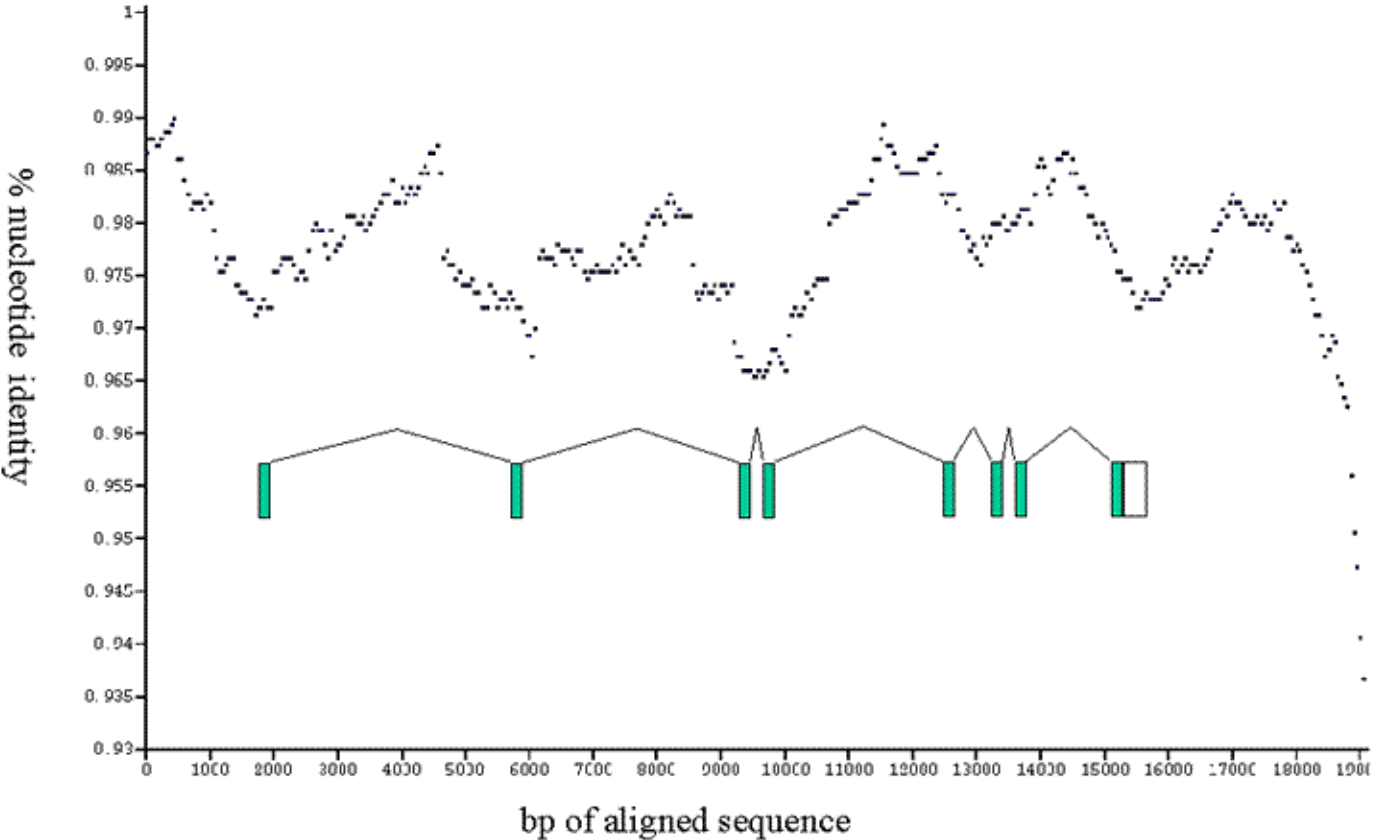# Hitch-hiking: evidence from Drosophila

An example of hitch-hiking in man.

The Morpheus gene family – function unknown – found in a class of segmental duplications.

20x normal rate of amino acid substitution.

Non synonymous substitution rate > synonymous.

Sequence alignment of two human copies of morpheus gene family.

0   16 K bases

So what?

Deleterious SNPs at a high frequency are likely to be of interest.

One way they may rise in frequency is through hitch-hiking.

Therefore – look for footprints of hitch-hiking:

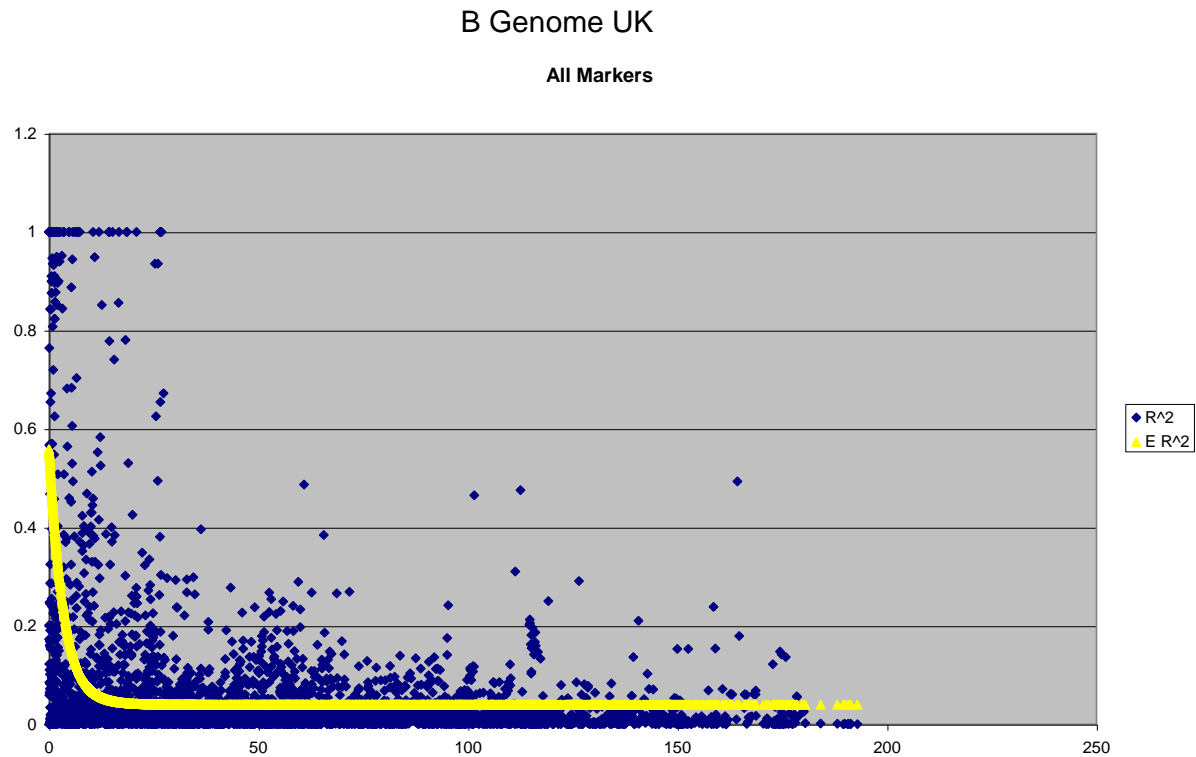   High LD / low recombination / gene rich regions
      Lower heterozygosity and freq. of neutral SNPs
      Higher heterozygosity and freq. of nsSNPs

Plotting and Modelling LD

$$E(\Delta^2) = 1/(1+4N_e\theta)$$

$$E(D') = L+(H-L)(1-\theta)^t$$

B Genome UK

All Markers

Haplotypes

Methods of determining phase:

        is AaBB:

                <u>AB,</u> <u>ab</u>

   or

                <u>Ab</u>, <u>aB</u>

Pedigree	CEPH families
Sequencing	short range
Clarke Algorithm	easy to understand
EM	much software - snphap
Evolutionary methods	Phase