GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

UNPLAB: Understand Plant Breeding - An Open-Access Series of Lecture Slides for Self-Study of Plant Breeding Principles and Methods

Wolfgang Link, Division of Plant Breeding Methodology, Carl-Sprengel, 1 University of Göttingen, Germany

With gratitude to my teachers - HCB, FWS, HHG, HFU, WE, TB.

UNPLAB (Understand Plant Breeding) aims to support you in deeply understanding the fundamental concepts of Plant Breeding.
UNPLAB is an open-access series of lecture slides designed as a self-study resource for students, breeders, lecturers, and researchers passionate about Plant Breeding.
UNPLAB emphasises conceptual understanding, the ability to recognise connection across different fields and to apply insight and knowledge to practical breeding situations.
In the era of AI, deep and independent understanding, including the ability to identify pitfalls and fallacies, will become even more important - for the individual breeder or researcher and for our shared field of Plant Breeding. UNPLAB is freely available online for self-directed learning and teaching purposes.

Overview of the UNPLAB chapters. See Serial Number and Chapters' names.

Serial Number #	Name of Chapter
#1	UNPLAB-Intro_Ch-1
#2	UNPLAB-PopGen_Ch-1[HWE I]
#3	UNPLAB-PopGen_Ch-2[HWE II]
#4	UNPLAB-PopGen_Ch-3[Inbreeding]
#5	UNPLAB-PopGen_Ch-4[Drift I]
#6	UNPLAB-PopGen_Ch-5[Drift II]
#7	UNPLAB-PopGen_Ch-6[Selecton]
#8	UNPLAB-PopGen_Ch-7[Δq]
#9	UNPLAB-PopGen_Ch-8[Equil]
#10	UNPLAB-PopGen_Ch-9[LD I]
#11	UNPLAB-PopGen_Ch-10[LD II]
#12	UNPLAB-PopGen_Ch-11[LD III]
#13	UNPLAB-h²_GxE_Ch-1[Broad Sense h²]
#14	UNPLAB-h²_GxE_Ch-2[Stabil]
#15	UNPLAB-h²_GxE_Ch-3[BLUP]
#16	UNPLAB-h²_GxE_Ch-4[Sel.Intensity]
#17	UNPLAB-h²_GxE_Ch-5[BreedersEquation]
#18	UNPLAB-h²_GxE_Ch-6[QTL-Detection]

Serial Number #	Name of Chapter
#19	UNPLAB-BrMeth_Ch-1[Intro]
#20	UNPLAB-BrMeth_Ch-2[Heterosis]
#21	UNPLAB-BrMeth_Ch-3[GCA I]
#22	UNPLAB-BrMeth_Ch-4[GCA II]
#23	UNPLAB-BrMeth_Ch-5[RecSel I]
#24	UNPLAB-BrMeth_Ch-6[RecSel II]
#25	UNPLAB-BrMeth_Ch-7[RecSel III]
#26	UNPLAB-BrMeth_Ch-8[TheFour I]
#27	UNPLAB-BrMeth_Ch-9[TheFour II]
#28	UNPLAB-BrMeth_Ch-10[TheFour III]
#29	UNPLAB-BrMeth_Ch-11[TheFour GS]
#30	UNPLAB-BrMeth_Ch-12[TheFour IV]
#31	UNPLAB-BrMeth_Ch-13[TheFour V]
#32	UNPLAB-CorrTraits_Ch-1[IndirSel]
#33	UNPLAB-CorrTraits_Ch-2[GenCorr]
#34	UNPLAB-CorrTraits_Ch-3[IndexSel]

Serial Number #	Name of Chapter
#35	UNPLAB-QuGen_Ch-1[Teaser]
#36	UNPLAB-QuGen_Ch-2[Transition I]
#37	UNPLAB-QuGen_Ch-3[Transition II]
#38	UNPLAB-QuGen_Ch-4[MetrMod]
#39	UNPLAB-QuGen_Ch-5[GenMeanAnal]
#40	UNPLAB-QuGen_Ch-6[StatMod I]
#41	UNPLAB-QuGen_Ch-7[StatMod II]
#42	UNPLAB-QuGen_Ch-8[Decomp-GenV I]
#43	UNPLAB-QuGen_Ch-9[Decomp-GenV II]
#44	UNPLAB-QuGen_Ch-10[GenV 0>F>1]
#45	UNPLAB-QuGen_Ch-11[QG Lines I]
#46	UNPLAB-QuGen_Ch-12[QG Lines II]
#47	UNPLAB-QuGen_Ch-13[GenCov I]
#48	UNPLAB-QuGen_Ch-14[GenCov II]

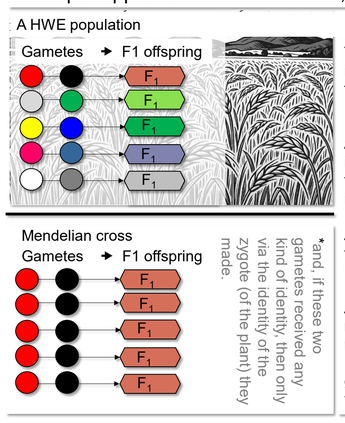
Let us get some orientation.
When producing seeds or tubers (...) to sell this to farmers for sowing or planting, then **FOUR REPRODUCTIVE SYSTEMS** exist .
These specify the **FOUR CATEGORIES OF PLANT BREEDING**.

- ► Self-fertilization
- ► Panmicitic outcrossing
- ► Vegetative propagation
- ► Controlled crossing between seed parent and pollen parent
- ► **LINE** Breeding
- ► **POPULATION** Breeding
- ► **CLONE** Breeding
- ► **HYBRID** Breeding

Four examples of the delightful luxury of seeing things clearly :-)
(1) **Heterozygosity and heterosis:** Increasing the inbreeding of the parents does **not** increase the heterozygosity of their F1-hybrid (nor its heterosis). What matters is the **genetic difference between the parents**. Simple, but oftentimes misunderstood.
(2) **Heritability and environmental control.** Increasing heritability by standardising environmental conditions only makes sense if the combine harvester will eventually harvest the farmer's crop under these very same standardised conditions ;-)
(3) Is **better-parent heterosis** really an **alternative** to **mid-parent heterosis?** If the question is how a hybrid performs against the best available material, why compare it only with its two parents? Look at the **best material in your own and your competitors' nurseries**. If the question is a genetic one, then use **mid-parent** heterosis.
(4) Are **mutation breeding**, **heterosis breeding**, or **genomic prediction breeding** (and so on) meaningful terms? Not really. These are neither breeding categories nor breeding methods. See: **Novel diversity** may be induced in all four categories; **heterosis** is relevant in three of the four; **genomic prediction** can be employed in all four. You know the **four** categories — and "bliblablu breeding" is none of them ;-)

Below are examples of UNPLAB 'pages'. The Serial Numbers of the chapter they belong to, as well as the page number therein, are indicated.

A deeper appreciation of what this is, a HWE-population. Read until you like it!



The HWE population is a mixture of non-inbred (name-them F1) individuals, each of them is a unique individual-genotype (if you just for instance self-fertilize any of them, you get a very normal Mendelian F2 generation).

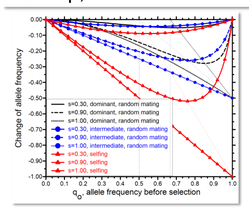
A Mendelian F1 is a generation with one-or-several-or-many individuals, all with same genotype. This F1 genotype is just one of the endless many non-inbred individual-genotypes that one can find in the HWE Reference Population.

A Mendelian F1 can be produced and be re-produced, because with the parents, we have **immortal gamete-re-producers**. A Mendelian F1 is called 'immortal' F1.

An individual of the HWE population (name it 'mortal' F1) can not be reproduced and not be produced at will, because the female and male gametes from which it came vanished in the mating*.
… it is even not easy to see, in the HWE-'mortal'-F1, exactly which genes and haplotypes came actually from its mother and from its father ;-)

#3, page 17

Sum up, rehearsal of what we have seen …



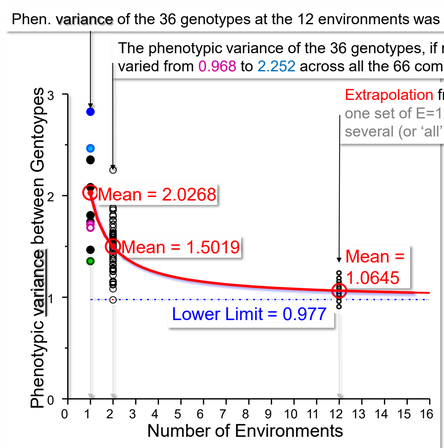
Maximize the result of one step of mass selection! To do so, choose your crop species' reproductive mode. Choose the trait to work with, according to predominant gene action. What is your choice then? Rather choose maize (outcrossing crop) or barley (selfing)? Rather choose grain yield (more dominant-recessive gene action) or protein digestibility (more intermediate gene action)? Rather favour a mono- or a oligo – or a polygenic trait? What should be the typical allele frequency at the loci of such trait ... to allow large gain from one step of selection?

The selection coefficient s should be large, this is more probably true in mono- or oligogenic traits than in polygenic traits.
Δq should be large, this is more probable in self-fertilizing crops than in cross-fertilizing crops.
Δq should be large. If ‚random-mating' and ‚a' is recessive, then Δq is larger at quite high frequencies of ‚a'; whereas if ‚a' is intermediate, Δq will be larger at somewhat lower frequencies // even better if the ‚bad' allele was dominant ;-)

If s per locus is very small, and if we have random mating and dominance of good alleles: Then allele frequencies of the unfavourable alleles should be at p=⅔. If gene action is intermediate (random mating) or if we have anyway self-fertilization, then allele frequency of ‚a' should be p=½.

#8, page 15

Phen. variance of the 36 genotypes at the 12 environments was small (σ^2_P =1.355, E12) or high (σ^2_P =2.826, E7).



The phenotypic variance of the 36 genotypes, if means across two envs. were used, was smaller, and varied from 0.968 to 2.252 across all the 66 combinations of 2 environments. The mean was 1.5019

Extrapolation from E=2-ANOVAs to E=12 and beyond. With only one set of E=12 at hand, you can not calculate variation between several (or 'all') E=12 sets. I did it again as ... extrapolation ;-)

From this visualization you may see that and how ANOVA is ‚just' accepting the size and pattern of variation between genotypes and due to their GxE interactions as deduced from the actual data set and then extrapolates ‚as if there were no more tomorrow'; unreserved, unconcerned, hurray-forward!

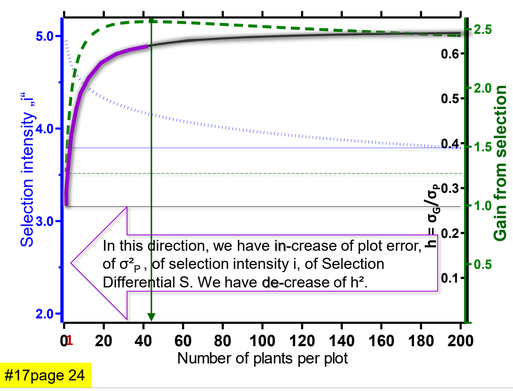
There is simply no other possibility as long as you don't know the true G values; you just should be aware of what happens; with a small number of environments, you need to be correspondingly cautious in employing the results.

With a fair number of envs., hence with high h², your phenotypic data well predicts the true h² and the G data and values.

#13, page 41

σ^2_G =1.00; σ^2_{GL} =0.3797; σ^2_{GY} =0.2132; σ^2_{GLY} =0.5178; σ^2_{GRLY} =0.2263; σ^2_{GPLYR} =12.4574

incl. Wang, Y., 2021



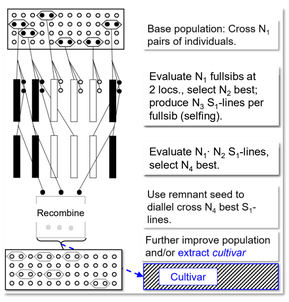
Single plot error is the larger, the less many plants per plot (the smaller each plot is).
Here, h² can not be larger than h²=0.4279 due to testing at only L=1, Y=1, R=1 (hence: h ≤ 0.6541; *cf.* UNPLAB-h²_GxE_Ch3[BLUP], p.10-12).

N= 5.000=> 200 p/plot => h=0.6541
N= 25.000=> 40 p/plot => h=0.6145
N=1.000.000=> 1 p/plot => h=0.2600
N=1.000.000=> 1 p/plot => h²=0.0676

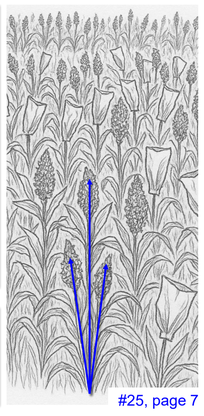
With less than about 40 plants per plot, h² decreases fast, dramatically.

#17page 24

Combined Fullsib (FS) and S1-line recurrent selection (Schipprack, 1993; https://s.gwdg.de/bzK9MF).

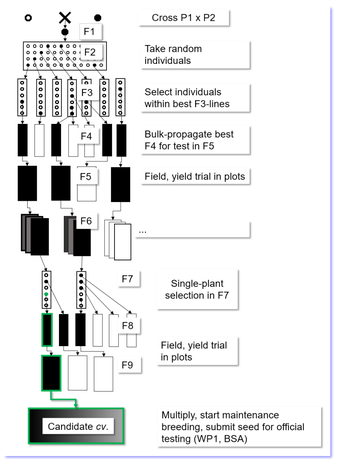


Choose at random (or with selection) pairs of plants for bagging-based crossing (technique depends: natural outcrossing; hot water technique, plastic bag technique for manual crossing, CMS, gametozide, MS). Test FS in plots at 2 locations and, from spare seed, self-fertilize 5 random individuals per FS. Select 45 (≙16%) best FS, test the 5 x 45 = 135 S1-lines that represent them at 3 locations. Select 35 (≙26%) best S1-lines. Recombine them from remnant seed (that was not sown for testing them).
How many individuals represent one S1-line when recombining (producing crossed seed for next season of new pair-crosses)?

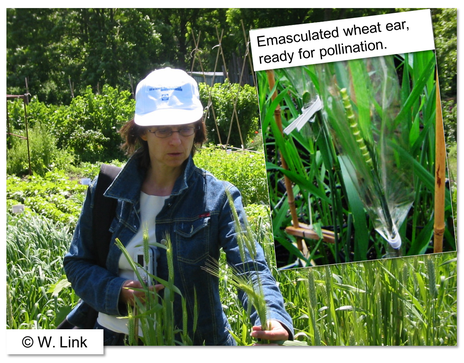


This method (one cycle = 4 seasons) was optimized with due attention to genetic parameters (variances, covariances) of a specific *Sorghum bicolor* germplasm and with yield as sole trait, and with data on cost for labour, for field, for lab etc.
Result for one of various modelled options:
N1=280; N2=45; N3=5; N4=35. Results for 'your' breeding will be different. Alas.

#25, page 7



An now, which method is this? Breeders breed – they do not stick to the textbook text ;-)

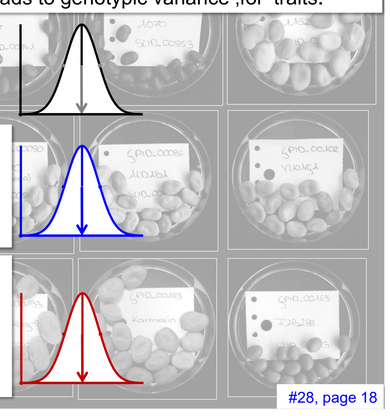


© W. Link

#27, page 9

En passant we detect here two further scales of genetic diversity and segregation.
Further: Beyond the ‚usual' segregation that leads to genotypic variance ‚for' traits.

Genotypes' segregation such as AA, Aa, aa; with genotype rations of 1 : 2 : 1.
Genetic variance: $\sigma^2_G=\sigma^2_A+\sigma^2_D$
Phenotypes' segregation such as resistant vs. susceptible: $\sigma^2_P=\sigma^2_G+\sigma^2_{GE}$



Genotypes in F2 fall in homozygosity levels.
- Fully homozygous: AABB; Aabb; aaBB; aabb
- Partly homozygous: AABb; AaBB
- Zero homozygous, i.e. fully heterozygous: AaBb

F2-Variance for level of homozygosity: $\sigma^2_H=1/(4N)$
(N=number of independently segregating genomic segments)

Genotypes differ for their share of genes inherited from P1
- All genes inherited from P1: AABB
- Most genes inherited from P1: AABb; AaBB
- Half of genes inherited from P1: Aabb; aaBB; AaBb
  Few genes inherited from P1: Aabb; aaBb
- No genes from P1: aabb

F2-Var. for genetic dose inherited from P1: $\sigma^2_{P_1} = 1/(8N)$

#28, page 18

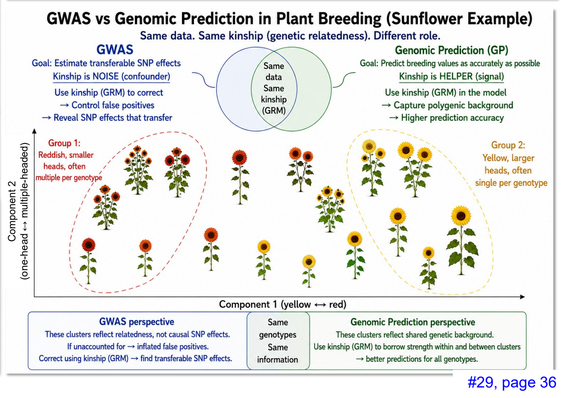
$y = X\beta + Zu + e$

y = Phenotypic value of trait
X = SNP genotype matrix
Xβ = SNP effects (fixed)
Z = Incidence matrix; links genotypes to Kinship Matrix
Zu = Polygenic background from Kinship (random)
e = residual

X, Z: Design-matrices
β; u: vectors of effects that have to be estimated (training) before employed (prediction)
GWAS: Zu = Noise || GP: Zu = Signal

In 'ideal' setting, marker scores from GWAS from all SNP (regardless of significance) would show how much of the prediction ability of a GP model is from transferable marker effects vs. how much is *via* exploitation of Kinship (Genomic Relationship Matrix GRM).

Is exploitation of pedigree (or of GRM) rather valuable aid ... or is it more like doping?



#29, page 36

Exercise, p(A1)=10% now! Schön, C.C. et al., 1993. Heredity 70, 648-659

#41, page 7

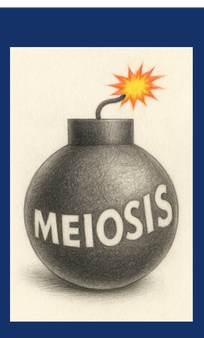
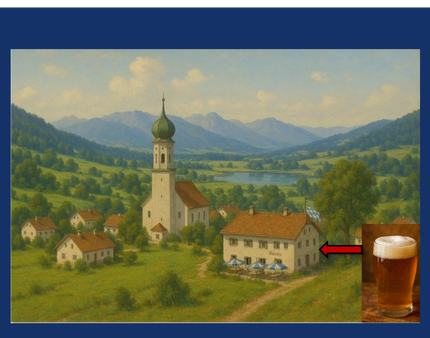
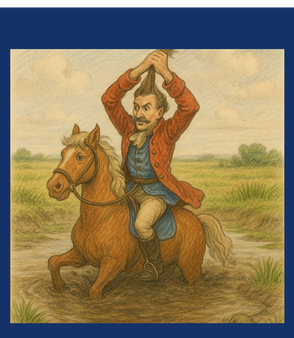
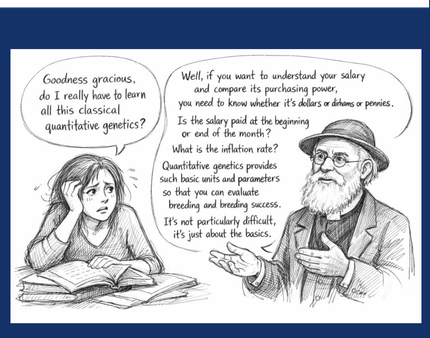
Putative QTLs for resistance against European corn borer in maize in the cross B73 x B52 (C.C. Schön. 1993). Here, the reference population has, as allele frequency, p(X1)=0.1; with X1 being the 'good' allele (shortening tunnels) at each of the 7 QTL.

			Metric model*			Statistical model					
Chrom.	# QLT	R²	a	d	μ	α_1	α_2	δ_{11}	δ_{12}	δ_{22}
1 short arm	1	3.4	1.7	-0.8	1.22	-2.11	0.23	1.30	-0.14	0.02
1 long arm	2	15.7	5.7	-2.2	4.16	-6.71	0.75	3.56	-0.40	0.04
2 short arm	3	3.9	2.2	+2.1	2.14	-0.47	0.05	-3.40	0.38	-0.04
2 long arm	4	13.5	3.4	-2.3	2.31	-4.72	0.52	3.73	-0.41	0.05
3 long arm	5	5.7	2.7	-3.8	1.48	-5.17	0.57	6.16	-0.68	0.08
7 long arm	6	3.7	2.4	+0.3	1.97	-1.94	0.22	-0.49	0.05	-0.01
10 long arm	7	4.9	2.7	-2.96	1.64	-4.52	0.50	4.70	-0.52	0.06
Sum	Σ	50.8	20.8	-9.6	65.81	-25.63	2.85	15.55	-1.73	0.19
Modelled	Σ	38.0								

*c=50.9cm

What is the Breeding Value **A** of the best-possible homozygous genotype ... which his:
A1A1/B1B2/C1C1/D1D1/…
E1E1/F1F1/G1G1
As value, it is
BV = A = $\mu + \sum_{l=1}^{k=7} 2 \cdot \alpha_1^l$
BV = A = 65.81cm - 2· 25.63cm = 14.55cm
As effect: 2· 25.63cm = -51.26cm.

The Breeding Value **A** of the best inbred line is quite large (absolute value: 51.26cm). Here! ... beneficial alleles are rare (10% at each locus) and anyway they are more or less dominant – hence, the seven α_1 are quite large.

The naïve sum of R² of the 7 QTL was $R^2_{naïve}$ =50.8%. The "% phenotypic variance 'explained' ", R², as resulting from fitting all 7 QTLs simultaneously, $R^2_{simult.}$ =38.0%. Ponder a bit. The estimate of a QTL's metric additive effect comes from phenotype-comparing individuals being A1A1 at that locus with those being A2A2. Half that difference estimates a of that locus. Having marker-analysed 300 F2-plants, how many may be A1A1, how many A2A2?

Some of UNPLAB's drawings and cartoons.
- Breeders are gamblers.
- Meiosis usually destroys genotypes.
- Church tower as marker.
- Variance within entries can not be exploited.
- Naïve circular thinking.