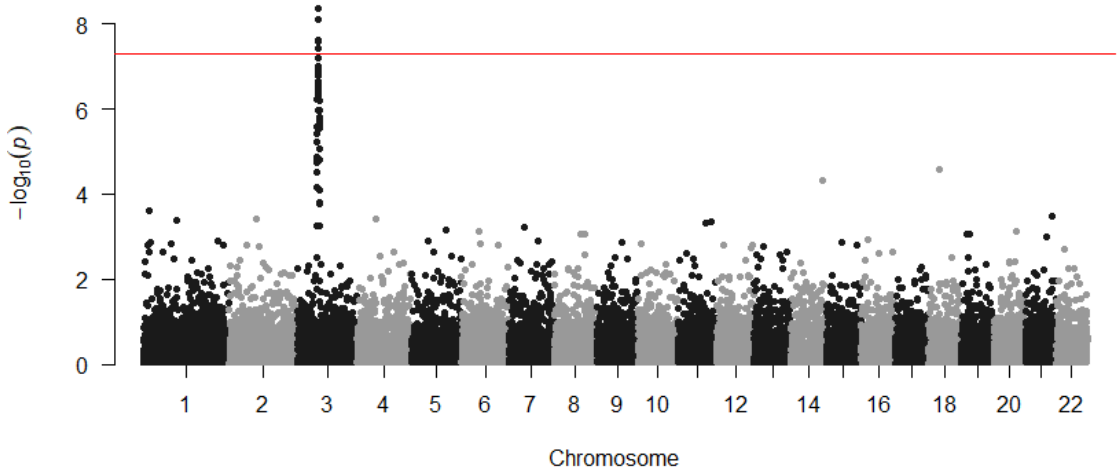
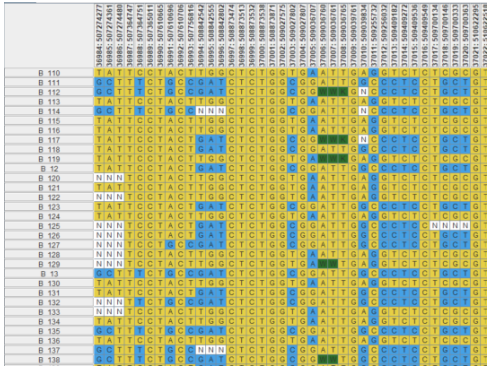
INSTITUTE OF PLANT SCIENCES

Sant'Anna
School of Advanced Studies – Pisa

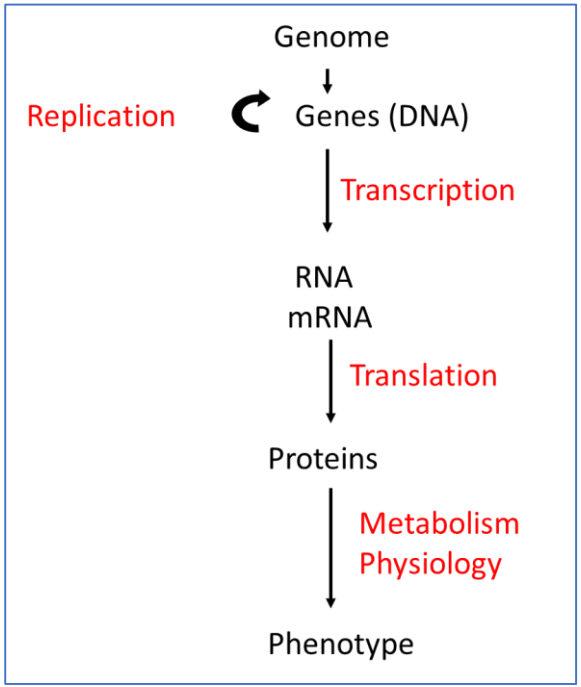
Map Genotype-Trait Associations in Plant Genetic Resources


8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
Chromosome

Some important concepts



Genomics provides the means to characterize the allelic diversity in large collections of plant genetic resources



The central dogma (Crick 1958):
information flows from DNA to phenotypes

Phenomics allows to measure a number of traits in different genetic backgrounds and environments

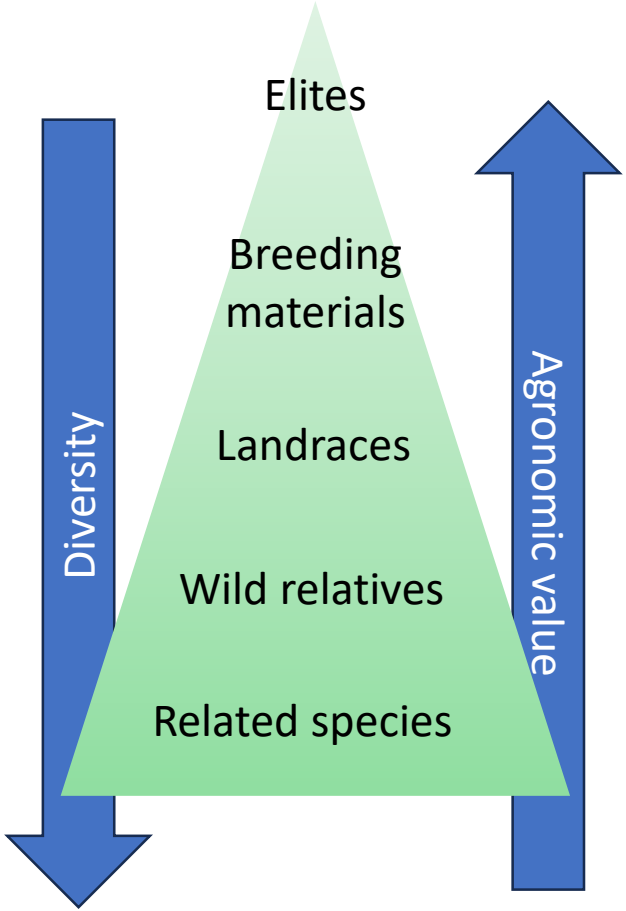
To map genotype-trait associations means to **detect** , **locate**, and **assess the importance** of «genetic factors» affecting traits

Mapping genotype-trait associations in plant genetic resources unlocks key information for **breeding**:

1. Identify new genetic factors/alleles that can be used to improve traits of agronomic performance and adaptation
2. Predict the genomic potential for any given individual with regards to a trait of interest
3. Understand the genetic basis of traits of interest

The value of diversity contributed by plant genetic resources

- Depending on the type of allele pool, there are different types and amounts of diversity available (and hence a different potential in association studies)



Agronomy

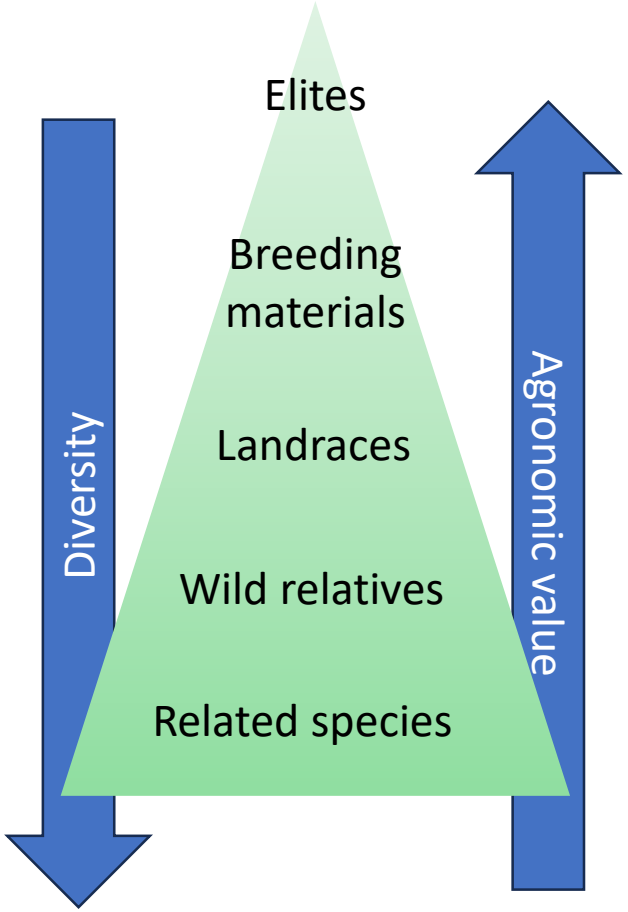
- Yield and yield component traits (e.g. spike size)
- Developemental traits (e.g. flowering time)
- Quality traits (e.g. micronutrients)
- Market traits (e.g. colour, taste)

Adaptation

- Resistance to abiotic stress (e.g. drought)
- Resistance to biotic stress (e.g. disease)

Future-proofing

- What is neeeded for adataptation to future climates (e.g. frequency and intensity of extreme events)

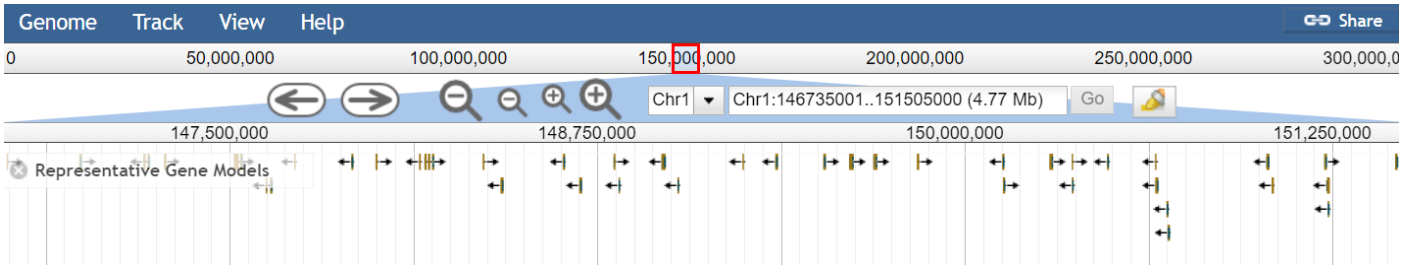


Our working hypothesis: there are one or more «genetic factors» somewhere on the genome affecting a trait of interest

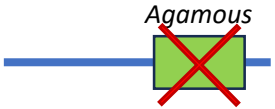
Gene X

We already know it's not an easy job:

- Most interesting traits are controlled by multiple genetic factors
- Most plant genomes are complex; loci may interact
- It is not really like finding a needle in a haystack; it is **finding a needle in pile of needles**



Reverse genetics



Gene(s)

What trait arises from the perturbation of a DNA sequence?

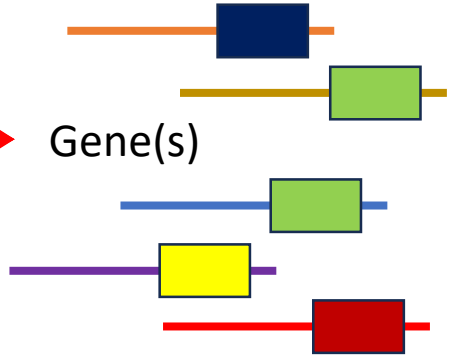
Trait(s)

Forward genetics

Trait(s)

Is variation of a trait associated with genotypic variation?

Gene(s)

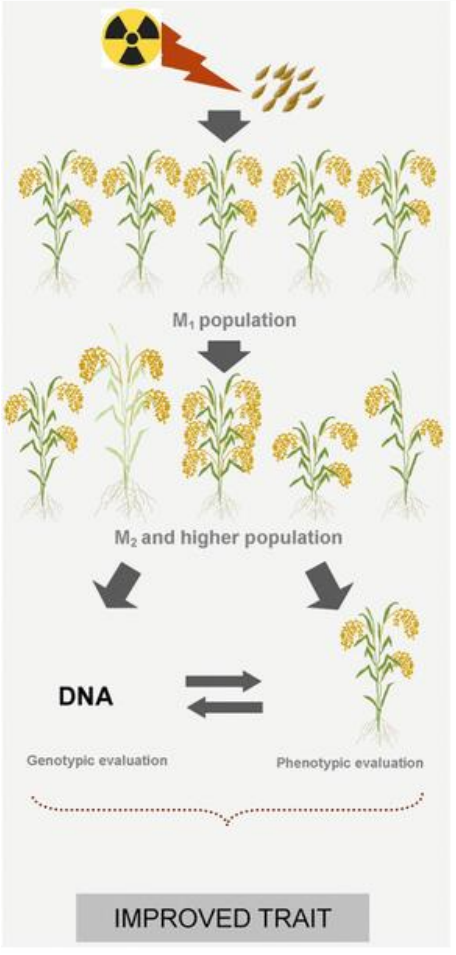


Reverse genetics in plant genetic resources

- Mutagenic agents (chemical/physical) are used to create new variation
- If you observe an interesting trait in an individual, and if the mutation is known, it is possible to link it to the trait to the genomic location
- Mutations can either be untargeted or *targeted* (more to come once we discuss genome editing)

Field Evaluation of Mutagenized Rice Material

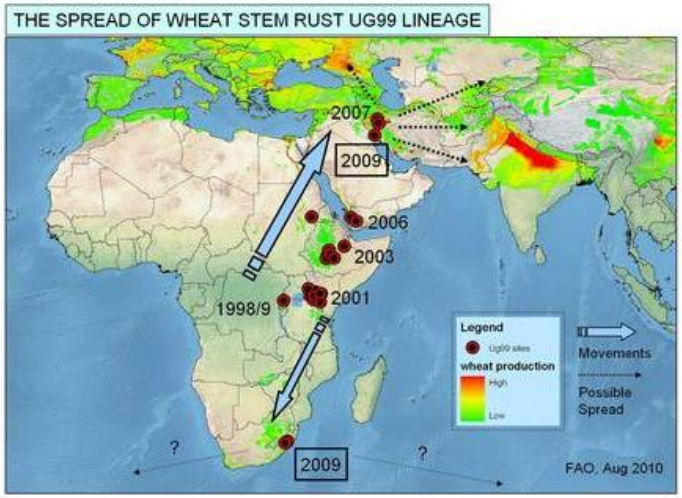
Sydney D. Johnson, Dennis R. Taylor, Thomas H. Tai, Joanna Jankowicz-Cieslak, Bradley J. Till & Alpha B. Jalloh



Forward genetics in plant genetic resources

- Stem rust (*Puccinia graminis* f. sp. *tritici*) is a devastating disease in wheat that may cause up to 100% losses
- A major target of breeding from the green revolution
- There is plenty of resistance genes that have been introgressed from wild species
  - *T. timopheevi*
  - *Ae. ventricosa*
  - *Ae. speltoides*
  - *Ae. tauschii*
  - *S. cereale*
  - *T. monococcum*
  - ….

- Race Ug99 (TTKSK) emerged and overcame resistance of many known genes
- New resistance alleles were found in african landrace materials and transferred to elite materials to confer resistance



Theor Appl Genet (2013) 126:1237–1256
DOI 10.1007/s00122-013-2050-8

ORIGINAL PAPER

Searching for novel sources of field resistance to Ug99 and Ethiopian stem rust races in durum wheat via association mapping

Tesfaye Letta · Marco Maccaferri · Ayele Badebo · Karim Ammar · Andrea Ricci · Jose Crossa · Roberto Tuberosa

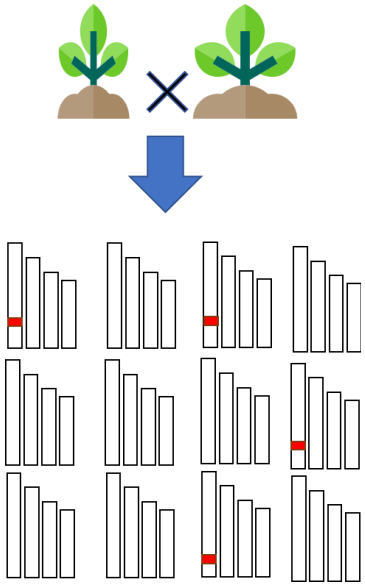
CrossMark

Stem Rust Resistance in a Geographically Diverse Collection of Spring Wheat Lines Collected from Across Africa

Renée Prins[1,2†], Susanne Dreisigacker[3†], Zakkie Pretorius[2], Hester van Schalkwyk[1,2], Elsabet Wessels[1], Corneli Smit[1], Cornel Bender[2], Davinder Singh[4] and Lesley A. Boyd[5]*

[1] CenGen (Pty) Ltd., Worcester, South Africa, [2] Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa, [3] International Maize and Wheat Improvement Centre, Mexico City, Mexico, [4] Faculty of Agriculture and Environment, Plant Breeding Institute Cobbitty, University of Sydney, Narellan, NSW, Australia, [5] Department of Genetics and Breeding, National Institute of Agricultural Botany, Cambridge, UK

Forward and Reverse genetics are not at odds



RESEARCH ARTICLE

Metabolic Engineering to Enhance Provitamin D_3 Accumulation in Edible Tomatoes

Sunmee Choi,[1,†] Min Kyoung You,[1,†] Yun-A Jeon,[1,†] Jaebok Lee,[1] Jinhwa Kim,[1] Young Jin Park,[2] Jeongmo Kim,[1] Jongjin Park,[1] Jae Kwang Kim,[2] and Sunghwa Choe[1,3,*]

Abstract

Ensuring adequate levels of vitamin D3 in the human diet has long been an important objective in crop breeding, as most crops have extremely low levels of this compound. To address this challenge, we have employed the CRISPR-Cas9 gene editing system in tomatoes to induce loss-of-function mutations in one of the two *DWARF5* genes, a homologue of the human dehydrocholesterol Δ^7-reductase gene. Lines with knocked out *SlDWF5A* gene exhibited visually indistinguishable phenotypes, yet remarkably accumulated provitamin D3 levels as high as 6 μg/g dry weight (DW) in the red fruits. As the daily recommended intake of vitamin D is 20 μg (800 IU), consuming a single ripe fresh tomato weighing 150 g (equivalent to 15 g DW) has the potential to significantly alleviate widespread vitamin D deficiencies worldwide.

A recipe for forward genetics: genome-wide association studies (GWAS)

Our ingredients:

1. **Genetic materials**, a set of plant genetic resources in which variation is present for certain traits
2. **Phenotypic values** measured on the set of genetic materials and representing variation of interest
3. **Molecular markers** typed on the set of genetic materials; most commonly SNPs, which are bi-allelic and distributed genome wide
4. **Appropriate statistics** to connect genotypes and phenotypes; many methods, same underlying reasoning

- Many different methods, same underlying reasoning: is there any given allele (marker) associated with the value of the trait of interest?
- In other words, we want to know whether our response variable (y, the phentoype) is associated with our explanatory variable (x, the marker)
- We can address this in a simple statistical framework based on a linear model

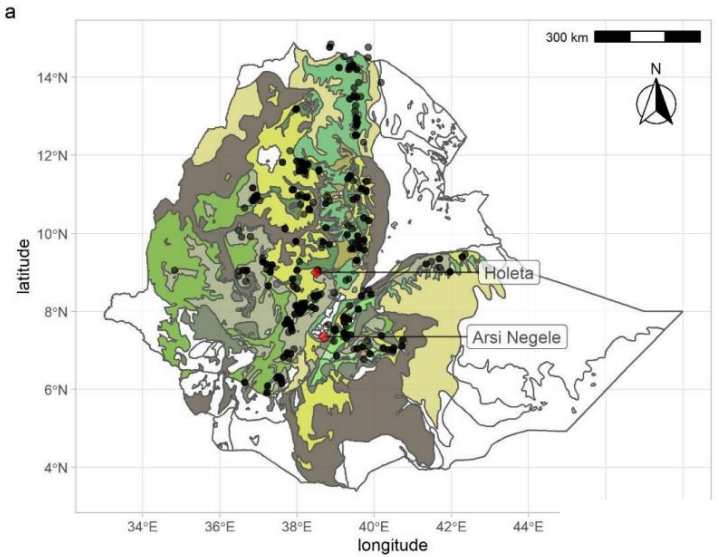
$$y = \beta_0 + \beta_1 x + \varepsilon$$

$$H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$$

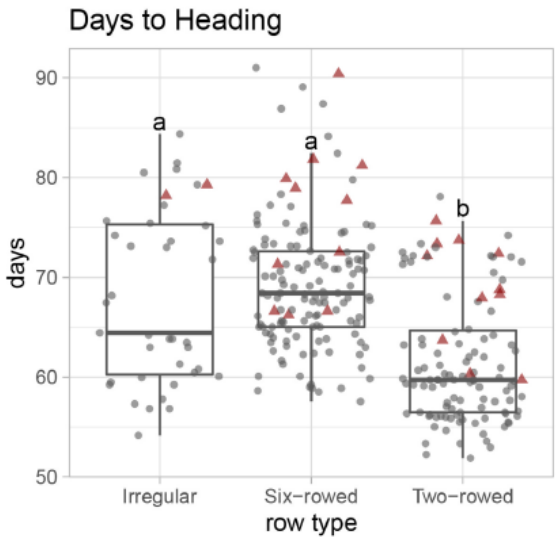
The recipe at work

Research question: climate change is affecting seasonal rainfall distribution in Ethiopia; there is the need to steer breeding towards early flowering genotypes to improve local adaptation; plant genetic resources may have useful alleles to contribute to this

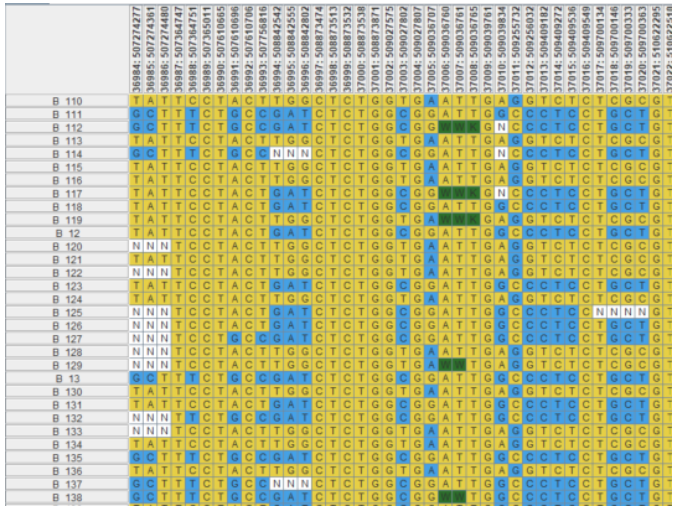
1. **Genetic materials**: A representative collection of 250 Ethiopian barley landraces and breeding lines

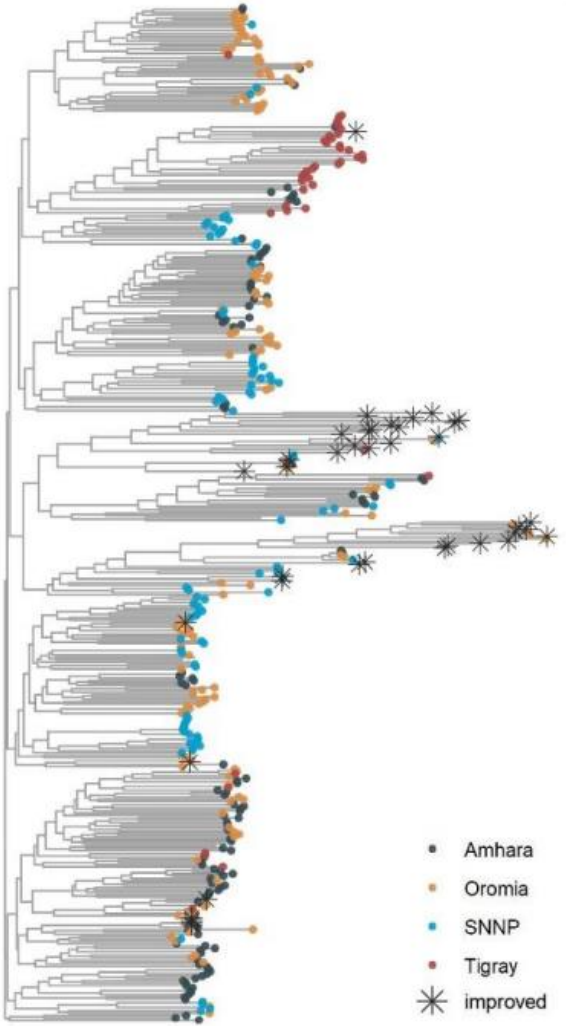


2. **Phentoypic values**: Days to flowering measured on all genetic materials for which genotypic data is also available



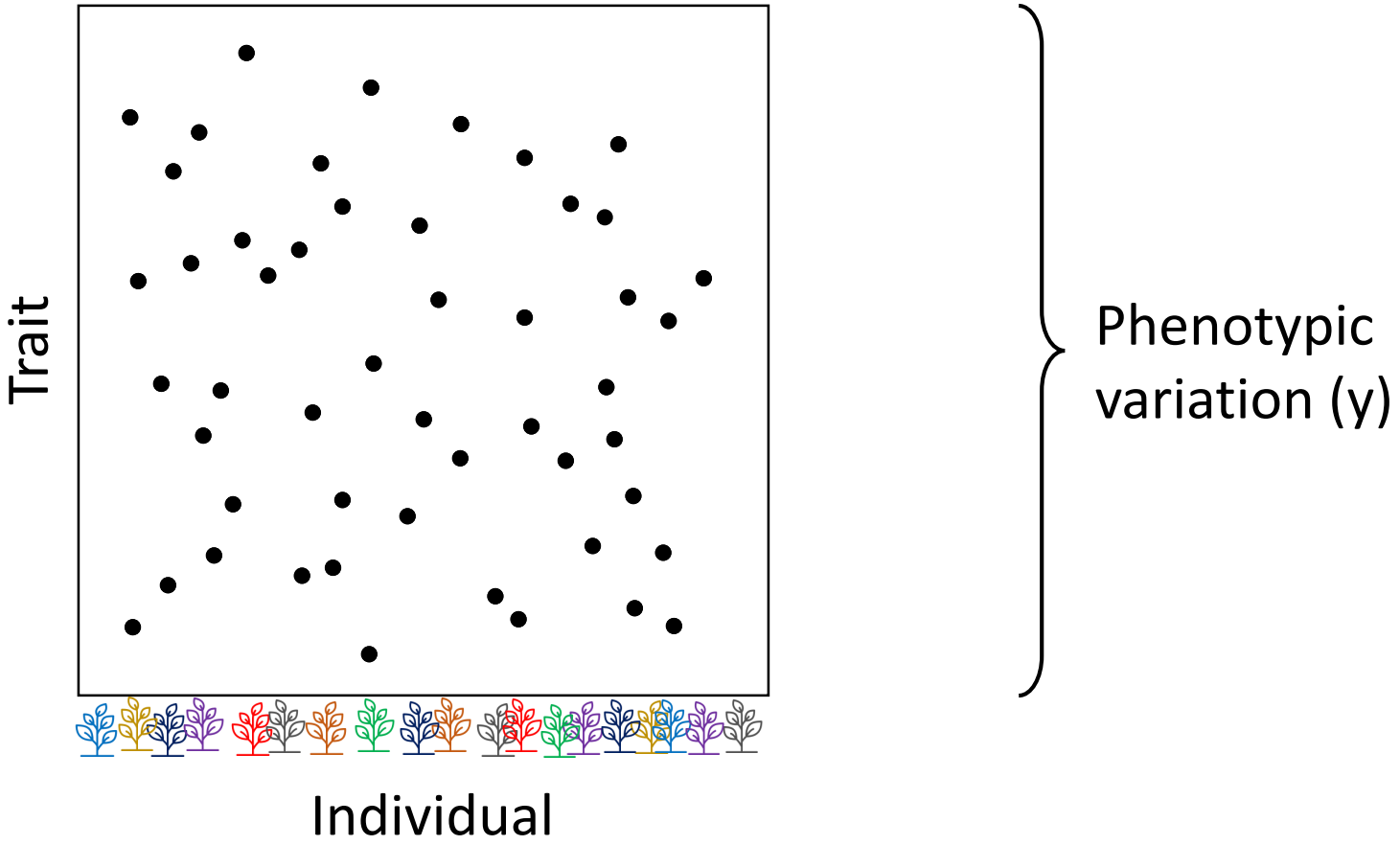
3. **Molecular markers**: 23K+ SNPs describing the diversity of genetic materials across the whole genome





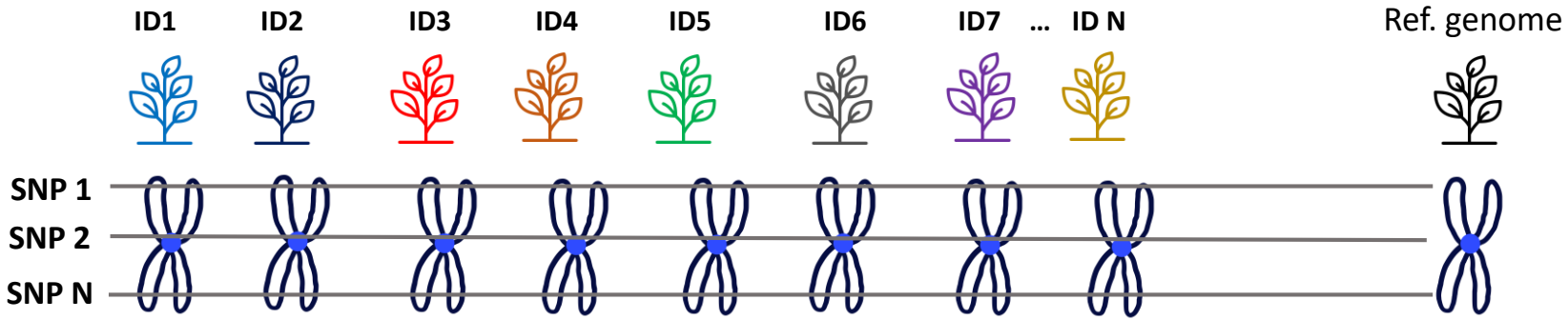
4. Appropriate statistics

4. Appropriate statistics



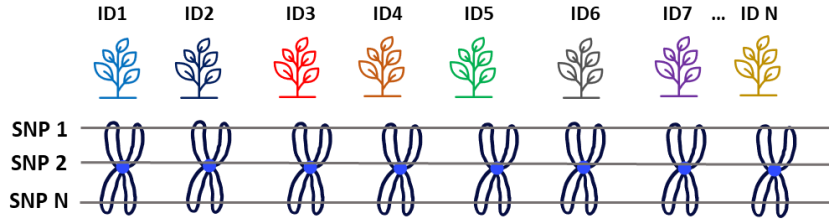
- Each individual is different from the others; when we genotype them with SNPs, we obtain biallelic markers at each locus, with different outputs depending on their allelic diversity
- We don't really need to worry about nucleotides; let's rather think in terms of alleles, and let's call the allele **0** when it is the same as the reference genome and **1** when it is different

Homozyogous reference: 00 Heterozygous: 01 Homozygous aternative: 11

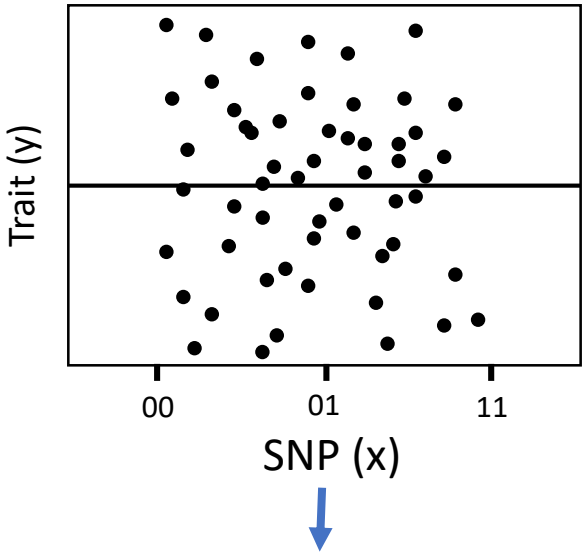


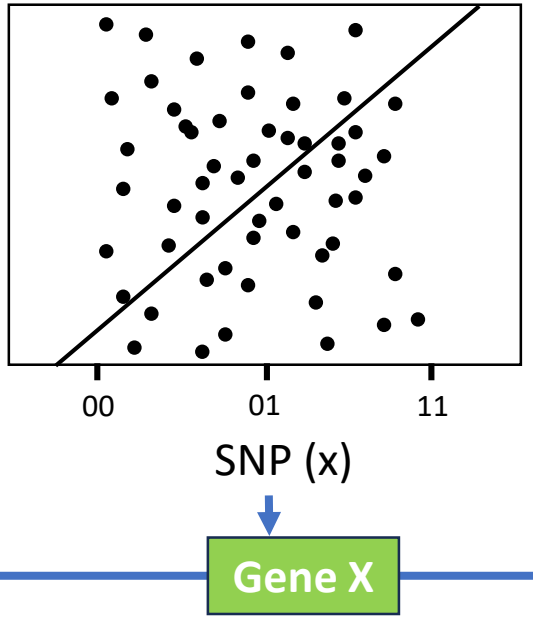
SNP 1	00	00	01	01	11	00	11	00
SNP 2	01	11	00	01	00	11	11	11
SNP *N*	11	00	00	11	11	11	00	01

Running a GWAS fitting a linear model to connect phenotypes and alleles at each locus



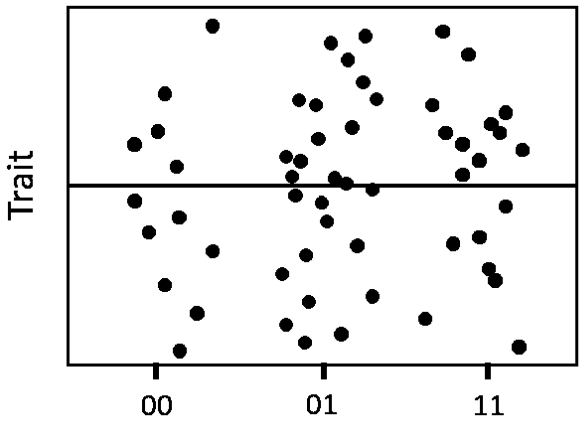
SNP 1	00	00	01	01	11	00	11	00
SNP 2	01	11	00	01	00	11	11	11
SNP *N*	11	00	00	11	11	11	00	01





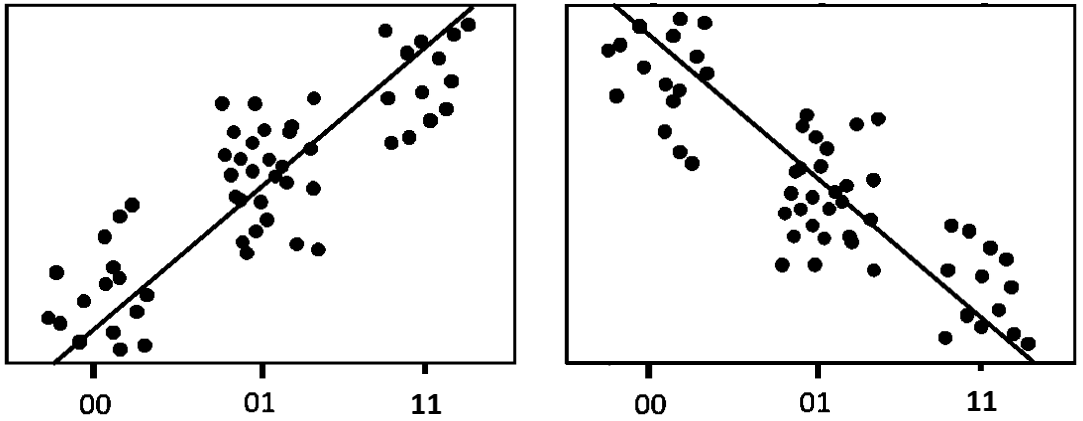
No association; this is the outcome expected on most tests (as most of the markers/loci have nothing to do with the trait)

$$y = \beta_0 + \beta_1 x + \varepsilon$$

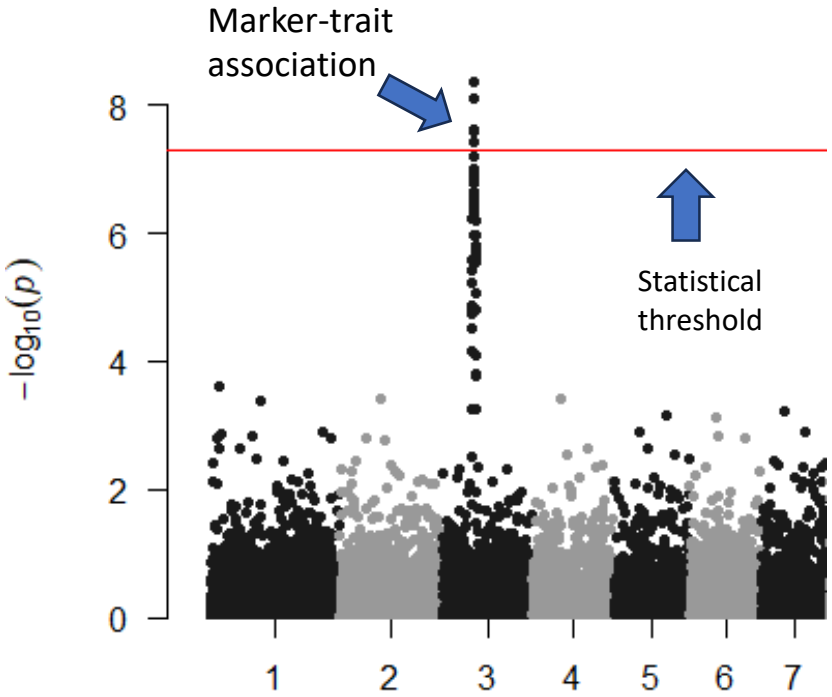


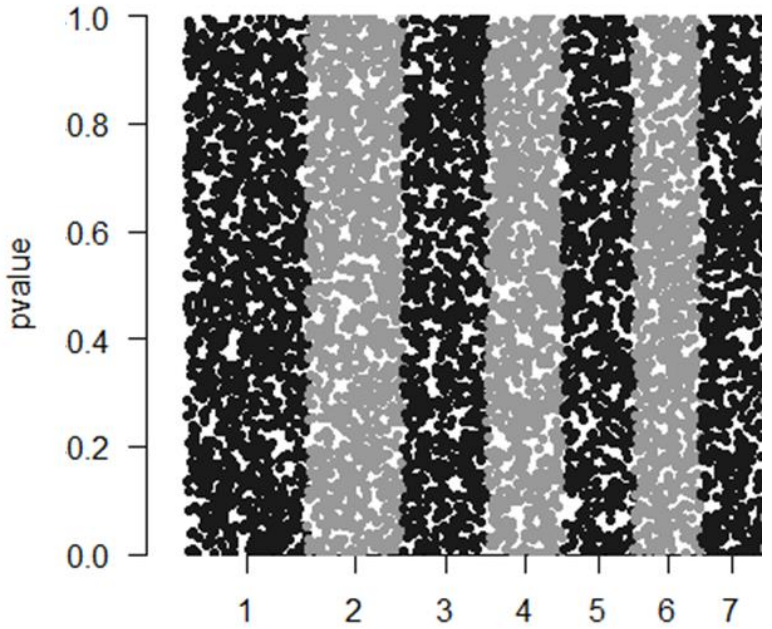
Association; it seems that the response variable is associated with the explanatory variable, and we expect it to happen rarely. To what extent the association is significant, the statistics tells us

$$y = \beta_0 + \beta_1 x + \varepsilon$$



- The model is tested on all markers; if you have 1M markers, that's 1M tests!
- Each test is specific to a marker, which is specific to a genomic location
- The common representation of the outcome is a **Manhattan plot** which puts together position on the genome (x) and significance of the associated test (y)

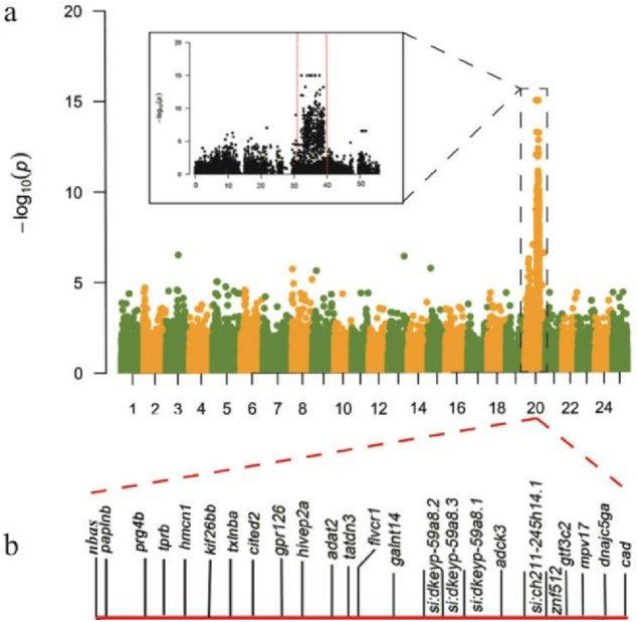




Remember that SNP markers, however many they may be, seldom represent the full extent of variation in the genome

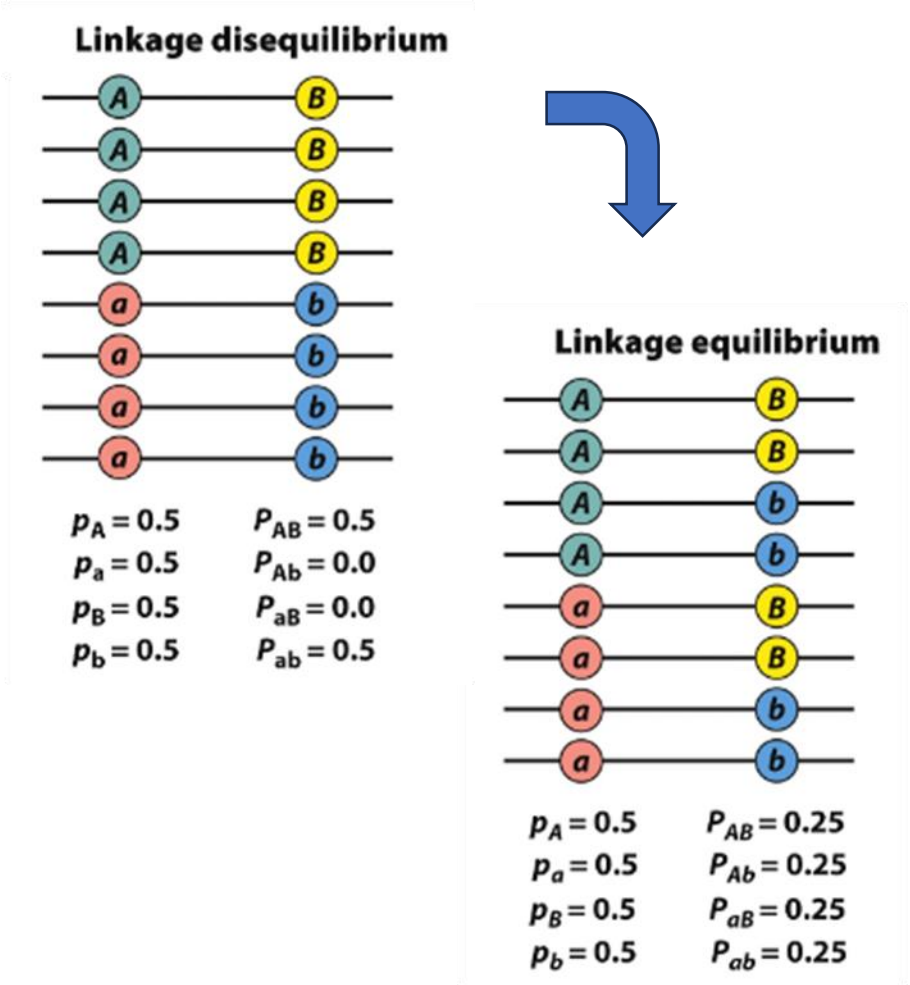
- Markers are our **proxy** to represent variation in the DNA level; they are the mean to an end and not the end itself

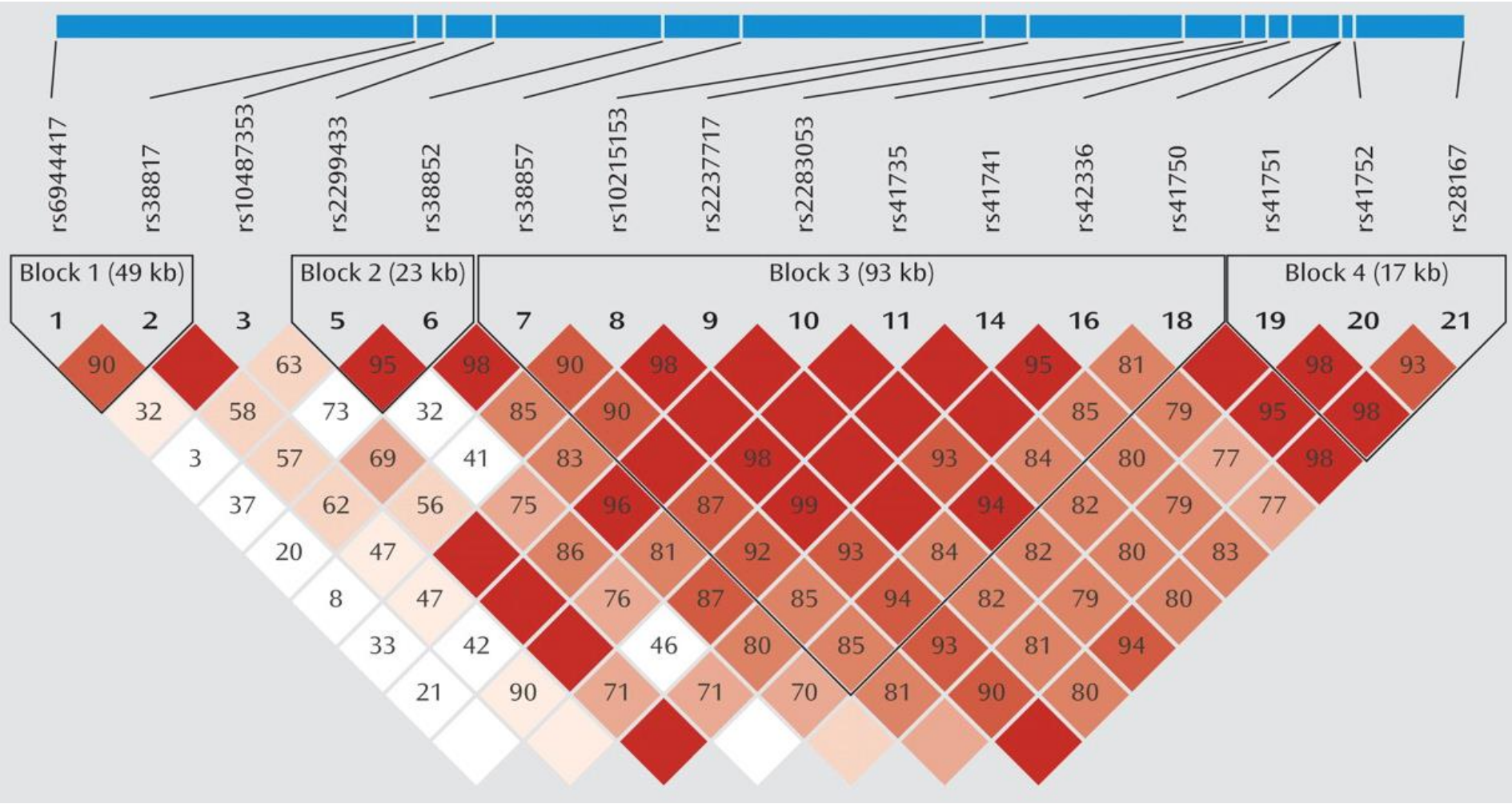
The reason why we capture the «effect» of a specific genetic factor on the value of the trait through GWAS is that **linkage disequlibrium (LD) exists between the marker and the causative variant**



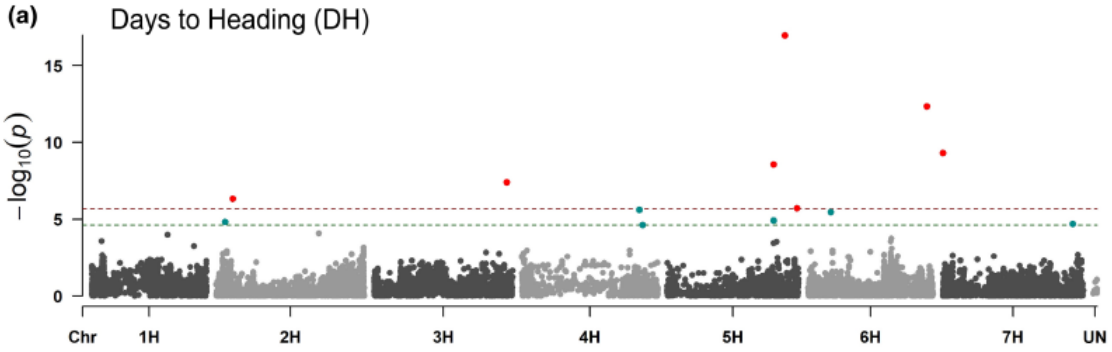
A key concept when it comes to mapping is that of linkage disequilibrium (LD)

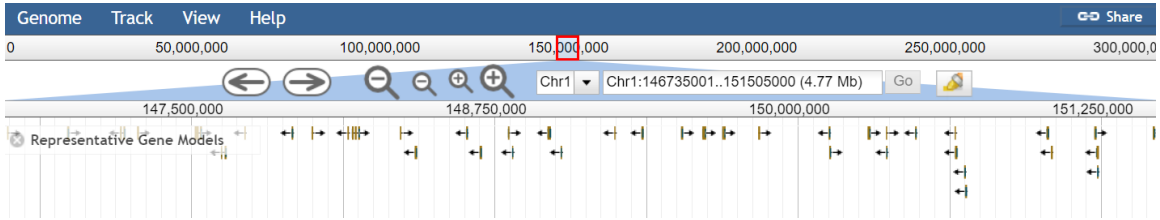
- LD is the non random association of alleles at different loci in a given population
- It occurs when alleles at different loci are inherited together more often than expected by chance
- Recombination decreases LD
- Throughout time, populations move from disequilibrium to equilibrium (assuming that recombination occurs)




rs6944417
rs38817
rs10487353
rs2299433
rs38852
rs38857
rs10215153
rs2237717
rs2283053
rs41735
rs41741
rs42336
rs41750
rs41751
rs41752
rs28167
Block 1 (49 kb)
Block 2 (23 kb)
Block 3 (93 kb)
Block 4 (17 kb)
1 2 3 5 6 7 8 9 10 11 14 16 18 19 20 21
90 63 95 98 90 98 95 81 98 93
32 58 73 32 85 90 85 79 95 98
3 57 69 41 83 98 93 84 80 77 98
37 62 56 75 96 87 99 94 82 79 77
20 47 86 81 92 93 84 82 80 83
8 47 76 87 85 94 82 79 80
33 42 46 80 85 93 81 94
21 90 71 71 70 81 90 80

Back to Ethiopian barley genetic resources now



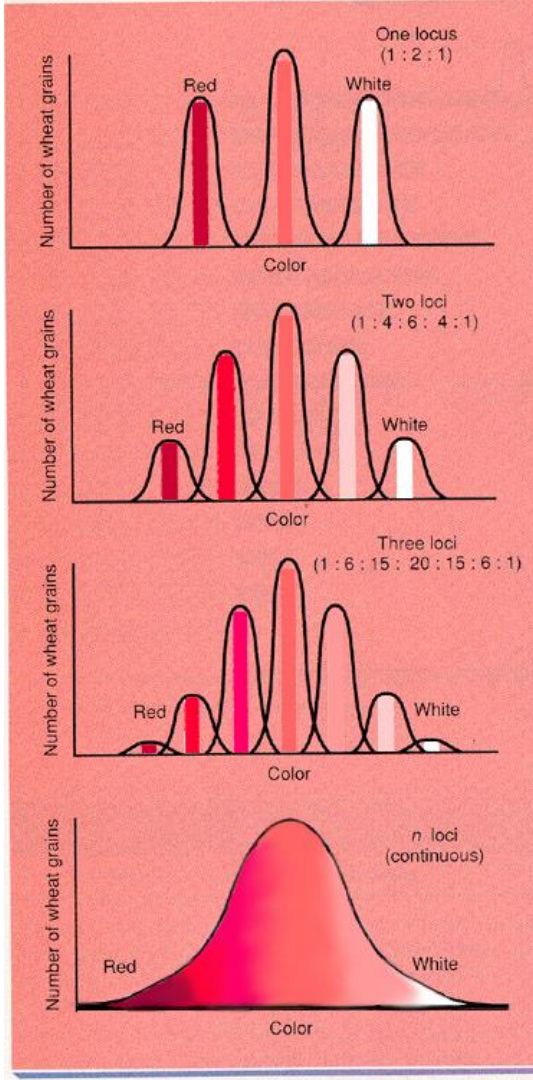


What's next?

- Characterize gene models in the region
- Develop segregating populations to fine map genetic elements
- Design cheap markers tagging loci of interest
- Derive sequences to be tested with **reverse genetics**

Trait determination complexity

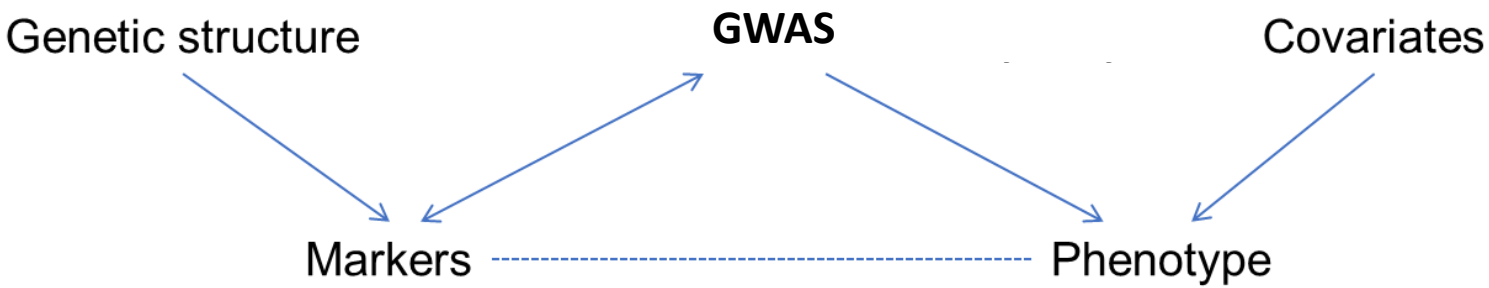
- Most traits of agronomic interest are complex
- They have a quantitative manifestation that is the result of the cumulative contribution and interaction of *n* loci, each with a fraction effect on the trait
- Hence, the term Quantitative Trait Loci (QTL) mapping



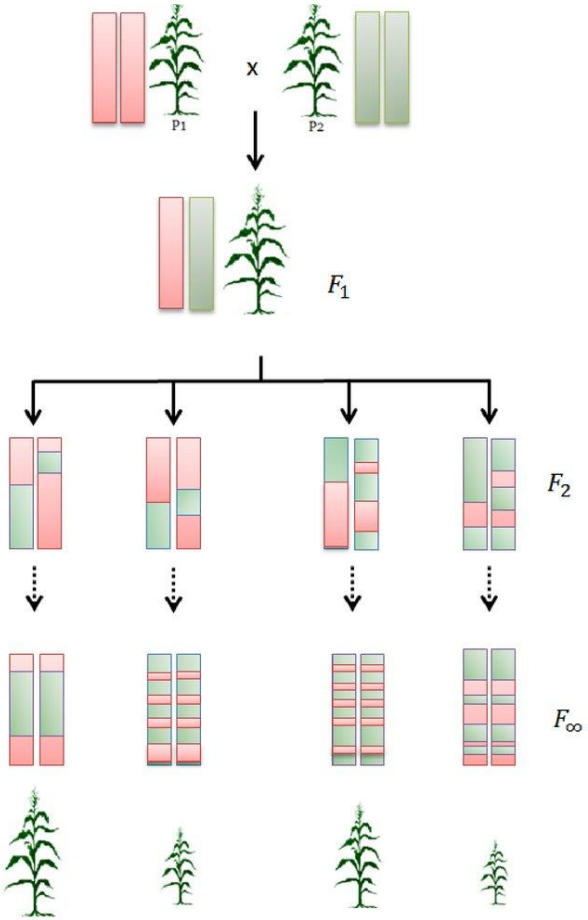
Caveats

The identification of genotype-trait association is a challenging effort depending on many variables, including:

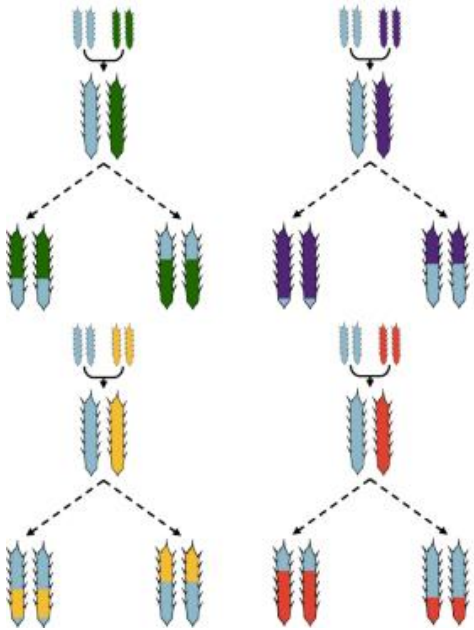
- Diversity available in plant genetic resources
- Sample size and experimental design (statistical power)
- Nature and extent of molecular characterization of the mapping panel / Frequency of recombination (linkage disequilibrium)
- Organization of the genetic diversity in the population (genetic structure)
- Complexity of the trait and heritability



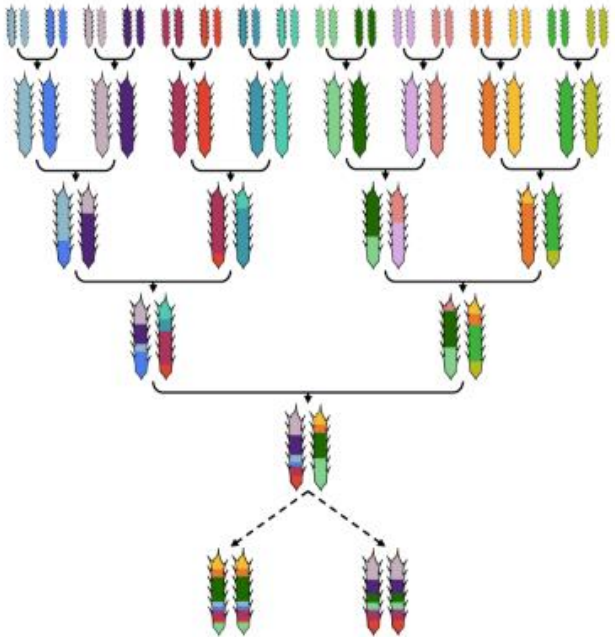
Developement of mapping populations



Nested Association Mapping (NAM) panel

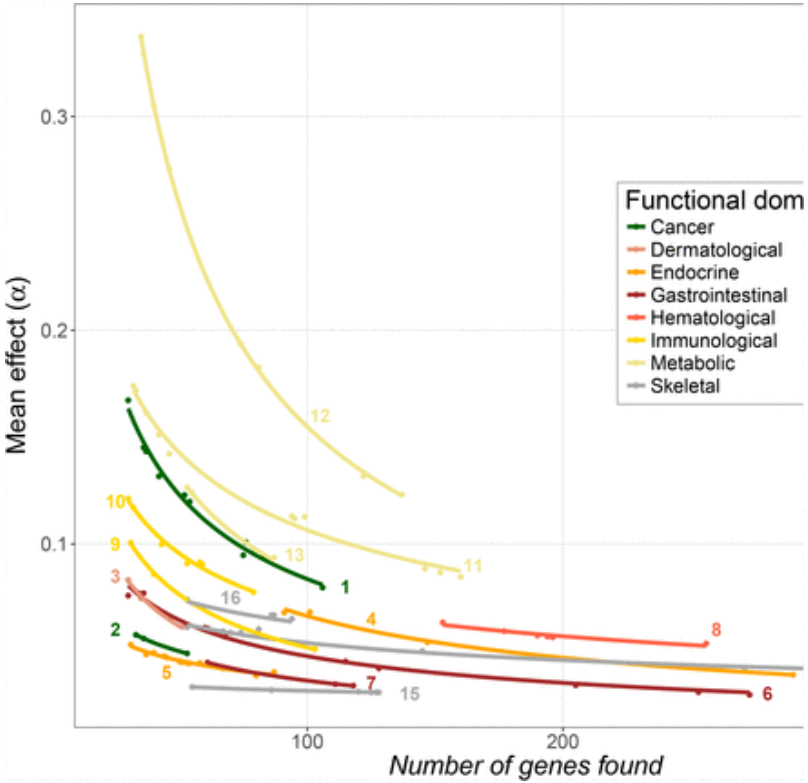


Multi-parent Advanced Generation Inter-Cross (MAGIC)

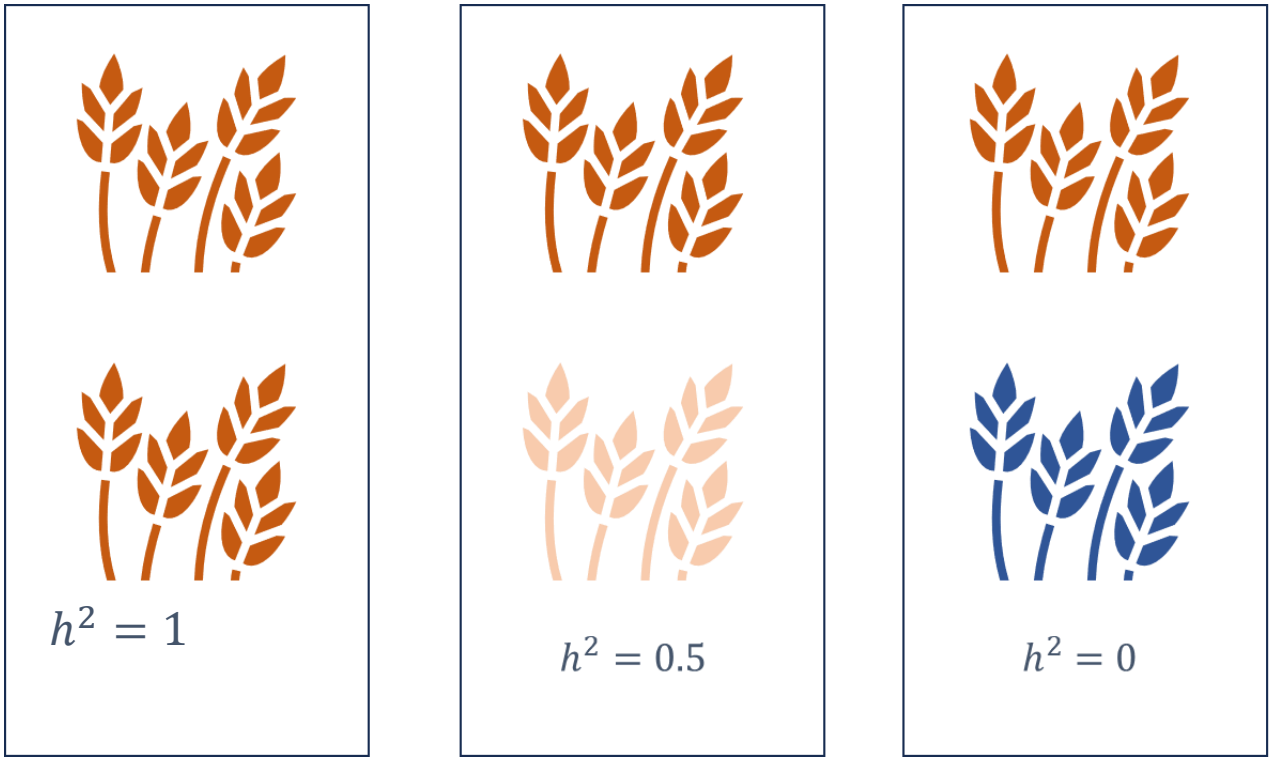


An issue with complexity

- It is becoming increasingly clear that traits are controlled by manifold, small effect loci
- Quantitative genetic mapping studies are tipically underpowered to capture small effects (few cases, many variables)
- Large human studies (e.g. UK BioBank) are filling in the gap



Heritability is the proportion of phenotypic variance that can be explained by genotypic variance; the higher, the easier to map

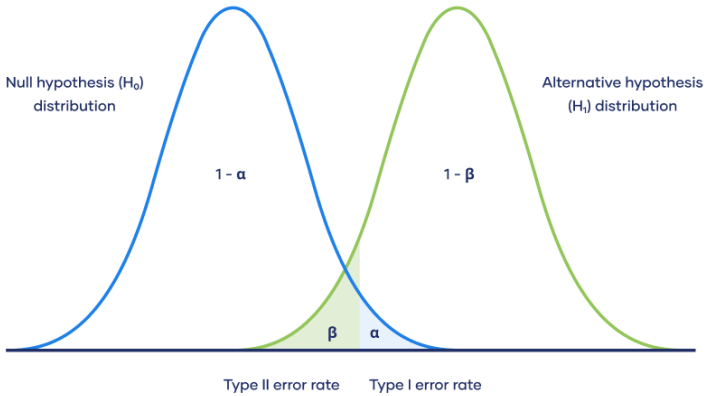


GWAS/forward genetics is a statistical excercise. Presence of a QTL/marker-trait assciation is defined on the basis of a significance threshold, and depending on it theres a certain chance of committing Type I and Type II errors

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

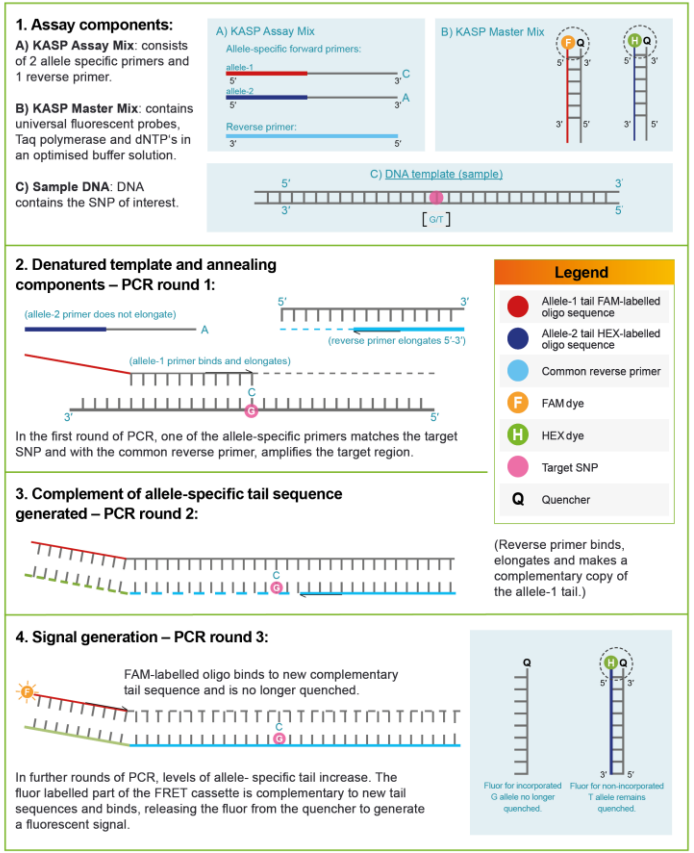
Probability of making Type I and Type II errors

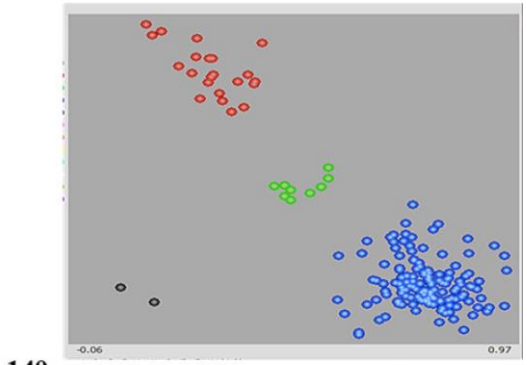


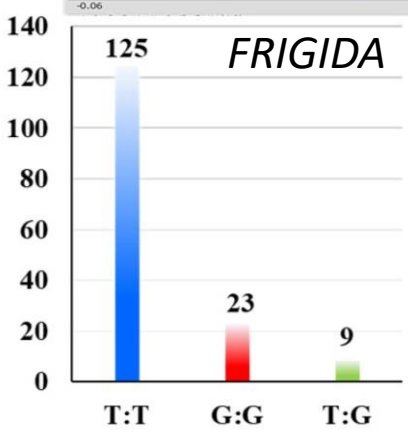
Reverse genetic approaches may be used to «validate» associations

Once marker-trait associations are identified

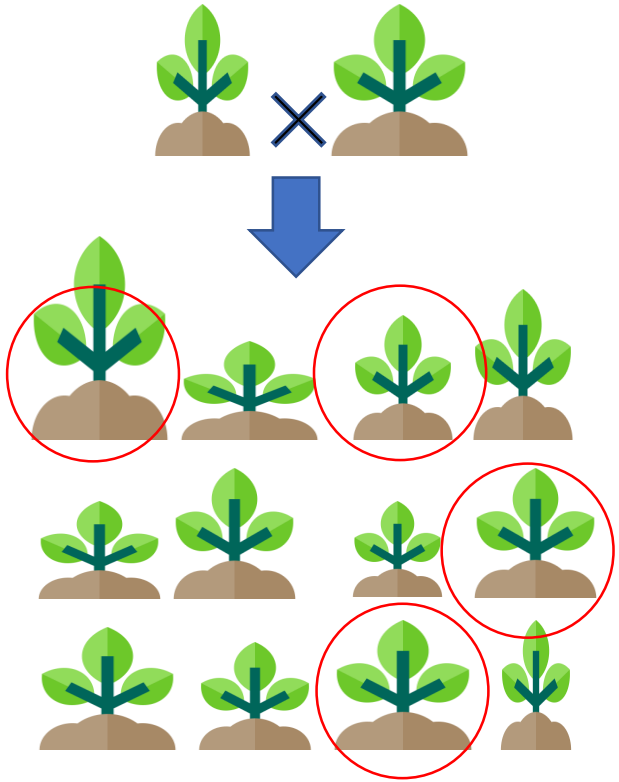
Design markers that can be used by breeders to follow the segregation of a trait of interest



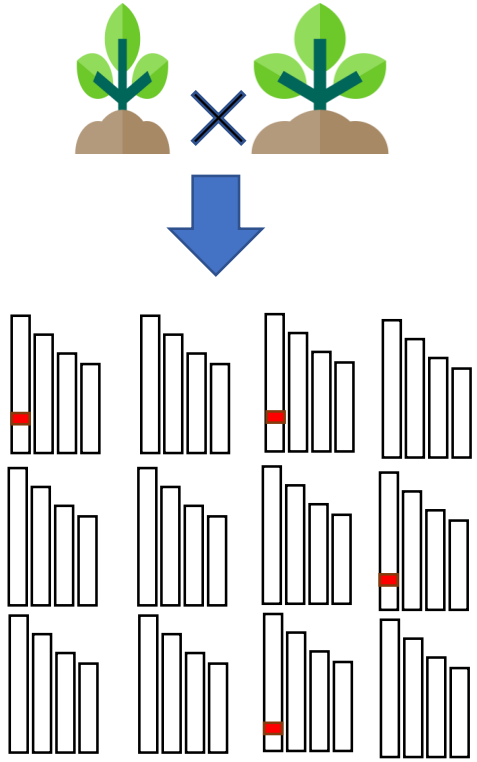




Use of mapping information for breeding



Selection based on traits



Selection based on markers

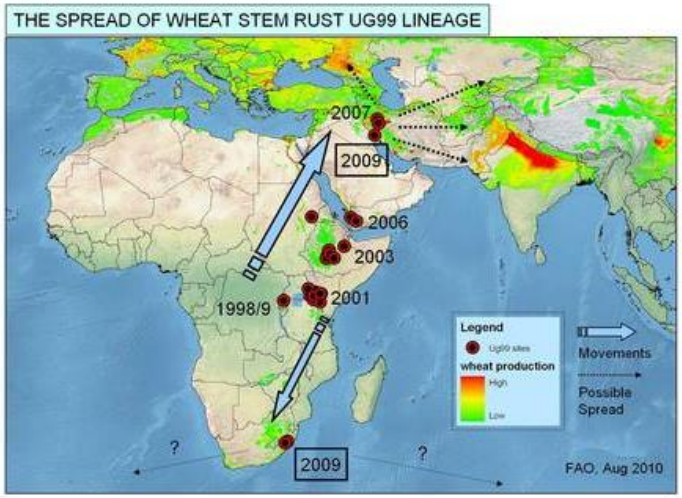
Once a marker-trait association is discovered, it can be used to accelerate the developement of new varieties with improved traits

Modifying genes

Forward genetics in plant genetic resources

- Stem rust (*Puccinia graminis* f. sp. *tritici*) is a devastating disease in wheat that may cause up to 100% losses
- A major target of breeding from the green revolution
- There is plenty of resistance genes that have been introgressed from wild species
  - *T. timopheevi*
  - *Ae. ventricosa*
  - *Ae. speltoides*
  - *Ae. tauschii*
  - *S. cereale*
  - *T. monococcum*
  - ….

- Race Ug99 (TTKSK) emerged and overcame resistance of many known genes
- New resistance alleles were found in african landrace materials and transferred to elite materials to confer resistance



Theor Appl Genet (2013) 126:1237–1256
DOI 10.1007/s00122-013-2050-8

ORIGINAL PAPER

Searching for novel sources of field resistance to Ug99 and Ethiopian stem rust races in durum wheat via association mapping

Tesfaye Letta · Marco Maccaferri · Ayele Badebo · Karim Ammar · Andrea Ricci · Jose Crossa · Roberto Tuberosa

CrossMark

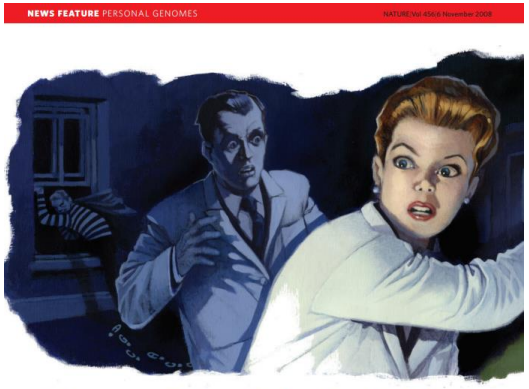
Stem Rust Resistance in a Geographically Diverse Collection of Spring Wheat Lines Collected from Across Africa

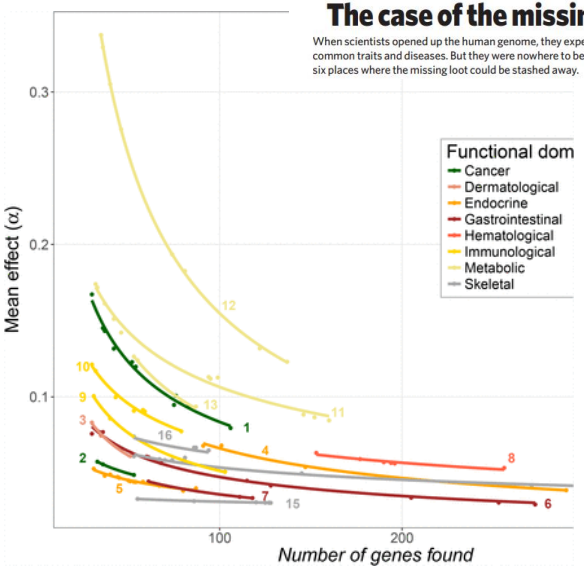
Renée Prins[1,2†], Susanne Dreisigacker[3†], Zakkie Pretorius[2], Hester van Schalkwyk[1,2], Elsabet Wessels[1], Corneli Smit[1], Cornel Bender[2], Davinder Singh[4] and Lesley A. Boyd[5]*

[1] CenGen (Pty) Ltd., Worcester, South Africa, [2] Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa, [3] International Maize and Wheat Improvement Centre, Mexico City, Mexico, [4] Faculty of Agriculture and Environment, Plant Breeding Institute Cobbitty, University of Sydney, Narellan, NSW, Australia, [5] Department of Genetics and Breeding, National Institute of Agricultural Botany, Cambridge, UK

- It is becoming increasingly clear that traits are controlled by manifold, small effect loci
- Forward genetics are tipically underpowered to capture small effects (few cases, many variables)
- Large studies are starting to fill in the gap

It is easier to map QTL for disease resistance than to map QTL for yield





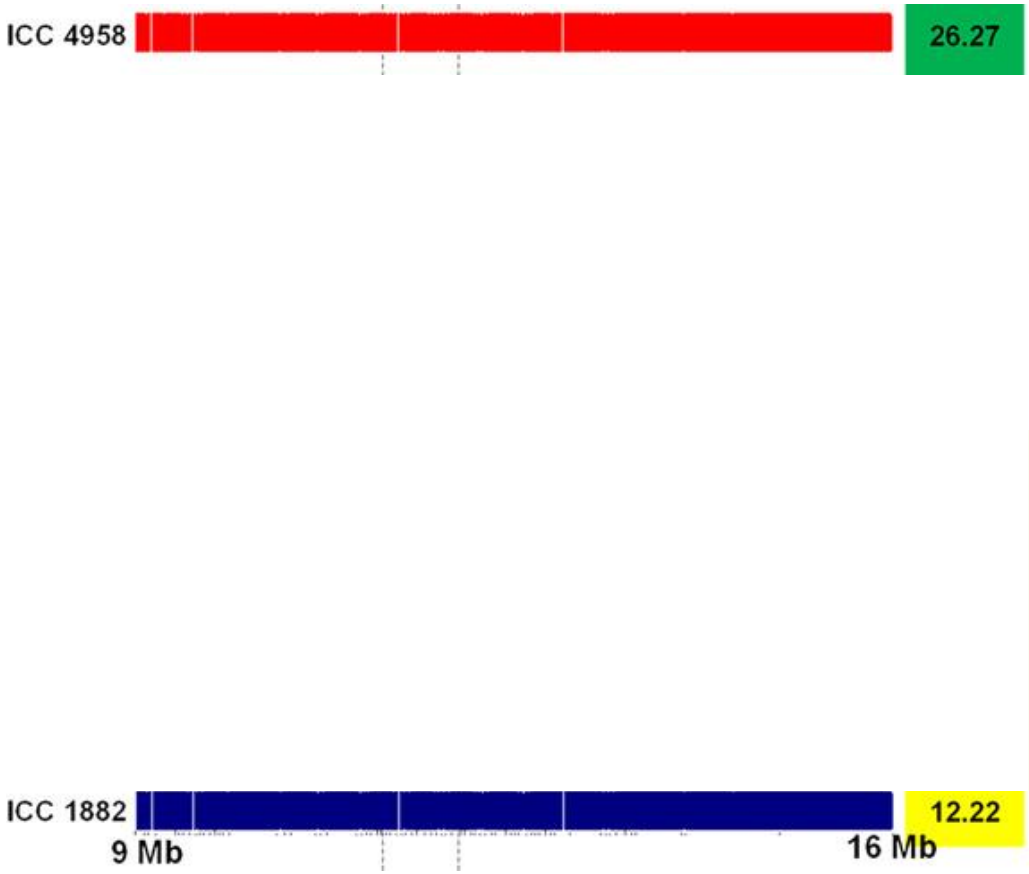
From López-Cortegano, Caballero 2018

Towards gene identification

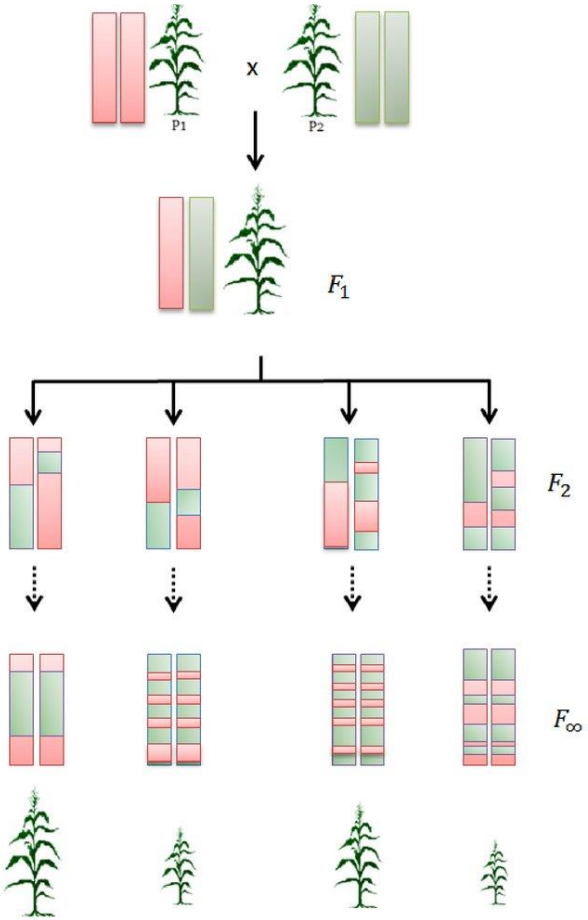
The mapping resolution depends on recombination density

It is very infrequent to be able to identify individual causative variants, and this depends on a number of factors:

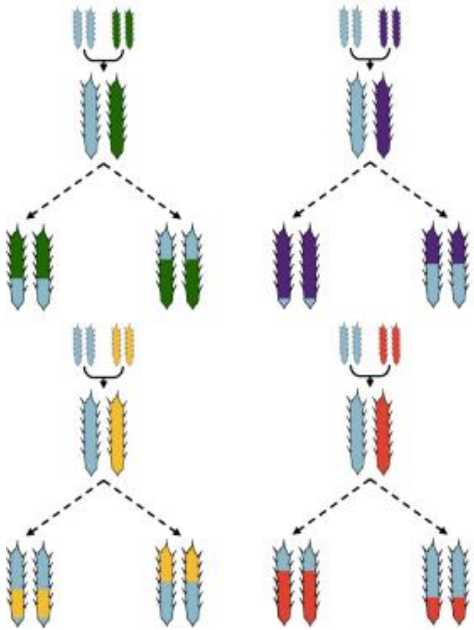
- Marker density
- Recombination density
- Complexity of the trait
- Quality of the annotation



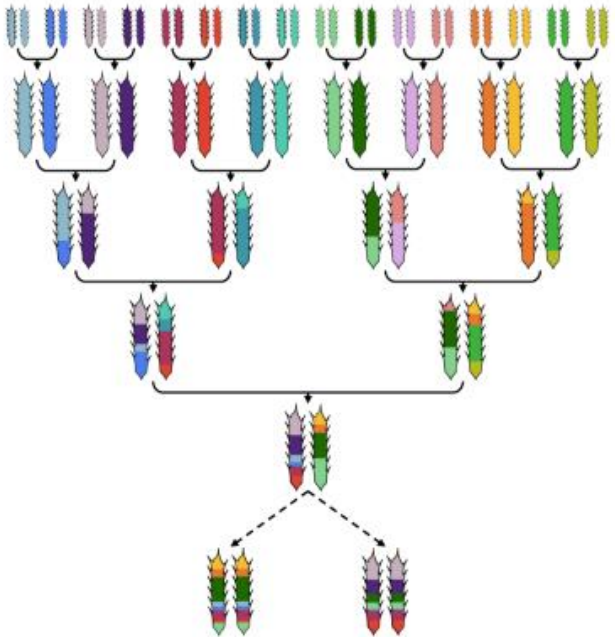
Developement of mapping populations



Nested Association Mapping (NAM) panel



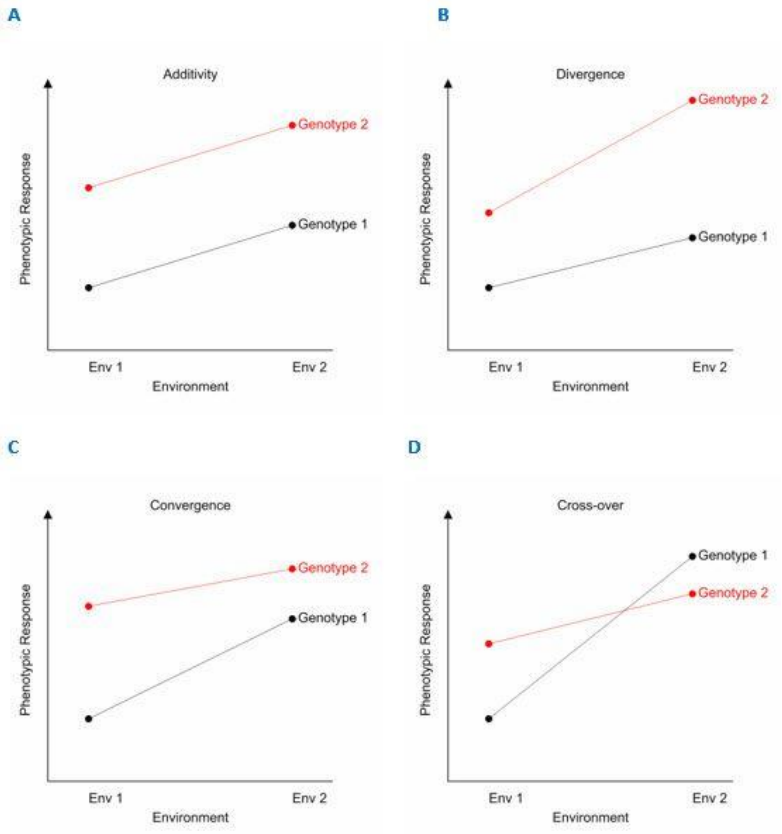
Multi-parent Advanced Generation Inter-Cross (MAGIC)



Decomposing phenotypic variance in quantitative traits

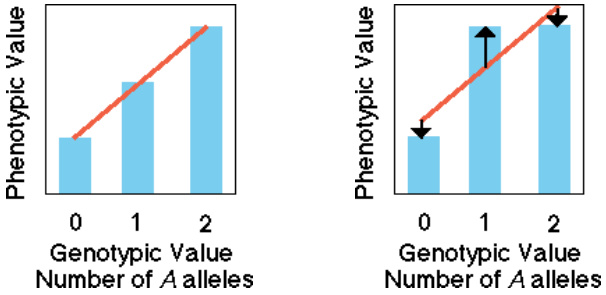
$$V_p = V_g + V_e + V_{ge}$$

- The performance of each individual is determined both by its genotype composition and by the environment
- The best performer in one environment may not be the best in another



Decomposing genotypic variance in quantitative traits

$$V_G = V_A + V_D + V_I$$



Genetic variance can also be decomposed in fundamental components:

- Additive genetic variance (A), refers to the deviation from the mean phenotype due to inheritance of a particular allele and this allele's relative effect on phenotype, i.e., relative to the mean phenotype of the population
- Dominance variance (D) due to interactions between alternative alleles at a specific locus
- Interaction or epistatic variance (I) due to interaction between alleles at different loci

The heritability of a given trait is calculated as the fraction of the trait variance that can be explained by genotypic variance

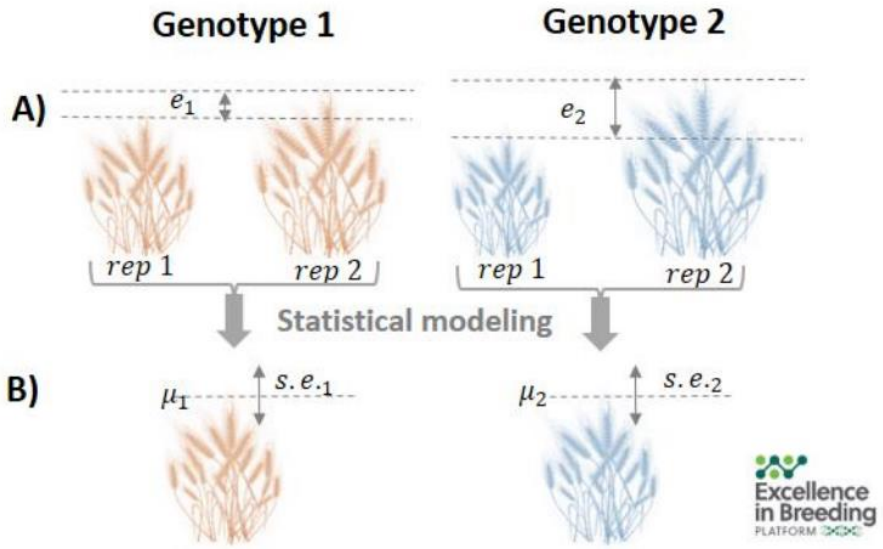
$$H^2 = \frac{\sigma_g}{\sigma_g + \sigma_e}$$

Broad sense heritability: all sources of genetic variance are considered

$$h^2 = \frac{\sigma_g}{\sigma_g + \sigma_e}$$

Narrow sense heritability: only additive genetic variance is considered

- If H^2 is 0, none of the phenotypic variation can be explained by the genetic variation, it is all due to variation in the environment
- If H^2 is small, the trait is strongly influenced by the environment (e.g., yield)
- If H^2 is large, the trait is only slightly influenced by the environment (e.g. flower colour).



Heritability:

Based on error plot variance

$$H^2_{Standard} = \frac{\sigma^2_g}{\sigma^2_g + (\sigma^2_e/n)}$$

Based on average genotype standard error

$$H^2_{Piepho} = \frac{\sigma^2_g}{\sigma^2_g + \bar{v}^{BLUE}_{\Delta}}$$

In a breeding perspective, heritability is quite important

Characters	Heritability	
	Broad sense (%)	Narrow sense (%)
Plant height (cm)	71	48
Number of panicles / plant	30	14
Number of spikelets / panicle	32	29
Number of fertile spikelets / panicle	36	27
Percentage spikelets fertility / plant	49	47
100 grain weight / plant (g)	67	50
Grain yield / plant (g)	32	19

Common misconceptions about heritability

- "**A heritability of .4 means that 40% of the trait is determined by genetics**". Nope. A heritability of 0.4 indicates that 40% of all the phenotypic variation for that trait is due to variation in genotypes for that trait (and not that in each plant 0.4 of the phenotype is determined by genes)
- "**A low heritability means that trait is not determined by genes**". Also wrong; low heritability may be due to low variance (in the population) or too much error
- "**A heritability is a fixed value**". It really is a population value and depends on genetic materials and experimental conditions in which variances are assessed
- "**A high heritability implies a major-effect QTL**". It could actually be due to a number of different QTL (each with small effects)

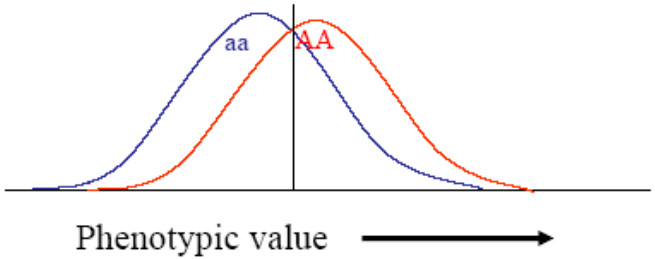
QTL mapping

- A QTL is a locus contributing to the phenotypic value of a complex (multigenic) trait
- QTL mapping aims at the dissection of complex traits into Mendelian factors: understand their location and their relative importance

QTL mapping to-dos

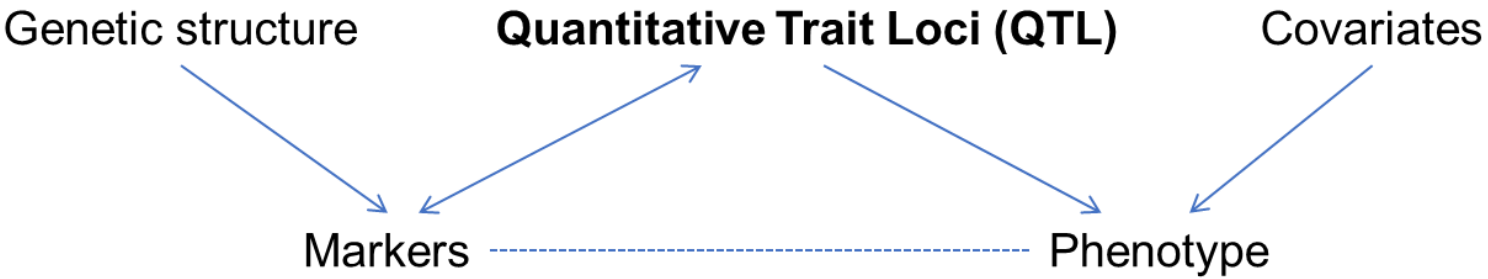
- Control environment & vary genetics
- Use a panel of genetically diverse plants with different trait levels
- Leverage statistical association between alleles and trait levels to find genomic loci and possibly genes

$$y = \beta_0 + \beta_1 x + \varepsilon$$



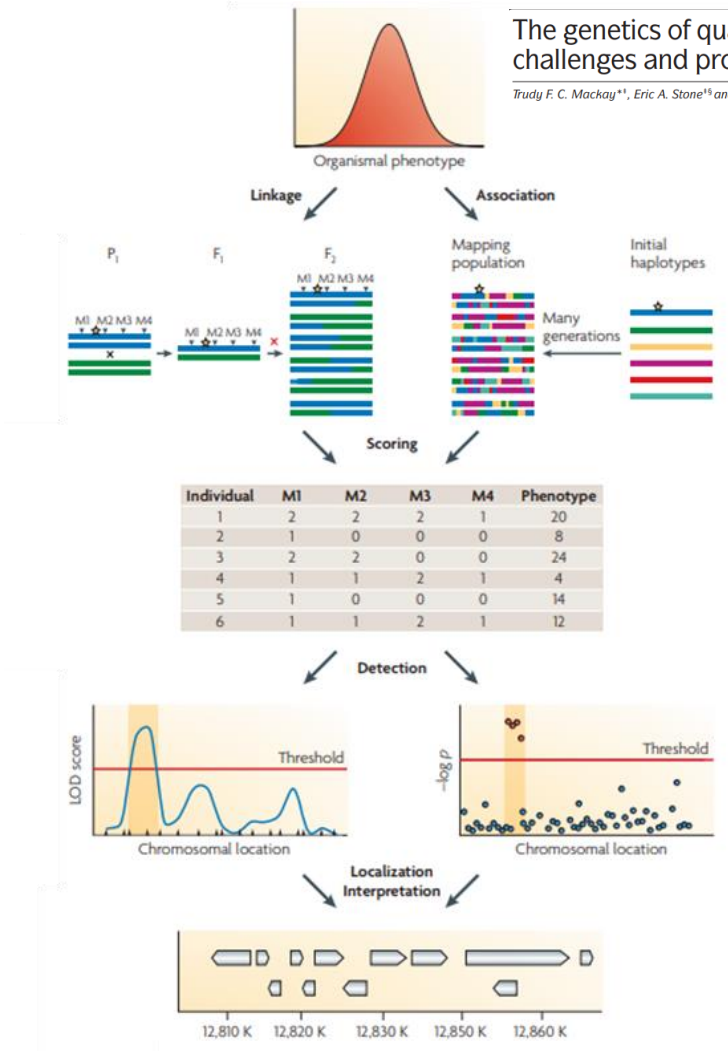
The identification of QTL is a challenging effort depending on many variables, including:

- Diversity in the mapping panel
- Frequency of recombination (linkage disequilibrium)
- Nature and extent of molecular characterization of the mapping panel
- Complexity of the trait (heritability)
- Sample size (statistical power)



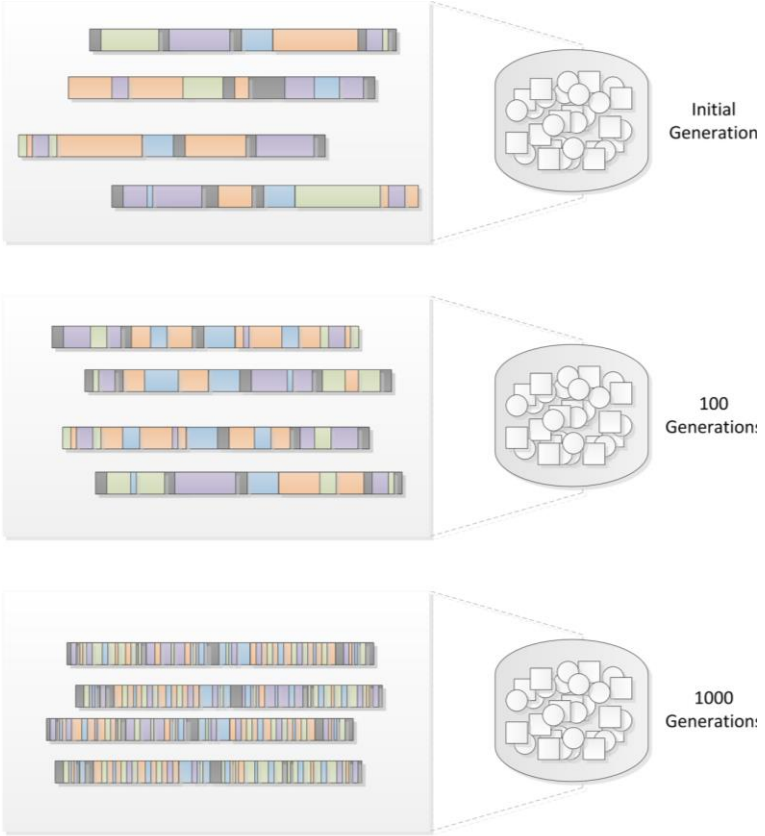
QTL mapping requires four things:

1. Segregating genetic materials
2. Genetic markers characterizing the mapping population
3. Consistent and reproducible phenotypic data
4. Appropriate statistics



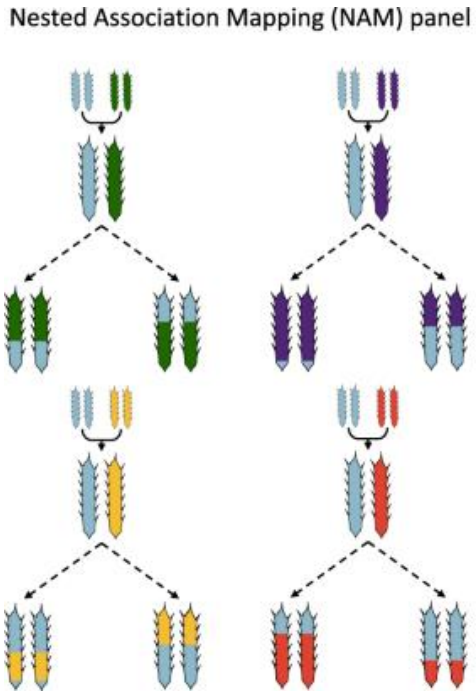
Diversity panels → groups of individuals collected from nature and resulting from an history of intermating

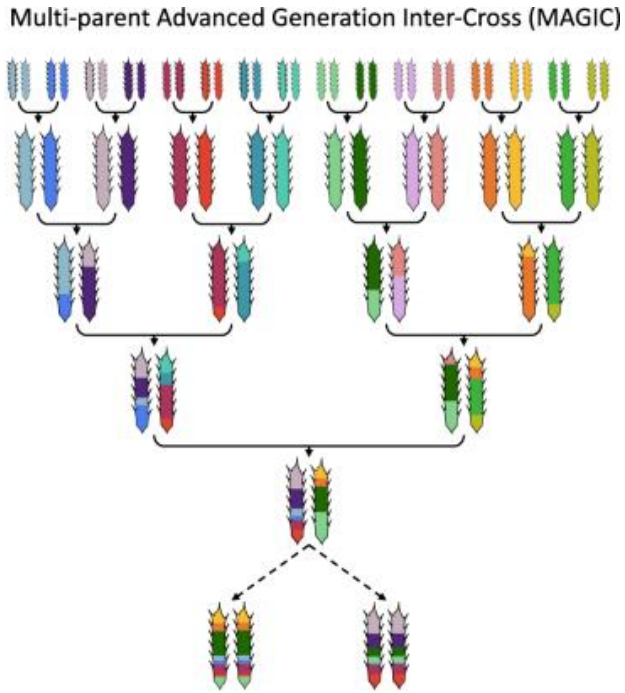
- Tipically high diversity
- High recombination density
- Differently from experimental crosses, the pedigree (derivation) of individuals is unkown



Multiparental mapping populations (MPPs) → artificial segregant populations developed intercrossing 2+ parental lines

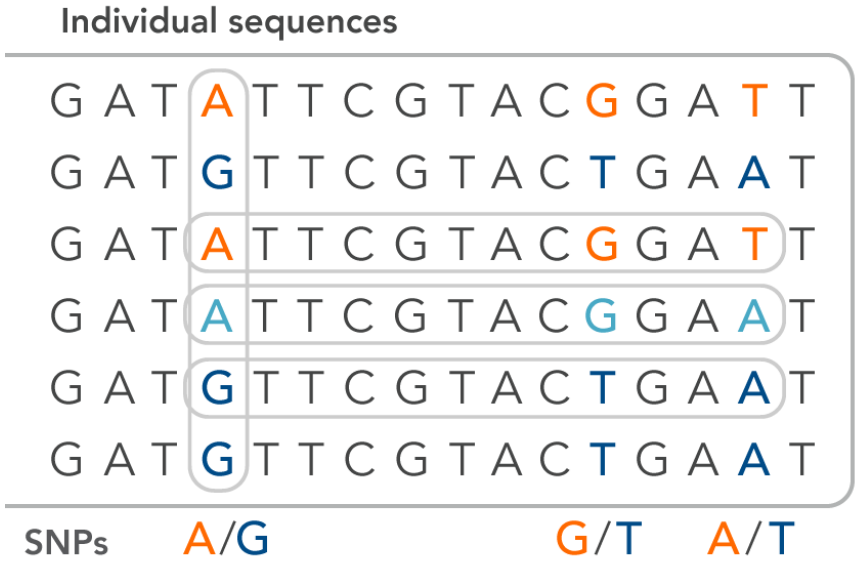
- Tipically high diversity
- High recombination density
- Known pedigree and balanced diversity





2. Molecular markers

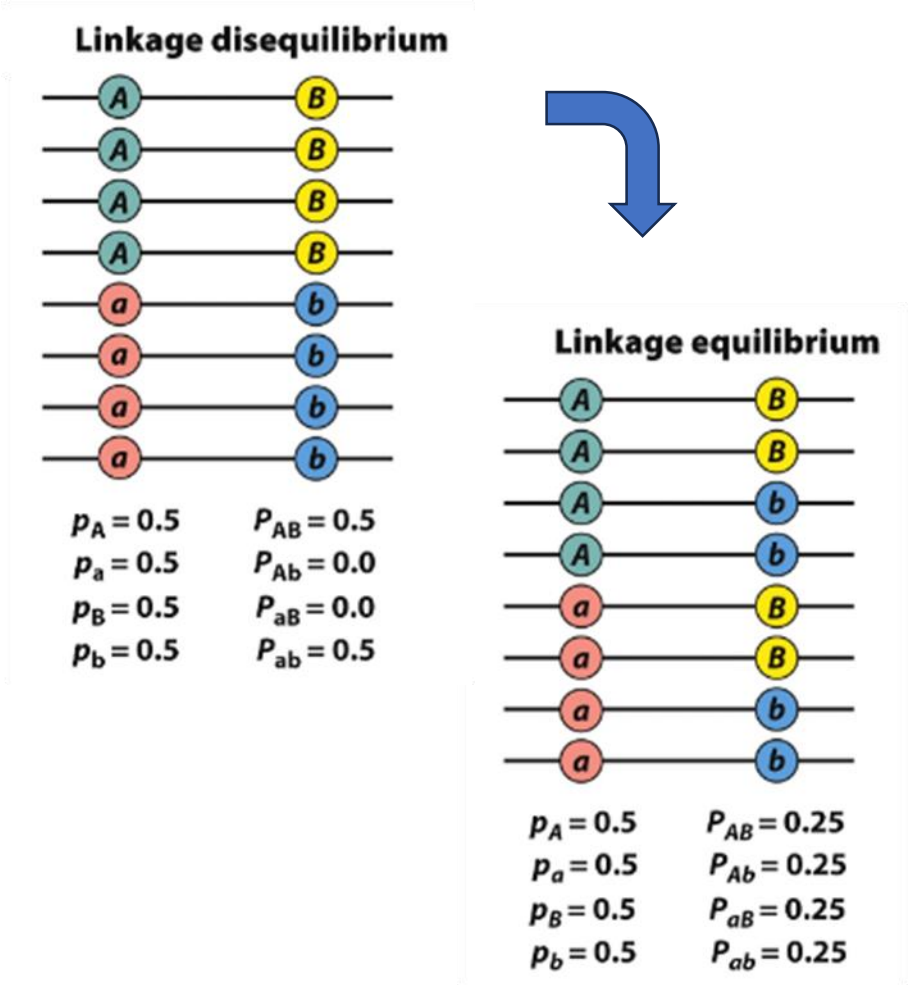
- Gentotyping information is necessary to characterize the genetic diversity in the mapping population
- You have all sort of molecular markers to choose from, the most accurate being single nucleotide polymorphisms (SNPs)

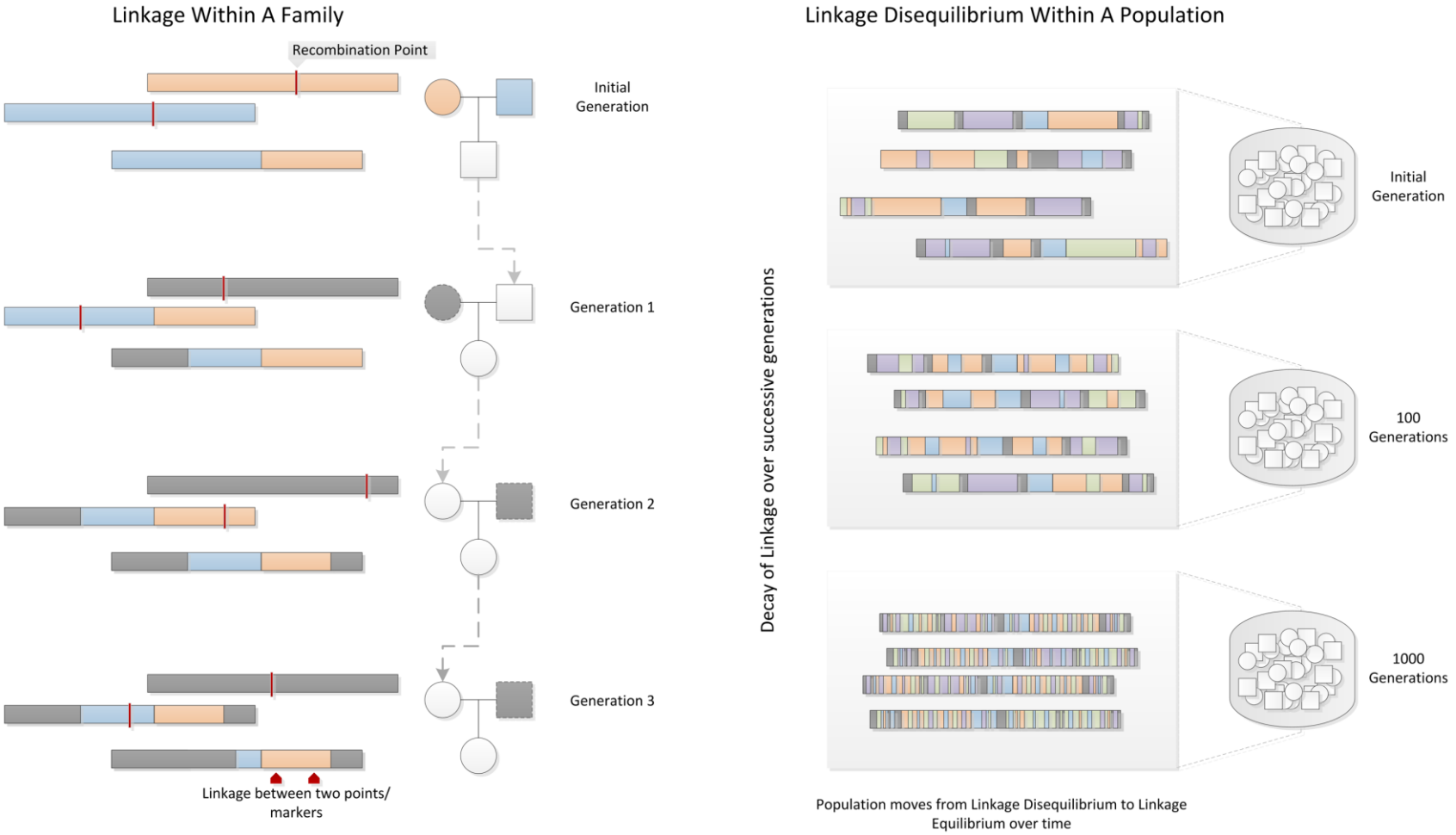


- A haplotype is a combination of alleles at multiple loci that are transmitted together on the same chromosome
- Looking back at genetic materials, a haplotype represents a group of loci within an organism that was inherited together from a single parent.

A key concept when it comes to mapping is that of linkage disequilibrium (LD)

- LD is the non random association of alleles at different loci in a given population
- It occurs when alleles at different loci are inherited together more often than expected by chance
- Recombination decreases LD
- Throughout time, populations move from disequilibrium to equilibrium (assuming that recombination occurs)

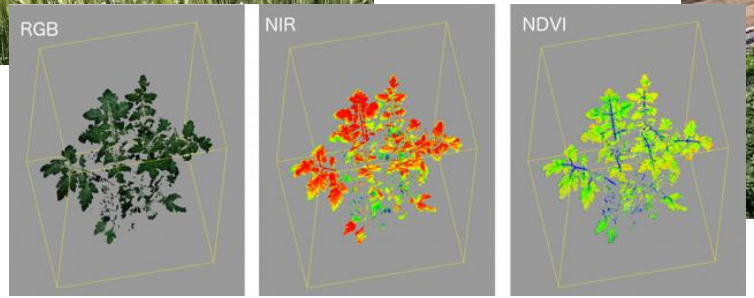




https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002822

3. Well measured phenotypes

- In order to support QTL mapping, measurements must be repeatible (remember G x E) and accurate (lower error)
- Nowadays, phenotyping comes in a –omics dimension



4. Appropriate statistics

- Many different methods exist, depending on population, distribution of traits, genetic mechanisms considered, molecular markers, …

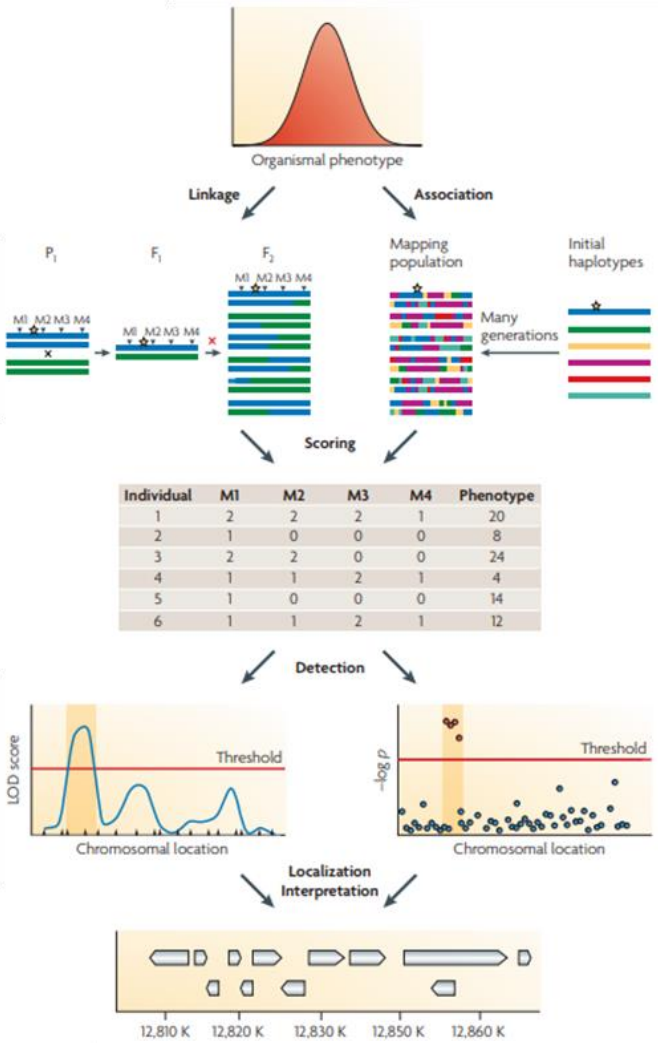
$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Two main avenues:
  - **Linkage / interval mapping;** When pedigree is known (artificial populations) and intervals of markers – rather than individual markers – is used to support mapping. The resulting statistic is the logarithm of odds (LOD), or the log of the probability of having a QTL in a specific location over the probability of not having it
  - **Genome-wide association studies / LD mapping**; a mapping conducted marker by marker, used in diversity panels. The resulting statistic is a p-value coming from testing the alternative hypothesis of a genotypic effect on the trait

Linkage mapping

- Low marker density required
- Fully known pedigree
- More robust

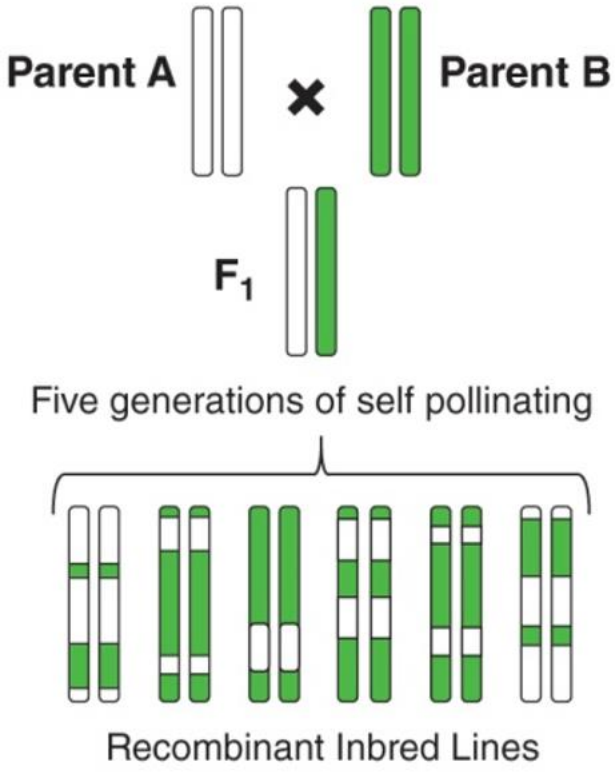
- Limited variation
- Low definition
- Time demanding

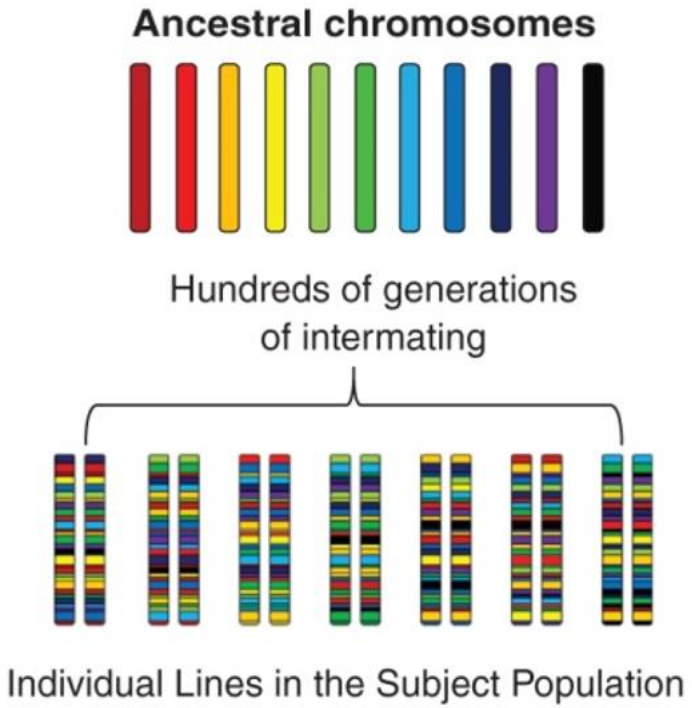


Association (GWAS)

- High marker density necessary
- Hidden structure, LD
- Higher false positive rate

- Broad variation
- High definition
- Faster, cheaper


Parent A
Parent B
F_1
Five generations of self pollinating
Recombinant Inbred Lines


Ancestral chromosomes
Hundreds of generations
of intermating
Individual Lines in the Subject Population

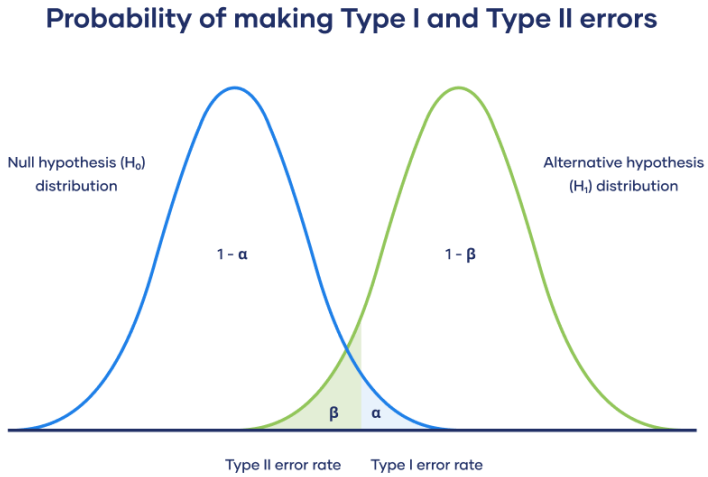
Forward genetics is a statistical excercise (LOD or pvalue). Presence of a QTL is defined on the basis of a significance threshold

$$y = \beta_0 + \beta_1 x + \varepsilon$$

$$H_0: \beta_1 = 0 \qquad H_A: \beta_1 \neq 0$$

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	**Type I error** False positive Probability = α	**Correct decision** True positive Probability = 1 - β
Not rejected	**Correct decision** True negative Probability = 1 - α	**Type II error** False negative Probability = β



Multiple testing problem: when conducting multiple statistical tests simultaneously, the chance of incorrectly rejecting a true null hypothesis (false positive) increases

- **Bonferroni**: the nominal test p-value (tipically 0.05) is divided by the number of independent tests performed
- **False Discovery Rate (FDR)**: an adjusted p-value distribution that is specific to each test and that takes in account the expected proportion of false positives among all significant tests
- **Permutations**: scrambling the phenotypic values and looking for QTL (expecting not to find any). Reapeat a large n of times and produce a distribution of statics that represents noise. Then pick a threshold according to the distribution

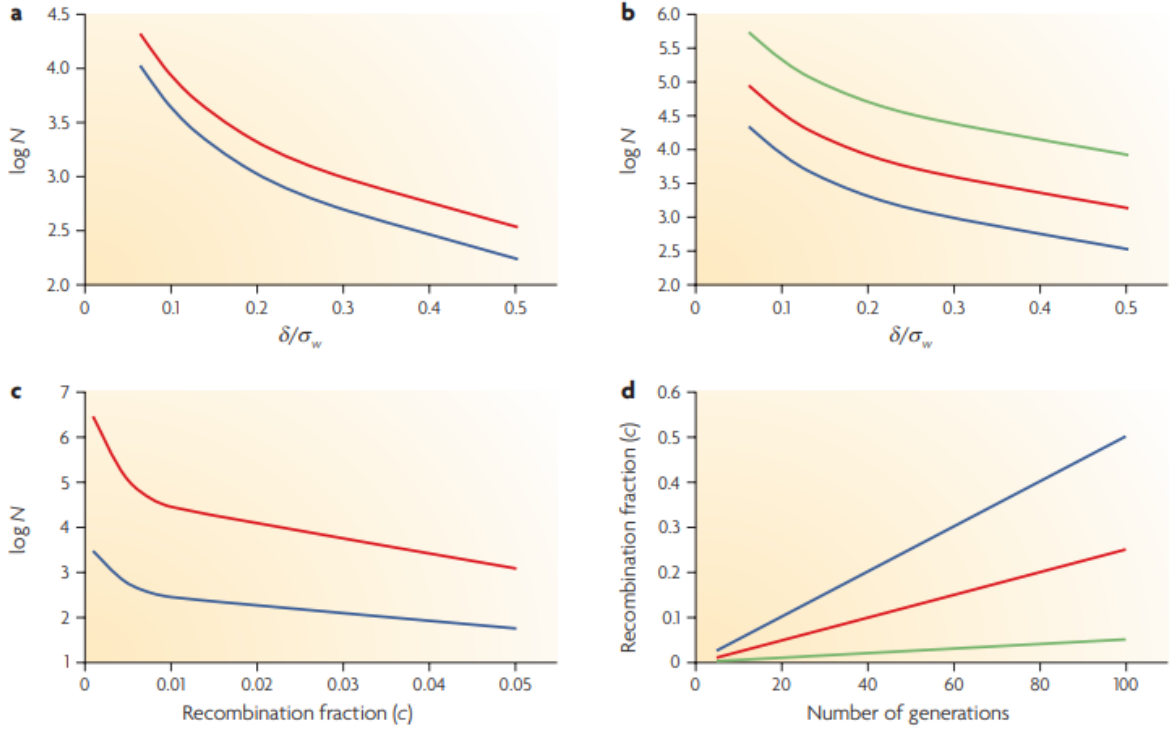
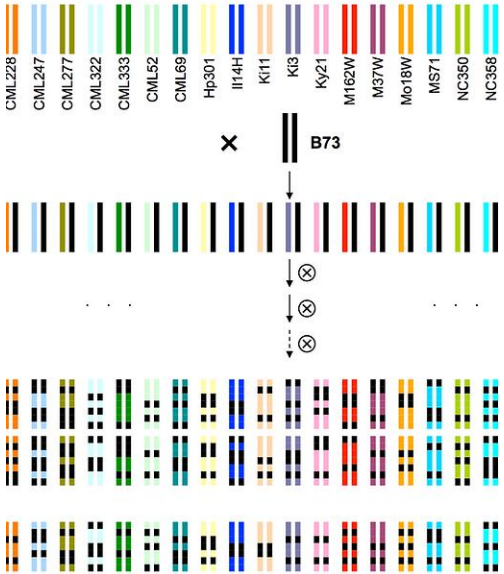


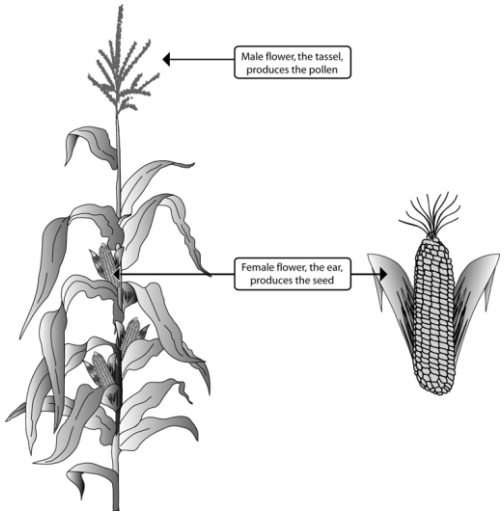
Figure 1 | **Power to localize and detect quantitative trait loci. a** | Numbers of individuals ($\log_{10}$ scale) required to detect quantitative trait loci (QTLs) for a range of effect sizes (δ/σ_w) in backcrossed (blue) and F_2 (red) linkage mapping populations. **b** | Numbers of individuals ($\log_{10}$ scale) required to detect QTLs for a range of effect sizes in association mapping populations in which the minor allele frequency is 0.5 (blue), 0.25 (red) and 0.1 (green). **c** | Log_{10} of the number of individuals required to detect at least one recombinant in an interval of size c (c = 100 centiMorgans; cM) (blue) and $\log_{10}$ of the number of marker genotypes needed to localize QTLs per 100 cM (red). **d** | The expected frequency of recombinants after t generations of recombination in a random mating population, for a per generation recombination fraction of c = 0.01 (blue), c = 0.005 (red) and c = 0.001 (green). δ, average difference in the trait phenotype between marker allele genotypes; σ_W, phenotypic standard deviation of the trait within marker genotype classes.

The Genetic Architecture Of Maize Height

Jason A. Peiffer,*,[1] Maria C. Romay,[†] Michael A. Gore,[‡] Sherry A. Flint-Garcia,[§,] Zhiwu Zhang,[†] Mark J. Millard,[††,‡‡] Candice A. C. Gardner,[††,‡‡] Michael D. McMullen,[§,**] James B. Holland,[§§,***] Peter J. Bradbury,[†††] and Edward S. Buckler[†,‡,†††]**

*Department of Genetics, Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina 27695, [†]Institute for Genomic Diversity and [‡]Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York, 14853, [§]United States Department of Agriculture -Agricultural Research Service, Columbia, Missouri 65211, **Division of Plant Sciences, University of Missouri, Columbia, Missouri 65211, [††]United States Department of Agriculture -Agricultural Research Service, Ames, Iowa, 50011, [‡‡]Department of Agronomy, Iowa State University, Ames, Iowa 50011, [§§]United States Department of Agriculture -Agricultural Research Service, Raleigh, North Carolina 27695, ***Department of Crop Science, North Carolina State University, Raleigh, North Carolina 27695, and [†††]United States Department of Agriculture -Agricultural Research Service, Ithaca, NY 14853





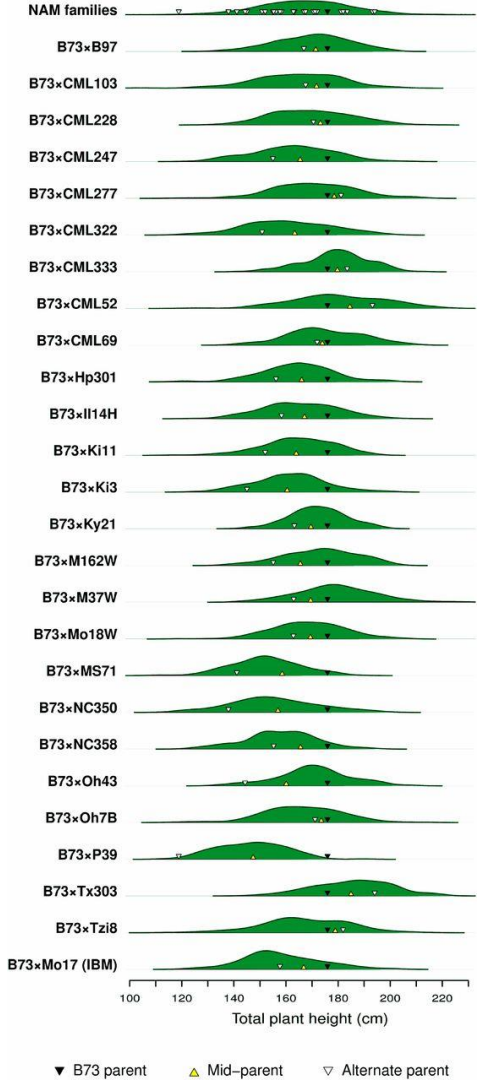


Table 1 Heritability estimated on a line mean basis

		Heritability			
Family/panel	Plots evaluated	PHT	EHT	DTA	NPH
NAM families	57,142	0.92	0.93	0.94	0.89
B73 × B97	1,924	0.93	0.93	0.84	0.85
B73 × CML103	1,914	0.91	0.91	0.87	0.91
B73 × CML228	1,770	0.92	0.93	0.94	0.89
B73 × CML247	2,006	0.93	0.93	0.93	0.87
B73 × CML277	2,020	0.92	0.92	0.94	0.88
B73 × CML322	1,859	0.92	0.92	0.91	0.90
B73 × CML333	1,892	0.93	0.93	0.94	0.91
B73 × CML52	1,860	0.92	0.92	0.92	0.88
B73 × CML69	1,911	0.93	0.94	0.89	0.86
B73 × Hp301	1,921	0.92	0.92	0.90	0.91
B73 × Il14H	1,805	0.93	0.93	0.91	0.93
B73 × Ki11	1,905	0.93	0.94	0.94	0.89
B73 × Ki3	2,041	0.92	0.92	0.93	0.91
B73 × Ky21	1,918	0.93	0.93	0.84	0.91
B73 × M162W	2,023	0.92	0.92	0.91	0.92
B73 × M37W	1,942	0.91	0.91	0.89	0.93
B73 × Mo18W	1,830	0.92	0.93	0.93	0.92
B73 × MS71	1,896	0.92	0.92	0.89	0.91
B73 × NC350	1,841	0.93	0.94	0.92	0.89
B73 × NC358	1,861	0.92	0.92	0.86	0.88
B73 × Oh43	1,920	0.95	0.94	0.81	0.90
B73 × Oh7B	1,890	0.94	0.95	0.90	0.91
B73 × P39	1,876	0.92	0.93	0.95	0.84
B73 × Tx303	1,678	0.94	0.94	0.92	0.89
B73 × Tzi8	2,107	0.94	0.95	0.92	0.93
B73 × Mo17(IBM)	1,989	0.93	0.94	0.92	0.91
NCRPIS diversity panel	7,471	0.87	0.86	0.92	NA

Plots evaluated detail the number of plots scored for PHT across all environments. The other surveyed traits possessed comparable values within each family or panel with the exception of NPH, which was not scored in the NCRPIS diversity panel. PHT, DTA, EHT, and NPH detail the proportion of variance between and within lines explained by between line variance after accounting for known environmental variation in the respective trait.

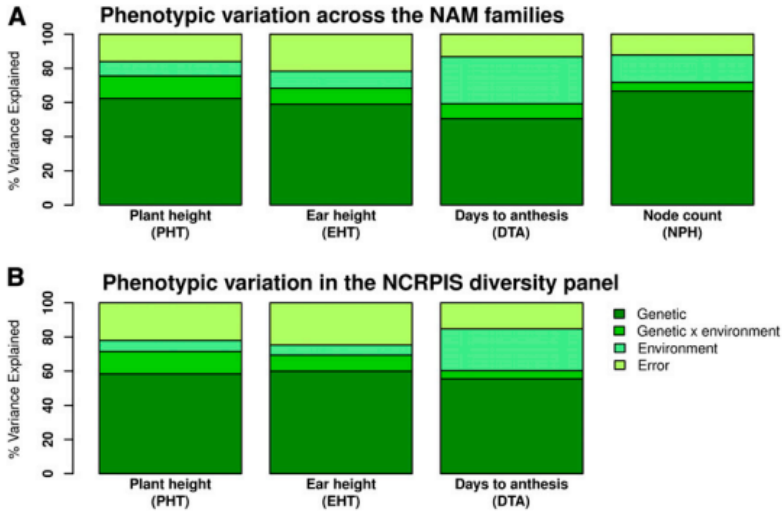
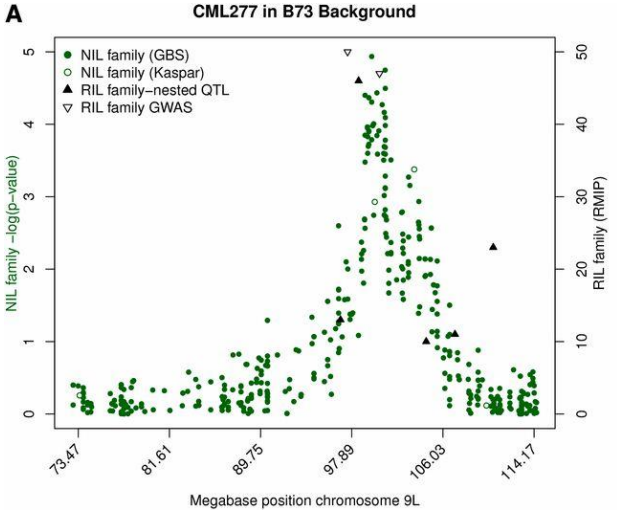
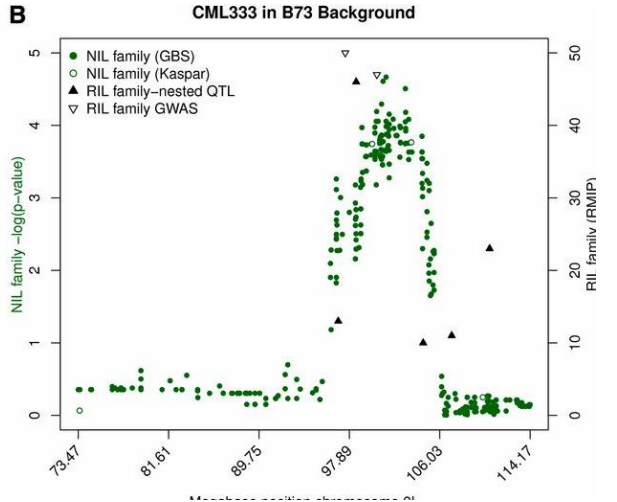


Table 2 Top height-associated family-nested QTL across RIL families

Chr	Mb	cM	Combined RMIP				Nearby annotations of interest
			PHT	EHT	DTA	NPH	
1	10	20	54	64	0	0	
1	29	47	36	0	0	0	
1	66	75	47	25	17	0	
1	83	82	45	64	95	51	
1	184	99	54	0	0	59	
1	204	117	43	44	40	0	
1	249	148	71	66	0	0	*brassinosteroid-deficient dwarf1* (Pettem 1956)
2	1	0	46	32	0	0	
2	3	7	40	15	0	0	
2	90	76	44	11	76	31	
3	5	21	12	0	0	0	*crinkly leaves1* (Beavis W *et al.* 1991)
3	10	34	34	23	0	0	
3	24	52	64	54	33	67	
3	160	73	67	26	78	78	
4	148	62	44	53	0	0	
4	235	115	52	43	12	0	
5	89	70	27	51	40	0	
5	201	109	69	91	0	11	
6	92	19	21	20	0	0	
6	96	22	77	51	12	0	
6	141	55	27	28	0	12	
6	147	58	21	0	0	0	
7	33	48	56	16	0	67	
7	135	73	52	61	37	17	
7	143	81	22	0	0	0	
7	152	89	27	0	0	0	
7	155	95	58	23	0	0	
8	22	49	24	19	0	0	
8	121	64	69	91	98	97	
9	99	50	83	96	0	15	
9	111	55	34	19	47	40	
9	133	69	64	17	0	0	
10	5	15	19	11	15	7	*crinkly leaves4* (Stinard and Robertson 1987)
10	140	69	26	0	0	50	
10	147	91	36	13	0	0	

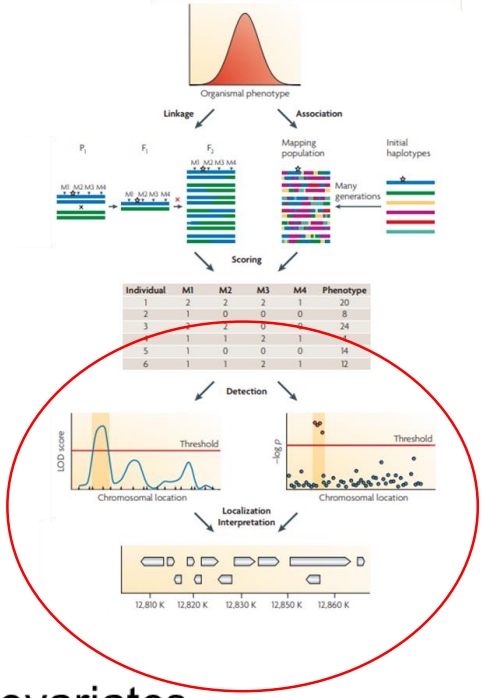
The combined resample model inclusion probability (RMIP) details the number of models one or more markers located within 3 cM of the stated association was selected out of the 100 models constructed for each trait (PHT, EHT, DTA, NPH). Each of the 100 models was calibrated from a family-stratified sampling of RILs during bootstrapped joint-linkage mapping. Mb denotes megabase positions in maize RefGenV1. cM denotes centimorgan positions of the composite NAM family genetic map.

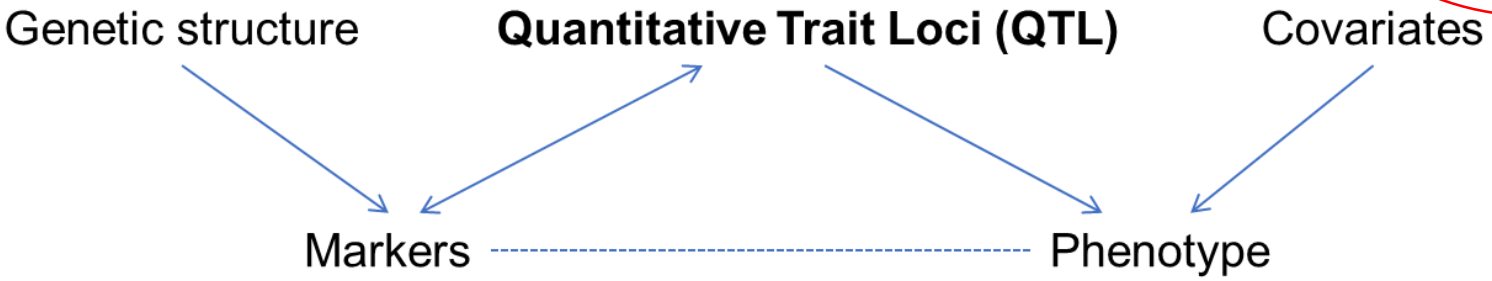




From QTL to genes

- Typically QTL regions identified contain many genes/genetic factors
- Molecular markers are a proxy of genetic factors to which they are associated through linkage disequilibrium (LD)





Euphytica **91:** 89–97, 1996.

Mapping QTLs in breeding for drought tolerance in maize (*Zea mays* L.)

Hesham A. S. Agrama[1,*] & Mounir E. Moussa[2]

[1]*Maize Research Department, Agricultural Research Centre, Field Crops Research Institute, Egypt;* [2]*Department of Genetics, Faculty of Agriculture, Alexandria University, Egypt; (*author for correspondence)*

Received 11 August 1995; accepted 15 March 1996

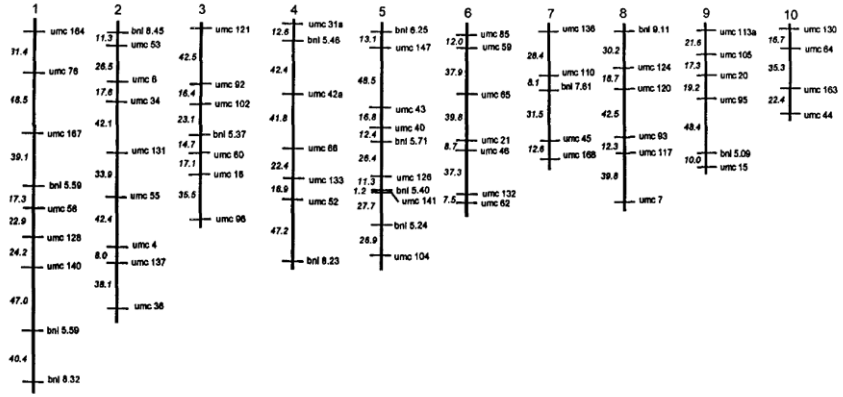


Figure 2. Genetic linkage map was generated using an F_3 population derived from SD34 x SD35. The map included 70 RFLP loci, scored in about 120 individulas, and linkage analysis was done using MAPMAKER v2.0.

Summary

Grain yield in the maize (*Zea mays* L) plant is sensitive to drought in the period three weeks either side of flowering. Maize is well-adapted to the use of restriction fragment length polymorphisms (RFLPs) to identify a tight linkage between gene(s) controlling the quantitative trait and a molecular marker. We have determined the chromosomal locations of quantitative trait loci (QTLs) affecting grain yield under drought, anthesis-silking interval, and number of ears per plant. The F_3 families derived from the cross SD34(tolerant) x SD35(intolerant) were evaluated for these traits in a two replicated experiment. RFLP analysis of the maize genome included non-radioactive DNA-DNA hybridization detection using chemiluminescence. To identify QTLs underlying tolerance to drought, the mean phenotypic performances of F_3 families were compared based on genotypic classification at each of 70 RFLP marker loci. The genetic linkage map assembled from these markers was in good agreement with previously published maps. The phenotypic correlations between yield and other traits were highly significant. In the combined analyses, genomic regions significantly affecting tolerance to drought were found on chromosomes 1,3,5,6, and 8. For yield, a total of 50% of the phenotypic variance could be explained by five putative QTLs. Different types of gene action were found for the putative QTLs for the three traits.

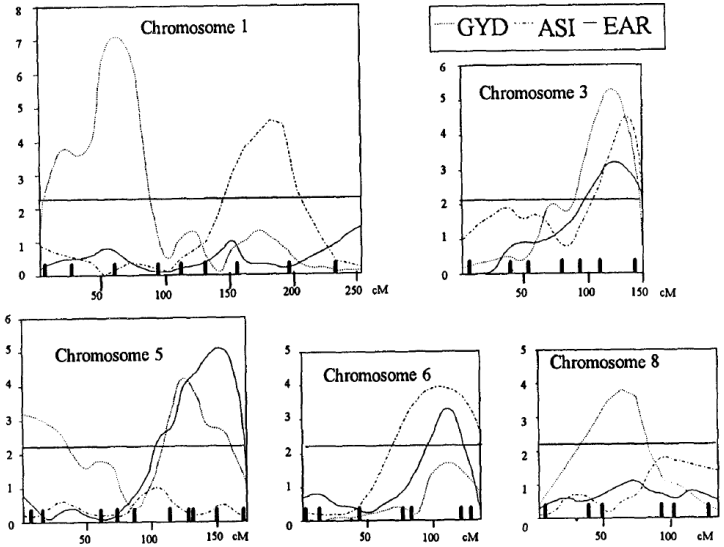


Figure 3. QTL likelihood maps indicating LOD score for grain yield under drought (GYD), anthesis-silking-interval (ASI), and ears per plant (EAR). The horizontal line at a height of 2.2 indicates the stringent threshold that the LOD score must cross to allow the presence of a QTL to be inferred.

RESEARCH Open Access

Using high-throughput mult phenotyping to decipher the architecture of maize drough

Xi Wu[1,2†], Hui Feng[1†], Di Wu[1], Shijuan Yan[3], Pei Zhang[1], Wenbin Yuan Fan[1], Weikun Li[1], Baoxing Song[4], Zedong Geng[1], Wanli Yan Michelle Stitzer[4], Lin Li[1], Lizhong Xiong[1,2], Jianbing Yan[1,2], Edwar Mingqiu Dai[1,2*]

Abstract

Background: Drought threatens the food supply the dynamic responses of plants to drought will tolerant crops, as the genetic controls of these re

Results: Here we develop a high-throughput mu to noninvasively phenotype 368 maize genotype over a course of 98 days, and collected multiple camera scanning, hyperspectral imaging, and X-r We develop high-throughput analysis pipelines t Of these i-traits, 10,080 were effective and herita internal drought responses. An i-trait-based genc 4322 significant locus-trait associations, represent (QTLs) and 2318 candidate genes, many that co- maize drought responsive QTLs. Expression QTL and distant regulatory variants that control the e We use genetic mutation analysis to validate two *ZmFAB1A*, which regulate i-traits and drought tol candidate genes as drought-tolerant genetic ma selection analysis, and 15 i-traits are identified as tolerance breeding.

Conclusion: Our study demonstrates that combi optical phenotyping and GWAS is a novel and e genetic architecture of complex traits and clone