INSTITUTE OF PLANT SCIENCES

Sant'Anna School of Advanced Studies – Pisa

Considerations on breeding for local adaptation

Experimental agriculture is a process based on development of *products*

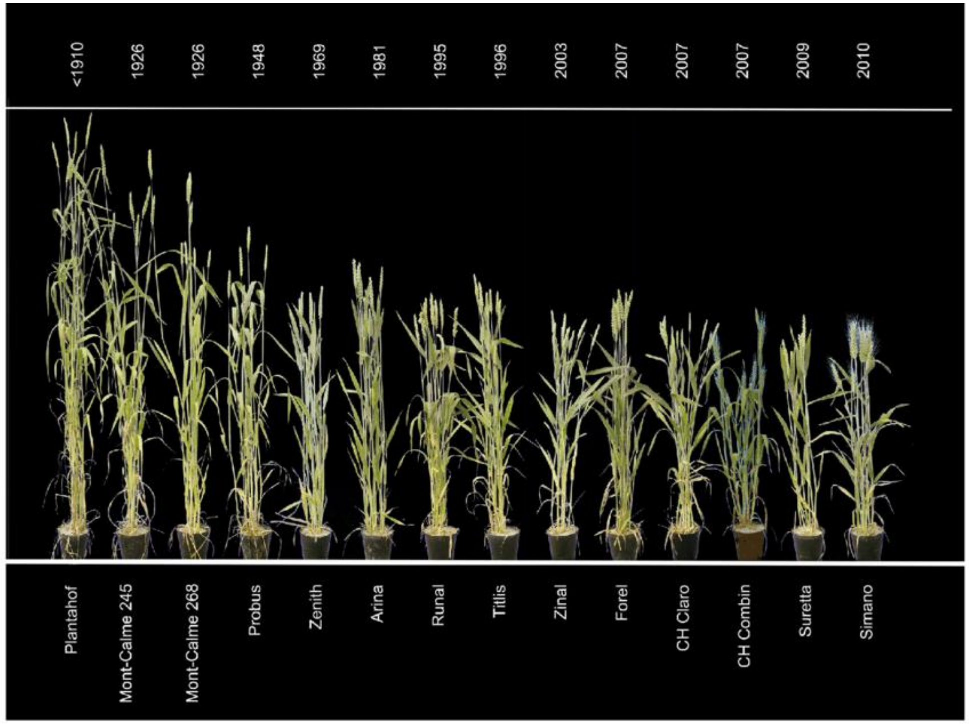
Innovation is to be directed to improvement of farming for the benefit of farmers, consumers, and environment

When thinking of a new variety or a new farming practice, the target users are **farmers**

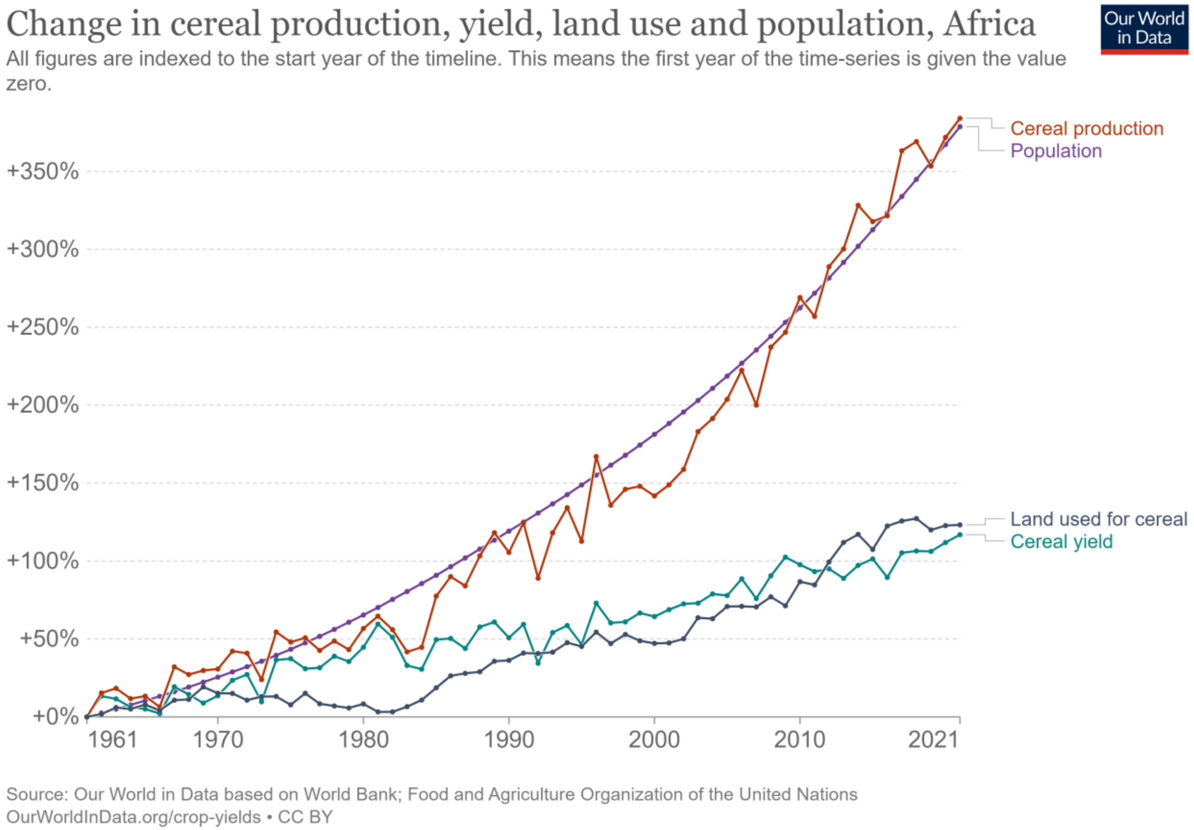
Product profiling (*i.e.* design) is a key issue to achieve impact

- Identify a need
- Consider who will use the product and why
- Is it clear how to use it?
- Is it engaging?

- Crop breeding is tasked with developiong producs – *varieties* – originating from new genetic combinations. A neverending quest for yield, resistance, quality…
- Modern crop breeding is largely a legacy of the Green Revolution

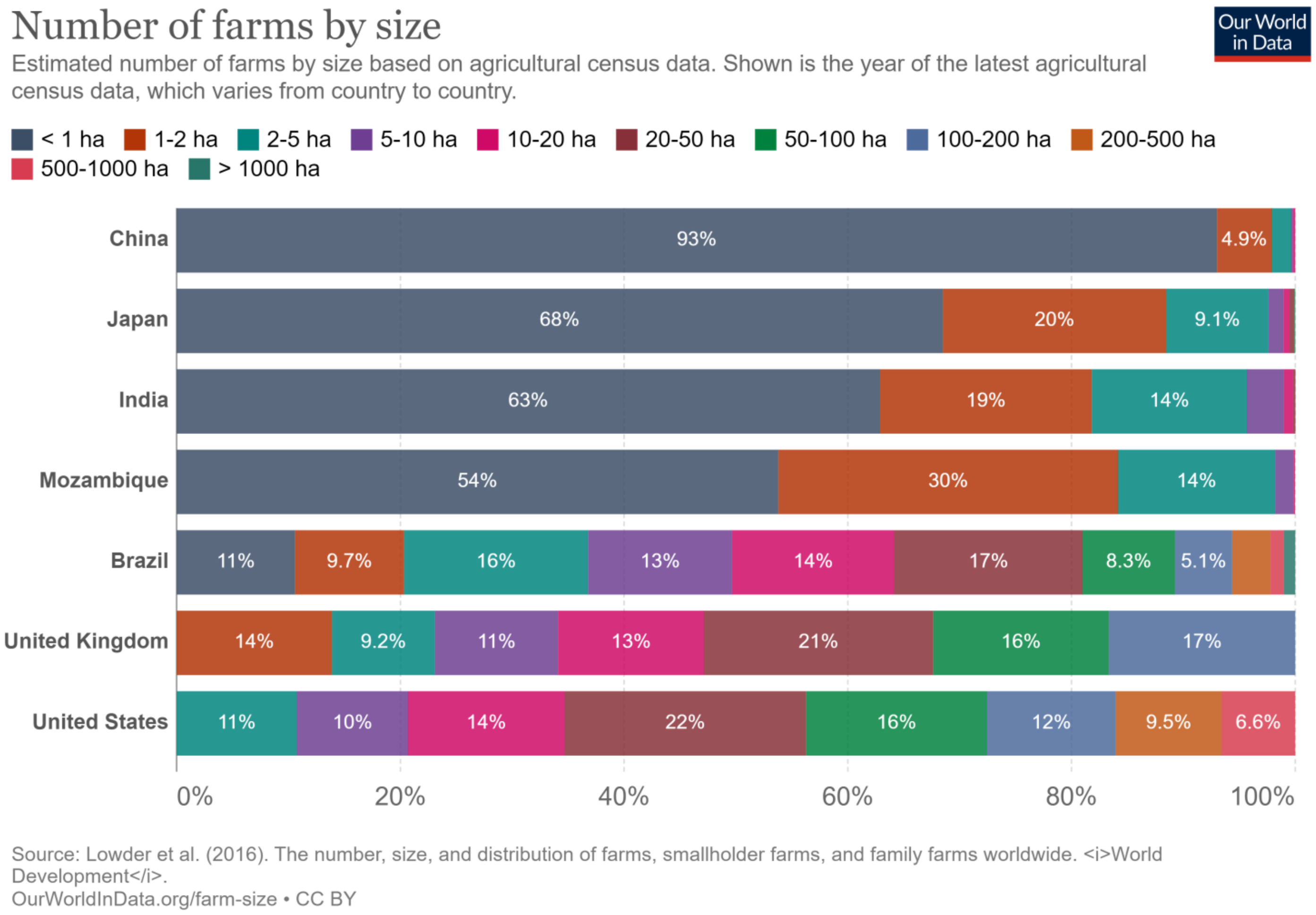


Friedli et al (2019)



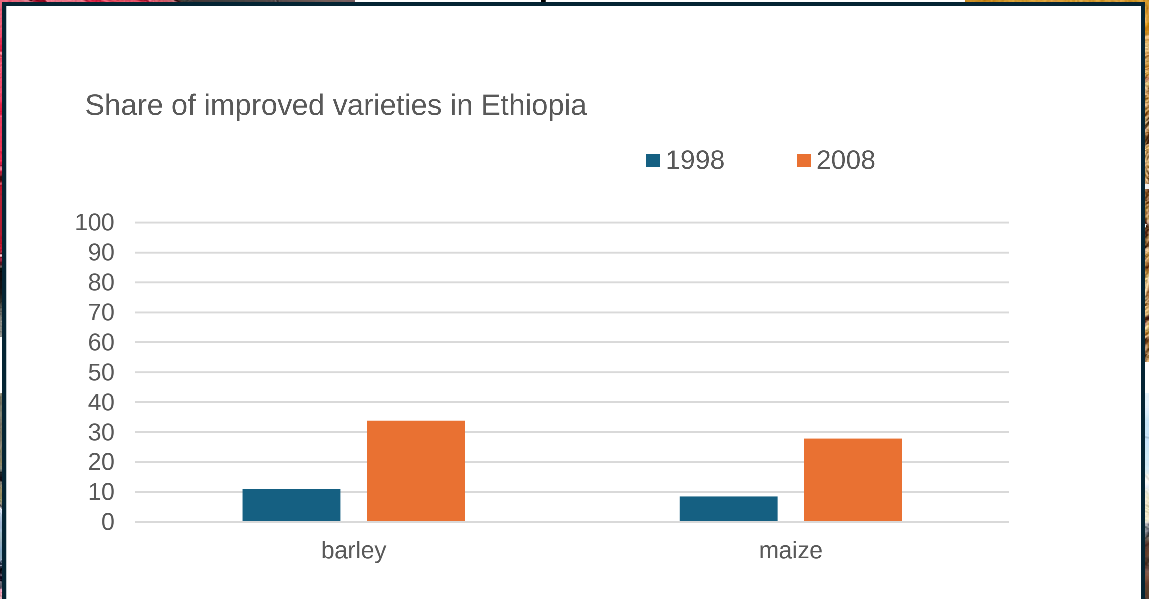
«Conventional» farming: uniform environment, high-input, high-tech

Smallholder farming: heterogeneous, low-input, low-tech



Globally, about 570M smallholder farmers support the livelihoods of 2B people; small farms produce 1/3 of global food

- Good looking and high performing
- Requires lots of



Shehan and Barrett (2014)

specifications

Molecular breeding for local adaptation

Identification and manipulation of genetic factors to speed up the production of new varieties for:

1. High yield potential (+ yield stability) under low inputs
2. Potential for adaptation to current and future climate
3. High nutritional value and cultural heritage

Agrobiodiversity

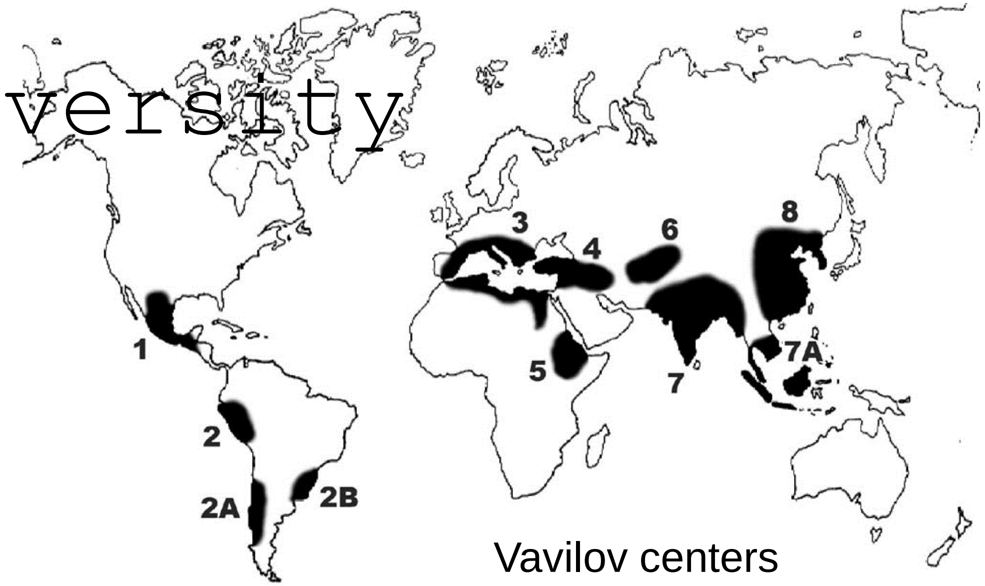
Genomics

Genetics

First pillar: agrobiodiversity

Cherinet's mother, serving coffe
Bahir Dar Ethiopia

No diversity? No improvement!

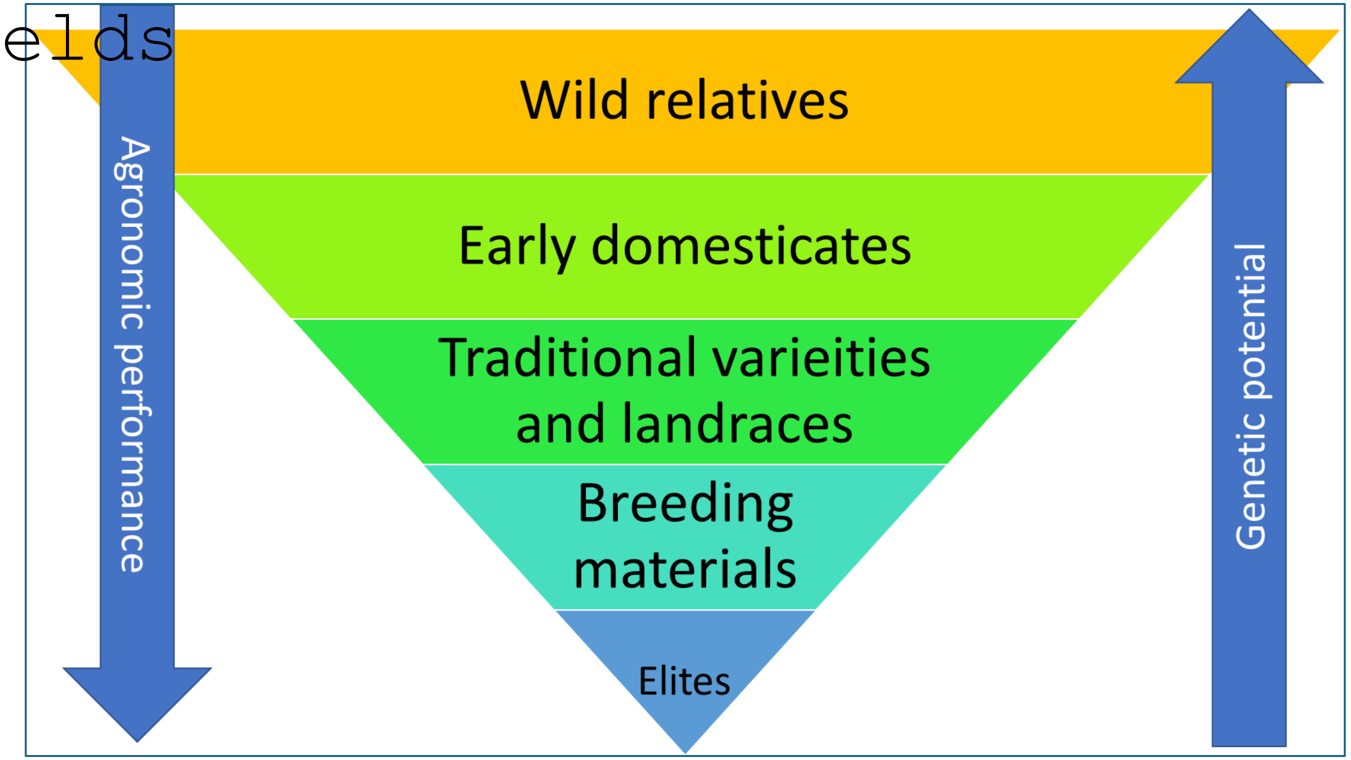


Vavilov centers

- Agrobiodiversity is nature + culture; biodiversity that has been shaped by human ingenuity
- Agrobiodiversity is the raw material that breeders shape into new and improved varieties
- The green revolution is really the most recent step

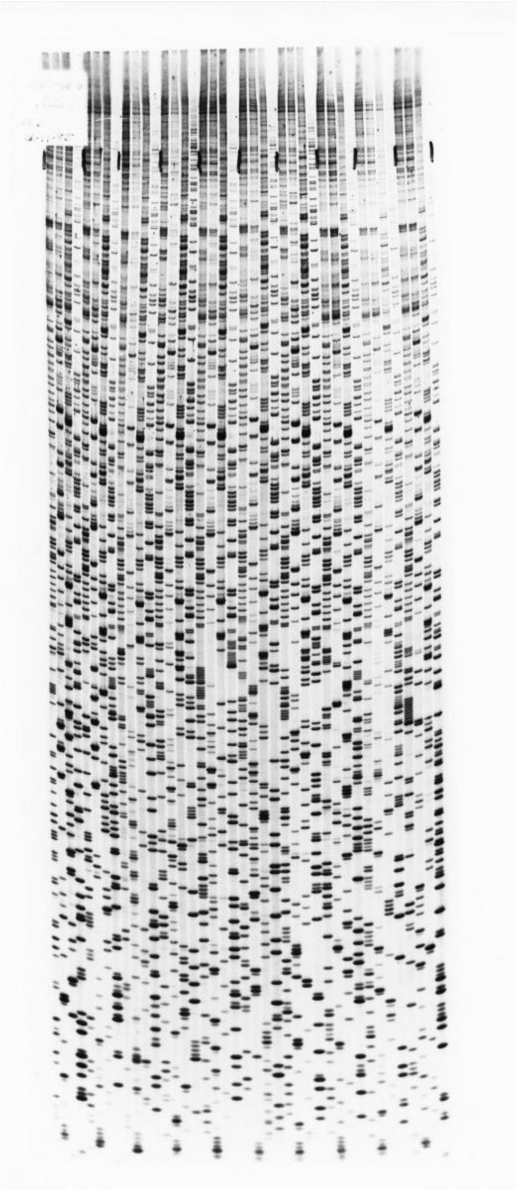
Norman Borlaugh
Nobel laureate

Current varieties are the result of thousands of years of selection, which limits the amount of diversity in current fields

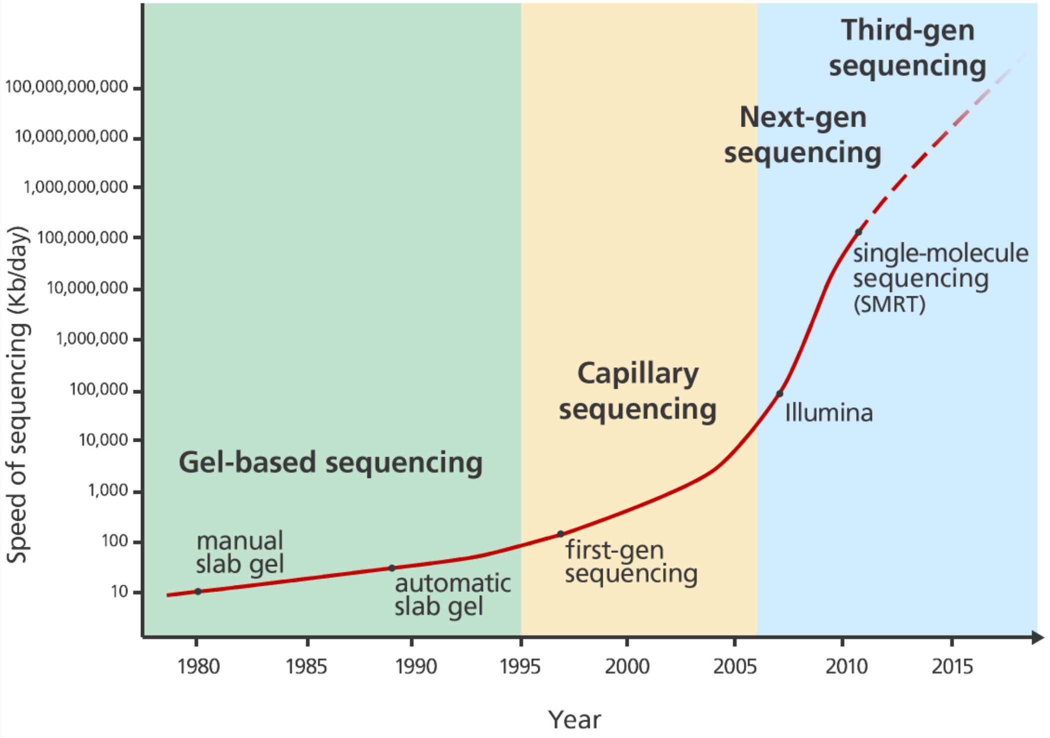


Untapped genetic agrobiodiversity can be sourced from seed banks and farmer fields

Second pillar: genomics



- Once you have diversity, you need the tools to characterize it
- Genomic technology evolved rapidily in the last 20+ years; DNA sequencing is an everyday task



It is now easy and cheap to produce tons of genomic data, and the future is bright!

BIOTECHNOLOGY

Multiple rereads of single proteins at single–amino acid resolution using nanopores

Henry Brinkerhoff[1], Albert S. W. Kang[1], Jingqian Liu[2], Aleksei Aksimentiev[2], Cees Dekker[1]*

A proteomics tool capable of identifying single proteins would be important for cell biology research and applications. Here, we demonstrate a nanopore-based single-molecule peptide reader sensitive to single–amino acid substitutions within individual peptides. A DNA-peptide conjugate was pulled through the biological nanopore MspA by the DNA helicase Hel308. Reading the ion current signal through the nanopore enabled discrimination of single–amino acid substitutions in single reads. Molecular dynamics simulations showed these signals to result from size exclusion and pore binding. We also demonstrate the capability to "rewind" peptide reads, obtaining numerous independent reads of the same molecule, yielding an error rate of $<10^{-6}$ in single amino acid variant identification. These proof-of-concept experiments constitute a promising basis for the development of a single-molecule protein fingerprinting and analysis technology.

Figure 1: Process of nanopore sequencing in the Ecuadorian Chocó rainforest. (A) Sampling endemic fauna; eyelash pitviper next to MinION. (B) Extraction of blood or tissue samples. (C) DNA extraction using the DNeasy kit and benchtop centrifuge, and PCR amplification with the MiniPCR. (D) Oxford nanopore library preparation

RESEARCH

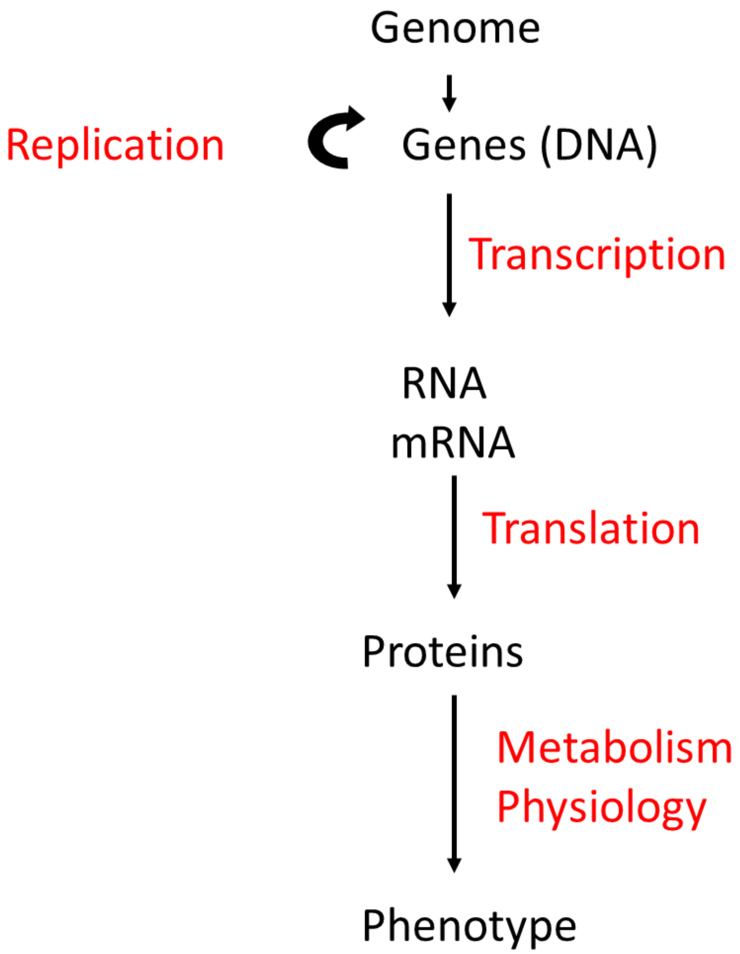
Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building

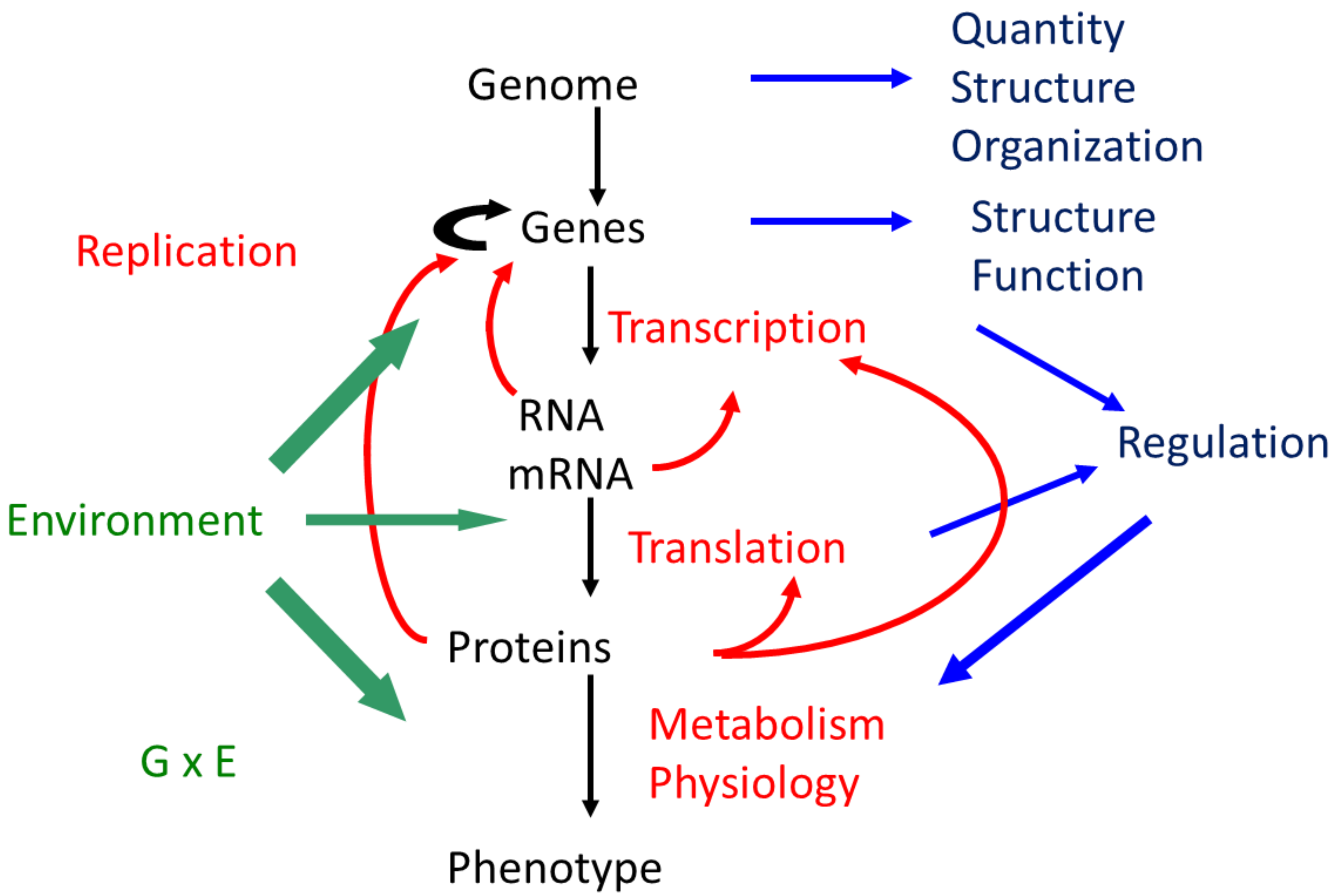
Aaron Pomerantz[1,*], Nicolás Peñafiel[2], Alejandro Arteaga[3,4,5], Lucas Bustamante[5], Frank Pichardo[5], Luis A. Coloma[6], César L. Barrio-Amorós[7], David Salazar-Valenzuela[2] and Stefan Prost[1,8,*]

[1]Department of Integrative Biology, University of California, Berkeley, CA, USA, [2]Centro de Investigación de la Biodiversidad y Cambio Climático (BioCamb) e Ingeniería en Biodiversidad y Recursos Genéticos, Facultad de Ciencias de Medio Ambiente, Universidad Tecnológica Indoamérica, Machala y Sabanilla, Quito, Ecuador, [3]Richard Gilder Graduate School, American Museum of Natural History, New York, USA, [4]Department of Herpetology, American Museum of Natural History, New York, USA, [5]Tropical Herping, Quito, Ecuador, [6]Centro Jambatu de Investigación y Conservación de Anfibios, Fundación Otonga, Quito, Ecuador, [7]Doc Frog Expeditions, Uvita, Costa Rica and [8]Program for Conservation Genomics, Department of Biology, Stanford University, Stanford, CA, USA

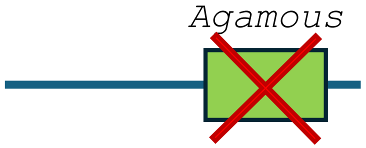
Third pillar: genetics

Once you can read DNA, the challenge is to understand what it is doing





Reverse genetics



Gene(s)

What trait arises from the perturbation of a DNA sequence?

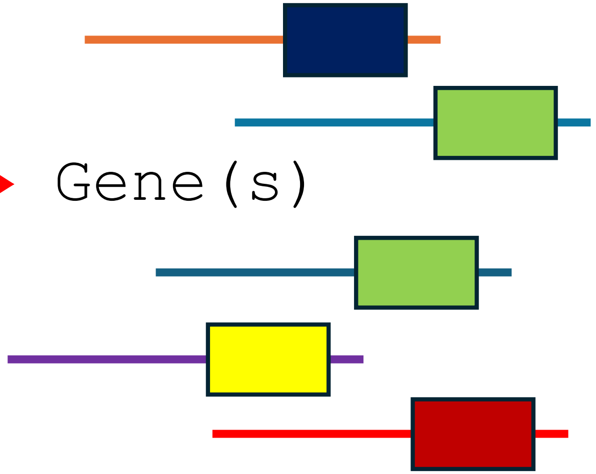
Trait(s)

Forward genetics

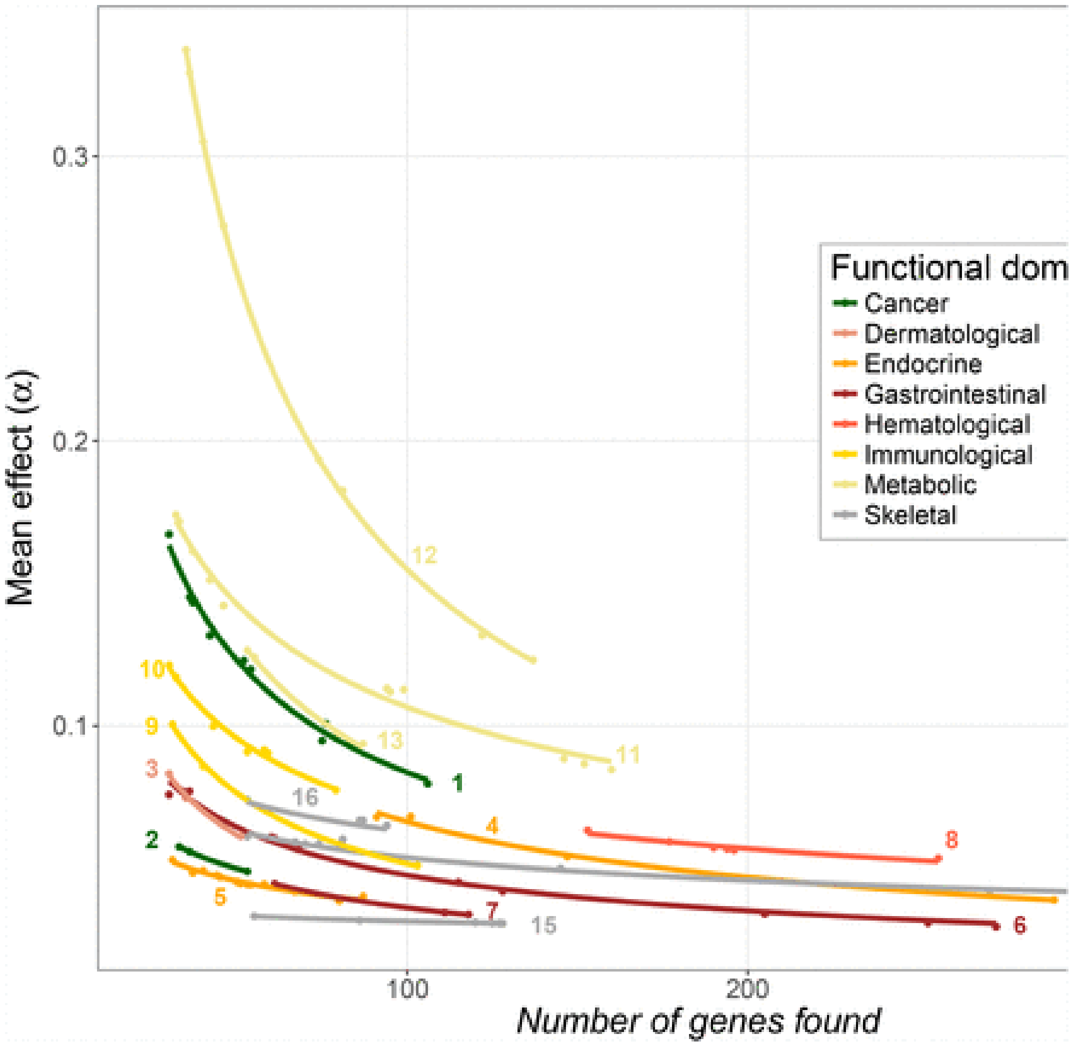
Trait(s)

Is variation of a trait associated with genotypic variation?

Gene(s)

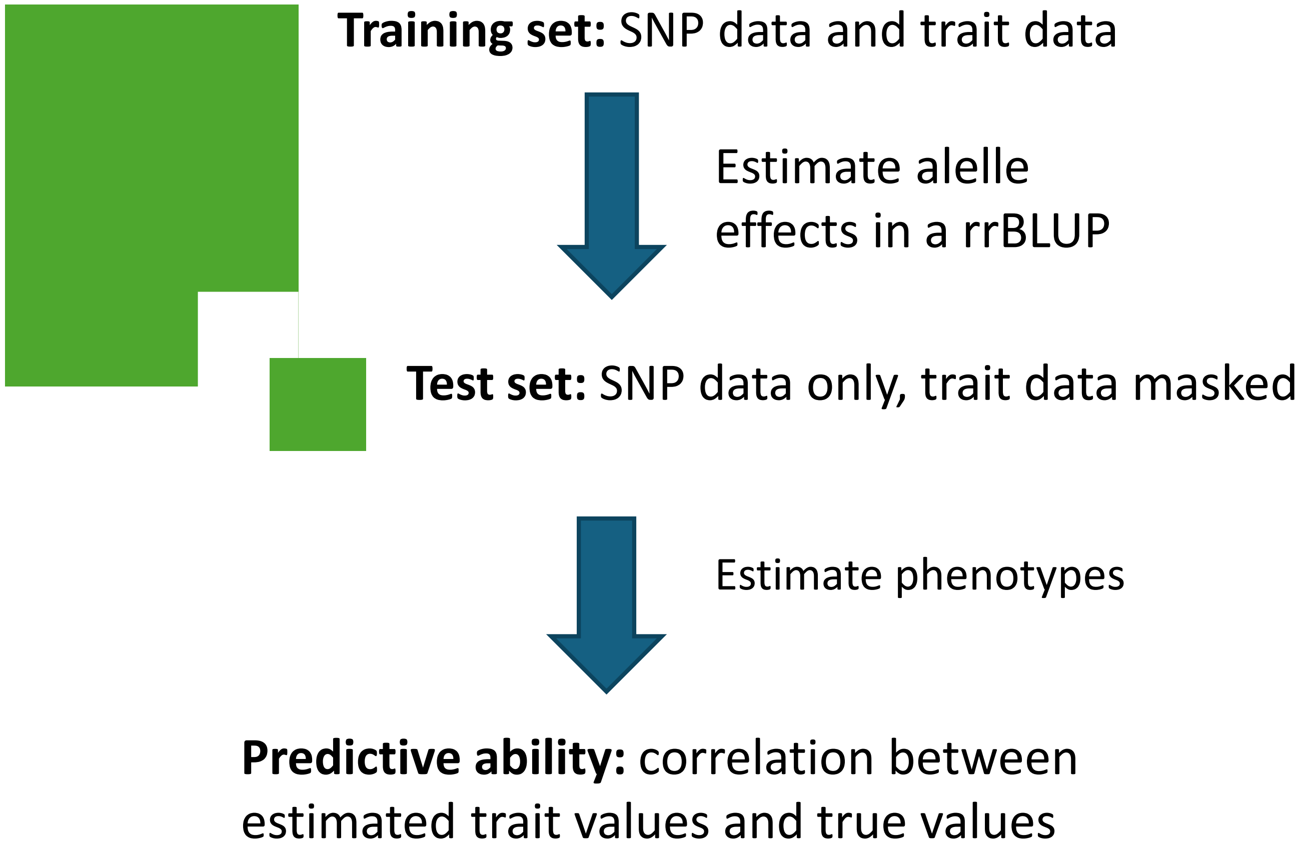


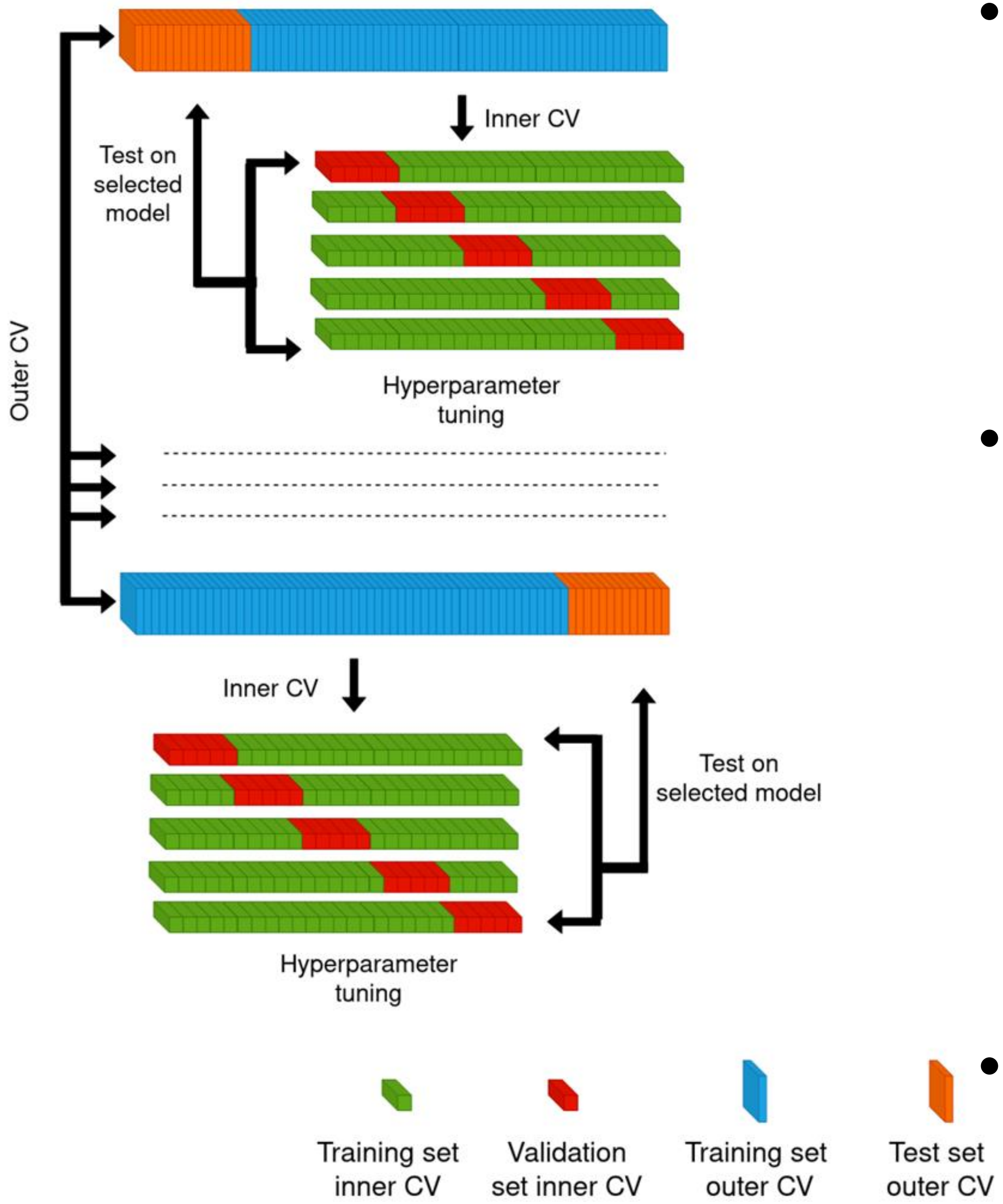
- It is becoming increasingly clear that traits are controlled by manifold, small effect loci
- Quantitative genetic mapping studies are tipically underpowered to capture small effects (few cases, many variables)
- Large human studies (e.g. UK BioBank) are filling in the gap



Predictive genomics

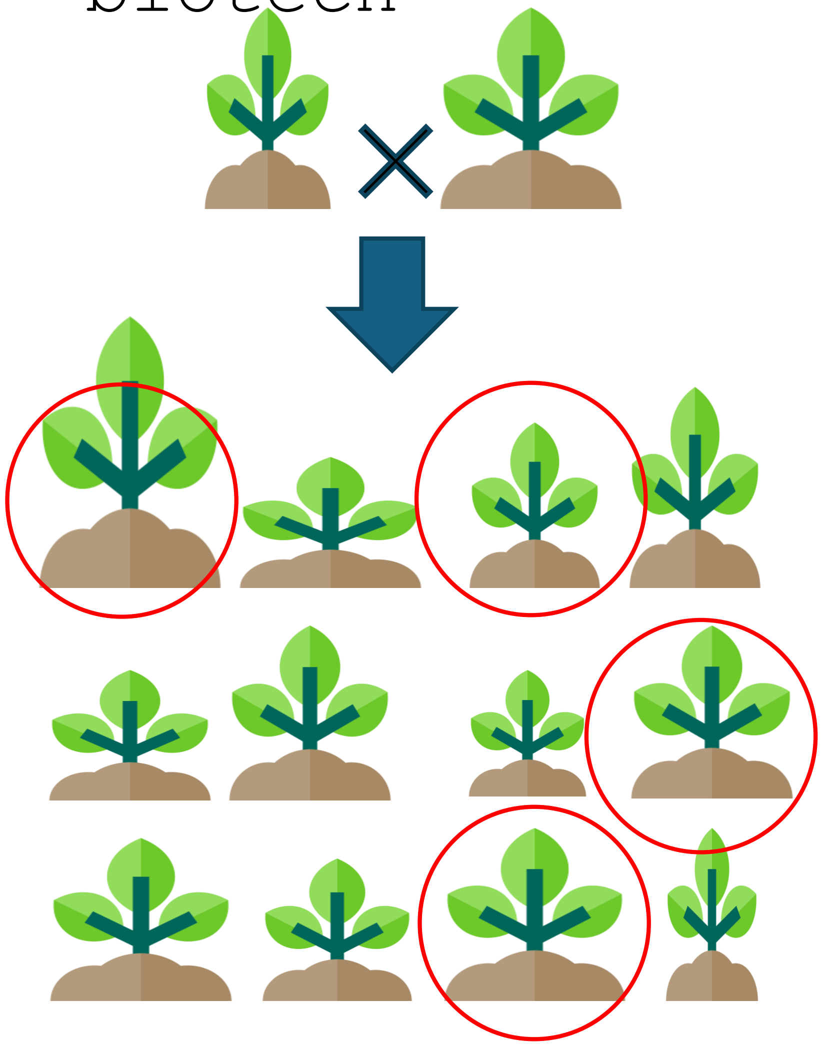
- It is possible to leverage big data to build simple models predicting outcomes (phenotypes) given a set of starting conditions (genotypes)
- Genomic selection / genomic prediction is usually based on:
  - Training set: in which individuals are genotyped AND phenotyped and a model is build to relate these quantities
  - Test set: in which

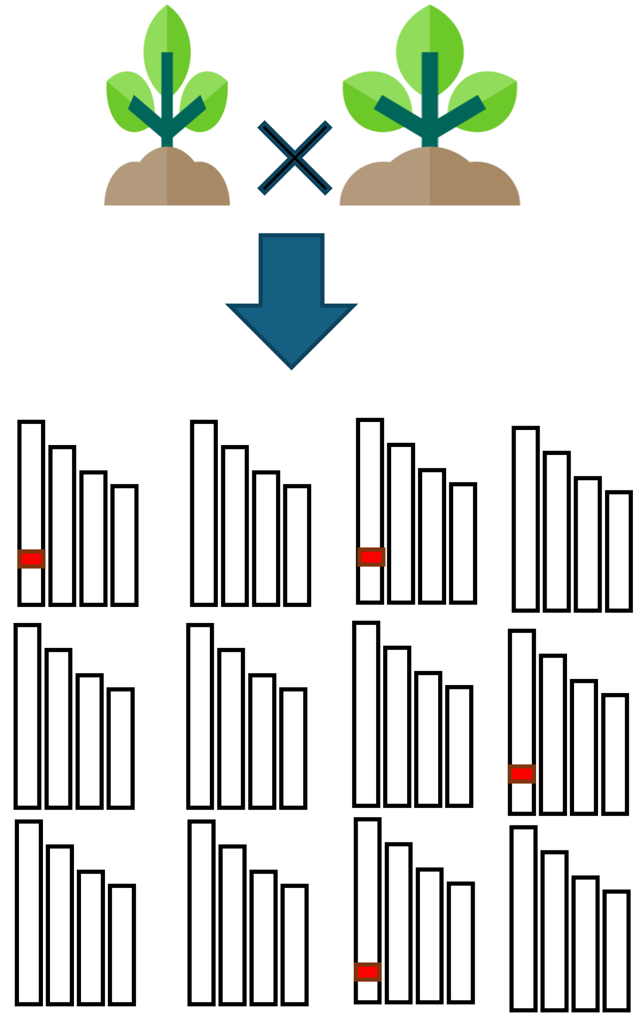


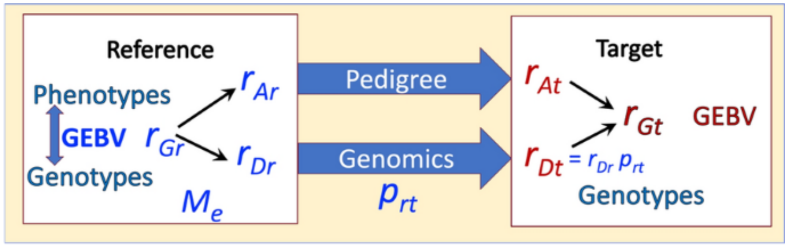


- In whole-genome predictions, each SNP value is associated with an estimated (infinitesimal) effect on the phenotype
- It’s a black box strategy: we don’t care about where genes are, what they do, what is the molecular basis of traits -> we just care about the numerical association between alleles and phenotypes
- The effectiveness of predictions is tipically assessed via cross

From phenotypic selection to molecular selection to biotech







Genomic selection

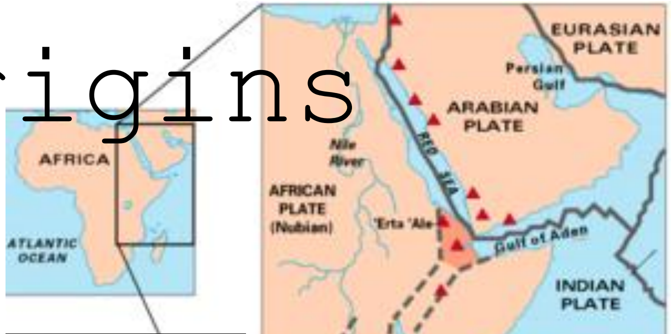
Modifying genes

Selection based on traits

Selection based on markers


How to better align breeding innovation to end user needs?

Ethiopia, the land of origins



- Many endemisms, wide topographic variation
- Main settlements b/w 2,000 and 2,500 masl
- 10 ecosystems, 49 agroecologies
- **43% of GDP** comes from farming
- **90% of farmers are smallholders** (80% of the total workforce)

Denakel depression (-115 m)

Broken plateau, canyons

Ras Dashen (4,533

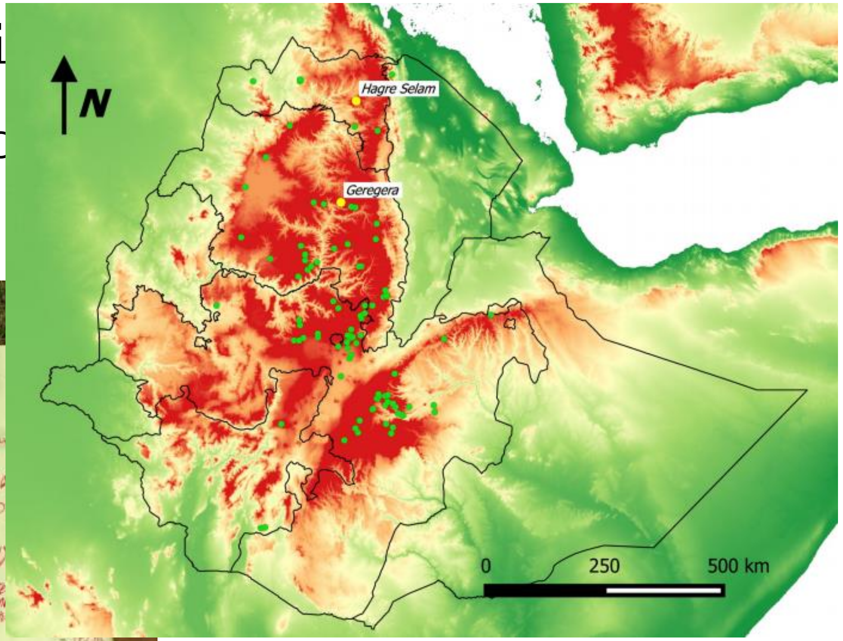
Durum wheat

Triticum turgidum, tetraploid wheat (subgenomes A, B)

- Independent domestication in Ethiopia? (debated)
- Cultivated for traditional preparations
- Estimated 4.2 Million farmers (13% of cereal growers)
- Low mean productivity **2.7 ton/ha**

Sampling Ethiopian durum wheat diversity

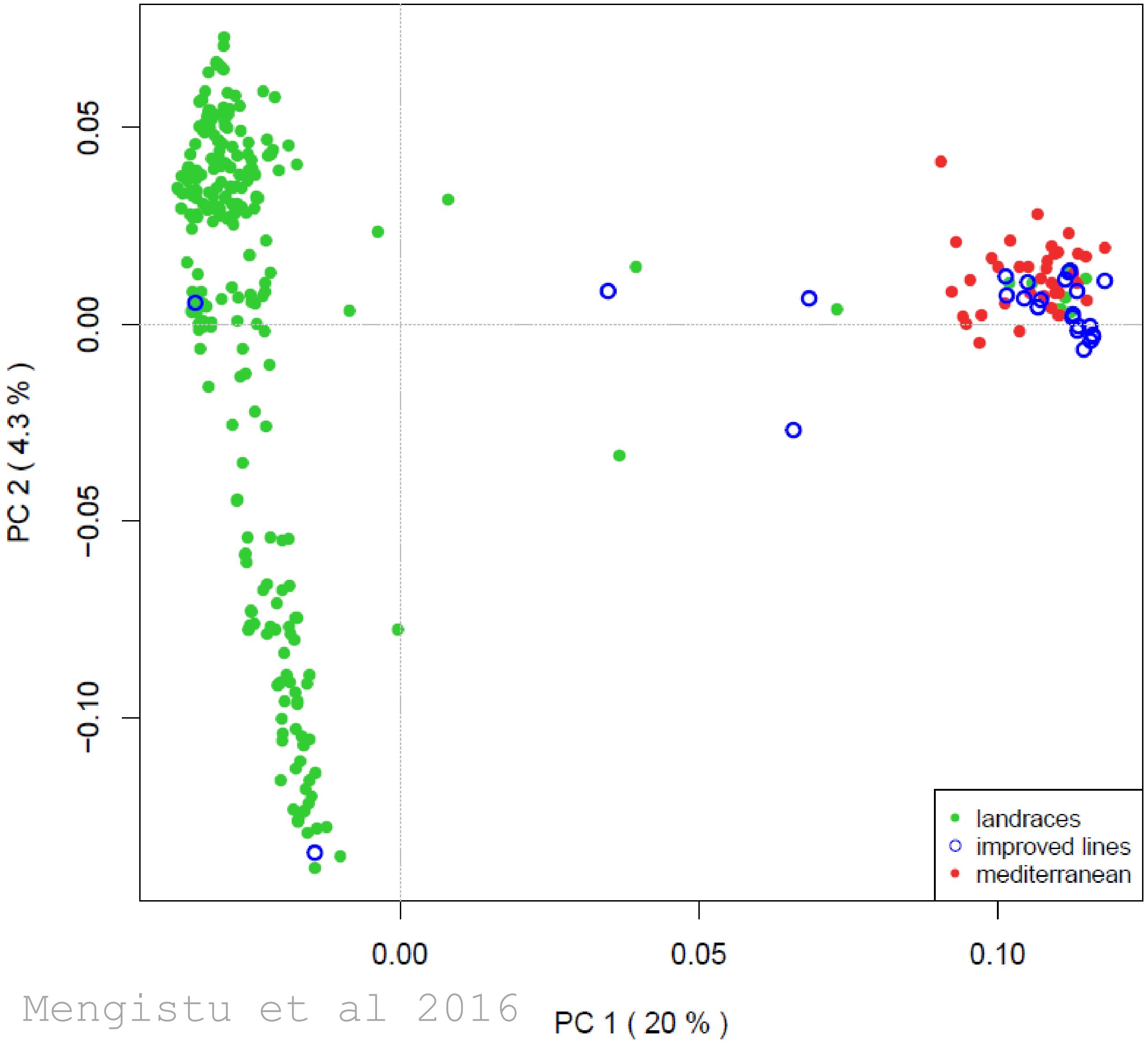
- Genetic materials selected on the basis of passport data of *ex situ* collections at the Ethiopian Biodiversty Institute (EBI)
- Purification plot prior characterizati
- Dominant types selected: accessions sp need
-



Dejene Mengistu, PhD

- Genotyping with the illumina 90K SNP array
- Diversity panel collection:
  - 298 traditional durum wheat landraces
  - 25 improved durum wheat lines release for cultivation
  - Mediterranean durum wheats

Breeding materials cultivated in Ethiopia lack Ethiopian heritage

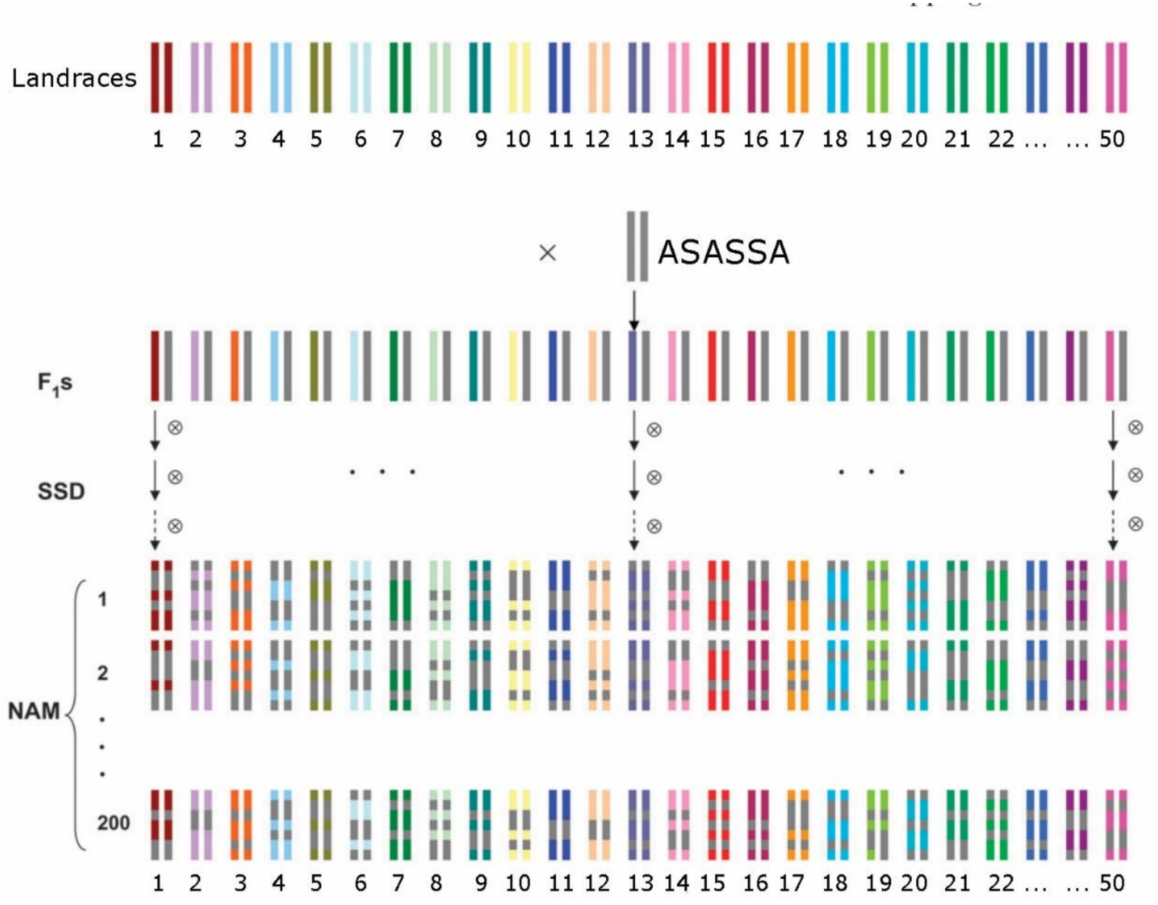


Mengistu et al 2016

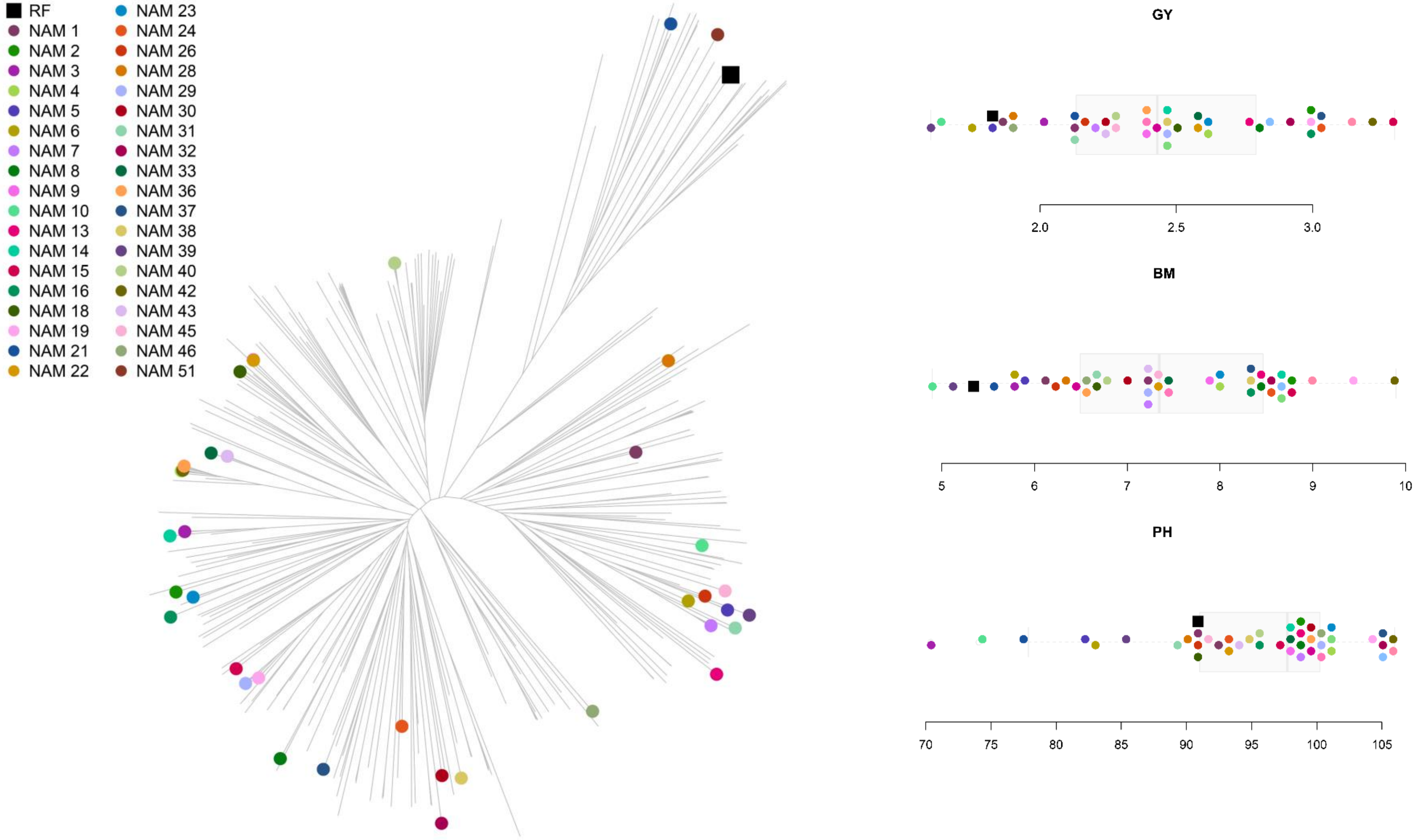
Moving closer to breeding: the Ethiopian NAM (EtNAM)

Yosef Kidane, PhD

- The Diversity of traditonal Ethiopian wheat may be useful for local and international breeding
- A nested association mapping (NAM) design may be used to recombine local diversity with an improved background, producing at once:
  1. Prebreeding materials
  2. A segregant



Kidane et al 2019

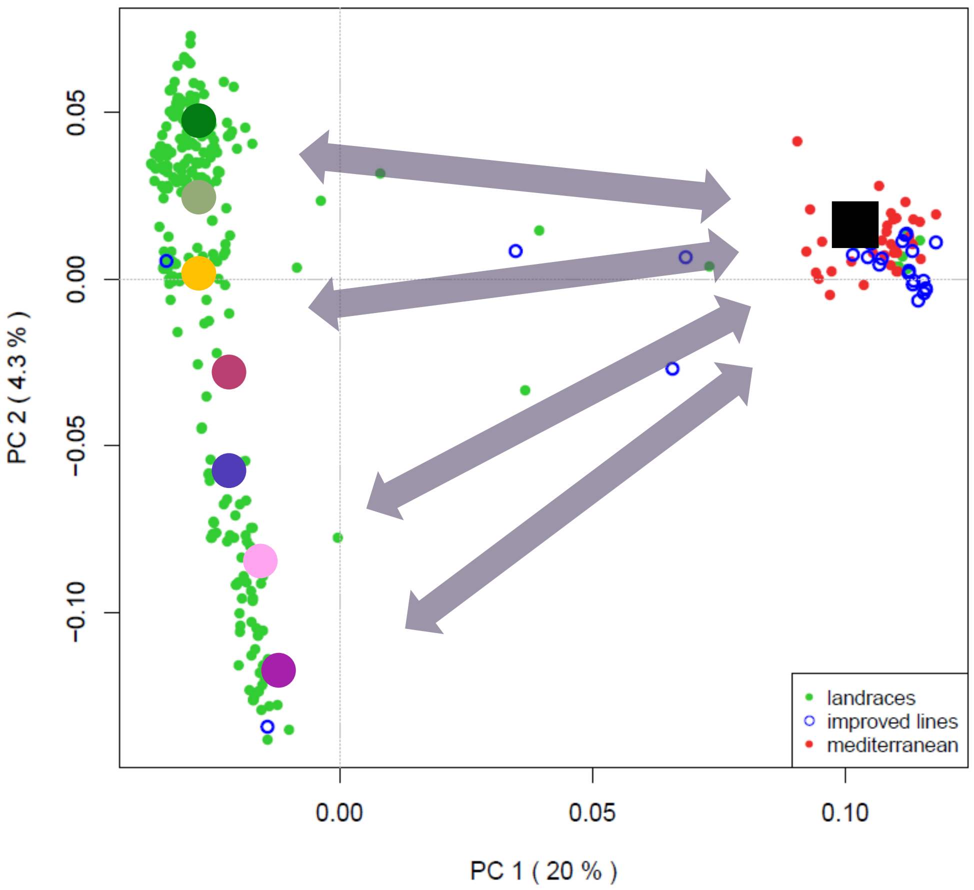


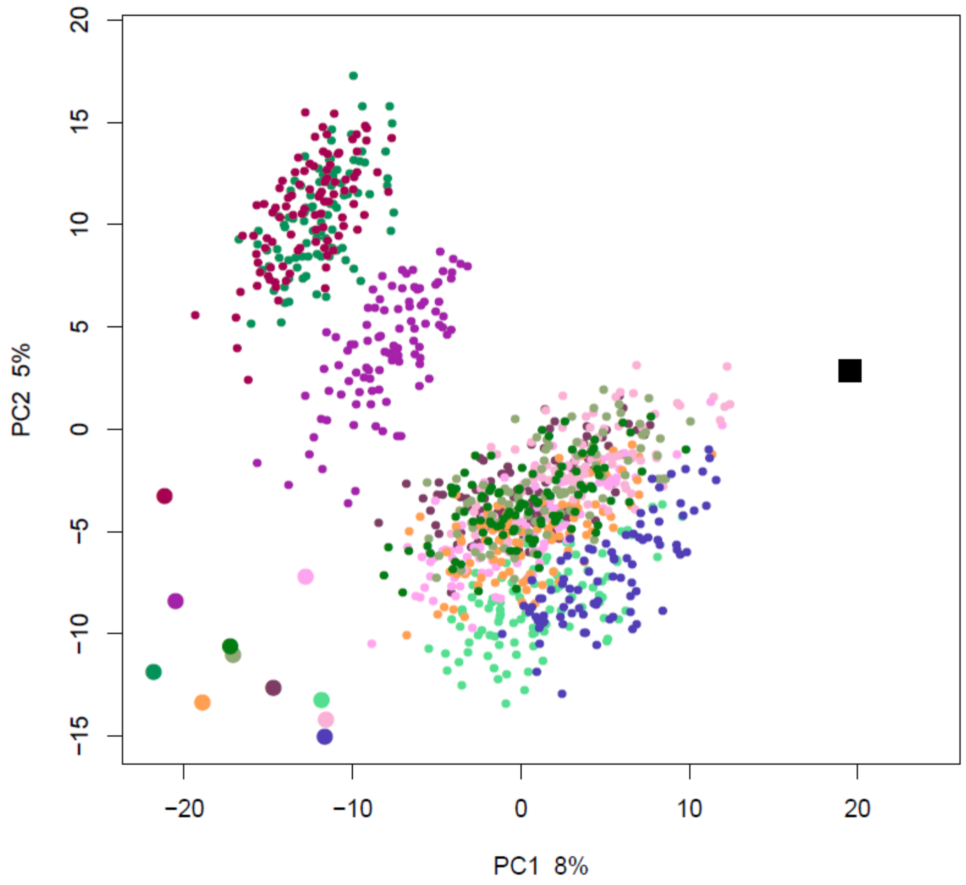
- 50 Ethiopian durum wheat landraces (+ 2 italian lines) chosen as female parents
- Recurrent male founder (Asassa) with international background selected on the basis of farmers'

Kidane et al 2019

12 NAM families, 100 RIL each (1,200 RILs) selected for initial characterization (total 6,500)

- Genotyping with the Illumina 15k SNP



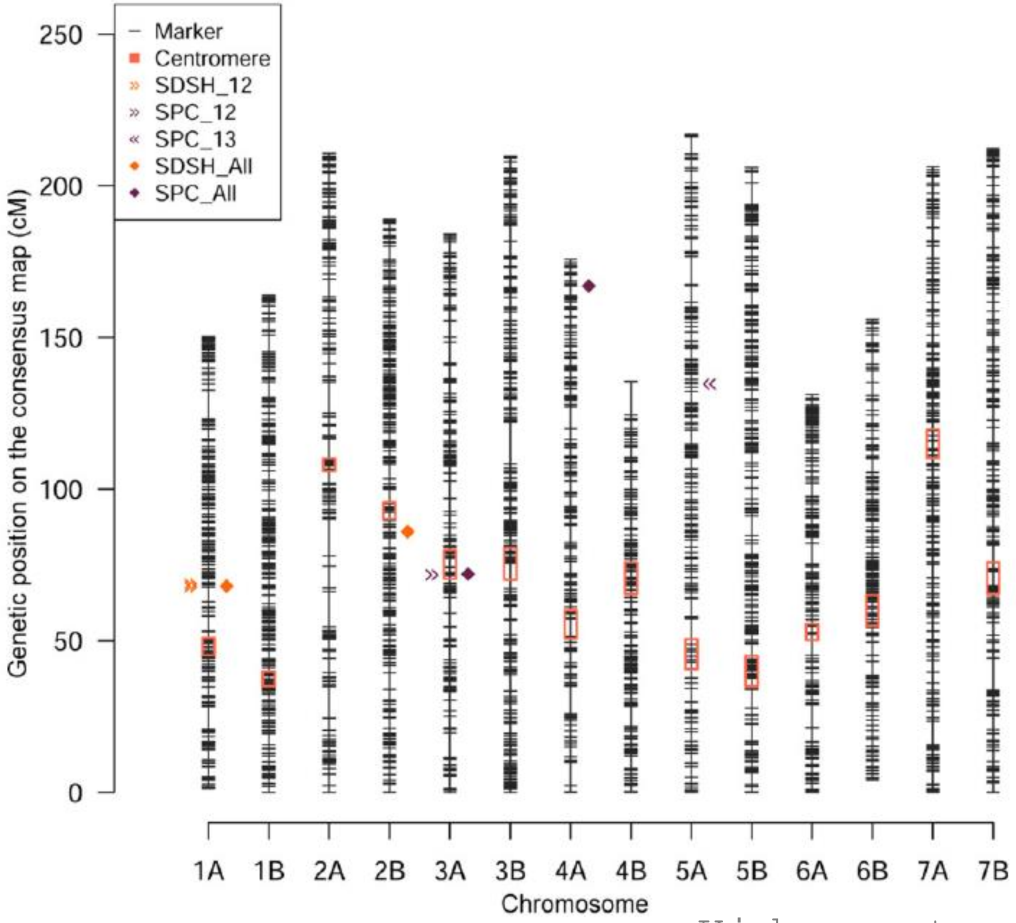


Kidane et al 2019

Plant materials: 400 landraces; 1,200 RILs

Genetic materials can be used in a genome wide association study (GWAS) to identify alleles of breeding relevance, *e.g.* resistance to Septoria

Bogale Nigir, PhD



Kidane et al 2017

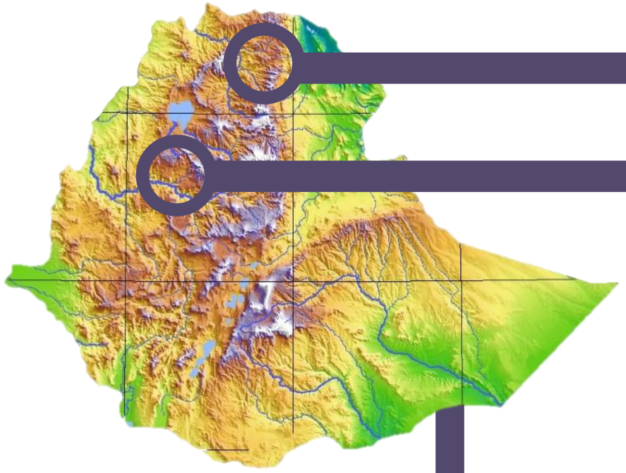
Bringing farmers into the picture

- Each family-village uses its own seeds, selected and maintained forward according to their preference
- Smallholder farmers must be efficient and knowledgeable: their environment is not very resilient. Their choice of genetic materials must be the right choice
- Participatory varietal selection (PVS) can help accessing this knowledge

see Ceccarelli et al for early work

Are PVS traits a quantitative phenotype? → Are PVS traits related with metric traits? → Do PVS traits have a genetic basis in wheat?

Different wheat agroecologies



Men farmers

Women farmers

Focus group discussions and survey to identify traits most relevant to farmers

Scores 1 to 5 given for overall appreciation (OA): how much do you like this wheat genotype?

In each, metric traits collected on hundreds of genotyped wheat accessions laid down in a replicated lattice design

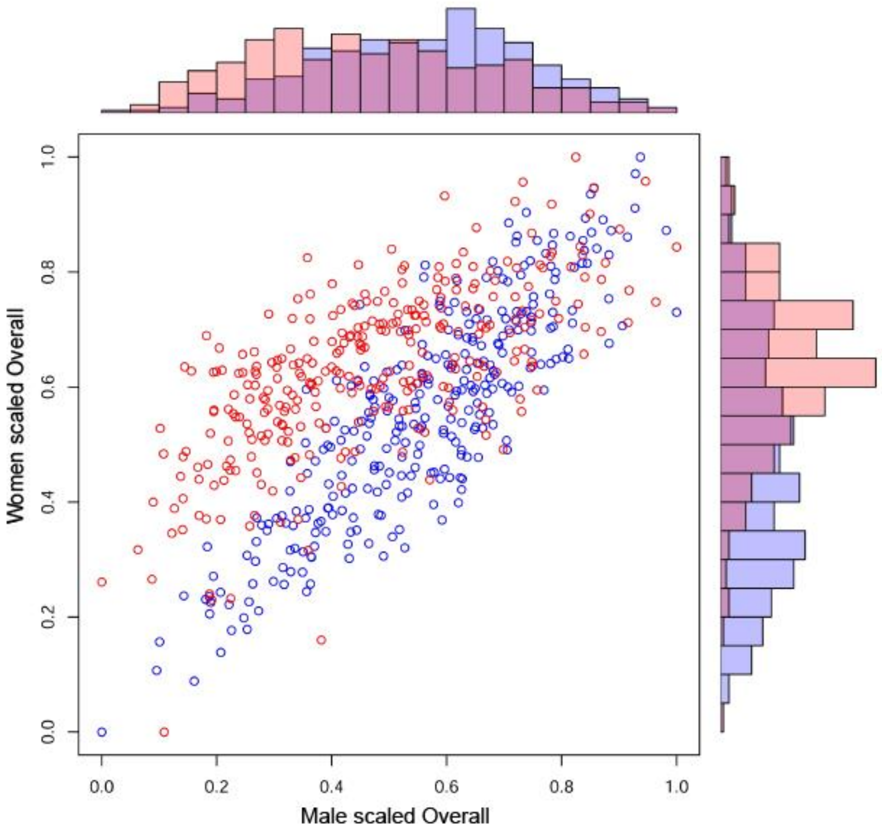
Evaluation given to each unlabeled plot, groups entering from random entry points, scoring system devised to avoid bias

Chiara Mancini, PhD

Scoring system

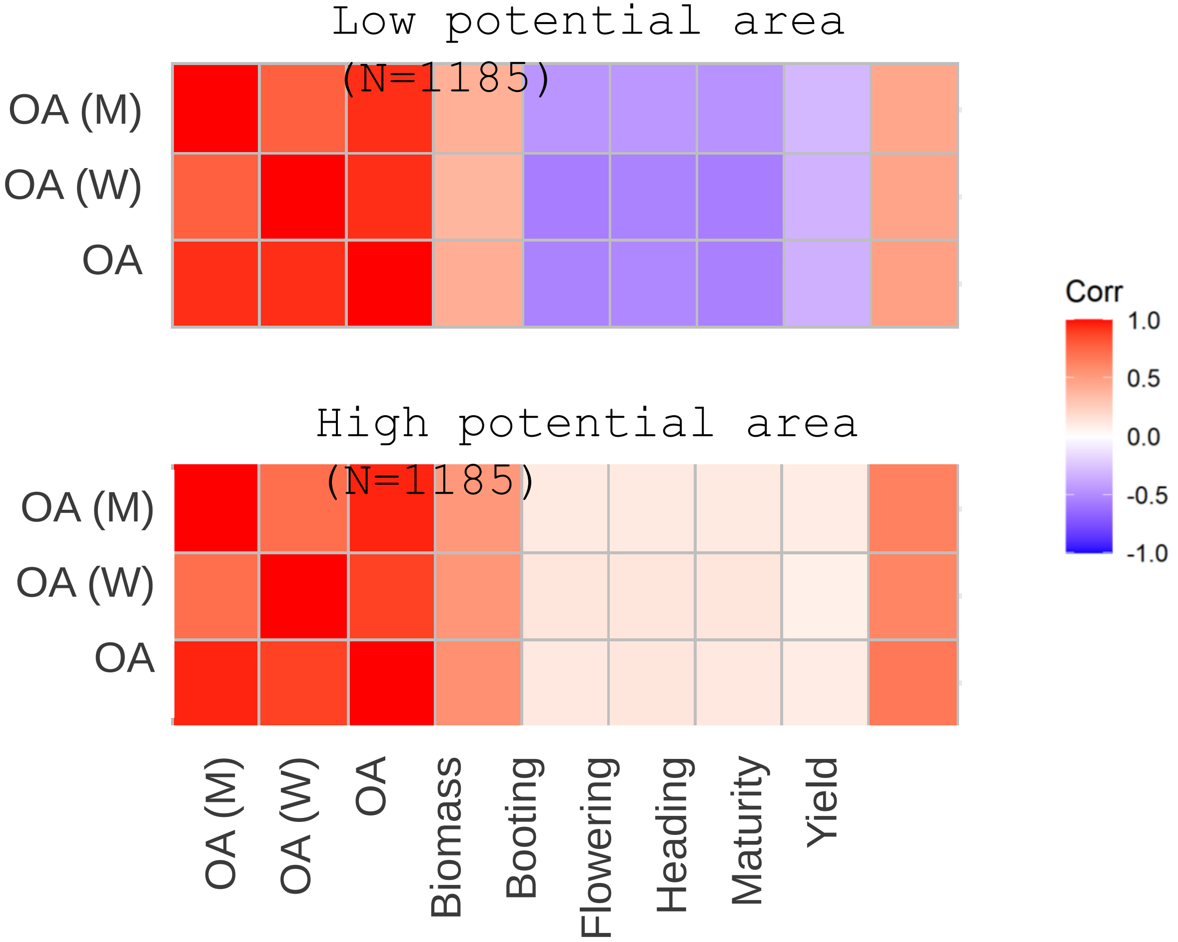
- Genetic materials never seen before
- Evaluations given eyeballing the field
- Individual scores recorded
- 400 landraces in 1200 plots (Geregera, Hagreselam)
- 1200 EtNAM RILs in 7200

Farmer scores are repeatible across genders and across locations



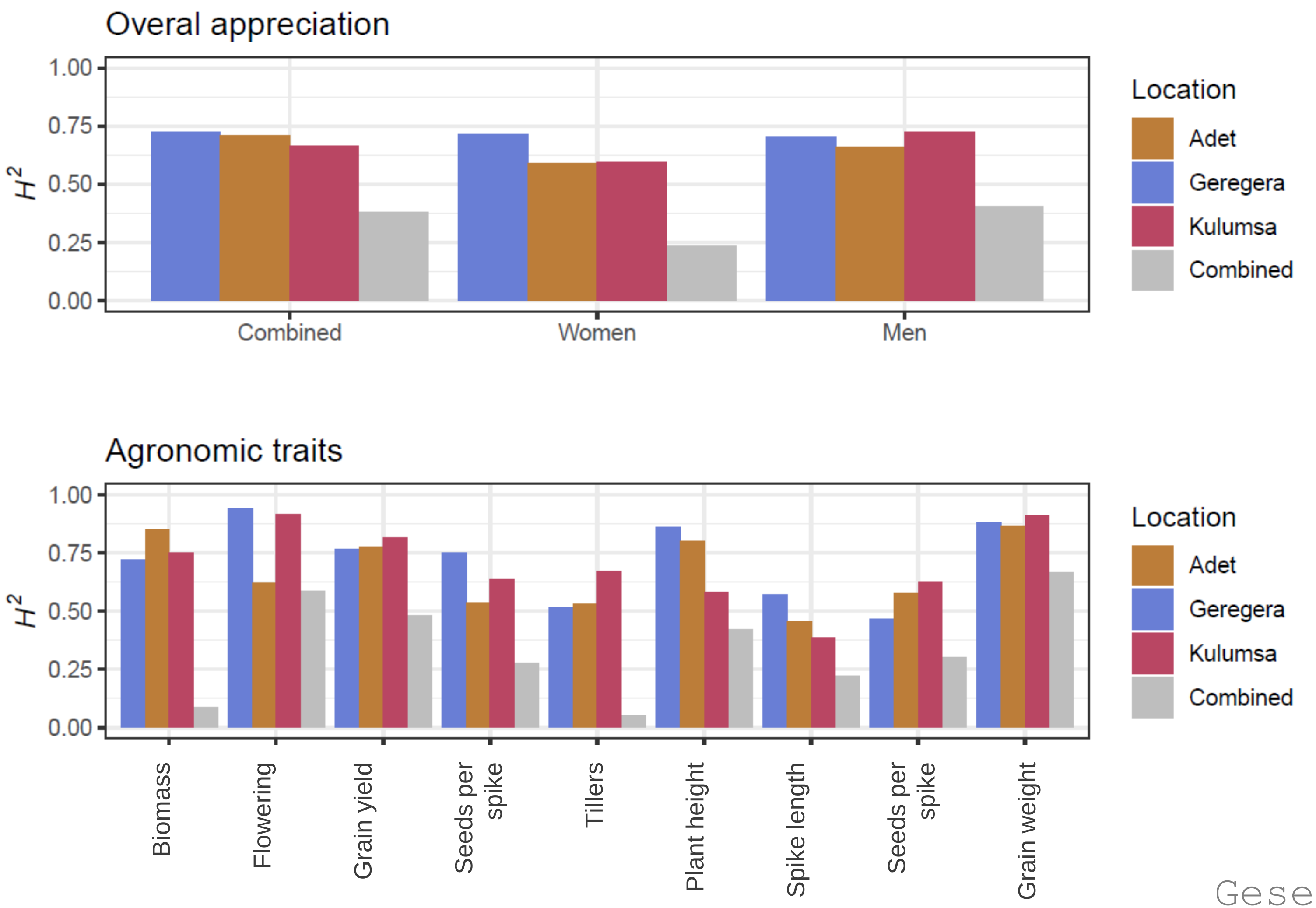
Mancini et al 2017

Farmer scores and metric traits are related

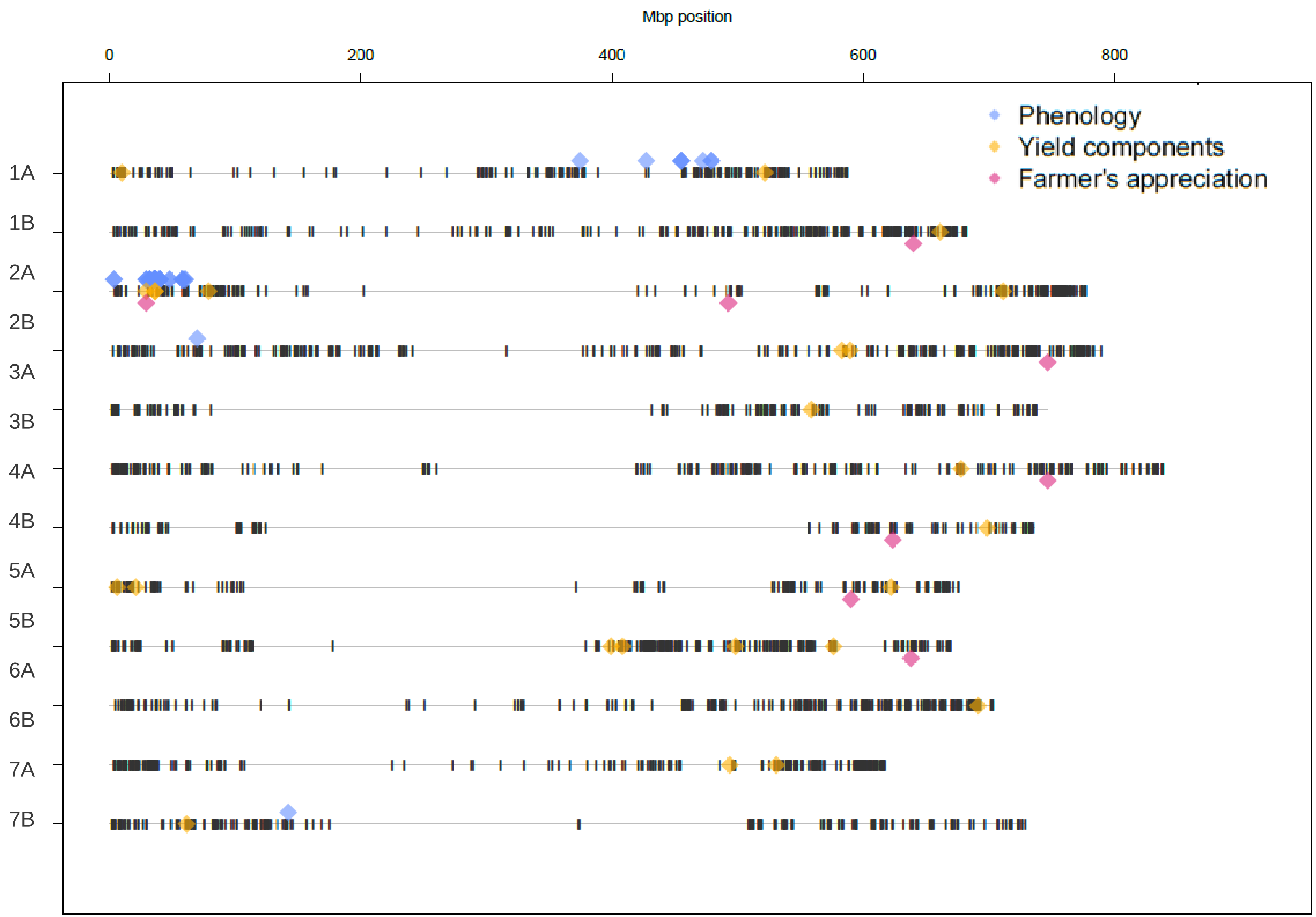


Gesesse et al 2023

Farmer evaluations of wheat performance are heritable over year, location, gender: they have a genetic basis in wheat



Gesesse et al 2023



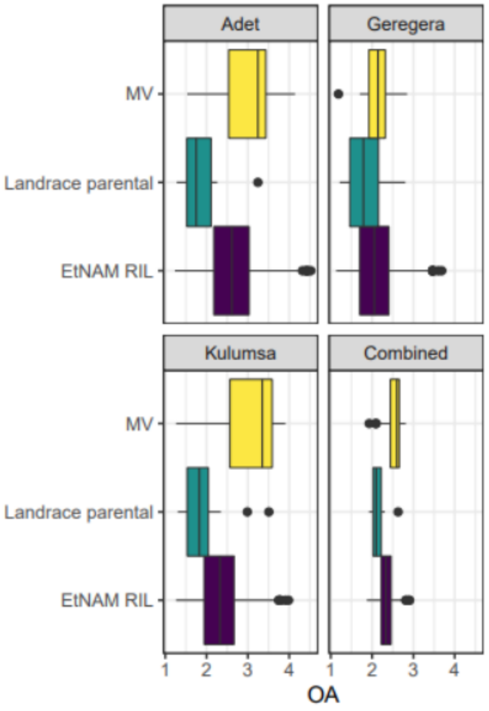
Gesesse et al 2023

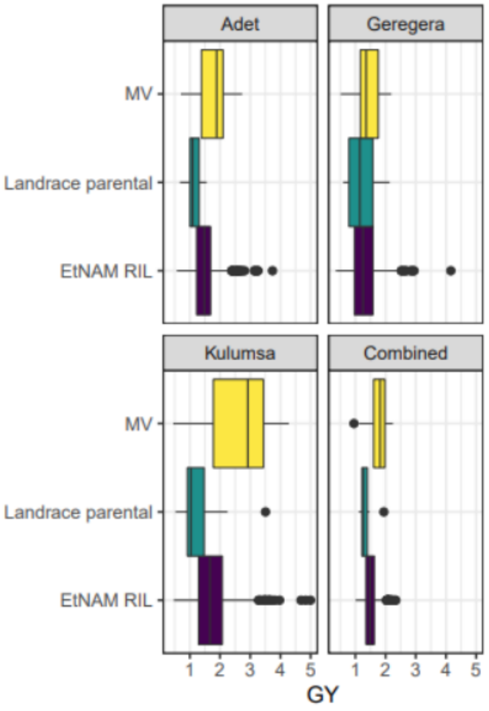
PVS farmer scores can identify QTL that partially overlap with those deriving from metric traits (*e.g.*, kernel size, phenology)

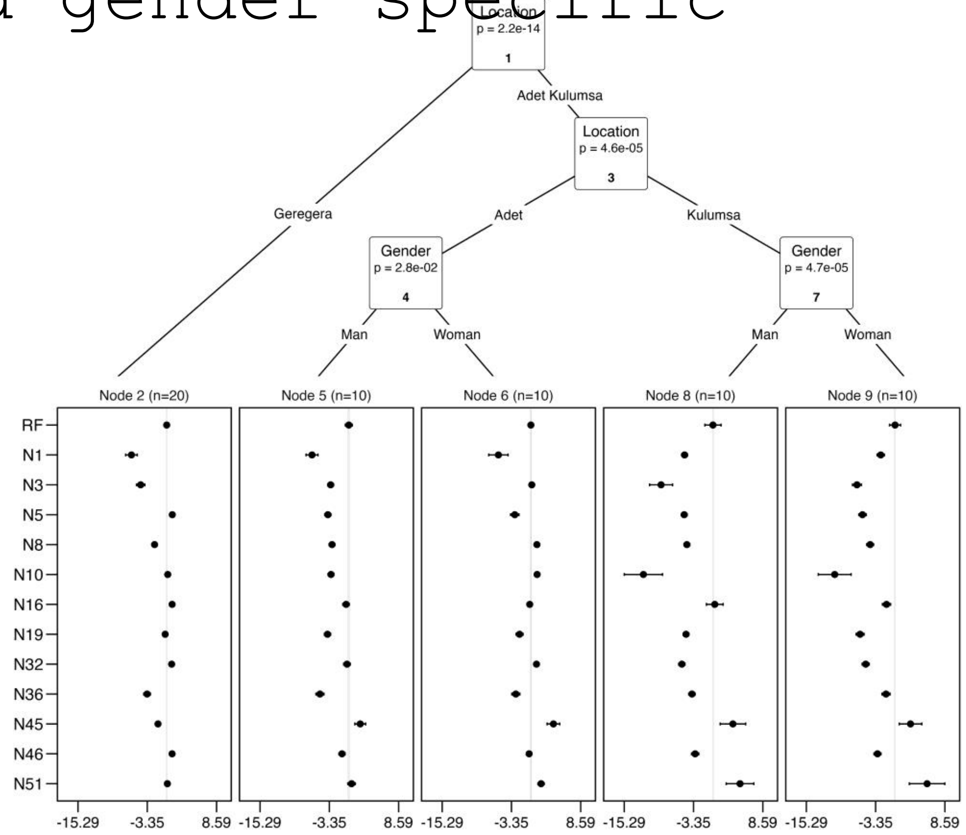
Some PVS QTL are consistent across germplasm, location, and

In a pre-breeding perspective:

- Several RILs were outperforming both EtNAM parental lines and improved genotypes checks
- The performance/preference for specific allelic combinations is location and gender-specific







Towards a quantitative integration of farmers' knowledge in genomic selection breeding

Cherinet Alem, PhD

Bogale Nigir, PhD

Training set: SNP data and trait data

Estimate alelle effects

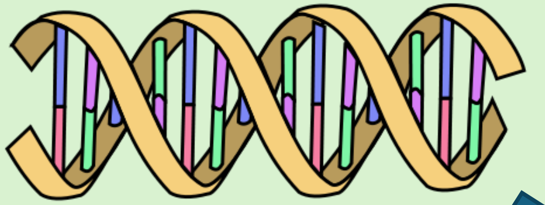
Validation set: SNP data only, trait data masked

Accuracy: correlation between estimated trait values and true values

- Landraces, 1600 plots
- EtNAM RILs, 7200 plots
- rrBLUP to perform genomic selection for yield (GY) and farmer appreciation(OA)
- Selection conducted on BLUPs measured across years and locations, accuracy monitored

Training set

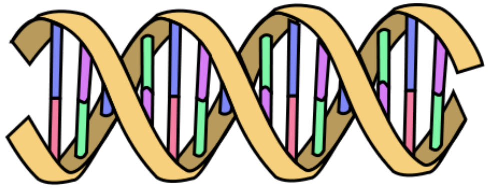
DNA markers



Phentoypes

Test set

DNA markers

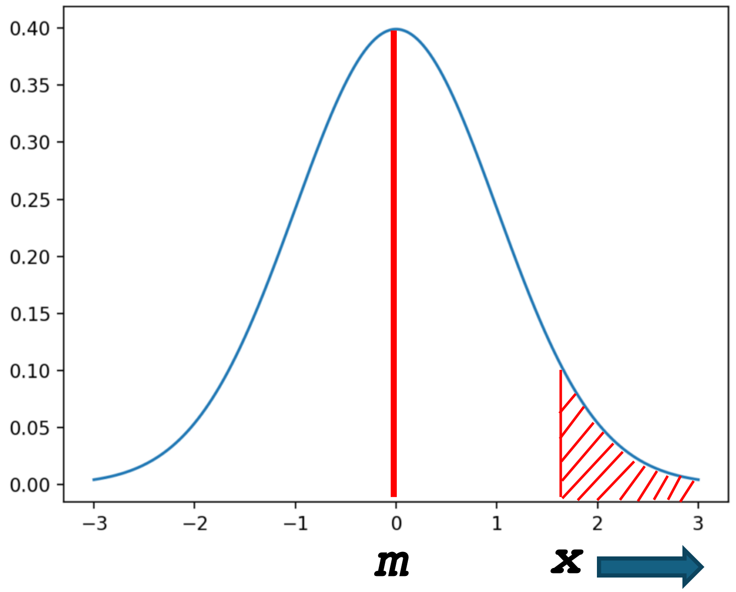


Model training

$$y = Xb + Zu + e$$

- **No** QTL mapping
- **No** testing for marker significance
- **No** effort to localize genes

Predict and select



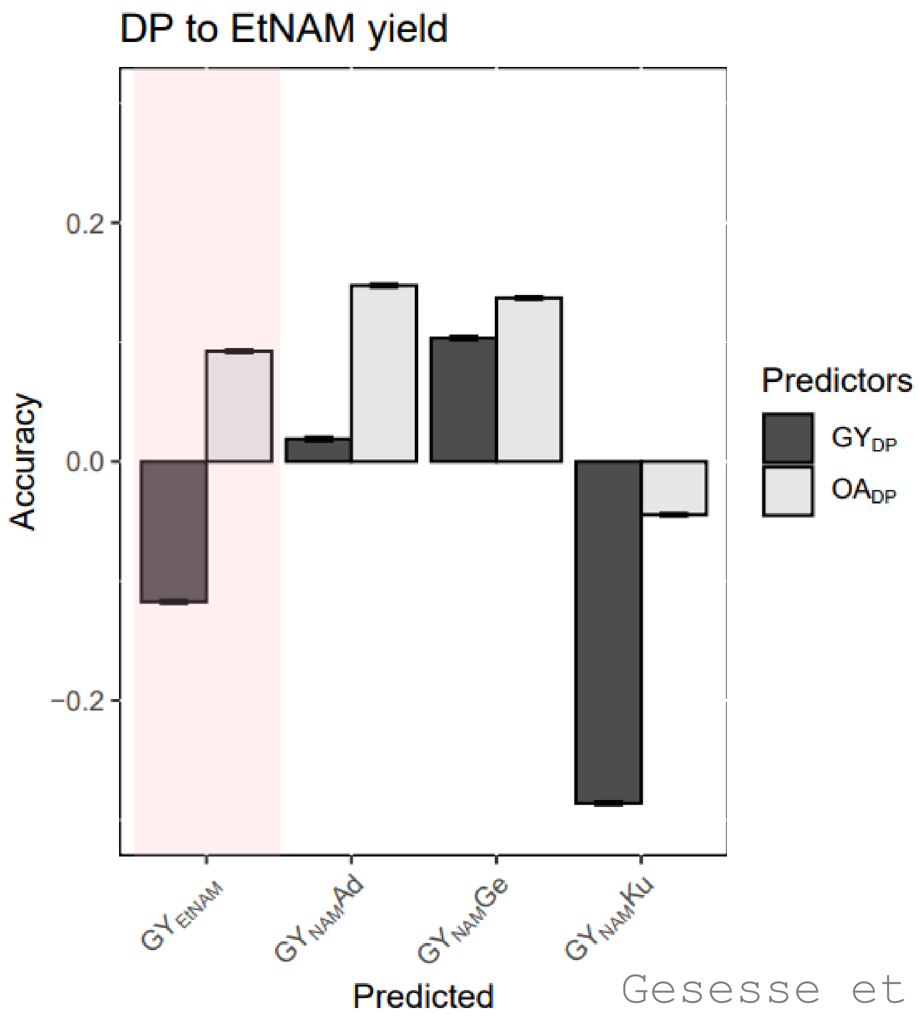
Best genotypes are chosen on the basis of GEBV and advanced in the breeding pipeline

Overall appreciation (OA) by PVS provides better accuracy in predicting grain yiled (GY) than GY itself

Training: landraces → **Testing:** NAM RILs

OA, GY → GY

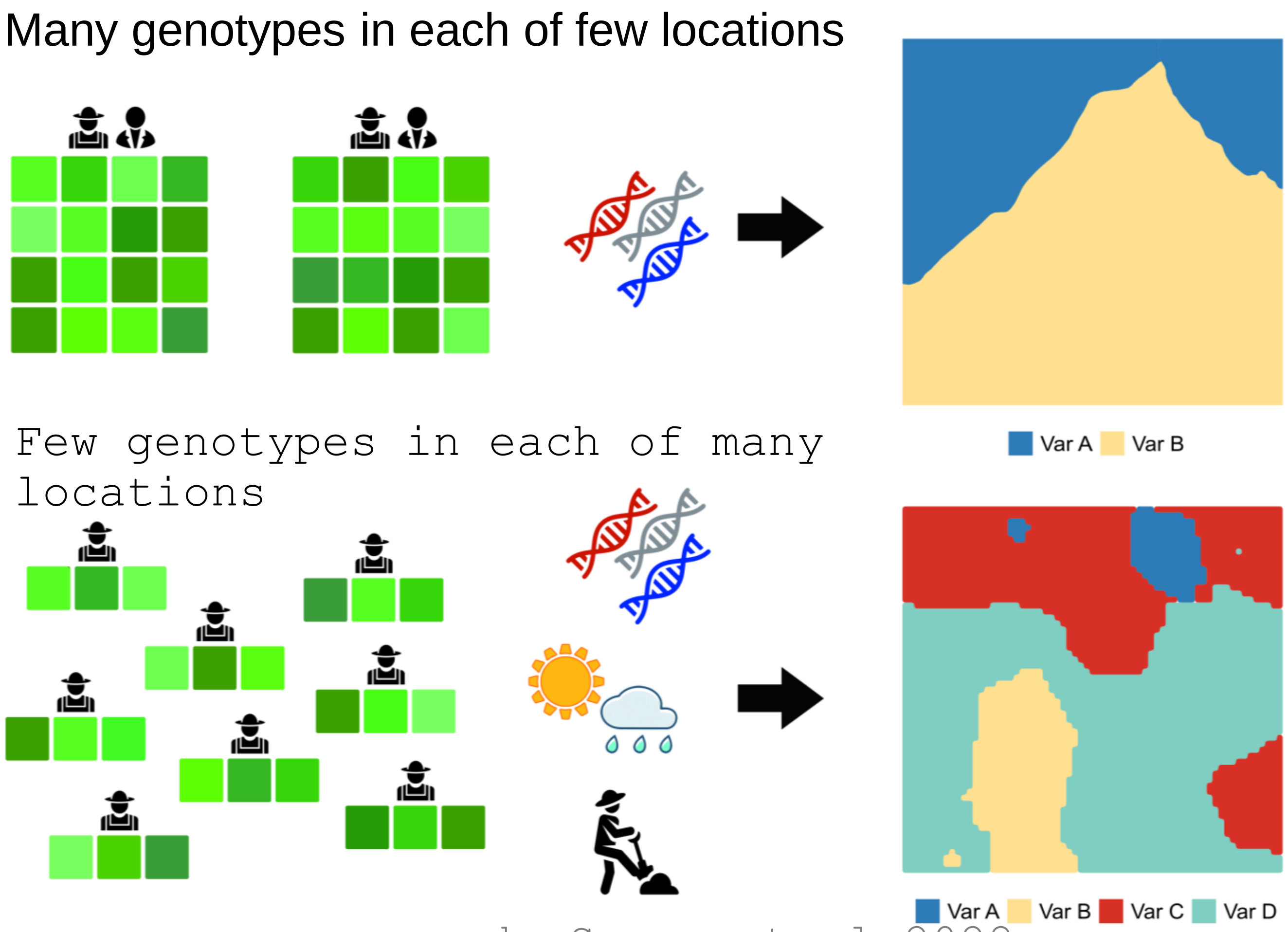
- Note that farmer groups providing evaluations are different and never seen the genetic materials before
- Are farmers able to capture the *expected* yield in local growing conditions?



Gesesse et al 20

Towards a decentralized breeding paradigm

- Selection moves to 1000+ farmer fields
- Varieties are grown in true farm condition and farmers are asked to rank varieties according to their preference
- The resulting accuracy for varietal development and recommendation is increased



de Sousa et al 2022

The decentralized evaluation of varieties is based on the *tricot* approach; incomplete ranking at each farmer field can be combined into a measure of *worthiness* – *i.e.*, the tendency in outperforming other varieties

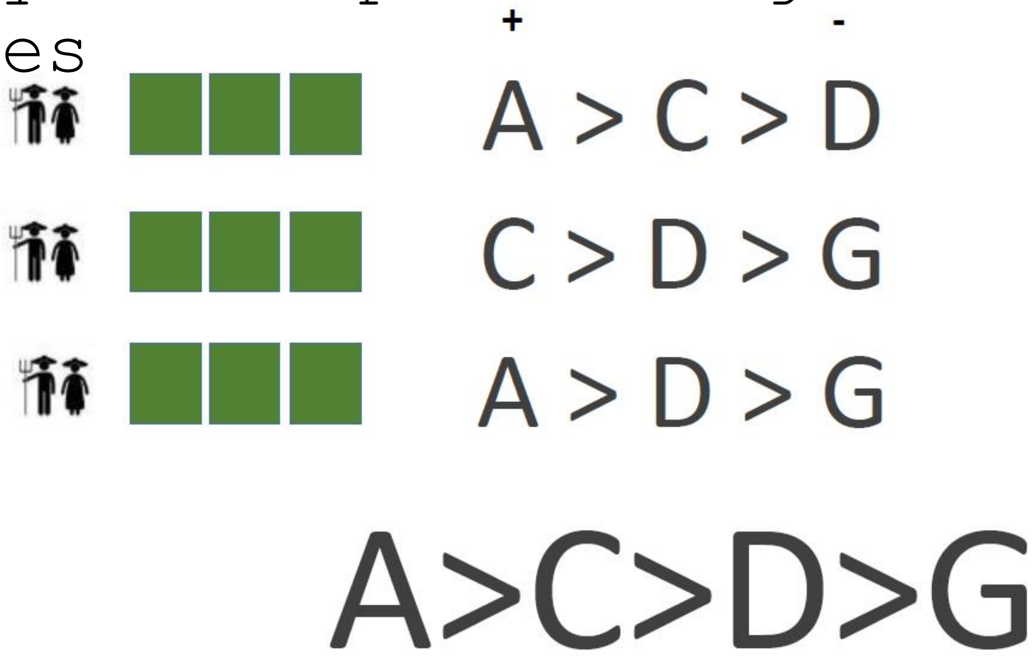
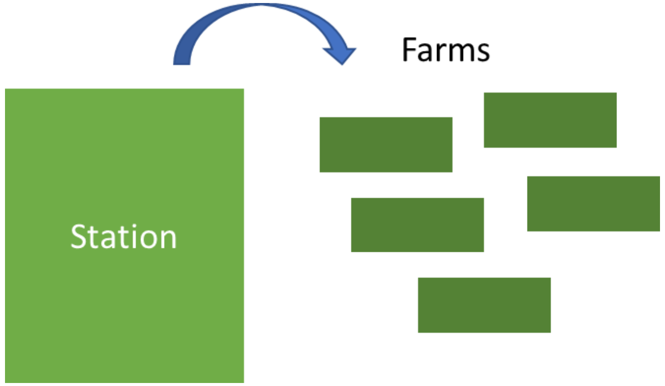
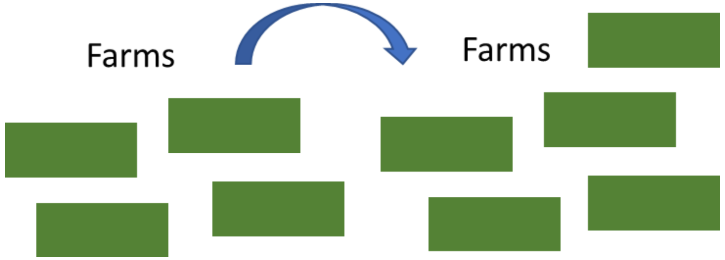


Image courtesy of Kaue de Sousa
For *tricot* work and conceptualization see van Etten et al 2019, 2020, 2021

The ranking derived from decentralized farms in combination with climate and genomic diversity predicts both GY and OA in untested environments with twice the accuracy of “conventional” genomic selection

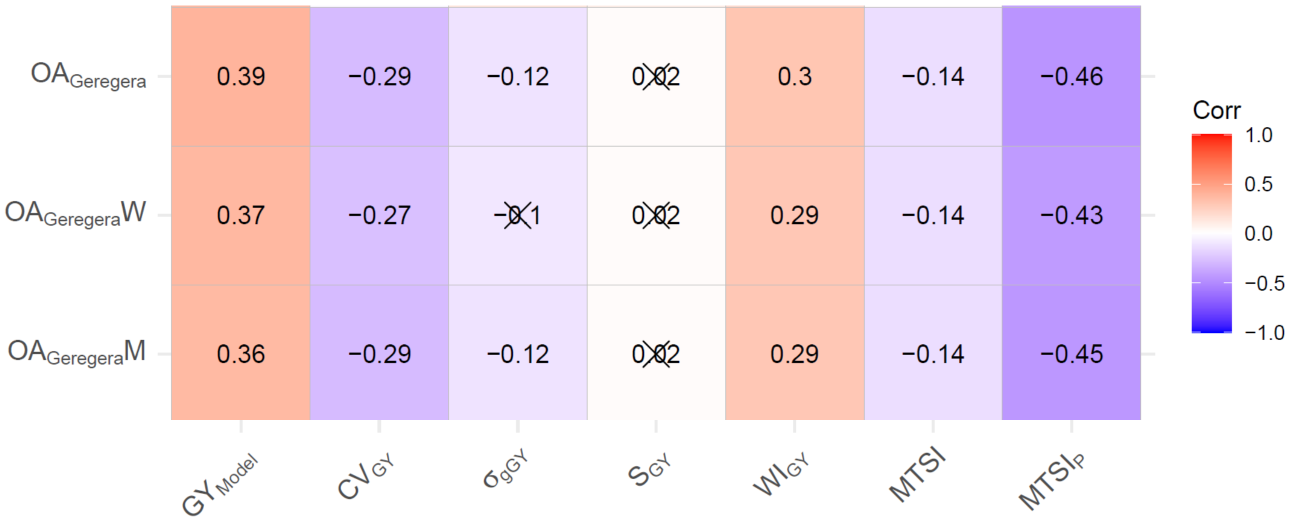
Approach	OA	GY
Centralized GS		
Season 1 (n=179)	0.134	-0.012
Season 2 (n=651)	0.105	0.076
Season 3 (n=335)	0.183	0.073
	0.141 (± 0.039)	**0.046 (± 0.049)**
3D-breeding		
Season 1 (n=179)	0.270	0.160
Season 2 (n=651)	0.276	0.078
Season 3 (n=335)	0.203	0.119
	0.251 (± 0.040)	**0.109 (± 0.041)**



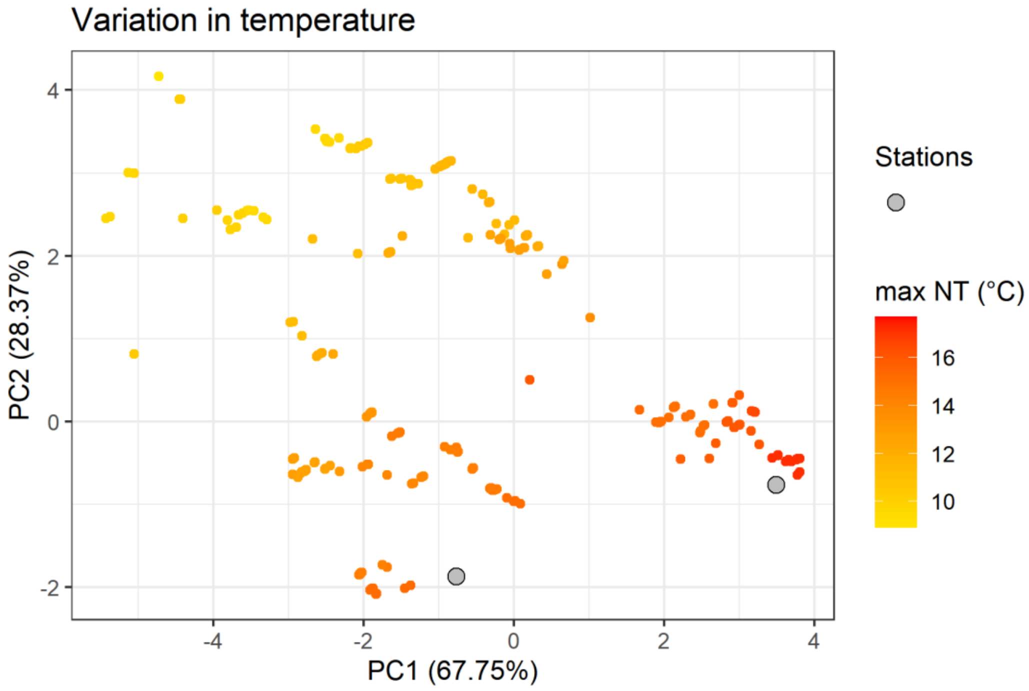


How comes?

- Farmers evaluate yield and yield components, increasing the heritability of the predicted trait
- Can farmers evaluate genotype stability across seasons?



A decentralized model captures environmental variation at farm sites, capturing GxE interactions



Wrapping up

- The innovation process flows both ways from researchers to end users;
- There is value in engaging in a conversation and open a space for everybody involved to bring their knowledge to the table;
- The challenge is to integrate genes, phenotypes, environment, and social sciences to tailor varieties for local adaptation to achieve sustainable intensification of local agriculture;
- Smallholder agriculture and traditional knowledge are not at odds with modern research; they may be a resource for breeding - a complement to current breeding strategies

Acknowledgements

Scuola Superiore Sant'Anna di Pisa

- Mario Enrico Pè
- Josef Kidane
- Dejene Mengistu
- Cherinet Alem
- Bogale Nigir
- Mara Miculan
- Chiara Mancini

Alliance of Bioversity International and CIAT

- Carlo Fadda
- Jacob van Etten
- Kaue de Sousa

Amhara Regional Agricultural Research Institute

- Ermias Abate

Mekelle University

Ethiopian Biodiversity Institute

Kansas State University

- Jesse Poland

University of Milan

- Luca Gianfranceschi
- Guido Gallo

Shortcomings of genomic prediction

- The genetic basis of some traits is still too complex or too feeble to be efficiently predicted (+ ethical considerations in medicine)
- Very little number of traits for which a single gene can be meaningful in predicticing outcomes
- The value of predictions depend on the representativeness of the training data
- Difficult to drive deterministic conclusions: $Y = G + E + G \times E$
- Good in capturing the mean, bad in capturing outliers

Molecular biology does magic

- It has been shown that modification of relatively simple molecular targets can improve yields from 10% to 68%
- Photosyntesis is one of the target of these modifications

RESEARCH ARTICLE

PLANT SCIENCE

A transcriptional regulator that boosts grain yields and shortens the growth duration of rice

Shaobo Wei[1]†, Xia Li[1]†, Zefu Lu[1], Hui Zhang[2], Xiangyuan Ye[1], Yujie Zhou[1], Jing Li[1], Yanyan Yan[1], Hongcui Pei[1], Fengying Duan[1], Danying Wang[3], Song Chen[3], Peng Wang[4], Chao Zhang[5], Lianguang Shang[5], Yue Zhou[6], Peng Yan[6], Ming Zhao[1], Jirong Huang[2], Ralph Bock[7], Qian Qian[1,3], Wenbin Zhou[1]*

Complex biological processes such as plant growth and development are often under the control of transcription factors that regulate the expression of large sets of genes and activate subordinate transcription factors in a cascade-like fashion. Here, by screening candidate photosynthesis-related transcription factors in rice, we identified a DREB (Dehydration Responsive Element Binding) family member, OsDREB1C, in which expression is induced by both light and low nitrogen status. We show that OsDREB1C drives functionally diverse transcriptional programs determining photosynthetic capacity, nitrogen utilization, and flowering time. Field trials with *OsDREB1C*-overexpressing rice revealed yield increases of 41.3 to 68.3% and, in addition, shortened growth duration, improved nitrogen use efficiency, and promoted efficient resource allocation, thus providing a strategy toward achieving much-needed increases in agricultural productivity.

A

... the yield and biomass of ... plants in field trials

... Zhang, Xueping Wang, Yingying Xu, Hong Yu, Yulong Li, Junbo Yang, Jun Tang, Hong-Chao Duan, Lian-Huan Wei, Haiyan Zhang, Jiangbo Wei, Qian Tang, Chunling Wang, Wutong Zhang, Ye Wang, Peizhe Song, Qiang Lu, Wei Zhang, Shunqing Dong, Baoan Song ✉, ... Guifang Jia ✉ + Show authors

Aren't breeders carried away by the magic?

Scale up trials to validate modified crops' benefits

Merritt Khaipho-Burch, Mark Cooper, Jose Crossa, Natalia de Leon, James Holland, Ramsey Lewis, Susan McCouch, Seth C. Murray, Ismail Rabbi, Pamela Ronald, Jeffrey Ross-Ibarra, Detlef Weigel & Edward S. Buckler

With a changing climate and a growing population, the world increasingly needs more-productive and resilient crops. But improving them requires a knowledge of what actually works in the field.

range of disciplines to work together much more than they currently do, and to use well-established yield-testing approaches.

Perspective is needed

Promising reports of the possible effects on crop yields of introducing a gene from another species, or of using the gene-editing technique CRISPR–Cas9 to modify a gene or multiple genes, attract considerable media attention. Yet, more-conventional plant-breeding approaches used over decades paint a very different picture of what genetic modifications are likely to achieve, in relation to yields, in the coming decades.

What breeders and quantitative geneticists

nature

nature › comment › article

COMMENT | 20 September 2023

Genetic modification can improve crop yields — but stop overselling it

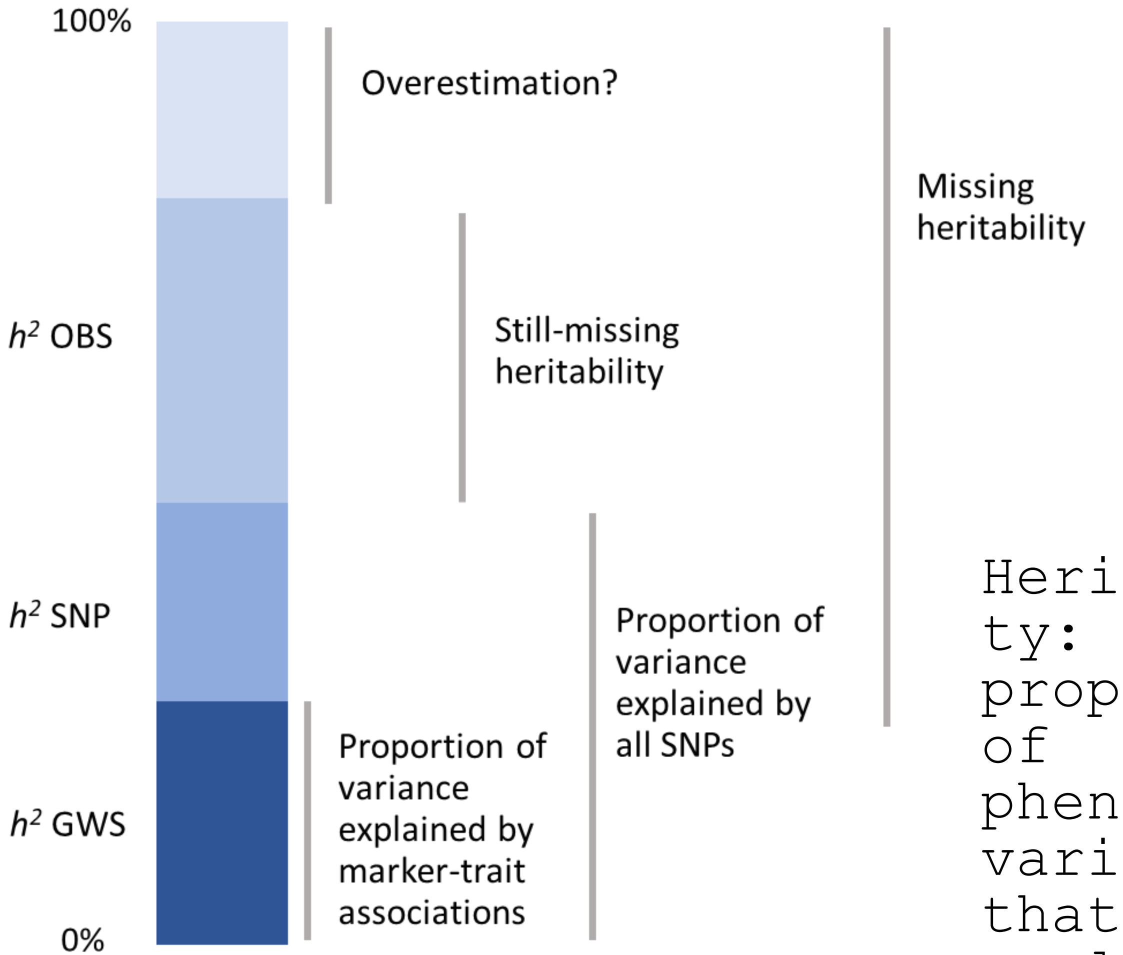
With a changing climate and a growing population, the world increasingly needs more-productive and resilient crops. But improving them requires a knowledge of what actually works in the field.

By Merritt Khaipho-Burch, Mark Cooper, Jose Crossa, Natalia de Leon, James Holland, Ramsey Lewis, Susan McCouch, Seth C. Murray, Ismail Rabbi, Pamela Ronald, Jeffrey Ross-Ibarra, Detlef Weigel & Edward S. Buckler

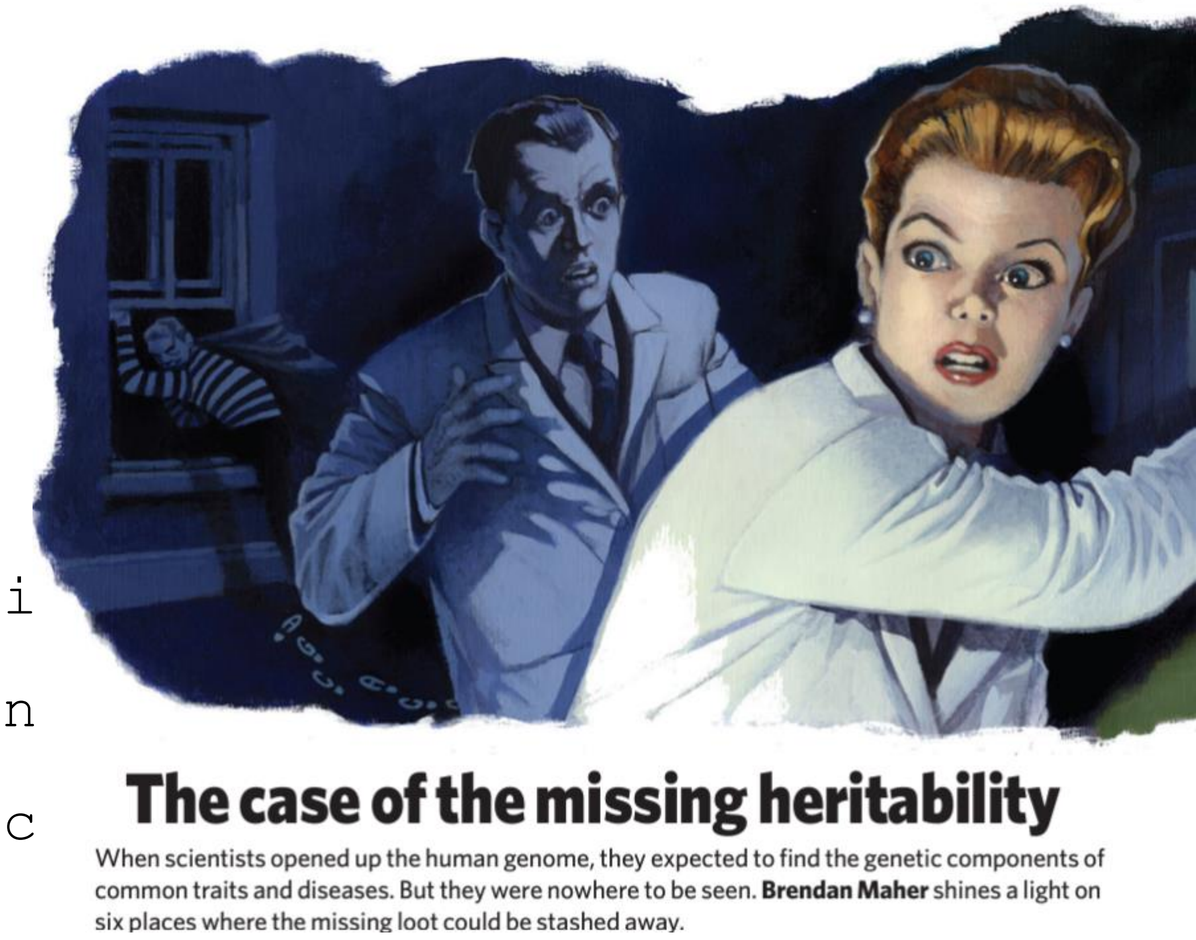
In breeders' hands, a yield increase of 1%-5% over generation is considered a breakthrough

- *E.g.* Corteva tested the effect of 1,671 genes, taken from 47 species, on complex traits in maize. Only 1% of these genes increased yield enough to warrant more investigation
- In subsequent rounds of testing, only *zmm28* (a TF) was validated for a 2% yield increase and this required the creation

Fact: a sizable portion of phenotypic diversity can not be traced back to genetic factors



Heritability: the proportion of phenotypic variance that is explained by genotypic



ELSEVIER

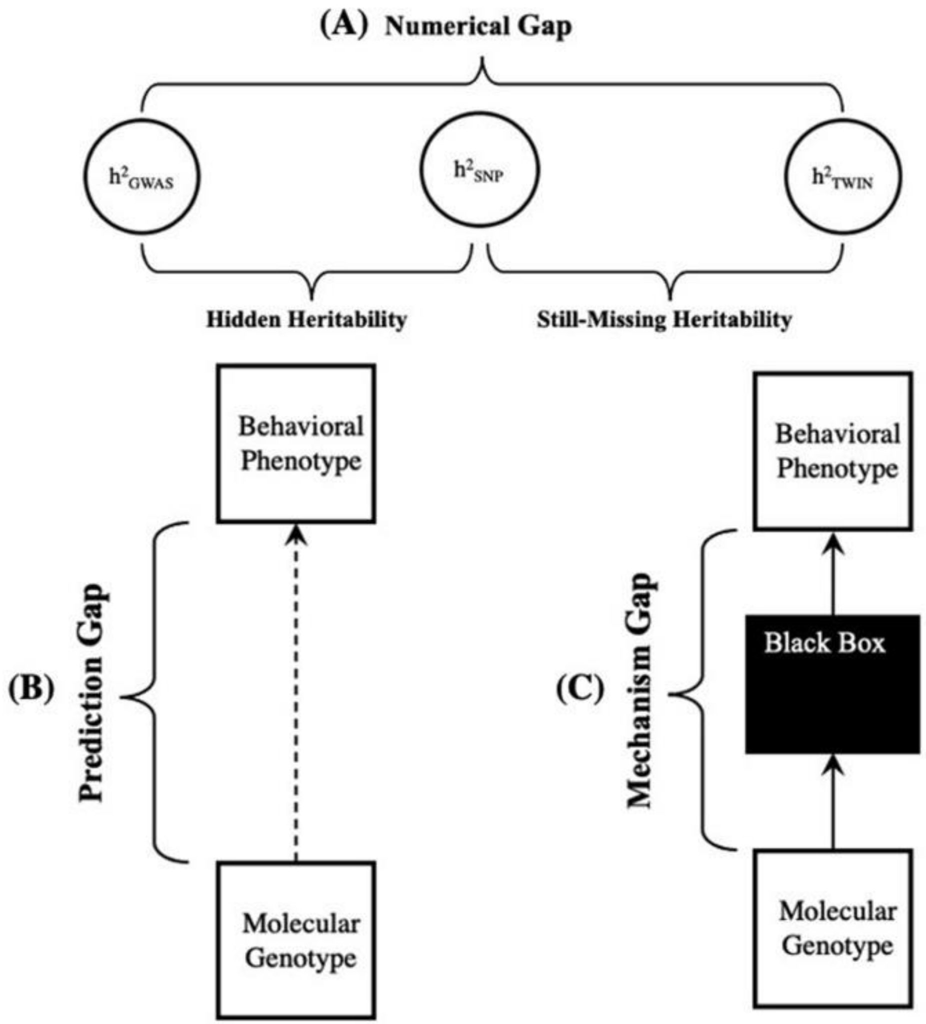
Studies in History and Philosophy of Science

Volume 93, June 2022, Pages 183-191

Three legs of the missing heritability problem

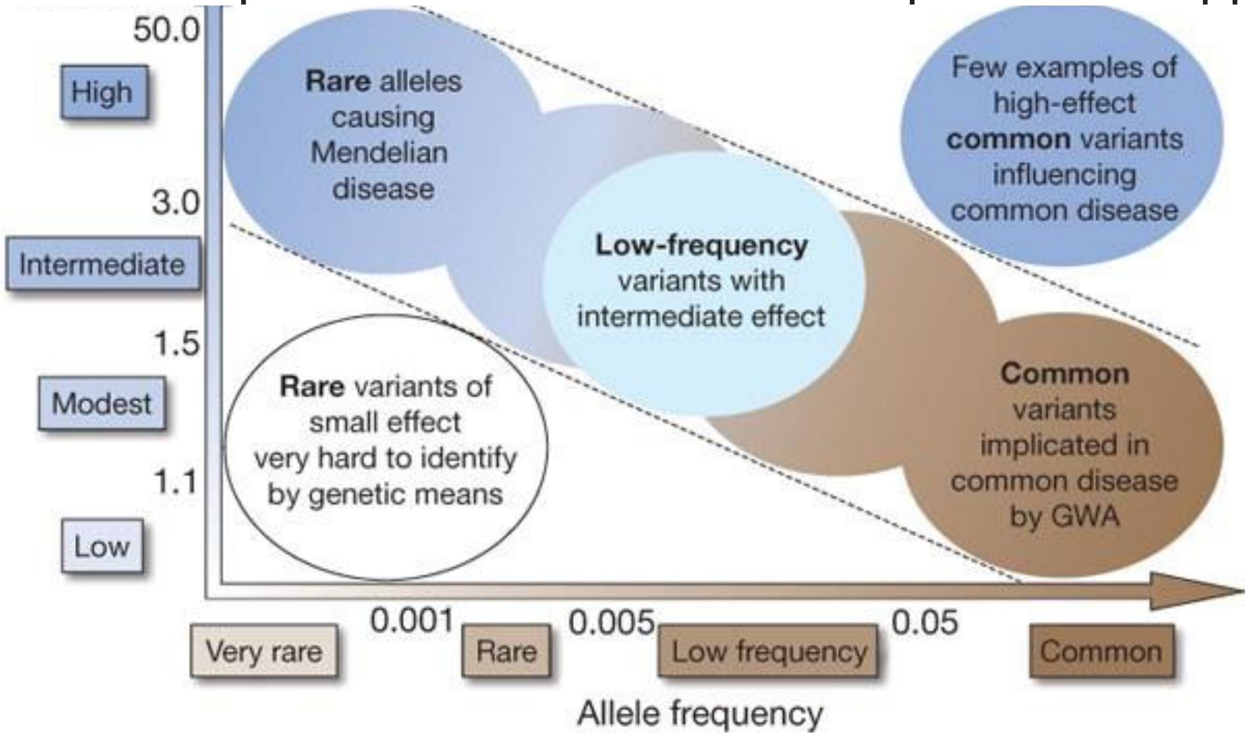
Lucas J. Matthews [a], Eric Turkheimer [b]

- **Numerical:** an issue with numbers, related with inadequacy in observational data (e.g. experiment size)
- **Predictive:** an issue with something that we cannot measure or that we are not measuring well enough
- **Mechanicistic:** an issue with our fundamental lack of understanding of the determinants of complex traits



The Numerical issue

- Complex traits are complex, and the contribution of each individual locus is small and hard to detect
- High-effect alleles are typically rare and may be missed when the size of the experiment is too small
- Small effect loci can escape detection if the statistical power of mapping is not sufficient

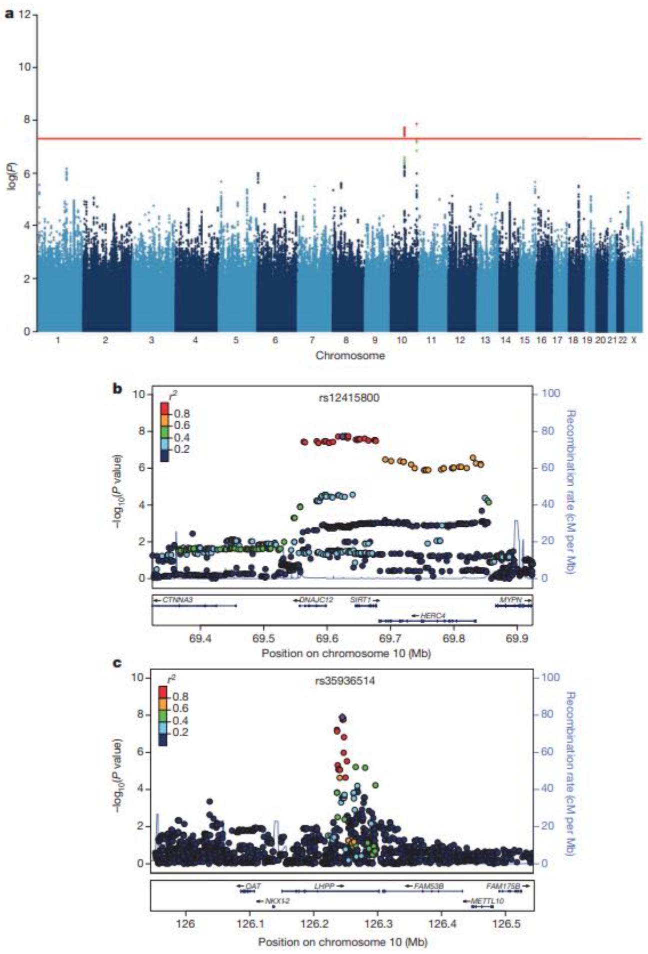


2015 – 11,670 Han Chinese women; 6.2M SNPs; 2 associations
2018 – 480,359 Europeans; 9.6M SNPs; 44 associations
2022 – 1,815,091 individuals from different cohorts; 22M+ SNPs and TWAS; 300+ associations

LETTER
doi:10.1038/nature14659

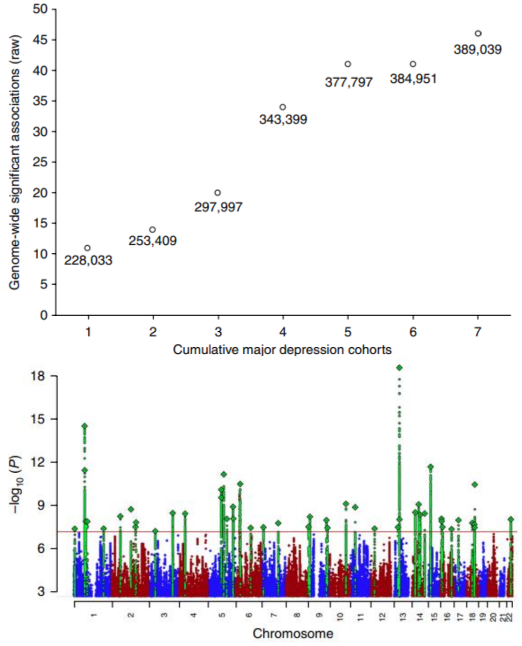
Sparse whole-genome sequencing identifies two loci for major depressive disorder

CONVERGE consortium*



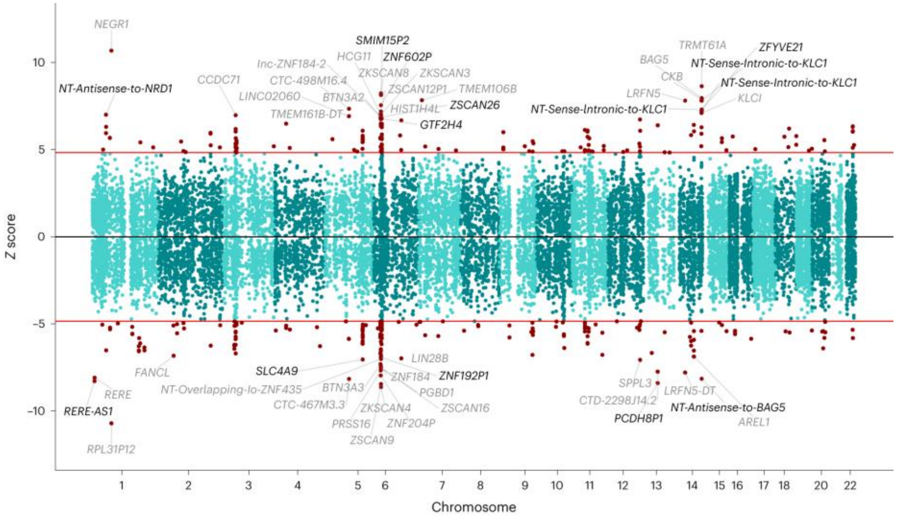
ARTICLES
https://doi.org/10.1038/s41588-018-0090-3
nature genetics

Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression



nature genetics
Article
https://doi.org/10.1038/s41588-023-01596-4

Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference



Gui *et al. Genome Biology* (2022) 23:178
https://doi.org/10.1186/s13059-022-02742-7

Genome Biology

RESEARCH Open Access

A pan-*Zea* genome map for enhancing maize improvement

Songtao Gui[1], Wenjie Wei[1], Chenglin Jiang[1], Jingyun Luo[1], Lu Chen[1], Shenshen Wu[1], Wenqiang Li[1], Yuebin Wang[1], Shuyan Li[1], Ning Yang[1,2], Qing Li[1,2], Alisdair R. Fernie[3] and Jianbing Yan[1,2*]

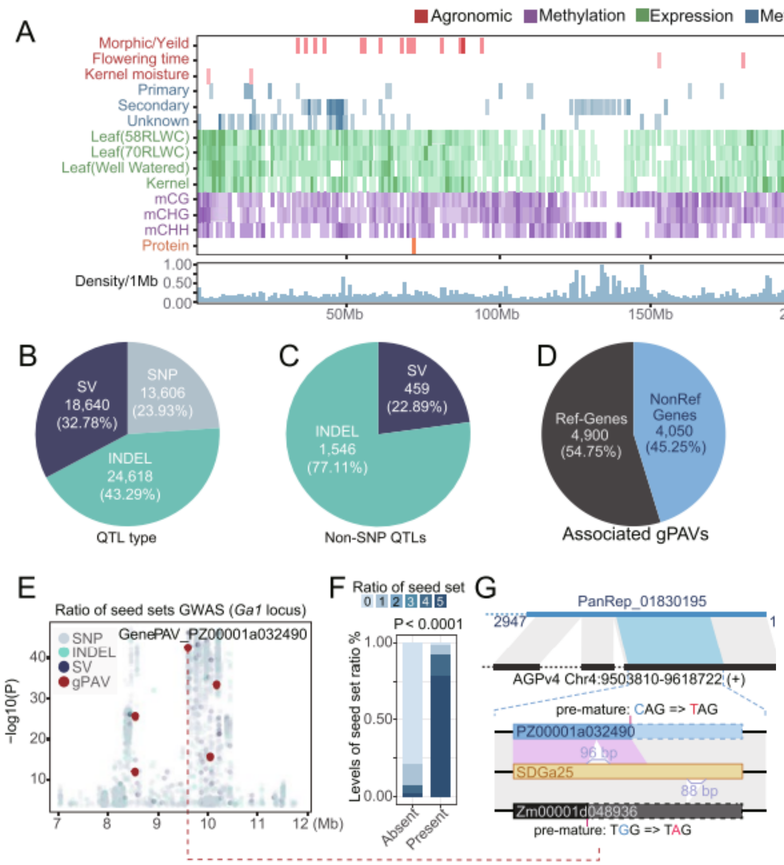
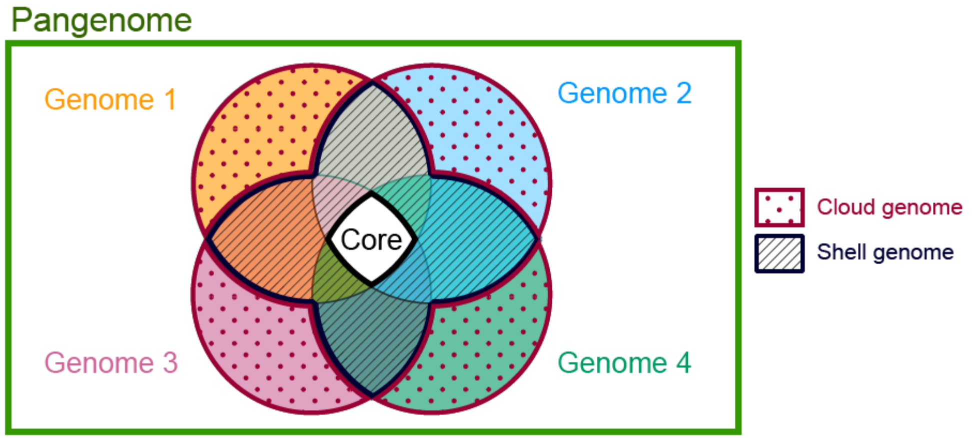
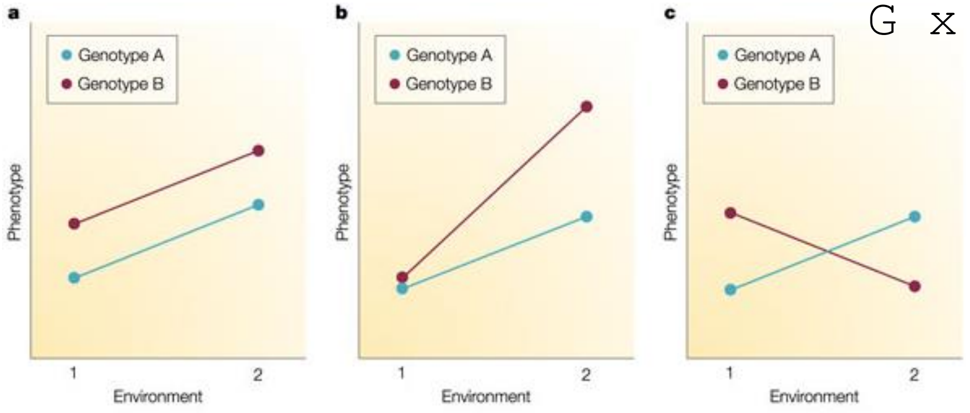


Fig. 4 Characteristics of the phenotype associated QTLs, genes, and causal variants. **A** Distribution of the associated QTLs, genes, and causal variations (example with chromosome 1, see Additional file 1: Fig. S14 for the whole genome). The heatmap represents the QTL density within each 1-Mb window, while the histograms indicate the density of the candidate causal variants that were normalized with the number of all variants within the 1-Mb window. **B** Proportions of QTL types (QTLs lead with different variant types). **C** Proportions of INDEL/SV-specific QTLs (QTLs that cannot be detected by SNPs). **D** Proportions of associated gPAVs for the reference genome genes (Ref-Genes) and non-reference genome genes (NonRef Genes). **E** Manhattan plot of the association result of the *Ga1* locus related to the ratio of seed sets, with gPAVs highlighted in red. **F** The proportion of different levels of seed set ratio related to the absence/presence of PZ00001a032490; the larger the number, the higher the seed set ratio. **G** The genome alignment indicates the anchoring of the NRS (PanRep_01830195) on the AGPv4 genome, and the schematic plot illustrates the differences between the three *PME* genes (PZ00001a032490, SDGa25, and Zm00001d048936). Solid rectangles indicate the gene coding sequence, while the dashed rectangles indicate the missing coding part related to SDGa25. Gray ribbons indicate the matched blocks. Pink ribbons indicate the matched CDS blocks. **H** Distribution of the PAV patterns (track 2) of the six *PME* genes and the ratio of seed sets (track 3) according to the structure tree of pan-*Zea* individuals (track 1). **I** Distribution of the number of presented *PME* genes (# PMEs) related to the levels of the seed set ratio. The gray histogram is the distribution of total sample numbers (count) according to the *X*-axis, while the colored histogram indicates the proportions. The *P*-value was calculated from 10,000 permutations of the Wilcoxon-Mann-Whitney test



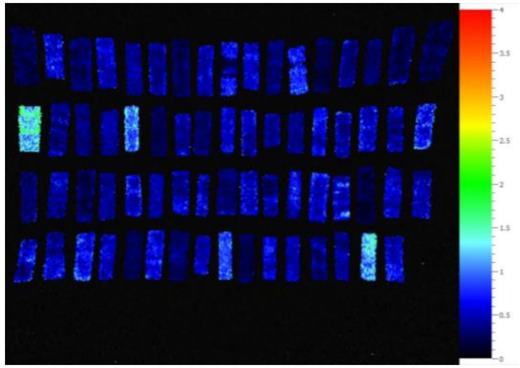
The Predictive issue

In crops: although single genes can affect complex traits, such genes typically operate in conjunction with soil and fertilizer management regimes, the hundreds of other genes involved in crop domestication and adaptation, and so on. Moreover, measurements of phenotypes are noisy

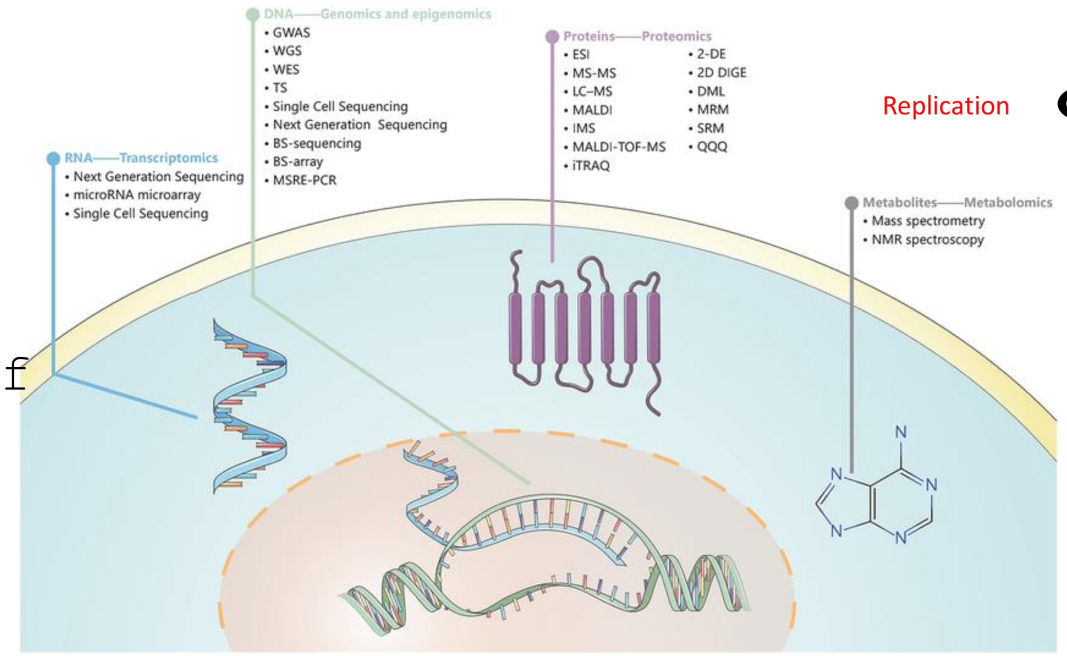


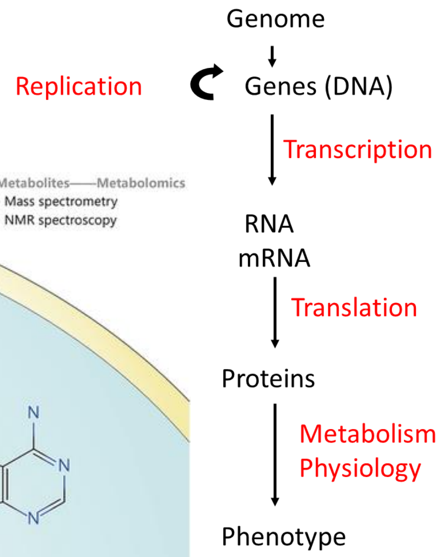
G x E x M

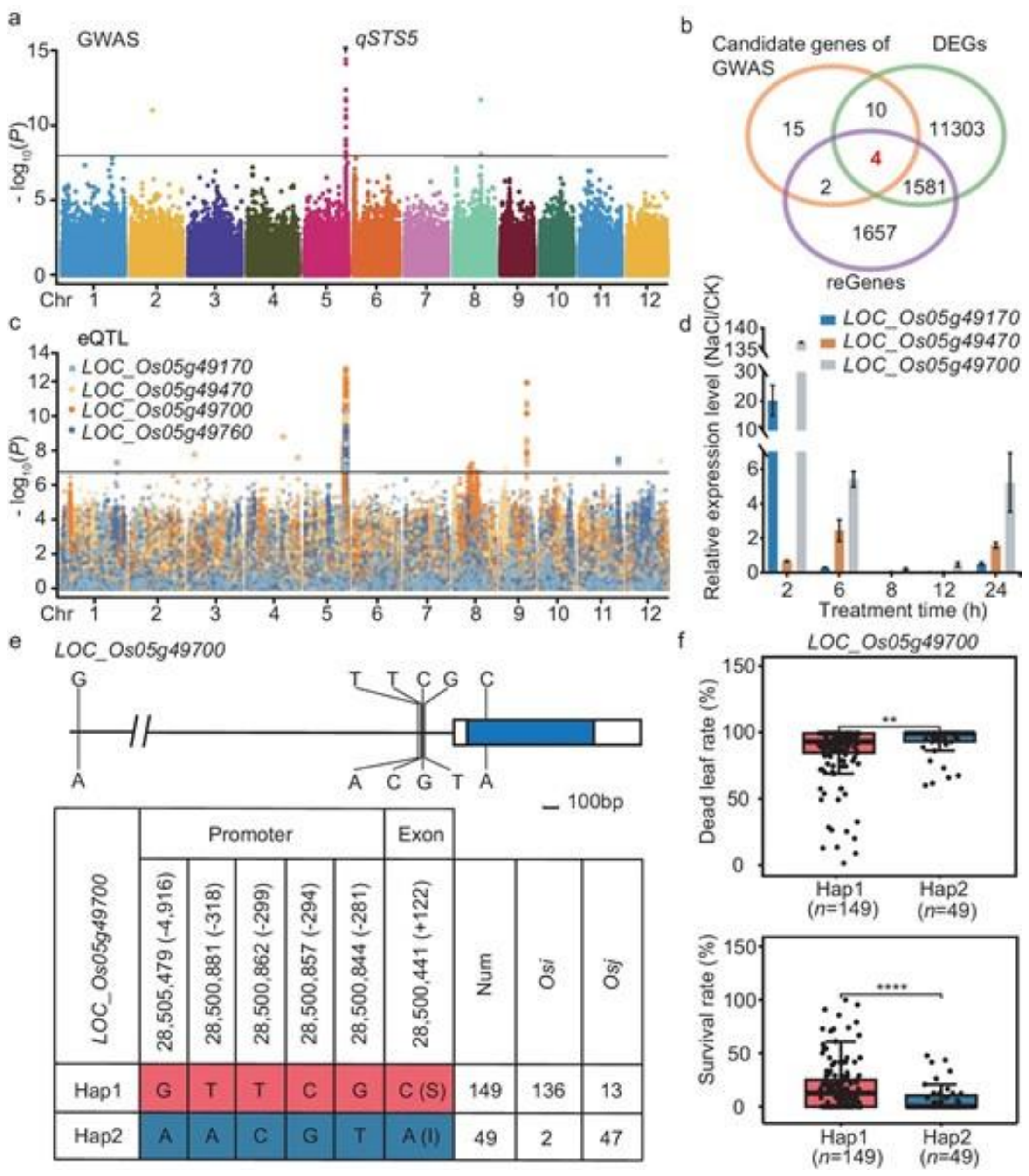
Complexity of determination of pheno



Accuracy of phenotypes







Uncovering key salt-tolerant regulators through a combined eQTL and GWAS analysis using the super pan-genome in rice

Hua Wei[1,†], Xianmeng Wang[1,†], Zhipeng Zhang[1,†], Longbo Yang[1,†], Qianqian Zhang[1,†], Yilin Li[1], Huiying He[1], Dandan Chen[1], Bin Zhang[1], Chongke Zheng[2], Yue Leng[1], Xinglan Cao[1], Yan Cui[1], Chuanlin Shi[1], Yifan Liu[1], Yang Lv[1,3], Jie Ma[1,3], Wenchuang He[1], Xiangpei Liu[1], Qiang Xu[1], Qiaoling Yuan[1], Xiaoman Yu[1], Tianyi Wang[1], Hongge Qian[1], Xiaoxia Li[1], Bintao Zhang[1], Hong Zhang[1], Wu Chen[1], Mingliang Guo[1], Xiaofan Dai[1], Yuexing Wang[3], Xiaoming Zheng[4], Longbiao Guo[3], Xianzhi Xie[2,*], Qian Qian[1,3,5,*] and Lianguang Shang[1,5,*]

202 fully sequenced rice lines with RNA seq available

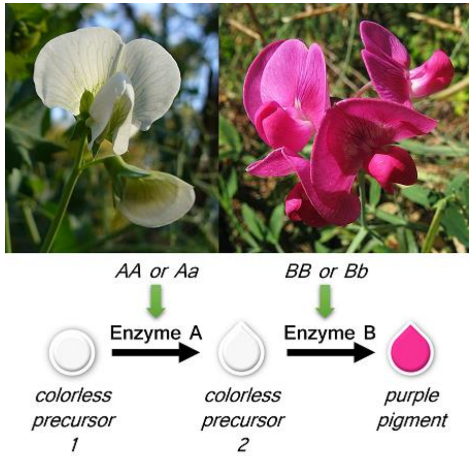
- GWAS finds association with salt resistance
- Additional associations are unveiled by the use of transcriptomic data (TWAS)
- Pangenome data enables the defintion of multiple alleles and haplotypes

The Mechanicistic issue

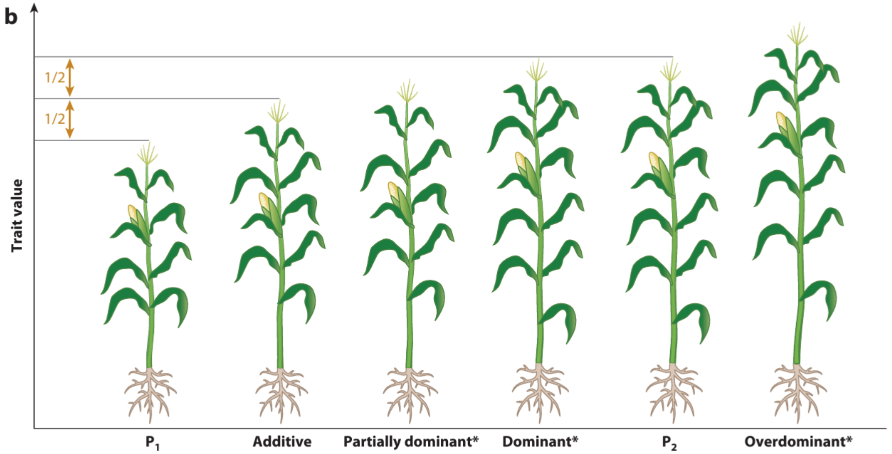
There are different mechanisms of contribution to traits, and many are still poorly understood and/or hard to model

Non-additive effects

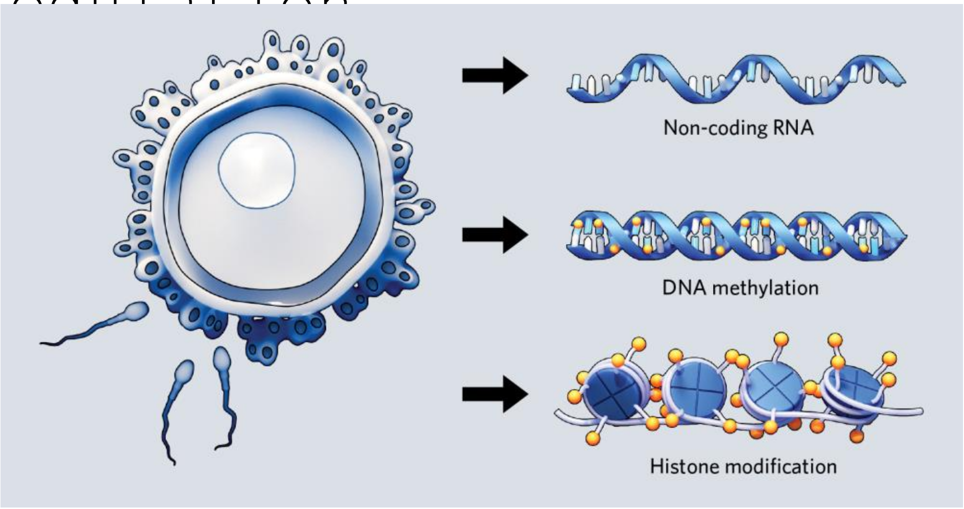
Epistasis



Heterosis



Epigenetics / post-transcriptional regulation



PNAS RESEARCH ARTICLE | AGRICULTURAL SCIENCES OPEN ACCESS

Epistatic QTLs for yield heterosis in tomato

Shai Torgeman[a] and Dani Zamir[a,1]

Edited by Loren Rieseberg, The University of British Columbia, Vancouver, Canada; received June 19, 2022; accepted October 14, 2022

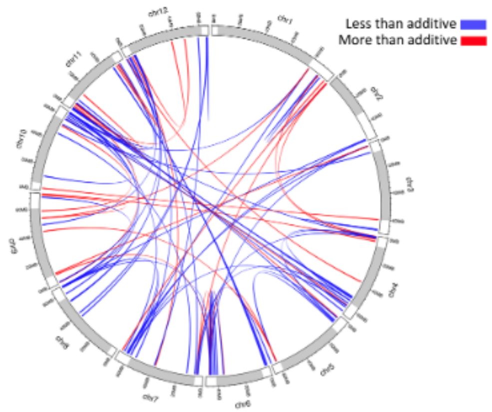
80 epistatic combinations, 19 more-than-additive

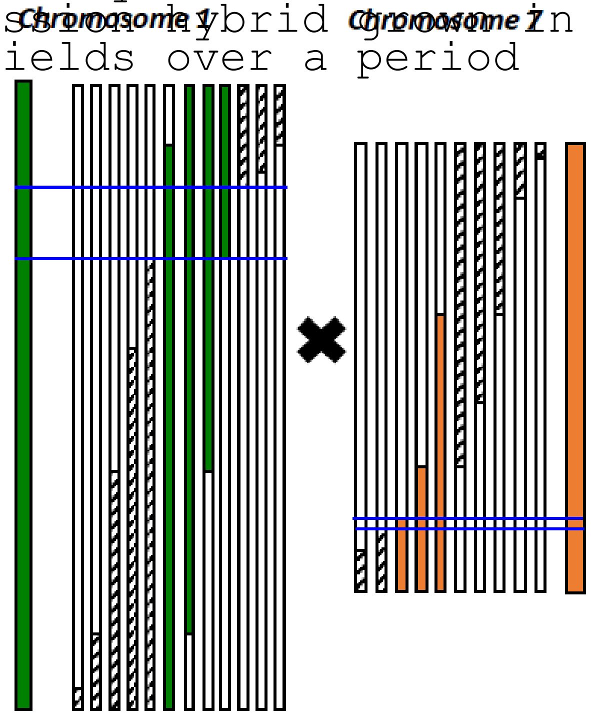
- Validated a single epistatic interaction involving *S. pennellii* QTLs on chromosomes 1 and 7, that independently did not affect yield, increased fruit yield by 20 to 50% in the double introgression hybrid genome in irrigated and dry fields over a period of 4 y

A

Genotypes & traits	Number of epistasis cases	Less than additive	More than additive
BILs PW(kg)	18	13	5
BILH PW(kg)	3	3	0
BILs FW(gr)	14	14	0
BILH FW(gr)	1	0	1
BILs BX(%)	4	3	1
BILH BX(%)	16	5	11
BILs TY(kg)	18	18	0
BILH TY(kg)	6	5	1
Total	80	61	19

B





ARTICLE

Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals

Valentin Hivert,[1] Julia Sidorenko,[1] Florian Rohart,[1] Michael E. Goddard,[2,3] Jian Yang,[1,4] Naomi R. Wray,[1,5] Loic Yengo,[1,6] and Peter M. Visscher[1,6,*]

Summary

Non-additive genetic variance for complex traits is traditionally estimated from data on relatives. It is notoriously difficult to estimate without bias in non-laboratory species, including humans, because of possible confounding with environmental covariance among relatives. In principle, non-additive variance attributable to common DNA variants can be estimated from a random sample of unrelated individuals with genome-wide SNP data. Here, we jointly estimate the proportion of variance explained by additive (h^2_{SNP}), dominance (δ^2_{SNP}) and additive-by-additive (η^2_{SNP}) genetic variance in a single analysis model. We first show by simulations that our model leads to unbiased estimates and provide a new theory to predict standard errors estimated using either least-squares or maximum likelihood. We then apply the model to 70 complex traits using 254,679 unrelated individuals from the UK Biobank and 1.1 M genotyped and imputed SNPs. We found strong evidence for additive variance (average across traits $\bar{h}^2_{SNP} = 0.208$). In contrast, the average estimate of $\bar{\delta}^2_{SNP}$ across traits was 0.001, implying negligible dominance variance at causal variants tagged by common SNPs. The average epistatic variance $\bar{\eta}^2_{SNP}$ across the traits was 0.055, not significantly different from zero because of the large sampling variance. Our results provide new evidence that genetic variance for complex traits is predominantly additive and that sample sizes of many millions of unrelated individuals are needed to estimate epistatic variance with sufficient precision.

Non-additive variance can be easily detected in experimental conditions (e.g. hybrid generation), however it is hard to estimate correctly in colletions of unrelated individuals

- UK biobank data, 1M SNPs and 250K individuals, different traits
- Additive variance has a clear contribution, not so much dominance and epistasis

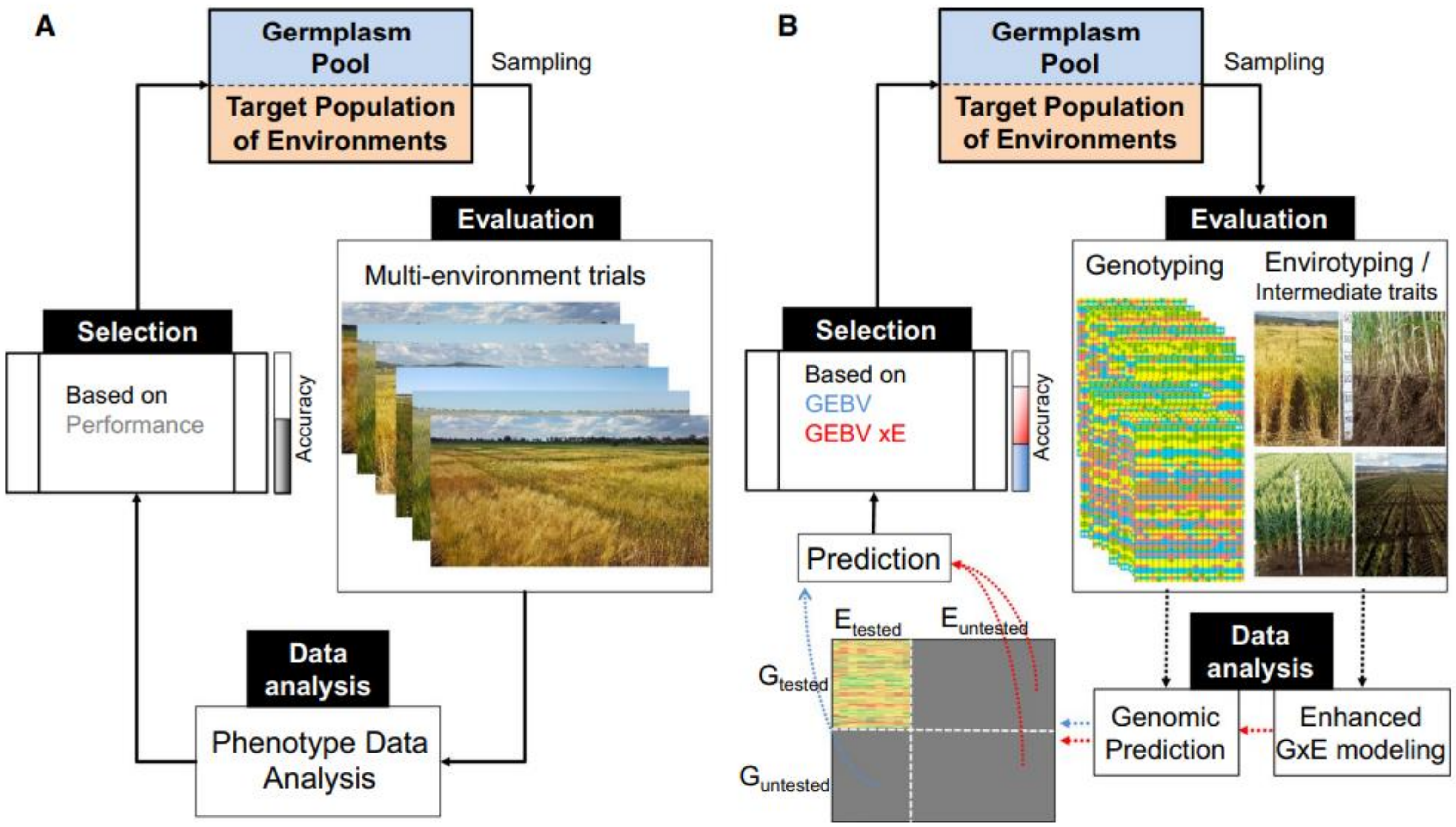
Acknowledging complexity – the breeders’ way

It doesn’t really matter which gene does what; based on large observational datasets it is possible to model the relation between genome-wide diversity and phenotypic outputs (and predict traits)

- Depending on the trait, the data, and the model, you may end up with different prediction accuracies

Accelerating crop genetic gains with genomic selection

Kai Peter Voss-Fels[1] · Mark Cooper[1] · Ben John Hayes[1]



GS models do a fairly good job capturing the complexity of the trait, and adding prior information (e.g. GWAS hits) makes a little difference

- However, GS models are sensitive to the representativeness of the training of the model (and bound to the allelic diversity and LD captured in the process)

RESEARCH ARTICLE Open Access

Exploring the potential of incremental feature selection to improve genomic prediction accuracy

Felix Heinrich[1*], Thomas Martin Lange[1], Magdalena Kircher[2], Faisal Ramzan[3], Armin Otto Schmitt[1,4†] and Mehmet Gültas[4,5*†]

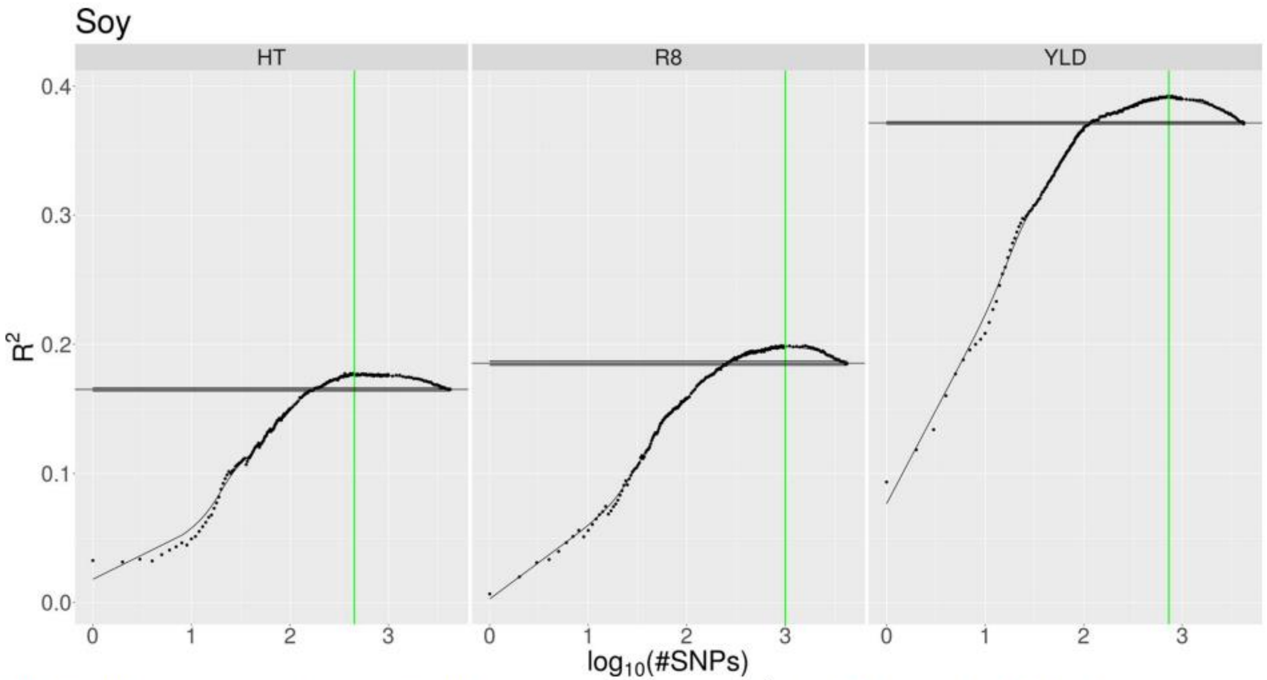
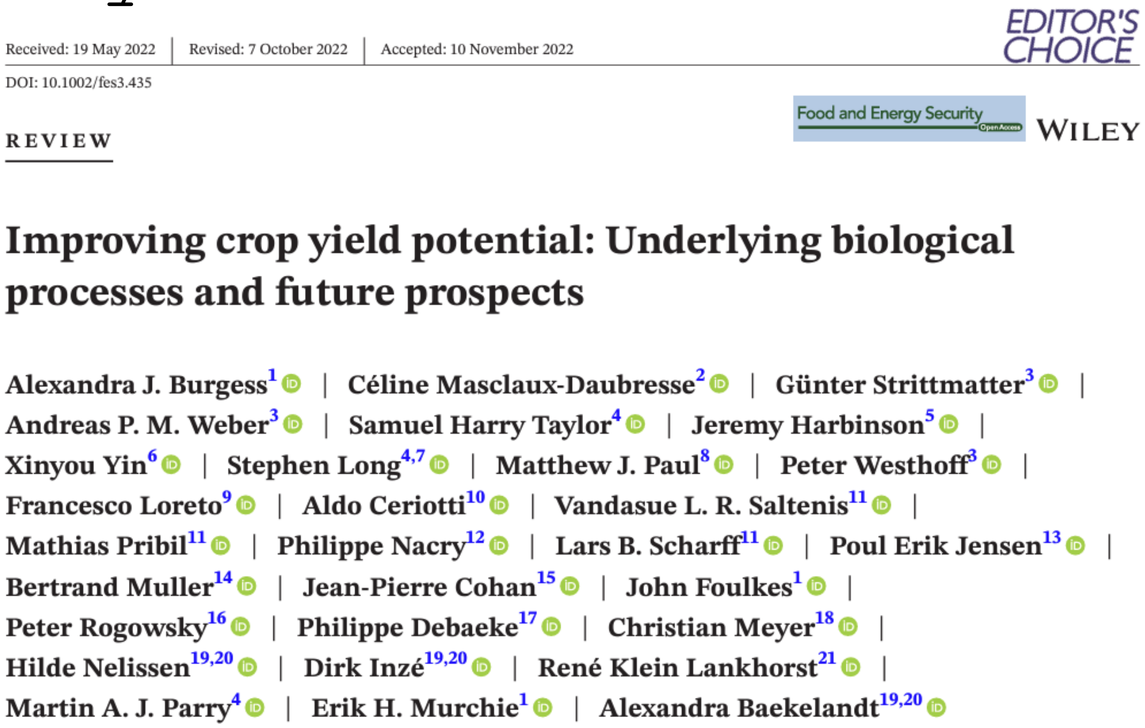


Fig. 2 Prediction accuracy of soy phenotypes. Prediction accuracy (measured as mean R^2) of soy phenotypes as a function of the number of SNPs used for the model (presented as logarithmic values) on the Φ data. The trend estimate, represented by the solid black curve, is obtained through smoothing. The maximum accuracy is indicated by the vertical green line. Mean performance of the model when trained on all SNPs is represented by the horizontal black line, with the shaded interval around it indicating the standard error of the mean of 10 cross-validation repetitions. The prediction accuracy of all three traits could be increased by up to 0.02 using IFS

Acknowledging complexity – the molecular geneticist way


EDITOR'S CHOICE

Received: 19 May 2022 | Revised: 7 October 2022 | Accepted: 10 November 2022

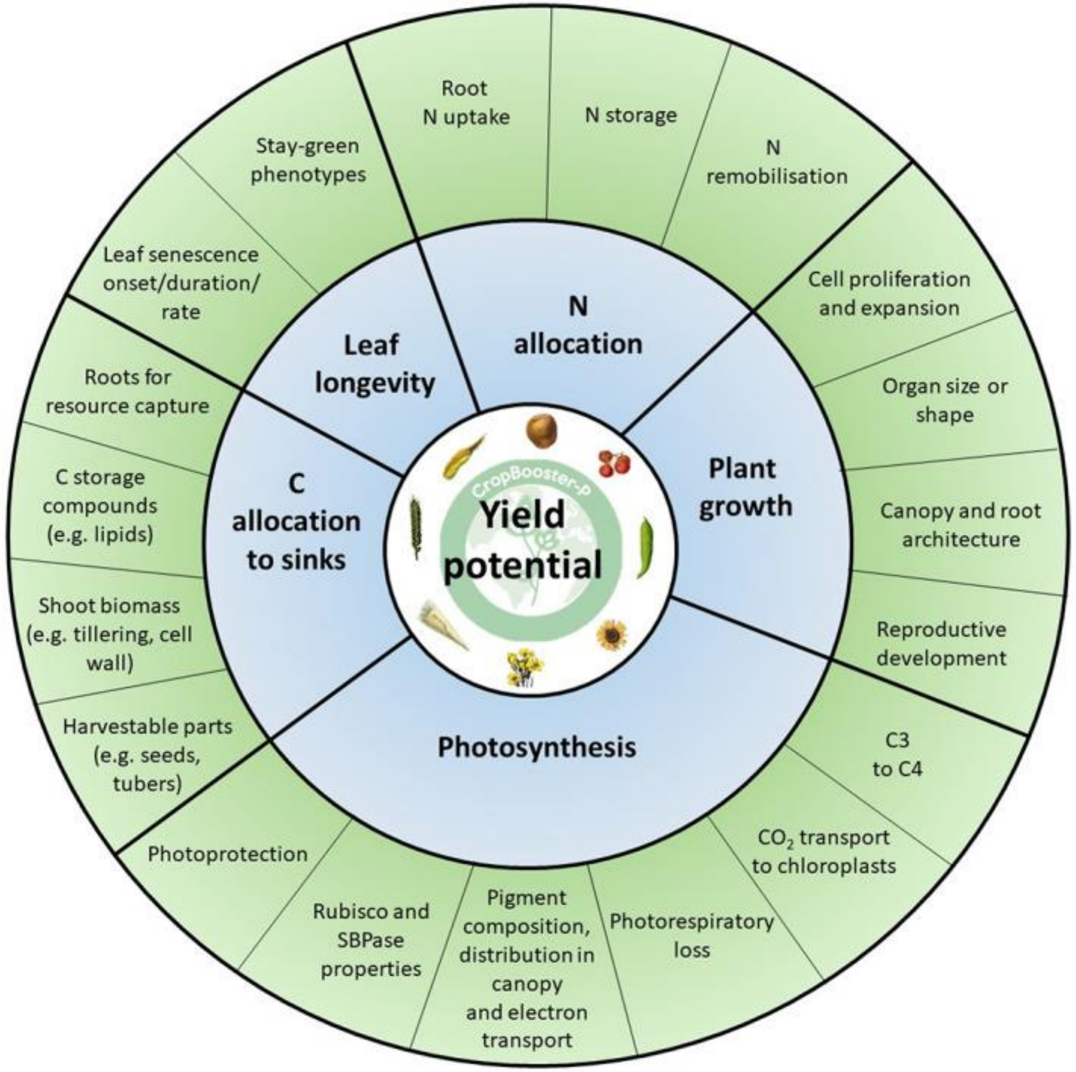
DOI: 10.1002/fes3.435

Food and Energy Security Open Access WILEY

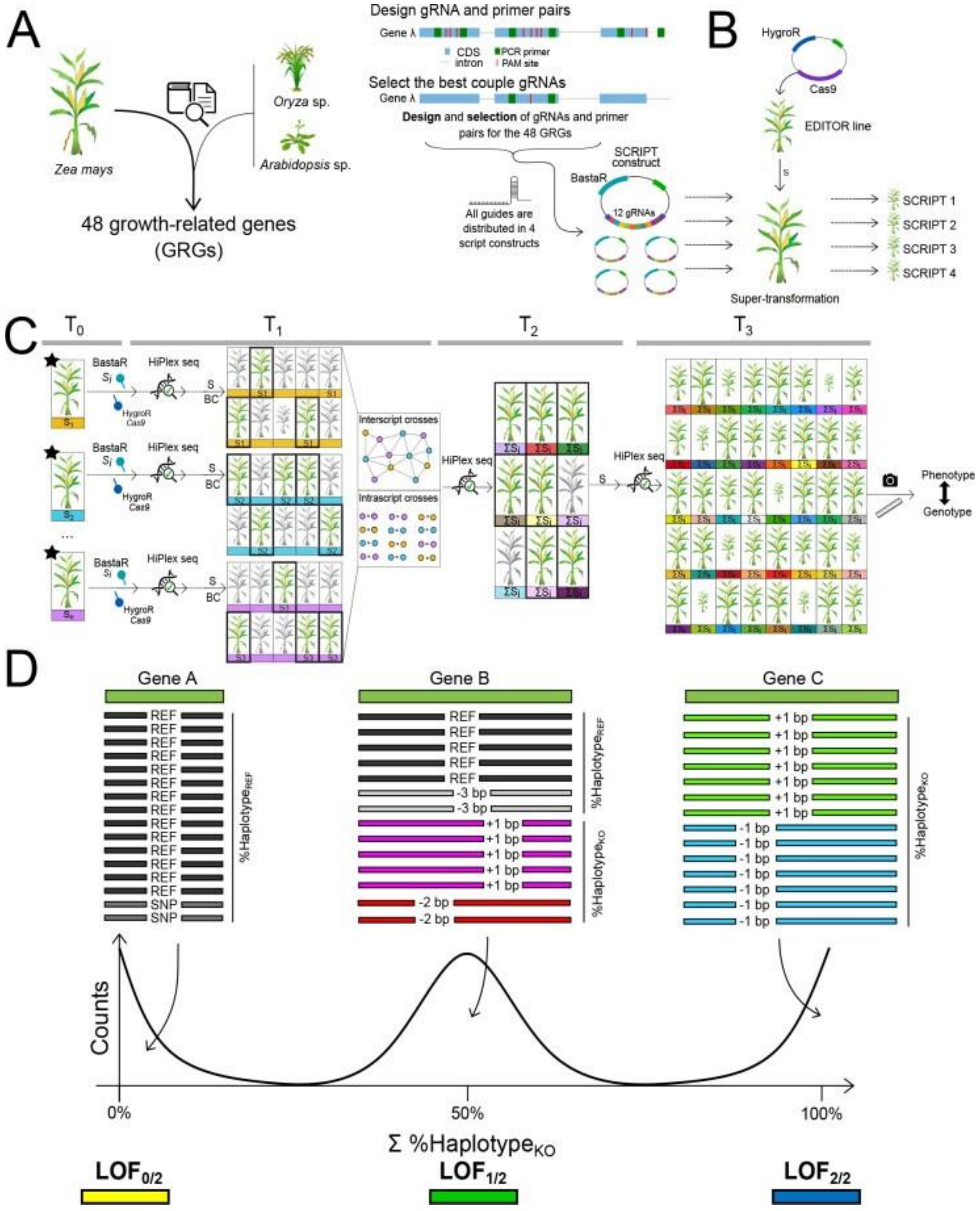
REVIEW

Improving crop yield potential: Underlying biological processes and future prospects

Alexandra J. Burgess[1] | Céline Masclaux-Daubresse[2] | Günter Strittmatter[3] | Andreas P. M. Weber[3] | Samuel Harry Taylor[4] | Jeremy Harbinson[5] | Xinyou Yin[6] | Stephen Long[4,7] | Matthew J. Paul[8] | Peter Westhoff[3] | Francesco Loreto[9] | Aldo Ceriotti[10] | Vandasue L. R. Saltenis[11] | Mathias Pribil[11] | Philippe Nacry[12] | Lars B. Scharff[11] | Poul Erik Jensen[13] | Bertrand Muller[14] | Jean-Pierre Cohan[15] | John Foulkes[1] | Peter Rogowsky[16] | Philippe Debaeke[17] | Christian Meyer[18] | Hilde Nelissen[19,20] | Dirk Inzé[19,20] | René Klein Lankhorst[21] | Martin A. J. Parry[4] | Erik H. Murchie[1] | Alexandra Baekelandt[19,20]



There's a finite number of molecular targets that can significantly impact yield, should there be allelic variation

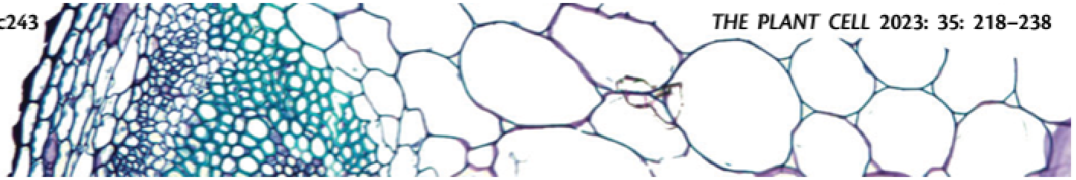


Nowadays, genome editing enables simultaneous targeting of multiple loci – the bottleneck being knowledge of what to edit

https://doi.org/10.1093/plcell/koac243

THE PLANT CELL 2023: 35: 218–238

THE PLANT CELL



BREEDIT: a multiplex genome editing strategy to improve complex quantitative traits in maize

Christian Damian Lorenzo [1,2] Kevin Debray [1,2] Denia Herwegh [1,2] Ward Develtere [1,2] Lennert Impens [1,2] Dries Schaumont [3] Wout Vandeputte [1,2] Stijn Aesaert [1,2] Griet Coussens [1,2] Yara De Boe [1,2] Kirin Demuynck [1,2] Tom Van Hautegem [1,2] Laurens Pauwels [1,2] Thomas B. Jacobs [1,2] Tom Ruttink [3] Hilde Nelissen [1,2] and Dirk Inzé [1,2,*,†]