Omics-based optimisation of hybrid performance and heterosis in winter oilseed rape

Rod Snowdon, Justus Liebig University, Giessen, Germany

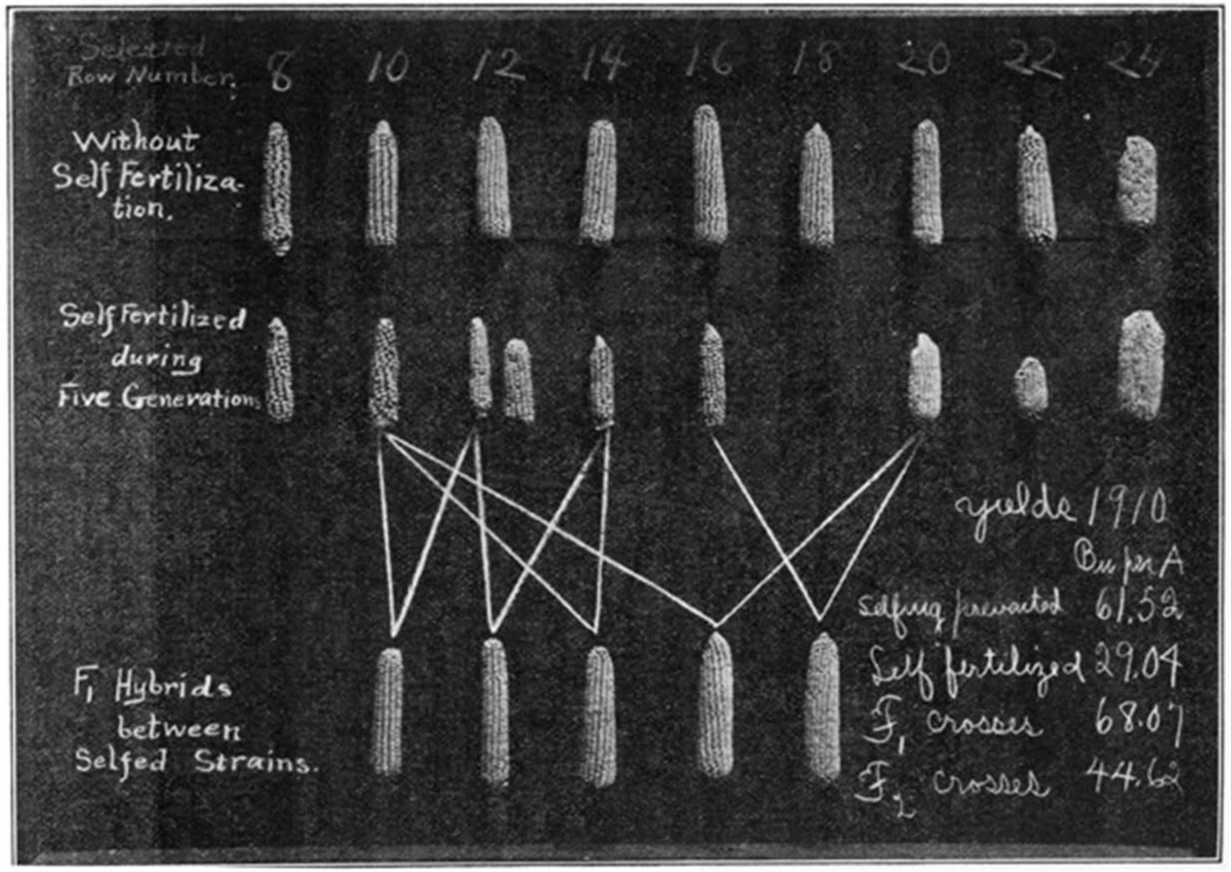


Image: George Harrison Shull (1874-1954)

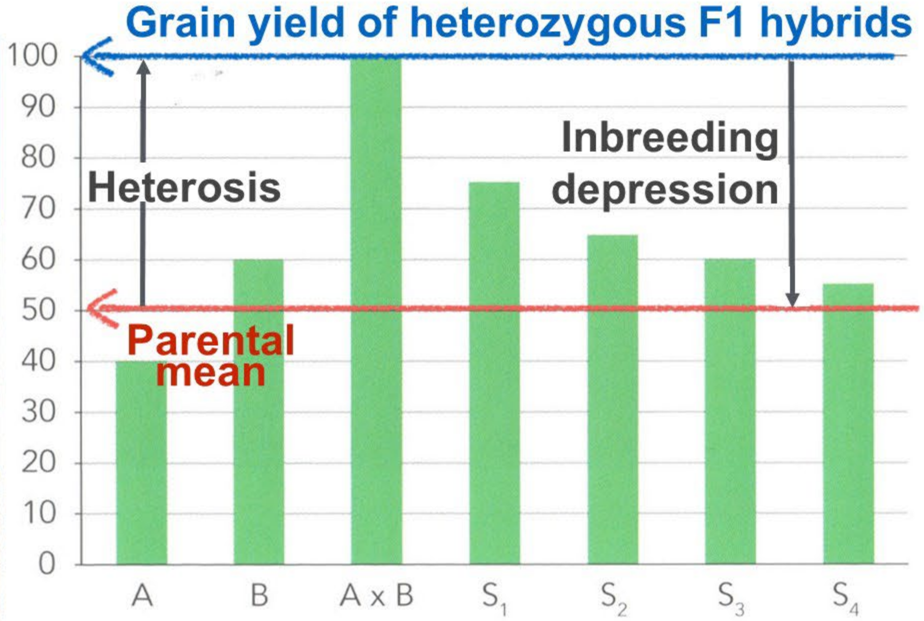
Hybrid breeding

Exploiting heterosis in allogamous crops

Heterosis: Discovered by Shull ~100 years ago, systematically exploited to establish hybrid breeding in maize

Maize: Easy to prevent self-pollination (i.e. to guarantee cross-pollination) by mechanical (manual) removal of the tassel

Result: Very strong and ongoing breeding progress worldwide with very high genetic gain



Adapted from Miedaner (2010) Basics of Plant Breeding

Exploiting heterosis in crop breeding

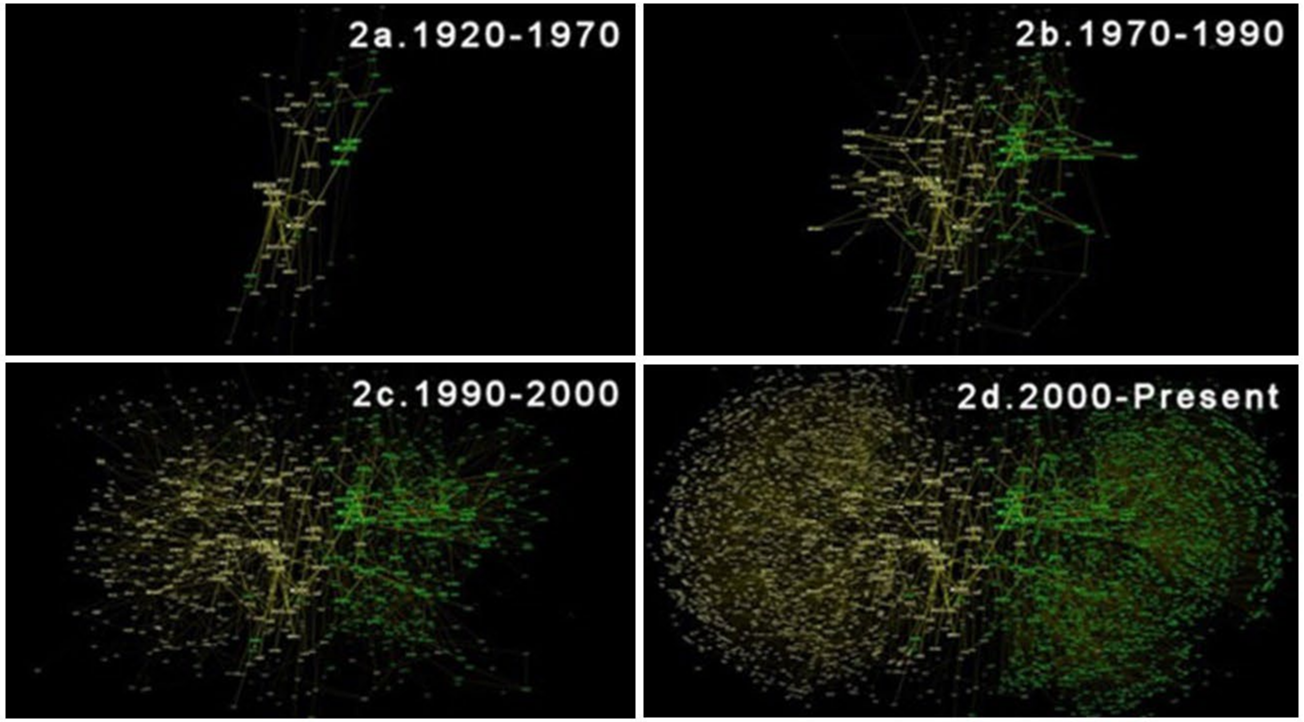
Maize sets the benchmark for successful hybrid breeding

1) Effective, distinct and diverse **heterotic pools** developed over ~100 years of hybrid breeding (exclusively intra-pool crossing)

2) Strong **heterotic dominance effects are effectively fixed** in opposite pools

3) Genetic gain now driven mainly by increasing ***per se* parental performance within pools**

Example: 100 years of US maize breeding, Du Pont Pioneer



Source: A. Rafalski 2007

Exploiting heterosis in other crops

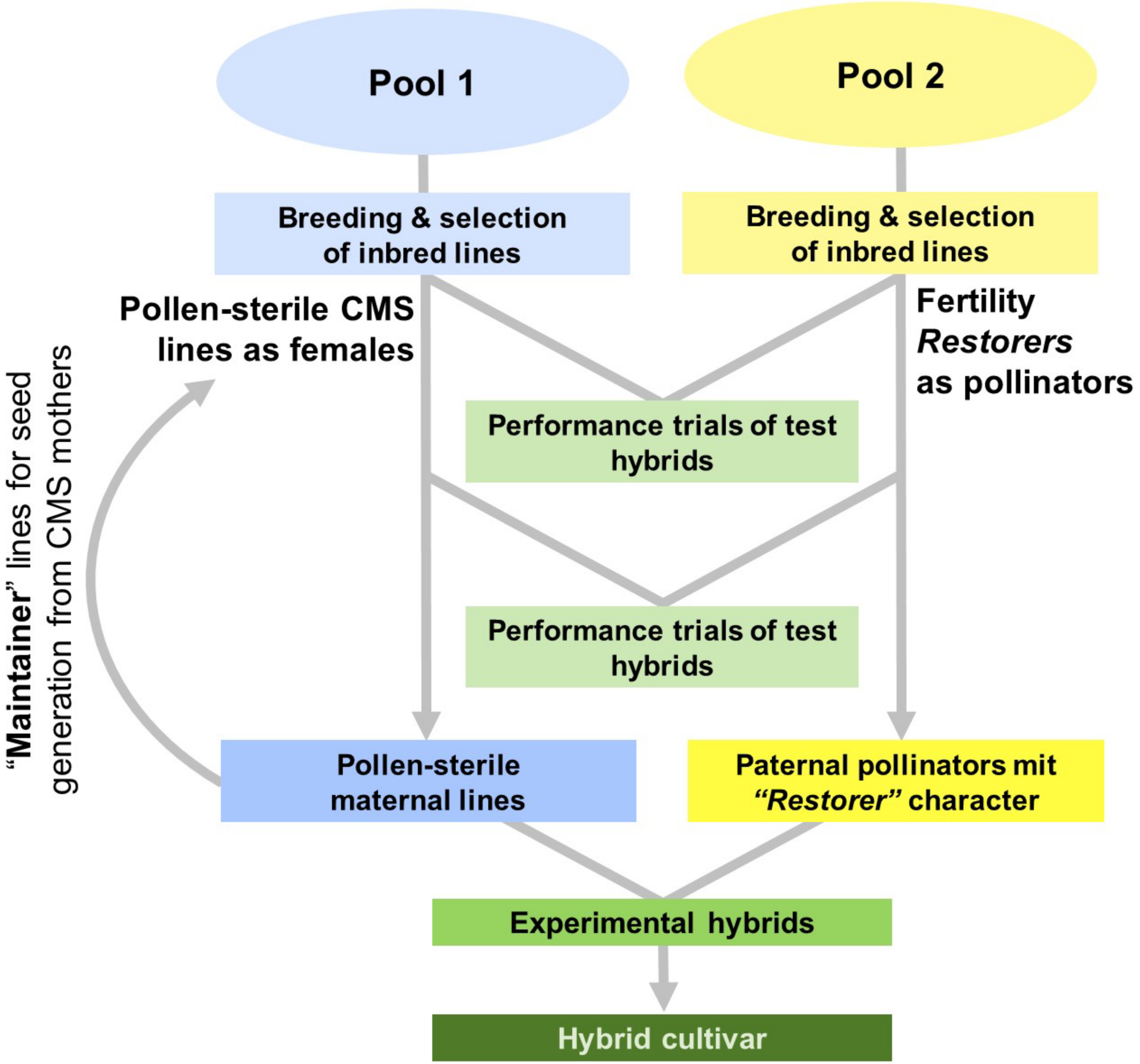
is not quite so easy…

We still need genetically distinct, diverse heterotic pools …

but we also need male-sterility & fertility restoration systems …

and effective procedures to select good-performing parents with high hybrid potential **(“general combining ability” – GCA)** …

and to identify the top potential hybrid combinations (with high **“specific combining ability” – SCA**)



The challenge

How to optimise hybrid breeding and heterosis in a facultative inbreeding crop after a century of pedigree breeding?

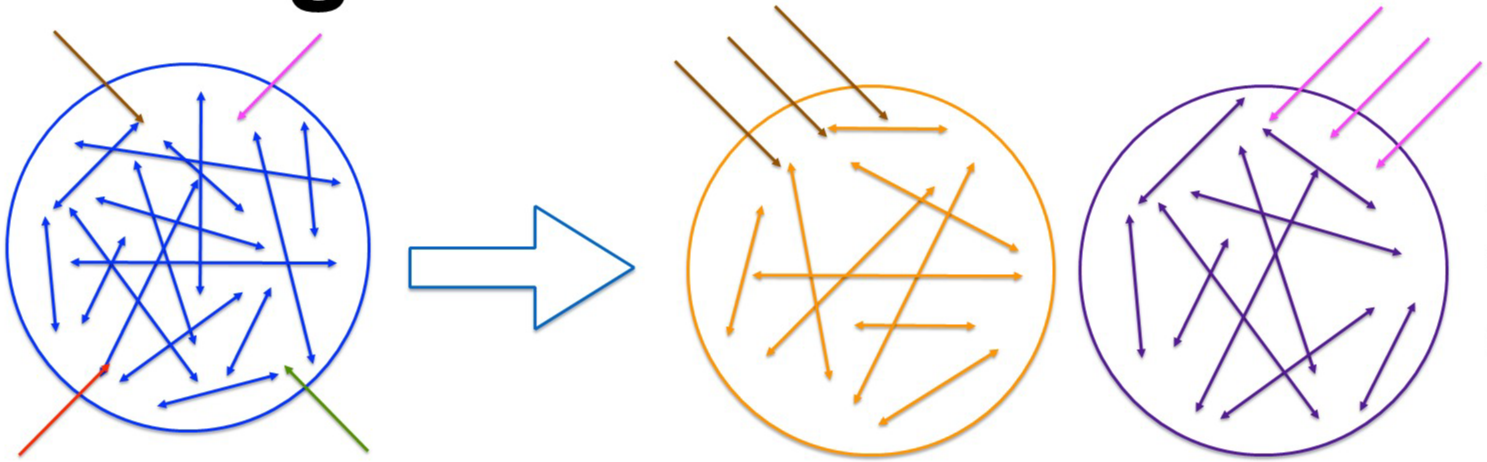
Winter rapeseed: No effective heterotic pools after >100 years of pedigree breeding with “best-by-best” crossing

Low level of heterosis reduces potential benefits in terms of hybrid genetic gain

High effort and cost for test hybrid production and evaluation

Transitioning from pedigree to hybrid breeding

Pedigree breeding
“My best” ⊗ “Your best”
>> even better



Hybrid breeding
“Pool 1” ⊗ “Pool 2”
>> better than both

Can *B. napus* “omics” data and advanced prediction methods help …

… to separate and recombine heterotic pools from a mixed gene pool?

… to predict and separate dominant heterotic patterns into opposite pools by exploiting **presence-absence** and **methylation** variants?

… to fix epistatic heterosis effects in opposite pools based on complementary patterns of **expression and regulatory elements?**

Participants & methods

JLU Giessen

Matthias Frisch, Eva Herzog: Simulation-based cross designs, analysis of haplotype structure and recombination/diversity patterns

Sven Weber, Agnieszka Golicz, Rod Snowdon: Genomics data generation and analysis, statistical and "omics"-based predictions

Forschungszentrum Jülich

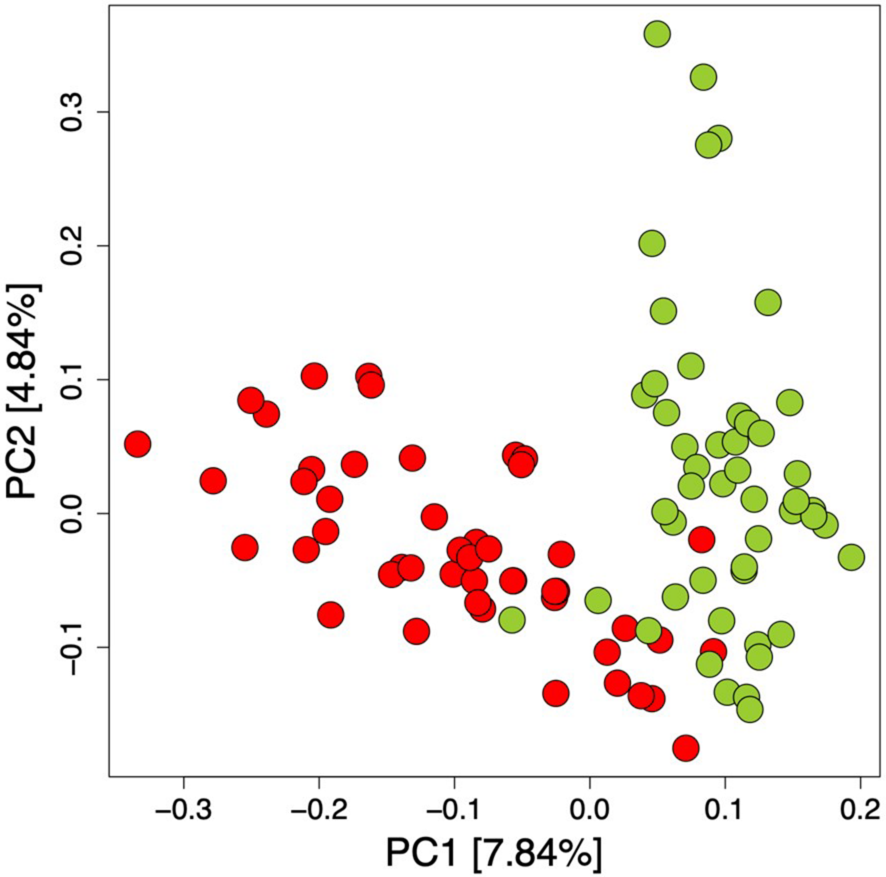
Björn Usadel: Database development, data analysis & visualisation

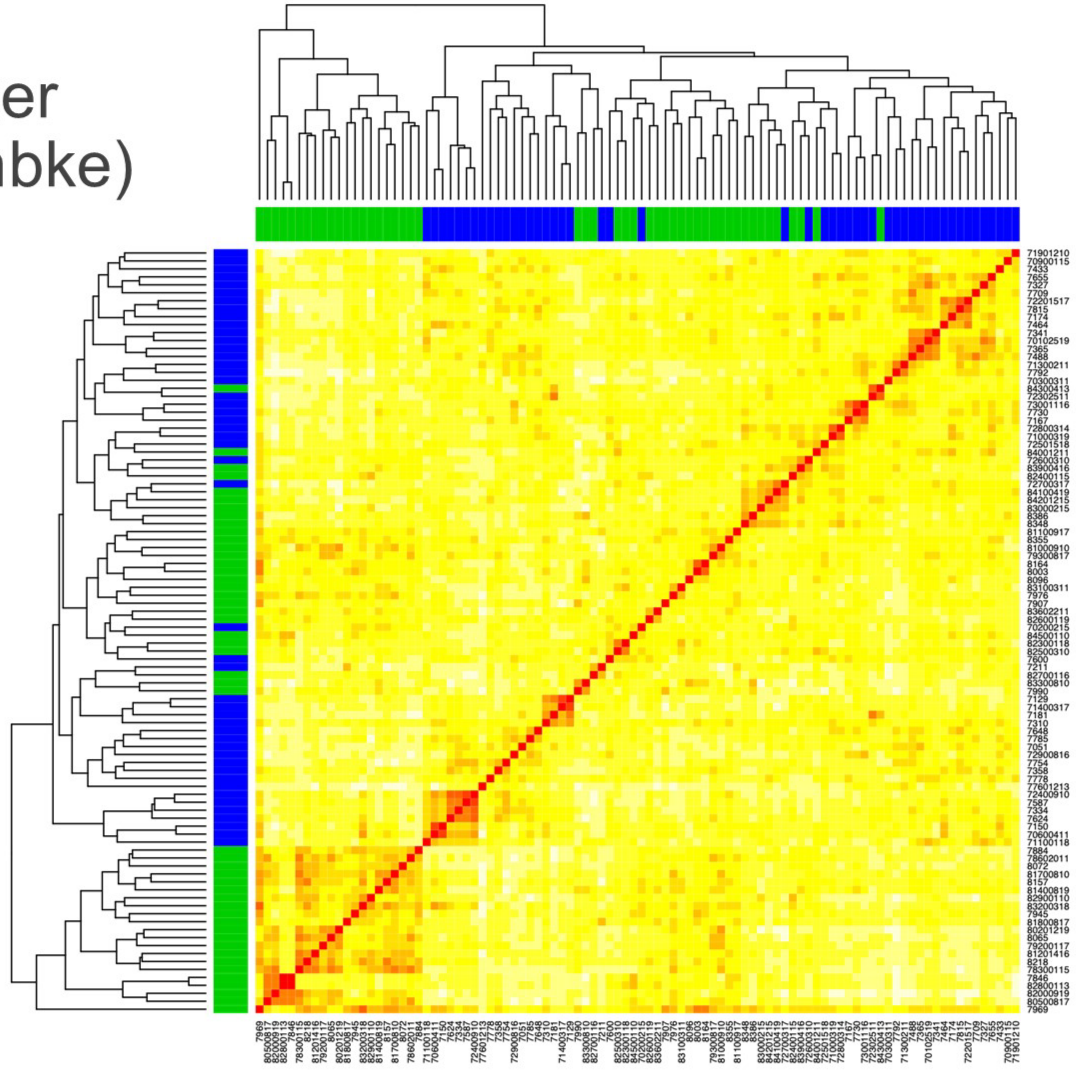
NPZ Innovation GmbH

Amine Abbadi, Tobias Kox, Christian Flachenecker: Crosses, development and field testing of plant pools and test hybrids

Starting materials

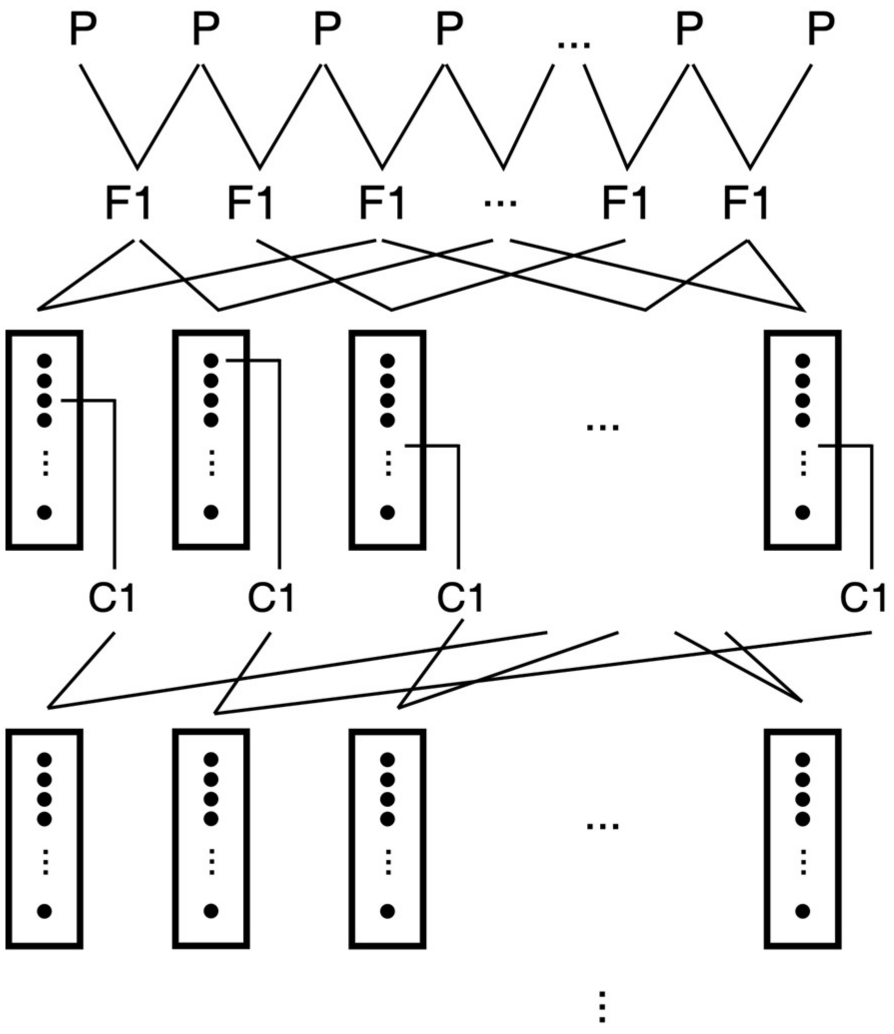
2 x 50 founder lines representing elite winter rapeseed hybrid breeding pools (NPZ Lembke)





Daniel Krenzer, Eva Herzog, Matthias Frisch

I: Promote within-pool recombination



50 Parental lines

R1

50 F1 genotypes

R2

50 C1 families
of size 10

R3

50 C2 families
of size 10

R4

50 C3 families
of size 10

Intra-pool chain crossing design

3 rapid-generation breeding cycles in greenhouse over ~3 years

Target: Generate and genotype 500 offspring per pool per cycle

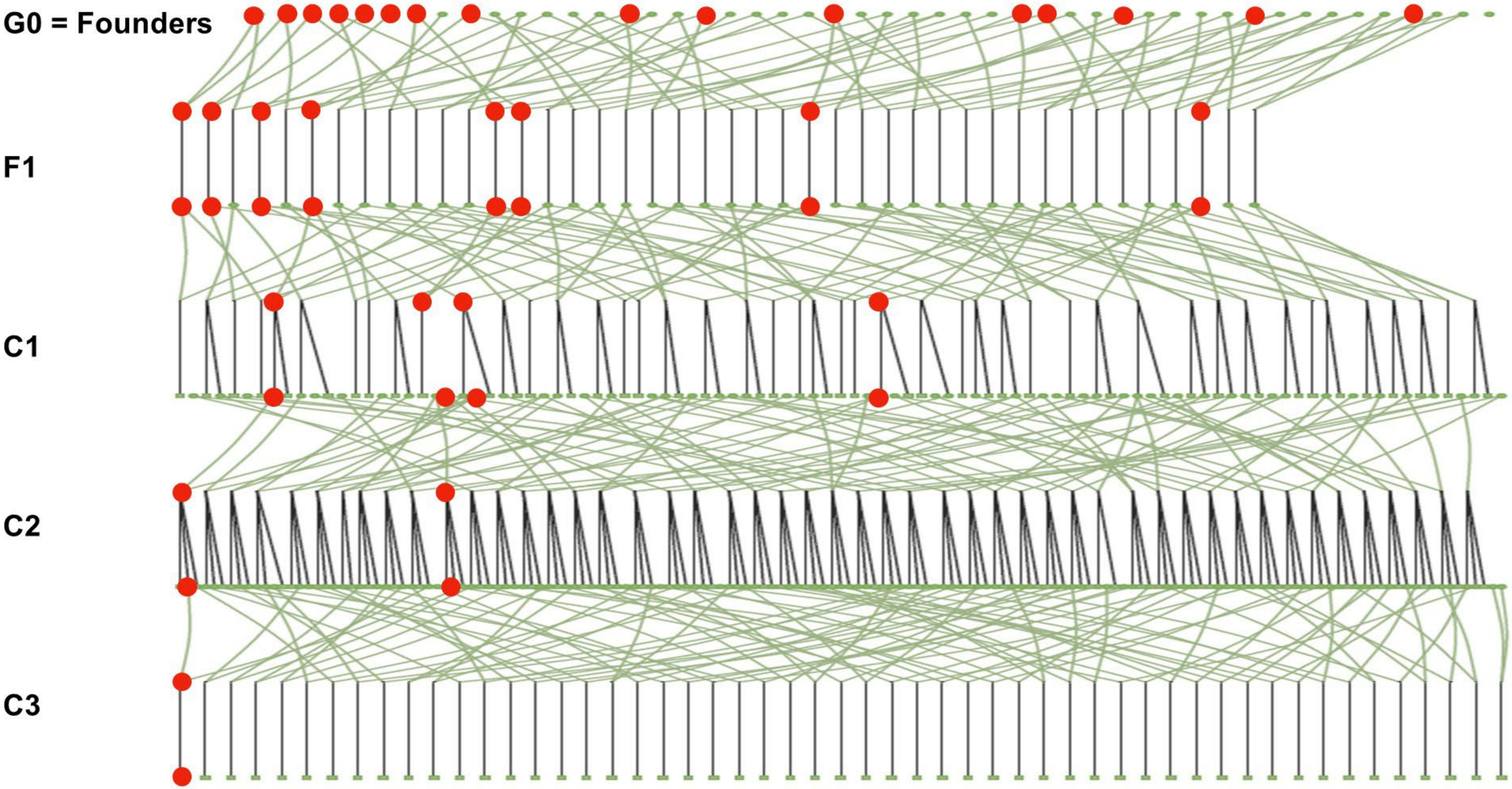
Aim: Maximise recombination but retain and fix the diversity of the 2 sets of 50 founders in each pool, generate RILs/DH lines and test hybrids from genomic-selected offspring with maximum predicted GCA

Daniel Krenzer, Eva Herzog, Matthias Frisch

I: Promote within-pool recombination

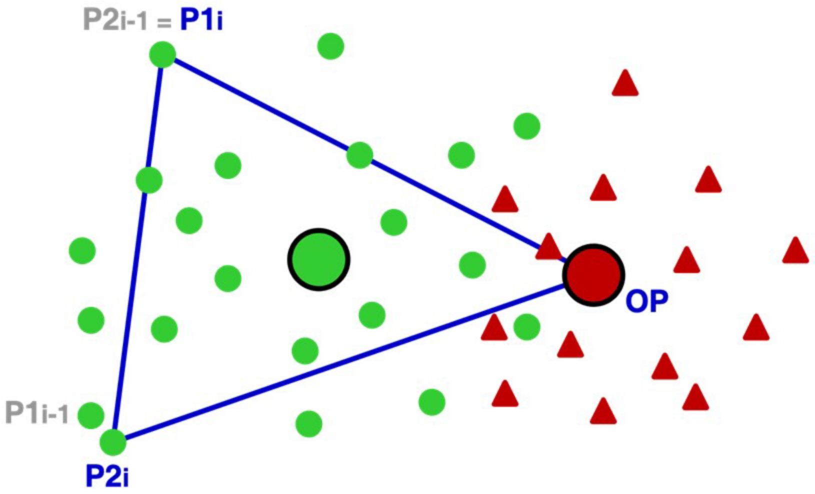
Example:

Pool 1

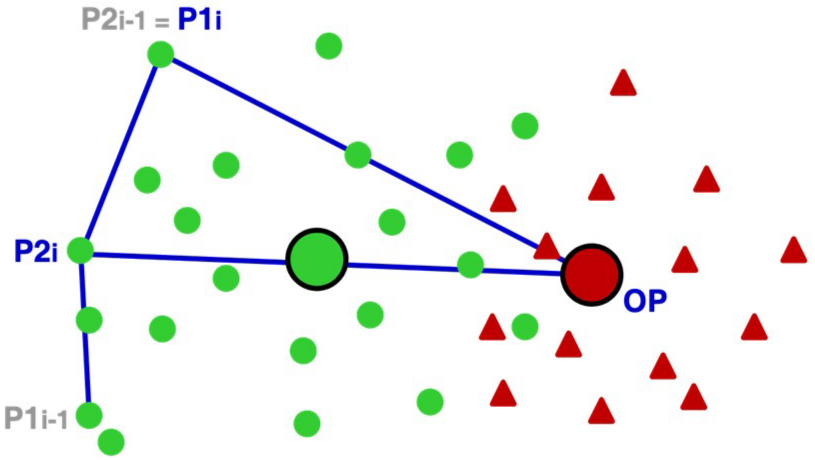


Nayyer Abdollahi, Agnieszka Golicz

II: Promote between-pool separation



Criterion $T' = d(P1_i,OP) + d(P2_i,OP) + d(P1_i,P2_i)$



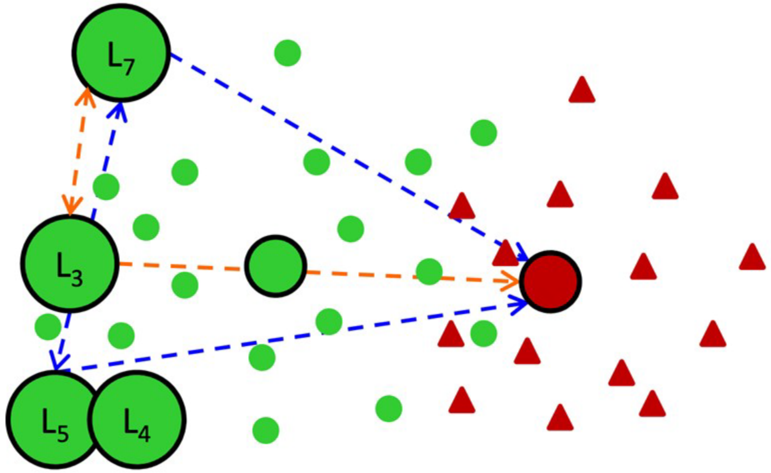
Criterion $T = d(P1_i,OP) + d(P2_i,OP) + d(P1_i,P2_i) + d(P2_i,P1_{i-1})$

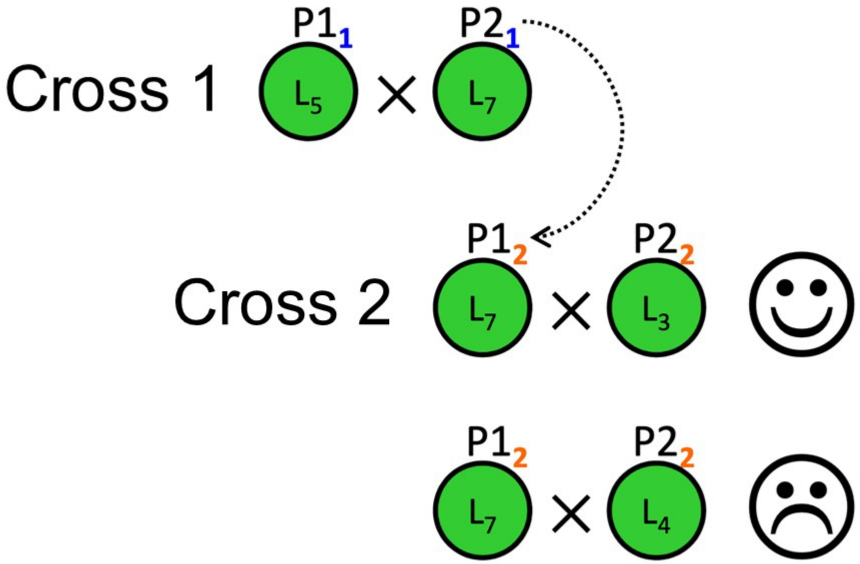
Computer simulations based on genome-wide diversity (assayed with a 15k SNP array) to select most suitable cross combinations

Aim: Maintain within-pool diversity but maximise between-pool diversity

Daniel Krenzer, Eva Herzog, Matthias Frisch

II: Promote between-pool separation



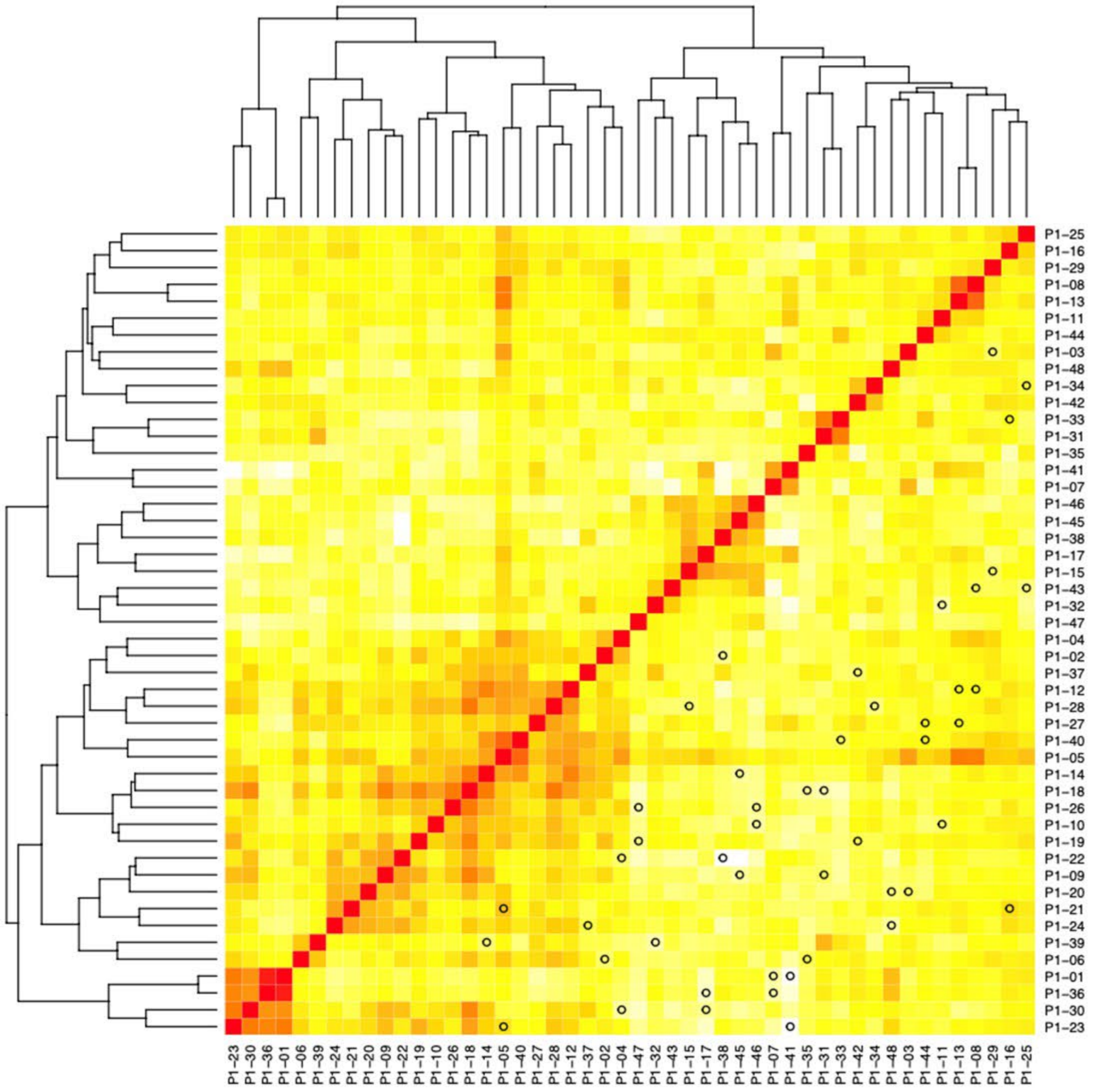


Computer simulations based on genome-wide diversity (assayed with a 15k SNP array) to select most suitable cross combinations

Aim: Maintain within-pool diversity but maximise between-pool diversity

Daniel Krenzer, Eva Herzog, Matthias Frisch

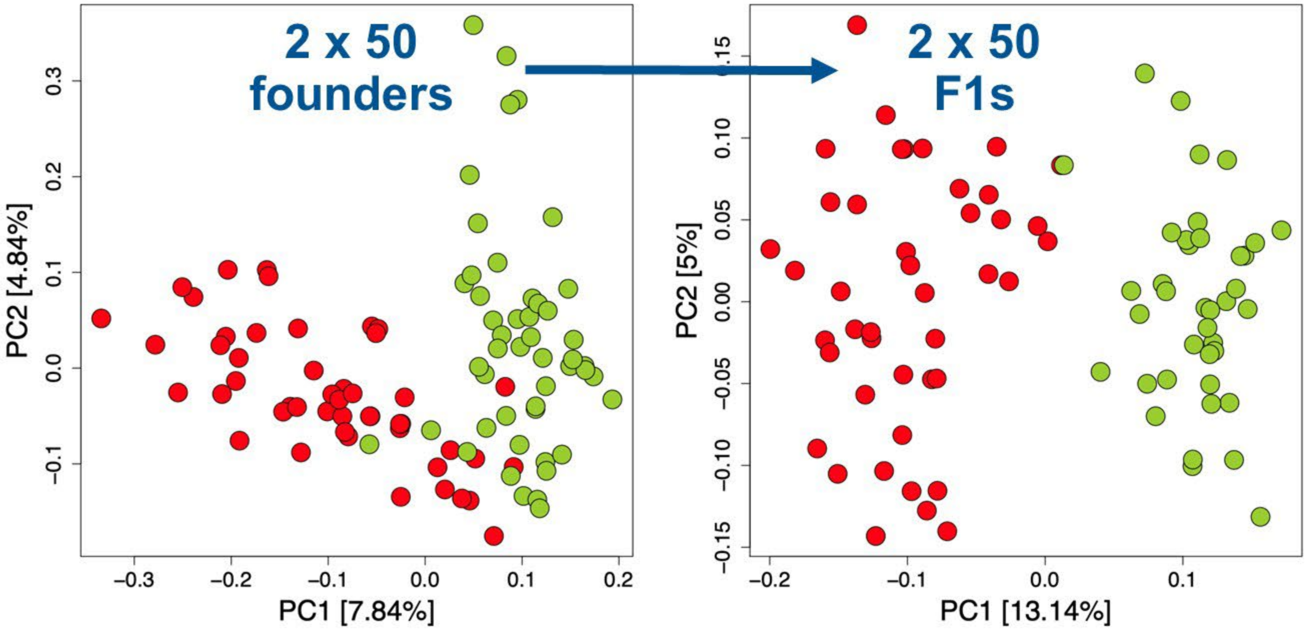
II: Promote between-pool separation



Cycle 1

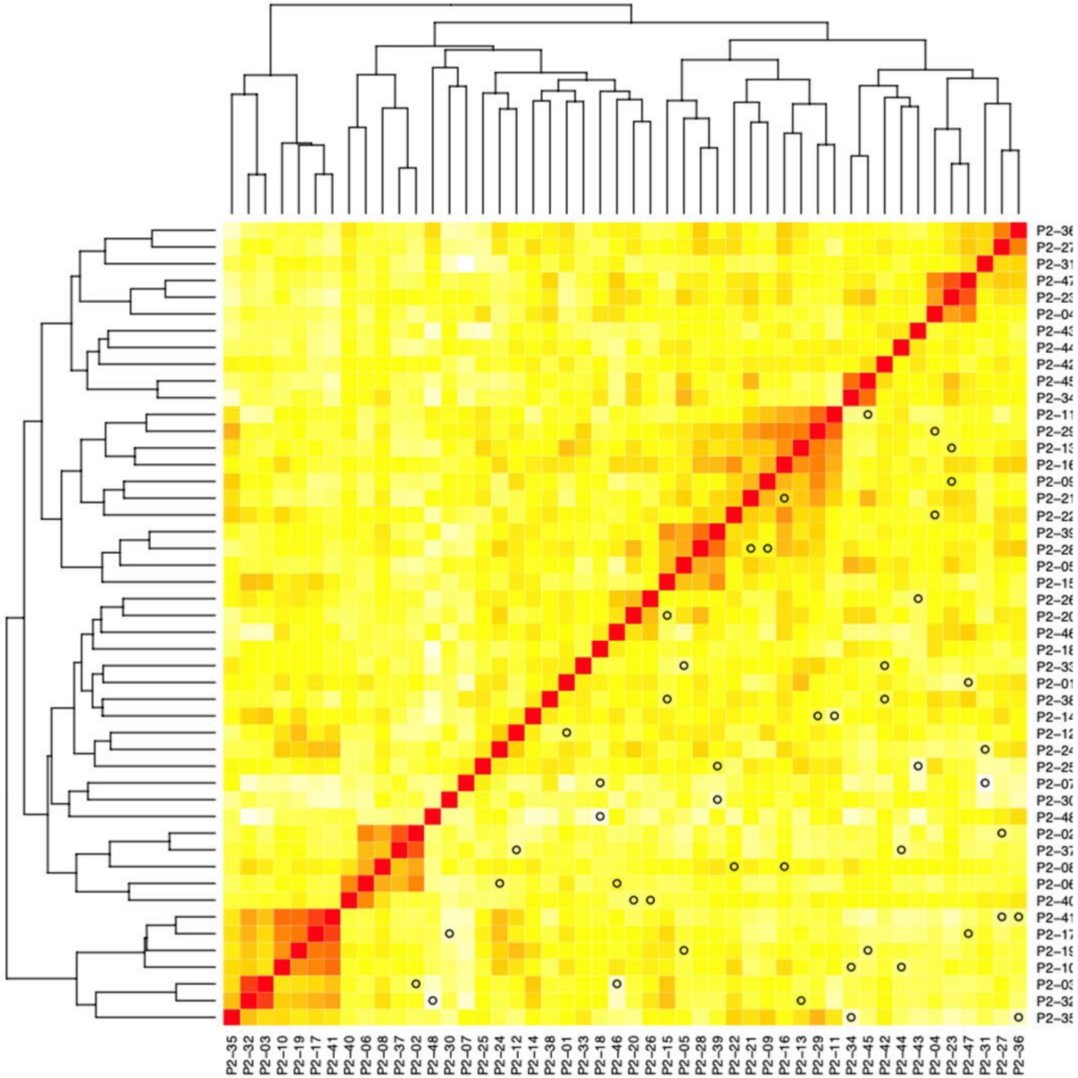
Most suitable simulated initial crosses

Result: Initial pool separation in F1:



Daniel Krenzer, Eva Herzog, Matthias Frisch

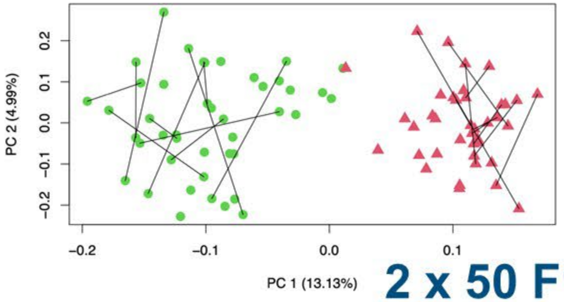
II: Promote between-pool separation



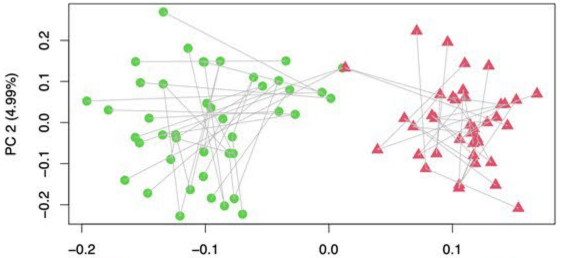
Cycle 1

Most suitable simulated complex crosses

“Best by best”



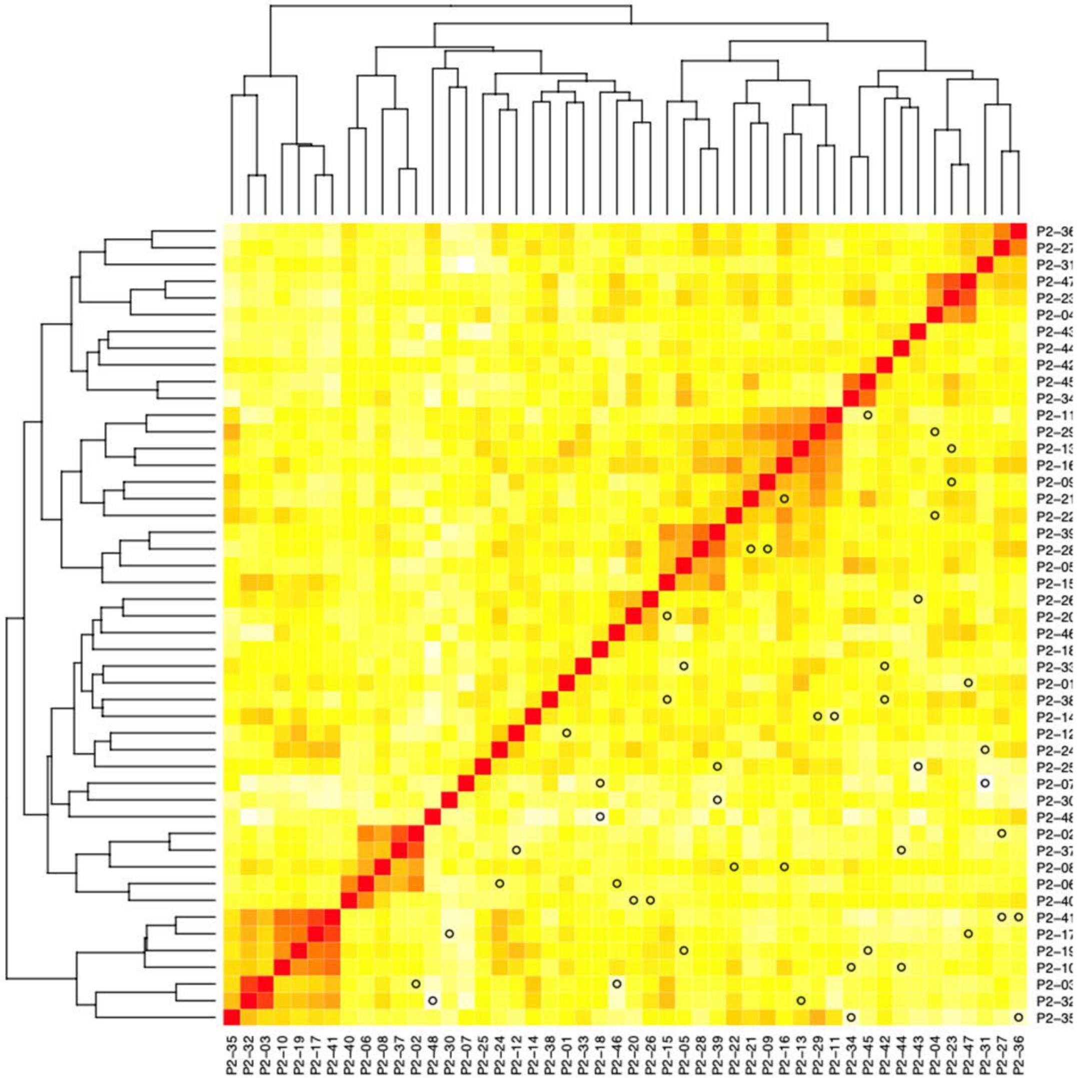
2 x 50 F1s



“Improve the worst”

Daniel Krenzer, Eva Herzog, Matthias Frisch

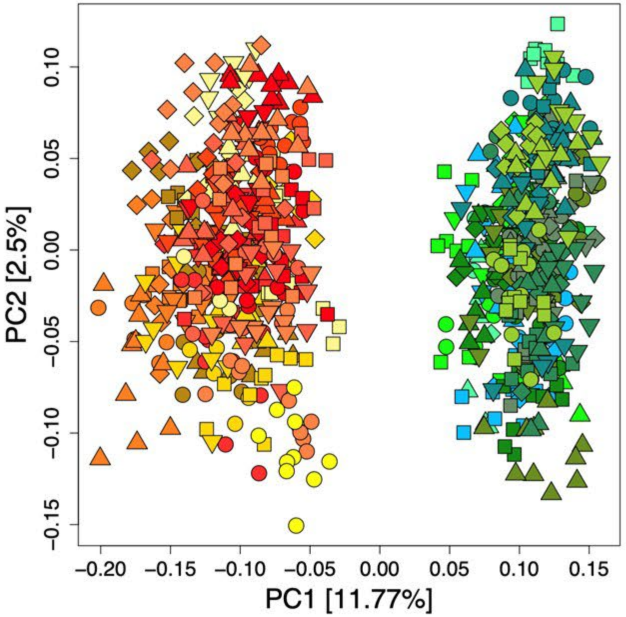
II: Promote between-pool separation



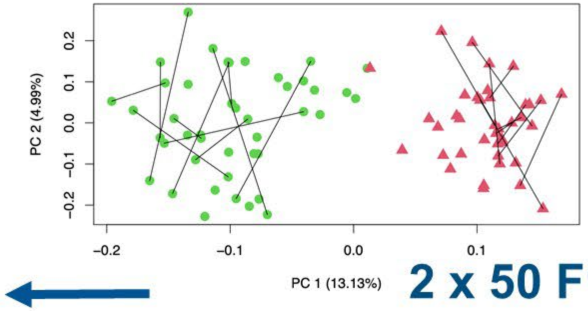
Cycle 1

Most suitable simulated complex crosses

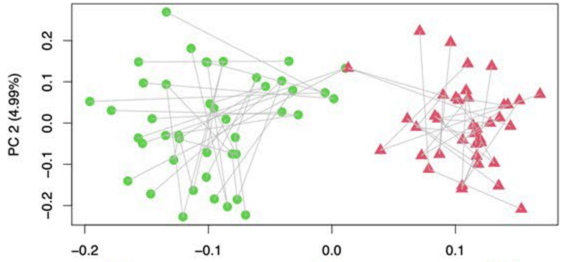
Simulated cycle 1 result



"Best by best"



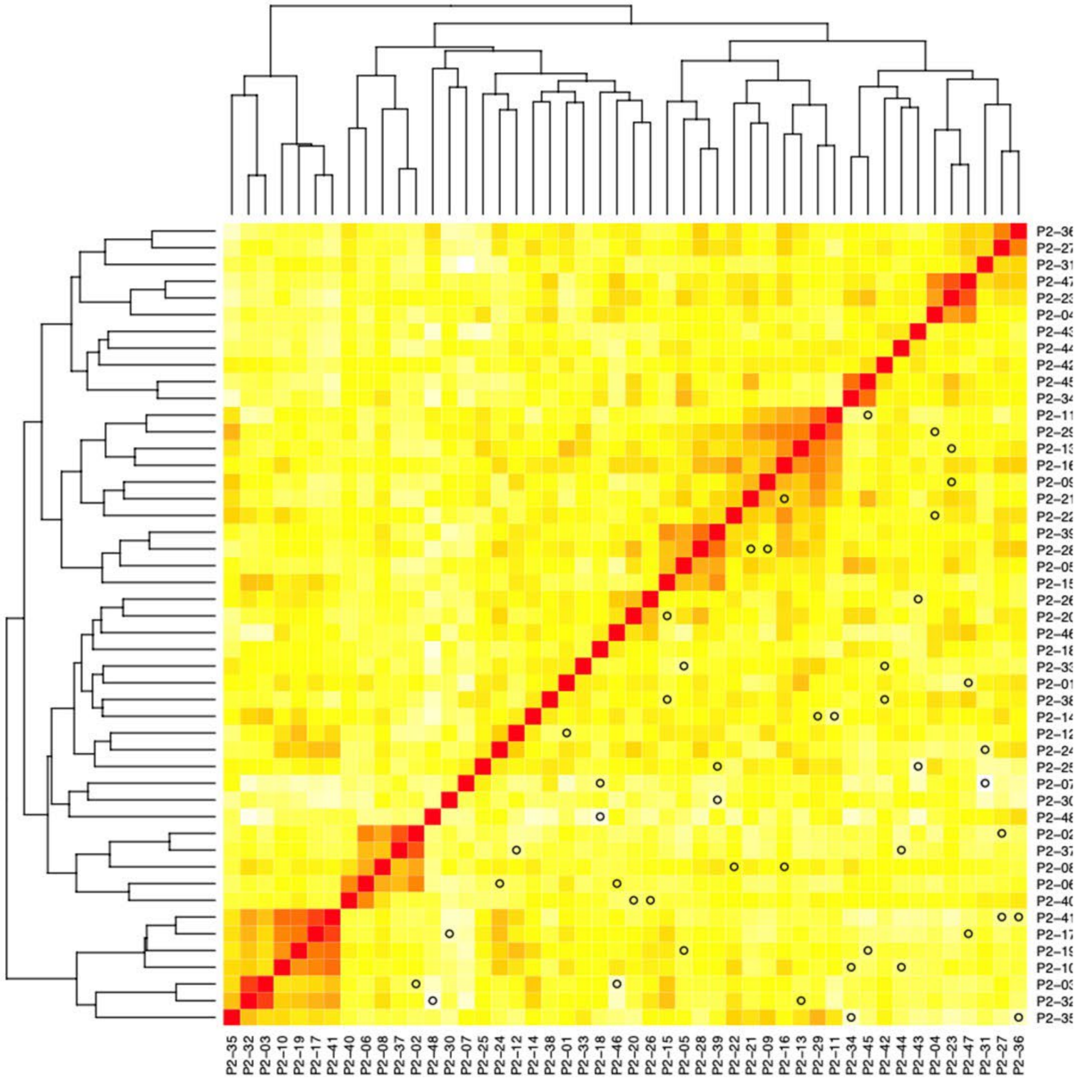
2 x 50 F1s



"Improve the worst"

Daniel Krenzer, Eva Herzog, Matthias Frisch

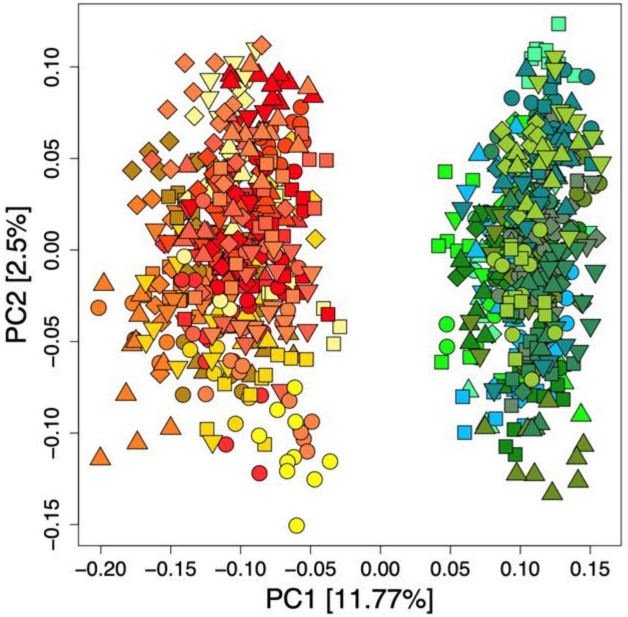
II: Promote between-pool separation



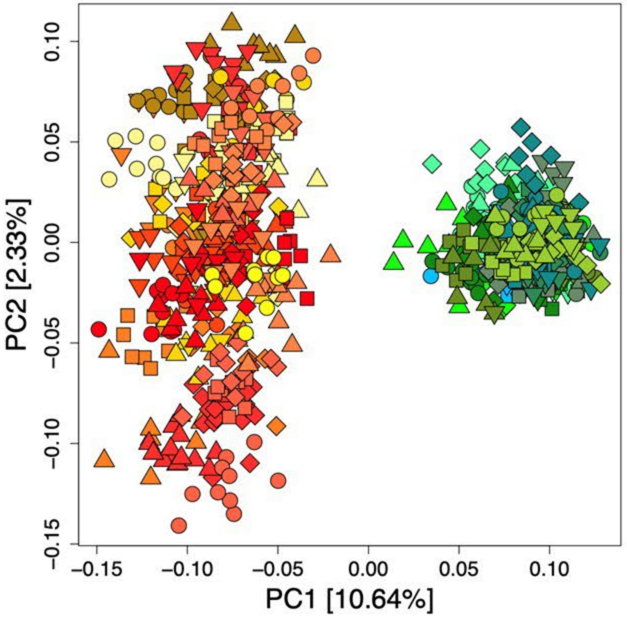
Cycle 1

Most suitable simulated complex crosses

Simulated cycle 1 result

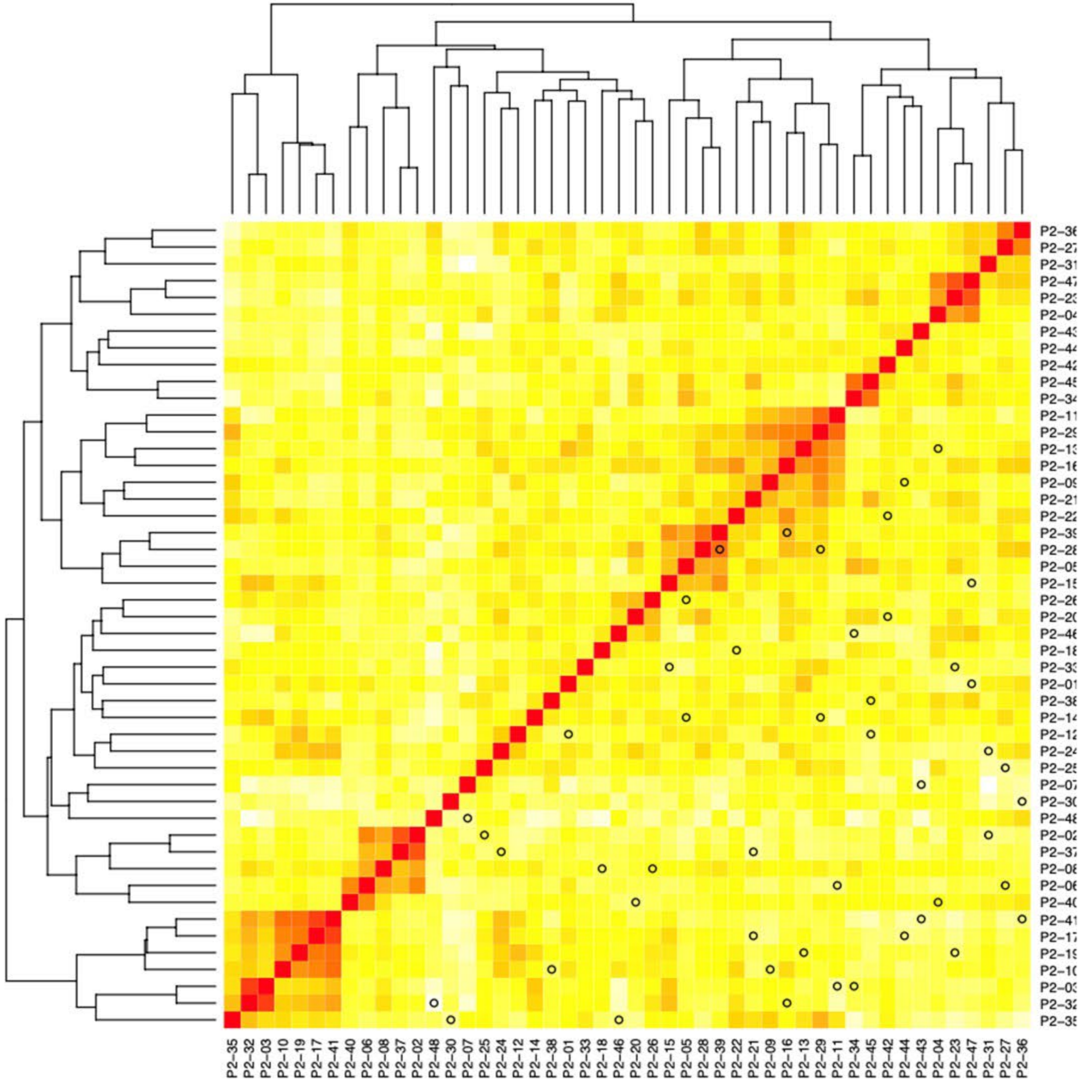


Real cycle 2 result



Daniel Krenzer, Eva Herzog, Matthias Frisch

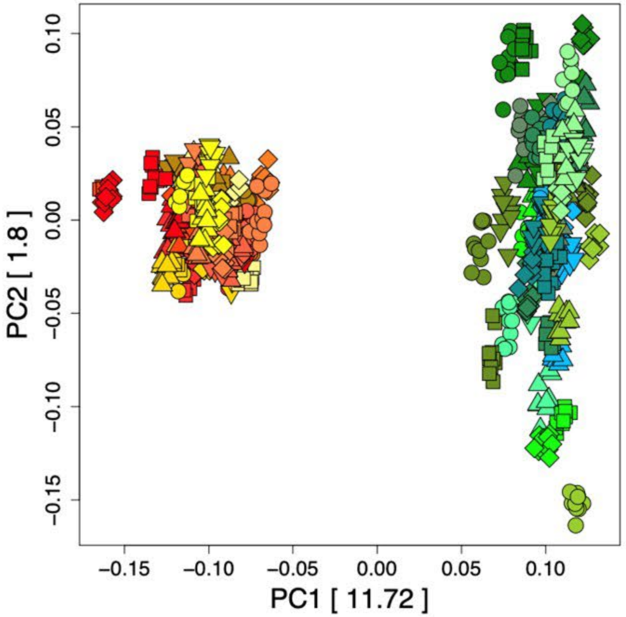
II: Promote between-pool separation



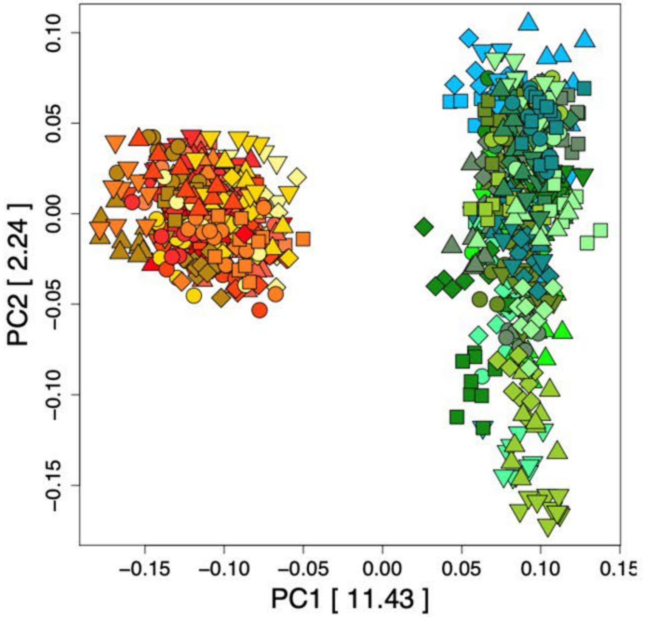
Cycle 2

Most suitable simulated complex crosses

Simulated cycle 2 result

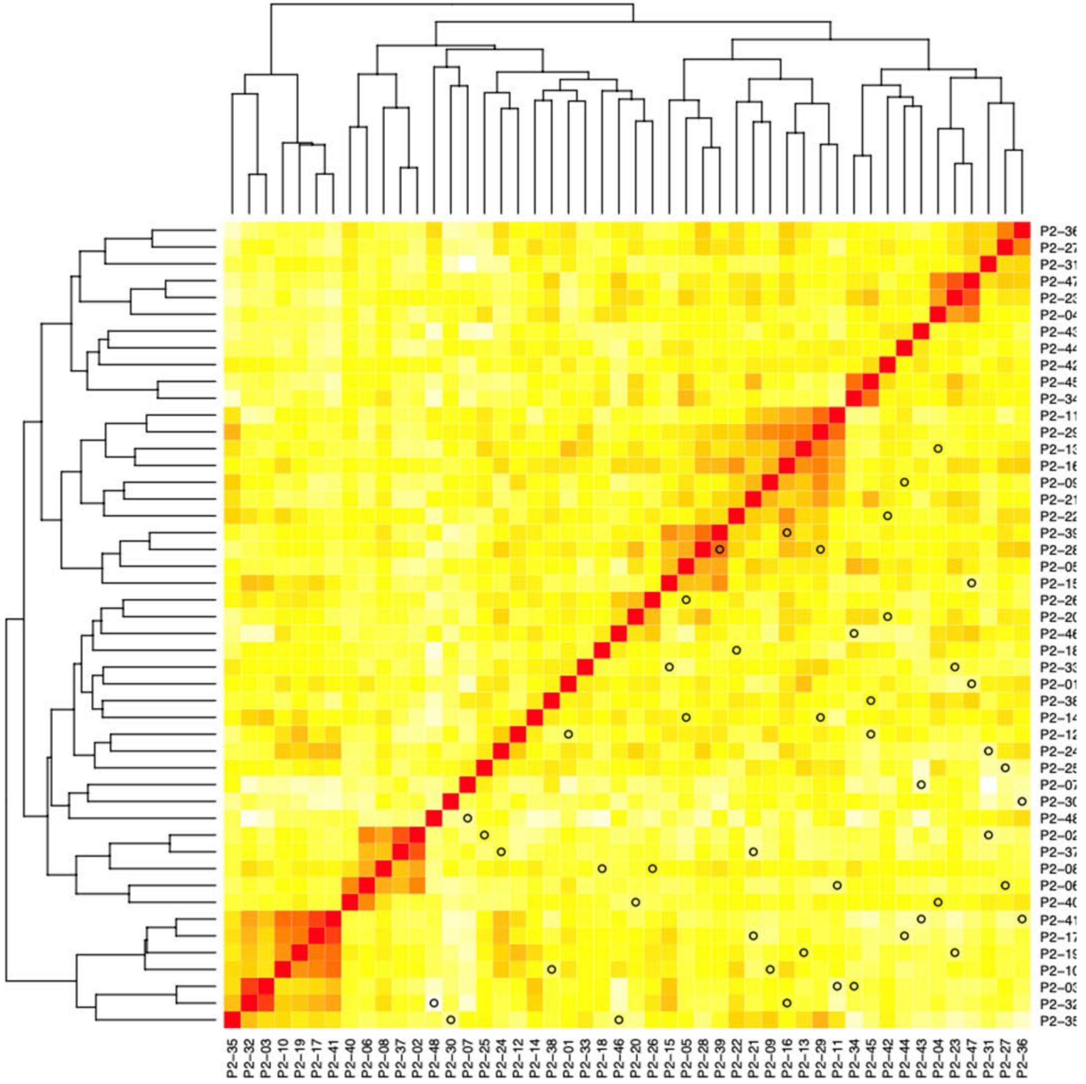


Real cycle 2 result



Daniel Krenzer, Eva Herzog, Matthias Frisch

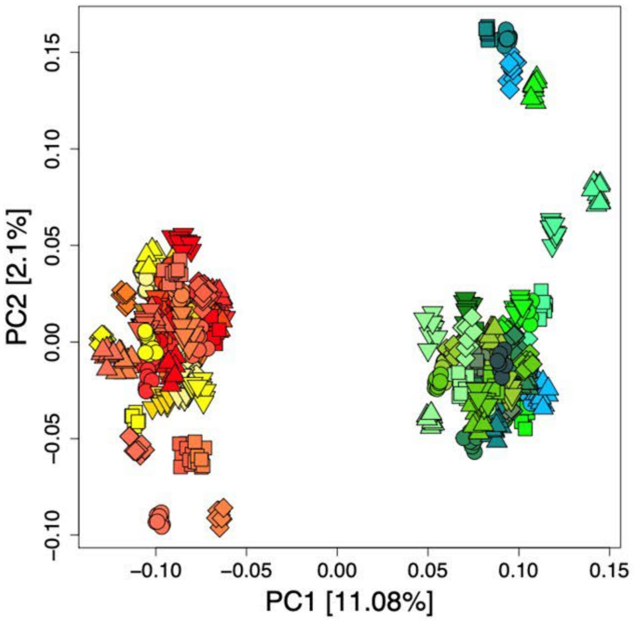
II: Promote between-pool separation



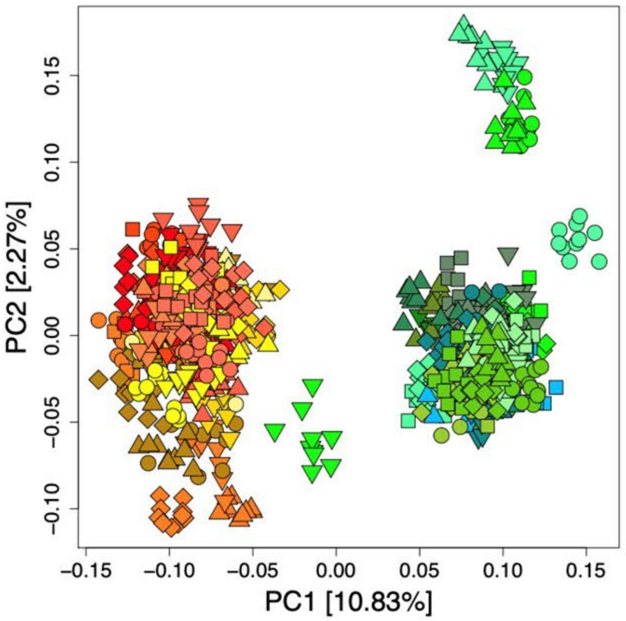
Cycle 3

Most suitable simulated complex crosses

Simulated cycle 3 result

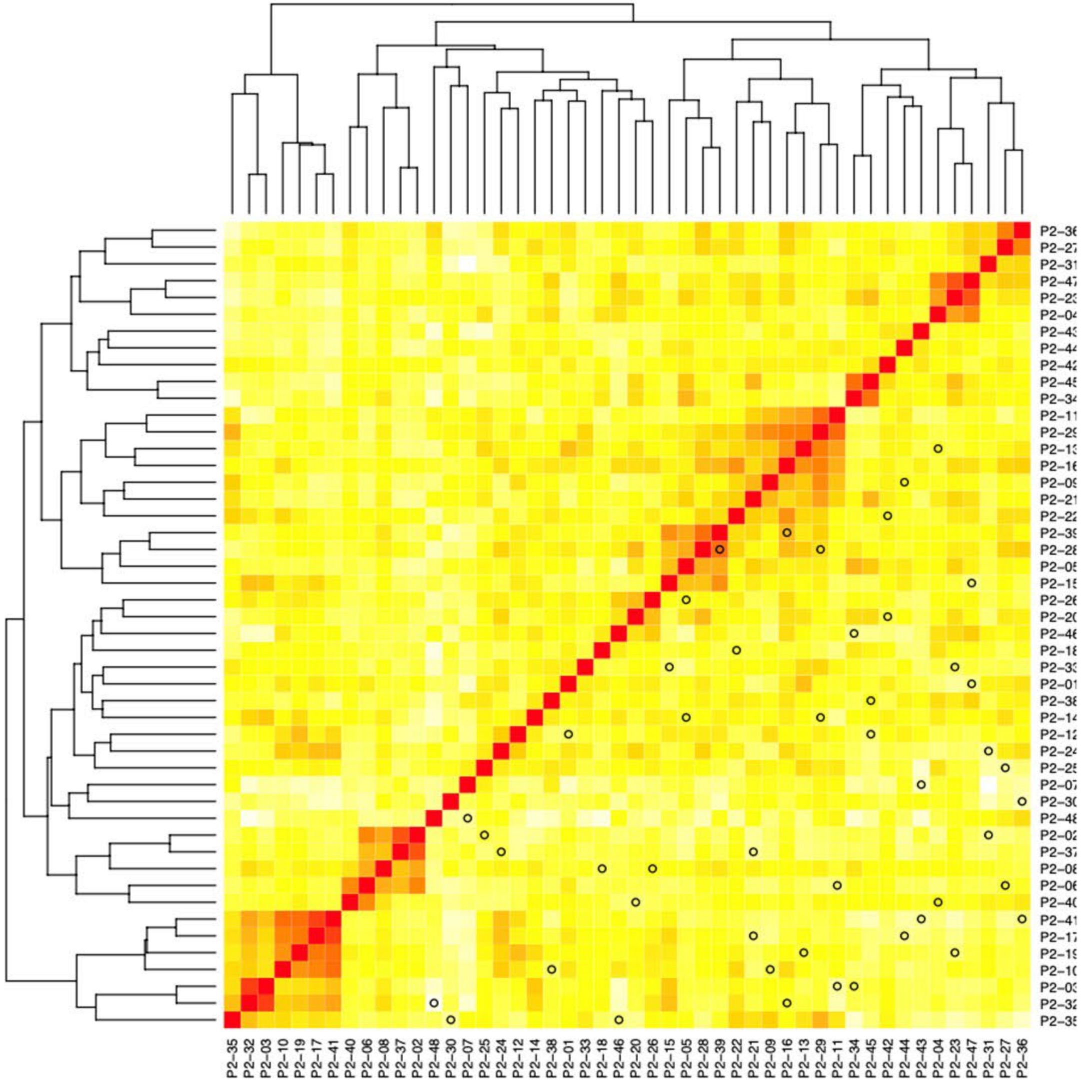


Real cycle 3 result



Daniel Krenzer, Eva Herzog, Matthias Frisch

II: Promote between-pool separation



After 3 cycles:

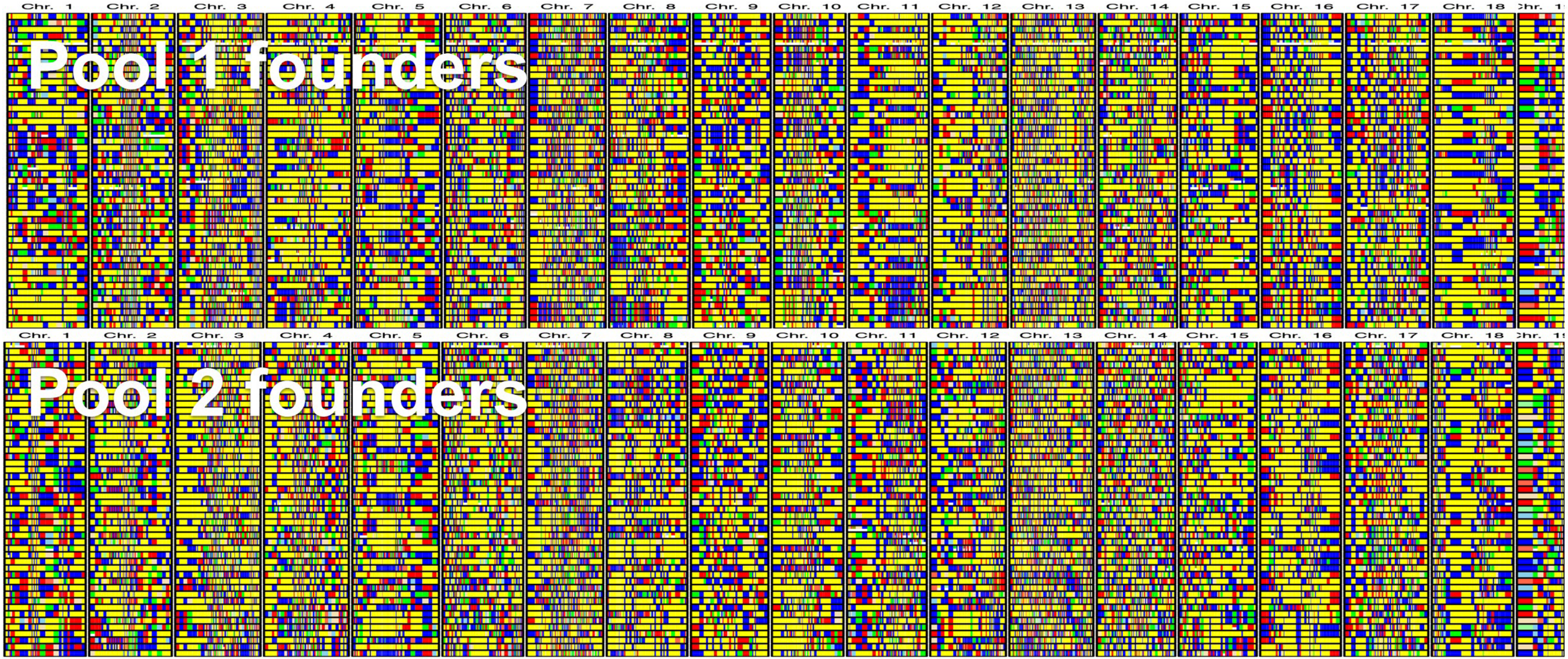
Two new, genetically distinct, recombined pools with 3*500 inbred or DH lines each, all genotyped with 15k SNP array

All **“must-have” genes** (resistance, seed quality, sterility & restorer properties) are **represented in both pools**

Daniel Krenzer, Eva Herzog, Matthias Frisch

III: Define founder-specific haplotype patterns

LD patterns in 100 founder lines based on ~15,000 genome-wide SNPs



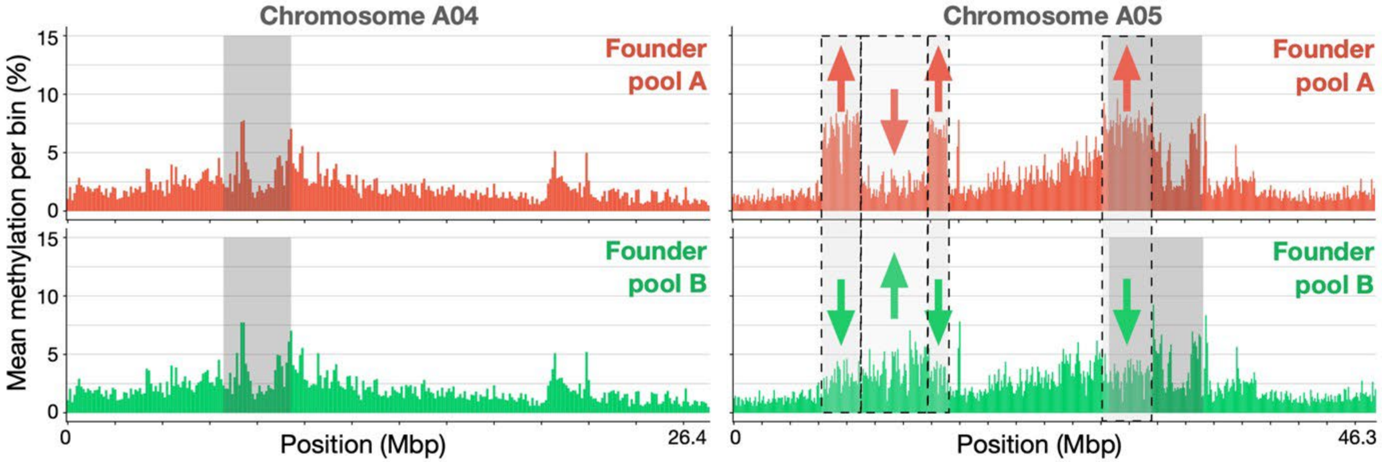
Daniel Krenzer, Eva Herzog, Matthias Frisch

IV: Identify pool-specific regulatory patterns

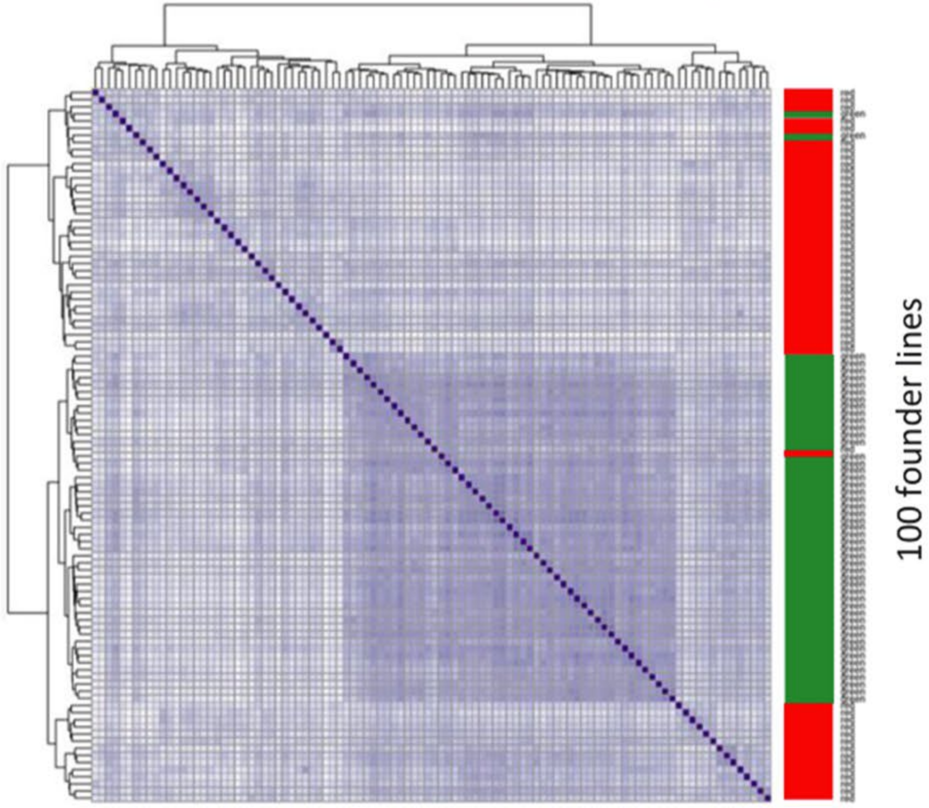
“Omics” data from 100 founder lines:

- ~30x Illumina short read WGS
- Whole genome methylation patterns (WGBS)
- Leaf RNAseq & miRNA
- Genome-wide structural variants (40x Nanopore)
- **All variants assigned to founder haplotypes**

Differentially methylated chromosome regions between pools



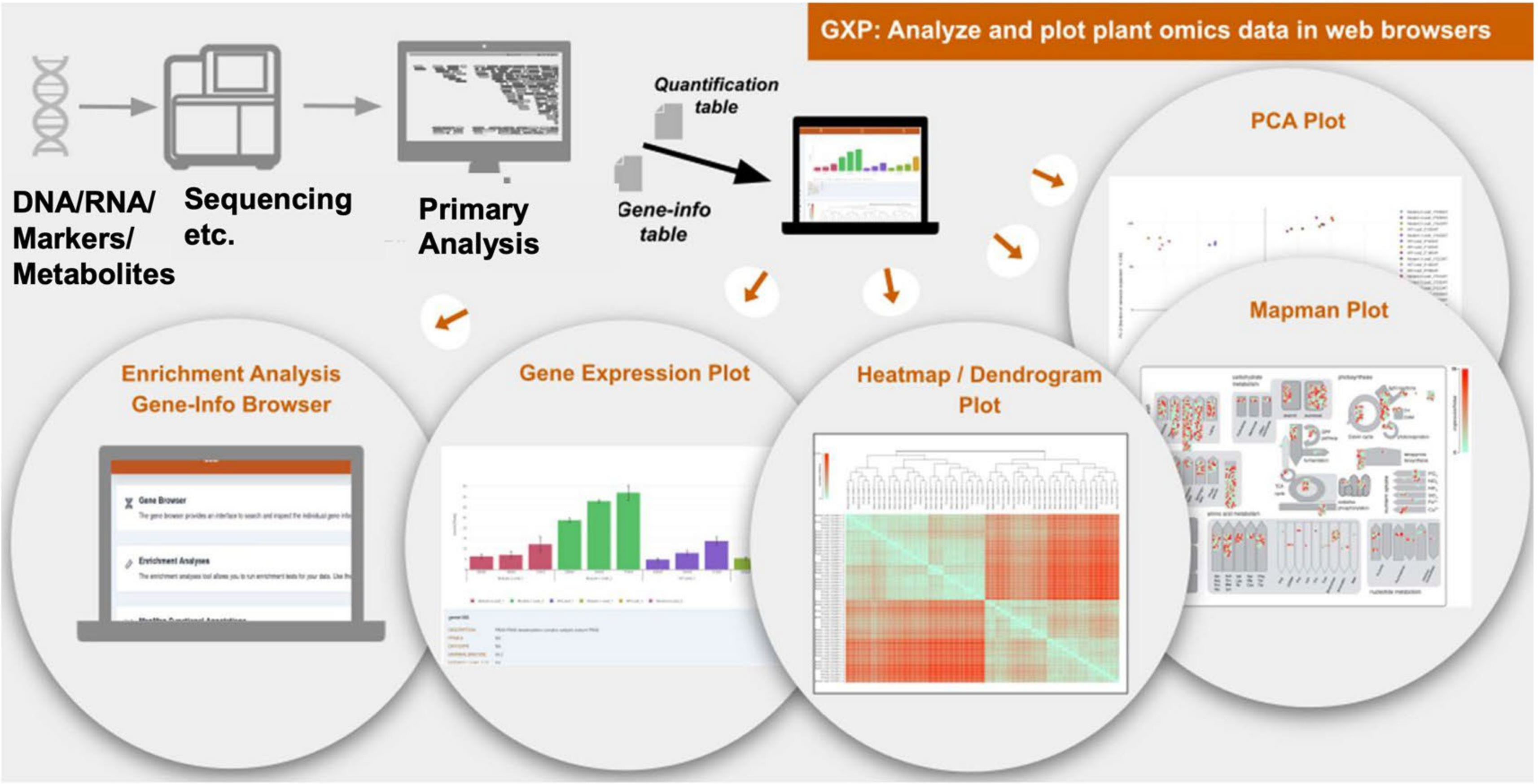
Differentially expressed genes between pools (n = 5797)



Jenny HueyTyng Lee

V: Make data accessible and useable

Dedicated *BreedPatH* database:



100 founder genomes

Representative assemblies

mRNA expression

miRNA expression

Cytosine methylation

Structural variants

GW-SNPs (3000 offspring)

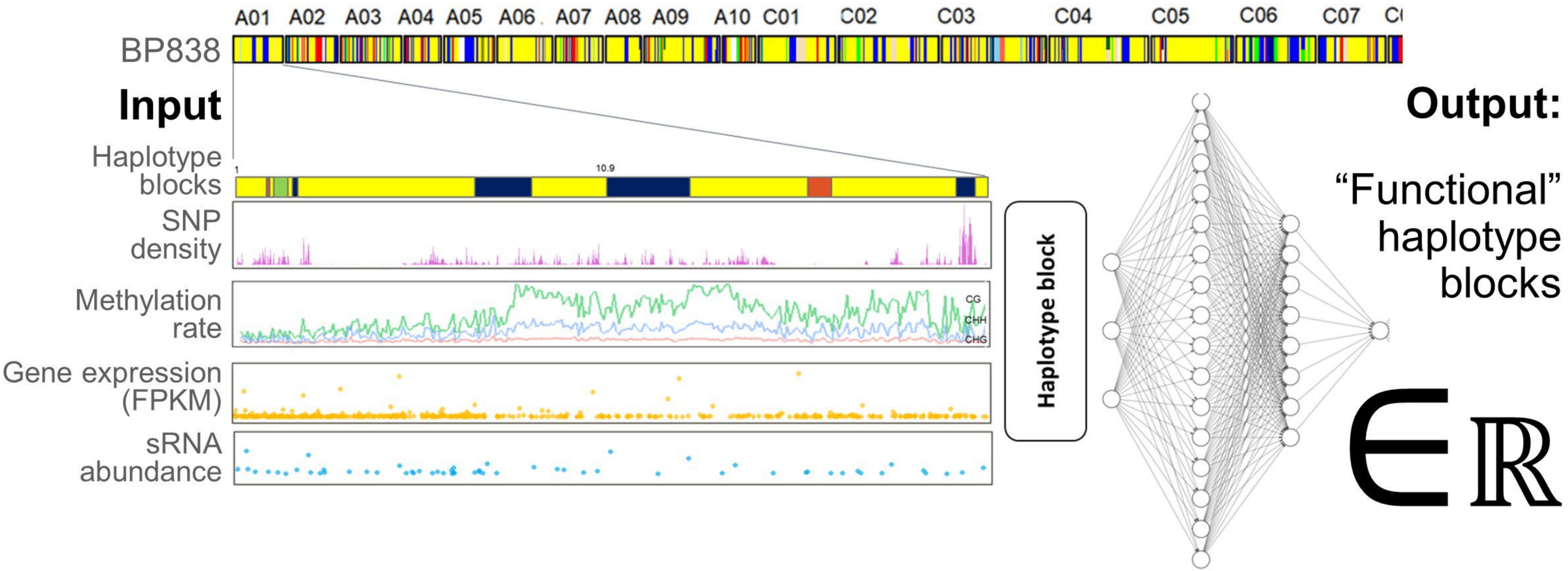
Data analysis and inspection

External access for analysis of public datasets

Asis Hellab, Björn Usadel, Forschungszentrum Jülich

VI: Index functional haplotype blocks

Information compression using deep-learning autoencoders



Input

Output:

"Functional" haplotype blocks

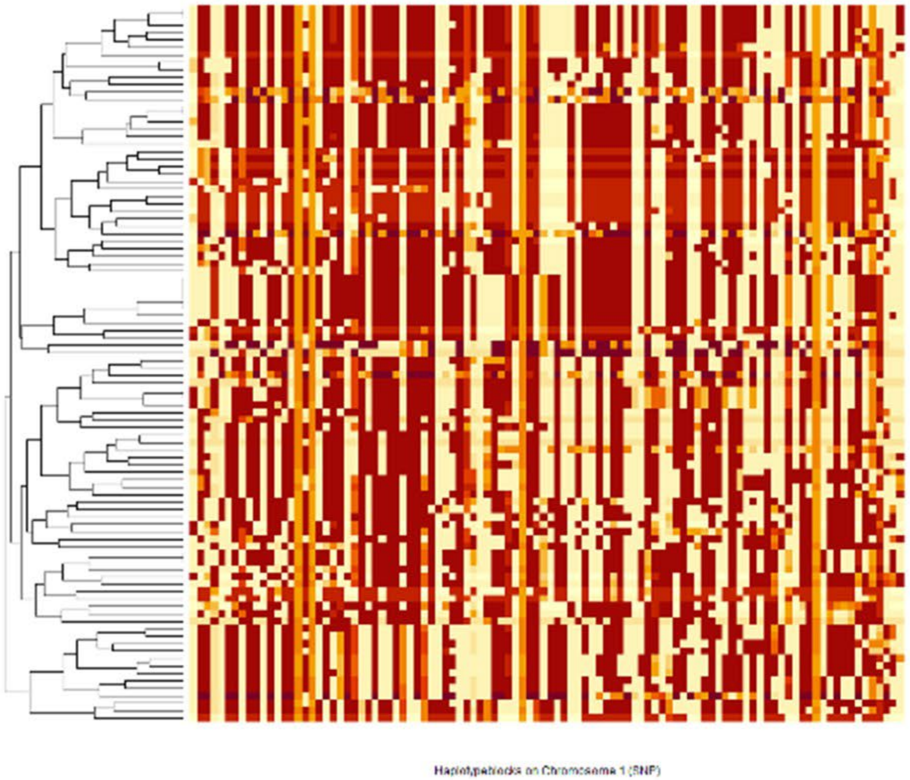
$\in \mathbb{R}$

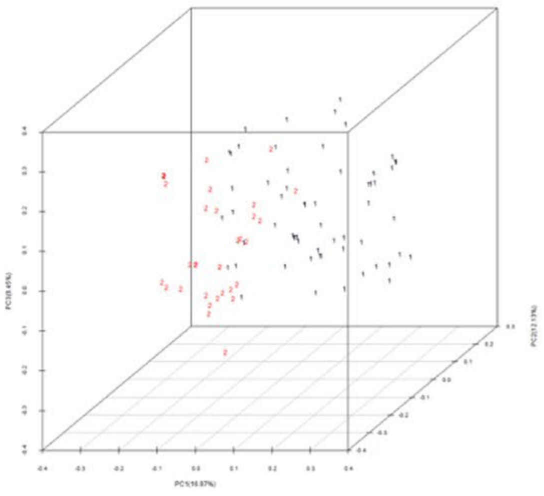
Reduction of genomic features and functional information to single predictors

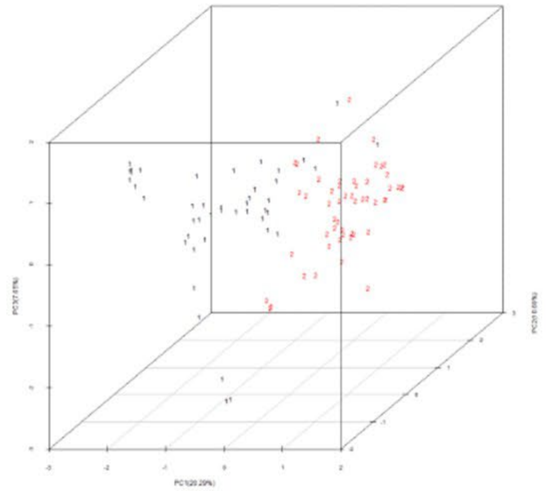
Sven Weber

VII: Re-define pools using functional data

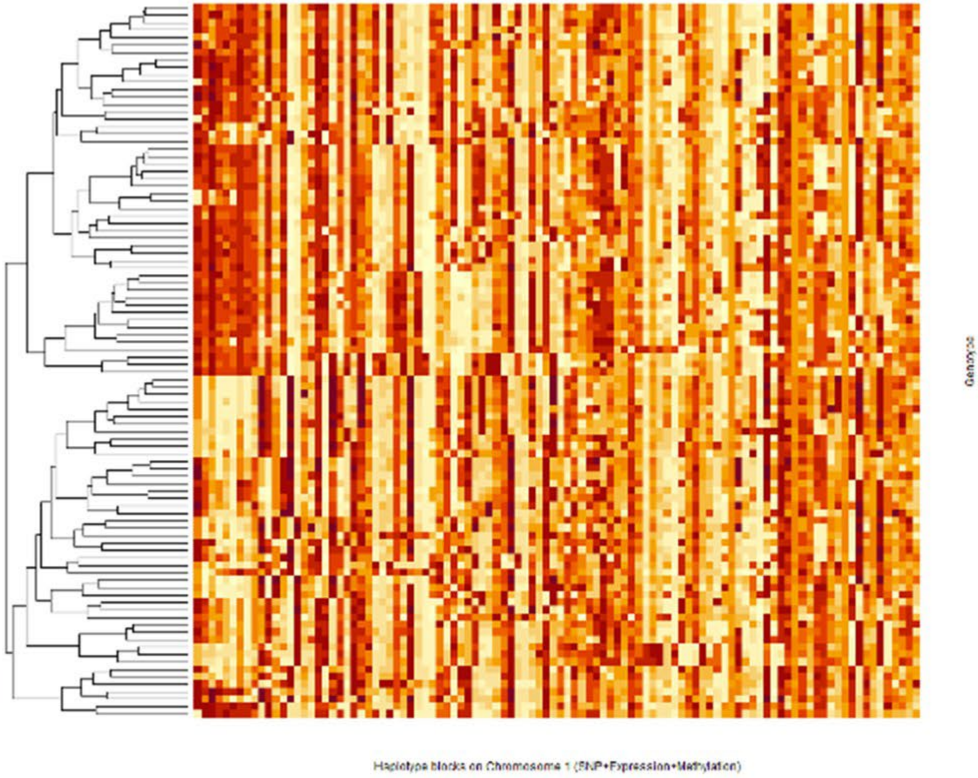
SNP-based pool definition


Haplotypeblocks on Chromosome 1 (SNP)





Pool definition via functional haplotypes


Haplotype blocks on Chromosome 1 (SNP+Expression+Methylation)

Sven Weber

VIII: Optimise GCA prediction accuracy

Hypothesis: *Non-linear prediction models* based on machine-learning methods should outperform linear predictions (e.g. BLUP) when large-scale regulatory datasets are available

Comparisons:

Effect of different **pool definitions** on predictions

Usefulness of **feature selection** for hybrid yield/GCA

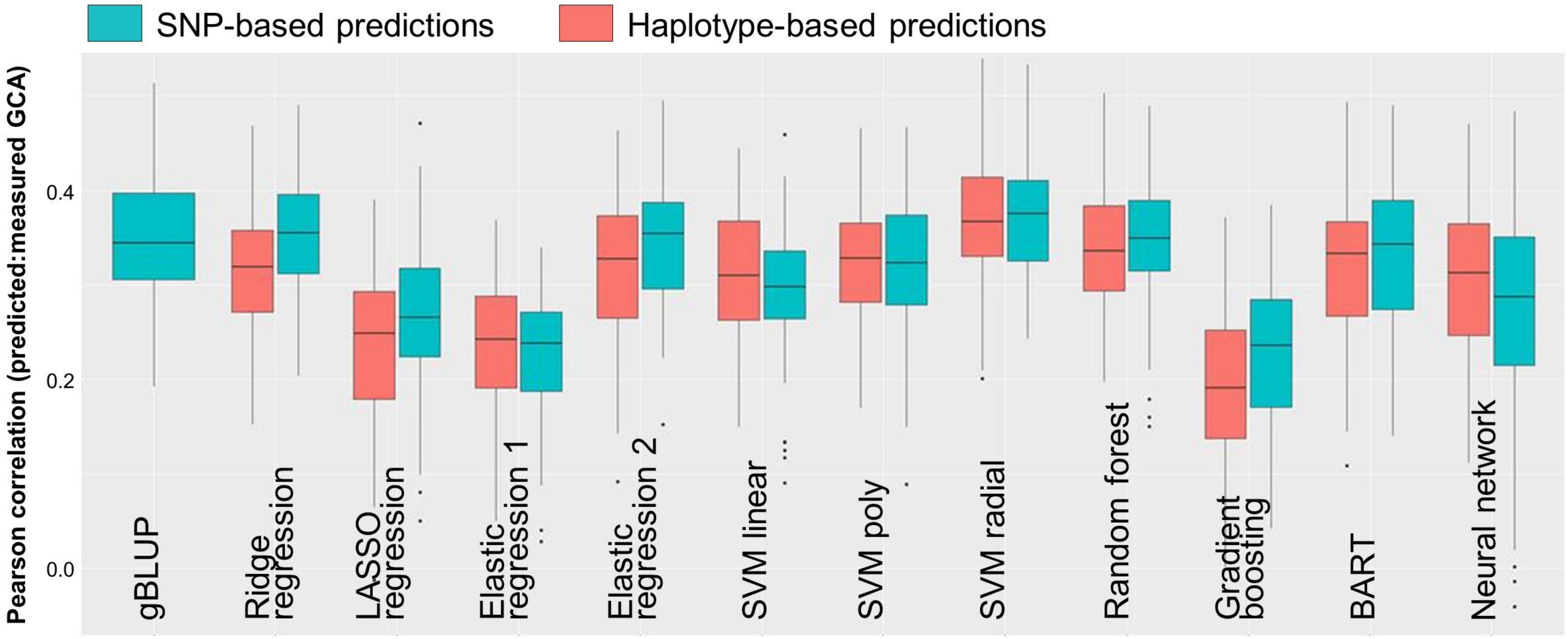
SNPs *vs.* (functional) **haplotypes** with/without **SV** info

Linear vs. **non-linear models**


BreedPatH test-hybrid data for fine-tuning of prediction models

Amine Abbadi, Christian Flachenecker, Sven Weber

IX: Predict hybrid yield performance



Sven Weber

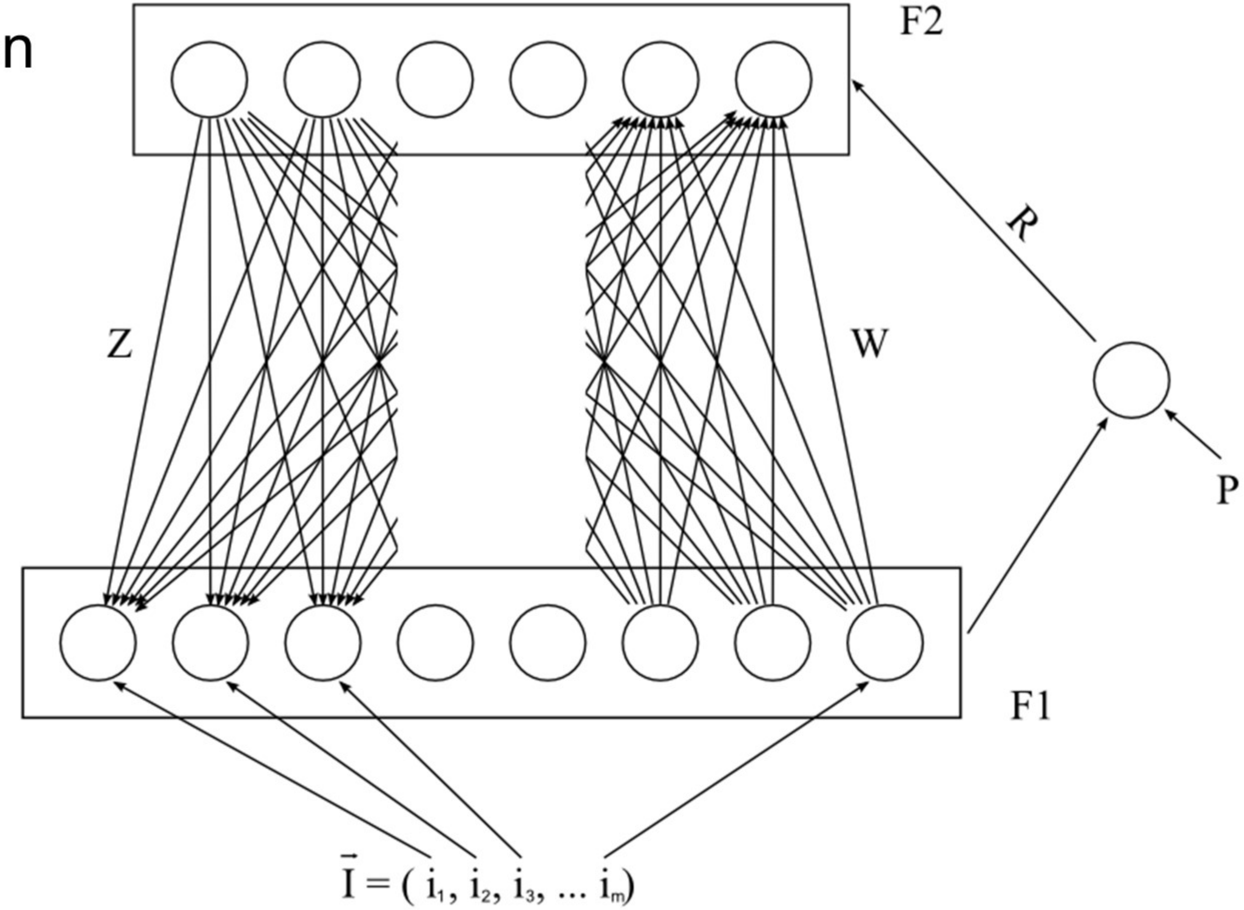
X: Improve ML-based predictions

“New” ML methods to potentially improve prediction accuracy

e.g. unsupervised neural networks based on **adaptive resonance theory (Fuzzy ART)**

“High adaptability for interactive feature prediction in applications analysing complex patterns with unpredictable environmental variables”

Work in progress



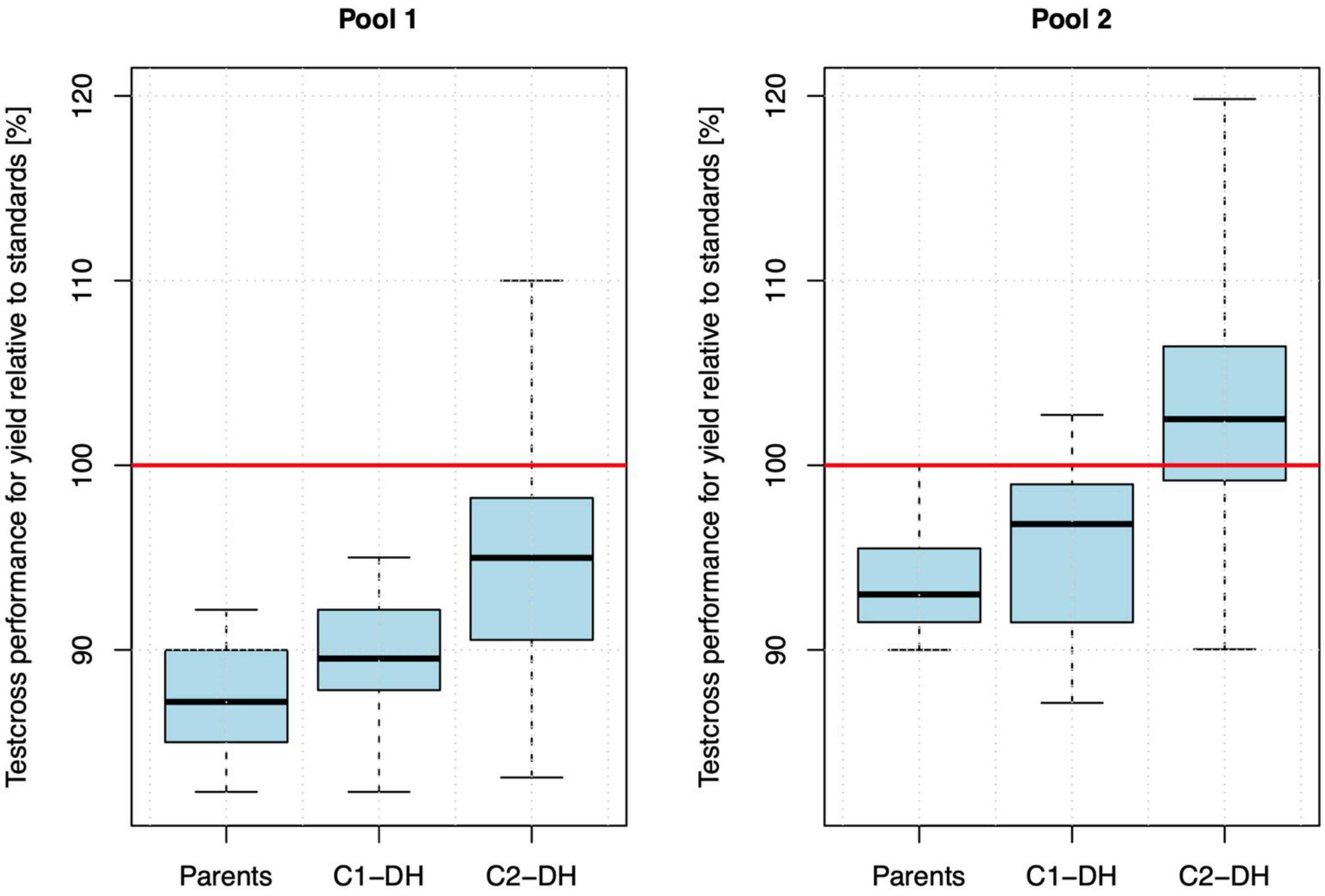
Sven Weber

The million-dollar question

So, is all that effort really worthwhile?

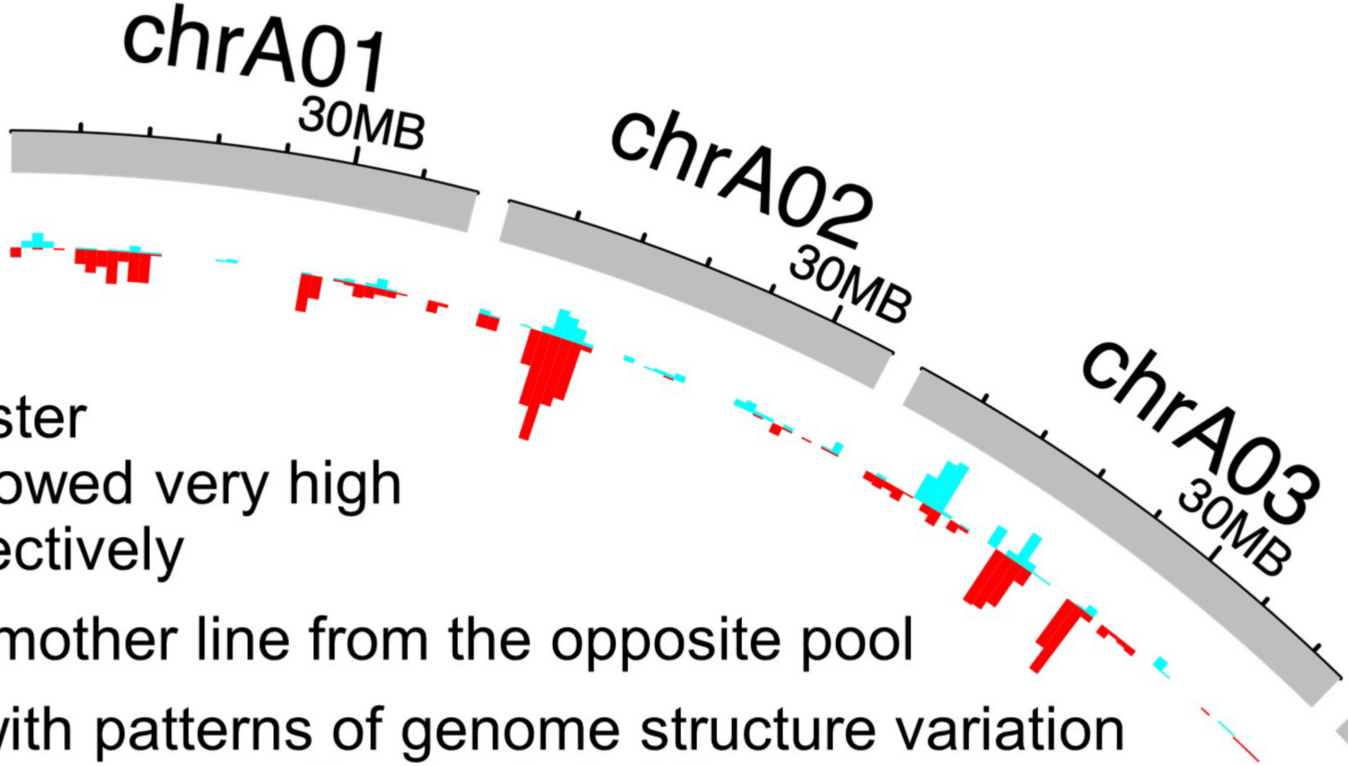
Answer: Yes!

- Hybrid performance improved substantially with each cycle
- Test-cross performance of Pool 2 restorers already outperformed commercial check hybrids after Cycle 2
- Best test hybrids from genomic-selected Cycle 2 DH lines yielded up to 120% of checks



A more scientific question

So why did it work?



Answer: We’re trying to find that out

- Whole-genome sequence data from sister DH lines (Cycle 2) in each pool that showed very high vs. poor test hybrid performance, respectively
- Genomes compared to the respective mother line from the opposite pool
- Identification of chromosome regions with patterns of genome structure variation that are unique to **all good hybrids** or unique to **all poor hybrids**, respectively
- Implementation of **“heterotic haplotypes”** provides a unique opportunity to inform future crossing designs in order to further improve heterosis and hybrid performance

JLU
NEUE WEGE. SEIT 1607.

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Sven Weber
Jenny Lee
Nayyer Abdollahi
Daniel Krenzer
Agnieszka Golicz
Eva Herzog
Matthias Frisch

Giessen Plant Breeding

BreedPatH

JÜLICH
Forschungszentrum

Asis Helab
Björn Usadel

PLANT 2030

NPZ INNOVATION
Wir forschen für Qualität.

Tobias Kox
Christian Flachenecker
Milka Malenica
Gunhild Leckband
Amine Abbadi

SPONSORED BY THE

Federal Ministry
of Education
and Research