UNIVERSITY OF HOHENHEIM

Power and precision of mapping genes in simulated F2 crosses using whole genome sequence data.

M. Schmid, R. Wellmann and J. Bennewitz

University of Hohenheim, Institute of Animal Science, Germany.

F2 data in the past

- Many porcine F2 crosses were established for gene mapping experiments

Divergent lineages as founder breeds for F2 designs

- Distantly related (**ASIA x EU**, e.g. Rückert & Bennewitz 2010)
- Closely related (**EU x EU**, e.g. Boysen et al. 2010)

Available datasets

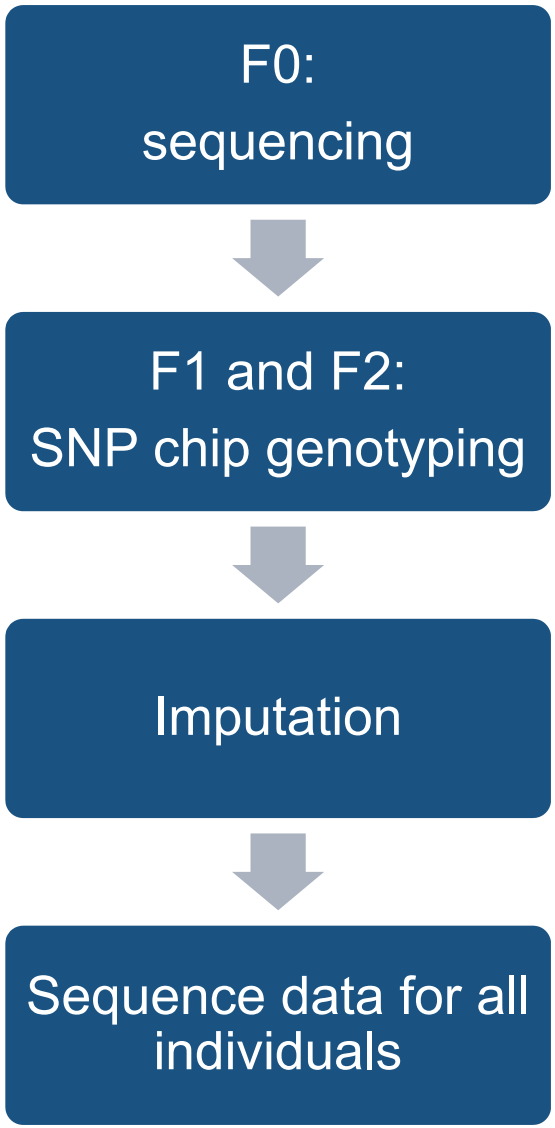
- Exact phenotypes for a lot of traits
- Mostly genotyped with microsatellite markers (low mapping resolution)

- Linkage mapping: no historical meioses can be considered

→ many QTL could be found, however, the confidence intervals were large

Can we use F2 data in the era of GENOMICS to precisely map genes?

F2 data today



Applying new next generation sequencing (NGS) techniques on F2 data might lead to powerful datasets for GWAS at reasonable costs.

Current simulation study: What's it about?

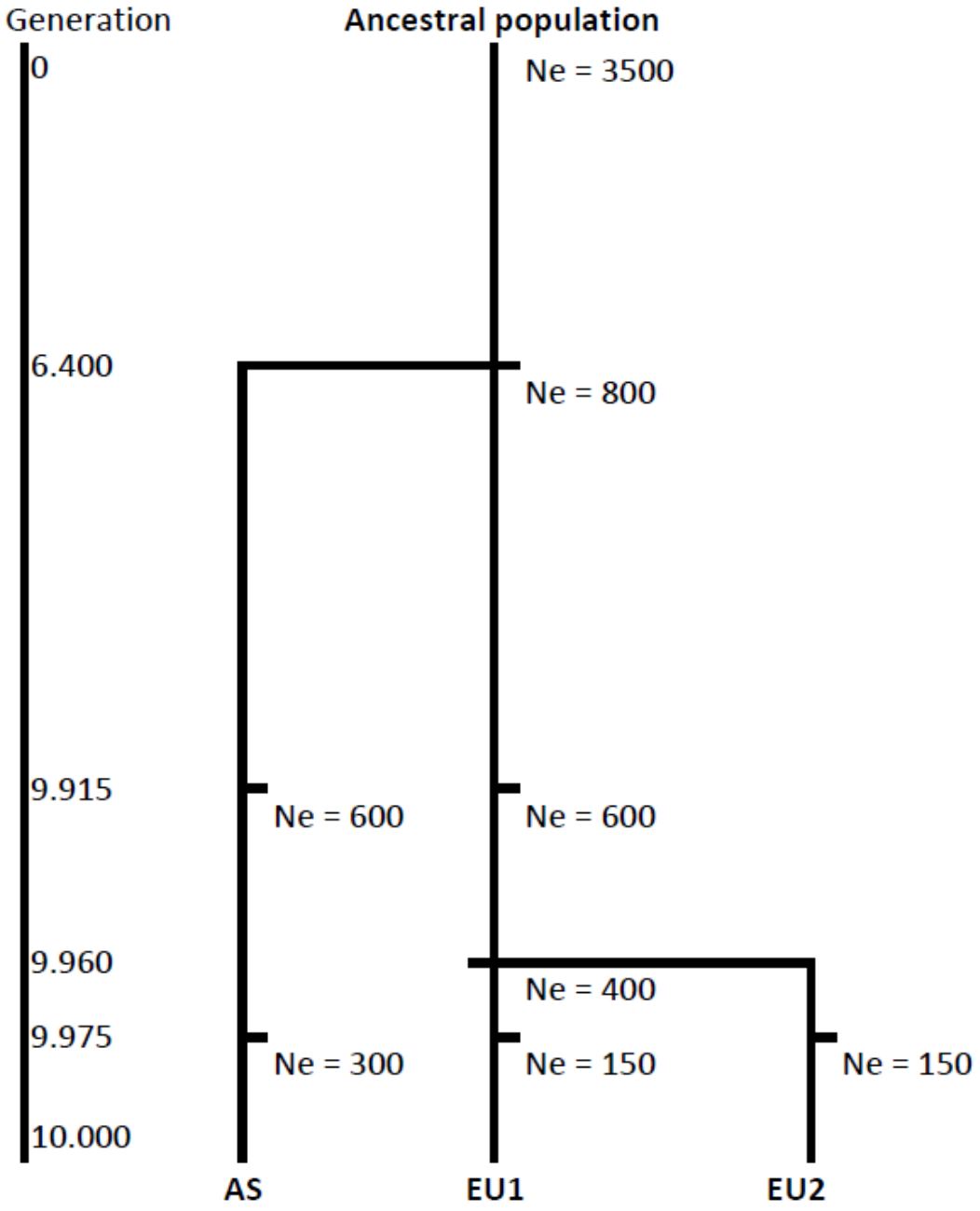
Investigation of power and precision in GWAS using simulated sequence data in

- F2 designs with closely related founder breeds (i.e. EU x EU),
- F2 designs with distantly related founder breeds (i.e. ASIA x EU),
- and the pooled data of both designs.

We compare:

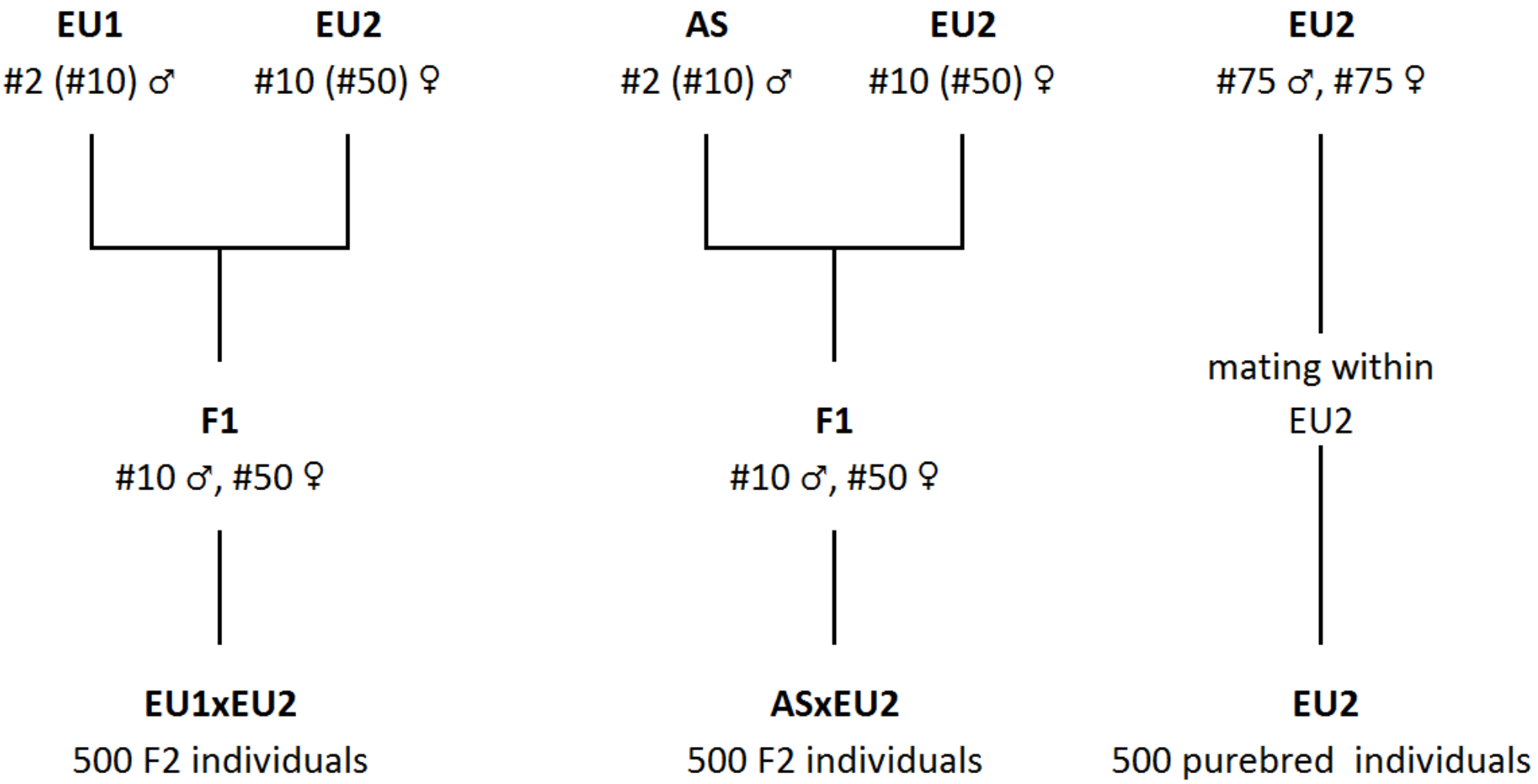
- F2 designs derived from closely and distantly related founder breeds
- Impact of pooling F2 data
- Purebred population vs. F2 crosses
- Small vs. large number of founder animals in F2 designs

Phylogeny of the founder breeds (drift model)

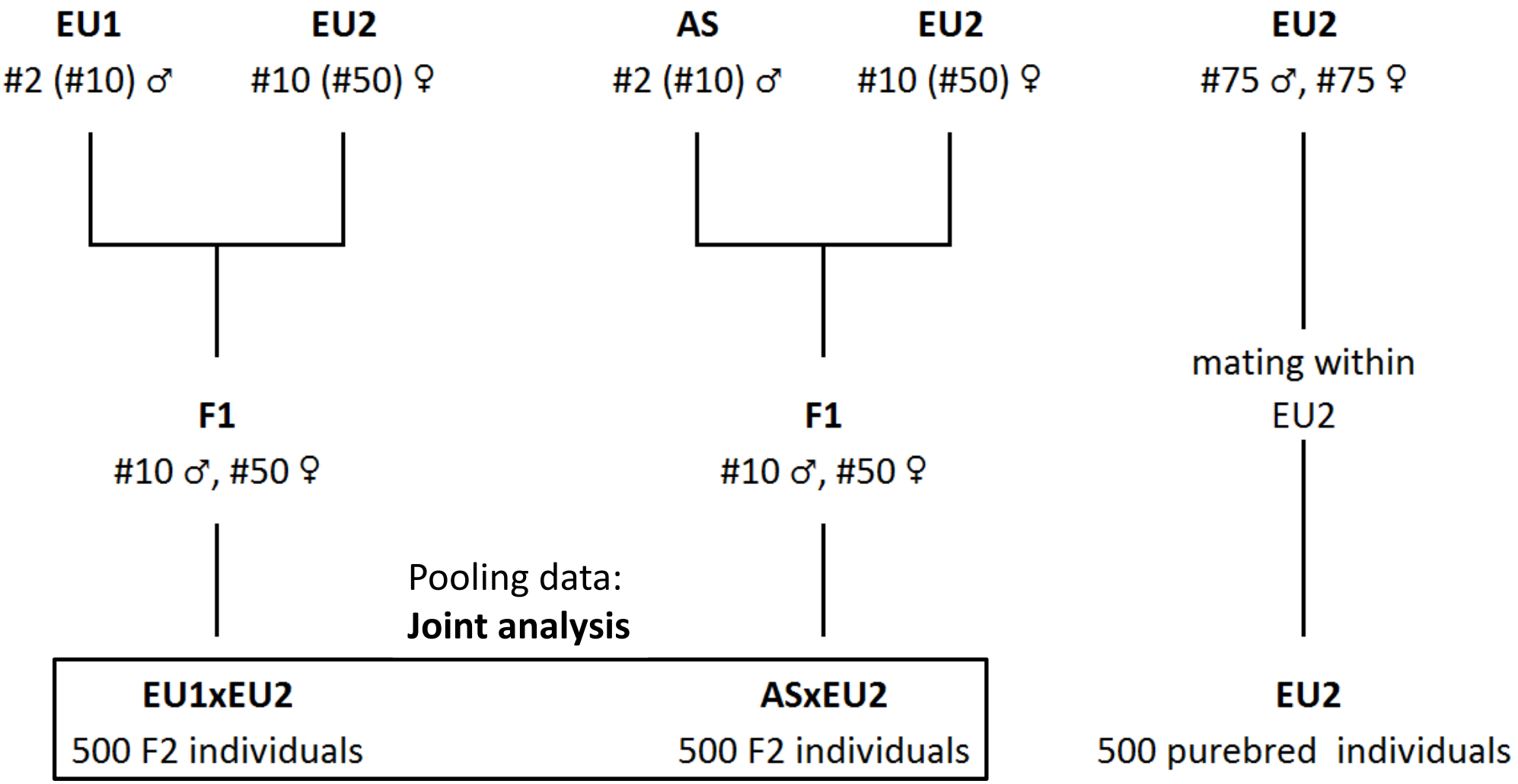


- Protocol is based on what is known about the phylogeny of pigs (Frantz et al. 2013).
- No selection → stepwise reduction of Ne to realize realistic population sizes
- Generation interval: 2.5 years
- Genomes:
  - 2 chromosomes à 1 M
  - 2 mutations/individual/generation

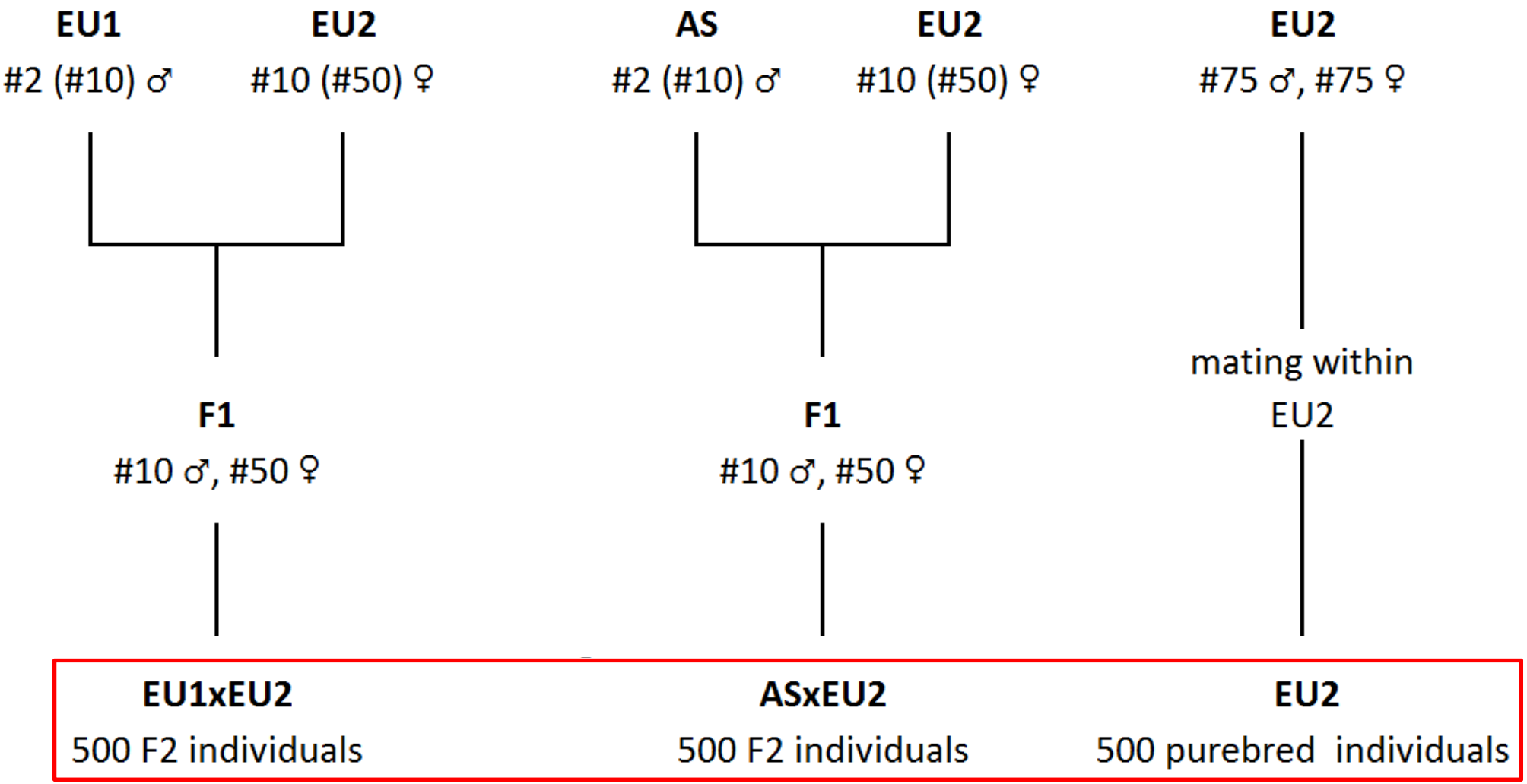
F2 crossing scheme



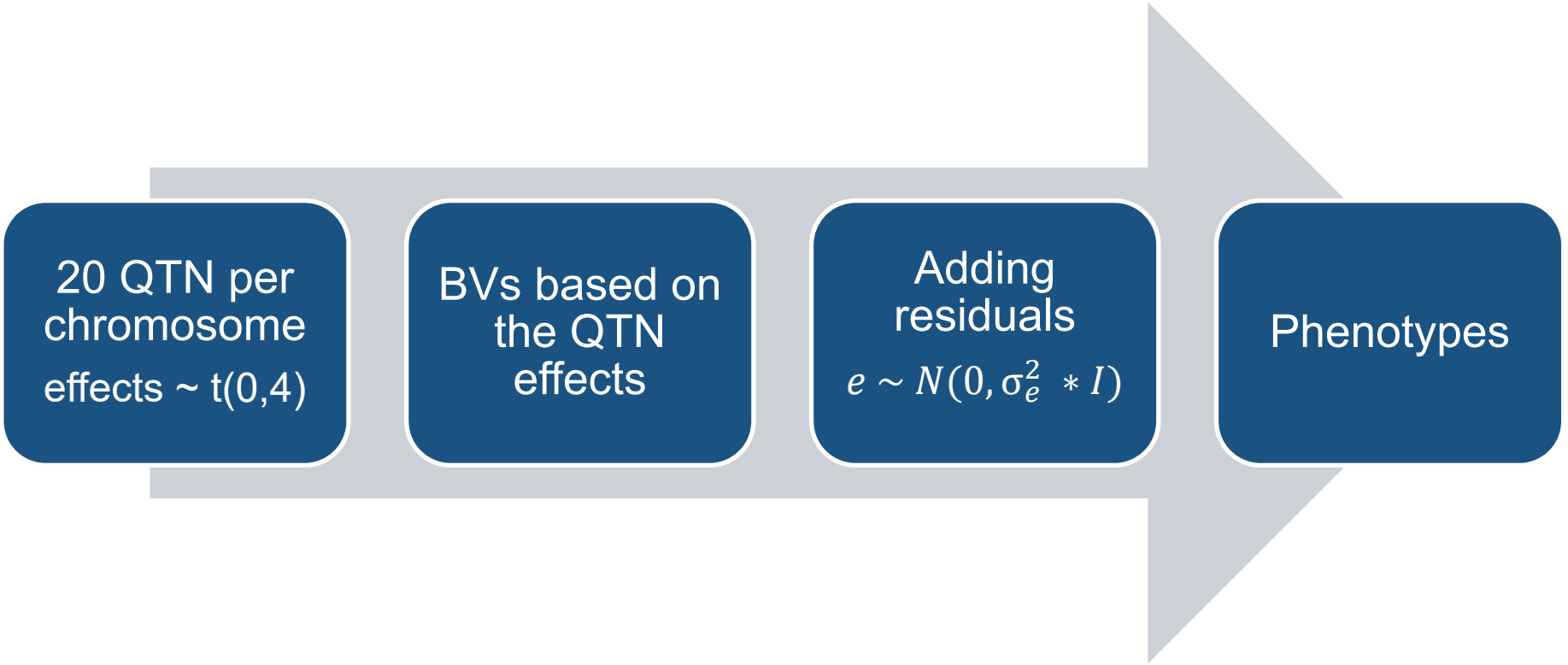
F2 crossing scheme



Individuals to be evaluated



Design of the traits



- Causative SNPs (QTN) were simulated in the pool of all individuals to be evaluated

  → Random but known positions

- Variance components are set so that $h^2 \approx 0.5$ is valid for the populations

Single marker regression using Genome-wide Complex Trait Analysis (Yang et al. 2014)

Mixed linear model:

$$y_i = \mu * b_j * SNP_{ij} + g_i + e_i$$

y_i	phenotype of individual i
μ	overall mean
b_j	regression coefficient of marker j
SNP_{ij}	gene content of SNP j of individual i
g_i	random polygenetic effect with $g \sim N(0, G\sigma_g^2)$ and G being the GRM
e_i	residual of individual i

- QTL mapping on chromosome 1 (2)
- Chromosome 2 (1) to model population structure (GRM)
  → SNPs to be tested are excluded from the GRM (MLMe)

- p-values were adjusted using Bonferroni correction, α = 0.01 (genome wide)

Results: Averaged values across all 50 replicates (10 simulations á 5 traits)

Parameter	QTN Power		QTL Power		QTN Precision	
Significance level	α = 0.01		α = 0.01		α = 0.01	
	mean	sd	mean	sd	mean	sd
EU2	0,21	0,08	0,29	0,11	0,90	0,10
EU1 x EU2 (small F0)	0,26	0,12	0,46	0,17	0,87	0,11
EU1 x EU2 (large F0)	0,22	0,09	0,32	0,10	0,89	0,11
ASIA x EU2 (small F0)	0,39	0,20	0,69	0,22	0,69	0,13
ASIA x EU2 (large F0)	0,36	0,18	0,65	0,23	0,70	0,14
Joint analysis (small F0)	0,44	0,16	0,82	0,16	0,71	0,11
Joint analysis (large F0)	0,41	0,14	0,76	0,18	0,75	0,13

Power increase in F2 data compared to purebred populations: Highest power in pooled data

Parameter	QTN Power		QTL Power		QTN Precision	
Significance level	α = 0.01		α = 0.01		α = 0.01	
	mean	sd	mean	sd	mean	sd
EU2	0,21	0,08	0,29	0,11	0,90	0,10
EU1 x EU2 (small F0)	0,26	0,12	0,46	0,17	0,87	0,11
EU1 x EU2 (large F0)	0,22	0,09	0,32	0,10	0,89	0,11
ASIA x EU2 (small F0)	0,39	0,20	0,69	0,22	0,69	0,13
ASIA x EU2 (large F0)	0,36	0,18	0,65	0,23	0,70	0,14
Joint analysis (small F0)	0,44	0,16	0,82	0,16	0,71	0,11
Joint analysis (large F0)	0,41	0,14	0,76	0,18	0,75	0,13

Precision: F2 designs with closely related founders almost reach the precision of purebred populations

Parameter	QTN Power		QTL Power		QTN Precision	
Significance level	α = 0.01		α = 0.01		α = 0.01	
	mean	sd	mean	sd	mean	sd
EU2	0,21	0,08	0,29	0,11	0,90	0,10
EU1 x EU2 (small F0)	0,26	0,12	0,46	0,17	0,87	0,11
EU1 x EU2 (large F0)	0,22	0,09	0,32	0,10	0,89	0,11
ASIA x EU2 (small F0)	0,39	0,20	0,69	0,22	0,69	0,13
ASIA x EU2 (large F0)	0,36	0,18	0,65	0,23	0,70	0,14
Joint analysis (small F0)	0,44	0,16	0,82	0,16	0,71	0,11
Joint analysis (large F0)	0,41	0,14	0,76	0,18	0,75	0,13

Pooling data increases the precision in F2 designs derived from distantly related founder breeds

Parameter	QTN Power		QTL Power		QTN Precision	
Significance level	α = 0.01		α = 0.01		α = 0.01	
	mean	sd	mean	sd	mean	sd
EU2	0,21	0,08	0,29	0,11	0,90	0,10
EU1 x EU2 (small F0)	0,26	0,12	0,46	0,17	0,87	0,11
EU1 x EU2 (large F0)	0,22	0,09	0,32	0,10	0,89	0,11
ASIA x EU2 (small F0)	0,39	0,20	0,69	0,22	0,69	0,13
ASIA x EU2 (large F0)	0,36	0,18	0,65	0,23	0,70	0,14
Joint analysis (small F0)	0,44	0,16	0,82	0,16	0,71	0,11
Joint analysis (large F0)	0,41	0,14	0,76	0,18	0,75	0,13

Where is the benefit to use (pooled) F2 sequence data in GWAS?

EUxEU

- Precise mapping results → short LD blocks

ASIAxEU

- High mapping power → increased gene frequencies
- Low precision → long LD blocks

Pooling data

- Increase in power → enlarged sample size
- Increase in precision compared to single analysis of ASIAxEU → reduced LD block length

This is in agreement with Toosi et al. (2009) and Bennewitz and Wellmann (2014).

Can we use F2 data in the era of GENOMICS to precisely map genes?

Something to reflect when taking a shower...

- Applying NGS techniques on F2 data leads to suitable datasets for GWAS at reasonable costs.

- F2 data at a maximum marker density provides a powerful and (cost) efficient possibility to (fine) map genes when the founder breeds are closely related or can be pooled.

Thank you!

This study was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).